Ultrasonic measurements and machine learning methods to monitor industrial processes

Alexander Lewis Bowler

Thesis submitted to the University of Nottingham for the degree of Doctor of Philosophy

September 2023

Abstract

The process manufacturing sector is increasingly using the collection and analysis of data to improve productivity, sustainability, and product quality. The endpoint of this transformation is processes that automatically adapt to demands in real-time. In-line and on-line sensors underpin this transition by automatically collecting the real-time data required to inform decision-making. Each sensing technique possesses its own advantages and disadvantages making them suitable for specific applications. Therefore, a wide range of sensing solutions must be developed to monitor the diverse and often highly variable operations in process manufacturing. Ultrasonic (US) sensors measure the interaction of mechanical waves with materials. They have benefits of being in-line, real-time, non-destructive, low in cost, small in size, able to monitor opaque materials, and can be applied non-invasively.

Machine Learning (ML) is the use of computer algorithms to learn patterns in data to perform a task such as making predictions or decisions. The correlations in the data that the ML models learn during training have not been explicitly programmed by human operators. Therefore, ML is used to automatically learn from and analyse data. There are four main types of ML: supervised, unsupervised, semi-supervised, and reinforcement learning. Supervised and unsupervised ML are both used in this thesis. Supervised ML maps inputs to outputs during training with the aim being to create a model that accurately predicts the outputs of data that was not previously used during training. In contrast, unsupervised learning only uses input data in which patterns are discovered. Supervised ML is being increasingly combined with sensor measurements as it offers several distinct advantages over conventional calibration methods, these include: reduced time for development, potential for more accurate fitting, methods to encourage generalisation across parameter ranges, direct correlations to important process information rather than material properties, and ability for continuous retraining as more data becomes available.

The aim of this thesis was to develop ML methods to facilitate the optimal deployment of US sensors for process monitoring applications in industrial environments. To achieve this, the thesis evaluates US sensing techniques and ML methods across three types of process manufacturing operations: material mixing, cleaning of pipe fouling, and alcoholic fermentation of beer. Two US sensing techniques were investigated: a non-invasive, reflection-mode technique, and a transmission-based method using an invasive US probe with reflector plate. The non-invasive, reflection-mode technique is more amenable to industrial implementation than the invasive probe given it can be externally retrofitted to existing vessels. Different feature extraction and feature selection methods, algorithms, and hyperparameter ranges were explored to determine the optimal ML pipeline for process monitoring using US sensors. This facilitates reduced development time of US sensor and ML combinations when deployed in industrial settings by recommending a pipeline that has been trialled over a range of process monitoring applications. Furthermore, methods to leverage previously collected datasets were developed to negate or reduce the burden of collecting labelled data (the outputs required during ML model training and often acquired by using reference measurements) for every new process monitoring application. These included unlabelled and labelled domain adaptation approaches.

Both US sensing techniques investigated were found to be similarly accurate for process monitoring. To monitor the development of homogeneity during the blending of honey and water the non-invasive, reflection-mode technique achieved up to 100 % accuracy to classify whether the materials were mixed or non-mixed and an R^2 of 0.977 to predict the time remaining (or time since) complete mixing was achieved. To monitor the structural changes during the mixing of flour and water, the aforementioned sensing method achieved an

accuracy of 92.5 % and an R^2 of 0.968 for the same classification and regression tasks. Similarly, the sensing method achieved an accuracy of up to 98.2 % when classifying whether fouling had been removed from pipe sections and R^2 values of up 0.947 were achieved when predicting the time remaining until mixing was complete. The non-invasive, reflection-mode method also achieved R^2 values of 0.948, Mean Squared Error (MSE) values of 0.283, and Mean Absolute Error (MAE) values of 0.146 to predict alcohol by volume percentage of alcohol during beer fermentation. In comparison, the transmission-based sensing method achieved R^2 values of 0.952, MSE values of 0.265, and MAE values of 0.136 for the same task. Furthermore, the transmission-based method achieved accuracies of up to 99.8 % and 99.9 % to classify whether ethanol production had started and whether ethanol production had finished during an industrial beer fermentation process.

The material properties that affect US wave propagation are strongly temperature dependent. However, ML models that omitted the process temperature were comparable in accuracy to those which included it as an input. For example, when monitoring laboratory scale fermentation processes, the highest performing models using the process temperature as a feature achieved R^2 values of 0.952, MSE values of 0.265, and MAE values of 0.136 to predict the current alcohol concentration, compared with R^2 values of 0.948, MSE values of 0.283, and MAE values of 0.146 when omitting the temperature. Similarly, when transferring models between mixing processes, accuracies of 92.2 % and R^2 values of 0.947 were achieved when utilising the process temperature compared with 92.1% and 0.942 when omitting the temperature. When transferring models between cleaning processes, inclusion of the process temperature as a feature degraded model accuracy during classification tasks as omitting the temperature produced the highest accuracies for 6 out of 8 tasks. Mixed results were obtained for regression tasks where including the process temperature increased model accuracy for 3 out of 8 tasks. Overall, these results indicate that US sensing, for some applications, is able to achieve comparable accuracy when the process temperature is not available. The choice of whether to include the temperature as a feature should be made during the model validation stage to determine whether it improves prediction accuracy.

The optimal feature extraction, feature selection, and ML algorithm permutation was determined as follows: Features were extracted by Convolutional Neural Networks (CNNs) followed by Principal Component Analysis (PCA) and inputted into deep neural networks with Long Short-Term Memory (LSTM) layers. The CNN was pre-trained on an auxiliary task using previously collected US datasets to learn features of the waveforms. The auxiliary task was to classify the dataset from which each US waveform originated. PCA was applied to reduce the dimensionality of the input data and enable the use of additional features, such as the US time of flight or measures of variation between consecutively acquired waveforms. This CNN and PCA feature extraction method was shown to produce more informative features from the US waveform compared to a traditional, coarse feature extraction approach, achieving higher accuracy on 65 % of tasks evaluated. The coarse feature method used commonly extracted parameters from US waveforms such as the energy, standard deviation, and skewness. LSTM units were used to learn the trajectory of the process features and so enable the use of information from previous timesteps to inform model prediction. Using LSTM units was shown to outperform neural networks with feature gradients used as inputs to incorporate information from previous timesteps for all process monitoring applications. Multi-task learning also showed improvements in learning feature trajectories and model accuracy (improving regression accuracy for 8 out of 18 tasks), however, at the expense of a greater number of hyperparameters to optimise. The choice to use multi-task learning should be evaluated during the validation stage of model development.

Unlabelled and labelled domain adaptation were investigated to transfer ML knowledge between similar processes. Unlabelled domain adaptation was used to transfer trained ML models between similar mixing and similar cleaning processes to negate the need to collect labelled data for a new task. Transfer Component Analysis was compared to a Single Feature transfer method. Transferring a single feature was found to be optimal, achieving classification accuracies of up to 96.0% and 98.4% to predict whether the mixing or cleaning processes were complete and R^2 of up to 0.947 and 0.999 to predict the time remaining for each process, respectively. The Single Feature method was most accurate as it was most representative of the changing material properties at the sensor measurement area. Training ML models across a greater process parameter range (a greater range of temperatures; 19.3 to 22.1°C compared with 19.8 to 21.2°C) or multiple datasets improved transfer learning to further datasets by enabling the models to adapt to a wider range of feature distributions. Labelled domain adaptation increased model accuracy on an industrial fermentation dataset by transferring ML knowledge from a laboratory fermentation dataset. Federated learning was investigated to maintain dataset privacy when applying transfer learning between datasets. The federated learning methodology performed better than the other methods tested, achieving higher accuracy for 14 out of 16 machine learning tasks compared with the base case model which was trained using data solely from the industrial fermentation. This was attributed to federated learning improving the gradient descent operation during network optimisation. During the federated learning training strategy, the local models were trained for a full epoch on each dataset before network weights were sent to the global model. In contrast, during the non-federated learning strategy, batches from each dataset were interspersed. Therefore, it is recommended that the order that the data is passed to the model during training should be evaluated during the validation stage.

Overall, there are two main contributions from this thesis: Development of the ML pipeline for process monitoring using US sensors, and the development of unlabelled and labelled domain adaptation methods for process monitoring using US sensors. The development of an ML pipeline facilitates reduced time for the deployment of US sensor and ML combinations in industrial settings by recommending a method that has been trialled over a range of process monitoring applications. The unlabelled and labelled domain adaptation methods were developed to leverage previously collected datasets. This negates or reduces the burden of collecting labelled data in industrial environments. Furthermore, the pipeline and domain adaptation methodologies are evaluated using a non-invasive, reflection-mode US sensing technique. This technique is industrially relevant as it can be externally retrofitted onto existing process equipment.

The novelty contained within this thesis can be summarised as follows:

- **The use of CNNs and LSTM layers for process monitoring using US sensor data**: CNNs were used to extract spatial-invariant features from US sensor data to overcome problems of features shifting in the time domain due to changes in temperature or sound velocity. LSTM units were used for their ability to analyse sequences and understand temporal dependencies, critical for monitoring processes that develop over time. Feature extraction using CNNs was shown to produce more informative features from the US waveform compared to a traditional, coarse feature extraction approach, achieving higher accuracy on 65 % of tasks evaluated. LSTM units were shown to outperform neural networks with feature gradients used as inputs to incorporate information from previous timesteps for all process monitoring applications.
- **Evaluating the omission of the process temperature as a feature for process monitoring using US sensor data**: This indicates whether the US sensor and ML

combinations could be used in industrial applications where measurement of the process temperature is not available. Overall, it was found that ML models which omitted the process temperature were comparable in accuracy to those which included it as an input (for example, R^2 values of 0.952, MSE values of 0.265, and MAE values of 0.136 when including temperature compared with R^2 values of 0.948, MSE values of 0.283, and MAE values of 0.146 were obtained when omitting the temperature to predict the current alcohol concentration during laboratory scale fermentation processes).

- **The use of labelled and unlabelled domain adaptation for US data for process monitoring**: Unlabelled domain adaptation was used to transfer trained ML models between similar mixing and similar cleaning processes to negate the need to collect labelled data for a new task. Labelled domain adaptation increased model accuracy on an industrial fermentation dataset by transferring ML knowledge from a laboratory fermentation dataset.
- **The use of labelled and unlabelled domain adaptation on features extracted from US waveforms**: This allows the domain adaptation methods to be used for diverse US waveforms as, instead of aligning the US sensor data, the US waveform features are used which provide information about the process being monitored as they develop over time.
- **The use of federated learning and multi-task learning with US data:** Federated learning was investigated to maintain dataset privacy when applying transfer learning between datasets. Multi-task learning was investigated to aid LSTM unit learning of the process trajectory. The federated learning methodology performed better than the other methods tested, achieving higher accuracy for 14 out of 16 ML tasks compared with the base case model. Multi-task learning also showed improvements in learning feature trajectories and model accuracy (improving regression accuracy for 8 out of 18 tasks evaluated), however, at the expense of a greater number of hyperparameters to optimise.
- **The use of data augmentation for US data for process monitoring applications**: Data augmentation was a component of the convolutional feature extraction method developed in this thesis. Data augmentation artificially increased the dataset size to train the convolutional feature extractor while ensuring that features specific to each waveform, rather than the position or magnitude of features, were learned. This improved the feature-learning auxiliary task the CNN was trained to perform which classified the dataset from which each previously collected US waveform originated.

Table of Contents

Abbreviations

ABV	Alcohol By Volume
ANN	Artificial Neural Network
CNN	Convolutional Neural Network
CWT	Continuous Wavelet Transform
DWT	Discrete Wavelet Transform
LSTM	Long Short-Term Memory
MAE	Mean Absolute Error
ML	Machine Learning
MMD	Maximum Mean Discrepancy
MSE	Mean Squared Error
PC	Principal Component
PCA	Principal Component Analysis
SF	Single Feature
SVM	Support Vector Machines
TCA	Transfer Component Analysis
US	Ultrasonic / Ultrasound
WPT	Wavelet Packet Transform

1 Introduction to thesis

1.1 Motivation

The manufacturing sector is increasingly utilising developments in digital technologies, especially advances in computing power and connectivity, to integrate processes (Ghobakhloo, 2020). This is often termed as Industry 4.0 or the fourth industrial revolution and is anticipated to culminate in entire value chains being optimised through autonomous data analysis (Kang et al., 2016). In-line and on-line sensors underpin this transformation, where in-line specifies techniques that directly measure the process stream and on-line techniques utilise automatic sampling methods (De Beer et al., 2011), by automatically collecting the real-time data required to inform decision-making. Discrete manufacturing is leading process manufacturing in the deployment of digital technologies owing to the simpler and less variable processes in operation (Fisher et al., 2018). Therefore, a wide range of sensing options must be developed for the process manufacturing sector. Low power (intensities below 1 Wcm^2), high frequency (greater than 100 kHz) ultrasonic (US) sensors monitor the interaction of materials with mechanical sound waves. Ultrasonic sensors have advantages of being low cost, small in size, able to monitor opaque materials, low in power consumption, able to operate non-invasively, non-destructive, real-time, in-line, and do not cause changes to the structure of the material through which they pass (Henning and Rautenberg, 2006). These attributes will ensure that US sensors are the optimal sensor choice for certain process monitoring applications and necessitates research to evaluate their use across a range of processes.

Machine Learning (ML) is the use of computer algorithms to learn patterns in data to perform a task such as making predictions or decisions (Shanthamallu et al., 2018). The correlations in the data that the ML models learn during training have not been explicitly programmed by human operators. There are four main types of ML: supervised, unsupervised, semi-supervised, and reinforcement learning (Ayodele, 2010), of which supervised and unsupervised learning are used in this thesis. Supervised ML maps inputs to outputs during training with the aim being to create a model that accurately predicts the outputs of data that was not previously used during training (Ge, 2017). In contrast, unsupervised learning only uses input data in which patterns are discovered (Ayodele, 2010). Supervised ML is being increasingly combined with sensor measurements as it offers several distinct advantages over conventional calibration methods: Firstly, ML can provide a reduced development time through collection of data over the natural variability of the process conditions opposed to creating calibration curves over a range for each process parameter (e.g., temperature, pressure, dissolved gas concentration etc.). Furthermore, the more complex algorithms used in ML can provide more accurate fitting and negate the reliance of additional measurements such as the process temperature. Correct selection of validation methods can encourage ML models to generalise outside of the parameter ranges it was trained on. Finally, the sensor data can be directly correlated to important process information opposed to material properties. For example, by creating ML models that predict the processing time remaining opposed to the density of the process material, the models may learn different features more useful to predict the desired task.

During application in industrial environments, ML models must be developed that are able to accurately predict information about the process being monitored from new, real-time US data. This requires two criteria to be met: Firstly, the US sensor and ML combination should achieve the level of accuracy desired by the manufacturer and, secondly, the collection of data should not present an unacceptable level of disruption to the manufacturer. For the first criteria, selection of feature extraction methodologies, hyperparameter values, and

algorithms utilised with US sensor data should be optimised to achieve the desired level of accuracy. For the second criteria, the sensing approach should be chosen according to the manufacturer's requirements, e.g., whether invasive sensing is acceptable or whether a non-invasive sensing approach must be used. For example, in the food and drink industry, invasive probes may be discounted for some applications due to risk of contamination between product batches (Bowler et al, 2020). During data collection, normal production schedules may have to be postponed in order to sample the process or to obtain data over a wider range of process parameters than observed during normal operation. Therefore, methods are required to minimise the volume of data required to collect whilst achieving the desired level of accuracy.

1.2 Aim and objectives

Aim: This thesis aimed to develop ML methods to facilitate optimal deployment of US sensors for process monitoring applications in industrial environments.

Objectives:

- **To collect US sensor data for process monitoring applications that enable the thesis conclusions to be expected to extend to industrial environments.** Three types of processes were monitored: material mixing, cleaning of pipe sections, and alcoholic fermentation covering the full range of impacts to US waveforms during process monitoring (see Section 3.1 Experimental datasets). Furthermore, the data was collected at sufficient timescale granularity and the US sensor data was shown to contain useful information about the processes (see Section 9 Discussion).
- **To evaluate different US sensing techniques to determine their benefits and limitations for industrial process monitoring.** Two US sensing techniques were investigated: a non-invasive, reflection-mode technique that can be externally retrofitted to existing processing equipment, and an invasive US probe with reflector plate.
- **To evaluate different feature extraction, feature selection, algorithm types and hyperparameter values to determine the optimal ML pipeline for process monitoring using US measurements.** This reduces time for ML model development in industrial environments by suggesting ML pipelines that achieve the highest accuracy for previous process monitoring tasks.
- **To develop unlabelled domain adaptation methods to utilise previously collected datasets and negate the data labelling burden for sensor deployment.** These methods can be used to transfer ML models between processes without requiring labelled data (the outputs required during ML model training and often acquired by using reference measurements) and therefore negate disruption to a manufacturing process during the data collection stage.
- **To develop labelled domain adaptation methods to utilise previously collected datasets, reduce the data labelling burden for sensor deployment, and improve ML model accuracy on target processes.** These methods can be used to reduce disruption to a manufacturing process during the data collection stage.

1.3 Contributions and novelty

Overall, there are two main contributions from this thesis: Development of the ML pipeline for process monitoring using US sensors, and the development of unlabelled and labelled domain adaptation methods for process monitoring using US sensors. To achieve this, the thesis evaluates US sensing techniques and ML methods across three types of process manufacturing operations: material mixing, cleaning of pipe fouling, and alcoholic fermentation of beer. Two US sensing techniques were also investigated: a non-invasive,

reflection-mode technique, and a transmission-based method using an invasive US probe with reflector plate. Extensive investigation of feature extraction and feature selection methods, algorithms, and hyperparameter ranges were used to optimise ML pipeline for process monitoring using US sensors. This facilitates reduced development time of US sensor and ML combinations when deployed in industrial settings by recommending a pipeline that has been trialled over a range of process monitoring applications. Furthermore, the unlabelled and labelled domain adaptation methods were developed to leverage previously collected datasets and negate or reduce the burden of collecting labelled data for every new process monitoring application.

The novelty contained within this thesis can be summarised as follows:

- The use of CNNs and LSTM layers for process monitoring using US sensor data. CNNs have previously been combined with US measurements for applications such as damage detection (Rai and Mitra, 2021; Rautela et al., 2021) and LSTMs for applications such as flow regime identification (Nguyen and Park, 2020; Ren et al., 2021).
- Evaluating the omission of the process temperature as a feature for process monitoring using US sensor data. This indicates whether the US sensor and ML combinations could be used in industrial applications where measurement of the process temperature is not available. Previous works have either inputted the process temperature into ML models (Wallhäußer et al., 2014), performed experiments at a constant temperature (Resa et al., 2004), or performed calibration procedures across a range of temperatures (Amer et al., 2015).
- The use of labelled and unlabelled domain adaptation for US data for process monitoring. Domain adaptation has been previously used with US measurements to monitor boreholes in oil fields (Gao et al., 2021) and for damage visualisation in plate structures (Alguri et al., 2021).
- The use of labelled and unlabelled domain adaptation on features extracted from US waveforms. Domain adaptation has previously been applied to the US waveforms opposed to the extracted features (Gao et al., 2021; Alguri et al., 2021).
- The use of federated learning and multi-task learning with US data.
- The use of data augmentation for US data for process monitoring applications. Data augmentation has previously been used with US sensor measurements for applications such as weldment defect classification (Munir et al., 2019; Munir et al., 2020).

1.4 Thesis structure

This thesis is constructed from journal publications (Figure 1). **Section 2** presents a literature review accompanying these publications. It is a review paper titled: “A review of ultrasonic sensors and machine learning methods to monitor industrial processes”. This review article was published in *Ultrasonics* on 28th May 2022. The article reviews process monitoring using US sensors and ML combined with US sensors. It reviews feature extraction, feature selection, algorithm choice, and hyperparameters for US waveforms, along with other topics such as data augmentation, domain adaptation, semi-supervised learning and machine learning interpretability. Recommendations for combining US measurements with ML for process monitoring are also provided to the reader using the ideas generated in this thesis. The review also includes the publications produced as part of this thesis, and therefore contextualises their position in the wider literature.

Section 3 presents a single methodology section for the work included in this thesis. The collection of the datasets used in these works are described and explanations of the ML methods utilised are included.

Overall, the articles included in this thesis are separated into two categories; one that optimises the ML pipeline for process monitoring using US sensors and the other that develops domain adaptation approaches to minimise the data collection burden in industrial settings (Figure 1).

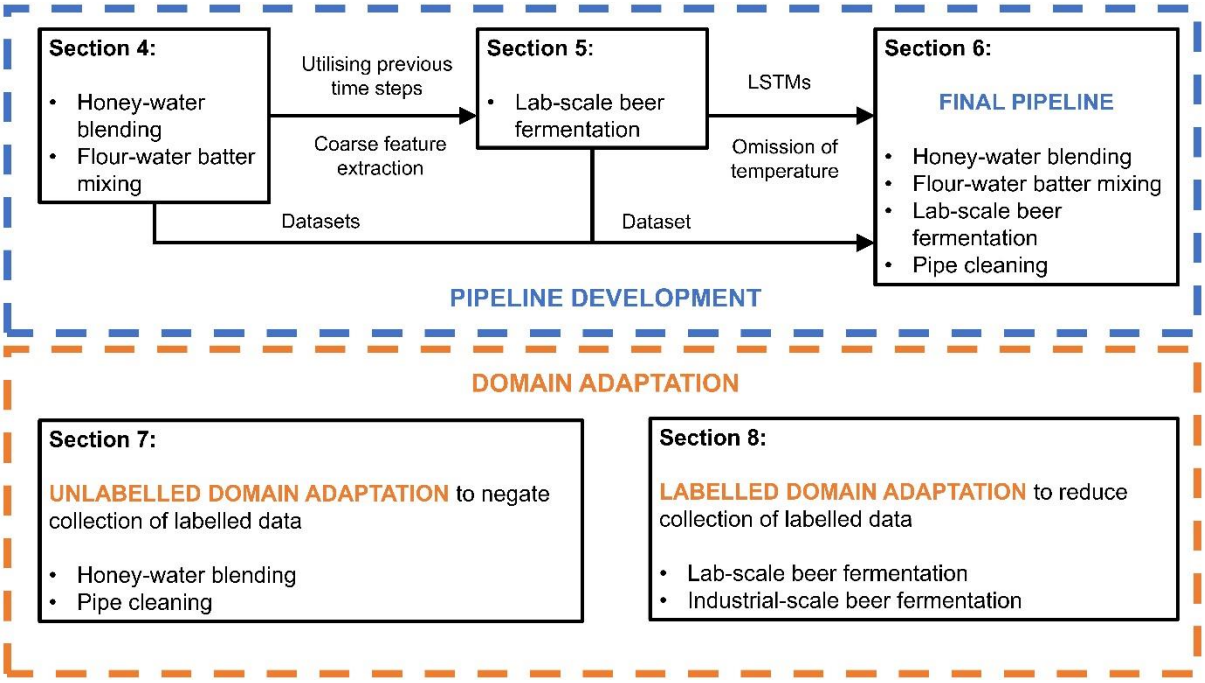


Figure 1: A schematic of the articles included in this thesis.

Section 4 presents a research article titled: "*Monitoring Mixing Processes Using Ultrasonic Sensors and Machine Learning*". This article was published in *Sensors* on 25th March 2020. Two mixing processes were monitored: namely, honey-water blending and flour-water batter mixing. Extensive feature extraction, feature selection, and algorithms were used for classification and regression tasks. The novelty of this work was the combination of an industrially relevant non-invasive, reflection-mode US sensing technique with ML to monitor mixing processes. Furthermore, the use of Long Short-Term Memory (LSTM) layers and Convolutional Neural Networks (CNN) for process monitoring using US sensors was novel. The main conclusions from this work were that using information from previous time steps was vital for accuracy on most tasks and, specifically, that flexible use of previous time steps was required. For example, LSTMs and CNNs, which were able to learn how far back in the process history to use previous US measurements, performed better than Artificial Neural Networks (ANNs) which used a fixed feature lag time.

Authors: Bowler A.L., Bakalis S., Watson N.J.

Author contributions (as published): Conceptualization, A.L.B., and N.J.W.; Methodology, A.L.B., N.J.W Software, A.L.B.; Validation, A.L.B.; Formal Analysis, A.L.B.; Investigation, A.L.B.; Resources, A.L.B., and N.J.W; Data Curation, A.L.B.; Writing—Original Draft Preparation, A.L.B.; Writing—Review and Editing, A.L.B., N.J.W., and S.B.; Visualization, A.L.B.; Supervision, A.L.B., N.J.W., and S.B.; Project Administration, A.L.B., N.J.W.; Funding Acquisition, N.J.W. All authors have read and agreed to the published version of the manuscript.

Section 5 presents a research article titled: "*Predicting Alcohol Concentration during Beer Fermentation Using Ultrasonic Measurements and Machine Learning*". This article was published in *Fermentation* on 4th March 2021. The novelty of this work was the combination of ML with a non-invasive, reflection-mode US sensing to monitor fermentation processes. Furthermore, the evaluation of ML models omitting the process temperature as a feature was also novel. The main conclusions from this work were that comparable ML accuracy could be achieved using the non-invasive, reflection-mode sensing method and omitting the process temperature as a feature. Furthermore, LSTM layers were again determined to be more accurate than using ANNs with a fixed feature time-lag.

Authors: Bowler, A.L., Escrig, J., Pound, M.P., Watson, N.J.

Author contributions (as published): Conceptualization, A.L.B., J.E. and N.J.W.; methodology, A.L.B., J.E., N.J.W. and M.P.P; software, A.L.B. and J.E.; validation, A.L.B.; formal analysis, A.L.B.; investigation, A.L.B. and J.E.; resources, A.L.B.; data curation, A.L.B. and J.E.; writing—original draft preparation, A.L.B.; writing—review and editing, A.L.B., N.J.W. and M.P.P; visualization, A.L.B.; supervision, N.J.W.; project administration, N.J.W.; funding acquisition, N.J.W. All authors have read and agreed to the published version of the manuscript.

Section 6 presents a research article titled: "*Convolutional feature extraction for process monitoring using ultrasonic sensors*". This article was published in *Computers & Chemical Engineering* on 28th August 2021. The novelty of the work was the convolutional feature extraction method presented which consisted of a CNN pre-trained on an auxiliary task followed by Principal Component Analysis (PCA) and the incorporation of additional features before being inputted into LSTM layers. The auxiliary task was to classify to which dataset previously collected US waveform originated from. In this way, the CNN could learn to identify features of US waveforms and be used as a feature extractor for new tasks. PCA was used to extract a reduced number of orthogonal features and the additional features included the variation between consecutively acquired waveforms and the US time of flight. Furthermore, the use of multi-task learning with US data and US data augmentation for process monitoring was novel. The main conclusions from the work were that the convolutional feature extraction method produced more informative features than a coarse feature extraction method. The coarse method used conventional features extracted from US waveforms such as energy, standard deviation, or skewness. Multi-task learning was shown to improve feature trajectory learning.

Authors: Bowler, A.L., Pound, M.P., Watson, N.J.

Author contributions (as published): A.L.B.: Conceptualization, Data curation, Formal analysis, Investigation, Methodology, Software, Validation, Visualization, Writing – original draft, Writing – review & editing. M.P.P.: Conceptualization, Supervision, Writing – review & editing. N.J.W.: Funding acquisition, Project administration, Resources, Supervision, Writing – review & editing.

Section 7 presents a research article titled: "*Transfer learning for process monitoring using reflection-mode ultrasonic sensing*". This article was published in *Ultrasonics* on 18th May 2021. This work investigated the transfer of ML models between similar mixing and similar cleaning processes using unlabelled domain adaptation. A method transferring a single waveform feature was compared with Transfer Component Analysis. The novelty of this work was the use of unlabelled domain adaptation with US sensors for process monitoring, the use of unlabelled domain adaptation on features of US waveform opposed to the waveform itself, and, finally, the investigation of omitting the process temperature as a

feature. The main conclusions from this work were that transfer of a single US waveform feature was found to be optimal as it was most representative of the changing material properties at the sensor measurement area. Further, training on a greater number of source datasets, or source datasets with larger variability in feature distributions, improved transfer learning.

Authors: Bowler, A.L., Watson, N.J.

Author contributions (as published): A.L.B.: Conceptualization, Formal analysis, Funding acquisition, Methodology, Project administration, Supervision, Visualization, Writing - review & editing. N.J.W.: Data curation, Investigation, Resources, Software, Validation, Writing - review & editing.

Section 8 presents a research article titled: "*Domain Adaptation and Federated Learning for Ultrasonic Monitoring of Beer Fermentation*". This article was published in *Fermentation* on 1st November 2021. This work compared three labelled domain adaptation methods to transfer knowledge for process monitoring using US sensors between different fermentation processes. The novelty of this work was the use of labelled domain adaptation for US sensor data for process monitoring, the use of labelled domain adaptation on features from US waveforms opposed to the waveform itself, the use of federated learning with US sensor data, and the use of multi-task learning with US sensor data. The main conclusion from this work was that federated learning produced the highest model accuracy. This was because the federated learning strategy improved the gradient descent operation during network optimisation. During the federated learning training, the local models were trained for a full epoch on each dataset before network weights were sent to the global model. In contrast, during the non-federated learning strategies, batches from each dataset were interspersed.

Authors: Bowler, A.L., Pound, M.P., Watson, N.J.

Author contributions (as published): Conceptualization, A.L.B., M.P.P. and N.J.W.; methodology, A.L.B., M.P.P. and N.J.W.; software, A.L.B.; validation, A.L.B.; formal analysis, A.L.B.; investigation, A.L.B. and N.J.W.; resources, A.L.B. and N.J.W.; data curation, A.L.B. and N.J.W.; writing—original draft preparation, A.L.B.; writing—review and editing, A.L.B., N.J.W. and M.P.P.; visualization, A.L.B.; supervision, N.J.W. and M.P.P.; project administration, N.J.W.; funding acquisition, N.J.W. All authors have read and agreed to the published version of the manuscript.

Section 9 presents the conclusions of the works and **Section 10** discusses directions for future work.

1.5 References

Alguri, K.S., Chia, C.C., Harley, J.B. (2021) 'Sim-to-Real: Employing ultrasonic guided wave digital surrogates and transfer learning for damage visualization' *Ultrasonics* 111, 106338. Doi:10.1016/j.ultras.2020.106338.

Amer, M.A., Novoa-Díaz, D., Puig-Pujol, A., Capdevila, J., Chavez, J.A, Turo, A., García-Hernandez, M.J., Salazar, J. (2015) 'Ultrasonic velocity of water-ethanol-malic acidlactic acid mixtures during the malolactic fermentation process' *Journal of Food Engineering* 149, 61–69. Doi:10.1016/j.jfoodeng.2014.09.042.

Ayodele, T.O., (2010). 'Types of machine learning algorithms' *New advances in machine learning*, 3, 19-48.

Bowler, A.L., Bakalis, S., Watson, N.J. (2020) ‘A review of in-line and on-line measurement techniques to monitor industrial mixing processes’ *Chemical Engineering Research and Design* 153, 463-495. Doi: 10.1016/j.cherd.2019.10.045.

De Beer, T., Burggraeve, A., Fonteyne, M., Saerens, L., Remon, J.P., Vervaet, C. (2011) ‘Near infrared and Raman spectroscopy for the in-process monitoring of pharmaceutical production processes’ *International Journal of Pharmaceutics* 417 (1-2), 32-47. Doi: 10.1016/j.ijpharm.2010.12.012.

Fisher, O., Watson, N., Porcu, L., Bacon, D., Rigley, M., Gomes, R.L. (2018) ‘Cloud manufacturing as a sustainable process manufacturing route’ *Journal of Manufacturing Systems* 47, 53-68. Doi: 10.1016/j.jmsy.2018.03.005.

Gao, X., Shi, Y., Zhu, Q., Li, Z., Sun, H., Yao, Z., Zhang, W. (2021) ‘Domain adaptation in intelligent ultrasonic logging tool: from microseismic to pulse-echo’ *IEEE Transactions on Instrumentation and Measurement* 70, 1–14. Doi:10.1109/TIM.2021.3050154.

Ge, Z., Song, Z., Ding, S.X., Huang, B. (2017) ‘Data mining and analytics in the process industry: the role of machine learning’ *IEEE Access* 5, 20590-20616. Doi: 10.1109/ACCESS.2017.2756872.

Ghobakhloo, M. (2020) ‘Industry 4.0, digitization, and opportunities for sustainability’ *Journal of Cleaner Production* 252, 119869. Doi: 10.1016/j.jclepro.2019.119869.

Henning, B., Rautenberg, J. (2006) ‘Process monitoring using ultrasonic sensor systems’ *Ultrasonics*, 44, e1395-e1399. Doi: 10.1016/j.ultras.2006.05.048.

Kang, H.S., Lee, J.Y., Choi, S., Kim, H., Park, J.H., Son, J.Y., Kim, B.H., Noh, S.D. (2016) ‘Smart manufacturing: Past research, present findings, and future directions’ *International Journal of Precision Engineering and Manufacturing - Green Technology* 3(1), 111-128. Doi: 10.1007/s40684-016-0015-5.

Munir, N., Kim, H.-J., Park, J., Song, S.-J., Kang, S.-S. (2019) ‘Convolutional neural network for ultrasonic weldment flaw classification in noisy conditions’ *Ultrasonics* 94, 74–81. DOI:10.1016/j.ultras.2018.12.001.

Munir, N., Park, J., Kim, H.-J., Song, S.-J., Kang, S.-S. (2020) ‘Performance enhancement of convolutional neural network for ultrasonic flaw classification by adopting autoencoder’ *NDT & E International* 111, 102218. Doi: 10.1016/j.ndteint.2020.102218.

Nguyen, T.H.L, Park, S. (2020) ‘Intelligent ultrasonic flow measurement using linear array transducer with recurrent neural networks’ *IEEE Access* 8, 137564–137573. Doi: 10.1109/ACCESS.2020.3012037.

Rai, A., Mitra, M. (2021) ‘Lamb wave based damage detection in metallic plates using multi-headed 1-dimensional convolutional neural network’ *Smart Materials and Structures* 30(3), 035010. Doi:10.1088/1361-665X/abdd00.

Rautela, M., Senthilnath, J., Moll, J., Gopalakrishnan, S. (2021) ‘Combined two-level damage identification strategy using ultrasonic guided waves and physical knowledge assisted machine learning’ *Ultrasonics* 115, 106451. Doi:1, E.0.1016/j.ultras.2021.106451.

Ren, W., Jin, N., OuYang, L., Zhai, L., Ren, Y. (2021) ‘Gas volume fraction measurement of oil-gas-water three-phase flows in vertical pipe by combining ultrasonic sensor and deep attention network’ *IEEE Transactions on Instrumentation and Measurement* 70, 1–9. Doi:10.1109/TIM.2020.3031186.

Resa, P., Elvira, L., Montero de Espinosa, F. (2004) ‘Concentration control in alcoholic fermentation processes from ultrasonic velocity measurements’ *Food Research International* 37(6), 587–594. Doi:10.1016/j.foodres.2003.12.012.

Shanthamallu, U.S., Spanias, A., Tepedelenlioglu, C., Stanley, M. (2018) ‘A brief survey of machine learning methods and their sensor and IoT applications’ *2017 8th International Conference on Information, Intelligence, Systems and Applications, IISA 2017*, 2018-January, 1-8. Doi: 10.1109/IISA.2017.8316459.

Wallhäußer, E., Sayed, A., Nobel, S., Hussein, M.A., Hinrichs, J., Becker, T. (2014) ‘Determination of cleaning end of dairy protein fouling using an online system combining ultrasonic and classification methods’ *Food and Bioprocess Technology* 7(2), 506–515. Doi:10.1007/s11947-012-1041-0.

2 A review of ultrasonic sensing and machine learning methods to monitor industrial processes

Article title: A review of ultrasonic sensing and machine learning methods to monitor industrial processes

Journal: Ultrasonics

Date published: 28/05/2022

DOI: https://doi.org/10.1016/j.ultras.2022.106776

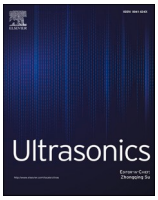
This section presents and literature review of US sensing for process monitoring and US sensing combined with ML methods. Recommendations are provided to the reader for combining US sensors and ML for process monitoring using the ideas development during the creation of this thesis. The published works present in this thesis are also included in the review to contextualise their position in the wider literature.

Ultrasonics 124 (2022) 106776

Contents lists available at ScienceDirect

Ultrasonics

journal homepage: www.elsevier.com/locate/ultras



A review of ultrasonic sensing and machine learning methods to monitor industrial processes

Alexander L. Bowler [a], Michael P. Pound [b], Nicholas J. Watson [a,*]

[a] Food, Water, Waste Research Group, Faculty of Engineering, University of Nottingham, University Park, Nottingham NG7 2RD, UK
[b] School of Computer Science, Jubilee Campus, University of Nottingham, Nottingham NG8 1BB, UK

ARTICLE INFO

Keywords:
Ultrasonic measurements
Machine learning
Deep learning
Industrial digital technologies
Transfer learning
Domain adaptation

ABSTRACT

Supervised machine learning techniques are increasingly being combined with ultrasonic sensor measurements owing to their strong performance. These techniques also offer advantages over calibration procedures of more complex fitting, improved generalisation, reduced development time, ability for continuous retraining, and the correlation of sensor data to important process information. However, their implementation requires expertise to extract and select appropriate features from the sensor measurements as model inputs, select the type of machine learning algorithm to use, and find a suitable set of model hyperparameters. The aim of this article is to facilitate implementation of machine learning techniques in combination with ultrasonic measurements for in-line and on-line monitoring of industrial processes and other similar applications. The article first reviews the use of ultrasonic sensors for monitoring processes, before reviewing the combination of ultrasonic measurements and machine learning. We include literature from other sectors such as structural health monitoring. This review covers feature extraction, feature selection, algorithm choice, hyperparameter selection, data augmentation, domain adaptation, semi-supervised learning and machine learning interpretability. Finally, recommendations for applying machine learning to the reviewed processes are made.

1. Introduction

The manufacturing sector is increasingly using the collection and interpretation of data to inform decision making and improve productivity, sustainability, and product quality [1]. This is part of the fourth industrial revolution, which is projected to culminate in Industry 4.0 and consist of fully interconnected supply chains, processes, and markets where intelligent, automatic decision-making adjusts to demands in real-time [2]. This transformation will be realised through the deployment of industrial digital technologies (IDTs) such as smart sensors, edge computing, cloud computing, the internet of things (IoT), and machine learning (ML). Sensors underpin this transition by acquiring the real-time data required to inform the decision-making process. This necessitates in-line and on-line sensors which do not require human operators, where in-line techniques directly measure the process stream and on-line measurements use automatic sampling methods [3]. Sensors can be adapted into smart sensors through additional functionalities such as wireless IoT connection or by providing some processing of the acquired data to reduce the complexity of the data being transferred [4]. Hardware solutions are required for process interconnectedness such as edge computing, where compute nodes are located close the end devices, or cloud computing, where data is transferred to a centralized cloud location [5]. ML can be used at all levels, from the individual sensors to the centralised data in the cloud, to analyse data and provide automatic decisions [6].

Discrete manufacturing is leading process manufacturing in IDT implementation owing to the much simpler processes to be monitored [7]. A wider range of sensor options is needed for process manufacturing to monitor more complex and often highly variable operations. The process analytical technology (PAT) initiative, first introduced to the pharmaceutical industry in 2004 and since spread to other sectors such as food, demonstrates the desire for greater process understanding [8,9]. PAT focuses on real-time sensor measurements, preferably in-line or on-line, which monitor critical process parameters that effect critical quality attributes of the products. There are many sensor techniques in development, each with different advantages and disadvantages making them suitable for specific applications [10]. Low power (intensities below 1 Wcm^2), high frequency (higher than 100 kHz) ultrasonic (US) sensors monitor the interaction of materials with mechanical sound waves. They benefit from being low cost, small in size, able to monitor

* Corresponding author.
E-mail address: nicholas.watson@nottingham.ac.uk (N.J. Watson).

https://doi.org/10.1016/j.ultras.2022.106776
Received 18 January 2022; Received in revised form 29 April 2022; Accepted 26 May 2022
Available online 28 May 2022

opaque materials, low in power consumption, able to operate non-invasively, non-destructive, real-time, in-line, and do not cause changes to the structure of the material through which they pass [11]. These attributes make US sensors the optimal sensor for certain applications and there use has been demonstrated for monitoring industrially relevant processes as reviewed in Section 3. Therefore, their appeal to industry can be expected to continue to grow.

The most commonly used US measurements include velocity, attenuation, and acoustic impedance [11]. The US velocity is calculated by measuring the time of flight and distance the sound wave has travelled. Attenuation is measured as a loss in sound wave energy as it passes through a material. Attenuation may be caused by absorption in homogeneous materials due to effects such as fluid viscosity, or by scattering due to encountering discontinuities in heterogeneous materials [12]. Acoustic impedance, the product of the sound velocity and material density, is typically monitored by measuring the proportion of a sound wave reflected from a boundary between two materials [13]. Pulse-echo sensing techniques utilise a single sensor to both transmit and receive a sound wave after reflection from an interface. Pitch-catch techniques use one sensor to produce the sound wave and another to receive it [13].

Traditionally, physical inversion models were developed from first principles to determine material properties from US measurements [14]. However, their development becomes challenging in real-life applications where the paths of the sound wave are often complex or the sound wave travels through multiple material interfaces. Furthermore, US properties are highly dependent on temperature and the presence of gas bubbles causes strong reflection of the sound wave, both of which must be accounted for [15,16]. As such, calibration procedures are commonplace that correlate ultrasonic measurements (such as the speed of sound, attenuation, or acoustic impedance) to desired material properties across a range of process parameters without defining the underlying paths of the sound wave. Calibration procedures also become complicated in industrial processes when many parameter ranges must be investigated, such as temperature, gas content, and the content of other heterogeneities [17]. ML uses algorithms to learn solutions to tasks without requiring explicit instructions. Supervised ML is a type of ML that maps inputs (or features) to outputs (or target variables) during training with the aim of producing a model that accurately predicts the outputs of previously unseen input data [18]. Supervised ML offers some distinct advantages over calibration methods: (1) The time investment for calibration procedures can be eliminated simply by monitoring the desired process across its natural parameter variations, so long as these are recorded and a reference measurement is available to label the sensor data with target variables. (2) ML typically uses a greater number of more complex US waveform features compared with calibration procedures allowing more US waveform information to be used in determining material properties. (3) ML models typically employ more complex fitting procedures to map input data to outputs. This allows more accurate predictions while minimising over-fitting to the training data through model regularisation and validation procedures. (4) Validation procedures can encourage development of ML models that accurately predict on new data from outside the range of process parameters that they were trained on. (5) ML models can be continuously retrained as more data becomes available to increase prediction accuracy. (6) Lastly, ML models can correlate sensor data directly to useful process information (such as classifying the state of a process or predicting the processing time remaining) rather than to material properties.

However, a lack of knowledge and experience in applying ML is a barrier to its deployment for US measurement analysis. To develop an adequate ML model, features must be extracted and selected from the US waveform, suitable ML algorithms must be identified and investigated, and a satisfactory set of hyperparameters must be chosen or found. Hyperparameters are any variables that may be selected by the ML model developer. The aim of this article is to facilitate the use of ML in combination with US sensors for in-line and on-line industrial process monitoring. This article first reviews the ability of US measurements to monitor processes before reviewing the combination of US measurements and ML including other areas such as structural health monitoring (SHM). This review covers feature extraction, feature selection and unsupervised learning, algorithm choice, hyperparameter selection, data augmentation, domain adaptation, semi-supervised learning and ML interpretability. Finally, recommendations are provided for combining ML and US measurements for the reviewed processes.

2. Machine learning background

This review includes supervised, unsupervised, and semi-supervised ML methods. Supervised learning uses features as inputs along with corresponding target variables as outputs (also known as labelled data) [18]. The ML algorithms then map the inputs to these outputs with the aim of accurately predicting the target variables for previously unseen data. This may be classification tasks, in which the target variables are discrete categories, or regression tasks where the targets are continuous variables. Unsupervised learning only uses input data for tasks such as finding patterns within the data or reducing its dimensionality. Semi-supervised learning is typically employed when a large volume of unlabelled data and a small volume of labelled data is available [19]. This may be due to the time and expense required to label each data point. Semi-supervised techniques may be used to pseudo-label previously unlabelled data points using knowledge from the labelled data. Then, a more accurate ML model can be constructed using the labelled and pseudo-labelled data compared with using the labelled data alone [18]. Fig. 1 displays a pipeline for supervised ML model development.

Labelled data is required to create the set of model outputs, or target variables, for the model inputs to be correlated to by the ML algorithm. Labelled data may be collected using: an alternative in-line or on-line sensing technique as a reference measurement such as imaging, density measurement, or particle sizer; off-line techniques where material samples are periodically collected during the process; stopping or sampling the process at stages, collecting data using a reference measurement technique, and using semi-supervised learning to pseudo-label the unlabelled data; using the US sensor measurements combined with prior process knowledge to infer process stages; or by transferring ML models between similar processes after domain adaptation [20]. Data labelling may be challenging in factory environments and its consideration should be taken into account throughout the different development stages of sensor and ML combinations.

During the supervised ML pipeline, the input data is usually divided in training, validation, and test sets. The training data is used for model training. Multiple trained models can be evaluated on the validation data to compare between algorithm choice, model architecture, hyperparameter and feature selection. Finally, the best performing models on the validation set are evaluated (or retrained and evaluated using the combined training and validation sets) on the test set to provide an assessment of the full ML pipeline. Several validation techniques are available such as holdout, k-fold cross validation, stratified k-fold cross validation, leave-one-out cross validation, leave-p-out cross-validation, and nested cross-validation. For a detailed comparison of these methods, readers are encouraged to visit [21]. Choice of the training, validation, and test data can also be used to evaluate the ML model's extrapolation capability beyond the process parameter bounds it was trained on [22]. This is a useful approach for applications with limited training data available or highly variable processes.

The success of ML tasks is in part dependent on the features used. A feature is any measurable property of the process being monitored that is inputted into an ML model [23]. Features can originate from the US signal or from other process parameters such as temperature or flowrate. Features may be extracted from the time domain US waveform, the frequency domain (for example after Fourier transformation [24,25]), or time–frequency domain following wavelet decomposition [26]. Wavelet

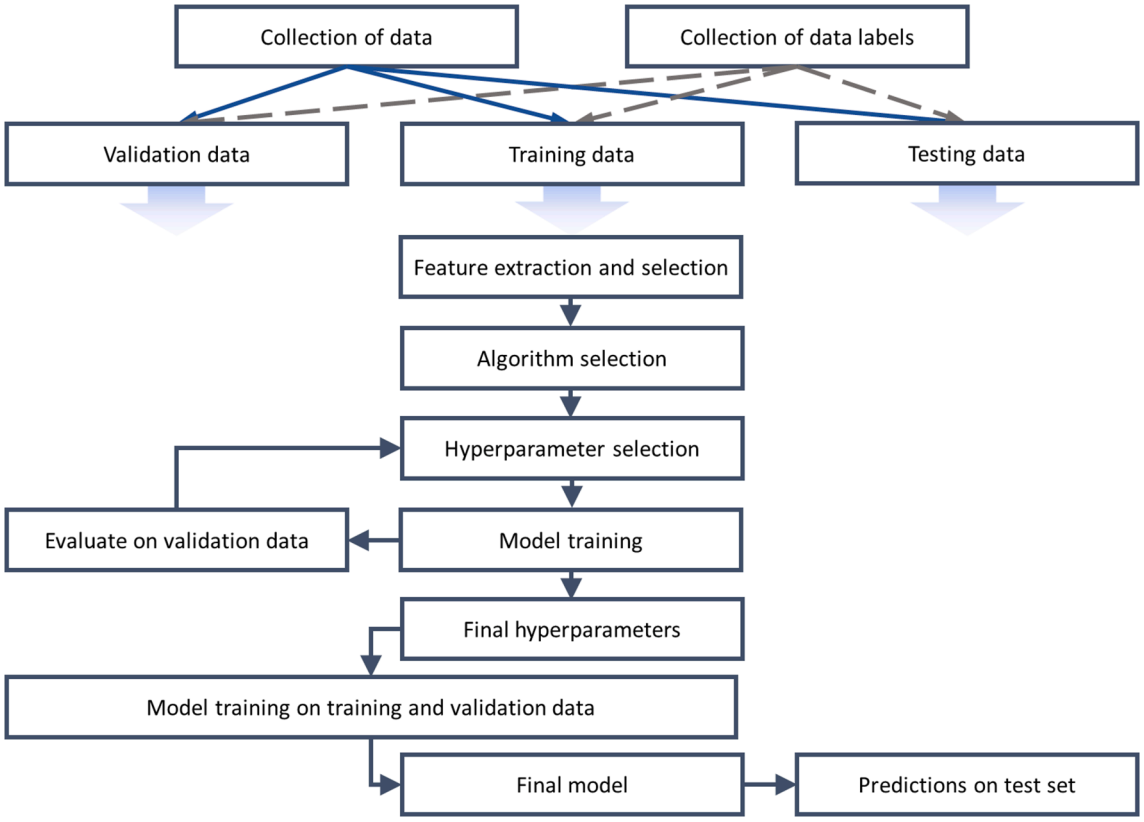


Fig. 1. An exemplar pipeline for developing supervised machine learning models.

analysis uses decaying waveforms as the transform function compared with the non-decaying sine or cosine waves used in the Fourier transform [27]. The continuous wavelet analysis uses a continuous range of frequencies to decompose the US signal whereas the discrete wavelet transform and wavelet packet transform use discrete frequencies at each decomposition. Wavelet packet decomposition performs successive decompositions on each branch of the original signal whereas the discrete wavelet transform only applies successive decompositions to the higher frequency signal content [15,27].

Feature selection encompasses methods of choosing which features to use in ML models or reducing the number of features by using algorithms. A common method is Principal Component Analysis (PCA) as used in [27–29] which is an unsupervised ML method that linearly transforms input variables into new, uncorrelated features called principal components (PCs) [30]. Feature selection can be used to improve ML model fitting by removing redundant information, reducing the likelihood of a model overfitting to its training data, providing simpler optimisation problems, and reducing the computational requirement to train the model [30,31]. However, some information from the input features could be lost leading to a reduction in ML model accuracy.

According to the No Free Lunch theorem, all optimisation techniques are equally as accurate when averaged over all possible problems [32]. Therefore, the optimal algorithm to use is dependent on the application. However, some knowledge of the procedure of each algorithm can help in identifying which to try. For classification tasks, support vector machines (SVMs) find a hyperplane that separates two classes of data by maximizing its distance from the closest data points from each category. In regression tasks, support vector regressors fit lines to continuous data by only accounting for the error from data points outside a set distance from the fitted line. SVMs generalise well to new data and, as they are effective with high dimensional feature spaces, make use of the kernel trick for non-linear fitting [33]. Decision and regression trees (DT) use conditions which are successively applied to the input data until an output decision is reached. They are simple, interpretable, have low computational cost, can be graphically represented, but typically have lower accuracy compared to other algorithms [34]. Random forests (RF) are an ensemble method that combines the predictive performance of multiple DTs by, for example, selecting the most common class predicted in classification tasks or the mean output for regression tasks [16]. K-nearest neighbours (KNN) uses the distance between data points in the feature space and a voting procedure of the K nearest training instances to determine the class or regression value of the queried data point [16]. Artificial neural networks (ANNs) can create new features in their hidden layers from combinations of input features to non-linearly fit model inputs to the outputs. In ANNs, information flows by feed-forward propagation from the input layer, through hidden layers, to the output layer. Weight and bias terms connect all units in the previous layer to all the units in the following layer. During training, the weight and bias terms are iteratively altered through backpropagation of the prediction error and gradient descent steps [34]. A deep neural network (DNN) is an ANN with more than one hidden layer. Convolutional neural networks (CNN) have convolutional layers as well as fully connected layers and are widely used in image recognition tasks [35]. The convolutional layers consist of filters of weights which perform cross-correlation on the input data. This enables CNNs to learn their own features from the input data at lower computational expense and with simpler optimisation than fully connected neural networks of similar size. Furthermore, CNNs are spatially invariant meaning that they are robust to changes in feature locations, unlike fully connected neural networks. The accuracy of ML methods is limited by the choice of features inputted into the models. Therefore, CNNs offer the advantage of negating the need for feature extraction or selection by automatically learning features important to the task (Fig. 2). Long short-term memory neural networks (LSTMs) are able to learn sequences of time series data and are widely used in natural language processing applications. LSTMs are a development of recurrent neural networks (RNNs) that reduce the likelihood of exploding or

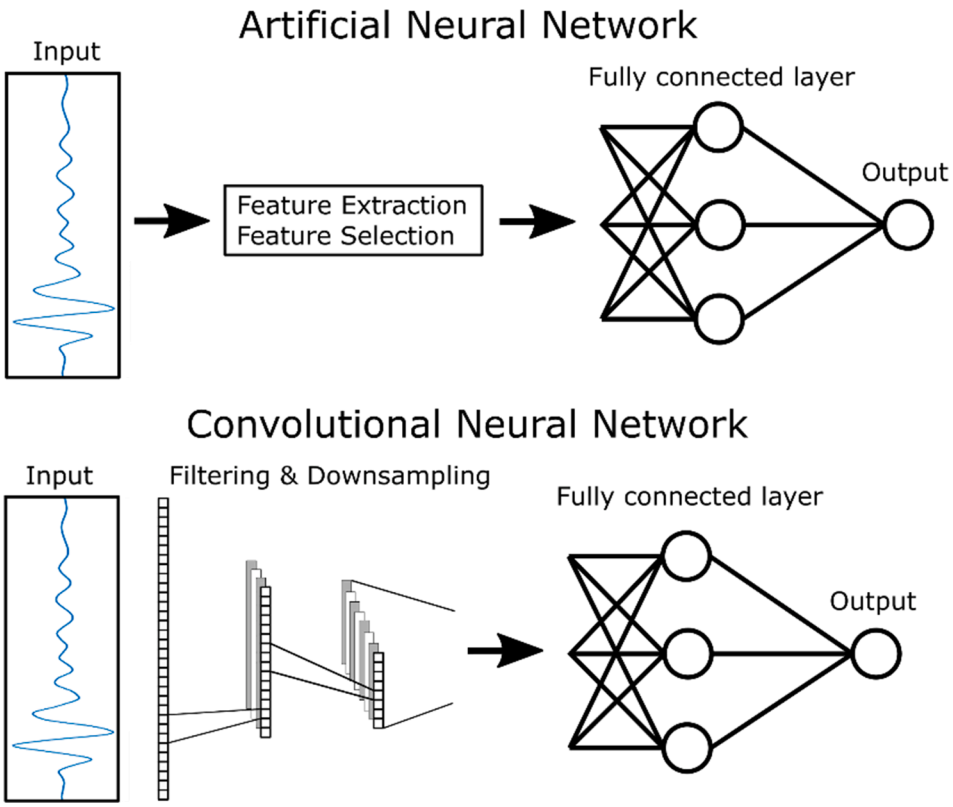


Fig. 2. A comparison between Artificial Neural Networks (ANNs) and Convolutional Neural Networks (CNNs). ANNs require feature extraction and, if necessary, feature selection before using the model. In contrast, CNNs use convolutional filters, and pooling layers to downsample the data, to automatically extract features. ANNs may be used with feature extraction or selection, i.e. the raw data is used as inputs. However, this network would not be robust against spatial variance of the features, unlike CNNs.

vanishing gradients, enabling the learning of long-term dependencies [36]. LSTMs can store representations of sequences by using gate units to update their internal network state. At each time step, LSTMs use the input features as well as information passed from the previous time step to make their prediction.

Hyperparameters are variables that may be selected by the developer. This may be through trial and error, using values previously employed in other works, through grid-searches of possible hyperparameter combinations, or through other procedures such as Bayesian optimisation. Hyperparameter selection can be evaluated using part of the data as a validation set. Hyperparameters may define the structure of the algorithm or how it trains. For example, neural networks (encompassing ANNs, DNNs, CNNs, and LSTMs) often require regularisation to prevent them from over-fitting to the training data and limiting their ability to generalise to new data. Common regularisation techniques include L1 and L2 penalties, early stopping, and dropout layers. L1 regularisation adds a penalty term to the error value that is the sum of all the weight magnitudes, whereas L2 sums the square of the weights [37]. Therefore, a model is penalised for having many or large weight values depending on the chosen magnitude of these regularisation penalties. Early stopping evaluates the current model on a validation set during training. After the prediction accuracy on this validation set decreases for a specified number of iterations, training is stopped [37]. Dropout layers randomly drop network nodes during training according to a specified probability. This effectively "thins" the network during training, allowing multiple input data propagation paths through the network and reducing co-adaptation of the hidden nodes [38].

3. Process monitoring using ultrasonic sensors

The section reviews the use of in-line and on-line ultrasonic measurement techniques to monitor processes including cleaning, fermentation, crystallisation, mixing, extrusion, injection moulding, curing, reactions, tabletting, and membrane fouling.

3.1. Cleaning

Cleaning is a process used to remove material from the internal surfaces of processing equipment in sectors such as food and drink, pharmaceutical and Fast-Moving Consumer Goods (FMCG) [39]. However, cleaning is usually carried out for a predetermined length of time which is designed to over-clean the equipment. With real-time monitoring of fouling removal, time and cleaning resource (e.g. water, energy, and chemicals) use can be minimised. This not only improves process economics but sustainability as well [40]. Furthermore, the build-up of fouling decreases the efficiency of heat exchangers and so a method to detect the presence of this fouling would allow for improved scheduling of heat transfer equipment cleaning [41].

Wallhäußer et al. [41] combined US measurements and an ANN to classify whether a model heat exchanger was fouled by dairy protein deposits. A single US sensor monitored waveforms reflected from the plate-fouling interface and from the far wall of the heat exchanger. The ANN achieved an accuracy of 98.6%. Wallhäußer et al. [42] used ANNs and SVMs to classify the presence of protein or mineral fouling. The SVMs achieved higher accuracies compared with the ANNs. It is suggested that this is due to the ability of SVMs to find global minima, opposed to local minima found by ANNs. In actuality, this may be due to the ability of SVMs to generalise well to new data and that not enough regularisation was applied to the ANNs. This is likely, as no validation set was used to evaluate hyperparameters for the ANN. Only features from the reflection from the plate-fouling interface was used. The SVMs achieved 97.6 % accuracy when the mineral and protein fouling datasets were combined, and 100 % and 98.2 % for the protein and mineral fouling, respectively, when each dataset was used individually. Wallhäußer et al. [43] monitored the cleaning process of protein fouling using a sodium hydroxide solution. Seven US waveform features were used along with the process temperature and cleaning fluid mass flow rate. The gradients of the US features were also monitored to identify a plateau that could indicate the end of cleaning. SVMs achieved greater than 94 % accuracy when classifying whether the heat exchanger was

fouled or cleaned. Úbeda et al. [40] used a single US transducer to monitor cleaning of milk fouling in a model plate heat exchanger. Milk protein deposits build up on heat transfer surfaces during thermal treatment, reducing their efficiency and necessitating cleaning. Cleaning was carried out for 95 min using sodium hydroxide and, subsequently, nitric acid solutions. Two sections of the received US waveforms were analysed: the start of the waveform corresponding to sound wave reflections from the plate-fouling interface, and the third echo reflecting from the far wall of the plate heat exchanger (after passing through the fouling and cleaning fluid or cleaning fluid alone). For the sound wave reflections from the plate-fouling interface, three US waveform features were monitored: the temporal spread, temporal roll-off, and temporal inertia. These features were used in an ANN to classify whether the plate was fouled or cleaned. For the third echo, four features were used: the maximum amplitude, spectral crest factor, spectral centroid, and temporal entropy. Classification accuracies of up to 98 % for the plate-interface features and 96% for the 3rd echo features were achieved.

Escrig et al. [39] monitored the removal of three types of food fouling (gravy, tomato paste, and malt) at two cleaning fluid temperatures (12 °C and 45 °C) using a single US sensor. The US sensing method monitored the US waveform reflected from the interface between the pipe wall and the fouling material. The experiments were conducted in a lab-scale pipe section. The pipe section was cuboid with a flat, stainless steel bottom plate where the US sensor was attached externally. Three features from the received US waveform were monitored: the waveform energy, peak-to-peak amplitude, and root mean square error of the amplitudes at every sample point in the waveform compared to that of a clean pipe. The US sensor could identify differences in the cleaning mechanisms between the mechanical removal of the tomato paste and gravy compared with the dissolution of the malt. The US technique was only sensitive to the area of fouling coverage not the fouling thickness. Escrig et al. [16] expanded this study by training classification machine learning models to predict whether the pipe section was fouled or cleaned. The highest model accuracy was attained when the amplitudes at each sample point in the waveforms were used directly instead of using any further feature extraction methodologies. A K-best feature selection methodology was used to select the number of amplitudes to be used. The classifiers used were KNN, SVM, RF and adaboost RF. It was found that combining multiple datasets from different fouling materials resulted in improved model accuracy. Accuracies up to 99 % were achieved. Simeone et al. [15] used the same pipe test section and materials to monitor cleaning by combining US and optical sensors. A three-level wavelet package transform using the 3 Daubechies mother wavelet was performed on the US waveforms. Afterwards, the mean, standard deviation, minimum, maximum, skewness, kurtosis, and energy were extracted from the decomposed signals. These features were then input into an ANN for a regression task to predict the surface area or volume of fouling remaining. Escrig et al. [44] used classification methods to monitor tomato paste and malt fouling in plastic (PMMA) and metal (stainless steel) cylindrical pipe sections. Accuracies up to 100% were achieved for both pipes. Finally, Chen et al. [45] used a single US sensor to monitor cleaning of wax deposits from a flat duct section over the course of 3 h. A decorrelation coefficient of coda waves compared with a clean plate was monitored and cleaning was completed after 2.2 h.

3.2. Fermentation

Fermentation processes are conventionally monitored through sampling and off-line analysis [46]. However, this has issues of requiring manual operation, risking contamination, and lacking timely results [17]. There are several types of fermentation that have been monitored using US sensors, such as alcoholic fermentation where yeast converts sugar into ethanol and carbon dioxide [46], lactic acid fermentation where lactose is converted to lactic acid through bacteria metabolism [47], and malolactic fermentation in red wines where malic acid is converted into lactic acid which is an important process for developing sensory characteristics [48].

Becker, Mitzscherling, and Delgado [49] used a single externally mounted sensor to monitor beer fermentation in a 300 m^3 tank under industrial conditions for 90 h. The US wave was transmitted across the 4.5 m diameter vessel and was reflected at the far wall before returning to the transducer. The US velocity was monitored and an ANN was used to compensate for the effects of temperature. Resa, Elvira, and De Espinosa [46] mounted two non-invasive US sensors to a square glass bottle (64 mm ID) to monitor the US velocity using a pitch-catch transmission method. A water bath was used to keep a constant temperature of 30 °C. The US velocity decreased with the decreasing density of the fermenting medium. Resa et al. [47,50] used similar experimental methods to monitor lactic acid fermentation. The US velocity decreased throughout fermentation despite no significant change in density. Resa et al. [17] monitored wine and beer fermentations and reported a decreasing US velocity with the decreasing density. The US amplitude was also used to monitor the beer wort fermentation and an increase in attenuation was obtained during the start of ethanol production due to the production of CO_2 bubbles. Similarly, Lamberti et al. [51] used two transducers for transmission across a 35 mm diameter square bottle to monitor wine fermentation. A decreasing US velocity during ethanol production was found.

Hussein, Hussein, and Becker [24] implemented a single US sensor on a circulation line for in-line monitoring of a 60 L (working volume) yeast fermentation process. The sensor used a reflector plate to transmit through the fermenting liquid. Frequency domain analysis, after Fourier transformation, and phase shift correction were used for the time-of-flight measurements. The US velocity was combined with nine waveform features and the temperature and was inputted into an ANN to predict the liquid density. The ANN produced a maximum error of 0.95%. The US velocity increased throughout the fermentation, contradicting the previously reported results presented above. Hoche et al. [52] also found the US velocity to increase during fermentation and the reflection coefficient to decrease. The reason for these results may be due to the larger scale process and industrial conditions monitored in [24,52] compared with [17,46,47,50,51]. At these specific combinations of temperature, along with the content of sugar, ethanol, yeast, and CO_2, the US velocity may increase during fermentation. [52] used invasive sensor probes with a 50 mm reflector plate transmission distance to monitor alcoholic fermentation in vessels up to 2140 L (maximum volume) in size. A sound velocity–density–temperature calibration model achieved an average root mean square error of 0.53% g/g sugar and 0.26% g/g ethanol during the fermentations. However, this does require a secondary measurement of the density of the fermenting wort. Bowler et al. [53] used LSTMs to predict the alcohol concentration during beer fermentation in a 30 L vessel. An invasive probe with a reflector plate was used. Accurate monitoring of the alcohol content was achieved without using the sound wave reflection that had passed through the wort or using the process temperature. This indicates that a non-invasive, reflection-mode US sensing technique could be possible. The energy of the reflection from the probe-wort interface increased throughout fermentation although no trend in the speed of sound was identified owing to variations in the process temperature.

Ogasawara et al. [54] used two invasive probes in pitch-catch mode with a 15 mm transmission distance to monitor yogurt fermentation in a beaker. The US velocity was used to detect phase changes of the yogurt along with variations in the internal temperature caused by the exothermic fermentation reaction. Meng et al. [55] used a single non-invasive sensor to monitor yogurt fermentation through the wall of a 250 ml stainless-steel reactor. The acoustic impedance was monitored using multiple reflections from the wall-yogurt interface. The acoustic impedance of the yogurt increased between pH 5.6 and 5.3.

Novoa-Díaz et al. [48] used an invasive US sensor probe with a reflector plate to monitor malolactic fermentation in red wine. The US velocity increased during the fermentation until reaching a stationary

phase after approximately six days. It was proposed that identification of the stationary phase could be used to determine the fermentation end point. However, temperature compensation would be required to uncover its masking effect on the velocity. Amer et al. [56] presented a temperature compensation strategy by determining the effect of temperature at different concentrations of alcohol and Amer et al. [57] presented two further temperature compensation methods. Çelik et al. [58] presented the design of an invasive sensor probe with reflector plate to be installed in the side of industrial vessels for malolactic acid fermentation.

Lastly, Keskinoğlu and Aydln [59] used two non-invasive transducers in pitch-catch mode to monitor cell growth in a small-scale vessel. From the US velocity, growth curves were obtained that could identify the lag, growth, and stationary phases.

3.3. Crystallisation

Crystallisation is a process predominantly used in the chemical and pharmaceutical industries [60]. The most important parameters to control are the mean crystal size and the crystal size distribution as these determine the properties of the product and effect downstream processing [61]. Mougin et al. [62] used an invasive probe consisting of two pairs of broadband US transducers to monitor the crystallisation of two organic compounds. US attenuation spectroscopy using the Epstein and Carhart and Allegra and Hawley (ECAH) scattering model was utilised. The frequency range of the probe spanned from 1 MHz to over 150 MHz and was capable of measuring particle sizes in the range from 0.01 to 1000 μm. The minimum crystal concentration for size characterisation was 0.1% vol. The technique had limited effectiveness for monitoring the crystallisation of urea owing to the formation of high aspect ratio needle crystals whose long axial length was beyond the sensor measurement range. Mougin, Wilkinson, and Roberts [63] went on to monitor the particle size of crystals in two different polymorphic forms. In this work, the US attenuation spectroscopy method was less sensitive than turbidometric measurements for the determination of the onset of crystallization. Mougin et al. [61] used a 470 ml stainless steel flow-through cell connected to the side-ports of a double-jacketted 2.6 L glass reactor to determine the crystal size and solid concentration during crystallisation. The method could identify secondary nucleation, growth and crystal breakage and the results were used to determine kinetic parameters such as the secondary nucleation rate and growth rate. Li et al. [64] inputted US attenuation spectroscopy measurements into two neural networks. The first ANN was used to predict the mean crystal size and crystal size standard deviation. The five inputs were the attenuation at four different frequencies along with the process temperature. The second neural network was used to predict the US attenuation at a reference frequency to determine solids concentration. Its inputs were the temperature and the mean crystal diameter and standard deviation predictions from the first ANN. Although the authors acknowledge that a single ANN could have been used to make both sets of predictions, two separate models were used to simplify the ANN training. ANNs were used as they did not need knowledge of the solid and liquid physical parameters required for the ECAH model. Furthermore, the ANNs could be used in-line where as the ECAH model must be completed off-line due to the long iterative process required. Lyall et al. [65] was able to monitor crystal breakage and the mass and linear crystal growth rates were determined from the US measurements. Shukla, Prakash, and Rohani [66] employed US attenuation spectroscopy using a single invasive sensor with a reflector plate in a jacketed glass reactor of 0.115 m diameter and 0.25 m height.

Pertig et al. [67] used an invasive probe and measurements of the US velocity and attenuation at a single frequency to determine the mean particle size and suspension density. Experiments were performed isothermally in a jacketed vessel with a diameter of 115 mm and a height of 200 mm. The method could measure particle sizes between 200 and 800 μm with solids content up to 40 wt%. This method was presented as faster, less expensive, and simpler than US attenuation spectroscopy which must be conducted across multiple frequencies. Stelzer, Pertig, and Ulrich [60] used two invasive probes in a 1 L jacketed glass crystallizer to monitor the suspension density, mean crystal size and liquid concentration. One sensor was surrounded by a mesh to prevent crystals from entering the measurement line so the liquid properties could be monitored. Frohberg and Ulrich [68] showed that the same two-sensor technique could be used for the determination of the metastable zone width, nucleation and growth kinetics, seeding events, and detection of phase transitions. Helmdach, Feth, and Ulrich [69] showed that calibration transfer using the same sensing technique was possible between lab and pilot scale processes so long as the influence of gas was minimal in the pilot-plant setup. Morris et al. [70] used a single US sensor with reflector plate in a 250 ml reactor. US attenuation spectroscopy and monitoring of multiple reflections was utilised.

3.4. Mixing

Mixing is a ubiquitous process across manufacturing, such as in the food, chemical, and pharmaceutical industries [10]. In many industries, mixing is typically carried out for a predetermined length of time without monitoring the product quality. Classification of whether the materials were mixed or a prediction of when the mixing process will finish would enable more consistent product quality, more efficient resource use, and better equipment scheduling. Bamberger and Greenwood [71] used an invasive probe to monitor slurry suspension in a 1.91 m diameter tank. The US attenuation was measured across three transmitter–receiver transducer pairs located at different heights along the probe and separated by a 10.2 cm distance. Fox, Smith, and Sahi [72] and Salazar et al. [73] used single, invasive sensors to monitor air incorporation into batters during mixing. Both sensors monitored the changing acoustic impedance of the batter by measuring the reflected sound wave. Tourbin and Frances [74] used a flow-through cell and US attenuation spectroscopy to monitor the suspension and aggregation of nanoparticles (mean diameter of 80 nm) in a 1 L capacity stirred tank. Liu et al. [75] also used a flow-through cell and US attenuation spectroscopy in combination with Electrical Resistance Tomography (ERT) to monitor crossflow membrane emulsification. The size distribution and concentration of droplets was determined using the ECAH inversion model. Hunter et al. [76] employed an in-situ, multi-frequency acoustic backscatter system to monitor high concentration particle dispersion. Homogeneous glass powder dispersions were monitored at small and large (2 m^3 mixing tank) scale. Transducers with central frequencies of 1, 2, 4, and 5 MHz were used in pulse-echo mode and the attenuation decay with time of the returning signal was dependent on the particle concentrations. Bowler, Bakalis, and Watson [27] used single, non-invasive US sensors in reflection-mode to monitor two model mixing systems: honey-water blending and flour–water batter mixing. Classification ML models were developed to predict if the materials were mixed or not, and regression models were trained to predict the time remaining until (or time passed since) the materials were fully mixed. ANNs, SVMs, LSTMs, and CNNs were all tested with extensive feature extraction in both the time and time–frequency domains (after applying the discrete or continuous wavelet transform). Multi-sensor fusion between two sensors was also investigated. Classification accuracies of up to 96.3% for the honey-water blending and 92.5% for the flour–water batter mixing were achieved, as well as R^2 values for the regression models of up to 0.977 for the honey-water blending and 0.968 for the flour–water batter mixing. Each prediction task achieved optimal accuracy using different ML algorithms and feature extraction methods.

3.5. Extrusion

Using ultrasonic sensors to monitor extrusion processes has applications in industries such as polymer processing and food production [77]. Coates et al. [78] used US and spectroscopic sensors to monitor the

blending of polyethylene and polypropylene pellets in a 38 mm single screw extruder. Two US transducers were used in pitch-catch mode across a path distance of 15 mm. The US velocity was more sensitive than the spectroscopic measurements in detecting a change of blend composition. Barnes et al. [79] monitored the blending of ethylene vinyl acetate (EVA) random co-polymers with varying vinyl acetate (VA) content. US probes were implemented into dies attached to the end of the extruder barrel and the time of flight was measured. Although, pressure fluctuations that effected the US velocity measurements did not affect the NIR probe measurements. Alig et al. [80] combined US, spectroscopic, and rheometry measurements, as well as US and dielectric measurements, in slit die extruders. US attenuation was used to monitor polymer filler blending. Sun et al. [81] monitored filler dispersion during extrusion using two US sensors. An ANN was trained using the US velocity and attenuation along with the pressure, temperature, filler type and feed rate for prediction of the dispersion index. An error of less than 5% was achieved.

Fischer et al. [82] combined US and spectroscopic measurements to monitor additive blending during extrusion. Fischer et al. [83] used NIR spectroscopy and US attenuation spectroscopy to monitor the extrusion of polymer nanocomposite blends. The measurements were used to determine the dispersion extent and the impact strength of the polymer product. Schober et al. [84] also used US attenuation spectroscopy to monitor dispersion and particle size during polymer melt extrusion. Wöckel et al. [85] monitored the reflected sound wave using a single US sensor. The standard deviation of consecutive signals was used to determine the filler concentration. Halmen et al. [86] used US tomography to determine filler distribution in polymer melts. The US velocity and attenuation from a 60 mm ID sensor ring consisting of 40 transducers were used.

3.6. Injection moulding

US sensors have been widely applied to monitor injection moulding of polymer materials. For a full review of this area, the interested reader is directed to [87]. Recently, Wu et al. [88] presented a T-shaped extension nozzle with two integrated high temperature US transducers in transmission mode for non-invasive monitoring of injection moulding. The US velocity and attenuation could follow the process stages and was also correlated to the polymer flow speed. Altmann, Praher, and Steinbichler [89] used three US transducers (10 MHz) in pulse-echo to monitor melting behaviour in injection moulding. Zhao et al. [90] used a single US transducer for in-line monitoring of microcellular injection moulding. The duration of the US signals and the change in US velocity could be used to monitor variations in cell size and thickness of the skin layer. Cheng and Wu [91] used two high temperature US sensors in transmission mode to non-invasively monitor injection moulding of two types of plastic. Each stage of the injection moulding process could be identified from the US measurements, and the effect of injection speed on the quality of the final product could be monitored using the US velocity. Zhao et al. [92] used the US velocity and pressure measurements for in-line temperature measurement during injection moulding. Finally, outside of polymer processing, Grob et al. [93] used four US transducers (two to transmit US waves and two to receive) to monitor the crystallisation, solidification, contraction, and mould wall detachment of chocolate products. Detachment from the mould wall produced a reduction in the US amplitude.

3.7. Curing

Rath et al. [94] used US sensors in a compression mould and measured the velocity and attenuation during curing to evaluate the effects of different mouldings compounds, elevated temperatures, and filler, moisture, and hardener content. Lionetto, Tarzia, and Maffezzoli [95] used two air-coupled US sensors to monitor the curing of resin. The US measurements were corrected for the variations in air temperature caused by the exothermic reaction by periodically switching the US sensors from pitch-catch to pulse-echo mode. Lionetto and Maffezzoli [96] used the US velocity and attenuation to monitor the curing processes of thermosetting resins. Both contact and air-coupled US techniques were used. Koissin, Demčenko, and Korneev [97] used a noncollinear US wave mixing approach to monitor curing of resin. This technique uses the interaction of two US waves to produce scattered waves with mixed frequencies. Ghodhbani, Maréchal, and Duflo [98] used a single US sensor to monitor the liquid viscous, glassy transition, and saturation solid stages during curing of an epoxy resin. Dominguez-Macaya et al. [99] used an air-coupled US sensor to monitor longitudinal and, after the gel point was reached, shear waves during ultraviolet curing of a vinyl ester resin. The US system was also used to monitor the change in thickness of the resin due to shrinkage. Finally, [100–103] measured the velocity and attenuation of US waves to monitor curing of carbon fibre-reinforced plastics.

3.8. Reaction monitoring

Pawelzyk, Toledo, and Willenbacher [104] monitored US velocity and attenuation during styrene emulsion polymerization. However, this was conducted using a through-transmission method across a small-scale sample volume (15 ml). Buckin and Atlas [105] and Buckin [106] demonstrate how non-invasive, through-transmission measurement of the US velocity and attenuation can be used to determine many phenomena of reactions at small scale. For example, [106] reviews the monitoring of substrate and product concentrations, degree of polymerisation, polymer molar mass, reaction rates, catalyst inhibition, reversible and irreversible thermal deactivation, and particle size changes in dispersions. Figueiredo et al. [107] used a single transmission-based US sensor to monitor the transesterification process of biodiesel at small scale (70 mm diameter vessel). Stabilisation of the US velocity and amplitude indicated the achievement of the maximum yield and that the process should be stopped at this point. Baêsso et al. [108] used the US velocity and attenuation to determine the content of contaminants or by-products in biodiesel samples. In this way, US sensors could eventually be used for in-line monitoring of transesterification final products. Schmachtl et al. [109] monitored the synthesis of zeolite A and zeolite X using a transmission US sensing method a small scale. Decreases in the US velocity and attenuation were correlated with gel formation at the start of the process. An increase in attenuation and corresponding peak in US velocity indicated zeolite crystallisation. Hums, Baser, and Schwieger [110] used an invasive US transducer and reflector plate to monitor nucleation and crystal growth during the hydrothermal synthesis of zeolite A and X from coal fly ash. Van Groenestijn et al. [111] used a US nanoparticle sizer probe to monitor the synthesis of spherical silica nanoparticles. The ECAH method was used to obtain the size and concentration of particles.

3.9. Tabletting

Tabletting involves the compaction of powders into tablet forms using punches. Stephens et al. [112] used a single, non-invasive, embedded US sensor to monitor the mechanical properties of tablets during compaction. The time of flight and reflection coefficient of the sound wave reflecting from the interface between the upper punch and the powder was monitored. Leskinen et al. [113] used two transducers implemented inside the upper and lower punches of a tableting machine. Through-transmission was used to measure the US velocity and frequency spectra to monitor the mechanical properties of tablets.

3.10. Membrane fouling

US measurements have been widely applied as a non-invasive technique to monitor membrane fouling during micro-, ultra- and nano-filtration separation processes [114,115]. The amplitudes of reflected

sound waves from the fouling and membrane layers are measured to monitor fouling formation. For example, Li et al. [116] used an in-situ US technique to monitor organic and colloidal fouling during nano-filtration. Differences in the fouling process were observed in the US measurements when using different mixtures of foulants.

4. Ultrasonic sensors and machine learning

This section reviews the use of ML with US sensor measurements applicable to industrial process monitoring and other similar applications. Feature extraction, feature selection, algorithm choice, hyper-parameter selection, data augmentation, and domain adaptation is reviewed. For reviews on similar areas, the interested reader is directed to [117–128].

4.1. Feature extraction

The choice of features to investigate may be decided through either understanding of US sensors and the process being monitored, plotting features over the course of the process and monitoring their trends, or preliminary experiments to evaluate the accuracy of ML models using different feature combinations. As explained in Section 2, features can be extracted from the US waveform in the time domain, frequency domain following the Fourier transform, time–frequency domain following wavelet transform, or after other transformations such as cosine [25], chirplet [129], or short-time Fourier [130]. Empirical mode decomposition has also been widely applied in the SHM community and a review of this area can be found in [131].

The choice of Mother wavelet and level of decomposition are important decisions when applying wavelet transformation to US waveforms [25]. For this reason, a review of wavelet methods to analyse US signals is provided in Table 1. Typically, the choice of mother wavelet is selected as that most visually similar to the received US waveform [27]. The number of decompositions and vanishing moments can be decided by evaluating ML models using different values [27].

Table 2 reviews features extracted from US waveforms and inputted into ML models. A US waveform in the time, frequency, or time–frequency domain is a function composed of amplitudes. Specific amplitudes in these functions can be monitored, such as the maximum, minimum, or peak-to-peak amplitudes. Other features use many of these function amplitudes in a single measure. For example: the standard deviation or variance monitor amplitude dispersion relative to an average magnitude, features such as the temporal slope monitor the rise or decrement of amplitudes in a function, features similar to the energy provide a measure of the overall magnitude of the function, the crest factor measures the dominance of the maximum amplitudes, and skewness and kurtosis provide measures of the shape of the function. Another set of features can provide measurements of the position of the function in its respective domain, such as the time centre, average frequency, temporal duration, or bandwidth. Also, all, or a subset of, function amplitudes may be used as features directly rather than incorporated into other measures [16,27,35,148]. The time of flight monitors the distance in the time domain between two waveforms and is used to measure the speed of sound in the process material. The variance between consecutive waveforms of any of the previously listed features may also be monitored for example to monitor the production of CO_2 during fermentation [53] or identify flow regimes [149,150]. Time-lagged features or feature gradients can be used to incorporate information from past time-steps into the ML models for processes that progress over time. Finally, additional features, such as the process temperature or material mass flowrate can be used to provide extra information to the ML models about the process being monitored.

Table 1
A review of US waveform decomposition through wavelet analysis as a feature extraction methodology for ML models.

Reference	Wavelet decomposition transform	Number of decompositions	Mother wavelet and number of vanishing moments
[15]	Wavelet packet	3	3 Daubechies
[27]	Discrete wavelet	3, 5, 7	Symlet
[27]	Continuous wavelet	N/A	Morlet
[132]	Wavelet packet	5	5 Daubechies
[133]	Discrete wavelet	4	Daubechies
[134]	Wavelet packet	4	5 Daubechies
[135]	Discrete wavelet	4	5 Coiflet
[136]	Continuous wavelet	N/A	Various investigated
[137]	Continuous wavelet	N/A	3 Morse
[138]	Discrete wavelet	4	1 Debuchet
[139]	Wavelet packet	3	4 Daubechies
[140]	Discrete wavelet	7	Created own mother wavelet
[141]	Wavelet packet	3	8 Symlet
[142]	Discrete wavelet	3	5 Coiflet
[143]	Discrete wavelet		10 Daubechies
[144]	Discrete wavelet	5	10 Daubechies
[145]	Discrete wavelet	5	8 Daubechies
[146]	Discrete wavelet	5	8 Symlet
[147]	Continuous wavelet		Gaus

4.2. Feature selection

This section reviews feature selection methods used with US measurements and ML. Feature selection encompasses methods to reduce the number of features inputted into ML models, however, it is not a mandatory step in the ML pipeline. Feature selection can improve ML model training by removing redundant information, reducing the likelihood of overfitting, providing an easier optimisation problem for the algorithm, and reducing the computational requirement for model training [30,31]. However, some information from the input features could be lost leading to a reduction in accuracy. As explained in Section 2.2, PCA is a common method of feature selection and was employed in [27–29,156–159]. Ref. [16] used a K-best predictors method to select the sample point amplitudes from waveforms. This involved using a grid search of the amplitudes used and an F-test to determine their importance. Ref. [28] used a Garson's method which calculated feature importance from weights of a previously trained ANN. The feature importance was scored between 0 and 1 and features scoring below a threshold value of 0.35 were discarded. Ref. [25] used the Wilcoxon-Mann-Whitney rank test to find class discriminant features. This method is usable in binary classification tasks and does not determine if features are redundant or not.

Autoencoders can also be used as unsupervised ML methods for feature selection. Autoencoders are a type of neural network that aims to reconstruct its inputs after having passed the data through a bottleneck, or latent space, in the network. For example, [160] used a convolutional autoencoder as a feature extraction methodology. During training, the information about the input signal contained in the latent space is maximised so that it may be reconstructed. After training, the encoder part of the network (from the inputs to the latent space) may be applied to new data as a feature extractor. Similarly, [161] used autoencoders as a feature extraction method for detecting fatigue damage in structures. Autoencoders have also been used for other applications when used with US sensor data. For example, [162] used a convolutional autoencoder to reconstruct noiseless US signals after artificial noise has been added. In this way, the trained autoencoder could then be used to denoise new US signals. The input signals consisted of 2048 datapoints and the latent space was 256 neurons. Ref. [163] used convolutional denoising autoencoders to remove the effects of temperature on US guided waves for structural health monitoring applications. Refs. [164,165] applied denoising autoencoders for US waveforms to improve the signal to noise ratio. As noise is random fluctuations overlaying a US waveform, autoencoders are unable to learn a relationship between the noise and

Table 2
A review of features extracted from US measurements in the time, frequency, or time–frequency domain for ML.

Category of features	Features	Description	References
Specific amplitudes in function	Maximum amplitude	The largest amplitude in the investigated function interval	[28,40,53,133,140]
	Minimum amplitude	The smallest (or largest negative) amplitude in the investigated function interval	[15,53,133,140]
	Peak-to-peak amplitude	The difference between the largest and smallest amplitude in the investigated function interval	[28,52,140]
Dispersion of amplitudes along the function	Standard deviation	A measure of the dispersion of amplitude values with respect to the mean	[15,140]
	Variance	A measure of the dispersion of amplitude values with respect to the mean	[28,29,133]
	Temporal spread	A measure of the dispersion of amplitude values with respect to the mean	[40]
	Temporal entropy	A measure of amplitude variability along the function	[24,40]
	Spectral standard deviation	A measure of the dispersion of frequency amplitudes with respect to the mean	[28]
	Spectral smoothness	The variability of frequency amplitudes with respect to their neighbouring amplitudes	[42,43]
	Spectral spread	The variance of frequency amplitudes with respect to the average	[24]
	Spectral entropy	The amplitude variability along a frequency domain function	[24]
Measures of the rise or descent of function amplitudes	Temporal roll-off	The time value at which 90 % of the signal energy is concentrated	[40]
	Logarithmic decrement	The logarithmic decrease of amplitudes in a function	[41]
	Temporal slope	A measure of the rate of decrease in function amplitudes	[28,43]
	Descent time	A time value. For example, the time at which the slope of amplitude descent crosses zero.	[28,29,43]
	Lower 25 % of power spectrum	Fraction of total energy between lower 25 % level and peak frequency amplitude	[28]
	Upper 25 % of power spectrum	Fraction of total energy between peak and upper 25 % level frequency amplitude	[28]
	Rising time	The time value for function increase from 25% level amplitude to peak	[29]
Energy	Temporal energy	The sum squared amplitude of the waveform interval investigated	[15,16,24,27,29,41-43,53,151,152]
	Spectral energy	The sum squared amplitude of the frequency domain interval investigated	[24]
	Temporal inertia	Weighted average of the signal amplitude in time domain	[40]
	Mean	The mean amplitude in a function	[15,28,133,140]
	Sum absolute amplitude	A measure that gives lesser weight to large amplitudes compared with the energy	[27]
	Median	The median amplitude in a function	[140]
Crest factor	Temporal crest factor	The magnitude of the maximum signal amplitude in the time domain compared to the average	[24,42,43]
	Spectral crest factor	The magnitude of the dominant frequency compared with the average	[24,40,42,43]
Average frequency	Mean frequency	The mean frequency value	[28,133]
	Spectral centroid	The frequency value where half of the waveform energy is contained	[24,28,29,40,133]
Temporal position	Time centre	The centre of the function in the time domain	[29]
Temporal duration	Pulse duration	The length of time between the start and end of the waveform	[29,151]
Bandwidth	Measured bandwidth	The range of the measured frequency values	[28,29,151]
Skewness	Temporal skewness	A measure of the lack of symmetry in the waveform	[15,151]
	Spectral skewness	A measure of the lack of symmetry in the frequency domain	[24,29]
Kurtosis	Temporal kurtosis	A measure of the tailedness of the waveform	[15]
	Spectral kurtosis	A measure of the tailedness of the frequency domain function	[29]
Amplitudes at sample points in the waveform		Using the amplitude at each sample point in a waveform as individual features	[16,27,35,148,154]
Variations in features between consecutive waveforms	Standard deviation of the energy	A measure of the dispersion of consecutive waveform energy values	[53]
	Frequency analysis of consecutive amplitudes	A measure of the dispersion of consecutive amplitude values	[149]
	Features extracted from velocity variations	Measures of the dispersion of consecutive velocity values	[150]
Time of flight		The length of time for a sound wave to travel through a material	[53,81,152,153,155]
Feature gradients		A measure of the current time step feature with respect to previous time steps	[27,53]
Other features	Temperature	The process temperature	[24,43,81]
	Mass flow rate	The process mass flow rate	[43,81]
	Pressure	The pressure of the process	[81]
	Material type	Information about the material being processed	[81]

Table 3
Hyperparameters used for ANNs with features from US measurements.

Reference	Number of input features	Number of neurons in each hidden layers	Training algorithm	Additional information
[172]	Varied between 128 and 512	128		Root mean square error goal of 0.01 during training Learning rate and momentum term varied
[173]		5	Levenberg–Marquardt	
[153]	2	10		Training stopped after mean squared error of 1×10^{-6} or 6,000 iterations reached Learning rate of 0.6 Momentum rate of 0.4
[152]	2	10		Training stopped after 20,000 iterations or 1×10^{-6} mean squared error reached Learning rate of 0.6 Momentum rate of 0.4
[64]	5	50	Levenberg–Marquardt	Training stopped after a maximum of 500 iterations or desired minimum mean square error of 5.0 $\times 10^{-4}$ reached
[64]	4	20	Levenberg–Marquardt	Training stopped after a maximum of 500 iterations or mean square error of 5 $\times 10^{-4}$ reached
[28]	24	40		Stopping criteria were: maximum epochs 500, minimum error gradient equal to 1×10^{-5}, minimum mean square error equal to 1×10^{-5}
[144]	Varied between 1 and 7	Varied between 6 and 22		Learning rate of 0.2 Additional momentum 0f 0.5
[145]		12		Learning rate of 0.05
[174]	2	3, 3, 1	Evolutionary optimisation	Learning rate of 0.5, Mutation rate of 0.04, population size of 50, and cross over rate of 0.2
[175]	10	10, 2	Levenberg–Marquardt	
[132]	32	100		Training stopped when accuracy of 1×10^{-3} reached
[143]		32, 12	Scaled conjugate gradient	
[176]	4	3 hidden layers	Evolutionary algorithm	Population of 50 individuals Crossover probability of 0.95 Mutation probability of 0.01 200 generations
[133]	8	8, 25	Scaled Conjugate Gradient	Training continued until error goal of 1×10^{-2} achieved
[41]	5	2		
[146]		10	Scaled Conjugate Gradient	
[24]	6	11		
[42]	5	2		
[43]	9	14		
[177]		10	Scaled Conjugate Gradient	
[40]	3	9		Training continued until error was below 0.1 % accuracy
[178]	4	5, 2	Bayesian regularization Levenberg–Marquardt	Network trained 100 times and weights with lowest score on the validation set were used Training was stopped once the Summation of Squared Errors reached below 10^2 or 10^1 depending on the prediction task, or 1000 epochs were reached
[139]	3	4	Levenberg–Marquardt	
[148]	502	980, 270		3 dropout layers with 0.5 probability Trained for 3000 epochs
[179]	151	10	Levenberg–Marquardt	
[180]		50	Levenberg–Marquardt	Trained for 5 epochs
[35]		3 hidden layers		Dropout layers with 0.5 probability Trained for 500 epochs
[154]	11,501	502		Trained for 500 epochs
[154]	11,501	4 hidden layers		
[181]	5	5		
[182]	3300	1300, 660, 330, 165	Scaled Conjugate Gradient	ReLu activation function Learning rate of 0.001 Trained for 1000 epochs
[15]	7	7	Bayesian Regularization	
[27]	Various	Determined through grid-search	Levenberg–Marquardt for regression Scaled Conjugate Gradient for classification	Early stopping applied with a validation patience of six iterations Ten networks were trained and their scores averaged A grid search determined L2 regularisation magnitude
[183]		1024, 512, 265, 128	Adam	ReLu activation function used Dropout layers with probabilities 0.2, 0.3, 0.4, 0.5 Learning rate of 0.0001 Batch size of 8 Trained for 10,000 epochs
[183]		2048, 2048, 1024, 1024	Adam	ReLu activation function Dropout layers with probabilities 0.2, 0.2, 0.2, 0.2 Learning rate of 0.0001 Batch size of 8 Trained for 10,000 epochs

(*continued on next page*)

Table 3 (*continued*)

Reference	Number of input features	Number of neurons in each hidden layers	Training algorithm	Additional information
[184]	5000	1000	Learning rate of 0.001	Dropout layers with 0.7 probability before and after each fully connected layer ReLu activation function
[184]	5000	1000, 1000	Learning rate of 0.001	Dropout layers with 0.7 probability before and after each fully connected layer ReLu activation function
[185]	24	6, 10, 2	Levenberg–Marquardt Algorithm	1000 epochs

the signal and so fail to reconstruct noise during training. Ref. [166] used simulations and a small number of experimentally collected samples to train an autoencoder to reconstruct full wavefield data from sparsely sampled US measurements. This allows US data to be collected at lower sampling frequencies and artificially reconstructed as higher sampling frequency waveforms, thus reducing measurement acquisition time. Ref. [167] trained autoencoders to reconstruct flawless US signals so that when a flaw is detected the autoencoder fails to reconstruct the waveform. This allows flaws to be identified even if they overlap the initial transducer pulse. Ref. [168] used stacked autoencoders to localise and classify acoustic emission sources in riveted panels.

As comparatively few feature selection methods have been used with US sensor measurements, the interested reader is directed to [169–171] for further information on techniques available such as wrapper, filter, hybrid, and embedded methods.

4.3. Algorithms

This section provides a review of the hyperparameters used in ML models with US measurements. Table 3 reviews the hyperparameters used for ANNs with US measurements. These can determine the structure of the network (e.g. the number of hidden layers and the number of neurons in each hidden layer) or the training of the network (e.g. training algorithm used and the learning rate).

Similar to ANNs, CNNs also require hyperparameter selection to decide the structure of the network and how it trains. However, the hyperparameters in the convolutional layers, which detect features, and pooling layers, which downsample the data, must also be chosen. Table 4 reviews the hyperparameters used for CNNs combined with US measurements. LSTMs also require selection of similar hyperparameters to ANNs in addition to the number of LSTM units to employ. Ref. [27] used 50 LSTM units, a fully connected layer of 50 neurons, a dropout layer with a probability of 0.5. The network was trained for 60 epochs using the Adam optimisation algorithm, a learning rate of 0.01, a batch size of 2 and a gradient threshold of 1 to prevent exploding gradients. Ref. [182] used the Scaled Conjugate Gradient optimisation algorithm and a learning rate of 0.01 for 400 epochs. Two dropout layers with probabilities of 0.1 and 0.2 were used. Ref. [186] used 32 LSTM units, two fully connected layers (with 512 and 128 neurons), and two dropout layers with probabilities of 0.25 and 0.2. Training was carried out for 500 epochs with a learning rate of 5×10^{-5} and a batch size of 8. Ref. [138] used 6 LSTM units and [187] used 7. Ref. [188] combined CNNs with two layers of 8 Gated Recurrent Units (GRU) along with dropout probabilities of 0.5 to extract temporal features from US waveforms. GRUs are similar mechanisms to LSTMs only simpler with two gates rather than three and therefore, generally, lower performance when learning long sequences. Ref. [189] trained a ConvLSTM encoder-decoder DNN on finite element simulations of 2-D US wave propagation. ConvLSTMs allow the learning of spatio-temporal dependence in input sequences by employing convolutional structures within the LSTM units. The trained model was comparable in accuracy to finite element simulations but faster to solve by approximately an order of magnitude through negating the computation of numerical calculations. Ref. [190] used an LSTM layer following a CNN for damage detection of copper pipelines using laser ultrasonic scanning.

CNNs have also been used for other applications such as for B-scan US images [208], combining multiple B-scan images [209], C-scan images [210], and guided waves [211]. Ref. [199] used a CNN to deconvolve overlapping US signals and extract the time of flight and amplitude. Ref. [212] used 3D CNN for defect detection by using US images of wave propagation from multiple time steps. Refs. [213,147] presented a 22-layer GFresNet and GFresNET-2D for guided-wave focusing defect signal classification, respectively. Ref. [214] employed a U-net style CNN for predicting the material thickness of plate-like structures using acoustic steady-state excitation spatial spectroscopy. Ref. [215] used a CNN for corrosion inspection on an aluminium plate using broadband Lamb waves.

For models using support vectors, the most commonly tuned hyperparameters include C, the penalty factor, γ or σ, the influence a single training example has, and the type of kernel used, e.g. linear, polynomial, or radial. In support vector regression, epsilon defines the distance from the fitted function where the error cost of datapoints is not counted. Hyperparameters used for support vector models and US measurements are reviewed in Table 5. Decision trees can require a choice in the number of trees used (500 [16], maximum tree depth (1 [16], 3 [44], 4 [16], the minimum number of leaf instances (10 [16]), the learning rate (1 [16]), or the maximum number of splits (4, 20, 100 [216]). The type of ensemble method can also be chosen, such as Adaboost used in [16]. Furthermore, K-nearest neighbour algorithms require a selection of the number of neighbours to use, such as 5 [216], 11 [217], 25 [44], 50 [44], or 105 [16].

Gaussian Processes (GP) are an algorithm for classification or regression which interpolates datapoints with normal distributions and thereby provides confidence intervals for its predictions. GPs have recently been applied to measure oil film thickness in journal bearings using US measurements [230] and to localise acoustic emission sources in SHM by measuring the time of flight with multiple sensor pairs [231]. Ref. [232] used data-driven GPs to model guided waves in composite materials. Physical knowledge of the system was inputted into the model by specifying constraints of the GP's kernels such as rotational symmetry, exponential decay for viscoelastic damping, and attenuation due to geometric spreading using a polynomial kernel. Ref. [233] used Bayesian linear regression to decompose guided wave signals into individual modes to enable damage sizing and localisation following two-dimensional Fourier transformation. Ref. [234] used a GP to predict thermal barrier coating porosity. Input features to the GP were first selected by evaluating a neural network using different feature combinations.

4.4. Out-of-distribution detection

Out-of-distribution (OOD) detection methods are used to identify datapoints that fall outside the range of normally expected values. For example, [235] used OOD detection to identify damage in wind turbine blades. During feature extraction, DWT, Fast Fourier Transform and PCA were used. One-class classification algorithms were used including support vector machine data description, K-means, and Euclidean distance measures. Autoencoders are commonly used for OOD detection. If anomalous data is passed through as encoder, the distribution of the latent space variables will be different to the training data and the

autoencoder fails to reconstruct the input. Ref. [167] used autoencoders to allows flaws to be identified even if they overlap the initial transducer pulse. Ref. [236] used autoencoders in a similar network structure for a Ganomaly approach. Furthermore, they used another generative model, normalizing flows, to learn transformations between normal and anomalous samples for OOD detection. Ref. [237] used variational AEs as an OOD detection system to detect defects from ultrasonic B-scans. In this study, a second encoder was added after the decoder and was found to provide increased accuracy in detecting defects. Ref. [238] compared autoencoders to one-class SVMs, isolation forests, and hidden Markov models.

4.5. Data augmentation

Data augmentation can be used to artificially increase the size of a dataset. This can be particularly useful when training deep learning models (e.g. DNNs, CNNs, and LSTMs) which can require many training instances to tune all the model parameters. Ref. [35] time-shifted US signals forwards and backwards by 5 and 10 μs forward to increase the dataset size by four times. Similarly, [162] time-shifted US signals forwards and backwards by 6×10^{-3}, 10×10^{-3}, 14×10^{-3}, and 20×10^{-3} μs. Ref. [239] laterally translated signals and used magnification to increase the dataset size by five times. Ref. [205] added six different levels of Gaussian white noise to US signals ranging from 20 to 30 dB. Ref. [240] rotated samples by 90°, 180°, and 270° for damage localisation in plate-like structures. Ref. [204] extracted parts of US waveforms received from measuring flawed samples and inserted them into flawless signals. In this way, virtual data could be created by implanting flaws into different locations of a test section. The choice of data augmentation techniques, such as lateral translation, magnification, or noise addition, must be decided based on the application it is being used for. For example, lateral translation could not be used if the time of flight is an important parameter to measure during the process being monitored. Refs. [35,162] were classifying weldment defects. As such, as shift in the time domain would only represent a change in depth of the flaw rather than its presence or type. Ref. [239] used data augmentation for US waveform feature learning in a CNN. The CNN was trained on an auxiliary task to classify the dataset membership of previously collected US measurements. The pretrained CNN weights were then used as a feature extractor on new US measurement datasets. Therefore, the lateral translation and magnification did not represent a change in any physical parameters of a system. Ref. [241] doubled the size of an experimental dataset by reversing US images of defects around the vertical axis. Ref. [207] used time shifting and the addition of white Gaussian noise for US flaw classification in weldments. Ref. [202] added white Gaussian noise to create three datasets with signal-to-noise ratios of 5, 10, and 20.

4.6. Semi-supervised learning

[242] used a hierarchical clustering algorithm to detect whether pipe sections were damaged or undamaged. This is traditionally an unsupervised learning method which divides the input data into the number of clusters specified (in this case, two). To label each of the clusters, only one labelled instance of an undamaged pipe was required to perform the classification. Similarly, [243] used a k-means clustering technique to monitor the growth of simulated cracks in pipes. An alarm threshold was developed to trigger when the size of the defect becomes critical based on the distance of the US measurement from each cluster. Ref. [244] presented a semi-supervised Gaussian mixture model which was updated through the expected maximisation algorithm over both the labelled and unlabelled data.

4.7. Active learning

Active learning uses methods to select unlabelled datapoints that would have the most benefit to model performance if labelled thereby minimising the total number of datapoints to be labelled. Ref. [245] used active learning to improve a probabilistic mixture model initially trained on a small sample of labelled data. This approach was evaluated on three datasets: Z24 Bridge data, a machining acoustic emission dataset, and data from ground vibration aircraft tests. Ref. [246] used the expected value of perfect information for SHM on a numerical case study and the Z24 Bridge benchmark.

4.8. Generative models

Generative Adversarial Networks (GAN) and Variational Autoencoders (VAE) are methods to produce realistic data from random inputs. The generator component of a GAN is trained by aiming to fool the discriminator component into determining whether its input data are real or synthetic. During training, a VAE learns a probability distribution of the input data in the latent space from which new samples can be drawn. Ref. [247] used a GAN for generating defects in US B-scan images and successfully increased defect detection from 70 % to 76 % when combining real and synthetic images. Ref. [248] compared two GAN structures to produce images of defects in US signals and confirmed that the generated images could not be identified by human experts. Ref. [249] used GANs to increase the size of Finite Element simulation datasets for welding defect detection. The highest defect detection accuracy was achieved by supplementing the generated data with noise derived from experiments and extracted using the sliding kernel approach. Ref. [250] also used GANs to generate B-scans from simulated US data for non-destructive evaluation applications. Ref. [251] used GANs to create synthetic acoustic emission spectrograms in the US range for detecting cavitation in hydraulic turbines. This increased cavitation detection accuracy from 94.2% using CNNs alone to 95.1%.

4.9. Transfer learning

Transfer learning encompasses methods which transfer knowledge learned from one task to different, but similar, tasks. Ref. [241] trained a Faster-CNN to detect defects in US images on simulated datasets first before training on a small set of experimental data. This greatly reduced the loss function compared with training on the experimental datasets alone. Ref. [252] presented an experimental dataset of 7004 ultrasonic images collected from 18 stainless steel plates and evaluate the performance of many pre-trained CNNs. They conclude that their dataset may be used by others for pre-training their own CNN models. Ref. [253] transferred fixed layers of a neural network trained on easier-to-classify tasks to more difficult cases for damage localisation on an aircraft wing. Ref. [254,255] used a pre-trained VGG16 to classify acoustic emission sources following CWT by training the last convolutional layer, two fully connected layers, and output layer, and the output layer only, respectively. Similarly, [256] updated the last layer of a pre-trained ResNet34 model for vibration data for SHM. Ref. [257] pre-trained a CNN on compressed vibration data in the form of a histogram of response thresholds for an SHM application. This CNN was then fine-tuned using extremely compressed, smoothed histogram data in the form of a mean, variance, and scale factor.

4.9.1. Domain adaptation

An ML model trained on one task (source domain) will predict poorly on a second task (target domain) if the feature distributions between the domains change. US waveform features may be different across domains due to differences in the path of the sound wave or the materials being monitored. Even for similar processes, differences in the sensor used, attachment procedure, or contact pressure may alter the feature distributions [20]. Domain adaptation is a subcategory of transfer learning which alters how a ML model trains so that it predicts accurately across both domains. Ref. [20] used unlabelled domain adaptation of a single-feature waveform feature to apply a trained ML model to new, similar

Table 4
A review of hyperparameters used for CNNs with US measurements.

Reference	Convolutional layers, number of filters, size of filters	Pooling layers	Fully connected layers	Training	Additional information
[134]	2 16 7 × 5 32 5 × 3	2 × 2 max pooling	128 neurons and an SVM final layer		
[135]	3 5 5 × 1 8 8 × 1 16 7 × 1	Max pooling: 2 × 1, stride 1 × 1	2 150 and 75 neurons	Adam optimisation algorithm Learning rate varied between 10^{-5} and 1 Trained for 100 epochs	ReLu activation function
[135]	2 6 10 × 2 16 10 × 2	2 × 2 max pooling	2 500 and 250 neurons	Adam optimisation algorithm Learning rate varied between 10^{-5} and 1 100 epochs	
[191]	2 16 10 × 1 32 10 × 1		2 200, 200		ReLu activation function
[192]	3 32 3 × 3 32 3 × 3 64 3 × 3	2 × 2 max pooling	64 neurons	Batch size of 8 3 epochs	Three dropout layers with 0.5 probability ReLu activation function
[188]	6 layers Combined with GRUs			Batch size of 512 300 epochs Adam optimisation function Initial learning rate of 0.001 Learning rate reduced every 50 epochs Final learning rate of 0.0001	ReLu activation function Batch normalisation used Dropout rate of 0.5
[35]	2 32 16 × 1 filter size, 8 × 1 stride 64 3 × 1 filter size, 2 × 1 stride	No pooling layer between 1st and 2nd convolutional layers Max pooling, 2 × 1 size with 2 × 1 stride	300 neurons	Trained for 500 epochs	Padding: "Same" for convolutional layers, "valid" for max pooling layer Activation function: Elu in convolutional layers, ReLu in fully connected layers Two dropout layers, 0.25 and 0.5 probability
[27]	2 8 5 × 5 16 5 × 5	2 × 2 max pooling		Adam optimisation algorithm Learning rate 0.01, drop factor of 0.33 after 4 epochs Trained for 8 epochs in total Batch size of 256, shuffled every epoch	"Same" padding Batch normalisation ReLu activation function Single dropout layer, probability varied between 0.1 and 0.5
[162]	2 32 25 × 1, 8 × 1 stride 64 3 × 1, 2 × 1 stride	No max pooling between 1st and 2nd convolutional layers Max pooling: 2 × 1 size, 2 × 1 stride			Three dropout layers with 0.7, 0.5 and 0.5 probability, respectively
[136]	2 16 3 × 3 512 3 × 3	2 × 2 max pooling	SVM final layer	Learning rate of 0.1	Batch normalisation ReLu activation function Single dropout layer with 0.5 probability
[193]	7 Structure based on VGGNet	3 max pooling layers			
[194]	3 32 5 × 1 64 5 × 1 96 5 × 1	Max pooling: 2 × 1, stride of 2	1000		
[195]	2 30 3 × 3 in both	Max pooling: 3 × 3, stride of 2	56, 28	Learning rate of $1e^{-6}$ Trained for 5000 iterations	ReLu activation function
[196]	21 convolutional layers across three channels Filters ranging in size from 1 × 1 to 12 × 20			Adam optimisation algorithm Trained for 6 epochs Learning rate of 0.001 Batch size of 32	Batch normalisation Leaky ReLu activation function
[197]	4 32 3 × 3 32 3 × 3 64 3 × 3 1 1 × 1	1 × 2 max pooling			Batch normalisation ReLu activation function
[198]	3 48 3 × 3 96 3 × 3 192 3 × 3	2 × 2 max pooling	64	Adam optimisation algorithm Learning rate of 0.001 Batch size of 128	10 % dropout rate before and after fully connected layer ReLu activation function

(*continued on next page*)

Table 4 (*continued*)

Reference	Convolutional layers, number of filters, size of filters	Pooling layers	Fully connected layers	Training	Additional information
				400 epochs Validation patience of 150 epochs	
[199]	1 64 filters			Stochastic gradient descent optimisation algorithm used Initial learning rate of 0.005 and decreased logarithmically 30,000 epochs	Dropout rate of 0.5
[200]	1 8 5 × 5	3 × 3 max pooling	2 1024 and 64 neurons	Adam optimisation algorithm	
[201]	5 Inception blocks with 1 × 1, 3 × 3, and 5x5 filters	3 × 3 max pooling Stride of 2	3	Momentum of 0.9 Weight decay of 0.0002 Stochastic gradient descent optimisation algorithm Initial learning rate of 0.001 160 epochs	Batch normalisation and ReLu activation function before each convolutional layer Dropout ratios of 0.5
[147]	4 ResBlocks			Adam optimisation algorithm Learning rate of 0.0001 20 epochs Batch size of 30	
[202]	3 convolutional blocks			Adam optimisation algorithm Learning rate of 0.01 Batch size of 4 500 epochs	
[137]	3 16 3 × 1 32 3 × 1 64 3 × 1	2x1 max pooling	16 neurons	Learning rate of $1e^{-4}$ Adam optimisation algorithm Trained for 250 epochs Batch size of 64	ReLu activation function
[137]	5 16 3 × 1 32 3 × 1 64 3 × 1 128 3 × 1 256 3 × 1	2x1 max pooling	2 512 and 128 neurons	Learning rate of $1e^{-5}$ Adam optimisation algorithm Trained for 500 epochs Batch size of 32	ReLu activation function Two dropout layers with probabilities of 0.25 and 0.2, respectively
[203]	5 16 3 × 1 32 3 × 1 64 3 × 1 128 3 × 1 256 3 × 1	2x1 max pooling	128 neurons	Adam optimisation algorithm Learning rate of $1e^{-5}$ Trained for 2500 epochs Batch size of 128	ReLu activation function Single dropout layer with 0.25 probability
[203]	5 16 3 × 1 32 3 × 1 64 3 × 1 128 3 × 1 256 3 × 1	2 × 1 max pooling	128 neurons	Adam optimisation algorithm Learning rate of $1e^{-5}$ 500 epochs Batch size of 32	ReLu activation function Single dropout layer with 0.25 probability
[204]	4 96 3 × 3 64 3 × 3 48 3 × 3 32 3 × 3	Max pooling with varying sizes of 7 × 1, 2x8, 3 × 4	14 neurons		ReLu activation function
[184]	32 50 × 1, stride of 5 × 1 64 4 × 1, stride of 2 × 1	2 × 1 max pooling, stride of 2 × 1	1000 neurons	Learning rate of 0.001	2 dropout layers, before and after full-connected layer, probabilities of 0.7 ReLu activation function
[205]	Dual-headed convolutional neural network 4 convolutional layers in each head 576 11 × 1 484 11 × 1 500 5 × 1 324 5 × 1	4 × 1 max pooling	1 fully connected layer in each head, 256 neurons in each Final fully connected layer with 196 neurons	Batch size of 24 Trained for 70 epochs	Two dropout layers in each head, 0.2 and 0.3 probabilities ReLu activation function L2 regularisation value of 0.01 Early stopping with patience of 10 epochs
[130]	1 64 3 × 3	2 × 2 max pooling	512 neurons	Adam optimisation algorithm Learning rate of 0.003 Trained for 30 epochs Batch size of 32 Momentum of 0.9	ReLu activation function
[206]	CNN base on WaveNet Dilation rate of 3 × 1 filters increased to 3, 9, 27, 81 through			Trained for 900 epochs	"Casual", "valid", and "same" padding used

(*continued on next page*)

Table 4 (*continued*)

Reference	Convolutional layers, number of filters, size of filters	Pooling layers	Fully connected layers	Training	Additional information
	four residual blocks 16 filters in each convolutional layer				
[207]	19-layer ResNet	1 × 2 max pooling	30o neurons	Adam optimisation algorithm 500 epochs	ReLu activation function Dropout probability of 0.5

Table 5
Hyperparameters used with support vector models and US measurements.

References	Information
[218]	Radial basis function (RBF) kernel C varied between 1 and 32 γ varied between 0.00049 and 0.5
[29]	C = 1000 γ = 10 Gaussian kernel Kernel Fisher discriminant used to optimise parameters
[219]	Linear kernel with C = 10 2nd order polynomial kernel with C = 0.1 RBF kernel with C = 10 and γ = 1
[220]	RBF and linear kernels γ = 20
[221]	Linear kernel, polynomial kernel, RBF kernel, and sigmoid kernel Range of C values tested: 1, 10^1, 10^2, 10^3, 10^4, 10^5
[42]	Third order polynomial kernels used
[43]	RBF kernel γ = 0.7
[222]	Epsilon, C, and γ determined through particle swarm analysis
[223]	Linear, polynomial, and RBF kernels used C varied between 0.001 and 100 in increasing powers of 10 Polynomial degree evaluated between 2 and 5 γ for RBF kernel evaluated between 0.01 and 100 in increasing powers of 10
[224]	Linear, quadratic, RBF, and polynomial kernels tested γ values tested: 4, 5, 6, 7 Polynomial orders tested: 2, 3, 4
[177]	Linear, quadratic, polynomial, multilayer perceptron and RBF kernels tested A 3rd degree polynomial kernel function was used
[225]	C = 2 γ = 10,000 RBF kernel
[129]	Binary tree SVM C and γ varied
[180]	Bias = 42.57 Box constraint = 9.4885 Epsilon = 0.9489 Number of iterations = 64 RBF kernel
[141]	RBF kernel C and γ parameters determined using particle swarm optimisation
[216]	Linear, polynomial and RBF kernel tested. Linear kernel performed best. The best box constraint was found to be 1.
[181]	Linear and RBF kernels investigated
[226]	RBF kernel C and γ optimised using Particle Swarm Optimisation
[227]	RBF kernel C = 0.1 γ = 0.3
[16]	C = 0.0001
[44]	C = 0.001
[27]	Bayesian optimisation for 60 evaluations to select box constraint value, kernel scale, kernel function, polynomial order, and whether the inputs were standardised The expected improvement acquisition function was used
[228]	C and γ optimised using Bat Optimisation Algorithm
[229]	Cross-validation to determine C and γ RBF kernel

processes (investigated for two mixing and three cleaning processes) without needing to label data in the new domain. This is therefore a method of eliminating the data labelling burden in a factory environment when applying a US sensor to a new, similar process. Combining multiple source datasets or using datasets collected from a wider range of process parameters (e.g. temperature variations) enabled the models to better adapt to changes in feature distributions. Ref. [258] used labelled domain adaptation, where a reference measurement is available in the target domain, to use previously collected data from a laboratory fermentation process to reduce the development time of accurate ML models for an industrial process. Three methods were investigated to train DNNs with LSTM layers: simultaneous training over both datasets, federated learning using both datasets, and fine-tuning of the previously trained models on the target domain dataset. Federated learning was investigated as a method for potentially using datasets from multiple companies while maintaining data privacy. Federated learning shares network weights between a local models from each dataset and a global model rather than sharing the real data. All methodologies provided an increase in prediction accuracy over solely using the industrial fermentation dataset. Federated learning provided the highest increase in accuracy by allowing further convergence to minima during training. Ref. [259] transferred knowledge from microseismic data of earthquake studies to acoustic logging tools for collecting borehole information in oil fields. The purpose of this study was to overcome the effects of noise on accurate determination of the time-of-flight using pulse-echo transducers. The maximum mean discrepancy was used to align feature spaces and a convolutional autoencoder was used to ensure class discriminant features were extracted. Ref. [260] used a dictionary learning method to use simulated US wavefields to isolate damage wavefields in experimental data from plate structures. Firstly, a dictionary is learned to optimally reconstruct the simulated wavefields and then transferred to the experimental data. Secondly, the reconstructed experimental data is aligned with the experimental data to account for changes in domain between the simulated and experimental datasets.

Ref. [261] used Transfer Component Analysis (TCA) to transfer damage detectors between experimental datasets from three aircraft tailplanes. The data from both domains were matched in a latent space and two Transfer Components were extracted. Ref. [262] used metric-informed joint domain adaptation to overcome the problem of pre- and post-repair changes in data distribution in SHM. Joint domain adaptation aligns both the marginal and conditional distributions in a latent space using pseudo-labels from the target domain. In this work, the Mahalanobis squared distance was used to select the data for pseudo-labelling. Ref. [263] used a domain-adapted Gaussian mixture model (DA-GMM) to transfer labelled information between two bridge datasets for SHM. A linear mapping was used to transform the target domain data and the model was optimised using an expectation maximisation technique. Ref. [264] presented a kernelised Bayesian transfer learning (KBTL) approach for SHM applications. The approach maps data from each domain onto a shared latent space where labelled data from the source domain is used to classify the data from the target domain. This method may be used to transfer labelled data from uncommon damage types from similar structures or simulations to reduce the burden of labelling these rare states in the target domain. Ref. [265] used Balanced

Table 6
An explanation of the recommendations presented in Table 7.

Recommendations	Explanation
Other sensors	Other in-line or on-line sensors as a reference measurement
Semi-supervised	Semi-supervised learning to pseudo-label the unlabelled data
Sampling	Sampling and off-line sensors as a reference measurement
US measurements	US measurements to infer process state
Unlabelled domain adaptation	Unlabelled domain adaptation from similar processes
Labelled domain adaptation	Labelled domain adaptation from similar processes
Coarse time dome features	E.g., specific amplitudes in a function, dispersion of amplitudes along the function, measures of the rise or descent of function amplitudes, energy, crest factor, kurtosis, skewness, temporal duration
Convolutional feature extraction	Convolutional feature extraction methodology as presented in [239]
Feature variations	Variations in features between consecutive waveforms

Distribution Adaptation to transfer damage localisation models between different types of aircraft wings. Three metrics were evaluated to determine the structural and data similarities between domains before domain adaptation. This was to minimise the risk of negative transfer where transfer learning leads to a decrease in model performance.

4.10. Object detection

Object detection is a type of image analysis where a CNN identifies the presence, location, and class of objects and has been used for defect detection in ultrasonic images. Ref. [241] used a Faster R-CNN to identify, locate, and size defects using simulated data and a small experimental dataset. Ref. [266] presented EfficientDet and a method to select anchor hyperparameters for the high aspect ratios expected. They achieved a 9 % increase in accuracy (89.6% accuracy overall) compared to the YOLOv3 architecture used in [267]. Ref. [268] presented DefectDet which uses a lightweight encoder-decoder feature extractor and a detection head with custom anchor box aspect ratio and stride to detect high aspect ratio defects. Ref. [269] presented two methods of using multiple ultrasonic B-scans for defect detection by merging feature maps using convolutional and convolutional LSTM layers.

5. Future directions - interpretability in machine learning

A particular barrier to implementation of sensor and ML combinations is the perceived lack of explainability and interpretability of ML models and therefore a lack of trust it their predictions. These issues must be addressed to increase buy-in from companies and operators, and, furthermore, to meet any potential regulation criteria that requires adequate transparency in the ML prediction process. There are three points in the supervised ML pipeline where these problems can be considered. The first is during feature extraction and selection. Inherently explainable features such as the waveform energy (typically a measure of the reflection coefficient, and therefore acoustic impedance, or attenuation in a system) or the time of flight (a measure of the speed of sound through a material) may be preferred over other, more abstract features acquired from a US waveform. Feature selection methods may be used to reduce the number of features to make models simpler or calculate the importance of each feature and thereby make the models more interpretable. Secondly, transparent algorithms could be used such as linear or logistic regression, decision trees, or k-nearest neighbour models [270]. Predictions from linear and logistic regression could be accompanied by the weights applied to each feature used to make the prediction. In this way, the prediction process can be made fully explainable. Similarly, decision trees could produce the hierarchical decision process used and k-nearest neighbour models could present the k nearest training points used to inform the prediction made. Finally, post-hoc explanation of individual predictions can be used to understand the decision-making process [271]. Local interpretable model-agnostic explanations (LIME) perturb training data around a particular query point and build a transparent model (e.g. decision tree) correlating the new synthetic training data with model predictions to understand the decision-making process around this particular point. Shapley Additive explanations (SHAP) calculate the change in predictions by varying feature values at a particular data point to understand the impact of each feature on the prediction being made. For CNNs, additional techniques such as, gradients, class activation mappings, saliency maps, or occlusions can be used to indicate the datapoints contributing to a particular prediction.

6. Recommendations

Tables 6 and 7 contain recommendations for combining ML and US sensors for the reviewed process applications. To obtain labelled data in industrial environments, other in-line or on-line sensors can be used as a reference measurement for all the reviewed processes. Periodic sampling combined with off-line analysis could be employed for all processes other than cleaning, curing, or membrane fouling. Instead, for cleaning, the process would need to be ended at different stages, the equipment dismantled, and the sensor data labelled. During curing, sampling would not be possible due to the toughening or hardening of the material. For membrane fouling, no non-disruptive sampling of the fouling material could be performed. As such, semi-supervised learning could be used in all processes to pseudo-label the unlabelled data except for curing or membrane fouling processes where single data points cannot be collected. For all the processes reviewed, the US measurements could be used to infer process stages (such as the attainment of homogeneity during mixing, or the start of ethanol production during fermentation), apart from for tabletting. This is because during tabletting the final product is monitored opposed to the compaction process. For every process, unlabelled and labelled domain adaptation from similar processes could be used to reduce the data labelling burden in a factory environment. The recommended features include the coarse time domain features and convolutional feature extraction methods as compared in [239]. Fig. 3 displays the convolutional feature extraction method as an unrolled DNN with an LSTM presented in [239]. The coarse time domain features include specific amplitudes in the function, dispersion of amplitudes along the function, measures of the rise or descent of function amplitudes, energy, crest factor, kurtosis, skewness, and temporal duration. These features do not misattribute waveform variations of narrow frequency band US sensors to changes in frequency content as do the Fourier or wavelet transformations [239]. These features also overcome the problem of lateral sample point shifting of waveforms due to temperature changes as suffered by using the amplitudes at sample points as features directly [239]. The time of flight can be used for all processes to monitor the speed of sound throughout the material. The variations between consecutively acquired US waveforms can be used to monitor product quality variation in extrusion and injection moulding, the degree of homogeneity in mixing processes, or the presence of CO_2 bubbles during fermentation. Feature gradients along with LSTMs can be used for all processes that progress over time to

Table 7
Recommendations for combining ML and US measurements with the reviewed processes.

	Cleaning	Fermentation	Crystallisation	Mixing	Extrusion	Injection moulding	Curing	Reaction monitoring	Tabletting	Membrane fouling
Obtaining labelled data	Other sensors Ending the process at different stages and using off-line sensors as a reference measurement Semi-supervised learning	Other sensors Sampling Semi-supervised learning	Other sensors Sampling Semi-supervised learning	Other sensors Sampling Semi-supervised learning	Other sensors Sampling Semi-supervised learning	Other sensors Sampling Semi-supervised learning	Other sensors	Other sensors Sampling Semi-supervised learning	Other sensors Sampling Semi-supervised learning	Other sensors
Reducing data labelling burden	US measurements Unlabelled domain adaptation Labelled domain adaptation	US measurements Unlabelled domain adaptation Labelled domain adaptation	US measurements Unlabelled domain adaptation Labelled domain adaptation	US measurements Unlabelled domain adaptation Labelled domain adaptation	US measurements Unlabelled domain adaptation Labelled domain adaptation	US measurements Unlabelled domain adaptation Labelled domain adaptation	US measurements Unlabelled domain adaptation Labelled domain adaptation	US measurements Unlabelled domain adaptation Labelled domain adaptation	Unlabelled domain adaptation Labelled domain adaptation	US measurements Unlabelled domain adaptation Labelled domain adaptation
Feature extraction	Time of flight Coarse time domain features Feature gradients Other features available (e.g. temperature, material type, mass flow rate) Convolutional feature extraction	Time of flight Coarse time domain features Feature gradients Other features available (e.g. temperature, material type) Convolutional feature extraction Feature variations	Time of flight Coarse time domain features Feature gradients Other features available (e.g. temperature, material type) Convolutional feature extraction	Time of flight Coarse time domain features Feature gradients Other features available (e.g. temperature, material type) Convolutional feature extraction Feature variations	Time of flight Coarse time domain features Other features available (e.g. temperature, material type, pressure, mass flow rate) Convolutional feature extraction Feature variations	Time of flight Coarse time domain features Other features available (e.g. temperature, material type, pressure, mass flow rate) Convolutional feature extraction Feature variations	Time of flight Coarse time domain features Feature gradients Other features available (e.g. temperature, material type) Convolutional feature extraction	Time of flight Coarse time domain features Feature gradients Other features available (e.g. temperature, material type) Convolutional feature extraction	Time of flight Coarse time domain features Other features available (e.g. temperature, material type) Convolutional feature extraction	Time of flight Coarse time domain features Feature gradients Other features available (e.g. temperature, mass flow rate, material type) Convolutional feature extraction
Algorithms	LSTMs	LSTMs	LSTMs	LSTMs	LSTMs for short time sequences of features ANNs or DNNs, using US waveforms	LSTMs for short time sequences of features ANNs or DNNs, using US waveforms	LSTMs	LSTMs	ANNs or DNNs	LSTMs

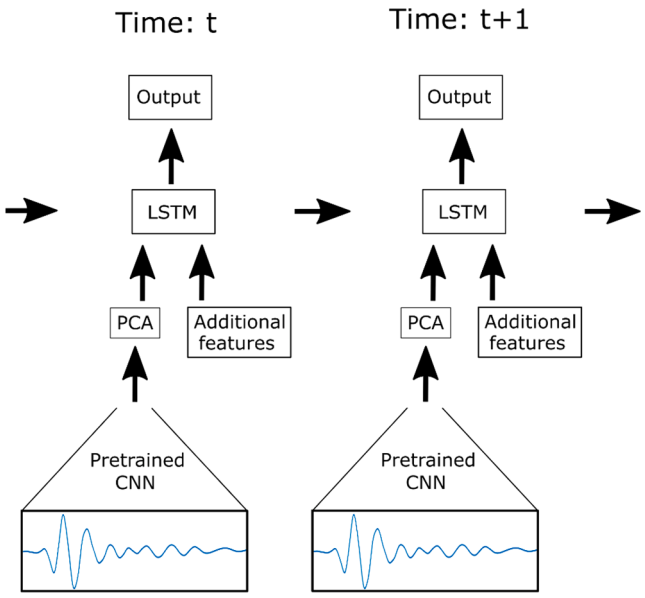


Fig. 3. A diagram of the convolutional feature extraction method presented in [239]. It is depicted as an unrolled DNN with an LSTM layer. A pretrained CNN extracts features from the US waveform which are then combined with additional features, such as the time of flight or deviations between consecutively acquired waveforms and inputted into the DNN.

incorporate knowledge from previous time steps. LSTMs can be used on short time sequences of features for extrusion and injection moulding to monitor variations in product qualities, or ANNs and DNNs can be used on single time-step features. ANNs or DNNs should be used for data that is not part of a time sequence due to their ability to construct new features from input data that correlate to output variables. This ability for complex fitting requires adequate regularisation to prevent over-fitting. ANNs and DNNs should also be used over CNNs as they allow for incorporation of other features such as the time of flight, mass flow rate, temperature, pressure, material type, or variations between consecutively acquired waveform features.

7. Summary

The manufacturing sector is increasingly using data to inform decision making. In-line and on-line sensors underpin this transition by automatically acquiring real-time data. Supervised ML techniques can be combined with US measurements and provide advantages over calibration procedures. However, their implementation is lagging due to expertise required to extract and select appropriate features from the sensor measurements, select the ML algorithm to use, and find a suitable set of model hyperparameters. The aim of this article is to facilitate the combination of ML and US measurements for in-line and on-line process monitoring or other similar applications. The article first reviews the use of US sensors for monitoring processes before reviewing the combination of US measurements and ML including literature from other sectors. This review covers feature extraction, feature selection, algorithm choice, hyperparameter selection, data augmentation, domain adaptation, semi-supervised learning and ML interpretability. Recommendations for applying ML methods for monitoring of the reviewed processes are also provided.

Funding: This work was supported by the Engineering and Physical Sciences Research Council (EPSRC) standard research studentship (EP/R513283/1).

CRediT authorship contribution statement

Alexander L. Bowler: Conceptualization, Data curation, Formal analysis, Investigation, Methodology, Software, Validation, Visualization, Writing – original draft, Writing – review & editing. **Michael P. Pound:** Conceptualization, Methodology, Software, Supervision, Visualization, Writing – review & editing. **Nicholas J. Watson:** Conceptualization, Funding acquisition, Methodology, Project administration, Resources, Supervision, Visualization, Writing – review & editing.

Declaration of Competing Interest

The authors declare that they have no known competing financial interests or personal relationships that could have appeared to influence the work reported in this paper.

References

[1] M. Ghobakhloo, Industry 4.0, digitization, and opportunities for sustainability, J. Clean. Prod. 252 (2020) 119869, https://doi.org/10.1016/j.jclepro.2019.119869.
[2] H.S. Kang, J.Y. Lee, SangSu Choi, H. Kim, J.H. Park, J.Y. Son, B.H. Kim, S.D. Noh, Smart manufacturing: Past research, present findings, and future directions, Int. J. Pr. Eng. Man.-GT. 3 (1) (2016) 111–128, https://doi.org/10.1007/s40684-016-0015-5.
[3] T. De Beer, A. Burggraeve, M. Fonteyne, L. Saerens, J.P. Remon, C. Vervaet, Near infrared and Raman spectroscopy for the in-process monitoring of pharmaceutical production processes, Int. J. Pharm. 417 (1-2) (2011) 32–47, https://doi.org/10.1016/j.ijpharm.2010.12.012.
[4] D.R. Sjödin, V. Parida, M. Leksell, A. Petrovic, Smart factory implementation and process innovation: a preliminary maturity model for leveraging digitalization in manufacturing moving to smart factories presents specific challenges that can be addressed through a structured approach focused on people, processes, and technologies, Res. Technol. Manag. 61 (5) (2018) 22–31, https://doi.org/10.1080/08956308.2018.1471277.
[5] J. Chen, X. Ran, Deep learning with edge computing: a review, Proc. IEEE 107 (8) (2019) 1655–1674.
[6] R.Y. Zhong, X. Xu, E. Klotz, S.T. Newman, Intelligent manufacturing in the context of industry 4.0: a review, Engineering 3 (5) (2017) 616–630, https://doi.org/10.1016/J.ENG.2017.05.015.
[7] O. Fisher, N. Watson, L. Porcu, D. Bacon, M. Rigley, R.L. Gomes, Cloud manufacturing as a sustainable process manufacturing route, J. Manuf. Syst. 47 (2018) 53–68, https://doi.org/10.1016/j.jmsy.2018.03.005.
[8] L.L. Simon, H. Pataki, G. Marosi, F. Meemken, K. Hungerbühler, A. Baiker, S. Tummala, B. Glennon, M. Kuentz, G. Steele, H.J.M. Kramer, J.W. Rydzak, Z. Chen, J. Morris, F. Kjell, R. Singh, R. Gani, K.V. Gernaey, M. Louhi-Kultanen, J. O'Reilly, N. Sandler, O. Antikainen, J. Yliruusi, P. Frohberg, J. Ulrich, R. D. Braatz, T. Leyssens, M. von Stosch, R. Oliveira, R.B.H. Tan, H. Wu, M. Khan, D. O'Grady, A. Pandey, R. Westra, E. Delle-Case, D. Pape, D. Angelosante, Y. Maret, O. Steiger, M. Lenner, K. Abbou-Oucherif, Z.K. Nagy, J.D. Litster, V. K. Kamaraju, M.-S. Chiu, Assessment of recent process analytical technology (PAT) trends: a multiauthor review, Org. Process Res. Dev. 19 (1) (2015) 3–62, https://doi.org/10.1021/op500261y.
[9] A. Gowen, C. Odonnell, P. Cullen, G. Downey, J. Frias, Hyperspectral imaging – an emerging process analytical tool for food quality and safety control, Trends Food Sci. Tech. 18 (12) (2007) 590–598, https://doi.org/10.1016/j.tifs.2007.06.001.
[10] A.L. Bowler, S. Bakalis, N.J. Watson, A review of in-line and on-line measurement techniques to monitor industrial mixing processes, Chem. Eng. Res. Des. 153 (2020) 463–495, https://doi.org/10.1016/j.cherd.2019.10.045.
[11] B. Henning, J. Rautenberg, Process monitoring using ultrasonic sensor systems, Ultrasonics 44 (2006) e1395–e1399, https://doi.org/10.1016/j.ultras.2006.05.048.
[12] D.J. McClements, Advances in the application of ultrasound in food analysis and processing, Trends Food Sci. Tech. 6 (9) (1995) 293–299, https://doi.org/10.1016/S0924-2244(00)89139-6.
[13] T.S. Awad, H.A. Moharram, O.E. Shaltout, D. Asker, M.M. Youssef, Applications of ultrasound in analysis, processing and quality control of food: a review, Food Res. Int. 48 (2) (2012) 410–427, https://doi.org/10.1016/j.foodres.2012.05.004.
[14] R.E. Challis, M.J.W. Povey, M.L. Mather, A.K. Holmes, Ultrasound techniques for characterizing colloidal dispersions, Rep. Prog. Phys. 68 (7) (2005) 1541–1637.
[15] A. Simeone, E. Woolley, J. Escrig, N.J. Watson, Intelligent industrial cleaning: a multi-sensor approach utilising machine learning-based regression, Sensors 20 (2020) 1–22, https://doi.org/10.3390/s20133642.
[16] J.E. Escrig, A. Simeone, E. Woolley, S. Rangappa, A. Rady, N.J. Watson, Ultrasonic measurements and machine learning for monitoring the removal of surface fouling during clean-in-place processes, Food Bioprod. Process. 123 (2020) 1–13, https://doi.org/10.1016/j.fbp.2020.05.003.
[17] P. Resa, L. Elvira, F. Montero de Espinosa, R. González, J. Barcenilla, On-line ultrasonic velocity monitoring of alcoholic fermentation kinetics, Bioproc. Biosyst. Eng. 32 (3) (2009) 321–331, https://doi.org/10.1007/s00449-008-0251-3.
[18] Z. Ge, Z. Song, S.X. Ding, B. Huang, Data mining and analytics in the process industry: the role of machine learning, IEEE Access 5 (2017) 20590–20616, https://doi.org/10.1109/ACCESS.2017.2756872.

[19] J.E. van Engelen, H.H. Hoos, A survey on semi-supervised learning, Mach. Learn. 109 (2) (2020) 373–440, https://doi.org/10.1007/s10994-019-05855-6.
[20] A.L. Bowler, N.J. Watson, Transfer learning for process monitoring using reflection-mode ultrasonic sensing, Ultrasonics 115 (2021) 106468, https://doi.org/10.1016/j.ultras.2021.106468.
[21] A. Joby, What Is Cross-Validation? Comparing Machine Learning Models, 2021. <https://learn.g2.com/cross-validation> (Accessed 17/11/2021).
[22] O.J. Fisher, N.J. Watson, J.E. Escrig, R. Witt, L. Porcu, D. Bacon, M. Rigley, R. L. Gomes, Considerations, challenges and opportunities when developing data-driven models for process manufacturing systems, Comput. Chem. Eng. 140 (2020) 106881, https://doi.org/10.1016/j.compchemeng.2020.106881.
[23] G. Chandrashekar, F. Sahin, A survey on feature selection methods, Comput. Electr. Eng. 40 (1) (2014) 16–28, https://doi.org/10.1016/j.compeleceng.2013.11.024.
[24] W.B. Hussein, M.A. Hussein, T. Becker, Robust spectral estimation for speed of sound with phase shift correction applied online in yeast fermentation processes, Eng. Life Sci. 12 (6) (2012) 603–614, https://doi.org/10.1002/elsc.201100183.
[25] F.C. Cruz, E.F. Simas Filho, M.C.S. Albuquerque, I.C. Silva, C.T.T. Farias, L. L. Gouvêa, Efficient feature selection for neural network based detection of flaws in steel welded joints using ultrasound testing, Ultrasonics 73 (2017) 1–8, https://doi.org/10.1016/j.ultras.2016.08.017.
[26] S.G. Mallat, A theory for multiresolution signal decomposition: the wavelet representation, IEEE T. Pattern Anal. 11 (1989) 674–693, https://doi.org/10.1109/34.192463.
[27] A.L. Bowler, S. Bakalis, N.J. Watson, Monitoring mixing processes using ultrasonic sensors and machine learning, Sensors 20 (2020) 1813, https://doi.org/10.3390/s20071813.
[28] F. Cau, A. Fanni, A. Montisci, P. Testoni, M. Usai, Artificial neural networks for non-destructive evaluation with ultrasonic waves in not accessible, IEEE Ind. Applic. Soc. 1 (2005) 685–692, https://doi.org/10.1109/IAS.2005.1518382.
[29] C. Miao, Y. Wang, Y. Zhang, J. Qn, M.J. Zuo, X. Wang, A SVM classifier combined with PCA for ultrasonic crack size classification, Can. Conf. Electr. Comput. Eng. (2008) 1627–1630, https://doi.org/10.1109/CCECE.2008.4564817.
[30] S. Khalid, T. Khalil, S. Nasreen, A survey of feature selection and feature extraction techniques in machine learning, in: Proceedings of 2014 Science and Information Conference, 2014, pp. 372–378, https://doi.org/10.1109/SAI.2014.6918213.
[31] R. Jenke, A. Peer, M. Buss, Feature extraction and selection for emotion recognition from EEG, IEEE T Affect. Comput. 5 (3) (2014) 327–339, https://doi.org/10.1109/TAFFC.2014.2339834.
[32] D. Gómez, A. Rojas, An empirical overview of the no free lunch theorem and its effect on real-world machine learning classification, Neural Comput. 28 (1) (2016) 216–228, https://doi.org/10.1162/NECO_a_00793.
[33] A.J. Smola, B. Schölkopf, A tutorial on support vector regression, Stat. Comput. 14 (3) (2004) 199–222, https://doi.org/10.1023/B:STCO.0000035301.49549.88.
[34] V. Rodriguez-Galiano, M. Sanchez-Castillo, M. Chica-Olmo, M. Chica-Rivas, Machine learning predictive models for mineral prospectivity: an evaluation of neural networks, random forest, regression trees and support vector machines, Ore Geol. Rev. 71 (2015) 804–818, https://doi.org/10.1016/j.oregeorev.2015.01.001.
[35] N. Munir, H.-J. Kim, J. Park, S.-J. Song, S.-S. Kang, Convolutional neural network for ultrasonic weldment flaw classification in noisy conditions, Ultrasonics 94 (2019) 74–81, https://doi.org/10.1016/j.ultras.2018.12.001.
[36] S. Hochreiter, J. Schmidhuber, Long short-term memory, Neural Comput 9 (8) (1997) 1735–1780, https://doi.org/10.1162/neco.1997.9.8.1735.
[37] J. Vaněk, J. Zelinka, D. Soutner, J. Psutka, A regularization post layer: an additional way how to make deep neural networks robust, Lect. Notes Comp. Sc. 10583 (2017) 204–214, https://doi.org/10.1007/978-3-319-68456-7_17.
[38] N. Srivastava, G. Hinton, A. Krizhevsky, I. Sutsever, R. Salakhutdinov, Dropout: a simple way to prevent neural networks from overfitting, J. Mach. Learn. Res. 15 (2014) 1929–1958.
[39] J.E. Escrig, E. Woolley, S. Rangappa, A. Simeone, N.J. Watson, Clean-in-place monitoring of different food fouling materials using ultrasonic measurements, Food Control 104 (2019) 358–366, https://doi.org/10.1016/j.foodcont.2019.05.013.
[40] M.A. Úbeda, W.B. Hussein, M.A. Hussein, J. Hinrichs, T.M. Becker, Acoustic sensing and signal processing techniques for monitoring milk fouling cleaning operations, Eng. Life Sci. 16 (1) (2016) 67–77, https://doi.org/10.1002/elsc.201400235.
[41] E. Wallhäußer, W.B. Hussein, M.A. Hussein, J. Hinrichs, T.M. Becker, On the usage of acoustic properties combined with an artificial neural network – a new approach of determining presence of dairy fouling, J. Food Eng. 103 (4) (2011) 449–456, https://doi.org/10.1016/j.jfoodeng.2010.11.015.
[42] E. Wallhäußer, W.B. Hussein, M.A. Hussein, J. Hinrichs, T. Becker, Detection of dairy fouling: combining ultrasonic measurements and classification methods, Eng. Life Sci. 13 (3) (2013) 292–301, https://doi.org/10.1002/elsc.201200081.
[43] E. Wallhäußer, A. Sayed, S. Nöbel, M.A. Hussein, J. Hinrichs, T. Becker, Determination of cleaning end of dairy protein fouling using an online system combining ultrasonic and classification methods, Food Bioprocess Tech. 7 (2) (2014) 506–515, https://doi.org/10.1007/s11947-012-1041-0.
[44] J. Escrig, E. Woolley, A. Simeone, N.J. Watson, Monitoring the cleaning of food fouling in pipes using ultrasonic measurements and machine learning, Food Control 116 (2020) 107309, https://doi.org/10.1016/j.foodcont.2020.107309.
[45] B. Chen, D. Callens, P. Campistron, E. Moulin, P. Debreyne, G. Delaplace, Monitoring cleaning cycles of fouled ducts using ultrasonic coda wave interferometry (CWI), Ultrasonics 96 (2019) 252–260, https://doi.org/10.1016/j.ultras.2018.12.011.
[46] P. Resa, L. Elvira, F. Montero de Espinosa, Concentration control in alcoholic fermentation processes from ultrasonic velocity measurements, Food Res. Int. 37 (6) (2004) 587–594, https://doi.org/10.1016/j.foodres.2003.12.012.
[47] P. Resa, L. Elvira, F.M. De Espinosa, T. Bolumar, G. Pérez, Ultrasonic velocity measurements in the ternary mixtures water-lactose-lactate, for the purpose of monitoring the lactic acid fermentation of lactose, Proc. – IEEE Ultrason. Symp. 3 (2004) P3U-Y-1, https://doi.org/10.1109/ULTSYM.2004.1418306.
[48] D. Novoa-Díaz, J.M. Rodríguez-Nogales, E. Fernández-Fernández, J. Vila-Crespo, J. García-Álvarez, M.A. Amer, J.A. Chávez, A. Turó, M.J. García-Hernández, J. Salazar, Ultrasonic monitoring of malolactic fermentation in red wines, Ultrasonics 54 (6) (2014) 1575–1580, https://doi.org/10.1016/j.ultras.2014.04.004.
[49] T. Becker, M. Mitzscherling, A. Delgado, Hybrid data model for the improvement of an ultrasonic-based gravity measurement system, Food Control 13 (4-5) (2002) 223–233, https://doi.org/10.1016/S0956-7135(01)00104-9.
[50] P. Resa, T. Bolumar, L. Elvira, G. Pérez, F.M. de Espinosa, Monitoring of lactic acid fermentation in culture broth using ultrasonic velocity, J. Food Eng. 78 (3) (2007) 1083–1091, https://doi.org/10.1016/j.jfoodeng.2005.12.021.
[51] N. Lamberti, L. Ardia, D. Albanese, M. Di Matteo, An ultrasound technique for monitoring the alcoholic wine fermentation, Ultrasonics 49 (1) (2009) 94–97, https://doi.org/10.1016/j.ultras.2008.06.003.
[52] S. Hoche, D. Krause, M.A. Hussein, T. Becker, Ultrasound-based, in-line monitoring of anaerobe yeast fermentation: Model, sensor design and process application, 51 (2016) 710–719. doi:10.1111/ijfs.13027.
[53] A. Bowler, J. Escrig, M. Pound, N. Watson, Predicting alcohol concentration during beer fermentation using ultrasonic measurements and machine learning, Fermentation 7 (2021) 34, https://doi.org/10.3390/fermentation7010034.
[54] H. Ogasawara, K. Mizutani, T. Ohbuchi, T. Nakamura, Acoustical experiment of yogurt fermentation process, Ultrasonics 44 (2006) e727–e730, https://doi.org/10.1016/j.ultras.2006.05.084.
[55] R. Meng, J. Zhou, X. Ye, D. Liu, On-line monitoring of yogurt fermentation using acoustic impedance method, Appl. Mech. Mater. 101–102 (2012) 737–742, https://doi.org/10.4028/www.scientific.net/AMM.101-102.737.
[56] M.A. Amer, D. Novoa-Díaz, A. Puig-Pujol, J. Capdevila, J.A. Chávez, A. Turó, M. J. García-Hernández, J. Salazar, Ultrasonic velocity of water-ethanol-malic acid-lactic acid mixtures during the malolactic fermentation process, J. Food Eng. 149 (2015) 61–69, https://doi.org/10.1016/j.jfoodeng.2014.09.042.
[57] M.A. Amer, D. Novoa-Díaz, J.A. Chávez, A. Turó, M.J. García-Hernández, J. Salazar, Temperature compensation of ultrasonic velocity during the malolactic fermentation process, Meas. Sci. Technol. 26 (12) (2015) 125602, https://doi.org/10.1088/0957-0233/26/12/125602.
[58] D.A. Çelik, M.A. Amer, D. Novoa-Díaz, J.A. Chávez, A. Turó, M.J. García-Hernández, J. Salazar, Design and implementation of an ultrasonic sensor for rapid monitoring of industrial malolactic fermentation of wines 46 (2018) 387–407. DOI:10.1080/10739149.2017.1394878.
[59] C. Keskinoğlu, A. Aydın, Ultrasound based noninvasive real-time cell proliferation process monitoring, J. Acoust. Soc. Am. 149 (5) (2021) 3345–3351, https://doi.org/10.1121/10.0004993.
[60] T. Stelzer, D. Pertig, J. Ulrich, Ultrasonic crystallization monitoring technique for simultaneous in-line measurement of liquid and solid phase, J. Cryst. Growth 362 (2013) 71–76, https://doi.org/10.1016/j.jcrysgro.2011.11.027.
[61] P. Mougin, A. Thomas, D. Wilkinson, G. White, K.J. Roberts, N. Herrmann, R. Jack, R. Tweedie, On-line monitoring of a crystallization process, AIChE J. 49 (2) (2003) 373–378, https://doi.org/10.1002/aic.690490209.
[62] P. Mougin, D. Wilkinson, K.J. Roberts, R. Tweedie, Characterization of particle size and its distribution during the crystallization of organic fine chemical products as measured in situ using ultrasonic attenuation spectroscopy, J. Acoust. Soc. Am. 109 (2001) 274–282, https://doi.org/10.1121/1.1331113.
[63] P. Mougin, D. Wilkinson, K.J. Roberts, In situ measurement of particle size during the crystallization of L-glutamic acid under two polymorphic forms: influence of crystal habit on ultrasonic attenuation measurements, Crystal Growth Des. 2 (2002) 227–234, https://doi.org/10.1021/cg0155752.
[64] M. Li, D. Wilkinson, K. Patchigolla, P. Mougin, K.J. Roberts, R. Tweedie, On-line crystallization process parameter measurements using ultrasonic attenuation spectroscopy, Cryst. Growth Des. 4 (5) (2004) 955–963, https://doi.org/10.1021/cg030041h.
[65] E. Lyall, P. Mougin, D. Wilkinson, K.J. Roberts, In situ ultrasonic spectroscopy study of the nucleation and growth of copper sulfate pentahydrate batch crystallized from supersaturated aqueous solutions, Ind. Eng. Chem. Res. 43 (16) (2004) 4947–4956, https://doi.org/10.1021/ie0342560.
[66] A. Shukla, A. Prakash, S. Rohani, Online measurement of particle size distribution during crystallization using ultrasonic spectroscopy, Chem. Eng. Sci. 65 (10) (2010) 3072–3079, https://doi.org/10.1016/j.ces.2010.01.034.
[67] D. Pertig, R. Buchfink, S. Petersen, T. Stelzer, J. Ulrich, Inline analyzing of industrial crystallization processes by an innovative ultrasonic probe technique, Chem. Eng. Technol. 34 (2011) 639–646, https://doi.org/10.1002/ceat.201000558.
[68] P. Frohberg, J. Ulrich, Single-frequency ultrasonic crystallization monitoring (UCM): Innovative technique for in-line analyzing of industrial crystallization processes, Org. Process Res. Dev. 19 (1) (2015) 84–88, https://doi.org/10.1021/op400362f.
[69] L. Helmdach, M.P. Feth, J. Ulrich, Application of ultrasound measurements as pat tools for industrial crystallization process development of pharmaceutical

compounds, Org. Process Res. Dev. 19 (1) (2015) 110–121, https://doi.org/10.1021/op4001803.
[70] L. Morris, E. Simeone, Z.J. Glover, H. Powell, S. Marty-Terrade, M. Francis, M. J. Povey, Dynamic monitoring of glycine crystallisation with low power ultrasound reflection spectroscopy, Chem. Eng. Res. Des. 170 (2021) 213–223, https://doi.org/10.1016/j.cherd.2021.04.003.
[71] J.A. Bamberger, M.S. Greenwood, Using ultrasonic attenuation to monitor slurry mixing in real time, Ultrasonics 42 (1-9) (2004) 145–148, https://doi.org/10.1016/j.ultras.2004.02.016.
[72] P. Fox, P.P. Smith, S. Sahi, Ultrasound measurements to monitor the specific gravity of food batters, J. Food Eng. 65 (3) (2004) 317–324, https://doi.org/10.1016/j.jfoodeng.2004.01.028.
[73] J. Salazar, A. Turó, J.A. Chávez, M.J. García, Ultrasonic inspection of batters for on-line process monitoring, Ultrasonics 42 (1-9) (2004) 155–159, https://doi.org/10.1016/j.ultras.2004.02.017.
[74] M. Tourbin, C. Frances, Monitoring of the aggregation process of dense colloidal silica suspensions in a stirred tank by acoustic spectroscopy, Powder Technol. 190 (1-2) (2009) 25–30, https://doi.org/10.1016/j.powtec.2008.04.067.
[75] L. Liu, R.F. Li, S. Collins, X.Z. Wang, R. Tweedie, K. Primrose, Ultrasound spectroscopy and electrical resistance tomography for online characterisation of concentrated emulsions in crossflow membrane emulsifications, Powder Technol. 213 (1-3) (2011) 123–131, https://doi.org/10.1016/j.powtec.2011.07.018.
[76] T.N. Hunter, L. Darlison, J. Peakall, S. Biggs, Using a multi-frequency acoustic backscatter system as an in situ high concentration dispersion monitor, Chem. Eng. Sci. 80 (2012) 409–418, https://doi.org/10.1016/j.ces.2012.06.038.
[77] G.M. Owolabi, M.N. Bassim, J.H. Page, M.G. Scanlon, The influence of specific mechanical energy on the ultrasonic characteristics of extruded dough, J. Food Eng. 86 (2) (2008) 202–206, https://doi.org/10.1016/j.jfoodeng.2007.09.029.
[78] P.D. Coates, S.E. Barnes, M.G. Sibley, E.C. Brown, H.G.M. Edwards, I.J. Scowen, In-process vibrational spectroscopy and ultrasound measurements in polymer melt extrusion, Polymer 44 (19) (2003) 5937–5949, https://doi.org/10.1016/S0032-3861(03)00544-5.
[79] S.E. Barnes, E.C. Brown, M.G. Sibley, H.G.M. Edwards, P.D. Coates, Vibrational spectroscopic and ultrasound analysis for the in-process monitoring of poly (ethylene vinyl acetate) copolymer composition during melt extrusion, Analyst 130 (2005) 286–292, https://doi.org/10.1039/b416244g.
[80] I. Alig, D. Fischer, D. Lellinger, B. Steinhoff, Combination of NIR, Raman, ultrasonic and dielectric spectroscopy for in-line monitoring of the extrusion process 230 (2005) 51–58. doi:10.1002/masy.200551141.
[81] Z. Sun, C.-K. Jen, J. Yan, M.-Y. Chen, Application of ultrasound and neural networks in the determination of filler dispersion during polymer extrusion processes, Polym. Eng. Sci. 45 (6) (2005) 764–772, https://doi.org/10.1002/pen.20328.
[82] D. Fischer, K. Sahre, M. Abdelrhim, B. Voit, V.B. Sadhu, J. Pionteck, H. Komber, J. Hutschenreuter, Process monitoring of polymers by in-line ATR-IR, NIR and Raman spectroscopy and ultrasonic measurements, CR Chim. 9 (2006) 1419–1424, https://doi.org/10.1016/j.crci.2006.06.006.
[83] D. Fischer, J. Müller, S. Kummer, B. Kretzschmar, Real time monitoring of morphologic and mechanical properties of polymer nanocomposites during extrusion by near infrared and ultrasonic spectroscopy, Macromol. Symp. 305 (2011) 10–17, https://doi.org/10.1002/masy.201000113.
[84] G. Schober, P. Heidemeyer, K. Kretschmer, M. Bastian, T. Hochrein, Degree of dispersion monitoring by ultrasonic transmission technique and excitation of the transducer's harmonics, AIP Conf. Proc. 1593 (2014) 24–28, https://doi.org/10.1063/1.4873727.
[85] S. Wöckel, h. Arndt, U. Steinmann, J. Auge, K. Dietl, G. Schober, C. Kugler, T. Hochrein, Statistical ultrasonic characterization of particulate filler in polymer compounds, IEEE Int. Ultra. Sym. 2016-November (2016) 7728747. doi:10.1109/ULTSYM.2016.7728747.
[86] N. Halmen, C. Kugler, T. Hochrein, P. Heidemeyer, M. Bastian, Ultrasound tomography for inline monitoring of plastic melts, J. Sens. Sens. Syst. 6 (1) (2017) 9–18, https://doi.org/10.5194/jsss-6-9-201710.5194/jsss-6-9-2017-supplement.
[87] M. Kariminejad, D. Tormey, S. Huq, J. Morrison, M. McAfee, Ultrasound sensors for process monitoring in injection moulding, Sensors 21 (2021) 5193, https://doi.org/10.3390/s21155193.
[88] Y.-L. Wu, C.-C. Cheng, M. Kobayashi, C.-H. Yang, Novel design of extension nozzle and its application on real-time injection molding process diagnosed by ultrasound, Sensor. Actuat. A-Phys. 263 (2017) 430–438, https://doi.org/10.1016/j.sna.2017.06.023.
[89] D. Altmann, B. Praher, G. Steinbichler, Simulation of the melting behavior in an injection molding plasticizing unit as measured by pressure and ultrasound measurement technology, AIP Conf. Proc. 2055 (2019) 040003, https://doi.org/10.1063/1.5084818.
[90] P. Zhao, Y. Zhao, H. Kharbas, J. Zhang, T. Wu, W. Yang, J. Fu, L.-S. Turng, In-situ ultrasonic characterization of microcellular injection molding, J. Mater. Process. Tech. 270 (2019) 254–264, https://doi.org/10.1016/j.jmatprotec.2019.03.012.
[91] C.-C. Cheng, Y.-L. Wu, Diagnosis of multi-stage injection molding process by ultrasonic technology at a T-shape extension nozzle, J. Mater. Process. Tech. 282 (2020) 116650, https://doi.org/10.1016/j.jmatprotec.2020.116650.
[92] P. Zhao, K. Ji, J. Zhang, Y. Chen, Z. Dong, J. Zheng, J. Fu, In-situ ultrasonic measurement of molten polymers during injection molding, J. Mater. Process. Tech. 293 (2021) 117081, https://doi.org/10.1016/j.jmatprotec.2021.117081.
[93] L. Grob, K. Papadea, P. Braun, E.J. Windhab, In-line detection method for crystallization, contraction and mold detachment during cooling of confectionery products, J Food Eng 292 (2021) 110322, https://doi.org/10.1016/j.jfoodeng.2020.110322.
[94] M. Rath, J. Döring, W. Stark, G. Hinrichsen, Process monitoring of moulding compounds by ultrasonic measurements in a compression mould, NDT&E Int. 33 (2) (2000) 123–130, https://doi.org/10.1016/S0963-8695(99)00029-8.
[95] F. Lionetto, A. Tarzia, A. Maffezzoli, Air-coupled ultrasound: a novel technique for monitoring the curing of thermosetting matrices, IEEE T. Ultrason. Ferr. 54 (7) (2007) 1437–1444, https://doi.org/10.1109/TUFFC.2007.404.
[96] F. Lionetto, A. Maffezzoli, Monitoring the cure state of thermosetting resins by ultrasound, Materials 6 (2013) 3783–3804, https://doi.org/10.3390/ma6093783.
[97] V. Koissin, A. Demčenko, V.A. Korneev, Isothermal epoxy-cure monitoring using nonlinear ultrasonics, Int. J. Ahes. Adhes. 52 (2014) 11–18, https://doi.org/10.1016/j.ijadhadh.2014.01.003.
[98] N. Ghodhbani, P. Maréchal, H. Duflo, Ultrasound monitoring of the cure kinetics of an epoxy resin: Identification, frequency and temperature dependence, Polym. Test. 56 (2016) 156–166, https://doi.org/10.1016/j.polymertesting.2016.10.009.
[99] A. Dominguez-Macaya, T.E.G. Álvarez-Arenas, I. Saenz-Dominguez, I. Tena, J. Aurrekoetxea, A. Iturrospe, Monitoring the evolution of stiffness during ultraviolet curing of a vinyl ester resin with quasi-normal air-coupled ultrasonic spectroscopy, Polym. Test. 80 (2019) 106112, https://doi.org/10.1016/j.polymertesting.2019.106112.
[100] J.S. Chilles, A.E. Koutsomitopoulou, A.J. Croxford, I.P. Bond, Monitoring cure and detecting damage in composites with inductively coupled embedded sensors, Compos. Sci. Technol. 134 (2016) 81–88, https://doi.org/10.1016/j.compscitech.2016.07.028.
[101] K. Mizukami, S. Yoshimoto, K. Ogi, In-process acquisition of cure-dependent viscoelastic properties of carbon fiber reinforced composites using micromechanics-based guided wave analysis, Polym. Test. 65 (2018) 459–467, https://doi.org/10.1016/j.polymertesting.2017.12.032.
[102] K. Mizukami, T. Ikeda, K. Ogi, Measurement of velocity and attenuation of ultrasonic guided wave for real-time estimation of cure-dependent anisotropic viscoelastic properties of carbon fiber-reinforced plastics, Ultrasonics 99 (2019) 105952, https://doi.org/10.1016/j.ultras.2019.105952.
[103] X. Liu, J. Li, J. Zhu, Y. Wang, X. Qing, Cure monitoring and damage identification of CFRP using embedded piezoelectric sensors network, Ultrasonics 115 (2021) 106470, https://doi.org/10.1016/j.ultras.2021.106470.
[104] P. Pawelzyk, M.L. Toledo, N. Willenbacher, Ultrasonic in-line monitoring of styrene miniemulsion polymerization, Chem. Eng. J. 219 (2013) 303–310, https://doi.org/10.1016/j.cej.2013.01.014.
[105] V. Buckin, M.C. Atlas, Ultrasonic monitoring of biocatalysis in solutions and complex dispersions, Catalysts 7 (2017) 336, https://doi.org/10.3390/catal7110336.
[106] V. Buckin, High-resolution ultrasonic spectroscopy, J. Sens. Sens. Syst. 7 (2018) 207–217, https://doi.org/10.5194/jsss-7-207-2018.
[107] M.K.K. Figueiredo, C.E.R. Silva, A.V. Alvarenga, R.P.B. Costa-Félix, Relating speed of sound and echo amplitude with biodiesel manufacture, Chem. Eng. Res. Des. 136 (2018) 825–833, https://doi.org/10.1016/j.cherd.2018.06.038.
[108] R.M. Baêsso, R.P.B. Costa-Felix, P. Miloro, B. Zeqiri, Ultrasonic parameter measurement as a means of assessing the quality of biodiesel production, Fuel 241 (2019) 155–163, https://doi.org/10.1016/j.fuel.2018.12.032.
[109] M. Schmachtl, T.J. Kim, W. Grill, R. Herrmann, O. Scharf, W. Schwieger, R. Schertlen, C. Stenzel, Ultrasonic monitoring of zeolite synthesis in real time, Ultrasonics 38 (2000) 809–812, https://doi.org/10.1016/S0041-624X(99)00201-2.
[110] E. Hums, H. Baser, W. Schwieger, In situ ultrasonic measurements: a powerful tool to control the synthesis of zeolites from coal fly ash, Res. Chem. Intermediat. 42 (10) (2016) 7513–7532, https://doi.org/10.1007/s11164-016-2550-7.
[111] G.J.A. Van Groenestijn, N.M.M. Meulendiiks, A.W.F. Volker, P.L.M.J. Van Neer, P. J.P. Buskens, Real-time monitoring of size and concentration of nanoparticles inside a reactor using ultrasound, IEEE Int. Ultra. Sym. (2018) 143804, https://doi.org/10.1109/ULTSYM.2018.8580198.
[112] J.D. Stephens, B.R. Kowalczyk, B.C. Hancock, G. Kaul, I. Akseli, C. Cetinkaya, In-die ultrasonic and off-line air-coupled monitoring and characterization techniques for drug tablets, AIP Conf. Proc. 1430 (2012) 1691–1698, https://doi.org/10.1063/1.4716416.
[113] J.T.T. Leskinen, S.-P. Simonaho, M. Hakulinen, J. Ketolainen, Real-time tablet formation monitoring with ultrasound measurements in eccentric single station tablet press, Int. J. Pharm. 442 (1-2) (2013) 27–34, https://doi.org/10.1016/j.ijpharm.2012.09.004.
[114] X. Li, Y. Mo, J. Li, W. Guo, H.H. Ngo, In-situ monitoring techniques for membrane fouling and local filtration characteristics in hollow fiber membrane processes: a critical review, J. Membrane Sci. 528 (2017) 187–200, https://doi.org/10.1016/j.memsci.2017.01.030.
[115] G. Rudolph, T. Virtanen, M. Ferrando, C. Güell, F. Lipnizki, M. Kallioinen, A review of in situ real-time monitoring techniques for membrane fouling in the biotechnology, biorefinery and food sectors, J. Membrane Sci. 588 (2019) 117221, https://doi.org/10.1016/j.memsci.2019.117221.
[116] X. Li, H. Zhang, Y. Hou, Y. Gao, J. Li, W. Guo, H.H. Ngo, In situ investigation of combined organic and colloidal fouling for nanofiltration membrane using ultrasonic time domain reflectometry, Desalination 362 (2015) 43–51, https://doi.org/10.1016/j.desal.2015.02.005.
[117] M. Mitra, S. Gopalakrishnan, Guided wave based structural health monitoring: a review, Smart Mater. Struct. 25 (5) (2016) 053001, https://doi.org/10.1088/0964-1726/25/5/053001.
[118] P. Gardner, R. Fuentes, N. Dervilis, C. Mineo, S.G. Pierce, E.J. Cross, K. Worden, Machine learning at the interface of structural health monitoring and non-

destructive evaluation: Machine Learning in SHM and NDE, Philos. Trans. Royal Soc. A 378 (2182) (2020) 20190581, https://doi.org/10.1098/rsta.2019.0581.
[119] M. Azimi, A.D. Eslamlou, G. Pekcan, Data-driven structural health monitoring and damage detection through deep learning: State-ofthe- art review, Sensors (Basel) 20 (10) (2020). doi:10.3390/s20102778.
[120] X.W. Ye, T. Jin, C.B. Yun, A review on deep learning-based structural health monitoring of civil infrastructures, Smart Struct. Syst. 24 (5) (2019) 567–585, https://doi.org/10.12989/sss.2019.24.5.567.
[121] S. Sony, K. Dunphy, A. Sadhu, M. Capretz, A systematic review of convolutional neural network-based structural condition assessment techniques, Eng. Struct. 226 (2021) 111347, https://doi.org/10.1016/j.engstruct.2020.111347.
[122] Y. Bao, Z. Chen, S. Wei, Y. Xu, Z. Tang, H. Li, The state of the art of data science and engineering in structural health monitoring, J. Eng. 5 (2) (2019) 234–242, https://doi.org/10.1016/j.eng.2018.11.027.
[123] M. Flah, I. Nunez, W. Ben Chaabene, M.L. Nehdi, Machine learning algorithms in civil structural health monitoring: a systematic review, Arch Comput. Methods Eng 28 (4) (2021) 2621–2643.
[124] J.B. Harley, D. Sparkman, Machine learning and NDE: Past, present, and future, in: AIP Conf. Proc. doi:10.1063/1.5099819.
[125] H. Ahmed, H.M. La, N. Gucunski, Review of non-destructive civil infrastructure evaluation for bridges: state-of-the-art robotic platforms, sensors and algorithms, Sensors (Basel) 20 (14) (2020) 1–38, https://doi.org/10.3390/s20143954.
[126] M. Gordan, S.-R. Sabbagh-Yazdi, Z. Ismail, K. Ghaedi, P. Carroll, D. McCrum, B. Samali, State-of-the-art review on advancements of data mining in structural health monitoring, Meas.: J. Int. Meas. Confed. 193 (2022) 110939, https://doi.org/10.1016/j.measurement.2022.110939.
[127] G. Toh, J. Park, Review of vibration-based structural health monitoring using deep learning, Appl. Sci. 10 (5) (2020) 1680, https://doi.org/10.3390/app10051680.
[128] R. Zhao, R. Yan, Z. Chen, K. Mao, P. Wang, R.X. Gao, Deep learning and its applications to machine health monitoring, Mech. Syst. Signal. Pr. 115 (2018) 213–237, https://doi.org/10.1016/j.ymssp.2018.05.050.
[129] S. Lu, L.i. Zhou, Fatigue crack monitoring of aerospace structure based on lamb waves and binary tree support vector machines, J. Vibroeng. 19 (5) (2017) 3271–3282, https://doi.org/10.21595/jve.2017.17528.
[130] A.K.U. Malikov, Y. Cho, Y.H. Kim, J. Kim, J. Park, J.-H. Yi, Ultrasonic assessment of thickness and bonding quality of coating layer based on short-time fourier transform and convolutional neural networks, Coatings 11 (2021) 909, https://doi.org/10.3390/coatings11080909.
[131] M. Barbosh, P. Singh, A. Sadhu, Empirical mode decomposition and its variants: a review with applications in structural health monitoring, Smart Mater. Struct. 29 (9) (2020) 093001, https://doi.org/10.1088/1361-665X/aba539.
[132] L. Zhang, H. Li, B. Gao, Combination of wavelet packet analysis with BPNN flaw type identification in concrete ultrasonic testing, in: 3rd International Conference on Innovative Computing Information and Control, 2008, p. 4603704, https://doi.org/10.1109/ICICIC.2008.195.
[133] S. Sambath, P. Nagaraj, N. Selvakumar, Automatic defect classification in ultrasonic NDT using artificial intelligence, J. Nondestruct. Eval. 30 (1) (2011) 20–28, https://doi.org/10.1007/s10921-010-0086-0.
[134] M. Meng, Y.J. Chua, E. Wouterson, C.P.K. Ong, Ultrasonic signal classification and imaging system for composite materials via deep convolutional neural networks, Neurocomputing 257 (2017) 128–135, https://doi.org/10.1016/j.neucom.2016.11.066.
[135] K. Virupakshappa, M. Marino, E. Oruklu, A multi-resolution convolutional neural network architecture for ultrasonic flaw detection, IEEE Int. Ultra. Sym. (2018) 8579888, https://doi.org/10.1109/ULTSYM.2018.8579888.
[136] Y. Yan, D. Liu, B. Gao, G.Y. Tian, Z.C. Cai, A deep learning-based ultrasonic pattern recognition method for inspecting girth weld cracking of gas pipeline, IEEE Sens. J. 20 (14) (2020) 7997–8006, https://doi.org/10.1109/JSEN.2020.2982680.
[137] M. Rautela, S. Gopalakrishnan, Ultrasonic guided wave based structural damage detection and localization using model assisted convolutional and recurrent neural networks, Exp. Syst. Appl. 167 (2021) 114189, https://doi.org/10.1016/j.eswa.2020.114189.
[138] S. Bosse, D. Weiss, D. Schmidt, Supervised distributed multi-instance and unsupervised single-instance autoencoder machine learning for damage diagnostics with high-dimensional data—a hybrid approach and comparison study, Computers 10 (2021) 34, https://doi.org/10.3390/computers10030034.
[139] J. Liu, G. Xu, L. Ren, Z. Qian, L. Ren, Defect intelligent identification in resistance spot welding ultrasonic detection based on wavelet packet and neural network, Int. J. Adv. Manuf. Tech. 90 (9–12) (2017) 2581–2588, https://doi.org/10.1007/s00170-016-9588-y.
[140] K. Sudheera, N.M. Nandhitha, V.B.V. Sai, N.V. Kumar, Deep learning techniques for flaw characterization in weld pieces from ultrasonic signals, Russ. J. Nondestruct+ 56 (10) (2020) 820–830, https://doi.org/10.1134/S1061830920100083.
[141] X. Wang, S. Guan, L. Hua, B. Wand, X. He, Classification of spot-welded joint strength using ultrasonic signal time-frequency features and PSO-SVM method, Ultrasonics 91 (2019) 161–169, https://doi.org/10.1016/j.ultras.2018.08.014.
[142] K. Virupakshappa, E. Oruklu, Ultrasonic flaw detection using hidden markov model with wavelet features, IEEE Int. Ultra. Sym. (2016) 7728491, https://doi.org/10.1109/ULTSYM.2016.7728491.
[143] Y. Lu, L. Ye, Z. Su, L. Zhou, L. Cheng, Artificial Neural Network (ANN)-based crack identification in aluminum plates with lamb wave signals, J. Intel. Mat. Syst. Str. 20 (2009) 39–49, https://doi.org/10.1177/1045389X07088782.
[144] F. Lanza di Scalea, P. Rizzo, S. Coccia, I. Bartoli, M. Fateh, E. Viola, G. Pascale, Non-contact ultrasonic inspection of rails and signal processing for automatic defect detection and classification, Insight: Non-Destruct. Test. Cond. Monitor. 47 (6) (2005) 346–353, https://doi.org/10.1784/insi.47.6.346.66449.
[145] P. Rizzo, I. Bartoli, A. Marzani, F.L. di Scalea, Defect classification in pipes by neural networks using multiple guided ultrasonic wave features extracted after wavelet processing, J. Press. Vess.-T. ASME 127 (2005) 294–303, https://doi.org/10.1115/1.1990213.
[146] S. Iyer, S.K. Sinha, B.R. Tittmann, M.K. Pedrick, Ultrasonic signal processing methods for detection of defects in concrete pipes, Autom. Constr. 22 (2012) 135–148, https://doi.org/10.1016/j.autcon.2011.06.012.
[147] H. Sun, S. Huang, S. Wang, W. Zhao, L. Peng, Quantification of defects with point-focusing shear horizontal guided wave EMAT using deep residual network, IEEE Intl. Conf. Ind. I (2021), https://doi.org/10.1109/INDIN45523.2021.9557567.
[148] N. Munir, H.-J. Kim, S.-J. Song, S.-S. Kang, Investigation of deep neural network with drop out for ultrasonic flaw classification in weldments, J. Mech. Sci. Technol. 32 (7) (2018) 3073–3080, https://doi.org/10.1007/s12206-018-0610-1.
[149] B.M. Abbagoni, H. Yeung, Non-invasive classification of gas-liquid two-phase horizontal flow regimes using an ultrasonic Doppler sensor and a neural network, Meas. Sci. Technol. 27 (8) (2016) 084002, https://doi.org/10.1088/0957-0233/27/8/084002.
[150] Y. Zhang, A.N. Azman, K.-W. Xu, C. Kang, H.-B. Kim, Two-phase flow regime identification based on the liquid-phase velocity information and machine learning, Exp. Fluids 61 (2020) 212, https://doi.org/10.1007/s00348-020-03046-x.
[151] S.-J. Song, H.-J. Kim, H. Cho, Development of an intelligent system for ultrasonic flaw classification in weldments, Nucl. Eng. Des. 212 (1-3) (2002) 307–320, https://doi.org/10.1016/S0029-5493(01)00495-2.
[152] M.B. Utomo, T. Sakai, S. Uchida, Use of neural network-ultrasonic technique for measuring gas and solid hold-ups in a slurry bubble column, Chem. Eng. Technol. 25 (2002) 293–299, https://doi.org/10.1002/1521-4125(200203)25:3<293::AID-CEAT293>3.0.CO;2-X/.
[153] M.B. Utomo, T. Sakai, S. Uchida, A. Maezawa, Simultaneous measurement of mean bubble diameter and local gas holdup using ultrasonic method with neural network, Chem. Eng. Technol. 24 (2001) 493–500, https://doi.org/10.1002/1521-4125(200105)24:5<493::AID-CEAT493>3.0.CO;2-L.
[154] Z. Tang, N. Munir, T.-G. Lee, Y.-T. Yeom, S.-J. Song, Lamb wave flaw classification in al plates using time reversal and deep neural networks, J. Korean Phys. Soc. 75 (12) (2019) 978–984, https://doi.org/10.3938/jkps.75.978.
[155] L.C. Silva, E.F. Simas Filho, M.C.S. Albuquerque, I.C. Silva, C.T.T. Farias, Segmented analysis of time-of-flight diffraction ultrasound for flaw detection in welded steel plates using extreme learning machines, Ultrasonics 102 (2020) 106057, https://doi.org/10.1016/j.ultras.2019.106057.
[156] D. Sen, K. Erazo, W. Zhang, S. Nagarajaiah, L. Sun, On the effectiveness of principal component analysis for decoupling structural damage and environmental effects in bridge structures, J. Sound Vib. 457 (2019) 280–298, https://doi.org/10.1016/j.jsv.2019.06.003.
[157] M.R. Azim, M. Gül, Data-driven damage identification technique for steel truss railroad bridges utilizing principal component analysis of strain response, Struct. Infrastruct. Eng. 17 (8) (2021) 1019–1035, https://doi.org/10.1080/15732479.2020.1785512.
[158] Z. Nie, Z. Shen, J. Li, H. Hao, Y. Lin, H. Ma, H. Jiang, Using a single sensor for bridge condition monitoring via moving embedded principal component analysis, Struct. Health Monit. 20 (6) (2021) 3123–3149, https://doi.org/10.1177/1475921720980516.
[159] Y. Yang, S. Nagarajaiah, Blind denoising of structural vibration responses with outliers via principal component pursuit, Struct. Control Health Monit. 21 (6) (2014) 962–978, https://doi.org/10.1002/stc.1624.
[160] X. Hong, B. Zhang, Y. Liu, H. Qi, W. Li, Deep-learning-based guided wave detection for liquid-level state in porcelain bushing type terminal, Struct. Control Health Monitor. 28 (1) (2021), https://doi.org/10.1002/stc.v28.110.1002/stc.2651.
[161] H. Lee, H.J. Lim, T. Skinner, A. Chattopadhyay, A. Hall, Automated fatigue damage detection and classification technique for composite structures using Lamb waves and deep autoencoder, Mech. Syst. Signal Pr. 163 (2022) 108148, https://doi.org/10.1016/j.ymssp.2021.108148.
[162] N. Munir, J. Park, H.-J. Kim, S.-J. Song, S.-S. Kang, Performance enhancement of convolutional neural network for ultrasonic flaw classification by adopting autoencoder, NDT E Int. 111 (2020) 102218, https://doi.org/10.1016/j.ndteint.2020.102218.
[163] M. Rautela, S. Jayavelu, J. Moll, S. Gopalakrishnan, Temperature compensation for guided waves using convolutional denoising autoencoders, P. Soc. Photo.-Opt. Ins. 11593 (2021) 1159319, https://doi.org/10.1117/12.2582986.
[164] F. Gao, B. Li, L. Chen, X. Wei, Z. Shang, C. He, Ultrasonic signal denoising based on autoencoder, Rev. Sci. Instrum. 91 (4) (2020) 045104, https://doi.org/10.1063/1.5136269.
[165] W. Xu, X. Li, J. Zhang, Z. Xue, J. Cao, Ultrasonic signal enhancement for coarse grain materials by machine learning analysis, Ultrasonics 117 (2021) 106550, https://doi.org/10.1016/j.ultras.2021.106550.
[166] K.S. Alguri, J.B. Harley, Transfer learning of ultrasonic guided waves using autoencoders: a preliminary study, AIP Conf. Proc. 2102 (2019) 050013, https://doi.org/10.1063/1.5099779.
[167] J.M. Ha, H.M. Seung, W. Choi, Autoencoder-based detection of near-surface defects in ultrasonic testing, Ultrasonics 119 (2022) 106637. doi:10.1016/j.ultras.2021.106637.

[168] A. Ebrahimkhanlou, B. Dubuc, S. Salamone, A generalizable deep learning framework for localizing and characterizing acoustic emission sources in riveted metallic panels, Mech. Syst. Signal. Pr. 30 (2019) 248–272, https://doi.org/10.1016/j.ymssp.2019.04.050.
[169] Y. Saeys, I. Inza, P. Larranaga, A review of feature selection techniques in bioinformatics, Bioinformatics 23 (19) (2007) 2507–2517, https://doi.org/10.1093/bioinformatics/btm344.
[170] Z.M. Hira, D.F. Gillies, A review of feature selection and feature extraction methods applied on microarray data, Adv. Bioinform. 2015 (2015) 1–13, https://doi.org/10.1155/2015/198363.
[171] V. Bolón-Canedo, N. Sánchez-Maroño, A. Alonso-Betanzos, A review of feature selection methods on synthetic data, Knowl. Inf. Syst. 34 (3) (2013) 483–519, https://doi.org/10.1007/s10115-012-0487-8.
[172] F.W. Margrave, K. Rigas, D.A. Bradley, P. Barrowcliffe, The use of neural networks in ultrasonic flaw detection, Meas.: J. Int. Meas. Confed. 25 (2) (1999) 143–154, https://doi.org/10.1016/S0263-2241(98)00075-X.
[173] S. Legendre, D. Massicote, J. Goyette, T.K. Bose, Neural classification of lamb wave ultrasonic weld testing signals using wavelet coefficients, IEEE T. Instrum. Meas. 50 (2001) 672–678, https://doi.org/10.1109/19.930439.
[174] M.A. Kewalramani, R. Gupta, Concrete compressive strength prediction using ultrasonic pulse velocity through artificial neural networks, Automat. Constr. 15 (3) (2006) 374–379, https://doi.org/10.1016/j.autcon.2005.07.003.
[175] Ó. Martín, M. López, F. Martín, Artificial neural networks for quality control by ultrasonic testing in resistance spot welding, J. Mater. Process. Tech. 183 (2-3) (2007) 226–233, https://doi.org/10.1016/j.jmatprotec.2006.10.011.
[176] R. Madandoust, R. Ghavidel, N. Nariman-zadeh, Evolutionary design of generalized GMDH-type neural network for prediction of concrete compressive strength using UPV, Comp. Mater. Sci. 49 (3) (2010) 556–567, https://doi.org/10.1016/j.commatsci.2010.05.050.
[177] S. Agarwal, M. Mitra, Lamb wave based automatic damage detection using matching pursuit and machine learning, Smart Mater. Struct. 23 (8) (2014) 085012, https://doi.org/10.1088/0964-1726/23/8/085012.
[178] M.M.F. Figueredo, J.L. Goncalves, A.M.V. Nakashima, A.M.F. Fileti, R.D. M. Carvalho, The use of an ultrasonic technique and neural networks for identification of the flow pattern and measurement of the gas volume fraction in multiphase flows, Exp. Therm. Fluid. Sci. 70 (2016) 29–50, https://doi.org/10.1016/j.expthermflusci.2015.08.010.
[179] Y. Nagatani, S. Okumura, S. Wu, Neural network based bone density estimation from the ultrasound waveforms inside cancellous bone derived by FDTD simulations, IEEE Int. Ultra. Sym. (2018) 8580010, https://doi.org/10.1109/ULTSYM.2018.8580010.
[180] J.Y. Park, Y.G. Yoon, T.K. Oh, Prediction of concrete strength with P-, S-, R-wave velocities by support vector machine (SVM) and artificial neural network (ANN), Appl. Sci.-Basel 9 9 (19) (2019) 4053, https://doi.org/10.3390/app9194053.
[181] K. Virupakshappa, E. Oruklu, Investigation of feature inputs for binary classification of ultrasonic NDT signals using SVM and neural networks, Midwest Symp. Circuit. (2019) 638–641, https://doi.org/10.1109/MWSCAS.2019.8884852.
[182] T.H.L. Nguyen, S. Park, Intelligent ultrasonic flow measurement using linear array transducer with recurrent neural networks, IEEE Access 8 (2020) 137564–137573, https://doi.org/10.1109/ACCESS.2020.3012037.
[183] S. Godfrey Nnabuife, B. Kuang, J.F. Whidborne, Z. Rana, Non-intrusive classification of gas-liquid flow regimes in an S-shaped pipeline riser using a Doppler ultrasonic sensor and deep neural networks, Chem. Eng. J. 403 (2021) 126401, https://doi.org/10.1016/j.cej.2020.126401.
[184] S.-H. Park, J.-Y. Hong, T. Ha, S. Choi, K.-Y. Jhang, Deep learning-based ultrasonic testing to evaluate the porosity of additively manufactured parts with rough surfaces, Metals-Basel 11 (2021) 290, https://doi.org/10.3390/met11020290.
[185] S. Lari, Y. Qian, H.-J. Kwon, Assessment of geometrical features of internal flaws with artificial neural network, Int. J. Precis. Eng. Man. 22 (5) (2021) 777–789, https://doi.org/10.1007/s12541-021-00515-z.
[186] W. Ren, N. Jin, L. OuYang, L. Zhai, Y. Ren, Gas volume fraction measurement of oil-gas-water three-phase flows in vertical pipe by combining ultrasonic sensor and deep attention network, IEEE T. Instrum. Meas. 70 (2021) 1–9, https://doi.org/10.1109/TIM.2020.3031186.
[187] Y. Qin, M. Ma, E. Zhu, Z. Mao, M. Haile, M. Shiao, T.-K. Chen, Temperature compensation of ultrasonic guided waves via recurrent neural network, Proceedings of the 7th Asia-Pacific Workshop on Structural Health Monitoring, APWSHM 2018 (2018) 402–409.
[188] Y. Guo, Z. Xiao, L. Geng, J. Wu, F. Zhang, Y. Liu, W. Wang, Fully convolutional neural network with GRU for 3D braided composite material flaw detection, IEEE Access 7 (2019) 151180–151188, https://doi.org/10.1109/ACCESS.2019.2946447.
[189] T. Gantala, K. Balasubramaniam, DPAI: A Data-driven simulation-assisted-Physics learned AI model for transient ultrasonic wave propagation, Ultrasonics 121 (2022) 106671, https://doi.org/10.1016/j.ultras.2021.106671.
[190] L. Huang, X. Hong, Z. Yang, Y. Liu, B. Zhang, CNN-LSTM network-based damage detection approach for copper pipeline using laser ultrasonic scanning, Ultrasonics 121 (2022) 106685, https://doi.org/10.1016/j.ultras.2022.106685.
[191] J. Melville, K.S. Alguri, C. Deemer, J.B. Harley, Structural damage detection using deep learning of ultrasonic guided waves, AIP Conf. Proc. 1949 (2018) 230004, https://doi.org/10.1063/1.5031651.
[192] M. de Oliveira, A. Monteiro, J. Vieira Filho, A new structural health monitoring strategy based on PZT sensors and convolutional neural network, Sensors (Basel) 18 (9) (2018) 2955, https://doi.org/10.3390/s18092955.
[193] C. Hu, B. Yang, J. Yan, Y. Xiang, S. Zhou, F.-Z. Xuan, Damage localization in pressure vessel by guided waves based on convolution neural network approach, J. Press. Vessel Technol. 142 (6) (2020), https://doi.org/10.1115/1.4047213.
[194] T. Zhang, Flow measurement of natural gas in pipeline based on 1d-convolutional neural network, Int. J. Comput. Int. Sys. 13 (2020) 1198–1206, https://doi.org/10.2991/ijcis.d.200803.002.
[195] H.J. Lim, H. Sohn, Online stress monitoring technique based on lamb-wave measurements and a convolutional neural network under static and dynamic loadings, Exp. Mech. 60 (2) (2020) 171–179, https://doi.org/10.1007/s11340-019-00546-8.
[196] Y. Li, K. Xu, Y. Li, F. Xu, D. Ta, W. Wang, Deep learning analysis of ultrasonic guided waves for cortical bone characterization, IEEE T Ultrason. Ferr. 68 (4) (2021) 935–951, https://doi.org/10.1109/TUFFC.2020.3025546.
[197] B. Filipovic, F. Milkovic, M. Subasic, S. Loncaric, T. Petkovic, M. Budimir, Automated ultrasonic testing of materials based on C-scan flaw classification, Int. Symp. Image. Sig. (2021) 230–234, https://doi.org/10.1109/ISPA52656.2021.9552056.
[198] R.J. Pyle, R.L.T. Bevan, R.R. Hughes, R.K. Rachev, A.A.S. Ali, P.D. Wilcox, Deep learning for ultrasonic crack characterization in NDE, IEEE T. Ultrason. Ferr. 68 (5) (2021) 1854–1865, https://doi.org/10.1109/TUFFC.2020.3045847.
[199] A. Chapon, D. Pereira, M. Toews, P. Belanger, Deconvolution of ultrasonic signals using a convolutional neural network, Ultrasonics 111 (2021) 106312, https://doi.org/10.1016/j.ultras.2020.106312.
[200] F. Gao, B. Li, L. Chen, Z. Shang, X. Wei, C. He, A softmax classifier for high-precision classification of ultrasonic similar signals, Ultrasonics 112 (2021) 106344, https://doi.org/10.1016/j.ultras.2020.106344.
[201] Z. Xiao, Q. Guo, Y. Guo, Y. Huang, Ultrasonic A-scan image detection for 3D braided composites based on convolutional neural network, ICCAI 2021 (2021) 8–14, https://doi.org/10.1145/3467707.3467709.
[202] K. Gopalakrishnan, M. Rautela, Y. Deng, Deep learning based identification of elastic properties using ultrasonic guided waves, Lect. Notes Civ. Eng 128 (2021) 77–90, https://doi.org/10.1007/978-3-030-64908-1_8.
[203] M. Rautela, J. Senthilnath, J. Moll, S. Gopalakrishnan, Combined two-level damage identification strategy using ultrasonic guided waves and physical knowledge assisted machine learning, Ultrasonics 115 (2021) 106451, https://doi.org/10.1016/j.ultras.2021.106451.
[204] I. Virkkunen, T. Koskinen, O. Jessen-Juhler, J. Rinta-aho, Augmented ultrasonic data for machine learning, J. Nondestruct. Eval. 40 (2021) 4, https://doi.org/10.1007/s10921-020-00739-5.
[205] A. Rai, M. Mitra, Lamb wave based damage detection in metallic plates using multi-headed 1-dimensional convolutional neural network, Smart Mater. Struct. 30 (3) (2021) 035010, https://doi.org/10.1088/1361-665X/abdd00.
[206] S. Mariani, Q. Rendu, M. Urbani, C. Sbarufatti, Causal dilated convolutional neural networks for automatic inspection of ultrasonic signals in non-destructive evaluation and structural health monitoring, Mech. Syst. Signal Pr. 157 (2021) 107748, https://doi.org/10.1016/j.ymssp.2021.107748.
[207] J. Park, S.-E. Lee, H.-J. Kim, S.-J. Song, S.-S. Kang, System invariant method for ultrasonic flaw classification in weldments using residual neural network, Appl. Sci.-Basel 12 (3) (2022) 1477, https://doi.org/10.3390/app12031477.
[208] T. Koskinen, I. Virkkunen, O. Siljama, O. Jessen-Juhler, The effect of different flaw data to machine learning powered ultrasonic inspection, J. Nondestruct. Eval. 40 (1) (2021), https://doi.org/10.1007/s10921-021-00757-x.
[209] O. Siljama, T. Koskinen, O. Jessen-Juhler, I. Virkkunen, Automated flaw detection in multi-channel phased array ultrasonic data using machine learning, J. Nondestruct. Eval. 40 (3) (2021), https://doi.org/10.1007/s10921-021-00796-4.
[210] Y. Cai, Y. Song, P. Ni, X. Liu, X. Li, Subwavelength ultrasonic imaging using a deep convolutional neural network trained on structural noise, Ultrasonics 117 (2021) 106552, https://doi.org/10.1016/j.ultras.2021.106552.
[211] R. Miorelli, C. Fisher, A. Kulakovskyi, B. Chapuis, O. Mesnil, O. D'Almeida, Defect sizing in guided wave imaging structural health monitoring using convolutional neural networks, NDT&E Int. 122 (2021) 102480, https://doi.org/10.1016/j.ndteint.2021.102480.
[212] J. Ye, N. Toyama, Automatic defect detection for ultrasonic wave propagation imaging method using spatio-temporal convolution neural networks, Struct. Health Monit. (2022), https://doi.org/10.1177/14759217211073503.
[213] H. Sun, L. Peng, S. Wang, S. Huang, K. Qu, Development of frequency-mixed point-focusing shear horizontal guided-wave EMAT for defect inspection using deep neural network, IEEE T. Instrum. Meas. 70 (2021) 1–14, https://doi.org/10.1109/TIM.2020.3033941.
[214] J.D. Eckels, E.M. Jacobson, I.T. Cummings, I.F. Fernandez, K. Ho, N. Dervilis, E. B. Flynn, A.J. Wachtor, Predicting local material thickness from steady-state ultrasonic wavefield measurements using a convolutional neural network, Ultrasonics 123 (2022) 106661, https://doi.org/10.1016/j.ultras.2021.106661.
[215] F. Gao, J. Hua, Damage characterization using CNN and SAE of broadband Lamb waves, Ultrasonics 119 (2022) 106592, https://doi.org/10.1016/j.ultras.2021.106592.
[216] A. Arcos Jiménez, C.Q. Gómez Muñoz, F.P. García Márquez, Dirt and mud detection and diagnosis on a wind turbine blade employing guided waves and supervised learning classifiers, Reliab. Eng. Syst. Safe 184 184 (2019) 2–12, https://doi.org/10.1016/j.ress.2018.02.013.
[217] L.F.M. Rodrigues, F.C. Cruz, M.A. Oliveira, E.F. Simas Filho, M.C.S. Albuquerque, I.C. Silva, C.T.T. Farias, Carburization level identification in industrial HP pipes using ultrasonic evaluation and machine learning, Ultrasonics 94 (2019) 145–151, https://doi.org/10.1016/j.ultras.2018.10.005.

[218] K. Lee, V. Estivill-Castro, Feature extraction and gating techniques for ultrasonic shaft signal classification, Appl. Soft Comput. J. 7 (1) (2007) 156–165, https://doi.org/10.1016/j.asoc.2005.05.003.
[219] M. Cacciola, S. Calcagno, F.C. Morabito, M. Versaci, Computational intelligence aspects for defect classification in aeronautic composites by using ultrasonic pulses, IEEE T. Ultrason. Ferr. 55 (4) (2008) 870–878, https://doi.org/10.1109/TUFFC.2008.722.
[220] S. Saechi, W. Kongprawechnon, R. Sahamitmongkol, Test system for defect detection in construction materials with ultrasonic waves by support vector machine and neural network, in: 6th International Conference on Soft Computing and Intelligent Systems, and 13th International Symposium on Advanced Intelligence Systems, 2012, pp. 1034–1039, https://doi.org/10.1109/SCIS-ISIS.2012.6505090.
[221] L.H. Lee, R. Rajkumar, L.H. Lo, C.H. Wan, D. Isa, Oil and gas pipeline failure prediction system using long range ultrasonic transducers and euclidean-support vector machines classification approach, Exp. Syst. Appl. 40 (6) (2013) 1925–1934, https://doi.org/10.1016/j.eswa.2012.10.006.
[222] H. Zhao, L. Peng, T. Takahashi, T. Hayashi, K. Shimizu, T. Yamamoto, Support vector regression-based data integration method for multipath ultrasonic flowmeter, IEEE T. Instrum. Meas. 63 (12) (2014) 2717–2725, https://doi.org/10.1109/TIM.2014.2326276.
[223] P. Yang, Q. Li, Wavelet transform-based feature extraction for ultrasonic flaw signal classification, Neural. Comput. Appl. 24 (3-4) (2014) 817–826, https://doi.org/10.1007/s00521-012-1305-7.
[224] H. Zamani Hosseinabadi, R. Amirfattahi, B. Nazari, H.R. Mirdamadi, S. A. Atashipour, GUW-based structural damage detection using WPT statistical features and multiclass SVM, Appl. Acoust. 86 (2014) 59–70, https://doi.org/10.1016/j.apacoust.2014.05.002.
[225] Y.-F. Shih, Y.-R. Wang, K.-L. Lin, C.-W. Chen, Improving non-destructive concrete strength tests using support vector machines, Materials 8 (2015) 7169–7178, https://doi.org/10.3390/ma8105368.
[226] Z. Yang, Q.u. Zhou, X. Wu, Z. Zhao, A novel measuring method of interfacial tension of transformer oil combined PSO optimized SVM and multi frequency ultrasonic technology, IEEE Access 7 (2019) 182624–182631, https://doi.org/10.1109/ACCESS.2019.2954899.
[227] H. Xiao, D. Chen, J. Xu, S. Guo, Defects identification using the improved ultrasonic measurement model and support vector machines, NDT&E Int. 111 (2020) 102223, https://doi.org/10.1016/j.ndteint.2020.102223.
[228] J.-G. Minonzio, B. Cataldo, R. Olivares, D. Ramiandrisoa, R. Soto, B. Crawford, V. H.C. De Albuquerque, R. Munoz, Automatic classifying of patients with non-traumatic fractures based on ultrasonic guided wave spectrum image using a dynamic support vector machine, IEEE Access 8 (2020) 194752–194764, https://doi.org/10.1109/ACCESS.2020.3033480.
[229] R. Miorelli, A. Kulakovskyi, B. Chapuis, O. D'Almeida, O. Mesnil, Supervised learning strategy for classification and regression tasks applied to aeronautical structural health monitoring problems, Ultrasonics 113 (2021) 106372, https://doi.org/10.1016/j.ultras.2021.106372.
[230] C.A. Lindley, S. Beamish, R.S. Dwyer-Joyce, N. Dervilis, K. Worden, A Bayesian approach for shaft centre localisation in journal bearings, Mech. Syst. Signal. Pr. 174 (2022) 109021, https://doi.org/10.1016/j.ymssp.2022.109021.
[231] M.R. Jones, T.J. Rogers, K. Worden, E.J. Cross, A Bayesian methodology for localising acoustic emission sources in complex structures mech, Syst. Signal. Pr. 163 (2022) 108143, https://doi.org/10.1016/j.ymssp.2021.108143.
[232] M. Haywood-Alexander, N. Dervilis, K. Worden, E.J. Cross, R.S. Mills, T.J. Rogers, Structured machine learning tools for modelling characteristics of guided waves, Mech. Syst. Signal. Pr. 156 (2021) 107628, https://doi.org/10.1016/j.ymssp.2021.107628.
[233] M. Haywood-Alexander, N. Dervilis, K. Worden, G.B. Dobie, T.J. Rogers, Decomposition of multi-mode signals using dispersion curves and Bayesian linear regression, Proc. SPIE (2021), https://doi.org/10.1117/12.2582967.
[234] Z. Ma, W. Zhang, Z. Luo, X.u. Sun, Z. Li, L.i. Lin, Ultrasonic characterization of thermal barrier coatings porosity through BP neural network optimizing Gaussian process regression algorithm, Ultrasonics 100 (2020) 105981, https://doi.org/10.1016/j.ultras.2019.105981.
[235] M.A. Oliveira, E.F. Simas Filho, M.C.S. Albuquerque, Y.T.B. Santos, I.C. da Silva, C.T.T. Farias, Ultrasound-based identification of damage in wind turbine blades using novelty detection, Ultrasonics 108 (2020) 106166, https://doi.org/10.1016/j.ultras.2020.106166.
[236] L. Posilović, D. Medak, F. Milković, M. Subašić, M. Budimir, S. Lončarić, Deep learning-based anomaly detection from ultrasonic images, Ultrasonics 124 (2022) 106737, https://doi.org/10.1016/j.ultras.2022.106737.
[237] F. Milkovic, B. Filipovic, M. Subasic, T. Petkovic, S. Loncaric, M. Budimir, Ultrasound anomaly detection based on variational autoencoders, Int. Symp. Image. Sig. (2021) 225–229, https://doi.org/10.1109/ISPA52656.2021.9552041.
[238] I. Kraljevski, F. Duckhorn, C. Tschope, M. Wolff, machine learning for anomaly assessment in sensor networks for NDT in aerospace, IEEE Sens. J. 21 (9) (2021) 11000–11008, https://doi.org/10.1109/JSEN.2021.3062941.
[239] A. Bowler, M. Pound, N. Watson, Convolutional feature extraction for process monitoring using ultrasonic sensors, Comput. Chem. Eng. 155 (2021) 107508, https://doi.org/10.1016/j.compchemeng.2021.107508.
[240] S. Zhang, C.M. Li, W. Ye, Damage localization in plate-like structures using time-varying feature and one-dimensional convolutional neural network, Mech. Syst. Signal Pr. 147 (2021) 107107, https://doi.org/10.1016/j.ymssp.2020.107107.
[241] T. Latête, B. Gauthier, P. Belanger, Towards using convolutional neural network to locate, identify and size defects in phased array ultrasonic testing, Ultrasonics 115 (2021) 106436, https://doi.org/10.1016/j.ultras.2021.106436.
[242] D. Sen, A. Aghazadeh, A. Mousavi, S. Nagarajaiah, R. Baraniuk, A. Dabak, Data-driven semi-supervised and supervised learning algorithms for health monitoring of pipes, Mech. Syst. Signal. Pr. 131 (2019) 524–537, https://doi.org/10.1016/j.ymssp.2019.06.003.
[243] A.E. Bouzenad, M.E. Mountassir, S. Yaacoubi, F. Dahmene, M. Koabaz, L. Buchheit, W. Ke, A semi-supervised based k-means algorithm for optimal guided waves structural health monitoring: a case study, Inventions 4 (2019) 17, https://doi.org/10.3390/inventions4010017.
[244] L.A. Bull, K. Worden, N. Dervilis, Towards semi-supervised and probabilistic classification in structural health monitoring, Mech. Syst. Signal. Pr. 140 (2020) 106653, https://doi.org/10.1016/j.ymssp.2020.106653.
[245] L.A. Bull, T.J. Rogers, C. Wickramarachchi, E.J. Cross, K. Worden, N. Dervilis, Probabilistic active learning: an online framework for structural health monitoring, Mech. Syst. Signal. Pr. 134 (2019) 106294, https://doi.org/10.1016/j.ymssp.2019.106294.
[246] A.J. Hughes, L.A. Bull, P. Gardner, R.J. Barthorpe, N. Dervilis, K. Worden, On risk-based active learning for structural health monitoring, Mech. Syst. Signal. Pr. 167 (2022) 108569, https://doi.org/10.1016/j.ymssp.2021.108569.
[247] L. Posilović, D. Medak, M. Subašić, M. Budimir, S. Lončarić, Generative adversarial network with object detector discriminator for enhanced defect detection on ultrasonic B-scans, Neurocomputing 459 (2021) 361–369, https://doi.org/10.1016/j.neucom.2021.06.094.
[248] L. Posilović, D. Medak, M. Subašić, M. Budimir, S. Lončarić, S., Generating ultrasonic images indistinguishable from real images using Generative Adversarial Networks. Ultrasonics 119 (2022). doi:10.1016/j.ultras.2021.106610.
[249] T. Gantala, K. Balasubramaniam, Automated defect recognition for welds using simulation assisted TFM imaging with artificial intelligence, J. Nondestruct. Eval. 40 (1) (2021), https://doi.org/10.1007/s10921-021-00761-1.
[250] K. Virupakshappa, E. Oruklu, Using generative adversarial networks to generate ultrasonic signals, IEEE Int. Ultra. Sym. (2020), https://doi.org/10.1109/IUS46767.2020.9251382.
[251] A. Look, O. Kirschner, S. Riedelbauch, Building robust classifiers with generative adversarial networks for detecting cavitation in hydraulic turbines, ICPRAM 2018 (2018) 456–462, https://doi.org/10.5220/0006636304560462.
[252] J. Ye, N. Toyama, Benchmarking deep learning models for automatic ultrasonic imaging inspection, IEEE Access 9 (2021) 36986–36994, https://doi.org/10.1109/ACCESS.2021.3062860.
[253] G.P. Tsialiamanis, D.J. Wagg, P.A. Gardner, N. Dervilis, K. Worden, On partitioning of an SHM problem and parallels with transfer learning, Conf. Proc. Soc. Exp. Mech. Ser. (2021) 41–50, https://doi.org/10.1007/978-3-030-47717-2_5.
[254] D.F. Hesser, S. Mostafavi, G.K. Kocur, B. Markert, Identification of acoustic emission sources for structural health monitoring applications based on convolutional neural networks and deep transfer learning, Neurocomputing 453 (2021) 1–12, https://doi.org/10.1016/j.neucom.2021.04.108.
[255] B. Ren, J. Chen, Fracture acoustic emission signals identification of broken wire using deep transfer learning and wavelet analysis, ACM Int. Conf. Proc. Ser. (2021), https://doi.org/10.1145/3469213.3470230.
[256] G. Reyes-Carmenaty, M.A. Pérez, Use of transfer learning for detection of structural alterations, Proc. Comput. Sci. 200 (2022) 1368–1377, https://doi.org/10.1016/j.procs.2022.01.338.
[257] M. Azimi, G. Pekcan, Structural health monitoring using extremely compressed data through deep learning, Comput.-Aid. Civil Infrastruct. Eng. 35 (6) (2020) 597–614.
[258] A.L. Bowler, M.P. Pound, N.J. Watson, Domain adaptation and federated learning for ultrasonic monitoring of beer fermentation, Fermentation 7 (2021) 253, https://doi.org/10.3390/fermentation7040253.
[259] X. Gao, Y. Shi, Q.i. Zhu, Z. Li, H.u. Sun, Z. Yao, W. Zhang, Domain adaptation in intelligent ultrasonic logging tool: from microseismic to pulse-echo, IEEE T. Instrum. Meas. 70 (2021) 1–14, https://doi.org/10.1109/TIM.2021.3050154.
[260] K.S. Alguri, C.C. Chia, J.B. Harley, Sim-to-Real: Employing ultrasonic guided wave digital surrogates and transfer learning for damage visualization, Ultrasonics 111 (2021) 106338, https://doi.org/10.1016/j.ultras.2020.106338.
[261] L.A. Bull, P.A. Gardner, N. Dervilis, E. Papatheou, M. Haywood-Alexander, R. S. Mills, K. Worden, On the transfer of damage detectors between structures: an experimental case study, J. Sound Vib. 501 (2021) 116072, https://doi.org/10.1016/j.jsv.2021.116072.
[262] P. Gardner, L.A. Bull, N. Dervilis, K. Worden, Overcoming the problem of repair in structural health monitoring: metric-informed transfer learning, J. Sound Vib. 510 (2021) 116245, https://doi.org/10.1016/j.jsv.2021.116245.
[263] P. Gardner, L.A. Bull, N. Dervilis, K. Worden, Domain-adapted Gaussian mixture models for population-based structural health monitoring, J. Civ. Struct. Health Monit. (2022), https://doi.org/10.1007/s13349-022-00565-5.
[264] P. Gardner, L.A. Bull, N. Dervilis, K. Worden, On the application of kernelised Bayesian transfer learning to population-based structural health monitoring, Mech. Syst. Signal Process. 167 (2022) 108519, https://doi.org/10.1016/j.ymssp.2021.108519.
[265] P. Gardner, L.A. Bull, J. Gosliga, J. Poole, N. Dervilis, K. Worden, A population-based SHM methodology for heterogeneous structures: Transferring damage localisation knowledge between different aircraft wings, Mech. Syst. Signal. Pr. 172 (2022) 108918, https://doi.org/10.1016/j.ymssp.2022.108918.

[266] D. Medak, L. Posilovic, M. Subasic, M. Budimir, S. Loncaric, Automated defect detection from ultrasonic images using deep learning, IEEE T. Ultrason. Ferr. 68 (10) (2021) 3126–3134, https://doi.org/10.1109/TUFFC.2021.3081750.

[267] L. Posilovic, D. Medak, M. Subasic, T. Petkovic, M. Budimir, S. Loncaric, Flaw detection from ultrasonic images using YOLO and SSD, Int. Symp. Image. Sig. (2019) 163–168, https://doi.org/10.1109/ISPA.2019.8868929.

[268] D. Medak, L. Posilović, M. Subašić, M. Budimir, S. Lončarić, DefectDet: a deep learning architecture for detection of defects with extreme aspect ratios in ultrasonic images, Neurocomputing 473 (2022) 107–115, https://doi.org/10.1016/j.neucom.2021.12.008.

[269] D. Medak, L. Posilovic, M. Subasic, M. Budimir, S. Loncaric, Deep learning-based defect detection from sequences of ultrasonic B-scans, IEEE Sens. J. 22 (3) (2022) 2456–2463, https://doi.org/10.1109/JSEN.2021.3134452.

[270] A.B. Arrieta, N. Díaz-Rodríguez, J. Del Ser, A. Bennetot, S. Tabik, A. Barbado, S. Garcia, S. Gil-Lopez, D. Molina, R. Benjamins, R. Chatila, F. Herrera, Explainable Artificial Intelligence (XAI): Concepts, taxonomies, opportunities and challenges toward responsible AI, Inform. Fusion 58 (2020) 82–115, https://doi.org/10.1016/j.inffus.2019.12.012.

[271] P. Linardatos, V. Papastefanopoulos, S. Kotsiantis, Explainable AI: a review of machine learning interpretability methods, Entropy 23 (2021) 18, https://doi.org/10.3390/e23010018.

3 Methodology

This section presents a cohesive methodology of the works in this thesis. Section 3.1 describes the experimental datasets used throughout the works and Section 3.2 explains the ML methodology used.

3.1 Experimental datasets

For all experiments, a US box (Lecoeur Electronique) was used to excite the transducers and digitise the received sound waves. Temperature sensors (RTD, PT1000) were connected to a PT-104 Data Logger (Pico Technology). The US box and temperature data logger were connected to a laptop and a bespoke MATLAB software controlled the hardware components and acquired the data.

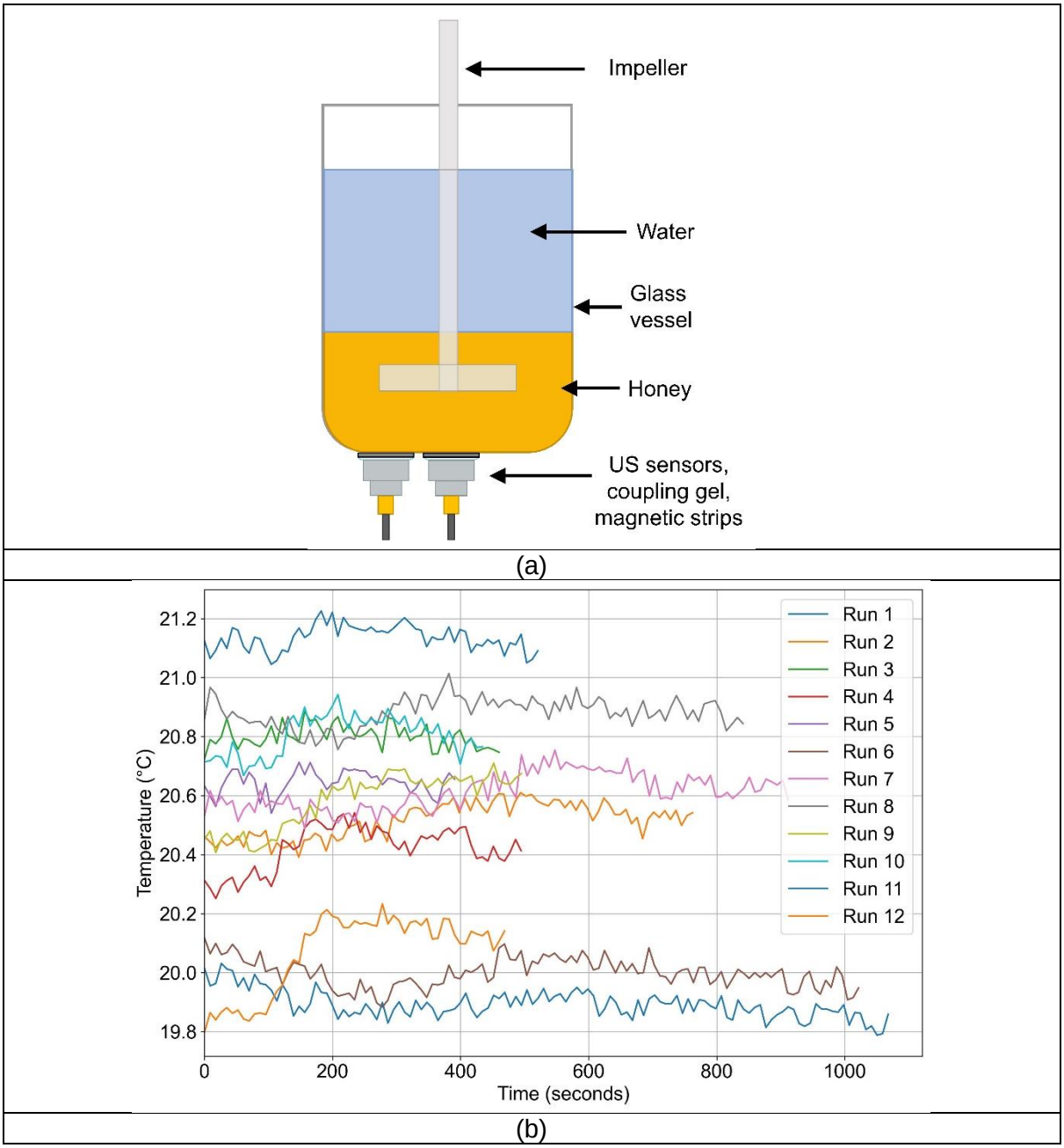
US waveforms can be affected by two phenomena: Firstly, the magnitude of the waveform can be altered by changing acoustic impedance or attenuation. Secondly, variations in sound velocity alter the displacement of the waveform in the time domain (Henning and Rautenburg, 2006). In this thesis, the processes monitored were honey-water blending, flour-water batter mixing, cleaning of food fouling from pipe sections, and alcoholic beer fermentation. During blending, mixing, and cleaning, the acoustic impedance at the measurement areas change as the material composition varies. Furthermore, conducting these processes over a range of temperatures alters the sound velocity through the materials. Throughout alcoholic beer fermentation, the acoustic impedance and speed of sound of a transmitted US wave change as the density of the wort decreases (Bowler et al., 2021). Moreover, attenuation increases as CO_2 bubbles are produced. Therefore, the works in this thesis investigate the full range of impacts to US waveforms during process monitoring. This means that the feature extraction methods developed in the ML pipeline and domain adaptation methods were created using all possible impacts to US waveforms. Furthermore, the methods in this thesis can be applied to both stationary and evolving processes. Processes which evolve over time, such as those investigated in this thesis, require information from previous timesteps as inputs to the ML model. In these instances, the LSTM layers used in the final ML pipeline may be used to learn process trajectories. However, if only information from the current timestep is required, then the feature extraction methodologies developed in this thesis can be applied into neural network architectures.

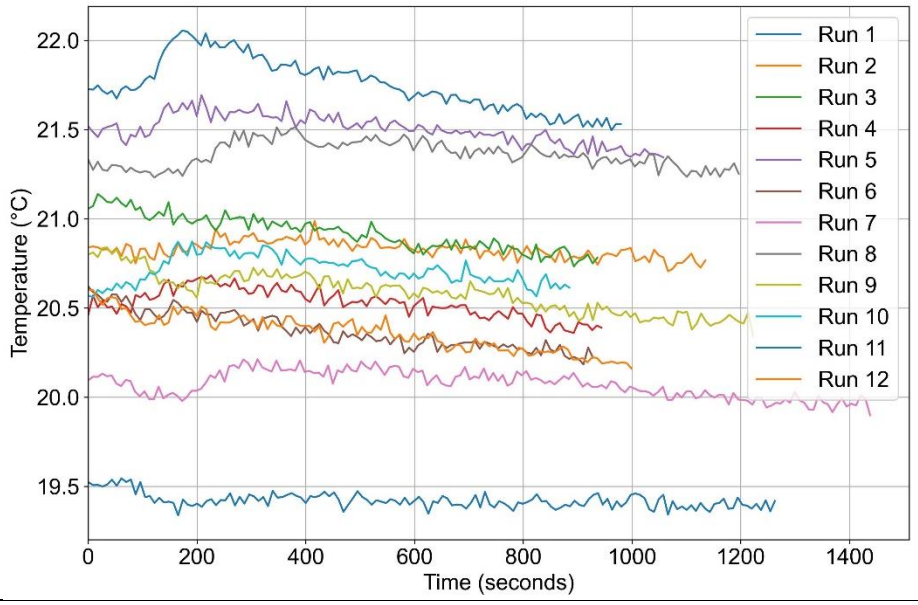
3.1.1 Honey-water blending

Two datasets for the honey-water blending experiments were collected to enable investigation of transfer learning between datasets from similar processes. These datasets were collected by Alex Bowler during 2019. Two magnetic US sensors (5 MHz central resonance, M1057, Olympus) were externally attached to the base of a mixing vessel (Figure 1). A reflection-mode, pulse-echo, non-invasive sensing technique was used to monitor the sound wave reflected from the interface between the vessel wall and the mixture. To the author's knowledge, only two works have used reflection-mode, non-invasive US sensing to monitor mixing processes. Buurman, Resoort, and Plaschkes (1986) used a non-invasive US Doppler sensor to detect whether particles were suspended at the bottom of a mixing vessel and Zhan et al. (2016) used a non-invasive pulse–echo transducer attached to the base of the vessel as part of their study to monitor particle suspension. Reflection-mode, non-invasive US sensing measures the change in acoustic properties at the vessel-material interface. It is therefore more industrially relevant than invasive sensing as it can be externally retrofitted to process equipment and does not require transmission of the sound wave through the mixture. The same sensing technique was used to monitor flour-water batter mixing, as outlined in Section 3.1.2.

A 250 ml glass mixing vessel was used. The sensing technique is viable to monitor the development of homogeneity for any scale of mixing process, however, only at the sensor measurement area. Therefore, during industrial deployment, the location of the sensors must be chosen to monitor useful mixing phenomena inside the vessel. The sensors were attached to adhesive magnetic strips on the outside of the vessel and coupling gel (Proceq ultrasound couplant) was applied between the sensor and strip. An overhead stirrer was used to stir the mixture. One sensor was attached in the centre of the vessel base and another sensor was attached approximately 2 cm offset from the centre as this was half the distance between the centre of the vessel and the vessel wall. This allowed monitoring of two areas of mixing phenomena to compare between the data obtained from different regions of the vessel. The sensors were attached to the base of the vessel as this was a flat surface in contrast with the curved vessel sides. This enabled enhanced transmission of the US wave through the vessel walls. The reduced curvature at larger scales would enable the attachment of the sensors to the vessel sides. The temperature sensor was also attached externally to the base of the vessel. US signals were acquired continuously for 1 s for each probe consecutively. Two different volumes of pure clear honey (Wm Morrison Supermarkets plc) were used: 20 and 30 ml. The volume of tap water used was 200 ml for all runs. Both the honey and water were loaded into the vessel before mixing was started. The impeller speed was set to either 200 or 250 rpm. These four parameter permutations were repeated three times, producing a set of 12 runs. The temperature ranged between 19.3 and 22.1 °C. Figures 1b and 1c display the degree of temperature variability between this range. The environmental temperature was varied throughout the data collection by controlling the laboratory thermostat set point. These parameter ranges were selected to create variability between the experiment repeats and encourage generalisation of the trained ML models whilst retaining a mixing time that enabled collection of US datapoints. The time length until mixing was complete varied between 200 (45 datapoints) and 1140 seconds (165 datapoints). This methodology was repeated across two days, producing the two datasets. Justification of the data collected in this thesis is provided in Section 9 Discussion. Between collection of the two datasets, the US sensors were removed and reattached to obtain different US waveforms. A video camera was used to film the mixing process. The labelled data was obtained by viewing the videos to determine the time where mixing was complete for each run. In an industrial setting, labelled data would need to be collected through using additional sensing techniques to follow the mixing trajectory or periodic sampling of the mixtures (Bowler et al., 2022). As honey is completely miscible in water, the honey-water blending dataset is representative of developing homogeneity in liquid–liquid blending.

The aim of this thesis was to develop ML methods to facilitate optimal deployment of US sensors for process monitoring applications in industrial environments. Therefore, the temperature range chosen, spanning from 19.3 to 22.1 °C, is used to draw conclusions about the relative strengths of each ML method investigated rather than to reflect actual industrial temperature spectra. The choice of temperature range and granularity is therefore justified as the emphasis of this thesis is comparative evaluation of ML methods rather than achieving high accuracy in an industrial application. The results drawn from this thesis are indicative of differing temperature ranges between the two mixing processes and provide insight into which ML methods may be most generalisable to varying temperature ranges in industrial scenarios.


Impeller
Water
Glass vessel
Honey
US sensors, coupling gel, magnetic strips
(a)
Temperature (°C)
Time (seconds)
Run 1
Run 2
Run 3
Run 4
Run 5
Run 6
Run 7
Run 8
Run 9
Run 10
Run 11
Run 12
(b)



(c)

Figure 1a: A diagram of the honey-water blending process. Adapted from Bowler et al. (2021). Figure 1b and 1c: The environmental temperature range and degree of temperature variability between the range for each of the honey-water blending datasets collected. The temperature during the first dataset (Figure 1b) ranges between 19.8 to 21.2°C and the temperature of the second dataset ranges between 19.3 to 22.1°C.

3.1.2 Flour-water batter mixing

The flour-water batter mixing dataset was collected by Alex Bowler during 2019. Two magnetic US sensors (5 MHz resonance, M1057, Olympus) were externally attached to a stand mixer glass mixing bowl (1000 W Kenwood kmix kmx754) using magnetic tape and silicone vacuum grease. A temperature sensor was also attached to the outside of the mixing bowl to monitor the environmental temperature. A reflection-mode, pulse-echo sensing technique monitored the sound wave reflected from the interface between the mixing bowl and the mixture. US signals were continuously acquired for 1 s for each probe consecutively and, on average, produced two waveforms. The quantity of strong white flour (Wm Morrison Supermarkets plc) and tap water was varied between 450 and 450 g, 500 and 450 g, and 500 and 400 g, respectively. Each combination was repeated three times producing a total of nine runs. These parameter combinations led to variability in process length between 720 and 1320 s. The variation in composition was used to ensure variation in the US measurements acquired for each run to enable the evaluation of ML model generalisability. Below 400 g of tap water, the batter began to pull away from the sides where the sensors were located. Above a water quantity of 450 g, the gluten network failed to develop. Measuring the power or torque supplied to the impeller is a common method of monitoring dough mixing. Mixing should be stopped at the maximum power input for optimal gluten network development and therefore bread properties (Perez Alvarado, Hussein, and Becker, 2016). Beyond this point of maximum resistance to extension, the gluten network begins to breakdown. The optimal mixing time was obtained by determining the time of maximum power input to the impeller. This was measured using a YouThink plug socket power meter.

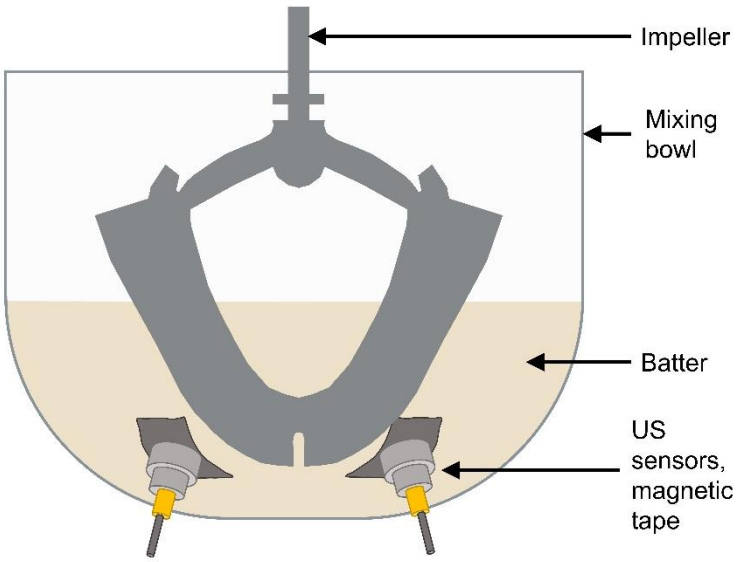


Figure 2: A diagram of the flour-water mixing process. Adapted from Bowler et al. (2021).

3.1.3 Cleaning of food fouling from pipe sections

These datasets were collected by Josep Escrig and have been used in Escrig et al. (2019, 2020a, and 2020b). Three pipe test sections were used: A rectangular rig with a SS340 base plate and clear, PMMA sides; a circular pipe section constructed from clear PMMA; and an opaque, circular pipe section constructed from SS316. This enabled collection of US waveform from different pipe materials and geometries including circular pipes as commonly used industrially. Three different food materials were used to foul the pipe test sections: tomato paste, concentrated malt, and gravy. This allowed for collection of US data for materials with different cleaning characteristics. Tomato paste and gravy were cleaned by mechanical forces whereas the malt extract dissolved into the cleaning solution (Escrig et al., 2019). The fouling material was then spread with a spatula to form a layer of approximately 5 mm thickness and left for 10 min to dry. It was placed in the centre of the base plate for the rectangular rig and 30 mm from the exit for the circular pipes. The temperature of the water used for cleaning was set at either 12 °C or 45 °C and a flowrate of 6 l/s was used. This produced mean cleaning times ranging between 2:12 (malt, 45°C) and 67:41 (gravy, 12°C) minutes. For the rectangular test section, a magnetic sensor (5 MHz resonance, M1057, Olympus) was externally attached to the base plate. For the circular pipe sections, the US transducers (2 MHz, Yushi, 2P10N) were glued externally to the bottom of the pipes in the location where the fouling material would be placed. The temperature sensors were also attached externally, in close proximity to the US sensors. A camera was used to determine the time at which all the fouling material was removed. The position of the camera was moved depending on whether the pipe section was clear or opaque. In an industrial setting, the labelled data could be acquired through either disassembling the pipe section to inspect cleanliness, using additional sensing techniques to monitor the concentration of fouling in downstream cleaning solution, or transfer learning from similar processes (Bowler et al., 2022). The US and temperature data was recorded every 4 s producing 4 US waveforms. A reflection-mode, pulse-echo sensing technique was used to monitor the waveform reflected from the interface between the pipe wall and the fouling material. To the author's knowledge only Chen et al. (2019) has used a non-invasive, reflection-mode US sensing technique as used to collect these datasets. The camera images were recorded every 20 s. A minimum of 7 repeats were conducted for every

permutation of pipe test section, fouling material and fluid temperature, producing 93 runs in total.

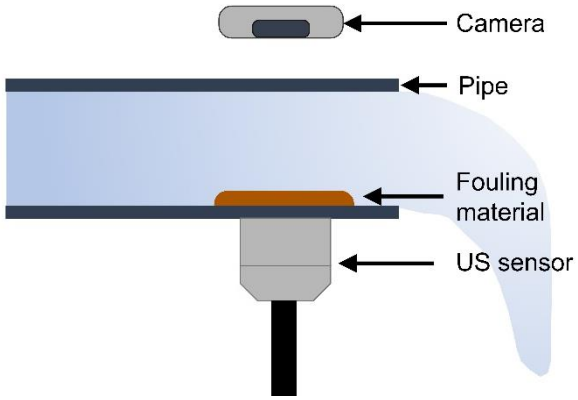


Figure 3: A diagram of the cleaning process to remove food fouling from pipe sections. Adapted from Bowler et al. (2021).

3.1.4 Laboratory scale fermentations

The laboratory scale fermentation dataset was collected by Josep Escrig. The fermentations were conducted in a 30 L cylindrical plastic vessel. A lid sealed the vessel to protect the wort from contamination. The lid contained an air lock to release the CO_2 produced during fermentation. A belt heater increased the temperature of the wort to facilitate fermentation. The wort was prepared in the vessel by dissolving and mixing 1.5 kg of malt (Coopers Real Ale, UK) and 1 kg of sugar (brewing sugar, the Home Brew Shop, UK) in 22 L of water. Once the ingredients were mixed, an invasive US probe was installed, consisting of a US transducer (Sonatest, 2 MHz central frequency, UK) and a temperature sensor. Coupling gel was applied between the US transducer and the probe, and a spring was used to maintain the contact pressure. The US sensor monitored up to three sound wave reflections. Firstly, a sound wave that travelled through the coupling gel, along the probe material, and reflected from the probe-wort interface. Secondly, a reverberation of this sound wave path. Thirdly, a reflection from the far probe wall where the sound wave has travelled through the fermenting wort. Using an invasive probe allowed comparison between machine learning models that used either the first or second and third waveform reflections.

A Tilt hydrometer was installed to provide real-time density measurements. This device was a small cylinder that floats in the liquid with its centre of gravity different from its centre of buoyancy. This causes an inclination of the device that is dependent on the specific gravity of the fermenting media. The inclination of the hydrometer was measured by a self-contained accelerometer and was transmitted by radio to a smartphone located outside of the vessel. A calibration procedure related the inclination to the specific gravity.

The yeast (Coopers Real Ale, UK) was distributed on the surface and the vessel sealed. The mixture was left for 4 to 7 days while the fermentation occurred. After this time, the fermentation equipment was cleaned and a new batch was prepared. In total, 13 batches were completed. During fermentation, data was collected from the three different sensors: the US sensor, the temperature sensor, and the hydrometer. The fermentation batches were conducted over a period of approximately 3 months. The environmental and water temperature in the laboratory changed during this time and the belt heater was only in contact with the lower section of the vessel. This resulted in the temperature ranging between approximately 20 and 30 °C. Sets of US and temperature data were collected periodically (at regular intervals where 200 s elapsed between each set of collected data).

Each of the sets consisted of 36 US waves and 36 temperature readings. The time between each waveform acquisition was 0.55 s.

To the author's knowledge, as reviewed in Section 2, only Meng et al. (2012) has used a non-invasive, reflection-mode US sensing technique to monitor fermentation. They used a single sensor to monitor yogurt fermentation by monitoring the change in acoustic impedance at the wall-yogurt interface. Although an invasive probe was used to monitor the alcoholic fermentation of beer during the collection of this dataset, this thesis (Section 5, Predicting Alcohol Concentration during Beer Fermentation Using Ultrasonic Measurements and Machine Learning) evaluates the potential of a non-invasive, reflection-mode US sensor by only using features reflected from the probe-wort interface and omitting the process temperature from the ML models.

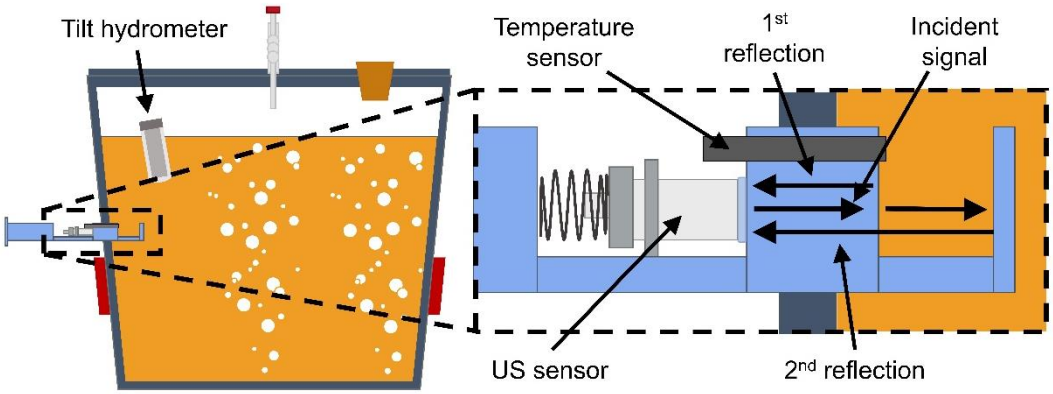


Figure 4: A diagram of the laboratory alcoholic beer fermentation process. Adapted from Bowler et al. (2021).

3.1.5 Industrial scale fermentations

The industrial scale fermentation dataset was collected by Nicholas Watson. Five fermentations were monitored in a 2000 L industrial scale fermenter at the Totally Brewed brewery in Nottingham, UK. Three different beers were monitored: three fermentations consisting of Slap in the Face, one Guardian of the Forest, and one 4 Hopmen of the Apocalypse. The same US probe was used to monitor both the laboratory and industrial scale fermentation processes. Samples were removed every two hours (except overnight) and the wort density was measured using a hydrometer. The temperature was decreased once the desired wort density was reached. Blocks of US and temperature data were collected periodically (at regular intervals where 200 s elapsed between each block of collected data). Each of the blocks consisted of 36 US waveforms and 36 temperature readings. The time between each waveform acquisition was 0.55 s.

3.1.6 Summary of experimental datasets

Table 1: A summary of the experimental datasets investigated in this thesis.

	Honey-water blending	**Flour-water batter mixing**	**Cleaning**	**Alcoholic beer fermentation**
Sensors	• Two • 5 MHz central frequency	• Two • 5 MHz central frequency	• One • 2 or 5 MHz central frequency	• One • 2 MHz central frequency
Sensor attachment	• Non-invasive	• Non-invasive	• Non-invasive	• Invasive probe

	• Magnetic sensors attached to magnetic adhesive strip	• Magnetic sensors attached to magnetic adhesive strip	• Magnetic sensors attached to metal pipe • Non-magnetic sensors glued to pipe	
Number of datasets	Two	One	One	Two
Equipment	• 250 ml glass mixing vessel • Overhead stirrer	• 5 l mixing bowl • Stand mixer	• Three pipe sections (rectangular with SS340 base plate, circular PMMA, and circular SS316)	• 30 l fermentation vessel • 2000 l fermentation vessel
Length of process	200 – 1140 s	720 – 1320 s	2:12 to 67:41 min	4 – 7 days
Number of runs	12	9	93	18
Number of datapoints	1493 and 921	999	22,207	61,220 and 11,920
Class balance	19.0 % mixed and 81.0 % not mixed, 28.9 % mixed and 71.1 % not mixed	30.3 % mixed and 69.7 % not mixed	56.0 % clean and 44.0 % not clean	Start of ethanol production: 89.6 % started and 10.4 % not started, 87.8 % started and 12.2 % not started End of ethanol production: 52.6 % ended and 47.4 % not ended, 54.9 % ended and 45.1 % not ended
Temperature range (°C)	19.3 - 22.1	19.4 - 21.3	12 – 45	10 – 30
Parameter range	• 20 – 30 ml honey • 200 – 250 rpm impeller speed	• 450 – 500 g flour • 400 – 450 g tap water	• Malt extract, gravy, tomato paste	-
Data labelling	Video camera used to determine the time at which the materials were fully mixed	Measurement of power output from impeller	Video camera used to determine the time at which the pipe was fully cleaned	Density measurements using in-line sensors and sampling

3.1.6.1 Dataset size for machine learning

For linear regression, a rule of thumb is that the size of the dataset should be at least 10 times larger than the number of features. It has been suggested that this should be increased to at least 50 times for neural networks (Alwosheel et al., 2018). Using the literature reviewed in Section 2 (Bowler et al., 2022), the median number of features used with neural networks and US data is seven (Section 2, Table 3). This suggests a dataset size of 350 samples is required to model the average task in the literature. All datasets used in this thesis have a greater number of samples than this (Table 1), suggesting that they are sufficient for training machine learning models. However, the number of data points required depends on many factors such as complexity of the task or desired model accuracy. Nonetheless, the machine learning models developed in this thesis are used to compare between machine learning techniques opposed to being deployed in final applications. Therefore, each model is trained on the same datasets to those with which it is compared. For further discussion about the data collection required to justify the conclusions developed in this thesis, see Section 9 'Discussion'.

3.2 Machine learning

This section outlines the feature extraction, feature selection, algorithm and hyperparameter choice used throughout the works in this thesis which are outlined in Table 2. Furthermore, labelled and unlabelled domain adaptation were investigated to transfer ML knowledge between datasets.

Table 2: A summary of the articles in each section of this thesis.

Section	Title of article	Novelty	Case studies
4	Monitoring Mixing Processes Using Ultrasonic Sensors and Machine Learning	ML was combined with non-invasive, reflection-mode US sensing to monitor mixing processes. A range of ML algorithms were trialled including LSTMs and CNNs.	• Honey-water blending • Flour-water batter mixing
5	Predicting Alcohol Concentration during Beer Fermentation Using Ultrasonic Measurements and Machine Learning	ML was combined with non-invasive, reflection-mode US sensing to monitor fermentation processes. Omission of the temperature as a feature was evaluated.	• Laboratory-scale beer fermentation
6	Convolutional feature extraction for process monitoring using ultrasonic sensors	A convolutional feature extraction method is presented. Multi-task learning and data augmentation were applied to US sensor data.	• Honey-water blending • Flour-water batter mixing • Pipe section cleaning • Laboratory-scale beer fermentation
7	Transfer learning for process monitoring using reflection-mode ultrasonic sensing	Unlabelled domain adaptation was applied to US sensor data for monitoring mixing and cleaning processes. Omission of the temperature as a feature was evaluated.	• Honey-water blending • Pipe section cleaning

8	Domain Adaptation and Federated Learning for Ultrasonic Monitoring of Beer Fermentation	Labelled domain adaptation was applied to US sensor data for fermentation monitoring. Federated learning and multi-task learning were combined with US sensor data.	• Laboratory-scale beer fermentation • Industrial-scale beer fermentation

3.2.1 Feature extraction

3.2.1.1 Wavelet transform

Wavelet analysis uses decaying waveforms as transform functions to analyse the frequency content of a waveform at each location in the time domain (Mallat and Mallat, 1999). The Continuous Wavelet Transform (CWT) uses a continuous range of frequencies to decompose the US signal whereas the Discrete Wavelet Transform (DWT) and Wavelet Packet Transform (WPT) use discrete frequencies at each decomposition. The WPT performs successive decompositions on each branch of the original signal whereas the DWT only applies successive decompositions to the lower frequency signal content (Mallat, 1989). Thereby the WPT has improved resolution of high frequency components of the signal. All these wavelet transform techniques have been widely applied for feature extraction from US waveforms (Bowler et al., 2022). The DWT was utilised in Section 4 and Section 7 (‘dwt’ function, MATLAB) as it has more commonly been utilised with US waveforms compared with the WPT (Bowler et al., 2022). A key parameter is the choice of the analytical wavelet shape, termed the mother wavelet (Safavian et al., 2005). The Symlet wavelet was chosen as the mother wavelet for both investigations due to it being the least asymmetric and therefore most visually similar to the waveforms studied. The number of vanishing moments was trialled between one and ten in Section 4 as is the range used in previous studies with US waveforms (Bowler et al., 2022). Six vanishing moments were selected in Section 7 after this was found to be optimal for the same application in Section 4 (honey-water blending). Three, five, and seven decomposition levels were evaluated in Section 4 owing to the common range of three to seven being used with US waveforms (Bowler et al., 2022). Different levels of decomposition worked best for each task. These results led to using five decomposition levels in Section 7 as the middle of the range tested. The CWT was also used in Section 4 (‘cwt’ function, MATLAB) as it retains greater information about the frequency content of the waveform than the DWT. In this case, it was used with CNNs to handle the greater number of input features compared to ANNs or LSTMs. The Morlet wavelet was used owing to its symmetricity.

3.2.1.2 Waveform features

The following section describes features calculated from the US waveform. The waveform may be in the time domain or time-frequency domain after applying the wavelet transform. The waveform in either domain may consist of multiple overlapping sound waves requiring multiple features to fully capture its variation with the properties of the measured materials.

The features investigated in this thesis were chosen to cover the full range of previous features extracted for use with US waveforms and machine learning determined in Bowler et al. (2022). For example, specific amplitudes in waveforms were monitored (3.2.1.2.4 Peak-to-peak amplitude, 3.2.1.2.5 Maximum amplitude, 3.2.1.2.6 Minimum amplitude), single measures of all waveform amplitudes (3.2.1.2.1 Energy, 3.2.1.2.2 Sum absolute amplitude, 3.2.1.2.3 Sum root amplitude, 3.2.1.2.7 Standard deviation, 3.2.1.2.8 Skewness, 3.2.1.2.9 Kurtosis), the position of the waveform (3.2.1.2.10 Position of maximum peak, 3.2.1.2.11 Position of minimum peak), using waveform amplitudes directly as features (3.2.2.1 Principal component analysis), the time of flight (3.2.1.2.12 Time of flight), the variance between

consecutive waveforms (3.2.1.4 Energy standard deviation), time-lagged features (3.2.1.3 Feature gradients), and additional process features (3.2.1.5 Temperature).

3.2.1.2.1 Energy

The Energy is the sum of all the squared amplitudes in a section of a waveform (Equation 1). It is a measure of the magnitude of the sound wave.

$$E = \sum_{i=start}^{i=end} A_i{}^2 \quad (1)$$

Where *E* is the Energy, *i* is the sample point in a waveform, *start* is the first sample point, *end* is the last sample point, A_i is the waveform amplitude at sample point *i* (Zhan et al., 2015).

3.2.1.2.2 Sum absolute amplitude

The Sum Absolute Amplitude (SAA) is the sum of all the absolute amplitudes in a section of a waveform (Equation 2). It is also a measure of the magnitude of a sound wave but it assigns larger weighting to smaller amplitudes than the Energy.

$$SAA = \sum_{i=start}^{i=end} |A_i| \quad (2)$$

Where *SAA* is the Sum Absolute Amplitude, *i* is the sample point in a waveform, *start* is the first sample point, *end* is the last sample point, A_i is the waveform amplitude at sample point *i* (Zhan et al., 2015).

3.2.1.2.3 Sum root amplitude

The Sum Root Amplitude (SRA) is the sum of the square root of the absolute amplitudes in a waveform section (Equation 3). It is another measure of the magnitude of a soundwave, however, gives larger weighting to smaller values than the SAA.

$$SRA = \sum_{i=start}^{i=end} \sqrt{|A_i|} \quad (3)$$

Where *SRA* is the Sum Root Amplitude, *i* is the sample point in a waveform, *start* is the first sample point, *end* is the last sample point, A_i is the waveform amplitude at sample point *i* (Zhan et al., 2015).

3.2.1.2.4 Peak-to-peak amplitude

The peak-to-peak amplitude measures the range between the maximum and minimum amplitudes in a waveform section (Equation 4).

$$PPA = max(A_{start:end}) - min(A_{start:end}) \quad (4)$$

Where *PPA* is the peak-to-peak amplitude, *start* is the first sample point, *end* is the last sample point, *max* and *min* indicate functions to find the maximum or minimum amplitudes in a waveform section, respectively.

3.2.1.2.5 Maximum amplitude

The maximum amplitude is the largest positive value in a waveform section (Equation 5).

$$A_{max} = max(A_{start:end}) \quad (5)$$

Where A_{max} is the maximum amplitude in a waveform section, *start* is the first sample point, *end* is the last sample point.

3.2.1.2.6 Minimum amplitude

The minimum amplitude is the largest negative value in a waveform section (Equation 6).

$$A_{min} = min(A_{start:end}) \quad (6)$$

Where A_{min} is the minimum amplitude in a waveform section, *start* is the first sample point, *end* is the last sample point.

3.2.1.2.7 Standard deviation

The standard deviation of sample point amplitudes along a waveform section is a measure of their dispersion relative to the mean value (Equations 7 and 8).

$$\mu = \frac{\sum_{i=\text{start}}^{i=end} A_i}{start-end} \quad (7)$$

$$STD = \sqrt{\frac{1}{start-end}\sum_{i=start}^{i=end}(A_i - \mu)^2} \quad (8)$$

Where μ is the mean amplitude in the waveform, *STD* is the standard deviation, *i* is the sample point in a waveform, *start* is the first sample point, *end* is the last sample point, A_i is the waveform amplitude at sample point *i* (Zhan et al., 2015).

3.2.1.2.8 Skewness

The skewness is a measure of the lack of symmetry in the waveform section (Equation 9).

$$S = \frac{\sum_{i=start}^{i=end}(A_i-\mu)^3}{(end-start)\times STD^3} \quad (9)$$

Where *S* is the skewness, *i* is the sample point in a waveform, *start* is the first sample point, *end* is the last sample point, A_i is the waveform amplitude at sample point *i* (Caesarendra and Tjahjowidodo, 2017).

3.2.1.2.9 Kurtosis

The kurtosis is a measure of the tailed-ness of the waveform section (Equation 10).

$$K = \frac{\sum_{i=start}^{i=end}(A_i-\mu)^4}{(end-start)\times STD^4} \quad (10)$$

Where *K* is the kurtosis, *i* is the sample point in a waveform, *start* is the first sample point, *end* is the last sample point, A_i is the waveform amplitude at sample point *i* (Caesarendra and Tjahjowidodo, 2017).

3.2.1.2.10 Position of maximum peak

The position of the maximum peak is the sample point where the maximum amplitude in a waveform section is located. This position may change during a process due to changes in material speed of sound, variations in temperature, or superposition of multiple overlapping sound waves.

3.2.1.2.11 Position of minimum peak

Similarly, the position of the minimum peak is the sample point where the minimum amplitude in a waveform section is located.

3.2.1.2.12 Time of flight

Three different methods to determine the time of flight (the length of time for a sound wave to travel through a material) were used in these works. Firstly, a thresholding method was used which determines the earliest sample point in a waveform which exceeds a fixed value (Zhu et al., 2017). However, with small signal to noise ratios, noise may sometimes rise above this threshold value or the threshold must be increased which can degrade the accuracy of the calculated waveform arrival time. Therefore, additional methods were also used. The second method was zero-crossing where the time of arrival is defined as the next sample point where the waveform crosses zero amplitude (Zhu et al., 2017). It therefore provides a more consistent selection of the arrival time as the threshold value may be

located along any part of a peak. However, it is still impacted by low signal to noise ratios. The third method was auto-correlation ('autocorr' function in MATLAB and 'correlate' function in NumPy) which performs a correlation analysis of the incident and received sound waves to determine the time delay (Khyam et al., 2017). This method is therefore unaffected by the signal to noise ratio but can become inaccurate if the shape of the incident and received waveform differ. Therefore, to improve robustness in this thesis, all three methods are used to provide time of flight estimates for the US waveform.

3.2.1.3 Feature gradients

Feature gradients provide a measure of the process trajectory by comparing feature values at the current timestep to their values at previous timesteps. In these works, the feature gradients were calculated after applying a backwards, one-sided moving mean so that only past data is used. The size of the sliding window to calculate the mean can be decided through using process knowledge (e.g., selected as 5 h in Section 5 or 5 % of the total process time in Section 7) or through testing multiple lengths of time (as used in Section 4).

3.2.1.4 Energy standard deviation

The Energy standard deviation is a measure of the dispersion in the Energy of consecutively acquired waveforms (Equation 11). This can be used to identify fluctuations in the materials being measured by the US sensor, for example, the production of CO_2 bubbles during fermentation.

$$ESTD = \sqrt{\frac{1}{W}\sum_{i=1}^{i=W}(E_i - \bar{E})^2} \tag{11}$$

Where *ESTD* is the Energy standard deviation, and *W* is the number of consecutively acquired waveforms investigated.

3.2.1.5 Temperature

Use of the process temperature as a feature, or its exclusion, were also investigated in the ML models. This is because properties effecting US wave propagation (e.g., compressibility, density) are highly sensitive to changes in temperature and therefore impact the acquired waveforms.

3.2.1.6 Convolutional feature extraction

Convolutional neural networks (CNNs) have convolutional layers as well as fully connected layers. The convolutional layers consist of filters that perform cross-correlation on the input data. This enables CNNs to automatically learn spatially invariant features from the input data (Bowler et al., 2022). Section 6 presents a convolutional feature extraction methodology for US waveforms (trained in Keras Python). A 1D CNN is pre-trained on an auxiliary task to classify waveform dataset membership of the experimental datasets previously outlined. Segments of 1000 sample points in length were selected from each waveform. The position of the 1000 sample point length window was chosen for each waveform by investigating the difference between the start and end waveforms of the individual process. The areas with the largest visual change throughout the process were used. To increase the training set size for training the CNN, and to improve informative feature learning in the convolutional layers, a 600 × 1 input to the CNN was used. Data augmentation using a sliding window, laterally translated by 100 sample points each time, produced five waveform segments of 600 sample points in length. Further data augmentation used separate normalisation of each waveform segment to differentially magnify the waveform. This was to ensure that the CNN learned features specific to each waveform, rather than the position or magnitude of features.

A grid search was used to select the learning rate, batch size and number of neurons in the fully connected layer. No padding was used. Training was performed with the Adam optimiser. The minimum number of neurons in the fully connected layer to achieve 100% accuracy for the dataset membership prediction was used to ensure feature identification in the convolutional layers rather than the fully connected layer. The pre-trained convolutional weights were then used to extract features on the full-size waveform for each dataset.

3.2.2 Feature selection

3.2.2.1 Principal component analysis

Principal Component Analysis (PCA) is an unsupervised ML method that linearly transforms input variables into new, uncorrelated features called principal components (PCs) (Khalid et al., 2014). Principal component analysis (PCA) was the only feature selection method used in these works and was applied in Section 4 ('pca' function, MATLAB) and Section 6 ('PCA' function scikit-learn). Despite only identifying linearly correlated variables, PCA was utilised instead of autoencoders because it offers advantages of not requiring model training or hyperparameter optimisation and the desired number of features can be selected without having to retrain a model.

There are many methods to estimate the number of PCs to retain. For example, choosing the most informative PCs that explain a predefined amount of variance in the data (typically larger than 90 % (Valle et al., 1999)), retaining PCs with eigenvalues greater than 1 (Braeken and Van Assen, 2017), using scree plot to visualise the impact of each PC (David and Jacobs, 2014), or selecting PCs that have a greater explained variance than a predefined threshold (e.g. 1 % (Cau et al., 2005) or 2 % (Miao et al., 2008)). In Section 4 the amplitude at each sample point in the waveform was used as the input variables for PCA. The PCs explaining > 95 % of the variance in the input variable were selected which is a commonly selected threshold. However, if this number was greater than ten, then only the first ten PCs were selected to reduce model dimensionality. In Section 6, PCA was applied to the waveform features extracted using the pre-trained convolutional filter weights. This was to aid LSTM layer training accuracy and stability in the deep neural network. PCA reduced the dimensionality of the data, minimised non-useful information inputted into the network, and enabled the use of additional features such as the US time of flight and standard deviation between consecutive waveforms. Five PCs were chosen for each dataset and justification is provided below. In this thesis, the use of PCA was compared to other ML methods. As such, the number of PCs was not optimised fully, only the estimation methods as discussed above were used. In practice, if using PCA, the number of PCs can be thought of as another hyperparameter to optimise during the model validation procedure.

Table 3 presents the percentage variability explained by each PC for the US waveform datasets and the number of PCs required to explain 95% of the variability, a commonly used method to determine the number of PCs to utilise (Valle et al., 1999). In Section 6, the first PC likely follows the common waveform changes across the full dataset caused by variations in the US properties of the materials being monitored (either due to changing composition or process temperature). Successive PCs will identify waveform changes more specific to each batch, most likely due to the different process temperatures. Therefore, it is anticipated that only a small number of PCs are required (i.e. greater than one) to monitor the changing material composition and account for changes in the monitoring US waveform at different temperatures. This is supported by Table 3 where the percentage variability explained drops off after the first two PCs. As shown in Table 3, the smallest number of PCs required to explain 95% of the variability in the dataset, is eight for the Plastic Cleaning dataset and nine for fermentation monitoring using only the first reflection. Therefore, using these two pieces of guidance (the primacy of the first and second PCs and the smallest

number of PCs to explain 95% of dataset variability), five PCs were selected to obtain useful waveform information while minimising noise in Section 6.

Table 3. A summary of the distribution of the explained variance by each PC for the US waveform datasets after convolutional feature extraction (Section 6).

Experimental dataset	Waveforms	Number of PCs to explain 95% of variability	Variability explained by 1st PC (%)	Variability explained by 2nd PC (%)	Variability explained by 3rd PC (%)	Variability explained by 4th PC (%)	Variability explained by 5th PC (%)
Fermentation	Reflection 1	9	56.4	23.1	9.2	2.1	1.5
	Reflection 2	18	30.4	21.6	14.9	9.0	6.1
Cleaning of food fouling from pipe sections	Flat rig	15	60.4	15.2	7.4	4.3	1.8
	Circular, plastic	8	56.7	14.3	12.4	6.3	1.9
	Circular, metal	32	50.9	12.3	8.5	4.6	3.7
Honey-water mixing 1	Central sensor	24	52.1	18.8	7.5	4.6	2.3
	Non-central sensor	41	51.4	17.0	4.7	3.8	2.8
Honey-water mixing 2	Central sensor	19	38.6	30.8	12.4	4.1	2.9
	Non-central sensor	25	41.6	36.8	4.6	2.7	2.4
Batter mixing	Sensor 1	42	49.1	15.1	14.3	4.5	2.6
	Sensor 2	16	60.5	16.3	7.5	3.1	1.5

3.2.3 Algorithms

3.2.3.1 Training, validation, and test procedures

In ML, datasets are split into training, validation, and test sets. The models are trained using the training set, the model hyperparameters are optimised by monitoring performance on the validation set, and a measure of the optimised model's accuracy on unseen data is estimated through evaluation on the test set. The size of each of these splits, the data used in each of these splits, and the validation procedure used must be decided based on the application of the ML model as the pipeline is chosen to maximise accuracy on the validation and test set data. The aim of this thesis was to develop ML methods to facilitate optimal deployment of US sensors for process monitoring applications in industrial environments. To achieve this purpose, the conclusions drawn from this thesis, such as the relative strengths and weaknesses of each US sensing or ML technique, must be representative of what may be expected in industrial environments. Therefore, the training, validation, and testing procedures must be representative of those that would be used in industrial settings so that the conclusions drawn from this thesis can be expected to transfer to industrial processes. K-fold cross validation splits the combined training and validation sets into *k* sections. Each of these sections is used as individual validation sets and an average metric for the validation accuracy is obtained (Jung and Hu, 2015). This is a robust method of validation as it repeats the procedure *k* number of times. The size of *k* typically ranges between 2 and 10 (Xu and Goodacre, 2018). However, in industrial environments, single-fold validation may be utilized to reduce ML model development time. Furthermore, drawing conclusions from test set results for models that have undergone single-fold validation biases for ML techniques with greater generalisation capabilities, a useful attribute in industrial scenarios (see Section 9.5 'Considerations for industrial application'). Therefore, either k-fold or single-fold validation methods are representative of industrial practices (Table 4). However, no

validation set was used for the LSTMs and CNNs in Section 4 or the LSTMs in Section 7. This represents a limitation although was not found to influence the conclusions drawn from this thesis. The reason for this is that the omission of a validation set disadvantages the hyperparameters selected for the LSTMs and CNNs compared to the other algorithms used in the articles (ANNs and SVMs in Section 4 and ANNs in Section 7) as they were not chosen to generalise to unseen data. Despite this, LSTMs and CNNs performed best in both studies. Further discussion is provided in Sections 4 and 7. The chosen test set size typically lies in the range of 10 to 50 % of the data with the validation set forming the same proportion of the total training and validation set size (Xu and Goodacre, 2018). However, in industrial applications, the desired level of model generalisability is dependent on the volume of representative data able to be collected which is dependent on the process complexity, variability, and level of disruption to the manufacturing process caused by labelled data collection (see Section 9.5 'Considerations for industrial application'). Therefore, all validation and test set sizes used in this thesis contribute to the thesis aim of to develop ML methods to facilitate optimal deployment of US sensors for process monitoring applications in industrial environments. For all validation procedures, the datasets were split by number of runs so that all datapoints for a single run were included in either the training, validation, or test set.

Table 4: The validation procedures used in the sections of this thesis. For all validation procedures, the datasets were splits by number of runs so that all datapoints for a run was included in either the training, validation, or test set.

Section	Algorithm	Validation procedure
4	ANN	Single-fold (~ 20% of data)
	SVM	5-fold
	LSTM	None
	CNN	None
5	ANN and LSTM	5-fold
6	LSTM	Single-fold (~ 20% of data)
7	ANN	Single-fold (~ 20% of data)
	LSTM	None
8	LSTM	k-fold (where k ranges from 1 to 4)

3.2.3.2 Metrics

In this thesis, several metrics were utilised to evaluate the predictive capabilities of the ML models. Firstly, for classification tasks, the accuracy was used. The accuracy is calculated by summing the number of correct predictions and dividing by the total number of predictions. This value is multiplied by 100 to convert into percentage terms. Other classification metrics include precision, specificity, and recall (or sensitivity). Precision is a measure of how precisely the model detects true positives compared to false positives, specificity is a measure of the model's accuracy to detect negative instances, and recall is a measure of the model's accuracy to detect positive instances (Towards Data Science, 2019). These additional metrics are typically used to evaluate models trained on data with imbalanced classes. For example, if few positive instances exist within the dataset, an artificially high accuracy can be obtained for a poorly performing model. However, in this thesis, the model accuracies are used to compare between ML pipelines and are not used to predict a final model's performance in application. Therefore, only the model accuracy was used as this provides a method of comparing between ML techniques. For regression tasks, a combination of the Mean Absolute Error (MAE, equation 12), Mean Squared Error (MSE, equation 13), and Coefficient of Determination (R^2, equation 14) was used. The Coefficient

of Determination is a commonly used method that provides the percentage variation between the predicted and true regression values. It therefore provides an interpretable metric with which to evaluate models and is commonly used to assess ML models (Hale, 2020). However, the R^2 score has several limitations, such as increasing as new variables are added to the model and being dependent on the variation in the data (Ford, 2015). Therefore, a combination of regression metrics should be used to provide an unbiased assessment of a model's performance. Similarly to the classification metrics, in this thesis, the regression metrics were used to compare between US and ML techniques rather than evaluate an ML model's performance in industrial application. Therefore, any of these metrics may be used to compare between models. However, in industrial application, choice of evaluation metric may alter the best choice of feature extraction method, algorithm, or hyperparameters to accomplish the desired task. For example, MSE provides larger weighting to high errors and therefore a lower MSE indicates a model more able to correctly predict over the full process length. In comparison, models with low MAE may provide increased accuracy at certain process stages compared to models with low MSE scores.

$$MAE = \frac{\sum_{i=1}^{n}|y_i - \hat{y}_i|}{n} \quad (12)$$

$$MSE = \frac{\sum_{i=1}^{n}(y_i - \hat{y}_i)^2}{n} \quad (13)$$

$$R^2 = 1 - \frac{\sum_{i=1}^{n}(y_i - \hat{y}_i)^2}{\sum_{i=1}^{n}(y_i - \bar{y})^2} \quad (14)$$

Where n is the number predictions, y is the true value, $\hat{y}$ is the predicted value, and $\bar{y}$ is the average of the true values.

3.2.3.3 Artificial neural networks

Artificial neural networks (ANNs) have the ability to create new features in their hidden layers from combinations of input features to perform the ML task. ANNs are formed of weight and bias terms that connect the model inputs to the outputs that are iteratively updated during the training procedure (Rodriguez-Galiano et al., 2015). ANNs were used in Sections 4 (MATLAB), 5 (Keras Python), and 7 (MATLAB). US waveform features from the current process timestep were inputted in to ANNs to predict the output value at the current timestep. In Section 4, the Levenberg-Marquardt optimisation algorithm was used for regression networks and the scaled conjugate gradient optimisation algorithm for classification, as recommended by MathWorks (2021a). Early stopping with a validation patience of 6 epochs was used to prevent overfitting. For each task, 10 ANNs were trained and the average validation error was used to account for the effects of random weight initialisation. To further prevent overfitting, single fold validation using 80% of the total training and validation set as the training set and 20% of the training and validation set as the validation set was used for hyperparameter optimisation. A grid search determined the optimal number of neurons in the hidden layer (varied between 1 and 10 in intervals of 1) and regularisation weight (varied between 0.1 and 0.5 in intervals of 0.1). The same methodology was used in Section 7, except 5 neurons were used in the hidden layer for all tasks as the maximum number of features used was 7 to reduce the likelihood of overfitting. Section 5 optimised ANNs using the Adam optimisation algorithm, which is the most common optimisation algorithm for deep learning tasks (Bowler et al., 2022). Cross-validation determined the optimal batch size, number of neurons in the hidden layer, learning rate, drop-out rate, L2 regularisation penalty, and number of epochs for training.

3.2.3.4 Support vector machines

Section 4 evaluated Support Vector Machines (SVMs) for classification tasks ('fitcsvm' function, MATLAB) and Support Vector Regression ('fitrsvm' function, MATLAB) for regression tasks. US waveform features from the current process timestep were inputted in to SVMs to predict the output value at the current timestep. Bayesian optimisation for 60 evaluations with the expected improvement acquisition function was used to select the box constraint value, kernel scale, kernel function, polynomial order, and whether the inputs were standardised, as suggested by MathWorks (2021b). 5-fold cross validation was used.

3.2.3.5 Convolutional neural networks

In Section 4, the CNNs consisted of 2 convolutional layers. The first convolutional layer was either 2D or 3D containing 8 5 × 5 pixel filters for each sensor input matrix depending on whether one or two sensor signals were being used as inputs. The second convolutional layer was a 2D convolutional layer containing 16 5 × 5 pixel filters. The "same" padding was applied to keep the input matrices the same size. Batch normalisation was applied after each convolutional layer to aid training and provide some regularisation. By normalising each mini batch for every layer in the network, each layer does not need to continuously adapt to changing input distributions (Ioffe, 2015). Batch normalisation was followed by the ReLu non-linearity function and 2 × 2 pixel max pooling. The Adam training function was used, the initial learning rate was selected as 0.01 with a drop factor of 0.33 after 4 epochs. Training was carried out for 8 epochs, with a mini batch size of 256. The training data was shuffled after every epoch. A dropout layer was added before the fully connected layer to further prevent overfitting, with the dropout factor varied between 0, 0.1, 0.3, and 0.5. Two types of input were used for the CNNs. Firstly, 25 stacked time-domain waveforms extending backwards approximately 10 s in time were used to provide historical timesteps to the network (Figure 5). Secondly, single waveforms after applying the CWT were used (Figure 6). Table 5 lists references which employ the same hyperparameters to those used in this study. Although the hyperparameter used in this model are similar to previous literature, ideally, they should be selected through a validation procedure. Hyperparameters used in previous studies may be used to set the bounds of the hyperparameter search to be fine-tuned on the validation data.

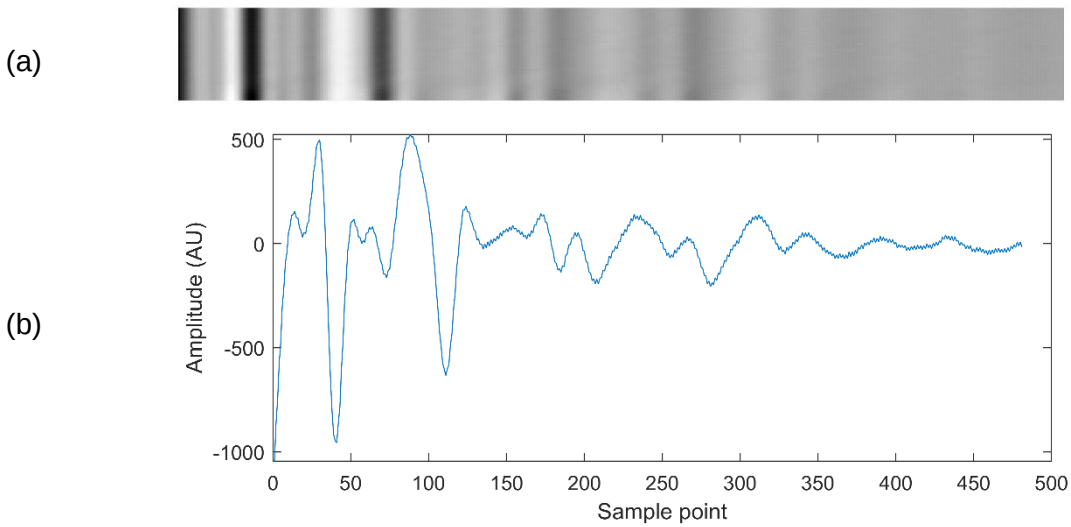


Figure 5. (a) A grey-scale representation of the input matrix to the time domain input CNNs in Section 4, constructed of 25 ultrasound waveforms similar to the waveform depicted below it in Figure 5b. The bright pixels correspond to the maximum amplitude values of the 25 waveforms, and the dark pixels correspond to the minimum amplitude values. (b) An example of an ultrasonic waveform used in each row of the time domain input matrix.

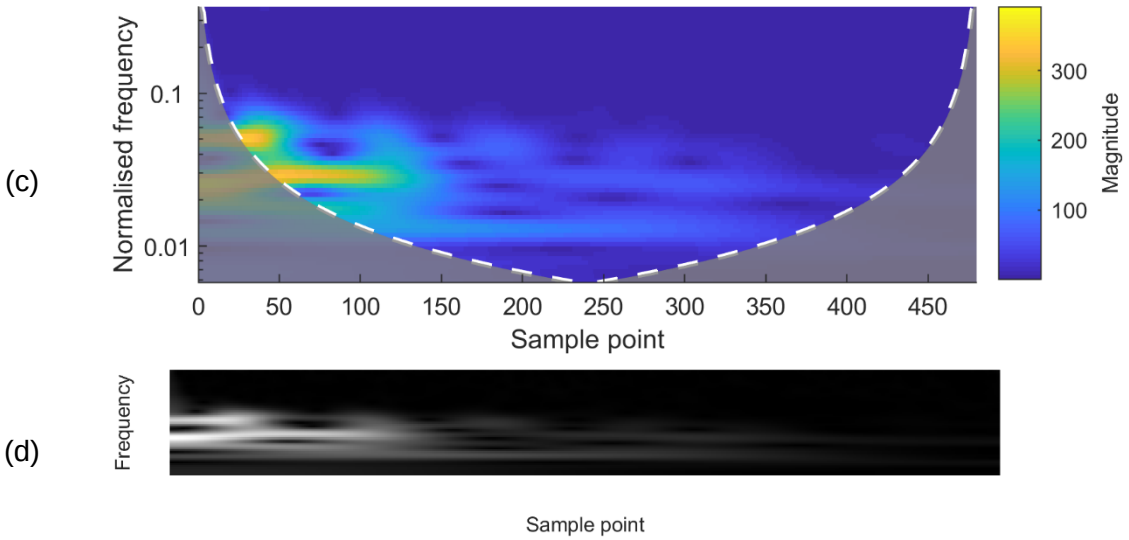


Figure 6. (a) A frequency-time domain magnitude scalogram after the CWT of a single waveform. The sample points correspond to the sample points of the original waveform. (b) A grey-scale image representation of the matrix used as an input to the CWT-input CNN in Section 4. It contains the absolute values after the CWT of the same waveform as used for Figure 6a. The bright pixels correspond to the maximum frequency amplitudes, and the dark pixels to the minimum.

Table 5: Previous studies that combine US sensors and CNNs that use the same hyperparameters as employed in this study.

Hyperparameter	**References**
2 convolutional layers	Lim and Sohn (2020) Yan et al. (2020)
8 filters in first layer	Gao et al. (2021)
Doubling number of filters between first and second layers	Rautela et al. (2021) Pyle et al. (2021)
5 x 5 filter size	Gao et al. (2021)
“Same” padding	Munir et al. (2019)
Batch normalisation	Xiao et al. (2021)
ReLu activation function	Rautela et al. (2021) Pyle et al. (2021)
2 x 2 max pooling	Pyle et al. (2021)
Adam optimisation algorithm	Rautela et al. (2021) Gao et al. (2021)
0.01 learning rate	Gopalakrishnan et al. (2021)

3.2.3.6 Long short-term memory neural networks

Long short-term memory neural networks (LSTMs) are able to learn sequences of time series data. LSTMs are a development of recurrent neural networks (RNNs) that reduce the likelihood of exploding or vanishing gradients and thereby enable the learning of long-term dependencies (Hochreiter and Schmidhuber, 1997). LSTMs store representations of sequences by using gate units to update their internal network state. At each time step, LSTMs use the input features at the current time step as well as information passed from the previous time steps to make their prediction. Therefore, they have the capability to learn feature trajectories during processes that evolve over time.

To the author's knowledge, from the literature review conducted in Section 2, the works included in this thesis are the first to use LSTM layers for process monitoring using US measurements. Previously, ANNs (Wallhäußer et al., 2011; Wallhäußer et al., 2013), SVMs (Wallhäußer et al., 2013; Wallhäußer et al., 2014), decision tree (Escrig et al, 2020a; Escrig et al, 2020b), and K-nearest neighbour algorithms (Escrig et al, 2020a; Escrig et al, 2020b) have been used to monitor cleaning processes with US sensors. However, this thesis is the first to use LSTM layers to monitor cleaning processes which provide the advantage of learning the trajectory of processes over time by incorporating knowledge from previous timesteps (Bowler et al., 2022). Similarly, only ANNs have been previously used to monitor fermentation processes using US sensors (Becker et al, 2002; Hussein et al., 2012). In this thesis, LSTMs are also used for this task owing to their ability to incorporate knowledge from previous time steps and monitor the progress of processes that evolve over time. This thesis also presents the first use of ML and US sensors to monitor mixing processes (Bowler et al., 2022). However, LSTMs have been used for US measurements for other applications such as flow regime monitoring (Nguyen and Park, 2020).

LSTM layers were evaluated in all works in this thesis. The LSTMs were used in sequence to sequence tasks, where the sequence of feature extracted from US sensor measurements were used to predict the sequence of output values. At each timestep the model predicted the output value at the current timestep. At the following timestep, the model obtains the new US sensor measurement features from this timestep with which to predict the output value. Therefore, only US sensor measurement features are used to make predictions and previous ground truth data or output predictions are not used as inputs. All timesteps for each dataset were used as a single sequence rather than being truncated into multiple sequences of shorter length. Long sequences (250–500 timesteps) are prone to producing vanishing gradients in LSTM layers when predicting a single output. The reason for this is that the gradient information from the output is backpropagated through all the timesteps to the initial input, leading to the attenuation of the gradient signal over time. As a result, the weights of the network do not get updated effectively, which can lead to poor performance and slow convergence. Predicting an output at every timestep is an approach to reduce the likelihood of vanishing gradients in LSTM layers, as used in the works in this thesis (Machine Learning Mastery, 2021). By doing so, the direct supervision of each timestep enables early US measurements to contribute to early timesteps during training by learning directly from the output at that time. This approach ensures that the gradient signal is not attenuated over time, and each output contributes to the updating of the network's weights. Moreover, by predicting an output at every timestep, later US measurements will contribute less to earlier timesteps during training. This is because the gradient signal attenuates over time, leading to a smaller contribution to the earlier timesteps. Conversely, at later timesteps, earlier US measurements will contribute a smaller gradient, while later US measurements will provide a more substantial contribution to the training.

Table 6 presents hyperparameters used in previous studies when combining US sensor data with LSTMs. Similar hyperparameters were used in Sections 4 and 7 as outlined in the following paragraphs. However, cross- and single-fold validation were used in Sections 5, 6, and 7 to optimise the hyperparameters used. Although the hyperparameters used in Sections 4 and 7 were selected based of values from literature, preliminary investigations, and prior experience, validation procedures should be used for fine-tuning of the parameter for the ML task. The Adam optimisation algorithm was used in all sections owing to it being most widely used for deep learning tasks with US data, most frequently with CNNs (Bowler et al., 2022).

Table 6: Hyperparameters used in previous studies that combine US sensors and LSTMs.

Hyperparameter	Value	Reference
Optimisation algorithm	Scaled Conjugate Gradient	Nguyen and Park (2020)
Learning rate	0.01	Nguyen and Park (2020)
	0.00005	Ren et al. (2021)
Number of LSTM units	32	Ren et al. (2021)
	6	Bosse et al. (2021)
	7	Qin et al. (2018)
Number of epochs	400	Nguyen and Park (2020)
	500	Ren et al. (2021)
Number of neurons in fully connect layer	512, 128	Ren et al. (2021)

In Section 4 (MATLAB), the training was carried out for 600 epochs with a batch size of 1, the Adam optimisation algorithm was used, a learning rate of 0.01, and a gradient threshold of 1 to prevent problems of exploding gradients. Only 5 hidden units were used in the LSTM layer, as the processes did not follow a complex sequence. Five neurons were used in the fully connected layer to ensure linear fitting of the feature combinations with the activation function.

In Section 5 (Keras Python), zero-padding was applied to the features to make every batch sequence an equal length. A masking layer specified that the LSTM units ignore this padding. The Adam optimisation algorithm was used and 5-fold cross-validation determined the optimal batch size, number of LSTM units, learning rate, drop-out rate, L2 regularisation penalty, gradient norm clipping value, and number of epochs. The optimal set of hyperparameters were used to retrain a model using all of the training and validation data.

In Section 6 (Keras Python), deep neural networks consisting of an LSTM layer followed by a fully-connected layer were used for all ML tasks. A fully-connected layer allows for the creation of modified features which better match the prediction task output while the LSTM layer learns the trajectories of the input features. The input features were normalised and zero-padding at the start extended the sequence lengths to that of the maximum. A masking layer specified the LSTM to disregard the zero-padding. The Adam optimisation algorithm and a gradient norm clipping value of 1 was used. A single-fold validation procedure determined the learning rate, number of LSTM units, dropout probability, L2 regularisation penalty, number of neurons in the fully-connected layer, and batch size. The optimal set of hyperparameters were used to retrain a model using all of the training and validation data. Multi-task learning was also investigated to aid LSTM learning of the process trajectory. Multi-task learning is an ML approach where a single model is trained on multiple related tasks simultaneously, sharing some or all of the model parameters between tasks to improve performance on all tasks (Caruana, 1997). By training on two correlated tasks, the shared LSTM layer may learn more effective feature trajectories while reducing redundant information being stored. This may have several benefits, such as increased model accuracy through global learning of feature trajectories important to the process being monitored, more stable model training by optimising for two combined losses, and reducing overfitting by preventing a single task from dominating the learning process (Zhang and Qiang, 2018).

In Section 7 (MATLAB), network training was carried out for 60 epochs using the Adam optimisation algorithm, an initial learning rate of 0.01, a batch size of 2, and a gradient threshold of 1 to prevent problems of exploding gradients. In the LSTM layer, 50 hidden units

were used and 50 neurons in the fully connected layer. A 0.5 probability dropout layer was used to prevent overfitting and improve algorithm generalisation performance.

In Section 8 (Keras Python), multi-task deep neural networks consisting of a fully connected layer followed by an LSTM layer were used for all ML tasks. Zero-padding was applied to the US features to make every fermentation sequence equal to the maximum sequence length of 1556 timesteps. A masking layer designated that the LSTM units ignore this padding. The Adam optimisation algorithm and a gradient norm clipping value of 1 was used to reduce the likelihood of exploding gradients. The order of the training sets was shuffled after every epoch. The regression losses were multiplied by 0.1 to ensure their magnitudes were similar to the classification losses. This aided the network in learning both the classification and regression tasks. After cross-validation, the optimal hyperparameters which resulted in the lowest average validation error were used to train a final model using the entire training and validation sets.

3.2.4 Domain adaptation

Transfer learning is a machine learning technique where a model trained on one task is used to improve performance on a different but related task (Torrey and Shavlik, 2009). The most common example is using pre-trained CNNs trained on large datasets to apply to new tasks (Yang et al., 2021). Domain adaptation is a subcategory of transfer learning where a model is trained on a source domain and transferred to a target domain where the data distributions may differ (Ben-David et al., 2010). Domain adaptations methodologies are therefore used to increase the accuracy of the model trained on the source domain to predict on the target domain (Figure 7). Unlabelled domain adaptation consists of domain adaptation methods where no labelled data is available in the target domain. In contrast, labelled domain adaptation methods use some labelled data from the target domain to aid transfer of the model. To the authors knowledge, as reviewed in Section 2, only Gao et al. (2021) has used domain adaptation for 1D US measurements (domain adaptation has been used for 2D US images in medical applications and is reviewed in Section 7). Gao et al. (2021) used microseismic datasets from earthquake studies to improve time-of-flight detection for acoustic logging tools for collecting borehole information in oil fields.

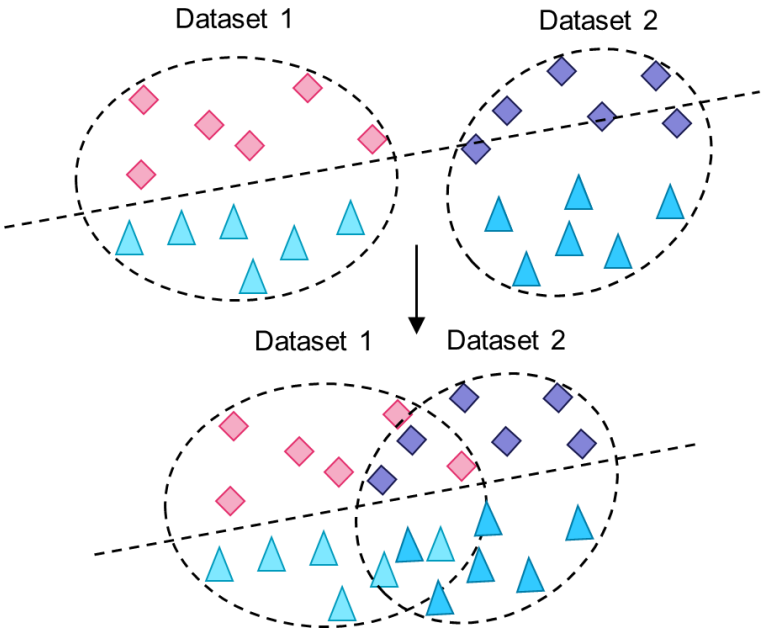


Figure 7: A depiction of domain adaptation where the position of Dataset 2 in the feature space has been shifted for accurate classification using the decision boundary separating Dataset 1.

3.2.4.1 Unlabelled domain adaptation

Section 7 compares two unlabelled domain adaptation techniques to transfer ML models to new processes where no labelled data is available. A Single Feature (SF) transfer method was compared to Transfer Component Analysis (TCA) (Pan et al., 2011). TCA minimises the distance between source and target domain feature spaces by using the Maximum Mean Discrepancy (MMD) and extracts transfer components that maximise the variance across this shared feature space. The MMD is a measure of the distance between feature distribution embeddings in a reproducing kernel Hilbert space (Tolstikhin et al., 2016). TCA is a highly cited method and MMD is common to use in domain adaptation problems (Li et al., 2019; Li, Zhang, and Deng, 2018; Guo et al., 2019; Lu et al., 2017; Geng, Tao, and Xu, 2011). Furthermore, TCA guarantees feature extraction with no training necessary unlike techniques such as domain adversarial networks (Tzeng et al., 2017). The TCA code provided in the MATLAB domain adaptation toolbox produced by Yan (2020) was used. Three dimensions, or transfer components, were selected to allow for comparison against the SF method.

The waveform Energy (section 3.2.1.2.1) was used for the SF method as it is a measure of the total magnitude of the sound wave and therefore is representative of the acoustic impedance of the process material at the vessel wall. Other features, such as the maximum or minimum peaks, position of peaks, skewness or kurtosis, are useful to monitor changes in the waveform shape and aid identifying multiple overlapping sound waves. Therefore, the trend in these other features does not follow changes in the process material. Features similar to the waveform Energy that monitor the process material properties include the Sum Absolute Amplitude or Sum Root Amplitude, however, these give greater or lesser weighting to larger amplitudes, respectively. Therefore, the discrepancy between the Energy and either of these features would be due to the shape of the waveform rather than the changing process material. As the shape of the waveform is unlikely to follow the same trends across domains, inclusion of these features may degrade model accuracy.

For the SF method, the features of each domain were standardised. This was to align and scale the feature spaces so that the ML model trained on the source domain could predict accurately on the target domain data. Furthermore, for the honey-water blending experiments, prior to standardisation, the waveform Energy of the first data point in each run was subtracted from all data points of that run so that they all began at a waveform Energy of 0. For this dataset, the process material being measured at the start of each run is known to be honey which has a greater density than the water, meaning it settles at the vessel base where the sensors are located. This is analogous to any industrial process where the same process material is located at the sensor measurement area at the start of each run. This procedure further aligned the feature spaces to combat the varying temperature range the honey-water mixing experiments were conducted over. As the laboratory set point temperature was not altered for the pipe section cleaning experiments, this additional operation was not performed. Although, no feature extraction was performed during the SF method, this is still classed as domain adaptation. Domain adaptation methodologies are used to increase the accuracy of the model trained on the source domain to predict on the target domain (Ben-David et al., 2010), and therefore common feature extraction to both domains is not necessitated. For example, the MMD is a commonly used processing step in domain adaptation methodologies (Li et al., 2019; Li, Zhang, and Deng, 2018; Guo et al., 2019; Lu et al., 2017; Geng, Tao, and Xu, 2011) where a function is found to minimise the

MMD metric between two domains. This therefore aligns the two feature distributions whilst not extracting common features to both domains.

3.2.4.2 Labelled domain adaptation

In Section 8, three domain adaptation investigations were conducted; network training on both datasets simultaneously (Simultaneous training), network training in a federated learning set-up (Federated learning), and fine-tuning (Fine-tuning) of the best performing previously trained networks on the target domain (industrial scale) dataset (Figure 8). The purpose was to improve ML model accuracy on the target fermentation dataset through transfer learning from the laboratory fermentation dataset.

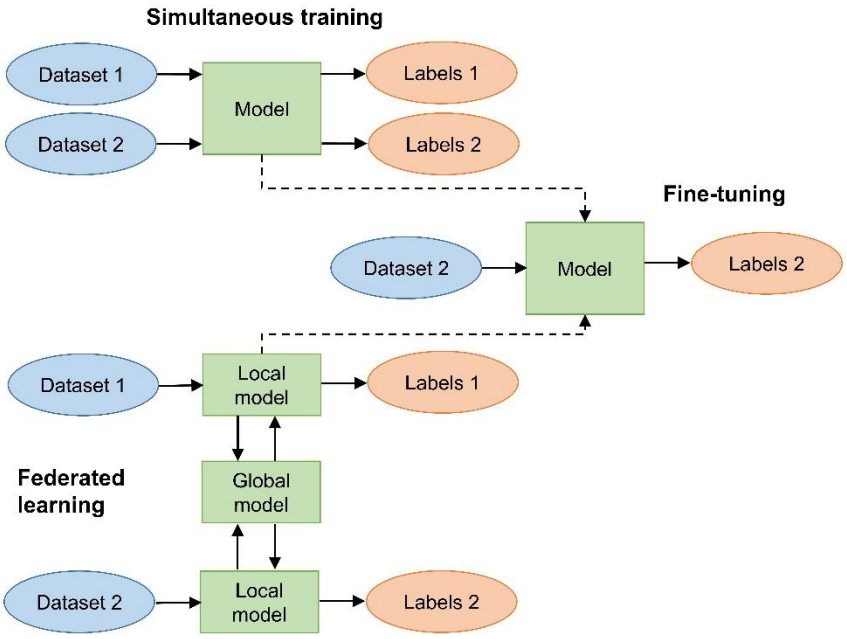


Figure 8: A depiction of the three labelled domain adaptation methodologies investigated.

Federated learning is a ML technique where multiple datasets can be used to collaboratively train a model without sharing the raw data with one another. Each dataset is used to train a local model and the model weights are aggregated to update a global model (McMahan et al., 2017). For the federated learning investigations, local models were trained on each dataset and a weighting factor was applied to the resulting local network weights before being summed to produce a global model. The global model weights were used as the initialisation weights for the next epoch of local network training. After training, the global model was evaluated on the test set. The weighting factors were changed depending on the number of industrial fermentation runs present in the training set.

Finally, fine-tuning the best performing models on the target domain data was assessed. As the models were used to monitor the industrial scale fermentations, the final models did not need to be accurate on the source domain laboratory scale fermentations. Therefore, after initial training to transfer knowledge from the source domain, fine-tuning on the target domain can increase model accuracy on the industrial scale data. All network weights were tuned. Preliminary investigations froze the model weights for the fully connected and LSTM layers and only tuned the output layers. However, this resulted in lower accuracy models on the validation sets than when all weights were allowed to be updated.

3.3 References

Alwosheel, A., Van Cranenburgh, S., Chorus, C.G. (2018) ‘Is your dataset big enough? Sample size requirements when using artificial neural networks for discrete choice analysis.’ *Journal of Choice Modelling* 28, 167-182. Doi:10.1016/j.jocm.2018.07.002.

Becker, T., Mitzscherling, M., Delgado, A. (2002) ‘Hybrid data model for the improvement of an ultrasonic-based gravity measurement system’ *Food Control* 13(4-5), 223–233. Doi:10.1016/S0956-7135(01)00104-9.

Ben-David, S., Blitzer, J., Crammer, K., Kulesza, A., Pereira, F., and Wortman Vaughan, J. (2010). A theory of learning from different domains. *Machine learning*, 79(1-2), 151-175. Doi: 10.1007/s10994-009-5152-4.

Bosse, S., Weiss, D., Schmidt, D. (2021) ‘Supervised distributed multi-instance and unsupervised single-instance autoencoder machine learning for damage diagnostics with high-dimensional data—a hybrid approach and comparison study’ *Computers* 10, 34. Doi:10.3390/computers10030034.

Bowler, A., Pound, M., Watson, N. (2021) ‘Convolutional feature extraction for process monitoring using ultrasonic sensors’ *Computers and Chemical Engineering* 155. Doi: 10.1016/j.compchemeng.2021.107508.

Bowler, A.L., Pound, M.P., Watson, N.J. (2022) ‘A review of ultrasonic sensing and machine learning methods to monitor industrial processes’ *Ultrasonics* 124. Doi:10.1016/j.ultras.2022.106776.

Braeken, J., Van Assen, M.A.L.M. (2017) ‘An empirical Kaiser criterion. Psychological methods’, 22(3), 450-466. Doi:10.1037/met0000074.

Buurman, C., Resoort, G., Plaschkes, A. (1986) ‘SCALING-UP RULES FOR SOLIDS SUSPENSION STIRRED VESSELS’ *Chemical Engineering Science* 41(11), 2865-2871. Doi:10.1016/0009-2509(86)80017-3.

Caesarendra, W., Tjahjowidodo, T. (2017). ‘A Review of Feature Extraction Methods in Vibration-Based Condition Monitoring and Its Application for Degradation Trend Estimation of Low-Speed Slew Bearing’, *Machines*, 5(4), 21. Doi:10.3390/machines5040021.

Caruana, R. (1997) ‘Multitask Learning’ *Machine Learning* 28, 41–75. Doi: https://doi.org/10.1023/A:1007379606734.

Cau, F., Fanni, A., Montisci, A., Testoni, P., Usai, M. (2005) ‘Artificial neural networks for non-destructive evaluation with ultrasonic waves in not accessible’ *IEEE Industry Applications Society* 1, 685–692. Doi:10.1109/IAS.2005.1518382.

Chen, B. Callens, D. Campistron, P., Moulin, E., Debreyne, P., Delaplace, G. (2019) ‘Monitoring cleaning cycles of fouled ducts using ultrasonic coda wave interferometry (CWI)’ *Ultrasonics* 96, 252-260. Doi:10.1016/j.ultras.2018.12.011.

David, C.C., Jacobs, D.J. (2014) ‘Principal component analysis: A method for determining the essential dynamics of proteins’ *Methods in Molecular Biology* 1084, 193-226. Doi: 10.1007/978-1-62703-658-0_11.

Escrig, J., Woolley, E., Rangappa, S., Simeone, A., Watson, N.J. (2019) ‘Clean-in-place monitoring of different food fouling materials using ultrasonic measurements’ *Food Control*, 104, 358-366. Doi:10.1016/j.foodcont.2019.05.013.

Escrig, J., Simeone, A., Woolley, E., Rangappa, S., Rady, A., Watson, N.J. (2020a) 'Ultrasonic measurements and machine learning for monitoring the removal of surface fouling during clean-in-place processes' *Food Bioproducts Process*, 123, 1-13. Doi:10.1016/j.fbp.2020.05.003.

Escrig, J., Woolley, E., Simeone, A., Watson, N.J. (2020b) 'Monitoring the cleaning of food fouling in pipes using ultrasonic measurements and machine learning' *Food Control* 116, Article 107309. Doi:10.1016/j.foodcont.2020.107309.

Ford, C. (2015) 'Is R-squared Useless?'. Available online: https://data.library.virginia.edu/is-r-squared-useless/ (accessed on 3 January 2023).

Gao, F., Li, B., Chen, L., Shang, Z., Wei, X., He, C. (2021) 'A softmax classifier for high-precision classification of ultrasonic similar signals' *Ultrasonics* 112, 106344. Doi:10.1016/j.ultras.2020.106344.

Gao, X., Shi, Y. Zhu, Q., Li, Z., Sun, H., Yao, Z., Zhang, W. (2021) 'Domain Adaptation in Intelligent Ultrasonic Logging Tool: From Microseismic to Pulse-Echo' *IEEE Transactions on Instrumentation and Measurement* 70, 9317751. Doi:10.1109/TIM.2021.3050154.

Geng, B., Tao, D., Xu, C., (2011) 'DAML: Domain adaptation metric learning'. *IEEE Transactions on Image Processing* 10, 2980–2989. Doi: 10.1109/TIP.2011.2134107.

Gopalakrishnan, K., Rautela, M., Deng, Y. (2021) 'Deep learning based identification of elastic properties using ultrasonic guided waves' *Lecture Notes in Civil Engineering* 128, 77–90. Doi:10.1007/978-3-030-64908-1_8.

Guo, L., Lei, Y., Xing, S., Yan, T., Li, N. (2019) 'Deep Convolutional Transfer Learning Network: A New Method for Intelligent Fault Diagnosis of Machines with Unlabeled Data' *IEEE Transactions on Industrial Electronics* 9, 7316–7325. Doi: 10.1109/TIE.2018.2877090.

Hale, J. (2020) Which Evaluation Metric Should You Use in Machine Learning Regression Problems? Available online: https://towardsdatascience.com/which-evaluation-metric-should-you-use-in-machine-learning-regression-problems-20cdaef258e (accessed on 17 April 2023).

Henning, B., Rautenberg, J. (2006) 'Process monitoring using ultrasonic sensor systems' *Ultrasonics* 44, e1395–e1399. Doi:10.1016/j.ultras.2006.05.048.

Hochreiter, S., Schmidhuber, J. (1997) 'Long short-term memory', *Neural Computation* 9(8), 1735–1780. Doi:10.1162/neco.1997.9.8.1735.

Hussein, W.B., Hussein, M.A., Becker, T., (2012) 'Robust spectral estimation for speed of sound with phase shift correction applied online in yeast fermentation processes' *Engineering in Life Sciences* 12(6), 603–614. Doi:10.1002/elsc.201100183.

Ioffe, S., Szegedy, C., (2015) 'Batch Normalization: Accelerating Deep Network Training by Reducing Internal Covariate Shift' *Proceedings of the 32nd International Conference on Machine Learning* 1, 448-456.

Jung, Y., Hu, J. (2015) 'A K-fold averaging cross-validation procedure' *Journal of Nonparametric Statistics* 27(2), 167-179. Doi:10.1080/10485252.2015.1010532.

Khalid, S., Khalil, T., Nasreen, S. (2014) 'A survey of feature selection and feature extraction techniques in machine learning' *Proceedings of 2014 Science and Information Conference* 372–378. Doi:10.1109/ SAI.2014.6918213.

Khyam, M.O., Ge, S.S., Li, X., Pickering, M.R. (2017) 'Highly Accurate Time-of-Flight Measurement Technique Based on Phase-Correlation for Ultrasonic Ranging' *IEEE Sensors Journal*, 17(2), 434-443. Doi: 10.1109/JSEN.2016.2631244.

Kouw, W.M., Loog, M. (2021) 'A Review of Domain Adaptation without Target Labels' *IEEE Transactions on Pattern Analysis and Machine Intelligence* 43(3), 766-785. Doi:10.1109/TPAMI.2019.2945942.

Li, X., Zhang, W., Ding, Q., Sun, J.-Q., (2019) 'Multi-Layer domain adaptation method for rolling bearing fault diagnosis' *Signal Processing* 157, 180–197. Doi:10.1016/j.sigpro.2018.12.005.

Li, X., Zhang, W., Ding, Q. (2018) 'A robust intelligent fault diagnosis method for rolling element bearings based on deep distance metric learning' *Neurocomputing* 310, 77–95. Doi: 10.1016/j.neucom.2018.05.021.

Lim, H.J., Sohn, H. (2020) 'Online stress monitoring technique based on lamb-wave measurements and a convolutional neural network under static and dynamic loadings' *Experimental Mechanics* 60 (2), 171–179. Doi:10.1007/s11340- 019-00546-8.

Lu, W., Liang, B., Cheng, Y., Meng, D., Yang, J., Zhang, T. (2017) 'Deep Model Based Domain Adaptation for Fault Diagnosis' *IEEE Transactions on Industrial Electronics* 3, 2296–2305. Doi: 10.1109/TIE.2016.2627020.

Machine Learning Mastery (2019). Available online: https://machinelearningmastery.com/handle-long-sequences-long-short-term-memory-recurrent-neural-networks/ (accessed on 12 January 2022).

Mallat, S.G. (1989) 'A Theory for Multiresolution Signal Decomposition: The Wavelet Representation' *IEEE Transactions on Pattern Analysis and Machine Intelligence*, 11, 674–693. Doi:10.1109/34.192463.

Mallat, S.G., Mallat, C. (1999) 'IV Time meets frequency' *A Wavelet Tour of Signal Processing* Elsevier Science & Technology: Amsterdam, The Netherlands, Volume 2, 67–124.

Mathworks (2021a). Available online: https://uk.mathworks.com/help/deeplearning/ug/choose-a-multilayer-neural-network-training-function.html (accessed 12 January 2022).

McMahan, H.B., Moore, E., Ramage, D., Hampson, S., and Agüera y Arcas, B. (2017) 'Communication-Efficient Learning of Deep Networks from Decentralized Data' *Proceedings of the 20th International Conference on Artificial Intelligence and Statistics (AISTATS)* 1273-1282.

Mathworks (2021b). Available online: https://uk.mathworks.com/help/stats/fitcsvm.html (accessed 12 January 2022).

Meng, R., Zhou, J., Ye, X., Liu, D. (2012) 'On-line monitoring of yogurt fermentation using acoustic impedance method' *Applied Mechanics and Materials* 101-102, 737-742. Doi:10.4028/www.scientific.net/AMM.101-102.737.

Miao, C., Wang, Y., Zhang, Y., Qn, J., Zuo, M.J., Wang, X. (2008) 'A SVM classifier combined with PCA for ultrasonic crack size classification' *Canadian Conference on Electrical and Computer Engineering* 1627–1630. Doi:10.1109/CCECE.2008.4564817.

Munir, N., Kim, H.-J., Park, J., Song, S.-J., Kang, S.-S. (2019) 'Convolutional neural network for ultrasonic weldment flaw classification in noisy conditions' *Ultrasonics* 94, 74–81. Doi:10.1016/j.ultras.2018.12.001.

Nguyen, T., Park, S. (2020) 'Intelligent Ultrasonic Flow Measurement Using Linear Array Transducer with Recurrent Neural Networks' *IEEE Access* 8, 137564 – 137573. Doi:10.1109/ACCESS.2020.3012037.

Pan, S.J., Tsang, I.W., Kwok, J.T., Yang, Q. (2011) 'Domain adaptation via transfer component analysis' *IEEE Transactions on Neural Networks* 22(2), 199-210. Doi: 10.1109/TNN.2010.2091281.

Perez Alvarado, F.A., Hussein, M.A., Becker, T. A (2016) 'Vision System for Surface Homogeneity Analysis of Dough Based on the Grey Level Co-occurrence Matrix (GLCM) for Optimum Kneading Time Prediction' *Journal of Food Process Engineering* 39, 166–177. Doi: 10.1111/jfpe.12209.

Pyle, R.J., Bevan, R.L.T, Hughes, R.R., Rachev, R.K., Ali, A.A.S, Wilcox, P.D. (2021) 'Deep learning for ultrasonic crack characterization in NDE' *IEEE Transactions on Ultrasonics, Ferroelectrics, and Frequency Control* 68 (5), 1854–1865. Doi:10.1109/TUFFC.2020.3045847.

Qin, Y., Ma, M., Zhu, E., Mao, Z., Haile, M., Shiao, M., Chen, T.-K. (2018) 'Temperature compensation of ultrasonic guided waves via recurrent neural network' *Proceedings of the 7th Asia-Pacific Workshop on Structural Health Monitoring, APWSHM 2018* 402–409.

Rautela, M., Senthilnath, J., Moll, J., Gopalakrishnan, S. (2021) 'Combined two-level damage identification strategy using ultrasonic guided waves and physical knowledge assisted machine learning' *Ultrasonics* 115, 106451. Doi:10.1016/j.ultras.2021.106451.

Ren, W., Jin, N., OuYang, L., Zhai, L., Ren, Y. (2021) 'Gas volume fraction measurement of oil-gas-water three-phase flows in vertical pipe by combining ultrasonic sensor and deep attention network' *IEEE Transactions on Instrumentation and Measurement* 70, 1–9. Doi:10.1109/TIM.2020.3031186.

Rodriguez-Galiano, V., Sanchez-Castillo, M., Chica-Olmo, M., Chica-Rivas, M. (2015) 'Machine learning predictive models for mineral prospectivity: an evaluation of neural networks, random forest, regression trees and support vector machines' Ore Geology Reviews 71, 804–818. Doi:10.1016/j. oregeorev.2015.01.001.

Safavian, L.S., Kinsner, W., Turanli, H. (2005) 'A QUANTITATIVE COMPARISON OF DIFFERENT MOTHER WAVELETS' *Canadian Conference on Electrical and Computer Engineering* 1453–1456.

Simeone, A., Woolley, E., Escrig, J., Watson, N.J. (2020) 'Intelligent industrial cleaning: a multi-sensor approach utilising machine learning-based regression' *Sensors* 20, 1–22. Doi:10.3390/s20133642.

Tolstikhin, I., Sriperumbudur, B.K., Schölkopf, B. (2016) 'Minimax estimation of maximum mean discrepancy with radial kernels' *Advances in Neural Information Processing Systems* 1938-1946.

Torrey, L., and Shavlik, J. (2009). 'Transfer Learning'. In Handbook of Research on Machine Learning Applications. Hershey, PA: IGI Global.

Towards Data Science (2019) ‘Data Science in Medicine — Precision & Recall or Specificity & Sensitivity?’. Available online: https://towardsdatascience.com/should-i-look-at-precision-recall-or-specificity-sensitivity-3946158aace1#:~:text=If%20we%20define%20a%20positive,refer%20to%20sensitivity%20as%20recall. (accessed on 3 January 2023).

Tzeng, E., Hoffman, J., Saenko, K.; Darrell, T. (2017) ‘Adversarial discriminative domain adaptation’ *Proceedings CVPR IEEE* 2017, 2962–2971.

Úbeda, M.A., Hussein, W.B., Hussein, M.A., Hinrichs, J., Becker, T.M. (2016) ‘Acoustic sensing and signal processing techniques for monitoring milk fouling cleaning operations’ Engineering in Life Sciences 16(1), 67–77. Doi:10.1002/elsc.201400235.

Valle, S., Li, W., Qin, S.J. (1999) ‘Selection of the number of principal components: The variance of the reconstruction error criterion with a comparison to other methods’ *Industrial and Engineering Chemistry Research* 38(11), 4389-4401. Doi:10.1021/ie990110i.

Wallhäußer, E., Hussein, W.B., Hussein, M.A., Hinrichs, J., Becker, T.M. (2011) ‘On the usage of acoustic properties combined with an artificial neural network – a new approach of determining presence of dairy fouling’ Journal of Food Engineering 103(4), 449–456. Doi:10.1016/j.jfoodeng.2010.11.015.

Wallhäußer, E., Hussein, W.B., Hussein, M.A., Hinrichs, J., Becker, T. (2013) ‘Detection of dairy fouling: combining ultrasonic measurements and classification methods’ Engineering in Life Sciences 13(3), 292–301. Doi:10.1002/elsc.201200081.

Wallhäußer, E., Sayed, A., Nobel, S., Hussein, M.A., Hinrichs, J., Becker, T. (2014) ‘Determination of cleaning end of dairy protein fouling using an online system combining ultrasonic and classification methods’ Food and Bioprocess Technology. 7(2), 506–515. Doi:10.1007/s11947-012-1041-0.

Yan, K. (2020) ‘A domain adaptation toolbox’. Available online: https://www.github.com/viggin/domain-adaptation-toolbox. (accessed on June 14 2020).

Yan, Y., Liu, D., Gao, B., Tian, G.Y., Cai, Z.C. (2020) ‘A deep learning-based ultrasonic pattern recognition method for inspecting girth weld cracking of gas pipeline’ *IEEE Sensors Journal* 20 (14), 7997–8006. Doi: 10.1109/JSEN.2020.2982680.

Yang, Y., Zhang, L., Du, M., Bo, J., Liu, H., Ren, L., Li, X., Deen, M.J. (2021) ‘A comparative analysis of eleven neural networks architectures for small datasets of lung images of COVID-19 patients toward improved clinical decisions’. Computers in Biology and Medicine 139, 104887. Doi: 10.1016/j.compbiomed.2021.104887.

Xiao, Z., Guo, Q., Guo, Y., Huang, Y. (2021) ‘Ultrasonic A-scan image detection for 3D braided composites based on convolutional neural network’ *ICCAI 2021* 8–14. Doi:10.1145/3467707.3467709.

Xu, Y., Goodacre, R. (2018) ‘On Splitting Training and Validation Set: A Comparative Study of Cross-Validation, Bootstrap and Systematic Sampling for Estimating the Generalization Performance of Supervised Learning’ *Journal of Analysis and Testing* 2(3), 249-262.

Zhan, X., Jiang, S., Yang, Y., Liang, J., Shi, T., Li, X. (2015) ‘Inline Measurement of Particle Concentrations in Multicomponent Suspensions using Ultrasonic Sensor and Least Squares Support Vector Machines’ *Sensors* 15, 24109–24124. Doi: 10.3390/s150924109.

Zhan, X., Yang, Y., Liang, J., Zou, D., Zhang, J., Feng, L., Shi, T., Li, X. (2016) ‘In-line mixing states monitoring of suspensions using ultrasonic reflection technique’ *Ultrasonics* 65, 43–50. Doi:10.1016/j.ultras.2015.10.024.

Zhang, Y., Qiang, Y. (2018) ‘An overview of multi-task learning’ *National Science Review* 5(1), 30-43. Doi:10.1093/nsr/nwx105.

Zhu, W.-J., Xu, K.-J., Fang, M., Shen, Z.-W., Tian, L. (2017) ‘Variable ratio threshold and zero-crossing detection based signal processing method for ultrasonic gas flow meter’ *Measurement: Journal of the International Measurement Confederation* 103, 343-352. Doi: 10.1016/j.measurement.2017.03.005.

4 Monitoring Mixing Processes Using Ultrasonic Sensors and Machine Learning

Article title: Monitoring Mixing Processes Using Ultrasonic Sensors and Machine Learning

Journal: Sensors

Date published: 25/03/2020

DOI: 10.3390/s20071813

Authors: Bowler A.L., Bakalis S., Watson N.J.

Author contributions (as published): Conceptualization, A.L.B., and N.J.W.; Methodology, A.L.B., N.J.W Software, A.L.B.; Validation, A.L.B.; Formal Analysis, A.L.B.; Investigation, A.L.B.; Resources, A.L.B., and N.J.W; Data Curation, A.L.B.; Writing—Original Draft Preparation, A.L.B.; Writing—Review and Editing, A.L.B., N.J.W., and S.B.; Visualization, A.L.B.; Supervision, A.L.B., N.J.W., and S.B.; Project Administration, A.L.B., N.J.W.; Funding Acquisition, N.J.W. All authors have read and agreed to the published version of the manuscript.

The aim of this thesis was to develop ML methods to facilitate optimal deployment of US sensors for process monitoring applications in industrial environments. This article contributed to this aim by monitoring two mixing processes: namely, honey-water blending and flour-water batter mixing. This contributed to the thesis objective of collecting US data from process monitoring applications that enable the thesis conclusions to be extended to industrial applications (see Section 1.2 Aims and Objectives). Extensive feature extraction, feature selection, and algorithms were used for classification and regression tasks to aid development of an optimal ML pipeline for process monitoring using US measurements. The novelty of this work was the combination of an industrially relevant non-invasive, reflection-mode US sensing technique with ML to monitor mixing processes. This contributed to the thesis objective of evaluating a non-invasive, reflection-mode US sensing technique that can be externally retrofitted to existing process equipment. Furthermore, the use of Long Short-Term Memory (LSTM) layers and Convolutional Neural Networks (CNN) for process monitoring using US sensors was novel. The main conclusions from this work were that using information from previous time steps was vital for accuracy on most tasks and, specifically, that flexible use of previous time steps was required. This influenced the decision to use LSTM algorithms for the remaining works in this thesis contributing to the thesis objective of determining an optimal ML pipeline for process monitoring using US measurements. LSTM layers were evaluated in all works in this thesis. The LSTMs were used in sequence to sequence tasks, where the sequence of feature extracted from US sensor measurements were used to predict the sequence of output values. At each timestep the model predicted the output value at the current timestep. At the following timestep, the model obtains the new US sensor measurement features from this timestep with which to predict the output value. Therefore, only US sensor measurement features are used to make predictions and previous ground truth data or output predictions are not used. All timesteps for each dataset were used as a single sequence rather than being truncated into multiple sequences of shorter length.

In this work, several feature extraction methods are utilised. No feature selection methods were used in this work. Instead, the methods developed in this article were used to inform feature engineering approaches for future works in this thesis. Firstly, the Sum Absolute

Amplitude and Energy are extracted from the waveforms as these are measures of the acoustic impedance at the vessel-wall interface and therefore follow material changes during mixing at this location (Bowler et al., 2022). The individual amplitude at each sample point in the waveform was also used to provide further information as utilised in Escrig et al. (2020) and Munir et al. (2019). However, the peaks forming a US waveform can shift in the time dimension as the speed of sound through materials changes. Therefore, in the article presented in Section 6 (titled “Convolutional feature extraction for process monitoring using ultrasonic sensors”), a CNN is used to extract features from individual sample point amplitudes as they are robust to spatially varying features (Lecun et al., 2015). Wavelet transformation was also used in this present work, as it has been widely applied with ML problems for US sensor data (Bowler et al., 2022). However, the waveforms in this work are composed of multiple overlapping sound waves. Therefore, the convolutional method presented in Section 6 is preferred to learn waveform features as the wavelet transform may misattribute waveform features to different frequency bands (Bowler et al., 2021). For clarity, the inputs to the CNNs in this work were matrices of sensor data opposed to images (Convolutional Neural Networks section, page 10 of 24).

A limitation of this article was the varying validation approaches utilised for each algorithm (single-fold validation with 20% of total training and validation set size for ANNs, 5-fold cross validation for SVMs, and no validation data for CNNs or LSTMs). This benefits the hyperparameters selected for the SVMs as these underwent the most comprehensive validation approach (the hyperparameters were selected based on five validation evaluations each on a different 20% of the training and validation data). This disadvantages the hyperparameters selected for the LSTMs and CNNs as they were not chosen to generalise to unseen data. However, despite this, SVMs performed worst for all prediction tasks whilst LSTMs performed best for honey-water blending classification (96.3 %) and CNNs performed best for honey-water blending regression (0.977 R^2), flour-water batter classification (92.5 %), and flour-water batter regression (0.976). Therefore, the conclusions drawn from this article were not influenced by this limitation. Furthermore, the use of LSTM layers and CNN for feature extraction were confirmed in Sections 5 and 6 in this thesis.

In this work, a video camera was used to film the honey-water blending process. The labelled data was obtained by viewing the videos to determine the time when mixing was complete for each run. For classification tasks, the time for complete mixing determined the first positive classification (i.e., a fully mixed system). For regression tasks, the time for complete mixing determined zero seconds remaining until a fully mixed system was obtained. Therefore, the labels for previous timesteps were assigned positive regression values for time until complete mixing was achieved. Later timesteps were assigned negative regression values for the time until complete mixing was achieved. A similar method was used for the flour-water batter mixing case study, but the power input to the impeller was used to determine the time for optimal mixing.

When applying this method to larger scale systems, the most important consideration is the location of the US sensors which determine the process phenomena able to be measured. For the honey-water blending datasets, two sensors were used. One sensor was located in the centre of the vessel base and the second was attached to the vessel base but offset from the sensor. The highest R^2 values (up to 0.977) were achieved by combining the inputs from both sensors. This is because the non-central sensor had better prediction ability nearer the beginning of the process as the honey was removed from this sensor measurement area first, and the central sensor had greater resolution nearer the end of the mixing process as the last of the honey was dissolved. Therefore, the location of the sensors should be chosen to obtain representative measurements of the process to be monitored.

Multiple sensors may require use on large vessels where the dynamics at different locations in the vessel vary.

4.1 References

Bowler, A., Pound, M., Watson, N. (2021) ‘Convolutional feature extraction for process monitoring using ultrasonic sensors’ *Computers and Chemical Engineering* 155. Doi: 10.1016/j.compchemeng.2021.107508.

Bowler, A.L., Pound, M.P., Watson, N.J. (2022) ‘A review of ultrasonic sensing and machine learning methods to monitor industrial processes’ *Ultrasonics* 124. Doi:10.1016/j.ultras.2022.106776.

Escrig, J., Simeone, A., Woolley, E., Rangappa, S., Rady, A., Watson, N.J. (2020a) ‘Ultrasonic measurements and machine learning for monitoring the removal of surface fouling during clean-in-place processes’ *Food Bioproducts Process*, 123, 1-13. Doi:10.1016/j.fbp.2020.05.003.

Lecun, Y., Bengio, Y., Hinton, G. (2015) ‘Deep learning’ *Nature* 521, 436–444. Doi:10.1038/nature14539.

Munir, N., Kim, H.-J., Park, J., Song, S.-J., Kang, S.-S. (2019) ‘Convolutional neural network for ultrasonic weldment flaw classification in noisy conditions’ *Ultrasonics* 94, 74–81. Doi:10.1016/j.ultras.2018.12.001.

sensors

MDPI

Article

Monitoring Mixing Processes Using Ultrasonic Sensors and Machine Learning

Alexander L. Bowler, Serafim Bakalis and Nicholas J. Watson *

Faculty of Engineering, University of Nottingham, University Park, Nottingham, NG7 2RD, UK; enxab17@exmail.nottingham.ac.uk (A.L.B.); ezzsb3@exmail.nottingham.ac.uk (S.B.)
* Correspondence: nicholas.watson@nottingham.ac.uk

Received: 19 February 2020; Accepted: 24 March 2020; Published: 25 March 2020

Abstract: Mixing is one of the most common processes across food, chemical, and pharmaceutical manufacturing. Real-time, in-line sensors are required for monitoring, and subsequently optimising, essential processes such as mixing. Ultrasonic sensors are low-cost, real-time, in-line, and applicable to characterise opaque systems. In this study, a non-invasive, reflection-mode ultrasonic measurement technique was used to monitor two model mixing systems. The two systems studied were honey-water blending and flour-water batter mixing. Classification machine learning models were developed to predict if materials were mixed or not mixed. Regression machine learning models were developed to predict the time remaining until mixing completion. Artificial neural networks, support vector machines, long short-term memory neural networks, and convolutional neural networks were tested, along with different methods for engineering features from ultrasonic waveforms in both the time and frequency domain. Comparisons between using a single sensor and performing multisensor data fusion between two sensors were made. Classification accuracies of up to 96.3% for honey-water blending and 92.5% for flour-water batter mixing were achieved, along with R^2 values for the regression models of up to 0.977 for honey-water blending and 0.968 for flour-water batter mixing. Each prediction task produced optimal performance with different algorithms and feature engineering methods, vindicating the extensive comparison between different machine learning approaches.

Keywords: food and drink manufacturing; industry 4.0; digital manufacturing; mixing; ultrasonic sensors; machine learning; convolutional neural networks; long short-term memory neural networks; wavelet transform

1. Introduction

The world is experiencing the fourth industrial revolution where digital technologies such as artificial intelligence, robotics, and the Internet of Things are used to improve the productivity, efficiency and sustainability of manufacturing processes [1,2]. This transformation is underpinned by the enhanced collection and use of data, and therefore sensors are one of the most important technologies in Industry 4.0 [3]. Although sensors exist for basic measurements such as temperature and pressure, there is a need for more advanced techniques that can monitor materials and processes. Mixing is one of the most common manufacturing processes. It is not only used for combining materials, but also for increasing heat and mass transfer, providing aeration, and suspending solids. Correct active ingredient dosing in the pharmaceutical industry is critical for patient safety and treatment effectiveness and effective mixing is essential to achieve this. In food manufacturing, mixing provides uniform heating and modifies material structure. In material manufacturing such as the polymer, cement, and rubber industries, final product qualities are determined by the level of homogeneity [4]. Sensors that provide automatic, real-time data acquisition capabilities are required to monitor critical processes such as mixing. These sensors are termed in- or on-line, where in-line methods directly

measure the process material with no sample removal, and on-line methods automatically take samples to be analysed without stopping the process [5]. Sensors able to characterise whether a mixture is non-mixed or fully mixed offer benefits of reducing off-specification products, early identification of process upset conditions, and reduced resource consumption from overmixing. Furthermore, techniques able to predict the required time remaining until mixing completion would improve batch scheduling and therefore process productivity.

There are numerous in-line and on-line techniques available to monitor industrial mixing processes, with the major categories of techniques being point property measurements, tomographic (e.g., electrical resistance tomography), and spectroscopic (e.g., Near Infrared Spectroscopy (NIRS)). Discussion of the aptitude of each technique to different mixing applications is provided in [4]. Active acoustic techniques introduce sound waves into a material or system by converting electrical signal pulses into pressure waves using piezoelectric transducers. Either a single transducer sends and receives the sound wave after reflection from an interface (pulse–echo mode) or a second transducer receives the sound wave after it has been transmitted through the material (pitch catch mode) [6]. Low power, high frequency sound waves in the ultrasonic frequency range are used for material characterisation, and do not affect the structure of the material [6]. Typical ultrasonic parameters measured to characterise a system include the speed of sound, sound wave attenuation, and the material's acoustic impedance. The speed of sound through the material is dependent on its density and compressibility, and is calculated by measuring the time of flight of the sound wave. The attenuation of the sound wave can be measured as a decrease in the signal amplitude, and is caused by sound wave scattering, reflection, or energy dissipation. The acoustic impedance is dependent on the speed of sound and density of the material, and the proportion of reflected sound wave from a material boundary is dependent on the magnitude of the acoustic impedance mismatch between the neighbouring materials [7]. Ultrasound sensors are low-cost, real-time, in-line, and capable of operating in opaque systems. However, the large changes in acoustic impedance when transmitting from liquid or solid to gas causes strong reflection of the sound wave, making transmission difficult in the presence of gas bubbles. Furthermore, the speed of sound in a material is strongly dependent on temperature [7]. Ultrasound has found application for material characterisation in industries such as food, chemicals, pharmaceuticals, and biotechnology [6–9].

Several studies have used ultrasonic measurements to monitor mixing. However, many of these require transmission of the sound wave through the mixture in order to measure the speed of sound or attenuation. Stolojanu and Prakash [10] used two invasive transducers in the pitch–catch mode to characterise glass bead suspensions up to concentrations of 45 wt % in a laboratory scale mixing system. The ultrasonic velocity, attenuation, and peak frequency shift were used to determine particle concentration and size. Both Ribeiro et al. [11] and Yucel and Coupland [12] used two non-invasive transducers in the pitch–catch mode to characterise laboratory scale systems. However, transmission-based measurements are unable to be used for most mixing systems at the industrial scale. Firstly, the increased distance that the sound wave must travel increases the attenuation of the signal. Secondly, industrial mixtures are typically more complex than simple model systems tested at laboratory scale. The number of materials being mixed in industrial mixers creates an increased number of heterogeneities causing scattering and reflection of the sound, or the presence of gas bubbles cause strong reflection of the sound wave. These also contribute to greater attenuation of the signal and transmission becomes more difficult without high power, high cost transducers.

Bamberger and Greenwood [13] mounted pitch–catch mode transducer pairs to a probe to monitor solids suspension in an industrial slurry mixing tank. However, this technique was invasive and the attenuation correlation with solids concentration was only possible over the short sound wave propagation distance. Sun et al. [14] monitored the dispersion homogeneity of calcium carbonate in polypropylene during extrusion. Two transducers in the pitch–catch mode measured the ultrasound attenuation. Again, this was invasive and transmission was only possible due to the short sound wave propagation distance. Fox et al. [15] and Salazar et al. [16] used the invasive pulse–echo

mode ultrasound probes to monitor air incorporation into aerated batters during mixing. Due to the strong reflectance of sound waves caused by gas bubbles, transmission was not possible. The acoustic impedance of the probe-batter interface was measured to determine the optimal mixing time. Hunter et al. [17] and Bux et al. [18] used intrusive pulse–echo transducers to monitor particle suspension. Acoustic backscatter techniques were used to measure speed of sound and attenuation, where the reflected sound wave from the particles was measured opposed to transmission through the suspension. Invasive techniques suffer from problems such as probe fouling, probe breakage, and difficulty in installation, thereby limiting their appeal in industrial settings. Ultrasound is applicable for non-invasive measurement by transmitting the sound wave through the wall of the vessel. Therefore, this current work uses a non-invasive, pulse–echo ultrasound technique to monitor mixing, which requires no sound wave transmission through the mixture being characterised. The only examples of non-invasive, no-transmission ultrasonic sensors for mixing processes are those used to monitor particle suspension. Buurman et al. [19] used non-invasive ultrasonic Doppler velocimetry to detect whether particles were suspended at the bottom of an opaque mixing vessel to monitor particle suspension. Zhan et al. [20] used a non-invasive pulse–echo transducer attached to the base of the vessel to monitor particle suspension by measuring the acoustic impedance of the base-suspension interface.

In this study, two laboratory-scale mixing systems are monitored: honey-water mixing and flour-water batter mixing. These two model systems were selected to show the application of ultrasonic sensors to monitor different mixing processes. As honey is completely miscible in water, this system is representative of the development of homogeneity in liquid–liquid blending. Flour-water batter was used in this study to monitor structural changes as the gluten proteins in the flour become hydrated and aligned into a network, as opposed to air incorporation as investigated in Fox et al. [15] and Salazar et al. [16]. Therefore, this flour-water batter system is similar to dough mixing, only with higher water content. This system was chosen as during dough mixing at atmospheric pressure, the dough pulls away from the mixer sides and is therefore not measurable using low-power ultrasound due to the created air gap. However, industrial dough mixing is typically performed at reduced pressure or vacuum pressure, where the dough will be in contact with the mixer sides. Furthermore, batter mixing has been shown to follow the same physical and chemical changes as dough during mixing, and is therefore representative of industrial dough mixing [21].

For in-line industrial process monitoring, suitable signal processing and interpretation is required for automatic process diagnosis. Supervised Machine Learning (ML) maps input data to output classes (classification) or values (regression) during training so that it may then be used to predict outputs from new input data. The advantage of ML is the ability to fit functions to input–output relationships without the need to define the often complex underlying physical models. The success of ML models is dependent on the input feature variables used to make predictions. A received ultrasonic waveform consists of an amplitude at each time period sample. From this waveform, useful features are typically manually engineered, e.g., selecting the maximum waveform amplitude, or monitoring the speed of the sound wave. This approach of using manually engineered features is termed shallow ML. Ultrasonic measurements have been combined with shallow ML algorithms such as Artificial Neural Networks (ANNs) [22–29] and Support Vector Machines (SVMs) [23,25,30,31], using waveform features from the time domain [23,25,27,31,32] and frequency domain [24,27,31,32] after analyses such as wavelet transforms [22,24]. These have been used for applications such as predicting sugar concentration during fermentation [33], measuring particle concentration in multicomponent suspensions [34], and classification of heat exchanger fouling in the dairy industry [23,25]. There are no examples of using ultrasonic measurements and ML to follow a mixing process; however, El-Hagrasy et al. [35] used the Soft Independent Modelling of Class Analogies (SIMCA) and Principal Component Modified Bootstrap Error-adjusted Single-sample Technique (PC-MBEST) algorithms to analyse NIRS spectra during pharmaceutical solids blending. Typically, shallow ML requires some expertise of the sensor signal to engineer useful features from the raw data. In contrast, Convolutional Neural Networks (CNNs) utilise representation learning, which requires no manual feature engineering by transforming the raw

data into higher, more abstract levels to automatically extract features [36]. CNNs use convolutional filters to measure the spatial relationship data values and have found application in image recognition tasks [37,38]. CNNs have also been used to improve ML prediction from ultrasonic signals in both the time [26] and frequency domain after the wavelet transform [39]. The focus of this study is to compare different feature engineering methods and ML algorithms to classify the mixture state and predict the time remaining until mixing completion for two model mixing systems. ANNs, SVMs, and Long Short-Term Memory (LSTM) neural network shallow ML algorithms are compared with CNNs. The wavelet transform will also be investigated to provide the frequency content of the waveforms as inputs to the ML models. The sensors used in this current work only characterise material close to the vessel wall and therefore the potential for non-representative readings must be investigated. This is achieved by comparing the results from multiple low-cost sensors distributed around the vessel along with data fusion between the sensors. Multisensor data fusion is the combination of measurements from multiple sensors to produce improved analysis over that which could be achieved by using the data from each sensor independently.

2. Materials and Methods

2.1. Experimental

2.1.1. Ultrasound

Two magnetic transducers of 1 cm^2 active element surface area with a 5 MHz resonance (M1057, Olympus, -6 DB bandwidth—116.43%) were externally mounted to the mixing vessels. The transducers were attached to adhesive magnetic strips on the outside of the vessels and coupling gel (Proceq ultrasound couplant) was applied between the sensor and strip (Figure 1a). The transducers were used in the pulse–echo mode to both transmit and receive the ultrasonic signal. The ultrasound wave is transmitted through the coupling gel, magnetic strip, adhesive, vessel wall, and mixture. At each interface between different materials, a part of the sound wave continues through to the second material and a part of it is reflected. The proportion of the sound wave reflected at the interface is dependent on the acoustic impedance mismatch between the two materials (Equation (1)) [40]. The ultrasound wave of interest is that reflected from the vessel–mixture interface. The acoustic impedance is a product of the material density and speed of sound (Equation (2)) [40]. Therefore, the acoustic impedance mismatch between the mixture and vessel materials is the parameter being measured; this is depicted by the wave returning from the vessel–mixture interface to the transducer in Figure 1a. During the mixing processes, the composition of the mixture in contact with the vessel wall at the sensor measurement area will change. This will change the acoustic impedance mismatch between the vessel wall and mixture, and therefore this technique is suitable to monitor the mixing process. The portion of the wave that continues through the mixture is either dissipated due to attenuation or reflects from additional boundaries present in the mixture. However, only reflections from boundaries approximately perpendicular to the direction of wave propagation will be received by the transducer. Due to the thin thickness of the vessel wall, the ultrasonic technique is operating in the near field. The near field is the region closest to the transducer element and has a complex structure due to constructive and destructive interference of multiple waves generated by the transducer surface. Although performing measurements in the near field is challenging, it is possible as highlighted by the work of Escrig et al. 2019 [8].

$$R = \frac{A_r}{A_i} = \frac{Z_1 - Z_2}{Z_1 + Z_2} \quad (1)$$

$$Z_i = C_i \rho_i, \quad (2)$$

where R is the reflection coefficient, A_r the amplitude of the reflected wave, A_i the amplitude of the incident wave, Z_1 and Z_2 the acoustic impedance of the material the wave is travelling from (material

1) and the material the wave is travelling into (material 2). C_i is the speed of sound in material i, and ρ_i the density of material i. These parameter values for the tested materials are provided in Table 1.

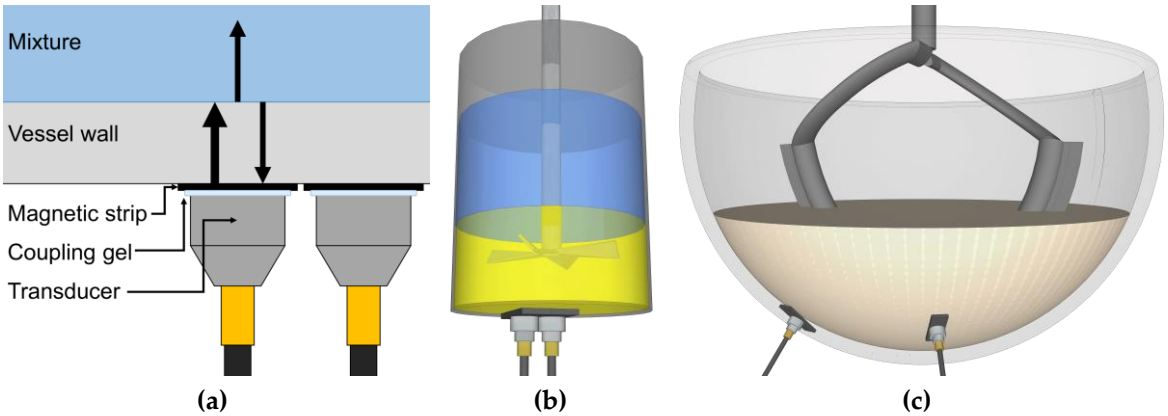


Figure 1. (**a**) At each boundary between two materials, part of the ultrasound wave is reflected. The proportion of the sound wave energy reflected is dependent on the acoustic impedance mismatch between the materials. The ultrasound wave of interest in this study is that reflecting back from the vessel–mixture interface to the transducer. (**b**) Diagram showing the position of the two ultrasound sensors on the vessel base to monitor honey-water blending. (**c**) Diagram showing the position of the two ultrasound sensors on the side of the mixing bowl to monitor the flour-water batter mixing.

Table 1. The speed of sound and density of the materials used in the honey-water blending experiments. The acoustic impedance is a product of these two values (Equation (2)). The reflection coefficient is the proportion of the sound wave reflected at the glass vessel wall and mixture interface.

Material	Speed of Sound (m/s)	Density (kg/m^3)	Acoustic Impedance ($\times 10^6$ Pa.s/m^3)	Reflection Coefficient
Water	1493 [41]	998	1.49	0.79
Honey	2125 [42]	1420 [43]	3.02	0.61
Well-mixed flour-water mixture (50 wt % water)	2000 [44,45]	1230 [46]	2.46	0.67
Glass, Pyrex	5640 [47]	2210 [48]	12.46	-

2.1.2. Honey-Water Blending

The ultrasonic transducers were mounted on the base of a 250 mL glass vessel. This was because a flat surface was needed to allow full contact of the sensors and the magnetic strip. It was not possible to mount the transducers on the side of the vessel due to its curvature. As the sensors only measure a small area of material properties in a single location, they may be designated as point property measurement techniques. Therefore, the positioning of the sensors is of paramount importance to obtain useful readings; for example, multiple NIR sensors have been used to monitor different mixing dynamics during particulate blending across different locations in a mixer [49,50]. Therefore, one sensor was located at the centre of the vessel base and the other was closer to the vessel sides (Figure 1b), allowing comparison between both sensor positions. Furthermore, sensor fusion could be explored by combining outputs from both sensors to improve ML prediction. An ultrasound box (Lecoeur Electronique) was used to excite the transducers by providing electrical pulses and digitisation of the received signals (2048 bit resolution). A temperature sensor was taped to the base of vessel and connected to a PT-104 Data Logger (Pico Technology) to monitor local temperature. The ultrasound box and temperature data logger were connected to a laptop and bespoke MATLAB software controlled the hardware components and acquired the data. Before beginning the experiments, the change in waveform in the presence of water was used to select the portion of the signal reflecting from the

vessel–material interface. The ultrasonic gain was then set to maximise the resolution of this portion of the reflected waveform. The number of samples was then also selected by monitoring the last sample point where a change in waveform could be seen visually using this test. Signals were acquired continuously for 1 s from each probe consecutively. Two ultrasonic waveforms were recorded during each 1 s time period. For the ANNs, SVMs, and LSTMs the signal was averaged over this 1 s interval before applying feature engineering to minimise the effect of signal noise. However, for the CNNs every waveform collected during the 1 s intervals was used in training to maximise the number of images that the network had available for training. The sampling frequency was set to 160 MHz to maximise waveform resolution. A mobile phone camera was used to film each mixing process to be later used to determine the time for mixing completion, defined as the time when the honey had fully dissolved. The location for the last part of the honey to be fully mixed was in the centre of the vessel base. This determination of time for mixing completion is the ground truth data to label the output of each ultrasonic waveform. From this labelled data, the ML models can be trained and tested to predict the mixture state (non-mixed or mixed) or predict the time remaining until mixing completion. Pure clear honey (Wm Morrison Supermarkets plc) and tap water were loaded into the vessel at the start of each mixing process. Two different volumes of honey were used for the experiments: 20 and 30 mL. A constant volume of 200 mL tap water was used throughout. An overhead stirrer with a cross-blade impeller was used to stir the mixture. The impeller speed was also set to values of either 200 or 250 rpm. These four parameter permutations were repeated three times across one day while varying the laboratory thermostat set point to produce a temperature variation from 19.3 to 22.1 °C. This induced variability in process parameters was performed to enable the ML models to generalise.

2.1.3. Flour-Water Batter Mixing

NIRS has previously been used to monitor the chemical and surface structure changes occurring during dough mixing [51–53] and image analysis of the dough surface has been used to determine optimal mixing time [54]. Measuring the power or torque supplied to the impeller is a common method of monitoring dough mixing. Mixing should be stopped at the maximum power input for optimal bread properties [54]. Beyond this point of maximum resistance to extension, the gluten network begins to breakdown. The standard deviation of the power measurement has also been found to peak at the optimal dough consistency [55]. Ultrasound has previously been used to characterise the effects of mixing on dough. The relevant literature is included in the Results section. However, none of these ultrasonic techniques have used an in-line monitoring system similar to that used in this investigation. The power supplied to the motor was monitored using a YouThink plug socket power meter to provide a reference measurement for the mixture's state. The optimal mixing time was determined by the time of maximum power drawn to the impeller. From this, an output value for each sensor signal can be labelled with the ground truth data to then train and test the ML models. The same transducers as used in the honey-water blending experiments were attached to the outside of a stand-mixer glass mixing bowl (1000 W Kenwood kmix kmx754). Due to the curvature of the vessel, the adhesive was not sufficient to hold the magnetic tape onto the sides. Therefore, the magnetic tape was attached to the vessel using silicone vacuum grease and electrical tape. The transducers were located close to the base of the vessel to reduce the likelihood of an air gap caused by the dough pulling away from the mixing bowl sides when close to the optimal mixing time. As both sensors were located at the same height on the mixing bowl, a comparison between different sensor positions could not be made. This is because the mixing was similar at all radial positions in the vessel for the same height position. Rather, the ability of one sensor versus two sensors to monitor the mixing process could be evaluated. The quantity of strong white flour (Wm Morrison Supermarkets plc) and tap water was varied between 450 and 450 g, 500 and 450 g, and 500 and 400 g, respectively. Each combination was repeated three times producing nine runs. Again, this induced variability between runs allowed the examination of the ability of the ML algorithms to generalise across process parameters. A creaming beater was chosen as the attachment to prevent non-mixed zones and the formation of a fouling layer at the surface of the

sensor measurement. The minimum impeller speed was used for 1 min to incorporate all of the flour into the water and was then increased to Speed 2 for the remainder of the process. The environmental temperature varied between 19.4 and 21.3 °C. The signal gain and number of sample points were selected in the same process as described in the honey-water blending section. Again, ultrasound signals were acquired continuously for 1 s from each probe sequentially at a sampling frequency of 160 MHz and treated in the same way as previously described.

2.2. Data Analysis

All data analysis and ML algorithms were completed in MATLAB R2019a.

2.2.1. Waveform Preprocessing

The ultrasonic signal was first windowed to select the useful information from the waveform to use in subsequent calculations and machine learning tasks. Visual inspection of the waveform was used to identify and remove the saturated part of the waveform (sample points 1 to approximately 210 in Figure 2a) corresponding to sound wave reflections prior to the vessel–mixture interface. The part of the waveform displaying no further amplitude change between non-mixed and well-mixed materials (sample point 700 to the end of the waveform in Figure 2a) was also identified and removed from the final waveform (Figure 2b).

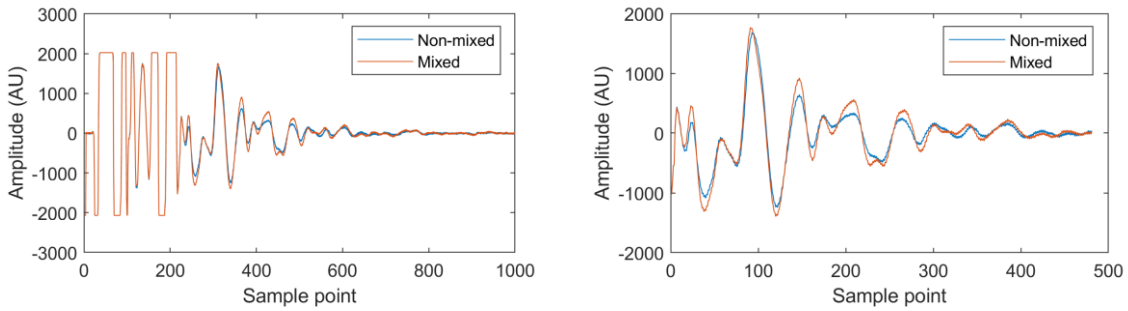


Figure 2. The example waveforms presented are the starting and final waveforms collected from the first run of the honey-water blending experiments. Corresponding to a non-mixed material state and a fully-mixed state, respectively. (**a**) The non-windowed waveform, with the signal saturated from sample points 1 to 210 and containing no further useful information from sample point 700 to the end of the waveform. (**b**) The windowed waveform.

2.2.2. Machine Learning Model Development

Feature Engineering

A particular focus of this study is to compare feature engineering methods for shallow ML in both the time and frequency domain. The following sections detail the methods used and justify the selection of each.

The waveform Sum Absolute Amplitude (SAA) is the summed amplitude magnitude of each sample point in a waveform. It is a measure of the sound wave proportion reflecting back from the vessel–mixture interface and is therefore dependent on changes in mixture acoustic impedance.

$$SAA = \sum_{i=1}^{i=SP} |A_i| \tag{3}$$

where SAA is the sum absolute amplitude, SP is the total number of sample points, and A_i is the amplitude at sample point i [34].

The waveform energy is the summed squared amplitudes of every sample point in a waveform. It is therefore a similar measure to the waveform *SAA*; however, it gives greater weight to larger amplitudes. Therefore, two waveforms of the same waveform *SAA* but having different shapes can produce different waveform energies.

$$E = \sum_{i=1}^{i=SP} {A_i}^2 \tag{4}$$

where E is the waveform energy [34].

Principal component analysis (PCA) was also used throughout the investigation as a feature dimension reduction method. The waveform amplitude at each sample point in the waveform was used as the input variables. Therefore, the obtained principal components (PC) were constructed from parts of the ultrasonic waveform that change at the same time, providing more information than only using the *SAA* or waveform energy. PCA is an unsupervised ML algorithm, meaning only the relationship of input variables to one another are investigated. PCA extracts a new set of orthogonal variables, or PCs, which are a combination of co-linear input variables [56]. When combining the data from both sensors, the sample points in both waveforms were used as input variables. In all cases, the PCs explaining >95% of the variance in the input data set were used as features. If this number happened to be greater than 10 PCs, then only the first 10 were taken as features to prevent overfitting.

The gradients of each feature (e.g., the waveform energy, *SAA*, or PC magnitudes) were investigated for use as additional features that represent previous process time-steps. The difference between consecutive parameter values were calculated after applying a backwards, one-sided moving mean of varying size. A backwards, one-sized gradient means that only past process data is used. The sizes of the moving mean chosen for each ML prediction task is presented in the relevant Results sections.

$$MMV_i = \frac{1}{N} \sum_{i}^{i-N} V_i \tag{5}$$

$$G = MMV_i - MMV_{i-1}, \tag{6}$$

where G is the gradient of a parameter, MMV is the moving mean value of a parameter, N is the size of backwards, one-sided moving mean, and V is the original parameter value [57,58].

While the Fourier transform uses non-decaying sine and cosine waves as transform functions, the wavelet transform uses decaying wavelets (small waves) to analyse the frequency content of a waveform at each location in the time domain [59]. Opposed to the Continuous Wavelet Transform (CWT) analysis, which uses continuous wavelet frequencies as transform functions and therefore produces much redundant information, the Discrete Wavelet Transform (DWT) performs successive decomposition of a waveform by halving the frequency of the orthogonal analytical wavelet, thereby retaining no redundant information after the transform [60]. A key parameter is the choice of the analytical wavelet shape, termed the mother wavelet [61]. The Symlet wavelet was selected owing to it being the least asymmetric, which is most visually similar to the expected waveform composition [62]. The number of vanishing moments and decompositions were investigated and the optimal results for each ML task are reported in the relevant Results sections. After performing the DWT, features for ML model development were engineered in similar ways as described previously. The waveform energy and *SAA* of each decomposition were used as features, or PCA was applied using the waveform amplitude at every sample point in all decompositions as input variables.

Training, Validation, and Test Data Set Splits

Throughout this investigation a k-fold testing procedure (where k is the number of runs undertaken for each mixing system; 12 in the case of honey-water blending and 9 for flour-water batter mixing) was carried out by holding one run back for the testing data set. The run held back was alternated between all runs and the models were retrained. The average test set performance was taken as the

result for that feature and algorithm combination. The training and validation splits are discussed in each individual algorithm section. The number of data sets obtained from each sensor for each ML method is provided in Tables 2 and 3.

Table 2. The number of data sets recorded for the honey-water blending process.

Run	The Number of Data Sets from Each Sensor		
	Shallow Learning	**Time Domain Input CNNs**	**CWT Input CNNs**
1	146	243	292
2	131	213	262
3	108	167	216
4	166	283	332
5	139	227	276
6	141	233	282
7	109	169	218
8	122	195	244
9	108	165	214
10	102	155	204
11	114	177	226
12	115	179	228
Total	1501	2406	2994

Table 3. The number of data sets recorded for the flour-water batter mixing process.

Run	The Number of Data Sets from Each Sensor		
	Shallow Learning	**Time Domain Input CNNs**	**CWT Input CNNs**
1	93	137	186
2	90	131	180
3	107	165	214
4	105	159	208
5	102	153	202
6	102	155	204
7	123	197	246
8	129	207	256
9	154	259	308
Total	1005	1563	2004

Artificial Neural Networks

Neural networks were investigated for their ability to create new features from input variables, which have a linear relationship with the outputs [36,63]. The constructed ANNs consisted of three layers—an input, hidden, and output layer. The trainlm training function was used for regression networks and the trainscg training function for classification [64]. The training was stopped once the validation loss had increased for 6 consecutive iterations to prevent overfitting. For each neural network, 10 networks were trained and the average performance value was used to account for the effects of random weight initialisation. To further prevent overfitting, the training and validation data set was further broken down into 70% training, 15% validation, and 15% test for an initial hyperparameter optimisation search. A grid search determined the optimal number of neurons in the hidden layer (varied between 1 and 10 in intervals of 1) and regularisation weight (varied between 0.1 and 0.5 in intervals of 0.1) by monitoring this new test set error. The optimal number of hidden neurons and regularisation weight were used for training the final networks. These used 80% of the original training and validation set as the training set and 20% of the original training and validation set as the validation set. The previously discussed k-fold testing procedure was then used to evaluate the final networks.

Support Vector Machines

SVMs were chosen for investigation owing to their ability to handle high dimensional feature spaces and thus use kernels functions for non-linear input–output fitting [65]. Bayesian optimisation for 60 evaluations was used to select the box constraint value, kernel scale, kernel function, polynomial order, and whether the inputs were standardised. The acquisition function was chosen to be the expected improvement [66]. The training and validation data set underwent 5-fold validation and was repartitioned after every evaluation to improve generalization performance.

Convolutional Neural Networks

The CNNs consisted of 2 convolutional layers [26,39]. The first was either a 2D or 3D convolutional layer containing 8 5 × 5 pixel filters for each sensor input image depending on whether one or two sensor signals were being used as inputs. The second convolutional layer was a 2D convolutional layer containing 16 5 × 5 pixel filters. Padding was applied to keep the input matrices the same size. Batch normalisation was applied after each convolutional layer to aid training, increase the learning rate, and to provide some regularisation. By normalising each mini batch for every layer in the network, each layer does not need to continuously adapt to changing input distributions [67]. Batch normalisation was followed by the ReLu non-linearity function and 2 × 2 pixel max pooling. The training function used was the "adam" function, the initial learning rate was selected as 0.01 with a drop factor of 0.33 after 4 epochs. Training was carried out for 8 epochs, with a mini batch size of 256. The training data was shuffled after every epoch to improve network generalisation. No validation data set was used to maximise the number of datasets that the network fits to. Therefore, a dropout layer was added before the fully connected layer to further prevent overfitting, with the dropout factor varied between 0, 0.1, 0.3, and 0.5. A dropout layer randomly forgets network nodes during training based on the probability specified by the dropout factor. The effect of this is to ensure all nodes contribute to the prediction and improve model generalisation.

Time Domain Input CNNs

The waveform amplitudes at every sample point were used for time domain inputs into CNNs. Input matrices for the CNN were created by stacking 25 windowed signal amplitudes (approximately 10 s of acquired ultrasonic signals) at each time domain sample point on one another, with the current time ground truth being equal to that of the last (bottom) signal (Figure 3).

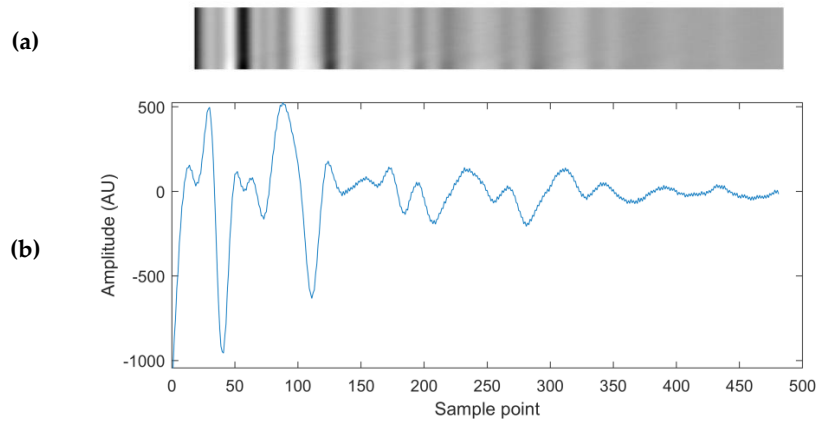


Figure 3. (**a**) A grey-scale representation of the input matrix to the time domain input Convolutional Neural Networks (CNNs), constructed of 25 ultrasound waveforms similar to the waveform depicted below it in Figure 3b. The bright pixels correspond to the maximum amplitude values of the 25 waveforms, and the dark pixels correspond to the minimum amplitude values. (**b**) An example of an ultrasonic waveform used in each row of the time domain input matrix.

Frequency-Time Domain Input CNNs

The absolute values of the frequency-time domain map after the Continuous Wavelet Transform (CWT) were used as an alternative input to the CNNs. An example of a single waveform after undergoing the CWT is presented in Figure 4. The Morlet wavelet was selected as the mother wavelet owing to the expected symmetry of returning sound waves.

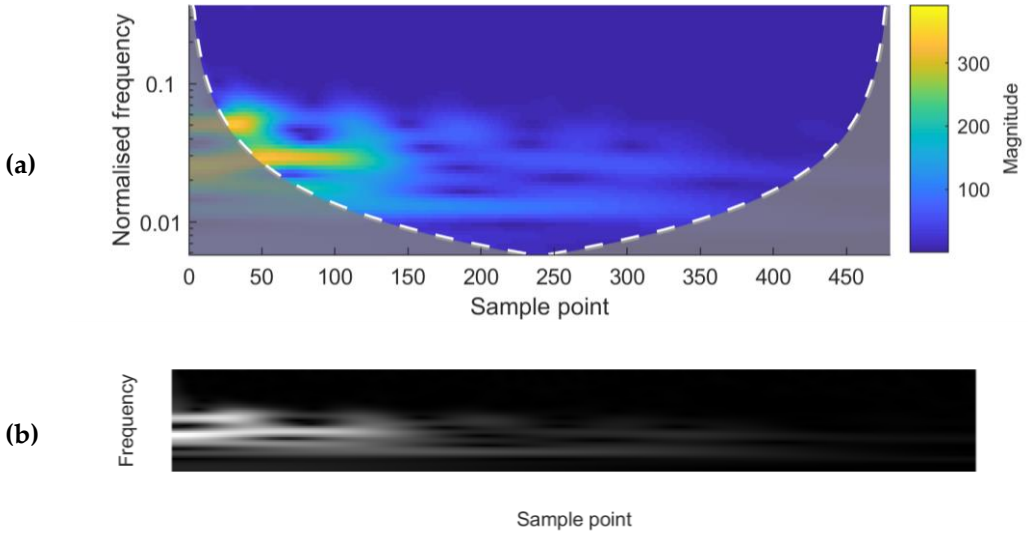


Figure 4. (**a**) A frequency-time domain magnitude scalogram after the Continuous Wavelet Transform (CWT) of a single waveform. The sample points correspond to the sample points of the original waveform. (**b**) A grey-scale image representation of the matrix used as an input to the CNN. It contains the absolute values after the CWT of the same waveform as used for Figure 4a. The bright pixels correspond to the maximum frequency amplitudes, and the dark pixels to the minimum.

Long Short-Term Memory Neural Networks

As this research was interested in monitoring time-evolving processes, LSTMs were an obvious choice for investigation due to their ability to store representations of all previous time-steps in a sequence. LSTM networks are a development of Recurrent Neural Networks (RNN) to overcome problems of exploding and vanishing error gradients. While RNNs use feedback connections to use the output from the previous time-step as an input for the current time-step, LSTMs have gate units to update the internal network state [68]. No validation set was specified to maximise the training data set size for the LSTM networks. The inputs were standardised to give a mean of zero and a standard deviation of 1. The mini-batch size was selected to be 2 runs and the sequence length of each run was sorted to minimise the amount of padding applied. The training was carried out for 60 epochs using the "adam" optimisation algorithm, an initial learning rate of 0.01, and a gradient threshold of 1 to prevent problems of exploding gradients. In the LSTM layer, 50 hidden units were used and 50 neurons in the fully connected layer. These relatively few hidden units and neurons were selected along with a 0.5 probability dropout layer to prevent overfitting and improve algorithm generalisation performance.

3. Results

3.1. Honey-Water Blending

Figure 5 displays the waveform energy profiles of both sensors during Run 1. The waveform energy was lowest at the beginning of the process as honey has a closer acoustic impedance to glass in comparison to water (Table 1, Equation (2)). Therefore, a greater proportion of the sound wave energy transferred into the honey. The waveform energy of the non-central sensor then increases as honey is removed and the water fraction increases at the sensor measurement area. This is because the action of

the impeller creates more mixing at the sides of the vessels, rather than directly below the impeller in the centre of the vessel. Finally a plateau is reached when honey has been completely removed from the non-central sensor measurement area. The waveform energy of the central sensor then increases to a plateau as the honey concentration at the measurement area decreases. The central sensor has less fluctuations in waveform energy as in the centre of the vessel base the glass was more flat, providing a vessel–mixture interface perpendicular to the direction of sound wave travel and therefore less variability in the proportion of the sound wave returning to the transducer. The waveform energy profiles for the central sensor during Run 1 and Run 12 are presented in Figure 6. Only these two runs are displayed to aid visibility of the parameter profiles. The differing waveform energy magnitudes are due to Run 1 being performed at an average temperature of 19.4 °C, Run 6 at 20.1 °C, and Run 12 at 20.7 °C. This causes different changes to the speed of sound in each material that the sound wave travels through, thus producing large changes in waveform energy in the reflected sound wave of interest. This highlights the need of using ML techniques to monitor the mixing process, as it can be seen that only using the waveform energy to monitor the mixing process would be insufficient. It can also be seen that the waveform energy during Run 12 continues to decline after peaking, indicating that monitoring the gradient of the waveform energy would also be insufficient for detecting the end of mixing.

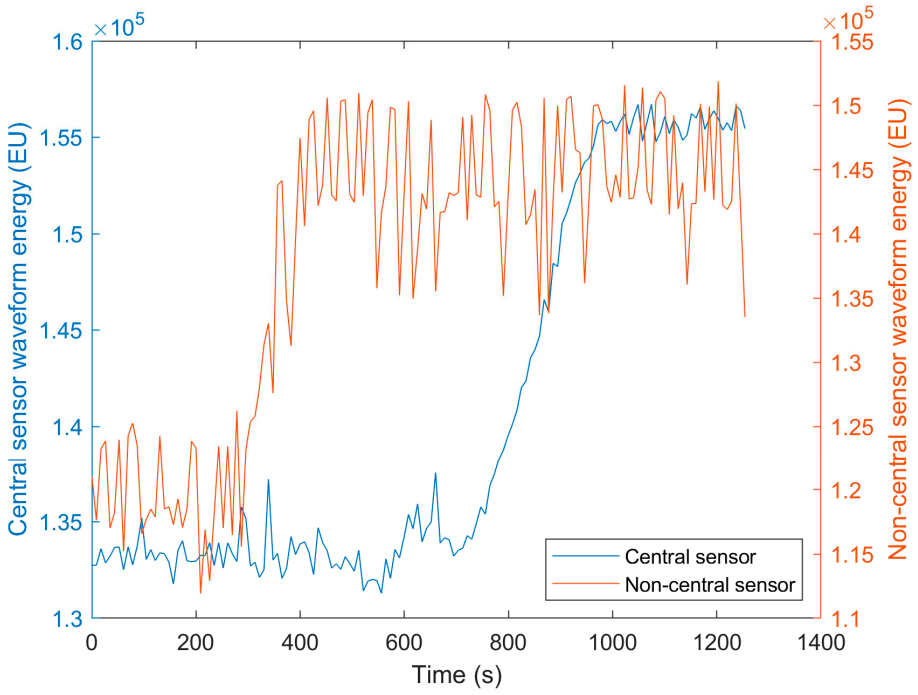


Figure 5. The waveform energy profile during the honey-water blending process for the central and non-central sensors.

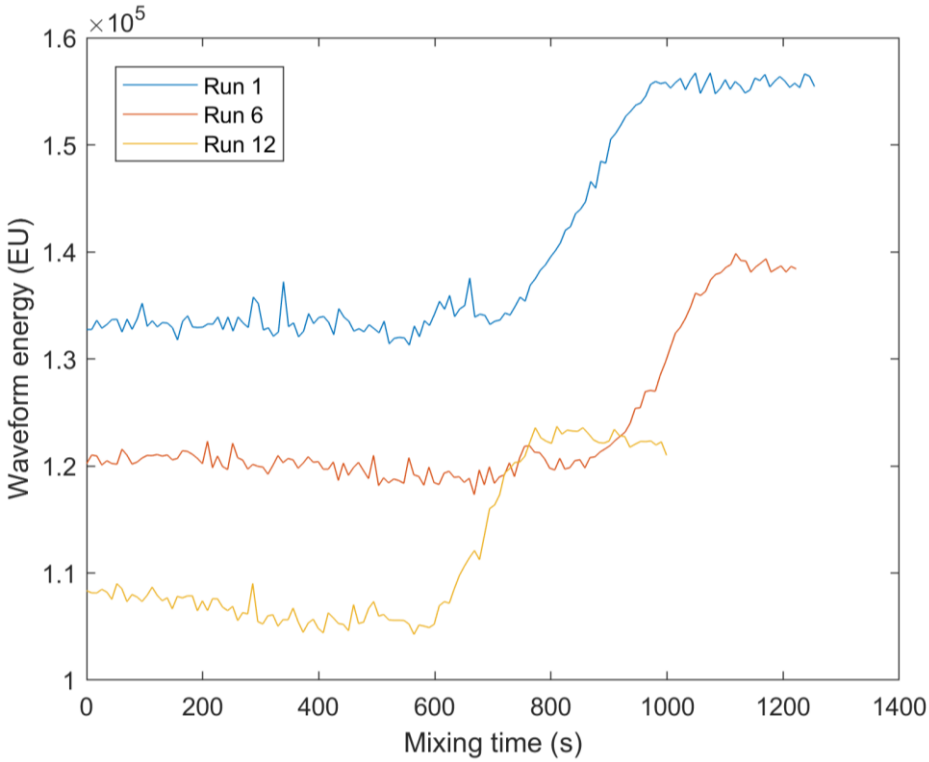


Figure 6. Waveform energy profiles for Run 1 (20 mL honey, 200 rpm impeller speed, 19.4 °C average temperature), Run 6 (30 mL honey, 200 rpm impeller speed, 20.1 °C average temperature), and Run 12 (30 mL honey, 250 rpm impeller speed, 20.7 °C average temperature). Only three runs are presented to aid visibility of the parameter profiles. This figure displays the variation in waveform energy levels in the data, highlighting the need of Machine Learning (ML) methods to monitor the mixing process.

3.2. Flour-Water Batter Mixing

In the batter mixing experiments both sensors were located at the same height on the mixing bowl but at different radial positions. The mixing dynamics were similar for all radial positions owing to the mixing bowl being circular and the impeller located in the centre. For this reason, both sensors displayed similar waveform energy profiles for all mixing runs as similar mixtures were present at each sensor measurement area throughout the mixing processes. Therefore, only information from Sensor 1 is presented in Figure 7. Only the waveform energy profiles and impeller power measurements Run 1 and Run 2 are presented in Figure 7 to aid visibility of the parameter profiles. Run 1 consisted of 500 g flour and 400 g water and Run 2 500 g flour and 450 g water. The increased water content of Run 2 delayed the gluten development process and therefore the time of maximum impeller power draw [21]. At the beginning of the process, water or flour may be present at the sensor's measurement area. If water was first present at the measurement area, the waveform energy initially increases as seen in Run 2. This is because water has a closer acoustic impedance to glass than a poorly mixed flour-water mixture would. The poorly mixed flour-water mixture has a high void fraction, which contains air and therefore produces a low average acoustic impedance over the sensor measurement area. In Run 1, flour was first present at the measurement area so the waveform energy initially decreases. This is because water is mixing into the flour, replacing the air between the flour particles. Ross et al. [44] found a peak in sound wave velocity in bubble free dough at optimal mixing time due to the aligned and fully hydrated glutenin polymers, whereas Létang et al. [45] found the speed of sound in high water content dough (>56% total weight) increases to a plateau at optimal mixing time. Ross et al. [44] found that despite the increasing alignment of the glutenin polymers, there were no significant density changes in bubble free dough during mixing. However, dough is a cellular structure of air incorporated into a viscoelastic matrix during mixing [45,69,70]. Therefore, dough density progressively decreases

with mixing time up until the optimum mixing time. Past optimal mixing it begins to increase again as the gluten matrix is broken down by shearing [71]. The decrease in waveform energy (increasing mixture acoustic impedance) up to optimal mixing time marked by the peak in the impeller power draw suggests that the increase in the speed of sound through the dough had a larger effect than this decreasing density (Equation (2)). It should be noted that although increased air entrainment reduces the speed of sound through the dough for lower frequencies [70], at the higher frequencies used here (5 MHz) the ultrasonic velocity approaches the velocity in bubble-free dough, i.e., it travels only through the viscoeleastic matrix [46]. Past the point of optimal mixing, the viscoelastic matrix begins to break down and the water binding capacity of the gluten declines, increasing the level of free water [72]. Although Ross et al. [44] show that this results in a decreasing sound velocity through a regular water content dough matrix, Létang et al. [45] found no change in speed of sound in highly hydrated doughs, suggesting that this increase in free water would not affect this overly hydrated batter. However, the results show a further increase in acoustic impedance, suggesting an increase in batter density past optimal mixing. This is due to the shearing action of the mixer breaking down the polymer matrix and therefore destroying the cellular structure of air pockets, thus increasing the batter density. Again, it can be seen that monitoring the magnitude of the waveform energy would not be sufficient to detect the optimal mixing time, further highlighting the need for ML techniques.

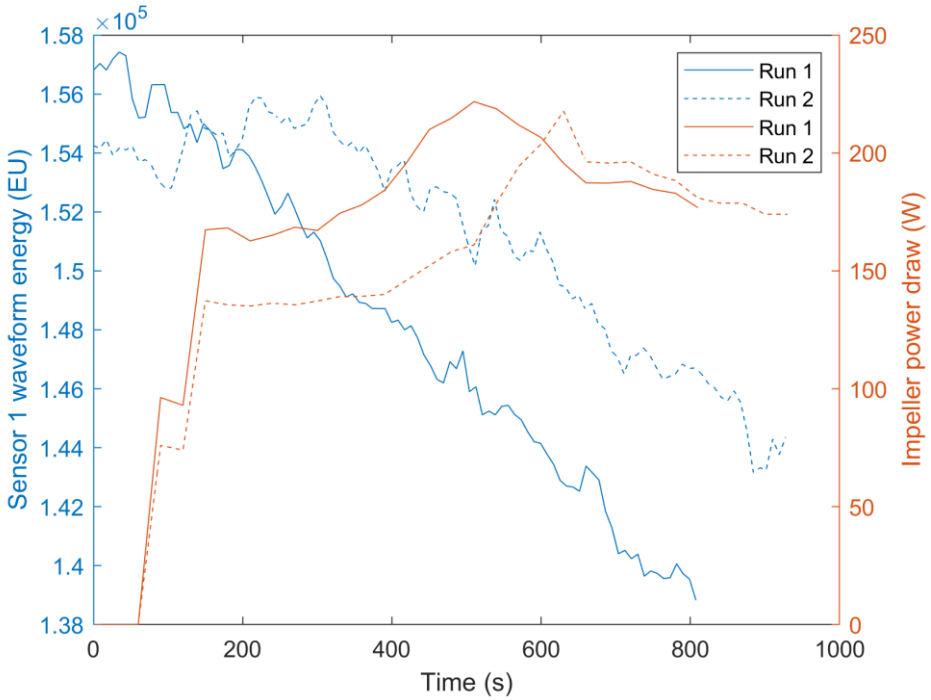


Figure 7. The waveform energy profiles of Sensor 1 and impeller power draw during two flour-water mixing processes containing different volumes of water. Run 1 consisted of 500 g flour and 400 g water and Run 2 500 g flour and 450 g water. Only the waveform energy profiles of Run 1 and Run 2 are provided to aid visibility of the parameter profiles.

3.3. Machine Learning Technique Comparison

3.3.1. Honey-Water Blending Classification

Initial investigations monitored the prediction accuracy of ANNs (as these required the least time to train) to determine the optimal moving mean gradient length, number of vanishing moments of the DWT mother wavelet, and number of DWT decompositions. The optimal size of the moving mean gradient was found to be 25 previous waveforms for the non-central sensor (approximately 25s) and 10 for the central sensor (approximately 10s). The optimal DWT mother wavelet was determined to be

Symlets 6 wavelet. This finding was carried forward for all remaining ML tasks. The optimal number of DWT decompositions was found to be 3 for both sensors.

The highest classification accuracy to predict whether the system was mixed was 96.3% and attained using the central sensor and an LSTM with the waveform energy, *SAA*, and gradients as features (Table 4). This superior classification accuracy demonstrates the efficacy of combining ultrasonic sensors with ML techniques. Performing data fusion between both sensors did not provide any benefit over results from using only the central sensor, sometimes even producing lower classification accuracy due to overfitting. This is because the time for mixing completion was defined as the moment the last remaining honey was dissolved. Owing to the motion of the impeller, the location of this event was the centre of the vessel base, where the central sensor was located. Therefore, the highest classification accuracy using the non-central sensor alone was only 89.8%. Although performing the DWT and PCA aided the performance of the ANNs and SVMs, this led to the LSTM neural networks beginning to overfit. Due to their ability to store representations of all previous time-steps, the LSTM neural networks were able to produce the highest classification accuracy using only the waveform energy, *SAA*, and their gradients. ANNs produced their highest classification accuracy when using feature gradients as additional features. However, despite using approximately 10 s of previous time-step waveforms, the time domain input CNNs produced lower classification accuracies than both LSTMs and ANNs (Table 5). This suggests that by using the amplitude at every sample point in the waveform, the time domain input CNNs began to overfit.

Table 4. Classification accuracies of shallow machine learning algorithms to predict whether the honey-water blending mixture was mixed or non-mixed. E—Energy, SAA—Sum Absolute Amplitude, G—Gradients of Features, PCs—Principle Components, DWT—Discrete Wavelet Transform.

	ANN (% Correct)			**SVM (% Correct)**			**LSTM (% Correct)**		
Features	Non-central	Central	Combined	Non-central	Central	Combined	Non-central	Central	Combined
E, SAA	65.5	85.0	83.4	80.2	80.1	86.0	89.5	93.0	88.0
E, SAA, G	76.1	91.1	92.5	77.2	91.1	89.1	89.8	96.3	95.4
PCs	76.2	90.7	83.1	75.1	81.8	80.4	82.9	86.1	91.5
PCs, G	79.3	93.0	92.7	71.8	83.7	86.7	86.2	89.4	93.7
DWT, E, SAA	79.2	91.7	90.6	72.1	82.3	86.4	77.8	95.1	92.9
DWT, E, SAA, G	80.9	92.4	94.6	82.9	92.1	91.9	80.8	94.5	90.8
DWT, PCs	71.9	88.7	90.0	76.2	82.7	80.1	82.5	84.5	89.5
DWT, PCs, G	80.5	95.0	93.9	75.0	85.3	91.0	79.7	86.1	90.4

Table 5. Classification accuracies of CNN algorithms to predict whether the honey-water blending mixture was mixed or non-mixed.

CNNs	**Time Domain Input (% Correct)**			**CWT Input (% Correct)**		
Dropout layer probability	Non-central	Central	Combined	Non-central	Central	Combined
0	71.7	92.2	93.0	73.5	88.6	88.2
0.1	75.3	92.6	93.1	75.4	89.1	87.0
0.3	72.9	92.7	91.3	73.1	88.2	85.4
0.5	76.3	93.1	93.0	74	90.0	85.4

3.3.2. Honey-Water Blending Regression

The optimal moving mean gradient length was found to be 25 previous waveforms for both sensors, and the optimal number of DWT decompositions was found to be three for both sensors.

The highest R^2 values were achieved by combining the inputs from both sensors. R^2 values of 0.972, 0.973, and 0.977 could be reached using ANNs, LSTMs, or CNNs, respectively (Tables 6 and 7). This is because the non-central sensor has better prediction ability nearer the beginning of the process as the honey is removed from this sensor measurement area first, and the central sensor has greater resolution nearer the end of the mixing process as the last remnants of honey are dissolved. This is presented in Figure 8, where the combined prediction is more accurate than either single sensor from approximately 600 s before mixing completion until 200 s afterwards. However, reasonable accuracy can be attained using only a single sensor should only one sensor position be available. ANNs using the DWT decompositions energy and *SAA* was able to generalise well and produced an R^2 value of 0.960 for the central sensor. Similarly, LSTMs using the same features with the non-central sensor achieved an R^2 value of 0.963, increasing to 0.965 when incorporating feature gradients. Again, these high prediction accuracies illustrate the effectiveness of combining ultrasonic sensors with ML to monitor mixing. As observed when classifying the mixture state, the ability to use data from previous time-steps was vital for accurate prediction of the mixing time remaining. LSTMs, which store representations of all previous time-steps; ANNs using feature gradients; and time domain input CNNs, which use the previous 10 s of acquired waveforms, all produced the greatest R^2 values with the true mixing time remaining. Time domain input CNNs, which use the amplitude at every sample point in the waveform, displayed the greatest prediction performance. This indicates that, unlike for classifying the mixture state, some useful waveform information is not represented by the time domain and DWT decomposition energies, SAAs, and PCs. However, the CWT input CNNs performed poorer, because only a single waveform was used for prediction. SVMs performed worst overall, most likely because of overfitting due to their convex optimisation leading to a global minima, as opposed to ANNs, which only converge to local minima. Although global cost minimisation is desirable to fit training and validation data, it may lead to poor prediction ability when the test data process parameters lies outside of the bounds of the training data. As in this investigation each monitored run is individually held back for testing, testing on data lying outside of the process parameter space used in training is unavoidable.

Table 6. Regression accuracies of shallow machine learning algorithms to predict the mixing time remaining of the honey-water blending process. E—Energy, SAA—Sum Absolute Amplitude, G—Gradients of Features, PCs—Principle Components, DWT—Discrete Wavelet Transform.

	ANN (R^2)			**SVM (R^2)**			**LSTM (R^2)**		
Features	Non-central	Central	Combined	Non-central	Central	Combined	Non-central	Central	Combined
E, SAA	0.751	0.788	0.920	0.276	0.223	0.795	0.852	0.818	0.954
E, SAA, G	0.875	0.806	0.922	0.894	0.663	0.780	0.755	0.853	0.969
PCs	0.810	0.882	0.956	0.622	0.503	0.817	0.705	0.584	0.959
PCs, G	0.910	0.904	0.972	0.771	0.190	0.721	0.939	0.936	0.969
DWT, E, SAA	0.862	0.960	0.949	0.713	0.570	0.907	0.963	0.895	0.957
DWT, E, SAA, G	0.758	0.892	0.948	0.827	0.722	0.899	0.965	0.865	0.959
DWT, PCs	0.786	0.881	0.957	0.552	0.551	0.806	0.857	0.751	0.973
DWT, PCs, G	0.916	0.914	0.972	0.787	0.503	0.854	0.914	0.930	0.972

Table 7. Regression accuracies of CNN algorithms to predict the mixing time remaining of the honey-water blending process.

CNNs	**Time Domain Input (R^2)**			**CWT Input (R^2)**		
Dropout layer probability	Non-central	Central	Combined	Non-central	Central	Combined
0	0.825	0.936	0.975	0.786	0.866	0.950
0.1	0.828	0.943	0.974	0.790	0.869	0.955
0.3	0.827	0.933	0.974	0.793	0.869	0.951
0.5	0.828	0.932	0.977	0.789	0.866	0.949

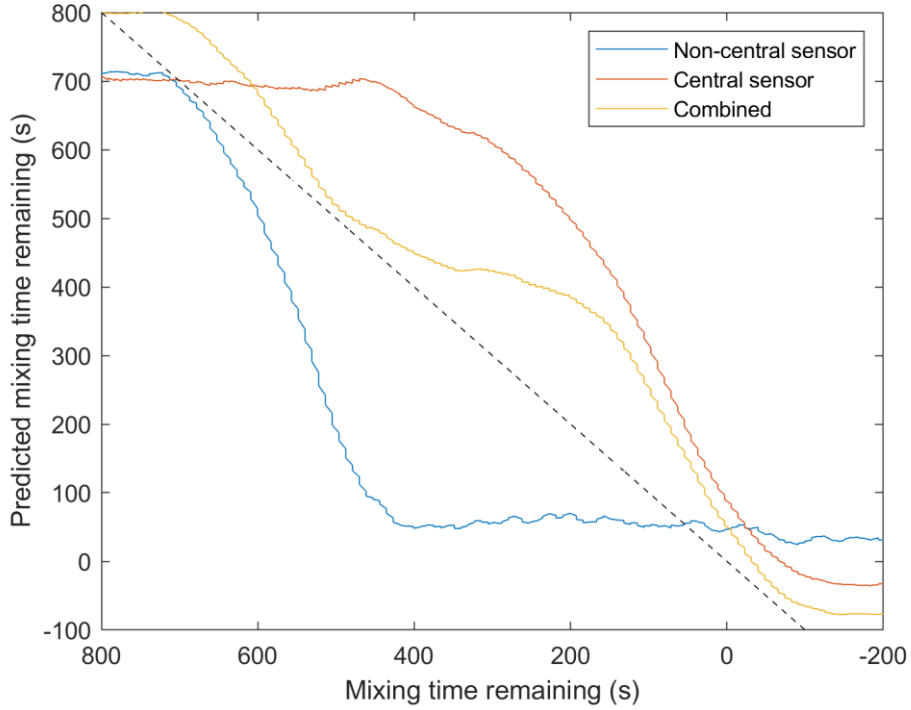


Figure 8. A comparison between regression predictions during honey-water blending for the central sensor, non-central sensor, and combining the outputs from both sensors. A time domain input CNN was used as the learning algorithm. R^2 values: 0.825 non-central sensor, 0.936 central sensor, and 0.975 multisensor data fusion.

3.3.3. Flour-Water Batter Mixing Classification

The optimal size of the moving mean gradient was found to be 25 previous waveforms for both sensors. The optimal number of DWT decompositions was 7 for Sensor 1 and 5 for Sensor 2.

Both sensors performed equally well and combining sensor inputs did not provide any advantage (Tables 8 and 9). This shows that the signal received from each sensor was adequate for determining when the system was fully mixed. ANNs and CWT-input CNNs produced the greatest classification accuracies of up to 91.3% and 92.5%, respectively. However, decomposition of the original time domain waveform using wavelet analysis was required to obtain optimal classification. This suggests there was a change in the waveforms only noticeable through wavelet analysis that marked the transition between non-mixed and mixed. Furthermore, the incorporation of previous time-steps in the prediction was not necessary for classification accuracy and may have led to the LSTMs and time domain input CNNs performing poorly due to overfitting. The only highly performing algorithm using past time-steps was ANNs with DWT decomposition energy, SAAs, and gradients as features. The lower classification accuracies for predicting the state of the batter mixture compared with the honey-water blending may be caused by limitations in the frequency analysis employed during this investigation. Further decompositions using the DWT may yield more information in the frequency content of the ultrasonic signals, and an incorporation of an LSTM layer into the CWT input CNNs would provide them with the ability to incorporate previous process states into their prediction.

Table 8. Classification accuracies of shallow machine learning algorithms to predict whether the flour-water batter mixture was fully mixed or non-mixed. E—Energy, SAA—Sum Absolute Amplitude, G—Gradients of Features, PCs—Principle Components, DWT—Discrete Wavelet Transform.

	ANN (% Correct)			**SVM (% Correct)**			**LSTM (% Correct)**		
Features	Sensor 1	Sensor 2	Combined	Sensor 1	Sensor 2	Combined	Sensor 1	Sensor 2	Combined
E, SAA	78.5	77.3	81.7	69.9	66.2	78.2	76.9	77.9	80.9
E, SAA, G	79.9	80.6	82.8	79.2	84.2	81.4	85.2	89.6	85.5
PCs	86.9	90.4	89.7	75.7	86.7	83.1	76.1	85.3	86.3
PCs, G	85.5	90.3	86.5	74.0	87.3	76.9	74.6	83.0	84.2
DWT, E, SAA	90.7	90.8	90.8	84.3	88.5	87.7	87.0	85.2	89.3
DWT, E, SAA, G	91.1	90.4	90.0	84.7	89.5	82.0	83.7	88.1	88.3
DWT, PCs	91.3	89.6	91.1	80.4	87.9	84.2	77.7	80.5	88.1
DWT, PCs, G	85.0	88.2	88.4	70.4	72.0	80.9	75.8	84.2	83.4

Table 9. Classification accuracies of CNN algorithms to predict whether the flour-water batter mixture was fully mixed or non-mixed.

CNNs	**Time Domain Waveforms (% Correct)**			**CWT (% Correct)**		
Dropout layer probability	Sensor 1	Sensor 2	Combined	Sensor 1	Sensor 2	Combined
0	82.6	84.4	83.4	91.5	92.5	90.3
0.1	86.6	85.0	78.7	90.3	90.6	92.2
0.3	85.3	82.4	85.8	87.2	91.6	92.2
0.5	85.7	86.0	78.9	88.4	92.3	92.4

3.3.4. Flour-Water Batter Mixing Regression

The optimal size of the moving mean gradient and number of DWT decompositions was the same as for the classification of the batter mixture. Despite combined sensor outputs producing the highest prediction accuracies, using both sensors was not required to achieve adequate performance (Tables 10 and 11). LSTMs using time domain waveform energy, SAAs, and feature gradients achieved an R^2 value of 0.966, and LSTMs using PCA of the DWT decompositions achieved an R^2 value of 0.968, both using a single sensor. However, time domain input CNNs produced the highest R^2 values of up to 0.976 using a single sensor. These results show that performing wavelet analysis was not necessary for high prediction accuracy, unlike for batter mixing classification. Instead, the ability to use previous time-step data and the amplitude of every sample point in the waveform as features is required. However, ANNs preformed worse than both LSTMs and time domain input CNNs, suggesting that a more flexible incorporation of previous time-step data was required for regression accuracy at different process stages rather than using fixed feature gradient lengths throughout the process. Again, the single waveform CWT input CNNs did not produce high prediction performance due to not incorporating previous time-step data as features.

Table 10. Regression accuracies of shallow machine learning algorithms to predict the mixing time remaining for the flour-water batter mixture. E—Energy, SAA—Sum Absolute Amplitude, G—Gradients of Features, PCs—Principle Components, DWT—Discrete Wavelet Transform.

	ANN (R^2)			**SVM (R^2)**			**LSTM (R^2)**		
Features	Sensor 1	Sensor 2	Combined	Sensor 1	Sensor 2	Combined	Sensor 1	Sensor 2	Combined
E, SAA	0.633	0.697	0.868	0.208	0.430	0.658	0.937	0.935	0.936
E, SAA, G	0.576	0.819	0.835	0.712	0.817	0.873	0.912	0.966	0.974
PCs	0.846	0.879	0.855	0.464	0.775	0.810	0.732	0.822	0.848
PCs, G	0.947	0.91	0.946	0.485	0.713	0.629	0.641	0.760	0.754
DWT, E, SAA	0.831	0.872	0.876	0.392	0.781	0.535	0.940	0.950	0.912
DWT, E, SAA, G	0.666	0.766	0.743	0.815	0.917	0.631	0.771	0.953	0.955
DWT, PCs	0.844	0.898	0.824	0.489	0.665	0.485	0.932	0.968	0.958
DWT, PCs, G	0.840	0.930	0.906	0.662	0.774	0.654	0.786	0.911	0.939

Table 11. Regression accuracies of CNN algorithms to predict the mixing time remaining for the flour-water batter mixture.

CNNs	Time Domain Waveforms (R^2)			CWT (R^2)		
Dropout layer probability	Sensor 1	Sensor 2	Combined	Sensor 1	Sensor 2	Combined
0	0.961	0.976	0.975	0.945	0.920	0.958
0.1	0.962	0.973	0.980	0.944	0.922	0.962
0.3	0.959	0.970	0.982	0.938	0.922	0.958
0.5	0.960	0.975	0.977	0.940	0.915	0.961

4. Discussion

Although ML algorithms were able to achieve similar regression accuracy for both the honey-water blending and flour-water batter mixing, the classification accuracy was lower for the flour-water batter mixing. This is because despite the waveform energies of both processes changing by a similar proportion throughout the mixing processes (Figures 5 and 7), the honey-water blending waveform energy profile has a sharper change during the time of mixing completion. The waveform energy increased as the honey was removed from the measurement area of the central sensor, giving greater resolution of this sensor around the end of the mixing process.

Different ML approaches performed best on each prediction task. To classify the honey-water mixture state, predict honey-water mixing time remaining, and predict the flour-water batter mixing time remaining, the use of previous time-steps as features was useful for prediction accuracy. However, the ability of LSTMs to represent all previous time-states in the internal network, and time domain input CNNs ability to use the previous 10 s of acquired waveforms, performed better than the fixed feature gradient lengths used for the ANNs. In contrast, to classify the mixture state of flour-water batter, no previous time-steps were required. Instead, decomposition of the time domain waveform by the wavelet transform was needed to monitor a state change signature in the frequency domain. Time domain input CNNs were the best performing algorithm to predict the mixing time remaining of both the honey-water blending and flour-water batter mixing. This suggests that the ability to use the amplitude at every sample point in the waveform was better equipped to predict the mixing time remaining than using the waveform energy, *SAA*, or PCs. However, the time domain input CNNs began to overfit when classifying the state of the honey-water mixture. Therefore, the ANN and LSTM prediction accuracy may only sometimes be improved by using the amplitude of all sample points in a waveform. The use of only one acquired waveform for prediction hindered the CWT input CNNs ability to predict the mixing time remaining for both systems, and classified the state of the honey-water blending. Therefore, the addition of an LSTM layer would aid the prediction performance of the CNNs by storing representations of previous time-step data. The only ML task that required combining two sensor outputs was predicting the mixing time remaining for the honey-water blending. This is because the different sensor positions gave increased resolution at different stages of the mixing process. SVMs performed the worst for all prediction tasks. This is likely due to overfitting causing low prediction accuracy on test data outside the parameter bounds of the training and validation data. This is because SVM have convex optimisation functions that produce a global minima. In comparison, ANNs only converge to local minima, which may have aided their ability to generalise to test data outside the parameter space of training.

The application of the combined sensor and ML techniques to monitor processes relies on attaining ground truth data to label the outputs of all sensor signals. In industrial settings, product quality evaluations are typically conducted off-line and require considerable time, expense, or manual operations. This can mean ground truth values to produce labelled data are difficult to obtain, and therefore only a small set of labelled data is available for ML model development. In this case, additional techniques must be considered. For example, semisupervised learning can be used to first perform unsupervised learning on the combined set of labelled and unlabelled data to extract

features. Supervised learning models using the labelled data can then be used to predict the class or value of the unlabelled data [73]. Subsequently, active learning can be employed to automatically select data, which would be most useful to the model development if labelled rather than employing annotation of random samples [74]. For example, data points close to classification boundaries or those, which expand the model training space. Transfer learning is another technique that can help overcome the limitation of small labelled data sets. It has found particular application for transferring pretrained CNNs for image recognition tasks or for NIR spectroscopy calibration transfer across spectrometers [75,76]. A model trained on another system, for example a laboratory or pilot scale model system, can be used to aid in the prediction of the state of the target system. For example, the optimised signal processing, network weights, or ML hyperparameter values from the first system can be used as initial training values for the target system. Alternatively, the outputs of the previously trained model applied to the target system may be used as inputs to a second model [75].

5. Conclusions

This work studied the potential of using an industrially applicable ultrasonic sensing technique combined with ML to monitor the mixing of two model systems. Two ultrasonic sensors were used for data acquisition, and different ML and feature engineering methods were compared. This work has shown the potential of using ultrasonic sensors and ML to predict the time remaining until mixing is complete and when a system is mixed. The superior prediction accuracies of up to 96.3% for honey-water blending and 92.5% for flour-water batter mixing, along with R^2 values of up to 0.977 for honey-water blending and 0.968 for flour-water batter mixing, highlight the efficacy of combining ultrasonic sensors and ML to monitor mixing processes.

Author Contributions: Conceptualization, A.L.B., and N.J.W.; Methodology, A.L.B., N.J.W Software, A.L.B.; Validation, A.L.B.; Formal Analysis, A.L.B.; Investigation, A.L.B.; Resources, A.L.B., and N.J.W; Data Curation, A.L.B.; Writing—Original Draft Preparation, A.L.B.; Writing—Review and Editing, A.L.B., N.J.W., and S.B.; Visualization, A.L.B.; Supervision, A.L.B., N.J.W., and S.B.; Project Administration, A.L.B., N.J.W.; Funding Acquisition, N.J.W. All authors have read and agreed to the published version of the manuscript.

Funding: This work was supported by the Engineering and Physical Sciences Research Council (EPSRC) standard research studentship (EP/R513283/1).

Conflicts of Interest: The authors declare no conflict of interest.

References

1. Zhong, R.Y.; Xu, X.; Klotz, E.; Newman, S.T. Intelligent Manufacturing in the Context of Industry 4.0: A Review. *Engineering* **2017**, *3*, 616–630. [CrossRef]
2. Diez-Olivan, A.; Del Ser, J.; Galar, D.; Sierra, B. Data fusion and machine learning for industrial prognosis: Trends and perspectives towards Industry 4.0. *Inf. Fusion* **2019**, *50*, 92–111. [CrossRef]
3. Kang, H.S.; Lee, J.Y.; Choi, S.; Kim, H.; Park, J.H.; Son, J.Y.; Kim, B.H.; Noh, S. Do Smart manufacturing: Past research, present findings, and future directions. *Int. J. Precis. Eng. Manuf. Green Technol.* **2016**, *3*, 111–128. [CrossRef]
4. Bowler, A.L.; Bakalis, S.; Watson, N.J. A review of in-line and on-line measurement techniques to monitor industrial mixing processes. *Chem. Eng. Res. Des.* **2019**, *153*, 463–495. [CrossRef]
5. De Beer, T.; Burggraeve, A.; Fonteyne, M.; Saerens, L.; Remon, J.P.; Vervaet, C. Near infrared and Raman spectroscopy for the in-process monitoring of pharmaceutical production processes. *Int. J. Pharm.* **2011**, *417*, 32–47. [CrossRef]
6. Awad, T.S.; Moharram, H.A.; Shaltout, O.E.; Asker, D.; Youssef, M.M. Applications of ultrasound in analysis, processing and quality control of food: A review. *Food Res. Int.* **2012**, *48*, 410–427. [CrossRef]
7. Henning, B.; Rautenberg, J. Process monitoring using ultrasonic sensor systems. *Ultrasonics* **2006**, *44*, 1395–1399. [CrossRef]
8. Escrig, J.; Woolley, E.; Rangappa, S.; Simeone, A.; Watson, N.J. Clean-in-place monitoring of different food fouling materials using ultrasonic measurements. *Food Control* **2019**, *104*, 358–366. [CrossRef]

9. Al-Aufi, Y.A.; Hewakandamby, B.N.; Dimitrakis, G.; Holmes, M.; Hasan, A.; Watson, N.J. Thin film thickness measurements in two phase annular flows using ultrasonic pulse echo techniques. *Flow Meas. Instrum.* **2019**, *66*, 67–78. [CrossRef]
10. Stolojanu, V.; Prakash, A. Characterization of slurry systems by ultrasonic techniques. *Chem. Eng. J.* **2001**, *84*, 215–222. [CrossRef]
11. Ribeiro, M.M.M.; Gonçalves, C.; Regueiras, P.F.; Guimarães, M.M.L.; Cruz Pinto, J.J.C. Measurements of toluene–water dispersions hold-up using a non-invasive ultrasonic technique. *Chem. Eng. J.* **2006**, *118*, 47–54. [CrossRef]
12. Yucel, U.; Coupland, J.N. Ultrasonic attenuation measurements of the mixing, agglomeration, and sedimentation of sucrose crystals suspended in oil. *JAOCS, J. Am. Oil Chem. Soc.* **2011**, *88*, 33–38. [CrossRef]
13. Bamberger, J.A.; Greenwood, M.S. Using ultrasonic attenuation to monitor slurry mixing in real time. *Ultrasonics* **2004**, *42*, 145–148. [CrossRef] [PubMed]
14. Sun, Z.; Jen, C.K.; Yan, J.; Chen, M.Y. Application of ultrasound and neural networks in the determination of filler dispersion during polymer extrusion processes. *Polym. Eng. Sci.* **2005**, *45*, 764–772. [CrossRef]
15. Fox, P.; Smith, P.P.; Sahi, S. Ultrasound measurements to monitor the specific gravity of food batters. *J. Food Eng.* **2004**, *65*, 317–324. [CrossRef]
16. Salazar, J.; Turó, A.; Chávez, J.A.; García, M.J. Ultrasonic inspection of batters for on-line process monitoring. *Ultrasonics* **2004**, *42*, 155–159. [CrossRef]
17. Hunter, T.N.; Darlison, L.; Peakall, J.; Biggs, S. Using a multi-frequency acoustic backscatter system as an in situ high concentration dispersion monitor. *Chem. Eng. Sci.* **2012**, *80*, 409–418. [CrossRef]
18. Bux, J.; Paul, N.; Hunter, T.N.; Peakall, J.; Dodds, J.M.; Biggs, S. In Situ Characterization of Mixing and Sedimentation Dynamics in an Impinging Jet Ballast Tank Via Acoustic Backscatter. *Am. Inst. Chem. Eng.* **2017**, *63*, 2618–2629. [CrossRef]
19. Buurman, C.; Resoort, G.; Plaschkes, A. SCALING-UP RULES FOR SOLIDS SUSPENSION STIRRED VESSELS. *Chem. Eng. Sci.* **1986**, *41*, 2865–2871. [CrossRef]
20. Zhan, X.; Yang, Y.; Liang, J.; Zou, D.; Zhang, J.; Feng, L.; Shi, T.; Li, X. In-line mixing states monitoring of suspensions using ultrasonic reflection technique. *Ultrasonics* **2016**, *65*, 43–50. [CrossRef]
21. Aït Kaddour, A.; Morel, M.H.; Cuq, B. Description of batter mixing using near infrared spectroscopy. *J. Cereal Sci.* **2008**, *48*, 698–708. [CrossRef]
22. Zhang, L.; Li, H.; Gao, B. Combination of wavelet packet analysis with BPNN flaw type identification in concrete ultrasonic testing. In Proceedings of the 2008 3rd International Conference on Innovative Computing Information and Control, Dalian, China, 18–20 June 2008; p. 515. [CrossRef]
23. Wallhäußer, E.; Hussein, W.B.; Hussein, M.A.; Hinrichs, J.; Becker, T. Detection of dairy fouling: Combining ultrasonic measurements and classification methods. *Eng. Life Sci.* **2013**, *13*, 292–301. [CrossRef]
24. Sambath, S.; Nagaraj, P.; Selvakumar, N. Automatic defect classification in ultrasonic NDT using artificial intelligence. *J. Nondestruct. Eval.* **2011**, *30*, 20–28. [CrossRef]
25. Wallhäußer, E.; Sayed, A.; Nöbel, S.; Hussein, M.A.; Hinrichs, J.; Becker, T. Determination of cleaning end of dairy protein fouling using an online system combining ultrasonic and classification methods. *Food Bioprocess Technol.* **2014**, *7*, 506–515. [CrossRef]
26. Munir, N.; Kim, H.-J.; Park, J.; Song, S.-J.; Kang, S.-S. Convolutional neural network for ultrasonic weldment flaw classification in noisy conditions. *Ultrasonics* **2019**, *94*, 74–81. [CrossRef]
27. Cau, F.; Fanni, A.; Montisci, A.; Testoni, P.; Usai, M. Artificial neural networks for non-destructive evaluation with ultrasonic waves in not accessible. In Proceedings of the Fourtieth IAS Annual Meeting. Conference Record of the 2005 Industry Applications Conference, Hong Kong, China, 2–6 October 2005; Volume 1, pp. 685–692.
28. Cruz, F.C.; Simas Filho, E.F.; Albuquerque, M.C.S.; Silva, I.C.; Farias, C.T.T.; Gouvêa, L.L. Efficient feature selection for neural network based detection of flaws in steel welded joints using ultrasound testing. *Ultrasonics* **2017**, *73*, 1–8. [CrossRef]
29. Munir, N.; Kim, H.J.; Song, S.J.; Kang, S.S. Investigation of deep neural network with drop out for ultrasonic flaw classification in weldments. *J. Mech. Sci. Technol.* **2018**, *32*, 3073–3080. [CrossRef]

30. Lee, L.H.; Rajkumar, R.; Lo, L.H.; Wan, C.H.; Isa, D. Oil and gas pipeline failure prediction system using long range ultrasonic transducers and Euclidean-Support Vector Machines classification approach. *Expert Syst. Appl.* **2013**, *40*, 1925–1934. [CrossRef]
31. Miao, C.; Wang, Y.; Zhang, Y.; Qn, J.; Zuo, M.J.; Wang, X. A SVM classifier combined with PCA for ultrasonic crack size classification. In Proceedings of the 2008 Canadian Conference on Electrical and Computer Engineering, Niagara Falls, ON, Canada, 4–7 May 2008; pp. 1627–1630. [CrossRef]
32. Song, S.J.; Kim, H.J.; Cho, H. Development of an intelligent system for ultrasonic flaw classification in weldments. *Nucl. Eng. Des.* **2002**, *212*, 307–320. [CrossRef]
33. Krause, D.; Hussein, W.B.; Hussein, M.A.; Becker, T. Ultrasonic sensor for predicting sugar concentration using multivariate calibration. *Ultrasonics* **2014**, *54*, 1703–1712. [CrossRef]
34. Zhan, X.; Jiang, S.; Yang, Y.; Liang, J.; Shi, T.; Li, X. Inline Measurement of Particle Concentrations in Multicomponent Suspensions using Ultrasonic Sensor and Least Squares Support Vector Machines. *Sensors* **2015**, *15*, 24109–24124. [CrossRef] [PubMed]
35. El-Hagrasy, A.S.; Delgado-Lopez, M.; Drennen, J.K. A process analytical technology approach to near-infrared process control of pharmaceutical powder blending: Part II: Qualitative near-infrared models for prediction of blend homogeneity. *J. Pharm. Sci.* **2006**, *95*, 407–421. [CrossRef] [PubMed]
36. Lecun, Y.; Bengio, Y.; Hinton, G. Deep learning. *Nature* **2015**, *521*, 436–444. [CrossRef] [PubMed]
37. Hu, F.; Zhou, M.; Yan, P.; Li, D.; Lai, W.; Bian, K.; Dai, R. Identification of mine water inrush using laser-induced fluorescence spectroscopy combined with one-dimensional convolutional neural network. *RSC Adv.* **2019**, *9*, 7673–7679. [CrossRef]
38. Wang, K.; Shang, C.; Liu, L.; Jiang, Y.; Huang, D.; Yang, F. Dynamic Soft Sensor Development Based on Convolutional Neural Networks. *Ind. Eng. Chem. Res.* **2019**, *58*, 11521–11531. [CrossRef]
39. Meng, M.; Chau, Y.J.; Wouterson, E.; Ong, C.P.K. Ultrasonic signal classification and imaging system for composite materials via deep convolutional neural networks. *Neurocomputing* **2017**, *257*, 128–135. [CrossRef]
40. McClements, D.J. Advances in the application of ultrasound in food analysis and processing. *Trends Food Sci. Technol.* **1995**, *6*. [CrossRef]
41. Speed of Sound in common Liquids. Available online: https://www.engineeringtoolbox.com/sound-speed-liquids-d_715.html (accessed on 10 December 2019).
42. González-Mohino, A.; Jiménez, A.; Paniagua, M.J.; Perez-Palacios, T.; Rufo, M. New contributions of ultrasound inspection to the characterization of different varieties of honey. *Ultrasonics* **2019**, *96*, 83–89 101016/jultras201902010. [CrossRef]
43. Bulk Density - Food Products. Available online: https://www.engineeringtoolbox.com/foods-materials-bulk-density-d_1819.html (accessed on 10 December 2019).
44. Ross, K.A.; Pyrak-Nolte, L.J.; Campanella, O.H. The use of ultrasound and shear oscillatory tests to characterize the effect of mixing time on the rheological properties of dough. *Food Res. Int.* **2004**, *37*, 567–577. [CrossRef]
45. Létang, C.; Piau, M.; Verdier, C.; Lefebvre, L. Characterization of wheat- flour-water doughs: A new method using ultrasound. *Ultrasonics* **2001**, *39*, 133–141. [CrossRef]
46. Koksel, F.; Scanlon, M.G.; Page, J.H. Ultrasound as a tool to study bubbles in dough and dough mechanical properties: A review. *FRIN* **2016**, *89*, 74–89. [CrossRef] [PubMed]
47. Speed of Sound in common Solids and Metals. Available online: https://www.engineeringtoolbox.com/sound-speed-solids-d_713.html (accessed on 10 December 2019).
48. Density of Selected Solids. Available online: https://www.engineeringtoolbox.com/density-solids-d_1265.html (accessed on 10 December 2019).
49. Scheibelhofer, O.; Balak, N.; Koller, D.M.; Khinast, J.G. Spatially resolved monitoring of powder mixing processes via multiple NIR-probes. *Powder Technol.* **2013**, *243*, 161–170. [CrossRef]
50. Togashi, D.; Alvarez-Jubete, L.; Rifai, H.; Cama-Moncunill, R.; Cruise, P.; Sullivan, C.; Cullen, P.J. Evaluation of diffuse reflectance near infrared fibre optical sensors in measurements for chemical identification and quantification for binary granule blends. *J. Near Infrared Spectrosc.* **2015**, *23*, 133–144. [CrossRef]
51. Wesley, I.J.; Larsen, N.; Osborne, B.G.; Skerritt, J.H. Non-invasive monitoring of dough mixing by near infrared spectroscopy. *J. Cereal Sci.* **1998**, *27*, 61–69. [CrossRef]

52. Dempster, E.; Olewnik, M.C.; Smail, V.W. Development of a controlled dough mixing system. In *Using Cereal Science and Technology for the Benefit of Consumers*; Woodhead Publishing: Sawston, UK, 1998; pp. 304–314. [CrossRef]
53. Aït Kaddour, A.; Barron, C.; Morel, M.; Cuq, B. Dynamic Monitoring of Dough Mixing Using Near-Infrared Spectroscopy: Physical and Chemical Outcomes. *Cereal Chem.* **2007**, *84*, 70–79. [CrossRef]
54. Perez Alvarado, F.A.; Hussein, M.A.; Becker, T. A Vision System for Surface Homogeneity Analysis of Dough Based on the Grey Level Co-occurrence Matrix (GLCM) for Optimum Kneading Time Prediction. *J. Food Process Eng.* **2016**, *39*, 166–177. [CrossRef]
55. Fowler, E.R. New techniques for: Commercial bread dough mixing. *IEEE Instrum. Meas. Mag.* **2000**, *3*, 21–25. [CrossRef]
56. Abdi, H.; Williams, L.J. Principal component analysis. *Wiley Interdiscip. Rev. Comput. Stat.* **2010**, *2*, 433–459. [CrossRef]
57. Gradient. Available online: https://uk.mathworks.com/help/matlab/ref/gradient.html#bvhp8_i (accessed on 13 March 2020).
58. Movmean. Available online: https://uk.mathworks.com/help/matlab/ref/movmean.html#bu2yug_-1_seealso (accessed on 13 March 2020).
59. Mallat, S.G.; Mallat, C. IV Time meets frequency. In *A Wavelet Tour of Signal Processing*; Elsevier Science & Technology: Amsterdam, The Netherlands, 1999; Volume 2, pp. 67–124.
60. Mallat, S.G. A Theory for Multiresolution Signal Decomposition: The Wavelet Representation. *IEEE Trans. Pattern Anal. Mach. Intell.* **1989**, *11*, 674–693. [CrossRef]
61. Safavian, L.S.; Kinsner, W.; Turanli, H. A QUANTITATIVE COMPARISON OF DIFFERENT MOTHER WAVELETS. *Can. Conf. Electr. Comput. Eng.* **2005**, *2005*, 1453–1456.
62. Mallat, S.G.; Mallat, C. 7.2 CLASSES OF WAVELET BASES. In *A Wavelet Tour of Signal Processing*; Elsevier Science & Technology: Amsterdam, The Netherlands, 1999; pp. 241–254.
63. Jain, A.K.; Mao, J.; Mohiuddin, K.M. Artificial Neural Networks: A Tutorial. *Computer (Long. Beach. Calif).* **1996**, *29*, 31–44. [CrossRef]
64. Choose a Multilayer Neural Network Training Function. Available online: https://uk.mathworks.com/help/deeplearning/ug/choose-a-multilayer-neural-network-training-function.html;jsessionid=e378b9dfbf595a83f44348fc1e7c (accessed on 13 March 2020).
65. Widodo, A.; Yang, B. Support vector machine in machine condition monitoring and fault diagnosis. *Mech. Syst. Signal Process.* **2007**, *21*, 2560–2574. [CrossRef]
66. Fitcsvm. Available online: https://uk.mathworks.com/help/stats/fitcsvm.html (accessed on 13 March 2020).
67. Ioffe, S.; Szegedy, C. Batch Normalization: Accelerating Deep Network Training by Reducing Internal Covariate Shift. In Proceedings of the 32nd International Conference on Machine Learning, ICML, Lille, France, 6–11 July 2015; Volume 1, pp. 448–456.
68. Hochreiter, S.; Schmidhuber, J. Long Short-Term Memory. *Neural Comput.* **1997**, *9*, 1735–1780. [CrossRef] [PubMed]
69. Scanlon, M.G. What Has Low-Intensity Ultrasound Informed Us about Wheat Flour Dough Rheology? *Cereal Foods World* **2013**, *58*, 61–65. [CrossRef]
70. Elmehdi, H.M.; Page, J.H.; Scanlon, M.G. Ultrasonic investigation of the effect of mixing under reduced pressure on the mechanical properties of bread dough. *Cereal Chem.* **2004**, *81*, 504–510. [CrossRef]
71. Mehta, K.L.; Scanlon, M.G.; Sapirstein, H.D.; Page, J.H. Ultrasonic Investigation of the Effect of Vegetable Shortening and Mixing Time on the Mechanical Properties of Bread Dough. *Food Eng. Phys. Prop.* **2009**, *74*, 455–461. [CrossRef] [PubMed]
72. Lee, H.O.; Luan, H.; Daut, D.G. Use of an Ultrasonic Technique to Evaluate the Rheological Properties of Cheese and Dough. *J. Food Eng.* **1992**, *16*, 127–150. [CrossRef]
73. Kaneko, H. Illustration of merits of semi-supervised learning in regression analysis. *Chemom. Intell. Lab. Syst.* **2018**, *182*, 47–56. [CrossRef]
74. Shi, X.; Xiong, W. Approximate linear dependence criteria with active learning for smart soft sensor design. *Chemom. Intell. Lab. Syst.* **2018**, *180*, 88–95. [CrossRef]
75. Cheplygina, V. Cats or CAT scans: Transfer learning from natural or medical image source data sets? *Curr. Opin. Biomed. Eng.* **2019**, *9*, 21–27. [CrossRef]

76. Malli, B.; Birlutiu, A.; Natschläger, T. Standard-free calibration transfer - An evaluation of different techniques. *Chemom. Intell. Lab. Syst.* **2017**, *161*, 49–60. [CrossRef]

5 Predicting Alcohol Concentration during Beer Fermentation Using Ultrasonic Measurements and Machine Learning

Article title: Predicting Alcohol Concentration during Beer Fermentation Using Ultrasonic Measurements and Machine Learning

Journal: Fermentation

Date published: 4/03/2021

DOI: 10.3390/fermentation7010034

Authors: Bowler, A.L., Escrig, J., Pound, M.P., Watson, N.J.

Author contributions (as published): Conceptualization, A.L.B., J.E. and N.J.W.; methodology, A.L.B., J.E., N.J.W. and M.P.P; software, A.L.B. and J.E.; validation, A.L.B.; formal analysis, A.L.B.; investigation, A.L.B. and J.E.; resources, A.L.B.; data curation, A.L.B. and J.E.; writing—original draft preparation, A.L.B.; writing—review and editing, A.L.B., N.J.W. and M.P.P; visualization, A.L.B.; supervision, N.J.W.; project administration, N.J.W.; funding acquisition, N.J.W. All authors have read and agreed to the published version of the manuscript.

The aim of this thesis was to develop ML methods to facilitate optimal deployment of US sensors for process monitoring applications in industrial environments. This article contributed to this aim by combining ML with a non-invasive, reflection-mode US sensing to monitor an alcoholic fermentation process. This contributed to the thesis objectives of collecting data from process monitoring applications that allow the thesis conclusions to extend to industrial scenarios as well as comparing reflection-mode and transmission-based US sensing approaches (see Section 1.2 Aims and Objectives). LSTM layers were determined to be more accurate than using ANNs with a fixed feature time-lag, echoing the results attained in the previous section (Section 4, titled: “Monitoring Mixing Processes Using Ultrasonic Sensors and Machine Learning”) further contributing to the thesis objective of developing an optimal ML pipeline for process monitoring using US sensor measurements. The evaluation of ML models omitting the process temperature as a feature was also novel for US monitoring of fermentation processes. This can be used to indicate whether the US sensor and ML combinations trialled could be used in industrial applications where measurement of the process temperature is not available. The choice of whether to include the temperature as a feature should be made during the model validation stage to determine whether it improves prediction accuracy.

In this work, the inclusion of the process temperature was shown to degrade ML model accuracy when using only the first waveform reflection (increasing the mean squared error from 0.146 to 0.345 % alcohol by volume). This was most likely due to the high variability of this feature (ranging between approximately 20 and 30 °C) creating a more difficult optimisation problem for the model. However, in subsequent works (Sections 6 and 8) an additional fully connected layer was used in the deep neural networks to reduce the burden on the LSTM layer to extract features, learn feature trajectories, and fit to the model outputs. Therefore, this new network structure is expected to handle the extra complexity of the process temperature and thereby achieve an increase in model accuracy, although, this was not confirmed in this thesis. The number of fully-connected layers should be trialled during the model validation stage.

In this work, LSTMs achieved higher accuracies (R^2 = 0.952, MAE = 0.265, and MSE = 0.136) compared with ANNs (R^2 = 0.398, MAE = 1.01, and MSE = 1.942). This shows that the gradients of the features, as provided to the ANNs, is insufficient, and the enhanced memory of the feature history provided by the LSTM units is required for the fermentation process. This is unsurprising, as LSTMs have been shown to outperform ANNs in many applications such as transit-time measurement in US flowmeters (0.95 for an LSTM compared with 0.9 R^2 for ANNs (Nguyen and Park, 2020)), gas volume fraction prediction in three-phase flows using US sensors (0.99 R^2 for LSTM methods compared with 0.97 R^2 for non-recurrent methods (Ren et al., 2021)), and damage detection using US sensors (0.999 R^2 for LSTM methods compared with 0.992 for non-recurrent neural networks (Huang et al., 2022)). The reason why the discrepancy between the LSTM and ANN results in this study is larger than the previously cited examples is likely due to the different volumes of data each algorithm had access to when making a prediction. In this work, the LSTMs had access to the full data series (i.e., the total fermentation times of 4 to 7 days) compared with the ANNs which only had feature gradient inputs from the previous 5 hours. In Nguyen and Park (2020) and Huang et al. (2022), the ANNs and LSTMs had access to the same volume of data with which to make a prediction, most likely producing the more similar R^2 values between the algorithms. This approach would not have been possible in this work as the LSTMs had access to the full trajectory of US measurements since the beginning of the process on the account of being able to process the features sequentially and output a prediction at each timestep. If the ANNs had access to all US measurements during the fermentation, this would have resulted in a single datapoint per fermentation batch owing to the ANNs inability to process the data sequentially.

The previous section (Section 4, titled: “Monitoring Mixing Processes Using Ultrasonic Sensors and Machine Learning”) monitored honey-water blending and flour-water batter mixing processes. Ultrasonic signals can be affected in three ways: variations in material acoustic impedance that alters the magnitude of the waveforms, variations in attenuation that alter the magnitude of the waveforms, and variations in sound velocity that alter the displacement of the waveforms in the time domain (Henning and Rautenburg, 2006). During mixing, the acoustic impedance at the measurement areas changes as the material composition varies whilst the sound velocity through the materials changes by conducting these processes over a range of temperatures. In this section, the final phenomena is also monitored: increases in attenuation as CO_2 bubbles are produced during the fermentation process. Throughout alcoholic beer fermentation, the acoustic impedance and speed of sound of a transmitted US wave also change as the density of the wort decreases.

In this work, a Tilt hydrometer was used to provide real-time density measurements, which floats at the surface of the fermenting wort. Therefore, altering the location of the hydrometer would not affect the density readings so long as it was not in contact with the sides of the vessel. In industrial applications, the wort density during fermentation is typically measured once or twice per day (Controllo e Misura, 2021). Measuring wort density provides information into the sugar concentration and alcohol level. Brewers can measure specific gravity during fermentation for multiple purposes. Firstly, the fermentation rate can be monitored. If fermentation proceeds too slowly, the yeast may be inactive, and the brewer can respond by adding more yeast or adjusting temperature in real-time to ensure fermentation progresses as intended (Precision Fermentation, 2023). Conversely, if fermentation occurs too rapidly, off-flavours may develop (Beer & Brewing, 2016). Secondly, monitoring specific gravity helps brewers determine when fermentation is complete to produce consistent batches. If fermentation is halted too early, the beer may be under-fermented, resulting in a sweet or low-alcohol beer. Thirdly, specific gravity measurements enable brewers to calculate the final alcohol content (ABV) of the beer. By taking initial and

final readings, brewers can determine how much sugar was converted to alcohol, thus calculating the final ABV (Brewer's Friend, 2023). Again, this is required for consistency as well as regulatory compliance. Lastly, tracking specific gravity lets brewers confirm fermentation has begun. If the specific gravity does not decrease over time, this could indicate problems with the yeast or temperature. By identifying whether fermentation has not started, corrective action can be taken or the batch may be abandoned. To monitor the rate of fermentation and alcohol content, a labelled data collection rate of once or twice per day would suffice owing to the fermentation processes requiring 4 to 7 days to complete. On the other hand, increased labelling at the start and end of ethanol production may be required to increase resolution at these points. If the times of these phenomena are not known, techniques such as interpolation or semi-supervised learning may be used to increase the labelling frequency at these points.

In this work, the ML models were trained to predict the alcohol levels during fermentation which can be used to achieve all these aims of monitoring fermentation rate, determining the start and end of fermentation, and predicting the final alcohol content of beer. Therefore, the conclusions drawn from this work are relevant to industrial beer fermentation monitoring. However, none of the final models produced in this article would be able to be utilised for all these aims (see Figure 5). For example, only Model 2 and 3 may be able to provide information as to whether fermentation had started, whereas Model 1 and 4 would be more useful to identify the end of fermentation and final alcohol content. Moreover, Model 2 appears to be most accurate at predicting the fermentation rate. However, in subsequent works (Sections 6 and 8) an additional fully connected layer was used in the deep neural networks to reduce the burden on the LSTM layer to learn the feature trajectories and also fit to the target variables. Therefore, this new network structure is expected to handle the extra complexity of the process temperature and thereby achieve an increase in model accuracy, although, this was not confirmed in this thesis. Furthermore, multi-task learning (as utilised in Section 6) may be used to enable learning the four aims separately to improve the accuracy for each whilst extracting useful information common to all tasks. Lastly, US measurements could be used in conjunction with continued sampling to provide real-time measurements and reduce manual burden on operators. Therefore, it can be decided which of these four aims is most pertinent to complement sampling activities and models can be trained to predict the chosen aim more accurately or this aim can be assigned a higher error weighting factor during multi-task training.

5.1 References

Beer & Brewing (2016) Available online: https://beerandbrewing.com/slow-and-steady-wins-the-race/ (Accessed on 19th April 2023).

Brewer's Friend (2023) Available online: https://www.brewersfriend.com/abv-calculator/ (Accessed on 19th April 2023).

Controllo e Misura (2021) The Brewing Process Is under Control. Available online: https://www.publiteconline.it/controlloemisura/2021/09/10/the-brewing-process-is-under-control/ (Accessed 18th April 2023)

Henning, B., Rautenberg, J. (2006) 'Process monitoring using ultrasonic sensor systems' *Ultrasonics* 44, e1395–e1399. Doi: 10.1016/j.ultras.2006.05.048.

Huang, L., Hong, X., Yang, Z., Liu, Y., Zhang, B. (2022) 'CNN-LSTM network-based damage detection approach for copper pipeline using laser ultrasonic scanning' *Ultrasonics* 121, 106685. Doi: 10.1016/j.ultras.2022.106685.

Nguyen, T.H.L., Park, S. (2020) ‘Intelligent ultrasonic flow measurement using linear array transducer with recurrent neural networks’ *IEEE Access* 8, 137564–137573. Doi: 10.1109/ACCESS.2020.3012037.

Precision Fermentation (2023) Available online: https://www.precisionfermentation.com/blog/yeast-pitching-rates-beer-fermentation/ (Accessed on 19th April 2023).

Ren, W., Jin, N., OuYang, L., Zhai, L., Ren, Y. (2021) ‘Gas volume fraction measurement of oil-gas-water three-phase flows in vertical pipe by combining ultrasonic sensor and deep attention network’ *IEEE T. Instrum. Meas.* 70, 1–9. Doi: 10.1109/TIM.2020.3031186.

fermentation

MDPI

Article

Predicting Alcohol Concentration during Beer Fermentation Using Ultrasonic Measurements and Machine Learning

Alexander Bowler [1], Josep Escrig [2], Michael Pound [3] and Nicholas Watson [1,*]

[1] Food, Water, Waste Research Group, Faculty of Engineering, University Park, University of Nottingham, Nottingham NG7 2RD, UK; alexander.bowler@nottingham.ac.uk
[2] i2CAT Foundation, Calle Gran Capita, 2-4 Edifici Nexus (Campus Nord Upc), 08034 Barcelona, Spain; josep.escrig@i2cat.net
[3] School of Computer Science, Jubilee Campus, University of Nottingham, Nottingham NG8 1BB, UK; michael.pound@nottingham.ac.uk
* Correspondence: nicholas.watson@nottingham.ac.uk

Abstract: Beer fermentation is typically monitored by periodic sampling and off-line analysis. In-line sensors would remove the need for time-consuming manual operation and provide real-time evaluation of the fermenting media. This work uses a low-cost ultrasonic sensor combined with machine learning to predict the alcohol concentration during beer fermentation. The highest accuracy model (R^2 = 0.952, mean absolute error (MAE) = 0.265, mean squared error (MSE) = 0.136) used a transmission-based ultrasonic sensing technique along with the measured temperature. However, the second most accurate model (R^2 = 0.948, MAE = 0.283, MSE = 0.146) used a reflection-based technique without the temperature. Both the reflection-based technique and the omission of the temperature data are novel to this research and demonstrate the potential for a non-invasive sensor to monitor beer fermentation.

Keywords: machine learning; ultrasonic measurements; long short-term memory; industrial digital technologies

Citation: Bowler, A.; Escrig, J.; Pound, M.; Watson, N. Predicting Alcohol Concentration during Beer Fermentation Using Ultrasonic Measurements and Machine Learning. *Fermentation* **2021**, *7*, 34. https://doi.org/10.3390/fermentation7010034

Received: 16 February 2021
Accepted: 2 March 2021
Published: 4 March 2021

1. Introduction

During beer fermentation, yeast metabolism produces ethanol and carbon dioxide from a sugar-water mixture called wort [1,2]. The fermentation is conventionally monitored through off-line wort density measurements until a predetermined ethanol concentration is reached [3], after which the process is continued for a predefined time for development of flavour compounds [4]. This requires manual sampling, takes time, and wastes resources by disposing of the measured sample. In-line measurement techniques directly measure the process material and on-line methods use bypasses to automatically collect, analyse, and return samples to the process [5]. By providing real-time, automatic alcohol concentration measurements, in-line and on-line techniques would ensure product quality through early detection of anomalous batches, allow effective scheduling of production equipment by predicting fermentation endpoint, and reduce the burden of manual sampling by operators. Furthermore, real-time data is key to the Fourth Industrial Revolution, which will implement industrial digital technologies such as the Internet of Things, cloud computing, and machine learning (ML) to integrate entire processes, automatically make decisions, and improve manufacturing productivity, efficiency, and sustainability [6].

Several in-line and on-line methods to monitor alcoholic fermentation have been investigated, including in-situ transflectance near-infrared spectroscopy [7,8], and Raman spectroscopy probes [9]; automated flow-through mid-infrared spectroscopy [10], Fourier transform infrared spectroscopy [11], and piezoelectric MEMS resonators [12]; non-invasive Raman spectroscopy through transparent vessel walls [13]; and CO_2 emission monitoring [14]. Ultrasonic (US) sensors are an attractive monitoring technique owing to their low

cost and have previously been used to study fermentation, including as in-line methods on circulation lines [15], in-situ in tanks [16], and using non-invasive, through-transmission of the fermenting media [17,18]. US monitoring techniques use high frequency (>1 MHz) and low power (<1 Wcm^{-2}) pressure waves to characterise material properties whilst causing no alterations to the material in which they propagate [19]. However, US properties vary with temperature and the presence of gas bubbles causes attenuation of the sound wave [20]. Previous in-line, on-line, and off-line studies to monitor fermentation using US measurements have developed empirical or semi-empirical models from the speed of sound or acoustic impedance to determine alcohol content [16]. These methods require extensive calibration procedures to compensate for the effects of temperature, dissolved CO_2 [16,18,21], and yeast cell concentration [18]. Supervised ML uses data to train predictive algorithms for classification or regression problems. Through ML, compensation procedures are not required as the complexities caused by varying process parameters imbedded in the sensor data can be unravelled. Furthermore, procedures for accurate determination of the speed of sound are not necessary [15,16,22].

This work presents three novel contributions to US monitoring of alcoholic fermentations: Firstly, ML is used to predict alcohol concentration during lab-scale beer fermentations from US measurements. Secondly, although an in-situ sensor probe is used, the potential for non-invasive monitoring of fermentation is investigated by only using the US wave reflected from the interface between the probe and the wort. This technique is similar to previous work by our group [23–26]. Implementation of this technique would provide in-line, non-invasive process monitoring without the need for circulation or bypass lines. This method would also not require transmission through the total vessel contents, which would be impossible at industrial scale. Therefore, this technique could be inexpensively fitted to the outside of existing vessels. Finally, exclusion of the temperature as a feature in the ML models is evaluated. Effective monitoring without the need for an invasive temperature sensor would further reduce the cost and complexity of industrial implementation.

2. Materials and Methods

The fermentation was conducted in a 30 L cylindrical plastic vessel (Figure 1). A lid sealed the vessel to protect the wort from contamination. The lid contained an air lock to release the CO_2 produced during fermentation. A belt heater increased the temperature of the wort to facilitate fermentation. The wort was prepared in the vessel by dissolving and mixing 1.5 kg of malt (Coopers Real Ale, UK) and 1 kg of sugar (brewing sugar, the Home Brew Shop, UK) in 22 L of water. Once the ingredients were mixed, a US probe was installed, consisting of a US transducer (Sonatest, 2 MHz central frequency, UK) and a temperature sensor (RTD, PT1000, UK). The US transducer was connected to a Lecouer Electronique US Box (France) that excited the transducer and digitised the received US signal. The temperature sensor was connected to a Pico electronic box (PT-104 Data Logger, UK). The two electronic boxes were connected to a laptop that controlled the data acquisition. Coupling gel was applied between the US transducer and the probe, and a spring was used to maintain the contact pressure. A Tilt hydrometer was installed to provide real-time density measurements. The real-time density measurements were required as the ground truth data of the wort alcohol concentration to train the ML models. This device was a small cylinder that floats in the liquid with its centre of gravity different from its centre of buoyancy. This causes an inclination of the device that is dependent on the specific gravity of the fermenting media. The inclination of the hydrometer was measured by a self-contained accelerometer and was transmitted by radio to a smartphone located outside of the vessel. A calibration procedure related the inclination to the specific gravity. It should be noted hydrometers are not suitable for in-line monitoring of industrial fermentations. Firstly, the balance of the device can be easily distorted by foam or solids floating on the surface, or by bubbles produced during fermentation. Secondly, as the hydrometer floats on the wort surface, it would need manual removal at the end of each fermentation batch. The most accurate method of specific gravity measurement is to extract

samples and use a portable density meter. However, this would require manual sample withdrawal at least every 2 h and would decrease the volume of liquid in the vessel, affecting the fermentation process. Furthermore, this would only produce sparse ground truth measurements of the density to train the ML models.

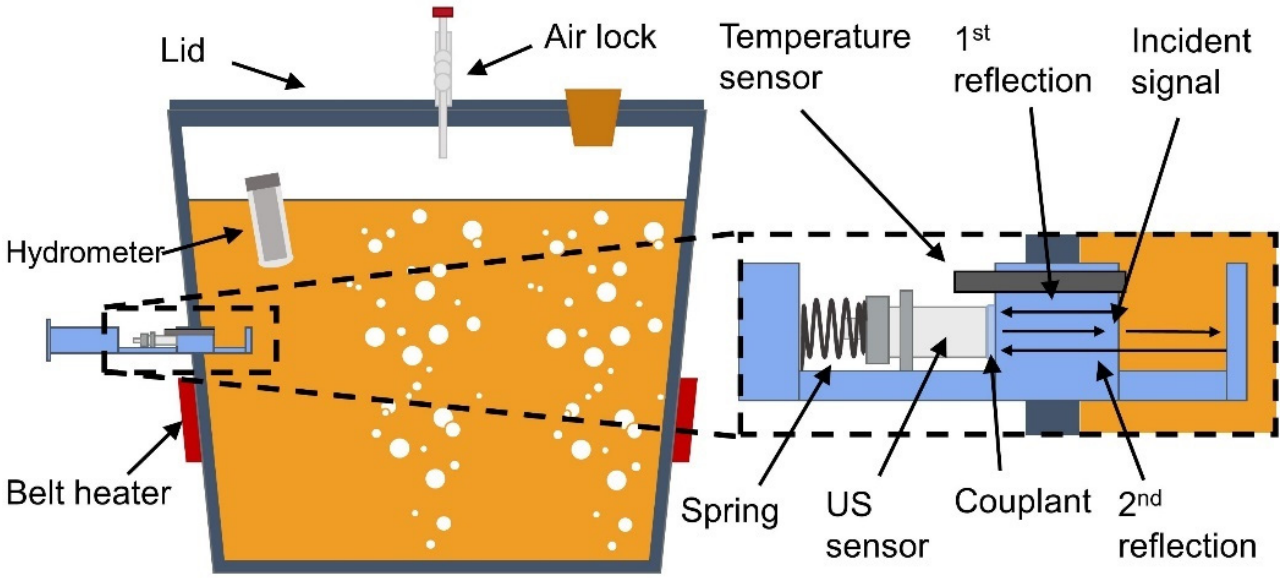


Figure 1. Experimental apparatus and measured US wave reflections.

The yeast (Coopers Real Ale, UK) was distributed on the surface and the vessel sealed. The mixture was left for 4 to 7 days while the fermentation occurred. After this time, the fermentation equipment was cleaned and a new batch was prepared. In total, 13 batches were completed. During fermentation, data was collected from the three different sensors: the US sensor, the temperature sensor, and the hydrometer. The time of each measurement was also recorded. The fermentation batches were conducted over a period of approximately 3 months. This meant that the environmental and water temperature in the laboratory changed during this time. Furthermore, the belt heater was only in contact with the lower section of the vessel. This produced temperature variations from around 20 to 30 °C. However, this temperature variation is beneficial to our ML evaluation as each model must be able to generalise across a wide range of process temperatures.

Sets of US and temperature data were collected periodically. Each of the sets consisted of 36 US waves and 36 temperature readings. For the US signal, 7000 sampling points were collected at 80 MHz sampling frequency. The time between each wave acquisition was 0.55 s. Between each set of data collection, 200 s elapsed. As depicted in Figure 1, the US transducers emitted sound waves which travelled along a PMMA buffer. At the interface between the buffer material and the wort, part of the sound wave is reflected back to the transducer (the 1st reflection). The rest of the sound wave continues through the wort, reflects at the opposite probe wall, and travels back to the transducer to be recorded (the 2nd reflection). An example of the signal recorded by the transducer is presented in Figure 2a. Close-ups of each reflection are presented in Figure 2b,c. The first section of the waveform (sample points < 500) is reflected back to the transducer before contacting the buffer material and wort interface and therefore contains no useful information about the process. The 1st reflection is identified between sample point 900 and 1500, and the 2nd reflection between 6000 and 6500, as shown in Figure 2.

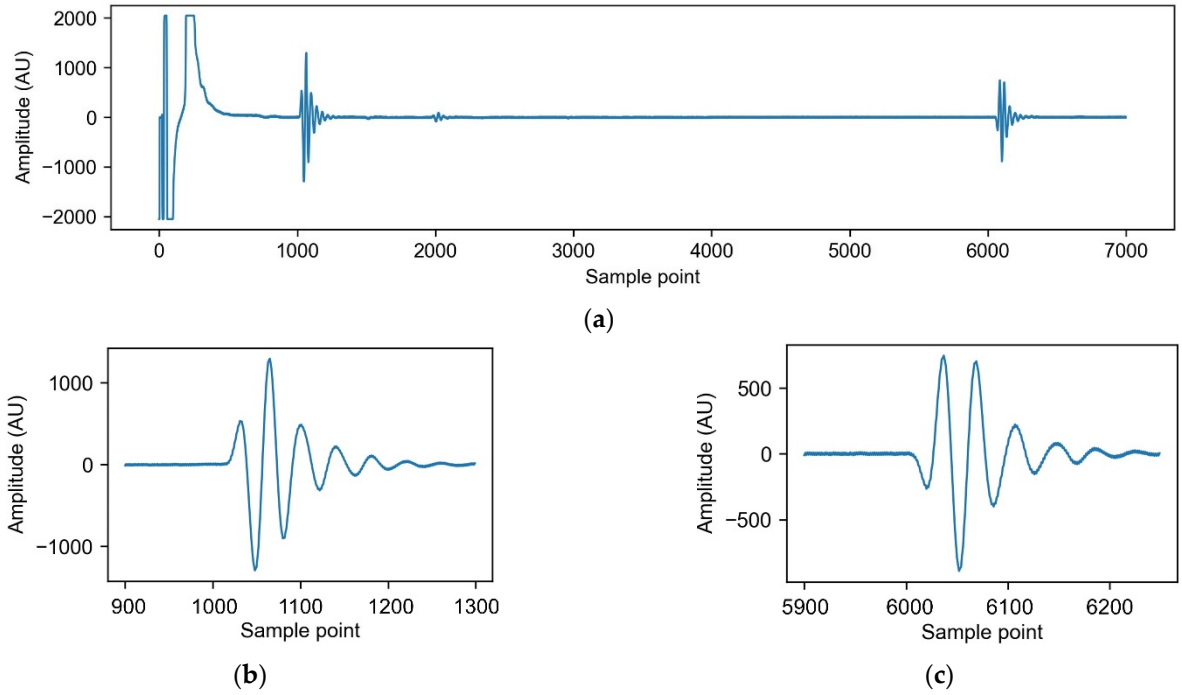


Figure 2. Example ultrasonic waveform obtained: (**a**) The 1st reflection is located around sample point 1000, the 2nd reflection is located around sample point 6000; (**b**) a close-up of the 1st reflection; (**c**) a close-up of the 2nd reflection.

2.1. Volume of Alcohol Calculation

The volume of alcohol (%) can be calculated from the specific gravity of the fermenting media using Equation (1) [27].

$$\text{ABV} = (\text{SG}_{\text{in}} - \text{SG}) \times 131.25, \quad (1)$$

where ABV is the alcohol by volume (%), SG_{in} is the starting specific gravity of the liquid before the yeast was added, and SG is the current specific gravity of the fermenting liquid. The multiplier of this equation is based on the stoichiometric relationship of the fermentation reaction, where the decreasing density is due to CO_2 production and escape through the air lock [28].

2.2. Ultrasonic Wave Features

The following features were calculated from the obtained US waveform to use in the ML models. These are common features extracted from US waveforms [29]. The theory behind the selection of each feature is presented in their respective sections. Different combinations of these features were tested during ML model optimisation. The optimal feature combinations are presented in Table 1, Section 3.1.

2.2.1. Energy

The waveform energy is a measure of the size of the waveform received by the transducer. For the 1st reflection, this is a measure of the proportion of the sound wave reflected from the interface between the buffer material and the wort. This is dependent on the change in acoustic impedance between these two materials [30]. Monitoring the waveform energy of the 2nd reflection offers additional information on the level of sound wave attenuation in the wort. This is caused by viscous losses in the media and scattering due to heterogeneities such as bubbles and yeast cells [30].

$$E = \sum_{i=start}^{i=end} {A_i}^2, \quad (2)$$

where E is the waveform energy, A_i is the waveform amplitude at sample point i, and *start* and *end* denote the range of samples points for the reflection of interest [29].

Table 1. Results for the long short-term memory neural network (LSTM) models. MAE: mean absolute error; MSE: mean squared error; ABV: alcohol by volume. The regression metrics for the models evaluated on the test set are highlighted in bold at the bottom of the table.

Model	1	2	3	4
Reflections	1st and 2nd	1st and 2nd	1st	1st
Temperature	Yes	No	Yes	No
Optimal features	• 1st reflection energy • 2nd reflection energy • 1st reflection energy standard deviation • 2nd reflection energy standard deviation • Time of flight • Temperature	• 1st reflection energy • 2nd reflection energy • 1st reflection energy standard deviation • 2nd reflection energy standard deviation • Time of flight	• 1st reflection energy • 1st reflection energy standard deviation • Temperature	• 1st reflection energy • 1st reflection energy standard deviation • 1st reflection peak-to-peak amplitude • 1st reflection maximum amplitude • 1st reflection minimum amplitude
Feature gradients	Yes	Yes	Yes	Yes
Batch size	2	2	4	4
Learning rate	0.01	0.01	0.033	0.033
LSTM units	2	2	4	4
L2 regularisation	0.0001	0.0001	0.00001	0.0001
Dropout rate	0	0	0	0
Epochs	100	100	100	100
Clip norm value	1	1	1	1
R^2	**0.952**	**0.939**	**0.878**	**0.948**
MAE (% ABV)	**0.265**	**0.355**	**0.426**	**0.283**
MSE (% ABV)	**0.136**	**0.173**	**0.345**	**0.146**

2.2.2. Peak-to-Peak Amplitude, Maximum Amplitude, and Minimum Amplitude

The peak-to-peak amplitude, maximum amplitude, and minimum amplitude provide additional information as to how the energy is distributed in the waveform. Changes in wort composition or temperature may affect how the sound wave travels and reflects from boundaries, presenting differences in the shape of the received waveform. These three features were calculated for both the 1st and 2nd reflections.

$$\mathrm{PPA} = \max(\mathrm{A_{start:end}}) - \min(\mathrm{A_{start:end}}), \quad (3)$$

$$\mathrm{A_{max}} = \max(\mathrm{A_{start:end}}), \quad (4)$$

$$\mathrm{A_{min}} = \min(\mathrm{A_{start:end}}), \quad (5)$$

where PPA is the peak-to-peak amplitude, $\mathrm{A_{max}}$ is the maximum amplitude, and $\mathrm{A_{min}}$ is the minimum amplitude.

2.2.3. Energy Standard Deviation

A total of 36 US waves were collected during each acquisition block. Phenomena in the process—e.g., the presence of bubbles at different times during fermentation—may cause fluctuations in the energy of the received waveforms. Therefore, the standard deviation of the energy in a block of acquired waveforms was investigated as a feature. The standard deviation of the energy was calculated for both the 1st and 2nd reflections.

$$STD = \sqrt{\frac{1}{W}\sum_{i=1}^{i=W}\left(E_i - \overline{E}\right)^2} \quad (6)$$

where *STD* is the standard deviation, *W* is the number of waveforms collected in the block, *i* is an individual waveform, and $\overline{E}$ is the mean waveform energy in the block.

2.2.4. Time of Flight

The time of flight was calculated using a thresholding method, i.e., the waveform sample point where the second reflection amplitude rises above the signal noise. This is a measure of the speed of sound in the wort that is dependent on its density and compressibility [20].

2.2.5. Feature Gradients

A one-sided, backwards moving mean was applied to obtain lagged feature representations over the previous 5 h. For the artificial neural networks (ANNs), this allows the use of past process information. For the long short-term memory neural networks (LSTMs), this allows for a way of storing past process information in some features, reducing the burden on the LSTM units to remember all feature trajectories.

2.3. Machine Learning

The ground truth data for the percentage volume of alcohol during fermentation was calculated from the portable density meter and hydrometer measurements. In total, 13 fermentation batches were monitored. The final two batches were selected as the test set to provide an unbiased assessment of the experimental methodology used. In an industrial setting, the final ML models would be deployed after collecting the training set runs. The remaining 11 batches were used in a 5-fold cross-validation procedure to optimize the ML models' hyperparameters. Long short-term memory neural networks (LSTMs) are able to retain information from previous time-steps in a sequence. LSTMs are a type of recurrent neural network that reduces the likelihood of vanishing or exploding gradients by using gate units. This enables their use over much longer sequences [31]. To evaluate the utility of using LSTMs to predict alcohol concentration, they were compared with artificial neural networks (ANNs) which are unable to store past process information. ANNs combine input features to produce new features which can approximate the relationship with the target variable given enough neurons in the hidden layer [32,33].

For the LSTMs, zero-padding was applied to the US features to make every fermentation batch sequence an equal length. A masking layer specified that the LSTM units ignore this padding. Each sequence consisted of 4646 timesteps. All timesteps for each batch were used as a single sequence rather than being split into multiple sequences of shorter length. While long LSTM sequences (250–500 timesteps) are prone to produce vanishing gradients when predicting a single output, this is not a problem when predicting an output at every timestep as used in this task [34].

For the ANNs, a single hidden layer and the Adam optimisation algorithm was used. Cross-validation determined the optimal batch size, number of neurons in the hidden layer, learning rate, drop-out rate, L2 regularisation penalty, and number of epochs for training. For the LSTMs, the Adam optimisation algorithm was used and the cross-validation procedure determined the optimal batch size, number of LSTM units, learning rate, drop-out rate, L2 regularisation penalty, gradient norm clipping value, and number of epochs. After cross-validation, the set of hyperparameters which resulted in the lowest average validation error were used to train a final model using all of the training set. The networks were trained using TensorFlow 2.3.0. The coefficient of determination (R^2), mean squared error (MSE) and mean absolute error (MAE) were used as performance metrics to evaluate the ML models. Multiple metrics produce a comprehensive assessment of a model's ability to fit to the test set and improve comparison between models.

3. Results

Figure 3 displays selected features from all the fermentation batches. It is shown that the energy of 1st reflection (Figure 3a), energy of the 2nd reflection (Figure 3b), and the

time of flight of the sound wave through the wort (Figure 3c) start at different values for each batch. There are several explanations for this. Firstly, as presented in Figure 3d, the process temperature is not the same at the start of each batch. As the speed of sound is highly dependent on temperature, the US properties begin from different magnitudes. Secondly, the US probe required manual removal and repositioning when disposing each batch after fermentation. This disturbed the spring maintaining the contact pressure of the US transducer, which affects the sound energy transferred through the materials from the sensor.

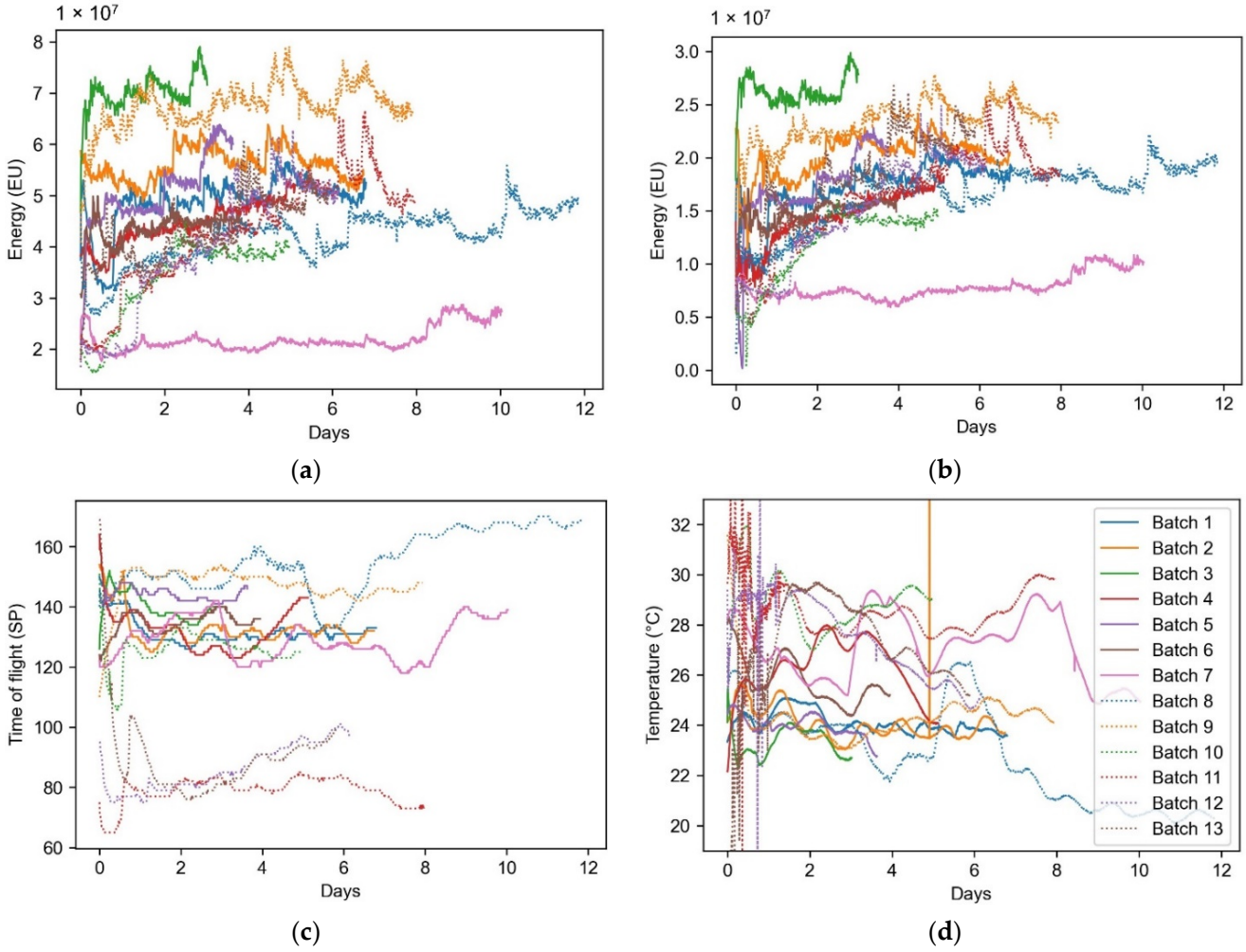


Figure 3. US waveform features for all fermentation batches: (**a**) The energy of the 1st reflection; (**b**) the energy of the 2nd reflection; (**c**) the time of flight for the 2nd reflection; (**d**) the process temperature.

The trajectories of the waveform features are also not smooth. Again, this is partly due to the oscillating process temperature. In addition, bubbles of CO_2 produced during the fermentation were observed to attach to the surface of the probe material, which would cause scattering and reflection of the sound wave. During the fermentation, as further CO_2 bubbles were produced, the new bubbles would replace the previous ones on the surface. This is likely to cause fluctuations in the waveform energy transferring through the interface between the probe and the wort.

The energy of the 1st reflection increases throughout the fermentation (Figure 3a). The energy of the 1st reflection is proportional to the change in acoustic impedance at the buffer-wort interface, with the acoustic impedance being a product of the material density and speed of sound [20]. As the density of the wort decreases during fermentation, the speed of sound also decreases as found in [17,18,35]. As the solid buffer material has a

greater density and speed of sound than the starting wort, the proportion of sound wave reflected at the buffer-wort material increases throughout the fermentation. However, the time of flight (Figure 3c), the inverse of the speed of sound, shows no general trend. This contrasts with the results obtained in [17,18,35], which suggests that it should increase. This is likely due to the changing process temperature masking an increasing time of flight. The results in [17,18,35] were all obtained at a constant temperature. The reduced time of flight for the last three batches (Batches 11, 12, and 13 in Figure 3c) is most likely due to a disturbance of the sensor positioning after Batch 10. These batches were kept in order to provide an unbiased assessment of the experimental methodology used. In an industrial setting because the test set data (batches 12 and 13) would not be available to analyse prior to the ML model training. The energy of the 2nd reflection is diminished compared with the 1st reflection at the beginning of the fermentation until approximately Day 3, as presented in Figure 4. A similar result was found in [17] and is due to the fermentation being most vigorous at the start of the process. This causes more CO_2 bubbles to be produced and therefore greater attenuation of the sound wave travelling through the wort.

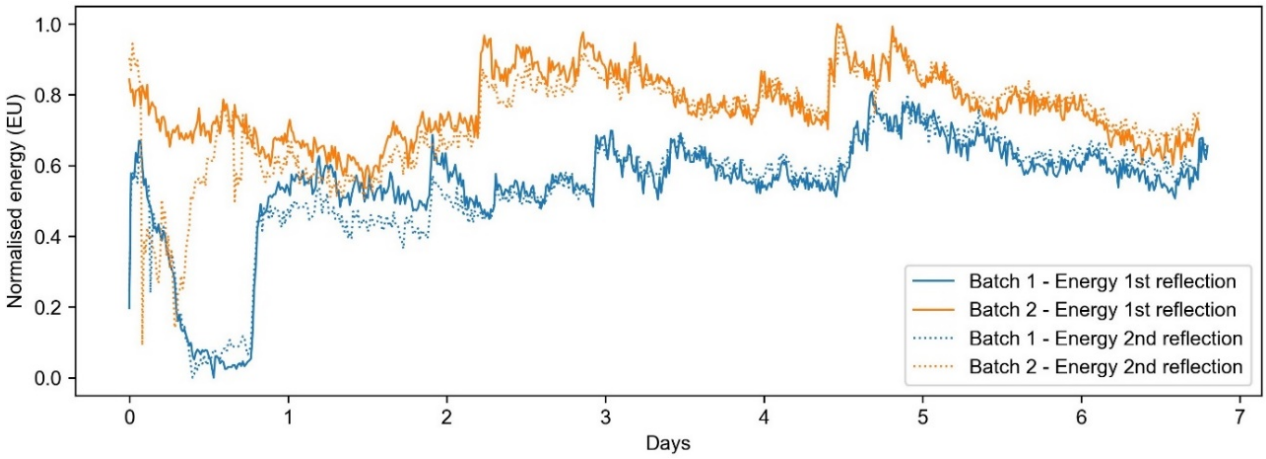


Figure 4. A comparison between the waveform energy of the 1st and 2nd reflections for batches 1 and 2. The energy of the 2nd reflection is diminished compared to the 1st until approximately Day 3. This is due to sound wave attenuation because of CO_2 bubbles being produced.

3.1. Machine Learning

The ANN model with the highest accuracy only achieved an R^2 of 0.398 (MAE = 1.010% ABV, MSE = 1.942% ABV). As such, only results from the LSTM models are included in Table 1. This shows that the gradients of the features, as provided to the ANNs, is insufficient, and the enhanced memory of the feature history provided by the LSTM units is required for this process. The results of four final LSTM models are presented in Table 1, which either use the 1st reflection or both reflections, and either use the temperature as a feature or not. The optimal features, optimal hyperparameters, and performance metrics are included. The most accurate LSTM model (Model 1) used features from both the 1st and 2nd reflections and the process temperature. Interestingly, the second most accurate model (Model 4) only used features from the 1st reflection, excluding the process temperature. The third most accurate model (Model 2) used features from the 1st and 2nd reflections without the process temperature. Finally, the least accurate model (Model 3) combined features from the 1st reflection and the process temperature. Graphical representations of these predictions are shown in Figure 5a–h.

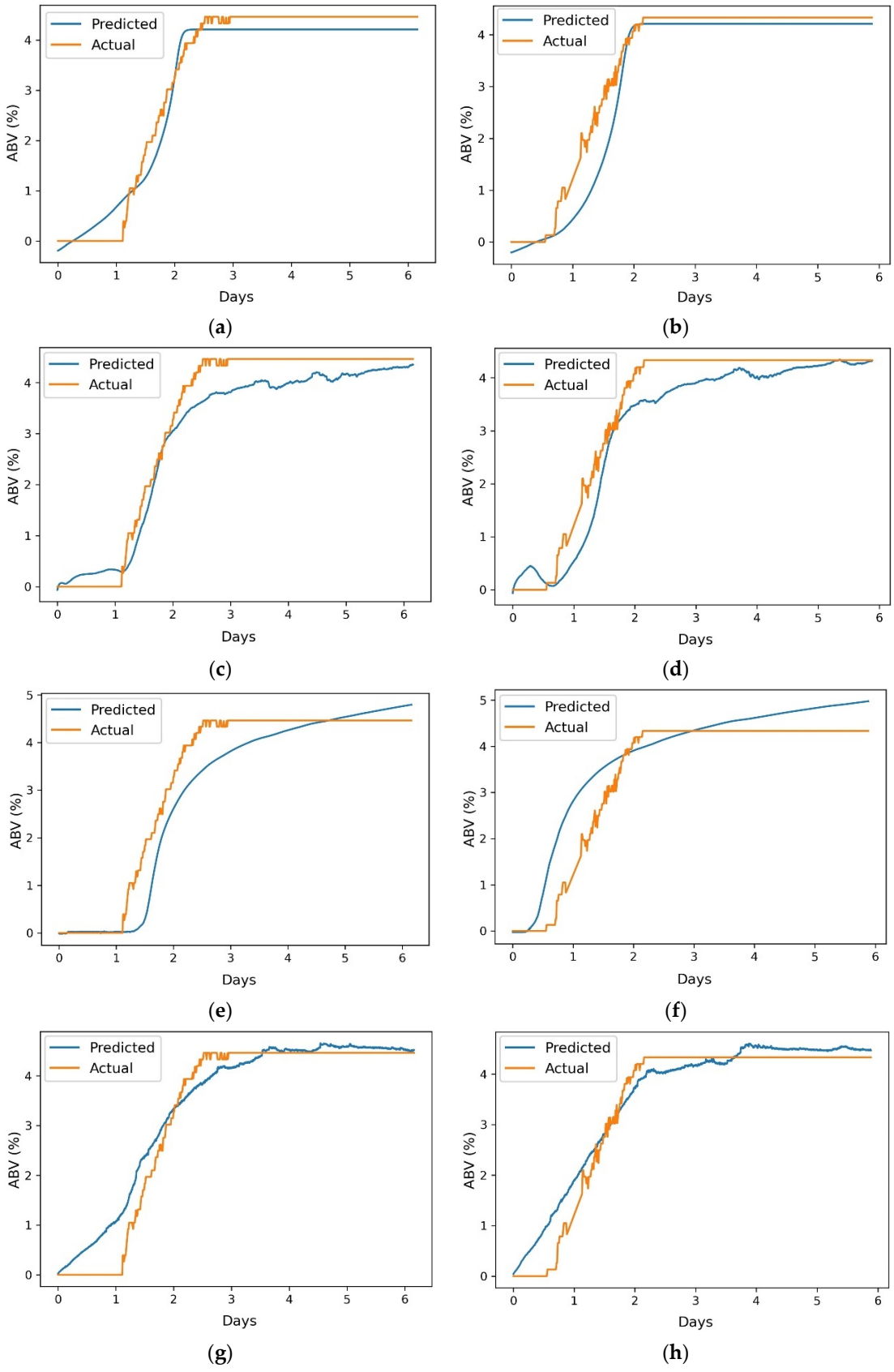


Figure 5. Predicted alcohol by volume percentage from the final LSTM models for the two test set batches (batches 12 and 13): (**a**) Model 1, Batch 12; (**b**) Model 1, Batch 13; (**c**) Model 2, Batch 12; (**d**) Model 2, Batch 13; (**e**) Model 3, Batch 12; (**f**) Model 3, Batch 13; (**g**) Model 4, Batch 12; (**h**) Model 4, Batch 13.

4. Discussion

The most accurate model (Model 1) uses features from both the 1st and 2nd reflections and the process temperature. This shows the potential of US sensors to predict the endpoint of fermentation and, as demonstrated in Figure 5a,b, accurately predict the alcohol concentration throughout the fermentation process. However, industrial implementation of this model would require the use of an invasive probe in order to obtain the 2nd reflection. In addition, an invasive temperature probe would be required to monitor the changing temperature of the fermentation media. Interestingly, the second most accurate model (Model 4) only used features from the 1st reflection and excluded the process temperature. The use of only the 1st reflection indicates that accurate results could be obtained using a non-invasive, no-transmission US sensor, similar to the techniques used in previous works by our group [23–26]. This is advantageous as it allows the alcohol volume to be accurately predicted by easily mounting a US sensor externally to an existing vessel. Therefore, it can be easily implemented into existing industrial settings at low effort and cost. The use of Model 4 would also remove the requirement for an invasive process temperature measurement. Furthermore, the performance metrics for Model 4 (R^2 = 0.948, MAE = 0.283, MSE = 0.146) are similar to those of Model 1 (R^2 = 0.952, MAE = 0.265, MSE = 0.136) indicating that no prediction accuracy would be lost through using a non-invasive and no-transmission sensor approach. This US sensing technique would also not require a hole to be bored into the vessel side, as used in this work. Instead, the US wave could be transmitted through the vessel wall.

In Model 3, the features from the 1st reflection combined with the process temperature produces a reduced accuracy. This is likely because the additional features required in Model 4 (the peak-to-peak amplitude, maximum amplitude, and minimum amplitude of the 1st reflection) contained more pertinent information about the temperature at the probe-wort interface than the non-local temperature sensor. The suggestion that the temperature sensor measured the temperature of the bulk wort instead of the region through which the 1st reflection passes is supported by the results from Model 2. When the temperature was removed as a feature, Model 2 produced a reduced accuracy compared with Model 1. This indicates that for accurate prediction using the 2nd reflection, the bulk wort temperature measurement is required as the sound wave travels through this region. The reduced accuracy obtained when combining the temperature data with the 1st reflection for Model 3 is most likely caused by the temperature at the probe-wort interface not closely following the trend of the bulk wort temperature. Therefore, using the temperature measurement as a feature increases the model complexity with little benefit, meaning it is more difficult for the network to find an optimal solution. This further supports the aforementioned point that accurate, invasive temperature measurement would not be required with a non-invasive, no-transmission US sensing technique.

Figure 5 displays the predicted ABV percentage from the trained LSTM models for the two batches used for the test set (batches 12 and 13). Model 1 (Figure 5a and b) accurately determines the fermentation endpoint. However, the final ABV prediction is not as accurate as Model 4, indicating that it may not be sensitive enough to determine differences in final ABV between batches. Whilst Model 3 appears to have no utility, Models 1, 2, and 4 all accurately followed the ABV trajectory. Owed to the real-time data acquisition of US sensors, these models suggest that the obtained data could be used to train additional anomaly detection models to provide early warning of undesired process trajectories within a batch.

Several locations in the prediction require improvement; for example, the detection of ABV plateau for Model 2 (Figure 5c,d around the 2nd day), the settling at a final ABV for Model 3 (Figure 5e,f), and the detection of the initial ABV rise for Model 4 (Figure 5g,h around the first day). This is likely due to the varying temperature throughout the fermentation having a large effect on the US properties of the wort compared with the changing density. There are also locations of decreasing ABV prediction (Figure 5d during the first day) or sudden increases in ABV prediction (Figure 5h at the end of the fourth

day). This is likely due to the temperature variations being different for each batch and the particular temperature variations during the test set causing these effects. These problems would likely be reduced through obtaining more training data.

In this work, ML models were trained to predict the ABV throughout the fermentation. However, in industrial settings this may not be the most appropriate output value with which to fit a model. For example, ML models could be trained to predict the final ABV of each batch, the time remaining until the ABV plateaus, classify the end of fermentation, or provide early detection of anomalous batches. In each of these cases, the models would be trained for a more specific purpose, as such the models may perform better than indicated by Figure 5. This work is therefore demonstrative of the efficacy of real-time fermentation monitoring using US sensors and ML, and increased accuracy may be achieved through predictions of more specific outputs.

If only the 1st reflection was used in an industrial monitoring system, the sound wave could be transferred through the vessel wall. Alternatively, if the 2nd reflection was also to be used, the probe could be fitted through existing ports common to industrial fermenters. This work monitored a laboratory scale fermentation process. At industrial scale, agitation methods are uncommon in beer fermentation to prevent damage to the yeast [36]. Therefore, radial variations in alcohol concentration exist and there would be a difference in the alcohol concentration at the sensor measurement area and the bulk wort [36]. However, previous work from our group showed that a non-local probe could accurately monitor a mixing process [23]. This is because, through machine learning, the sensor data is correlated to the location of the ground truth data, rather than the sensor. In this case, a sensor would be trained to predict the alcohol concentration at the location of the hydrometer measurements or sample collection.

Future Research Directions

The largest barrier to industrial implementation of sensors and ML combined technologies is the burden of obtaining labelled data. Labelled data is used as the targets for training supervised ML models. To obtain the ground truth to label data requires another analysis method. In this work, the hydrometer readings were used due to the sample density measurements producing insufficient data points and disturbing the fermentation. In an industrial setting, the hydrometer may only be able to be used for a small number of batches for ML model development. In this case, semi-supervised learning may be required to train high accuracy models. Semi-supervised learning uses both labelled and unlabelled samples to train a model [37–39]. Firstly, unsupervised learning techniques, such as principal component analysis or autoencoders, can be used on the total dataset to learn relationships between features across the labelled and unlabelled samples. Then traditional supervised learning can be used on the new features using just the labelled samples. Secondly, a self-training (or pseudo-labelling) approach may be used to predict the labels of the unlabelled data from the trained model. These pseudo-labels may then be added to the labelled data set and the procedure repeated to improve the label predictions or to train a final model.

Alternatively, conventional sample extraction and density measurement may be used to obtain the labelled data. Either a curve may be fitted to these sparse density measurements to produce interpolated data points, or a similar semi-supervised learning procedure can be implemented. Active learning may also be used to identify data points for labelling that may be the most useful to the model [40,41]. These datapoints may be during a sparsely sampled time in the fermentation or be in a particular temperature and composition range. Operators could then analyse these samples to provide the most benefit to the ML model at the lowest investment in effort.

5. Conclusions

The transition to Industry 4.0 promises increased manufacturing efficiency, sustainability, and productivity. By implementing digital technologies such as the Internet of

Things, Cloud Computing and ML, not only can entire processes be integrated, but supply chains as well. Sensors are a key technology in this revolution by providing the real-time data to inform automatic, intelligent decision-making. Currently, beer fermentation is monitored through periodic manual sampling and off-line wort density measurements. This work has presented an in-line, low-cost US sensing technique combined with ML, which would remove the need of operator sampling. This work has shown that US sensor data combined with LSTM models are able to accurately predict the volume of alcohol during beer fermentation. The highest accuracy model (R^2 = 0.952) used a transmission-based ultrasonic sensing technique along with the process temperature. Importantly, the second most accurate model (R^2 = 0.948) only used a reflection-based technique without measurement of the temperature. This demonstrates the potential for a non-invasive, no-transmission US technique, which doesn't require invasive measurement of the process temperature. This sensing technique could be easily and inexpensively retrofitted onto existing fermentation vessels.

Author Contributions: Conceptualization, A.B., J.E. and N.W.; methodology, A.B., J.E., N.W. and M.P.; software, A.B. and J.E.; validation, A.B.; formal analysis, A.B.; investigation, A.B. and J.E.; resources, A.B.; data curation, A.B. and J.E.; writing—original draft preparation, A.B.; writing—review and editing, A.B., N.W. and M.P.; visualization, A.B.; supervision, N.W.; project administration, N.W.; funding acquisition, N.W. All authors have read and agreed to the published version of the manuscript.

Funding: This work was supported by the Engineering and Physical Sciences Research Council (EPSRC) standard research studentship (EP/R513283/1) and EPSRC network+ Connected Everything (EP/P001246/1).

Data Availability Statement: Researchers at the University of Nottingham can be contacted for access to data.

Conflicts of Interest: The authors declare no conflict of interest.

References

1. Schock, T.; Becker, T. Sensor array for the combined analysis of water–sugar–ethanol mixtures in yeast fermentations by ultrasound. *Food Control* **2010**, *21*, 362–369. [CrossRef]
2. Resa, P.; Elvira, L.; de Espinosa, F.M.; Gómez-Ullate, Y. Ultrasonic velocity in water–ethanol–sucrose mixtures during alcoholic fermentation. *Ultrasonics* **2005**, *43*, 247–252. [CrossRef]
3. Jan, M.V.S.; Guarini, M.; Guesalaga, A.; Pérez-Correa, J.R.; Vargas, Y.; Perez-Correa, J. Ultrasound based measurements of sugar and ethanol concentrations in hydroalcoholic solutions. *Food Control* **2008**, *19*, 31–35. [CrossRef]
4. Kucharczyk, K.; Tuszyński, T. The effect of wort aeration on fermentation, maturation and volatile components of beer produced on an industrial scale. *J. Inst. Brew.* **2017**, *123*, 31–38. [CrossRef]
5. De Beer, T.; Burggraeve, A.; Fonteyne, M.; Saerens, L.; Remon, J.; Vervaet, C. Near infrared and Raman spectroscopy for the in-process monitoring of pharmaceutical production processes. *Int. J. Pharm.* **2011**, *417*, 32–47. [CrossRef] [PubMed]
6. Oztemel, E.; Gursev, S. Literature review of Industry 4.0 and related technologies. *J. Intell. Manuf.* **2020**, *31*, 127–182. [CrossRef]
7. Vann, L.; Layfield, J.B.; Sheppard, J.D. The application of near-infrared spectroscopy in beer fermentation for online monitoring of critical process parameters and their integration into a novel feedforward control strategy. *J. Inst. Brew.* **2017**, *123*, 347–360. [CrossRef]
8. Corro-Herrera, V.A.; Gómez-Rodríguez, J.; Hayward-Jones, P.M.; Barradas-Dermitz, D.M.; Gschaedler-Mathis, A.C.; Aguilar-Uscanga, M.G. Real-time monitoring of ethanol production during Pichia stipitis NRRL Y-7124 alcoholic fermentation using transflection near infrared spectroscopy. *Eng. Life Sci.* **2018**, *18*, 643–653. [CrossRef] [PubMed]
9. Wang, Q.; Li, Z.; Ma, Z.; Liang, L. Real time monitoring of multiple components in wine fermentation using an on-line auto-calibration Raman spectroscopy. *Sens. Actuators B Chem.* **2014**, *202*, 426–432. [CrossRef]
10. Mazarevica, G.; Diewok, J.; Baena, J.R.; Rosenberg, E.; Lendl, B. On-Line Fermentation Monitoring by Mid-Infrared Spectroscopy. *Appl. Spectrosc.* **2004**, *58*, 804–810. [CrossRef]
11. Veale, E.; Irudayaraj, J.; Demirci, A. An On-Line Approach to Monitor Ethanol Fermentation Using FTIR Spectroscopy. *Biotechnol. Prog.* **2007**, *23*, 494–500. [CrossRef]
12. Toledo, J.; Ruiz-Díez, V.; Pfusterschmied, G.; Schmid, U.; Sánchez-Rojas, J. Flow-through sensor based on piezoelectric MEMS resonator for the in-line monitoring of wine fermentation. *Sens. Actuators B Chem.* **2018**, *254*, 291–298. [CrossRef]
13. Schalk, R.; Frank, R.; Rädle, M.; Methner, F.-J.; Beuermann, T.; Braun, F.; Gretz, N. Non-contact Raman spectroscopy for in-line monitoring of glucose and ethanol during yeast fermentations. *Bioprocess Biosyst. Eng.* **2017**, *40*, 1519–1527. [CrossRef] [PubMed]

14. Ete-Carmona, E.C.; Gallego-Martinez, J.-J.; Martin, C.; Brox, M.; Luna-Rodriguez, J.-J.; Moreno, J. A Low-Cost IoT Device to Monitor in Real-Time Wine Alcoholic Fermentation Evolution through CO_2 Emissions. *IEEE Sens. J.* **2020**, *20*, 6692–6700. [CrossRef]
15. Hussein, W.B.; Hussein, M.A.; Becker, T. Robust spectral estimation for speed of sound with phase shift correction applied online in yeast fermentation processes. *Eng. Life Sci.* **2012**, *12*, 603–614. [CrossRef]
16. Hoche, S.; Krause, D.; Hussein, M.A.; Becker, T. Ultrasound-based, in-line monitoring of anaerobe yeast fermentation: Model, sensor design and process application. *Int. J. Food Sci. Technol.* **2016**, *51*, 710–719. [CrossRef]
17. Resa, P.; Elvira, L.; De Espinosa, F.M. Concentration control in alcoholic fermentation processes from ultrasonic velocity measurements. *Food Res. Int.* **2004**, *37*, 587–594. [CrossRef]
18. Resa, P.; Elvira, L.; De Espinosa, F.M.; González, R.; Barcenilla, J. On-line ultrasonic velocity monitoring of alcoholic fermentation kinetics. *Bioprocess Biosyst. Eng.* **2008**, *32*, 321–331. [CrossRef]
19. Ojha, K.S.; Mason, T.J.; O'Donnell, C.P.; Kerry, J.P.; Tiwari, B.K. Ultrasound technology for food fermentation applications. *Ultrason. Sonochem.* **2017**, *34*, 410–417. [CrossRef]
20. Henning, B.; Rautenberg, J. Process monitoring using ultrasonic sensor systems. *Ultrasonics* **2006**, *44*, e1395–e1399. [CrossRef]
21. Schock, T.; Hussein, M.; Hitzmann, B.; Becker, T. Influence of dissolved carbon dioxide on the sound velocity and adiabatic compressibility in aqueous solutions with saccharose and ethanol. *J. Mol. Liq.* **2012**, *175*, 111–120. [CrossRef]
22. Hoche, S.; Hussein, W.B.; Hussein, M.A.; Becker, T. Time-of-flight prediction for fermentation process monitoring. *Eng. Life Sci.* **2011**, *11*, 417–428. [CrossRef]
23. Bowler, A.L.; Bakalis, S.; Watson, N.J. Monitoring Mixing Processes Using Ultrasonic Sensors and Machine Learning. *Sensors* **2020**, *20*, 1813. [CrossRef] [PubMed]
24. Escrig, J.; Woolley, E.; Simeone, A.; Watson, N. Monitoring the cleaning of food fouling in pipes using ultrasonic measurements and machine learning. *Food Control* **2020**, *116*, 107309. [CrossRef]
25. Escrig, J.E.; Simeone, A.; Woolley, E.; Rangappa, S.; Rady, A.; Watson, N. Ultrasonic measurements and machine learning for monitoring the removal of surface fouling during clean-in-place processes. *Food Bioprod. Process.* **2020**, *123*, 1–13. [CrossRef]
26. Escrig, J.; Woolley, E.; Rangappa, S.; Simeone, A.; Watson, N. Clean-in-place monitoring of different food fouling materials using ultrasonic measurements. *Food Control* **2019**, *104*, 358–366. [CrossRef]
27. Kitchn. Available online: https://www.thekitchn.com/how-to-check-and-control-alcohol-levels-the-kitchns-beer-school-2015-217260 (accessed on 14 January 2021).
28. BrewMoreBeer. Available online: http://www.brewmorebeer.com/calculate-percent-alcohol-in-beer/ (accessed on 24 February 2021).
29. Zhan, X.; Jiang, S.; Yang, Y.; Liang, J.; Shi, T.; Li, X. Inline Measurement of Particle Concentrations in Multicomponent Suspensions using Ultrasonic Sensor and Least Squares Support Vector Machines. *Sensors* **2015**, *15*, 24109–24124. [CrossRef]
30. McClements, D. Advances in the application of ultrasound in food analysis and processing. *Trends Food Sci. Technol.* **1995**, *6*, 293–299. [CrossRef]
31. Hochreiter, S.; Schmidhuber, J. Long Short-Term Memory. *Neural Comput.* **1997**, *9*, 1735–1780. [CrossRef]
32. LeCun, Y.; Bengio, Y.; Hinton, G. Deep learning. *Nature* **2015**, *521*, 436–444. [CrossRef]
33. Jain, A.; Mao, J.; Mohiuddin, K. Artificial neural networks: A tutorial. *Computer* **1996**, *29*, 31–44. [CrossRef]
34. Machine Learning Mastery. Available online: https://machinelearningmastery.com/handle-long-sequences-long-short-term-memory-recurrent-neural-networks/ (accessed on 9 February 2021).
35. Lamberti, N.; Ardia, L.; Albanese, D.; Di Matteo, M. An ultrasound technique for monitoring the alcoholic wine fermentation. *Ultrasonics* **2009**, *49*, 94–97. [CrossRef]
36. Nienow, A.W.; McLeod, G.; Hewitt, C.J. Studies supporting the use of mechanical mixing in large scale beer fermentations. *Biotechnol. Lett.* **2010**, *32*, 623–633. [CrossRef] [PubMed]
37. Kostopoulos, G.; Karlos, S.; Kotsiantis, S.; Ragos, O. Semi-supervised regression: A recent review. *J. Intell. Fuzzy Syst.* **2018**, *35*, 1483–1500. [CrossRef]
38. Forestier, G.; Wemmert, C. Semi-supervised learning using multiple clusterings with limited labeled data. *Inf. Sci.* **2016**, *361-362*, 48–65. [CrossRef]
39. Ge, Z.; Song, Z.; Ding, S.X.; Huang, B. Data Mining and Analytics in the Process Industry: The Role of Machine Learning. *IEEE Access* **2017**, *5*, 20590–20616. [CrossRef]
40. Burbidge, R.; Rowland, J.J.; King, R.D. Active Learning for Regression Based on Query by Committee. *Comput. Vis.* **2007**, *4881*, 209–218. [CrossRef]
41. Cai, W.; Zhang, Y.; Zhou, J. Maximizing expected model change for active learning in regression. In Proceedings of the 13th IEEE International Conference on Data Mining, Dallas, TX, USA, 7–10 December 2013; pp. 51–60.

6 Convolutional feature extraction for process monitoring using ultrasonic sensors

Article title: Convolutional feature extraction for process monitoring using ultrasonic sensors

Journal: Computers and Chemical Engineering

Date published: 28/08/2021

DOI: 10.1016/j.compchemeng.2021.107508

Authors: Bowler, A.L., Pound, M.P., Watson, N.J.

Author contributions (as published): A.L.B.: Conceptualization, Data curation, Formal analysis, Investigation, Methodology, Software, Validation, Visualization, Writing – original draft, Writing – review & editing. M.P.P.: Conceptualization, Supervision, Writing – review & editing. N.J.W.: Funding acquisition, Project administration, Resources, Supervision, Writing – review & editing.

The aim of this thesis is to develop ML methods to facilitate optimal deployment of US sensors for process monitoring applications in industrial environments. This article contributed to this aim by evaluating a convolutional feature extraction method and comparing it to the next best method: the extraction of coarse features such as energy, standard deviation, or skewness (Bowler et al., 2022). The methods were evaluated on material mixing, cleaning of pipe sections, and alcoholic fermentation. The convolutional feature extraction method produced more informative features than the coarse feature extraction method. Furthermore, the use of multi-task learning with US data and US data augmentation for process monitoring was novel. Multi-task learning was also investigated to aid LSTM learning of the process trajectory. Multi-task learning is an ML approach where a single model is trained on multiple related tasks simultaneously, sharing some or all of the model parameters between tasks to improve performance on all tasks (Caruana, 1997). By training on two correlated tasks, the shared LSTM layer may learn more effective feature trajectories while reducing redundant information being stored. This may have several benefits, such as increased model accuracy through global learning of feature trajectories important to the process being monitored, more stable model training by optimising for two combined losses, and reducing overfitting by preventing a single task from dominating the learning process (Zhang and Qiang, 2018). This contributed to the thesis objective of developing and optimal ML pipeline for process monitoring using US sensor measurements.

This work concluded the ML pipeline optimisation portion of this thesis, and the final method is depicted in Figure 1. The method used a pretrained 1D CNN as a feature extractor. The CNN was pre-trained on an auxiliary task to classify waveform dataset membership of the all the experimental datasets. The output of this CNN is therefore a multi-class classification of waveform dataset membership. Segments of 1000 sample points in length were selected from each waveform. The position of the 1000 sample point length window was chosen for each waveform by investigating the difference between the start and end waveforms of the individual process. The areas with the largest visual change throughout the process were used. To increase the training set size for training the CNN, and to improve informative feature learning in the convolutional layers, a 600 × 1 input to the CNN was used. Data augmentation using a sliding window, laterally translated by 100 sample points each time, produced five waveform segments of 600 sample points in length. Further data

augmentation used separate normalisation of each waveform segment to differentially magnify the waveform. This was to ensure that the CNN learned features specific to each waveform, rather than the position or magnitude of features. This data augmentation approach increased the dataset size five-fold. The CNN was used as a fixed weight feature extractor. PCA was used to extract a small set of orthogonal features from the CNN outputs to be combined with additional features such as the time of flight or deviations between consecutively acquired waveforms. These features are then used as inputs in a deep neural network with LSTM layers. This method is used to produce features from US waveforms for every timestep (Figure 1).

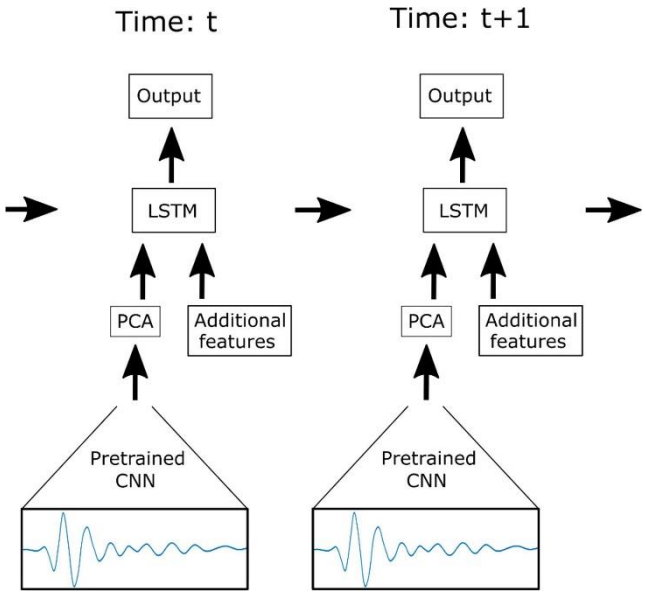


Figure 1: The optimised ML pipeline for process monitoring using US measurements (Bowler et al., 2022).

Table 1 presents the percentage variability explained by each PC for the US waveform datasets and the number of PCs required to explain 95% of the variability, a commonly used method to determine the number of PCs to utilise (Valle et al., 1999). In this work, the first PC likely follows the common waveform changes across the full dataset caused by variations in the US properties of the materials being monitored (either due to changing composition or process temperature). Successive PCs will identify waveform changes more specific to each batch, most likely due to the different process temperatures. Therefore, it is anticipated that only a small number of PCs are required (i.e. greater than one) to monitor the changing material composition and account for changes in the monitoring US waveform at different temperatures. This is supported by Table 1 where the percentage variability explained drops off after the first two PCs. As shown in Table 1, the smallest number of PCs required to explain 95% of the variability in the dataset, is eight for the Plastic Cleaning dataset and nine for fermentation monitoring using only the first reflection. Therefore, using these two pieces of guidance (the primacy of the first and second PCs and the smallest number of PCs to explain 95% of dataset variability), five PCs were selected to obtain useful waveform information while minimising noise.

Table 1. A summary of the distribution of the explained variance by each PC for the US waveform datasets after convolutional feature extraction.

Experimental dataset	Waveforms	Number of PCs to explain	Variability explained	Variability explained	Variability explained	Variability explained	Variability explained

		95% of variability	by 1st PC (%)	by 2nd PC (%)	by 3rd PC (%)	by 4th PC (%)	by 5th PC (%)
Fermentation	Reflection 1	9	56.4	23.1	9.2	2.1	1.5
	Reflection 2	18	30.4	21.6	14.9	9.0	6.1
Cleaning of food fouling from pipe sections	Flat rig	15	60.4	15.2	7.4	4.3	1.8
	Circular, plastic	8	56.7	14.3	12.4	6.3	1.9
	Circular, metal	32	50.9	12.3	8.5	4.6	3.7
Honey-water mixing 1	Central sensor	24	52.1	18.8	7.5	4.6	2.3
	Non-central sensor	41	51.4	17.0	4.7	3.8	2.8
Honey-water mixing 2	Central sensor	19	38.6	30.8	12.4	4.1	2.9
	Non-central sensor	25	41.6	36.8	4.6	2.7	2.4
Batter mixing	Sensor 1	42	49.1	15.1	14.3	4.5	2.6
	Sensor 2	16	60.5	16.3	7.5	3.1	1.5

As determined in the previous sections (Sections 4 and 5), LSTM layers are used to allow the ML model to learn the trajectory of the US features. The coarse feature extraction method as used in Section 5 (titled: “Predicting Alcohol Concentration during Beer Fermentation Using Ultrasonic Measurements and Machine Learning”) is compared with the convolutional method presented in this section. The method is designated as “coarse” owing to it extracting many features from the US waveform that encompass the waveform changes, compared with the convolutional method which directly monitors these changes by measuring the amplitude of every waveform sample point.

In previous literature, the ML methods combined with US sensor data for cleaning and fermentation have either used coarse waveform features (for example, the waveform energy (Wallhäußer et al., 2011), crest factor (Wallhäußer et al., 2013), or maximum amplitude (Úbeda et al., 2016)), the amplitude at every sample point in a waveform (Escrig et al., 2020), or wavelet decomposition (Simeone et al., 2020). The convolutional feature extraction method presented in this section provides advantages compared with coarse time domain features by directly measuring changes to the waveform, does not misattribute waveform variations of narrow frequency band US sensors to changes in frequency content as do the Fourier or wavelet transformations, and overcomes the problem of lateral sample point shifting of waveforms due to temperature changes as suffered by using the amplitudes at sample points as features directly (Bowler et al., 2022).

The use of all processes (honey-water blending, flour-water batter mixing, cleaning of food fouling from pipe sections, and alcoholic beer fermentation) in this work allow the method to be trialled on all possible impacting factors for a US waveform: changing acoustic impedance or attenuation that alter the magnitude of the waveform, or variations in sound velocity that alter the displacement of the waveform in the time domain (Henning and Rautenburg, 2006). This contributed to the thesis objective of evaluate US sensing and ML approaches on process monitoring applications that enable the thesis conclusions to extend to industrial scenarios. During blending, mixing, and cleaning, the acoustic impedance at the measurement areas changes as the material composition varies. Furthermore, conducting these processes over a range of temperatures alters the sound velocity through the materials. Throughout alcoholic beer fermentation, the acoustic impedance and speed of sound of a transmitted US wave change as the density of the wort decreases (Bowler et al.,

2022). Moreover, attenuation increases as CO2 bubbles are produced during the fermentation process. Finally, this article contributed to the thesis objective of comparing non-invasive reflection-mode and transmission-based US sensing by evaluating the feature extraction methodologies for both of these sensing approaches.

6.1 References

Bowler, A.L., Pound, M.P., Watson, N.J. (2022) 'A review of ultrasonic sensing and machine learning methods to monitor industrial processes' *Ultrasonics* 124. Doi:10.1016/j.ultras.2022.106776.

Caruana, R. (1997) 'Multitask Learning' *Machine Learning* 28, 41–75. Doi: https://doi.org/10.1023/A:1007379606734.

Escrig, J., Simeone, A., Woolley, E., Rangappa, S., Rady, A., Watson, N.J. (2020) 'Ultrasonic measurements and machine learning for monitoring the removal of surface fouling during clean-in-place processes' *Food Bioproducts Process*, 123, 1-13. Doi:10.1016/j.fbp.2020.05.003.

Henning, B., Rautenberg, J. (2006) 'Process monitoring using ultrasonic sensor systems' *Ultrasonics* 44, e1395–e1399. Doi:10.1016/j.ultras.2006.05.048.

Simeone, A., Woolley, E., Escrig, J., Watson, N.J. (2020) 'Intelligent industrial cleaning: a multi-sensor approach utilising machine learning-based regression' *Sensors* 20, 1–22. Doi:10.3390/s20133642.

Úbeda, M.A., Hussein, W.B., Hussein, M.A., Hinrichs, J., Becker, T.M. (2016) 'Acoustic sensing and signal processing techniques for monitoring milk fouling cleaning operations' Engineering in Life Sciences 16(1), 67–77. Doi:10.1002/elsc.201400235.

Valle, S., Li, W., Qin, S.J. (1999) 'Selection of the number of principal components: The variance of the reconstruction error criterion with a comparison to other methods' *Industrial and Engineering Chemistry Research* 38(11), 4389-4401. Doi:10.1021/ie990110i.

Wallhäußer, E., Hussein, W.B., Hussein, M.A., Hinrichs, J., Becker, T.M. (2011) 'On the usage of acoustic properties combined with an artificial neural network – a new approach of determining presence of dairy fouling' Journal of Food Engineering 103(4), 449–456. Doi:10.1016/j.jfoodeng.2010.11.015.

Wallhäußer, E., Hussein, W.B., Hussein, M.A., Hinrichs, J., Becker, T. (2013) 'Detection of dairy fouling: combining ultrasonic measurements and classification methods' Engineering in Life Sciences 13(3), 292–301. Doi:10.1002/elsc.201200081.

Zhang, Y., Qiang, Y. (2018) 'An overview of multi-task learning' *National Science Review* 5(1), 30-43. Doi:10.1093/nsr/nwx105.

Computers and Chemical Engineering 155 (2021) 107508

Contents lists available at ScienceDirect

Computers and Chemical Engineering

journal homepage: www.elsevier.com/locate/compchemeng

Convolutional feature extraction for process monitoring using ultrasonic sensors

Alexander Bowler[a], Michael Pound[b], Nicholas Watson[a,*]

[a] Food, Water, Waste Research Group, Faculty of Engineering, University of Nottingham, University Park, Nottingham, NG7 2RD, UK
[b] School of Computer Science, Jubilee Campus, University of Nottingham, Nottingham, NG8 1BB, UK

ARTICLE INFO

Article history:
Received 30 June 2021
Revised 5 August 2021
Accepted 26 August 2021
Available online 28 August 2021

Keywords:
Convolutional neural networks
Ultrasonic sensors
Transfer learning
Process monitoring
Industrial digital technologies, multi-task learning

ABSTRACT

Ultrasonic sensors are a low-cost and in-line technique and can be combined with machine learning for industrial process monitoring. However, training accurate machine learning models for process monitoring using sensor data is dependant on the feature selection methodology. This paper compares a convolutional feature extraction method to a traditional, coarse feature engineering approach. The convolutional method uses filter weights pre-trained on an auxiliary task to classify ultrasonic waveform dataset membership using previously obtained sensor data. The filter weights are used to extract features from the ultrasonic waveform. Principal component analysis is then applied to produce five principal components to be input into long short-term memory neural networks. The two approaches are compared on fermentation, mixing and cleaning datasets monitored using ultrasonic sensors. Overall, the convolutional feature method produced more informative waveform features than the coarse feature engineering approach, achieving higher model accuracy for datasets requiring substantial waveform information and for 65% of tasks overall. Multi-task learning also improved feature trajectory learning but led to reduced model accuracy for data points far from the classification decision boundaries. This can be overcome by further optimisation of neural network hyperparameters, though at increased model development time. Once trained, the convolutional feature extraction approach is a fast and convenient way of producing high quality features from ultrasonic waveforms using convolutional neural networks with little training data.

1. Introduction

The fourth industrial revolution, also termed Industry 4.0, has the potential to improve the productivity, efficiency, and sustainability of process manufacturing (Sjödin et al., 2018). This will be via the implementation of industrial digital technologies which include: The Internet of Things to enable connectivity between devices; Cloud, Fog, and Edge Computing to process large data stream (Chen and Ran, 2019; Wu et al., 2017); and machine learning (ML) to provide automatic data analysis and decision making. Industry 4.0 requires continuous data streams to enable real-time communication across processes, markets, and supply chains. Therefore, in-line and on-line sensors are a key technology in this transformation as they provide process data with no human intervention. In-line sensors directly measure the process stream while on-line sensors use automatic sampling systems (De Beer et al., 2011).

Ultrasonic (US) sensors have the benefits of being: low-cost, in-line, real-time, able to be non-invasive, small in size, low energy consuming, non-destructive, and able to characterise opaque materials. US sensors have been widely applied across manufacturing, such as fermentation (Ojha et al., 2017), polymerisation, crystallisation (Henning and Rautenberg, 2006), and food product analysis (Awad et al., 2012; Mohd Khairi et al., 2015). US sensors consist of a piezoelectric transducer which converts electrical pulses into sound waves and vice versa. Single sensors may be used in pulse-echo mode, where the sound wave is reflected back to the transducer from an interface between two neighbouring materials, or in pitch-catch mode where a second sensor receives the sound wave after it has been transmitted through a material (Awad et al., 2012). High frequency (>1 MHz), low power (<1 Wcm^{-2}) sound waves are used which do not affect the structure of the material that they pass through (Ojha et al., 2017). However, US properties are highly dependant on temperature and large changes in the acoustic impedance at a material interface (e.g. if gas bubbles are present in a liquid) causes strong reflection of the sound waves making transmission techniques difficult to use for many industrial applications (Henning and Rautenberg, 2006).

* Corresponding author at: Dr. Nicholas Watson, University of Nottingham, United Kingdom.
E-mail address: nicholas.watson@nottingham.ac.uk (N. Watson).

https://doi.org/10.1016/j.compchemeng.2021.107508
0098-1354/

Traditionally, either first principle or empirical correlations are used to determine material properties from US sensor data, or waveforms. However, first principle models soon become complex under industrial conditions, where the sound wave travels through multiple interfaces and process parameters (e.g. temperature) are changing. Similarly, empirical models require extensive calibration to account for all process parameter variations. In contrast, ML can be used to predict material properties without extensive calibration procedures by learning the relationships between these variations and the US waveform. ML also provides automatic interpretation of the sensor data. For example: training an ML model to predict the processing time remaining would enable improved batch scheduling; classifying the end of processing would reduce resource consumption; and anomaly detection methods would provide early warning of problems with batches and ensure product quality.

During training, ML models fit input data, or features, to the desired prediction outputs. The success of the ML models is partly dependant on the choice of features used for the model. For ultrasonic techniques, the speed of sound is commonly used as a feature as it is dependant on the density and compressibility of the material it passes through and is calculated by measuring the sound wave time of flight and distance travelled (Utomo et al., 2001; Utomo et al., 2002; Supardan et al., 2003; Sun et al., 2005). The changing amplitude between consecutively acquired waveforms can be used as a feature to identify process states and has been applied to determine flow regimes (Ren et al., 2021; Abbagoni and Yeung, 2016). Other process information can also be used to aid the prediction accuracy of the ML model, such as the temperature, material composition and concentration (Sun et al., 2005), or mass flow rate (Wallhäußer et al., 2014). Along with these features, measurements that describe the oscillations of the waveform are also required. The energy of the waveform (the sum of the squared amplitudes at each point in the waveform) may be used to monitor attenuation of the sound wave as it passes through a material (Utomo et al., 2001; Utomo et al., 2002; Supardan et al., 2003; Sun et al., 2005) or to monitor a change in acoustic impedance by measuring the proportion of the sound wave reflected from a material boundary (Wallhäußer et al., 2013; Wallhäußer et al., 2014; Figueiredo et al., 2016). However, the energy may not account for all the changes to the waveform, as some peaks may increase in amplitude while others decrease, or the waveform could be composed of multiple overlapping sound waves. Further features can be extracted which describe the shape of the waveform by monitoring information such as maximum amplitudes, variance in the amplitudes, the rising and falling slopes of the waveform, the duration of the waveform, and the relationship between all of these (Wallhäußer et al., 2013; Wallhäußer et al., 2014; Cau et al., 2005). Nevertheless, this is still a coarse method of monitoring waveform changes, which are indirectly measured rather than directly identified. Signal features similar to those previously listed can also be extracted in the frequency domain, commonly after using the discrete wavelet transform (Cau et al., 2005; Simeone et al., 2020). However, US transducers used for material characterisation typically have narrow frequency bands. Therefore, areas where the waveform changes or overlaps may be mis-identified as frequency changes. The amplitudes at each sample point in the time domain waveform can also be used as individual features (Escrig et al., 2020a; Escrig et al., 2020b; Munir et al., 2018). Though, should a peak translate along sample points, whether due to changes to the monitored materials or a change in temperature, the information regarding this part of the waveform is lost.

Convolution Neural Networks (CNNs) overcome these issues by using convolutional filters to measure spatial relationships in the waveform. CNNs use representation learning to automatically extract features by transforming the data into higher, more abstract levels (Lecun et al., 2015). CNNs have been used previously with US signals (Virupakshappa et al., 2018; Meng et al., 2017; Munir et al., 2019; Munir et al., 2020; Bowler et al., 2020). However, previous work has also shown that Long Short-Term Memory (LSTMs) neural network layers are required to accurately monitor time-evolving processes (Bowler et al., 2020 and 2021). LSTMs are able to retain process information from previous time-steps and are a type of recurrent neural network which uses gate units to reduce the likelihood of vanishing or exploding gradients. This enables them to be used over much longer sequences (Hochreiter and Schmidhuber, 1997). Previous time-step information could also be included in CNN inputs or even fully-connected neural networks; however, LSTMs are more memory efficient than fully connected structures and are better equipped to handle long sequences and sequences of varying length. In this work a pre-trained CNN is used to extract features from the waveform. The CNN is pre-trained on an auxiliary task using previously collected US data. The auxiliary task is to classify which dataset each US waveform belongs to. This is a transfer learning task, in which the CNN learns features of a US waveform in the auxiliary task which are then used to aid prediction on the main tasks. Augmentation of the waveforms for the auxiliary task is used to improve CNN feature learning. Furthermore, principal component analysis (PCA) is applied to these extracted features to enable the use of additional features (such as the speed of sound, changes between consecutively acquired waveforms, the process temperature, feature gradients, and time-lagged representations of waveform features) and to reduce the dimensionality of the extracted features to improve LSTM unit training accuracy and stability. The extracted Principal Components (PCs) and additional features are used as input features to the LSTM models. The novelty of this work can be summarised as: the use of CNN extracted features from US waveforms used as inputs to LSTM models, the pre-training of a CNN on an auxiliary task to identify features in US waveforms, using previously collected US datasets to improve ML model prediction through transfer learning, and applying PCA to CNN extracted US features to enable the use of additional features. The convolutional feature extraction method is compared to traditional, coarse features extracted from the time-domain waveform, such as the waveform energy, peak-to-peak amplitude or sample point position of the maximum peak. The benefits of another type of transfer learning, multi-task learning, to tasks which require multiple outputs is also evaluated throughout. The feature extraction and ML methods are compared on previously collected fermentation, cleaning, and mixing process monitoring tasks to provide a comprehensive evaluation of their advantages.

2. Method

2.1. Ultrasonic data collection

For all experiments, a US box (Lecoeur Electronique) was used to excite the transducers and digitise the received sound waves. The temperature sensors were connected to a PT-104 Data Logger (Pico Technology). The US box and temperature data logger were connected to a laptop and a bespoke MATLAB software controlled the hardware components and acquired the data.

2.1.1. Beer fermentation

Full experimental details are provided in Bowler et al. (2021). The fermentation batches were conducted in a 30 l cylindrical plastic vessel. A US probe consisting of a US transducer (Sonatest, 2 MHz central frequency) and a temperature sensor (RTD, PT1000) was installed into the vessel wall. A Tilt hydrometer provided real-time density measurements of the wort. 1.5 kg of malt (Coopers

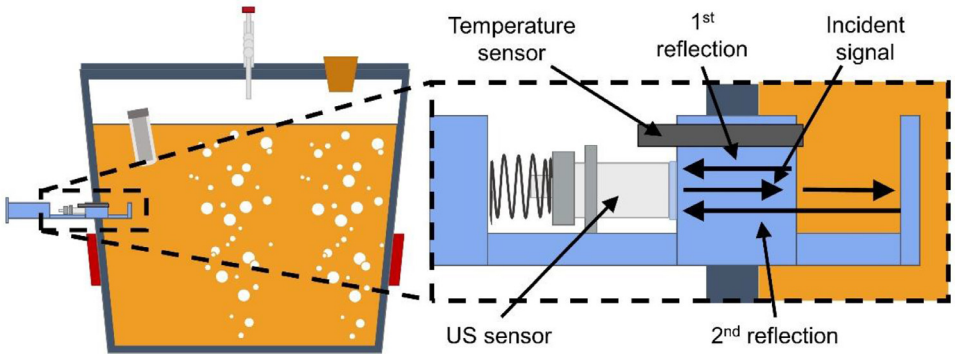


Fig. 1. The experimental apparatus and path of the received US sound wave reflections. Adapted from Bowler et al. (2021).

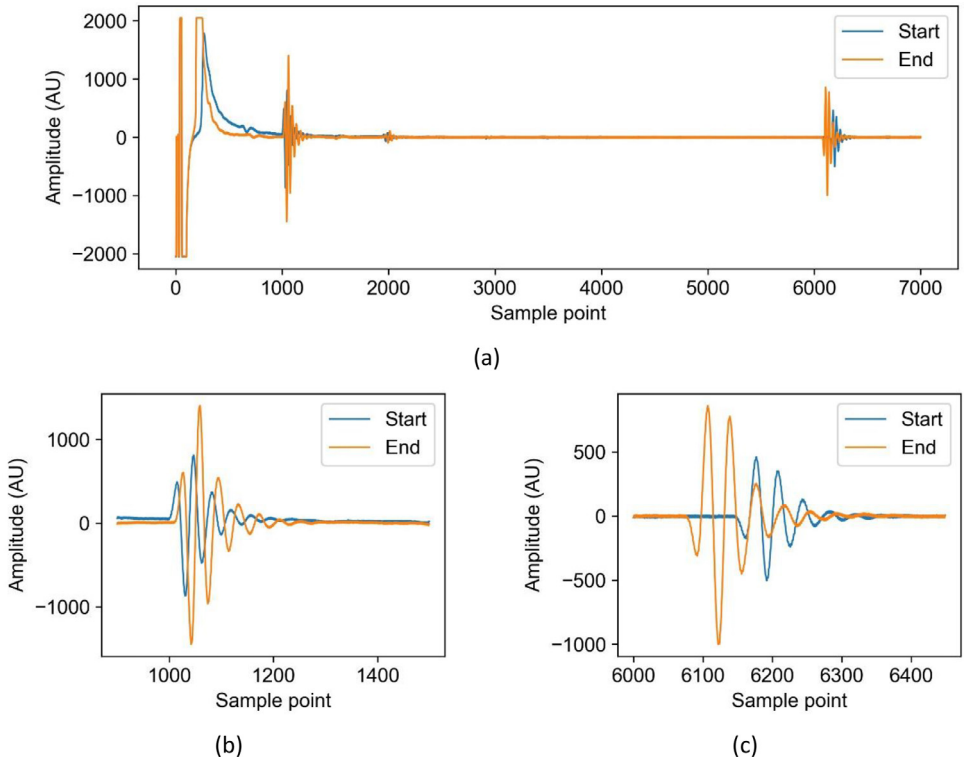


Fig. 2. (a) Example US waveforms obtained for the start and end of a fermentation batch. (b) The first reflection, located between sample points 900 and 1400. (c) The second reflection, located between sample points 6000 and 6500.

Real Ale), 1 kg of brewing sugar (The Home Brew Shop) and yeast (Coopers Real Ale) were used. In total, 13 batches were completed with the fermentation lasting between 4 and 7 days. The US waveform consisted of two sound wave reflections: the first from the interface between the probe material and the wort, and the second being transmitted through the wort and reflecting from the far probe interface (Fig. 1 and Fig. 2). The US and temperature data were collected periodically. Each set of collected data consisted of 36 US waveforms and temperature readings. The US waveforms were averaged for each set to minimise noise disturbance. Between the collection of each set of data, 200 s elapsed.

2.1.2. Cleaning of pipe fouling

Full experimental details are provided in Escrig et al. (2019, 2020a, and2020b). Three pipe test sections were used: A rectangular rig with a SS340 base plate and clear, PMMA sides; a circular pipe section constructed from clear PMMA; and an opaque, circular pipe section constructed from SS316. Three different food materials were used to foul the pipe test sections: tomato paste, concentrated malt, and gravy. The fouling material was spread onto the pipes and allowed to dry. It was placed in the centre of the base plate for the rectangular rig and 30 mm from the exit for the circular pipes (Fig. 3). The temperature of the water used for cleaning was set at either 12 °C or 45 °C and a flowrate of 6 l/s was used. For the rectangular test section, a magnetic sensor (5 MHz resonance, M1057, Olympus) was externally attached to the base plate. For the circular pipe sections, the US transducers (2 MHz, Yushi, 2P10N) were glued externally to the bottom of the pipes in the location where the fouling material would be placed. The temperature sensors were attached at the same locations. A camera was used to determine the time at which all the fouling material was removed. The position of the camera was moved depending on whether the pipe section was clear or opaque. The US and temperature data was recorded every 4 s producing 4 waveforms which were averaged. A reflection-mode, pulse-echo sensing technique was used to monitor the waveform reflected from the interface between the pipe wall and the fouling material (Fig. 4). The camera images were recorded every 20 s. A minimum of 7 repeats

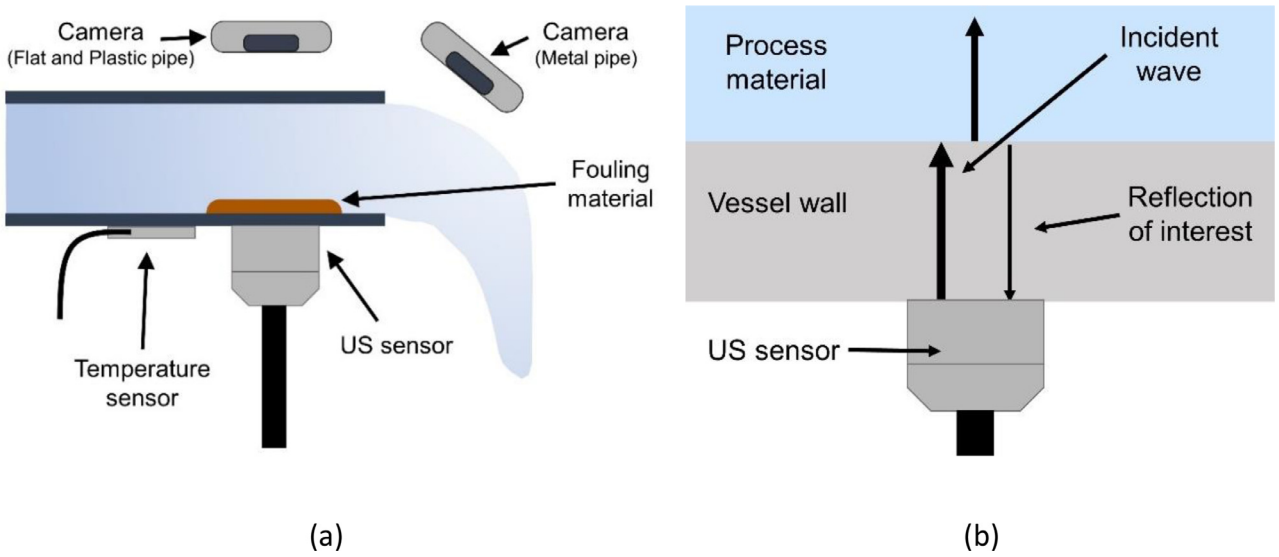


Fig. 3. (a) The experimental apparatus including the positions of the pipe section, US sensor, temperature sensor, and fouling material. (b) The paths of the received US reflections.

were conducted for every permutation of pipe test section, fouling material and fluid temperature, producing 93 runs in total.

2.1.3. Honey-water mixing

Full experimental details are provided in Bowler et al. (2020). Two US sensors (5 MHz resonance, M1057, Olympus) were externally attached to the base of a 250 ml glass mixing vessel (Fig. 5). An overhead stirrer was used to stir the mixture. One sensor (the central sensor) was attached in the centre of the vessel base. Another sensor (the non-central sensor) was attached approximately 2 cm offset from the centre. The temperature sensor was also attached to the base of the vessel. A reflection-mode, pulse-echo sensing technique was used to monitor the sound wave reflected from the interface between the vessel wall and the mixture. US signals were acquired continuously for 1 s for each probe consecutively. On average, this acquired two US waveforms which were then averaged to minimise noise disturbance. Two different volumes of pure clear honey (Wm Morrison Supermarkets plc) were used: 20 and 30 ml. 200 ml of tap water was used for all runs. The impeller speed was set to either 200 or 250 rpm. These four parameter permutations were repeated three times whilst varying the environmental temperature, producing a set of 12 runs. This methodology was repeated across two days, producing two datasets. Between, the US sensors were removed and reattached. The ground truth was obtained using a video camera to determine the time for complete mixing.

2.1.4. Batter mixing dataset

Full experimental details are provided in Bowler et al. (2020). Two US sensors (5 MHz resonance, M1057, Olympus) were externally attached to a stand mixer glass mixing bowl (1000 W Kenwood kmix kmx754). The temperature sensor was also attached to the outside of the mixing bowl. A reflection-mode, pulse-echo sensing technique monitored the sound wave reflected from the interface between the mixing bowl and the mixture (Fig. 5). US signals were continuously acquired for 1 s for each probe consecutively. On average, this produced 2 waveforms which were averaged to minimise disturbance from signal noise. The quantity of strong white flour (Wm Morrison Supermarkets plc) and tap water used was varied. A total of 9 runs were monitored. The optimal mixing time was obtained by determining the time of maximum power input to the impeller. This was measured using a YouThink plug socket power meter.

2.1.5. Feature extraction

Two feature extraction methodologies were compared: extracting coarse, time-domain signal features (Coarse method) and convolutional feature extraction using a CNN pre-trained on an auxiliary task (Convolutional method). The Coarse features method obtains coarse information about the changing waveform oscillations compared with the Convolutional method which can identify changing amplitudes at individual sample points in the waveform. The Coarse features method is designated as the next best approach (as justified in Section 1) and was the method used in Bowler et al. (2021). Therefore, a comparison between these two methods will evaluate the advantage of using convolutional feature extraction.

2.1.6. Coarse feature extraction

In total, 10 signal features were extracted from the waveform. The sum absolute amplitude (SAA), energy, sum root amplitude (SRA), standard deviation, skewness and kurtosis Eqs. (1)–(7) provide measurements of the distribution of amplitudes within the waveform. In addition, the amplitude and position of the maximum and minimum peaks were used as features to monitor the largest peaks in the waveform.

$$SAA = \sum_{i=1}^{i=SP} |A_i| \tag{1}$$

Where *SAA* is the sum absolute amplitude, *SP* is the number of sample points in the waveform, *A* is the waveform amplitude at sample point *i* (Zhan et al., 2015).

$$E = \sum_{i=1}^{i=SP} {A_i}^2 \tag{2}$$

Where *E* is the waveform energy (Zhan et al., 2015).

$$SRA = \sum_{i=1}^{i=SP} \sqrt{|A_i|} \tag{3}$$

Where *SRA* is the sum root amplitude (Zhan et al., 2015).

$$\mu = \frac{\sum_{i=1}^{i=SP} A_i}{SP} \tag{4}$$

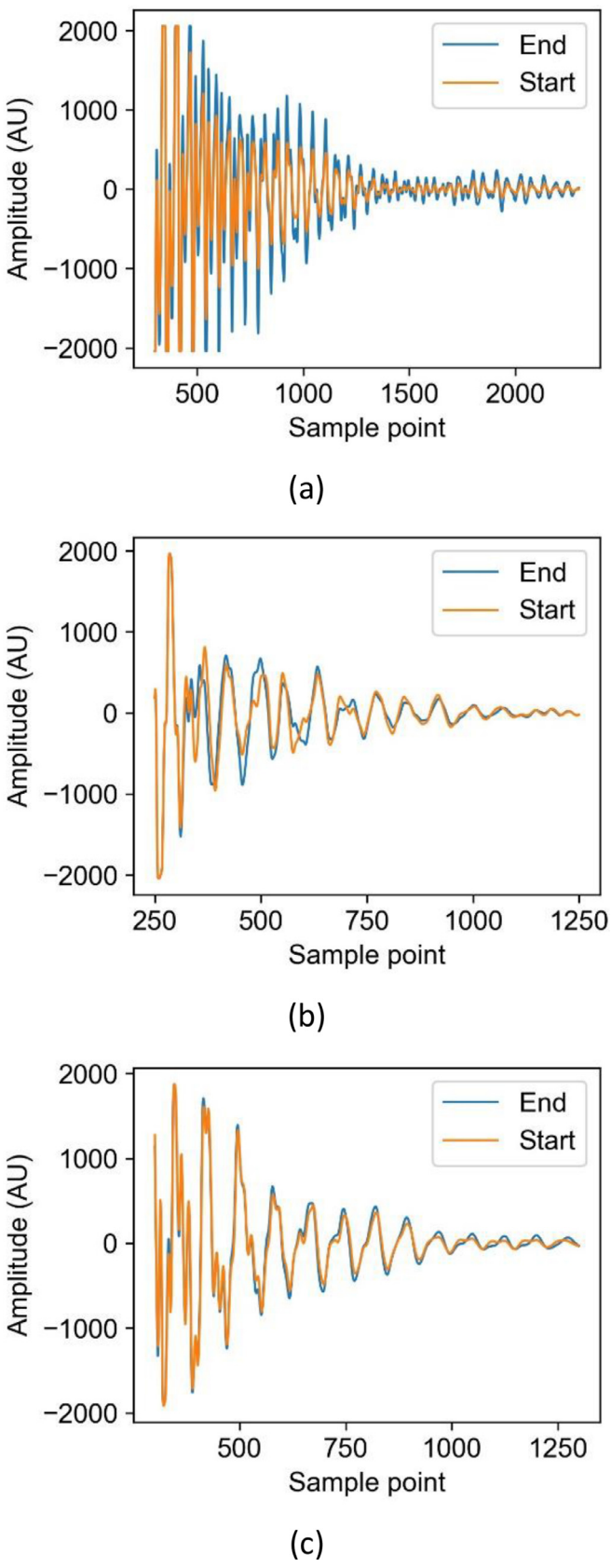


Fig. 4. The received US waveform at the start and end of the cleaning process for the (a) Flat, (b) Plastic, and (c) Metal pipe sections.

$$SD = \sqrt{\frac{1}{SP}\sum_{i=1}^{i=SP}(A_i - \mu)^2} \tag{5}$$

Where μ is the mean amplitude of the waveform, and SD is the standard deviation (Zhan et al., 2015).

$$S = \frac{\sum_{i=1}^{i=SP}(A_i - \mu)^3}{SP \times STD^3} \tag{6}$$

Where S is the waveform skewness (Caesarendra and Tjahjowidodo, 2017).

$$K = \frac{\sum_{i=1}^{i=SP}(A_i - \mu)^4}{SP \times STD^4} \tag{7}$$

Where K is the waveform kurtosis (Caesarendra and Tjahjowidodo, 2017).

2.1.7. Convolutional feature extraction

Previous work has determined LSTM layers are required for accurate time-series process monitoring. Training a convolutional neural network to the target data without an LSTM layer to obtain pre-trained convolutional filter weights would be a sub-optimal task due to the LSTM layer's ability to learn the important process feature trajectories. Therefore, the input waveforms would not be able to fit to the target data optimally without an LSTM layer and informative waveform features would not be learned (Bowler et al., 2020 and (2021). Training convolutional and LSTM layers simultaneously would also be a difficult task especially with long time sequences and limited training data used in the present case studies. This is because the many weights present in the convolutional filters and LSTM units would compete during the training process and likely fail to fit to the target data or make the training unstable. As such, to easily train convolutional layers that extract informative ultrasonic waveform features, this work trained a 1D CNN on an auxiliary task to predict waveform dataset membership. Table 1 summarises the 11 waveform datasets used. Segments of 1000 samples points in length were taken from each waveform. The position of the 1000 sample point length window was chosen for each waveform by investigating the difference between the start and end of the corresponding process. The areas with the largest visual change throughout the process were used. To increase the training set size for the network, and to improve meaningful feature extraction in the convolutional layers, a 600 × 1 input to the CNN was used. Data augmentation using a sliding window, laterally translated by 100 sample points each time, produced five waveform segments of 600 sample points in length. Further data augmentation through separate normalisation of each waveform segment was used to differentially magnify the waveform. This ensures that the network learns features specific to each waveform, rather than the position or magnitude of features.

A summary of the 1D CNN trained is presented in Table 2 which also presents CNN structures used in other previous works as a comparison. It should be noted that optimal CNN architectures are task-specific and should be chosen through validation procedures. CNN architectures for US sensor signals are included as a literature review for the interested reader. A grid search was used to select the learning rate, batch size and number of neurons in the fully connected layer. No padding was used. Training was performed with the Adam optimiser. The minimum number of neurons in the fully connected layer to achieve 100% accuracy for the dataset membership prediction was used to ensure feature identification in the convolutional layers rather than the fully connected layer. The designated training and validation sets for all datasets were used. A training accuracy of 100% was achievable after only 3

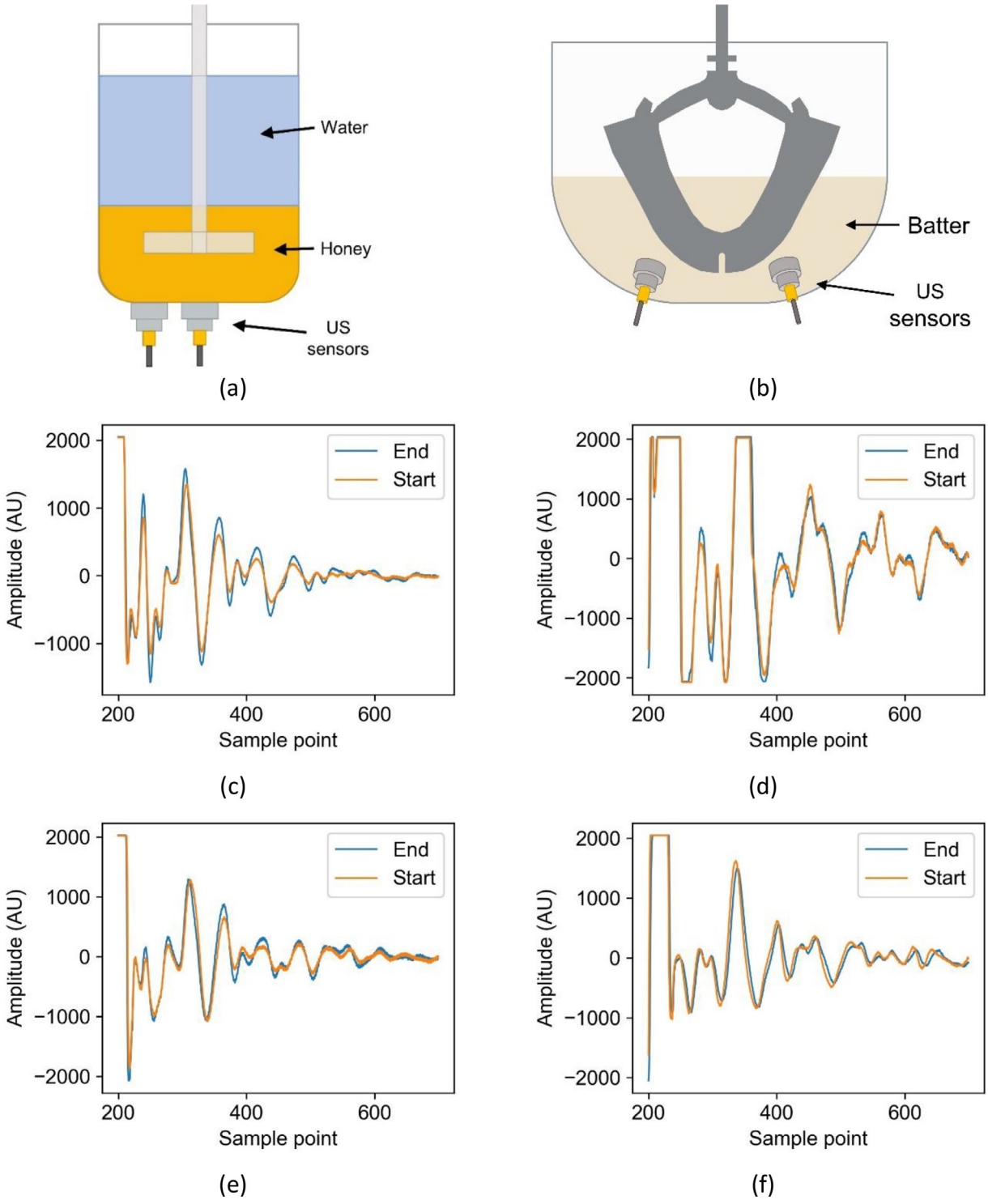


Fig. 5. The experimental apparatus for (a) the honey-water mixing experiments and (b) the flour-water batter mixing experiments. The received US waveforms reflections for (c) honey-water mixing probe 1, (d) batter mixing probe 1, (e) honey-water mixing probe 2, and (f) batter mixing probe 2.

epochs, highlighting the rapidity in developing our proposed convolutional feature extraction methodology. The pre-trained convolutional weights were then used to extract features on the full-size waveform for each dataset.

To reduce the dimensionality of the data, minimise non-useful information input into the network, aid LSTM unit training accuracy and stability, and enable the use of additional features such as the US time of flight and standard deviation between consecutive waveforms, PCA was applied to the waveform features extracted using the pre-trained convolutional filter weights. PCA extracts a set of orthogonal principal components (PCs) which are a combination of the co-linear original features (Abdi and Williams, 2010). Alternatively, a CNN feature extractor structure with more down-sampling or additional layers to reduce the number features extracted could have been used. However, preliminary investigations showed this method produced features too specific to the auxiliary training task. Furthermore, an autoencoder could have been used to learn non-linear feature relationships compared to the linear re-

Table 1
A summary of the datasets used to train the convolutional feature extractor on the auxiliary task and also evaluate the performance of the proposed feature extraction methodology.

Experimental dataset	ML task	Waveforms for CNN auxiliary task	Total number of runs (train/ validation/ test split)	Maximum sequence length
Fermentation	• Regression to predict alcohol concentration	Reflection 1 Reflection 2	13 (9/2/2)	3112
Cleaning of food fouling from pipe sections	• Classify the end of cleaning • Regression to predict cleaning time remaining	Flat rig Circular, plastic Circular, metal	35 (25/5/5) 30 (20/5/5) 28 (20/4/4)	400 300 200
Honey-water mixing 1	• Classify the end of mixing • Regression to predict mixing time remaining	Central sensor Non-central sensor	12 (8/2/2)	165
Honey-water mixing 2	• Classify the end of mixing • Regression to predict mixing time remaining	Central sensor Non-central sensor	12 (8/2/2)	123
Batter mixing	• Classify the end of mixing • Regression to predict mixing time remaining	Sensor 1 Sensor 2	9 (5/2/2)	153

Table 2
A summary of the feature extraction layers of the proposed convolutional neural network and a comparison with the other 1D CNN structure present in the literature for US sensor data. .

Layer	Proposed network	Virupakshappa et al., 2018	Meng et al., 2017	Munir et al., 2019	Munir et al., 2020
1	1D Convolutional layer 7 × 1 filter size 16 filters	1D Convolutional layer 5 × 1 filter size 5 filters	2D Convolutional layer 7 × 5 filter size 16 filters	1D Convolutional layer 16 × 1 filter size 32 filters 8 × 1 stride	1D Convolutional layer 25 × 1 filter size 32 filters 8 × 1 stride
2	Max Pooling layer 2 × 1 pool size	Max Pooling layer 2 × 1 pool size	Max Pooling layer 2 × 2 pool size	1D Convolutional layer 3 × 1 filter size 64 filters 2 × 1 stride	1D Convolutional layer 3 × 1 filter size 64 filters 2 × 1 stride
3	1D Convolutional layer 5 × 1 filter size 32 filters	1D Convolutional layer 8 × 1 filter size 8 filters	2D Convolutional layer 5 × 3 filter size 32 filters	Max Pooling layer 2 × 1 pool size 2 × 1 stride	Max Pooling layer 2 × 1 pool size 2 × 1 stride
4	Max Pooling layer 2 × 1 pool size	Max Pooling layer 2 × 1 pool size	Max Pooling layer 2 × 2 pool size	-	-
5	-	1D Convolutional layer 7 × 1 filter size 16 filters	-	-	-
6	-	Max Pooling layer 2 × 1 pool size	-	-	-

Table 3
A summary of the distribution of the explained variance by each PC for the US waveform datasets after convolutional feature extraction.

Experimental dataset	Waveforms	Number of PCs to explain 95% of variability	Variability explained by 1st PC (%)	Variability explained by 2nd PC (%)	Variability explained by 3rd PC (%)	Variability explained by 4th PC (%)	Variability explained by 5th PC (%)
Fermentation	Reflection 1	9	56.4	23.1	9.2	2.1	1.5
	Reflection 2	18	30.4	21.6	14.9	9	6.1
Cleaning of food fouling from pipe sections	Flat rig	15	60.4	15.2	7.4	4.3	1.8
	Circular, plastic	8	56.7	14.3	12.4	6.3	1.9
	Circular, metal	32	50.9	12.3	8.5	4.6	3.7
Honey-water mixing 1	Central sensor	24	52.1	18.8	7.5	4.6	2.3
	Non-central sensor	41	51.4	17	4.7	3.8	2.8
Honey-water mixing 2	Central sensor	19	38.6	30.8	12.4	4.1	2.9
	Non-central sensor	25	41.6	36.8	4.6	2.7	2.4
Batter mixing	Sensor 1	42	49.1	15.1	14.3	4.5	2.6
	Sensor 2	16	60.5	16.3	7.5	3.1	1.5

lationships assumed using PCA. However, as outlined in Section 1, the convolutional feature extraction methodology only needs to overcome a possible translation in waveform peaks by measuring spatial relationships between sample point amplitudes. Therefore, compared with autoencoders, owing to the sufficient feature extraction capability, elimination of hyperparameter optimisation, model training and convenient selection of the number of features extracted, PCA was identified as the optimal methodology. Table 3 includes the percentage variability explained by each PC for the US waveform datasets and the number of PCs required to explain 95% of the variability. The first PC likely follows the common waveform changes across the full dataset caused by variations in the US properties of the materials being monitored (either due to changing composition or process temperature). Successive PCs will identify waveform changes more specific to each batch, most likely due to the different process temperatures. Therefore, it is anticipated

that only a small number of PCs are required (i.e. greater than one) to monitor the changing material composition and account for changes in the monitoring US waveform at different temperatures. This is supported by Table 3 where the percentage variability explained drops off after the first two PCs. As shown in Table 3, the smallest number of PCs required to explain 95% of the variability in the dataset, a common method for selecting the number of PCs to use, is eight for the Plastic Cleaning dataset and nine for fermentation monitoring using only the first reflection. Therefore, using these two pieces of guidance (the primacy of the first and second PCs and the smallest number of PCs to explain 95% of dataset variability), five PCs were selected to obtain useful waveform information while minimising noise. The PCs were also combined with the standard deviation of the energy between consecutive waveforms in an acquisition block (where the number of waveforms was greater than two) to provide a measure of material differences between consecutive waveform acquisitions (Eq. (8)). In the case of the fermentation dataset using both the first and second waveform reflections, the sound wave time of flight was also added. The time of flight was calculated using a thresholding method, identifying the sample point where the waveform rises above the signal noise.

$$ESD = \sqrt{\frac{1}{W}\sum_{i=1}^{i5=W}\left(E_i - \bar{E}\right)^2} \tag{8}$$

Where *ESD* is the standard deviation in the energy of the waveforms in the acquired block, and *W* is the number of waveforms in the acquired block.

2.1.8. Model training and testing

Neural networks consisting of an LSTM layer followed by a fully-connected layer were used for all ML tasks. A fully-connected layer allows for the creation of modified features which better match the prediction task output while the LSTM layer learns the trajectories of the input features. The input features were normalised and zero-padding at the start extended the sequence lengths to that of the maximum. A masking layer specified the LSTM to disregard the zero-padding. The Adam optimisation algorithm and a gradient norm clipping value of 1 was used. A single-fold validation procedure determined the learning rate, number of LSTM units, dropout probability, L2 regularisation penalty, number of neurons in the fully-connected layer, and batch size. As many tasks and hyperparameters were investigated, only a single validation set was used to reduce the training time required. The optimal set of hyperparameters were used to retrain a model using all of the training data. The LSTMs were trained using TensorFlow 2.3.0. The coefficient of determination (R^2), mean squared error (MSE), and mean absolute error (MAE) were used as performance metrics to evaluate the regression ML models. The accuracy, precision, and recall were used to evaluate the classification models. Evaluation of multiple performance metrics allow for improved comparison between models. Multi-task learning was also investigated to aid LSTM learning of the process trajectory (Fig. 6). By training on two correlated tasks (in this case, both the classification and regression tasks for the mixing and cleaning datasets), the shared LSTM layer may learn more effective feature trajectories while reducing redundant information being stored (Li et al., 2016). This may have two benefits. The first being increased model accuracy through global learning of feature trajectories important to the process being monitored. The second being more stable model training by optimising for two combined losses. To reduce the model validation time, the number of neurons in the fully-connected layers and the dropout rate for the task-specific branches of the neural networks were fixed as the optimal hyperparameters determined from the single-task learning networks. Alternatively, a shared fully connected layer could have been used for the multi-task networks. However, to provide easier evaluation of multi-task learning utility compared with the single task learning networks, only the LSTM layer was shared. This allows for task specific feature combinations to be learned in the fully connected layers. A single-fold validation procedure optimised the number of LSTM units, dropout rate, L2 regularisation parameter, learning rate, batch size, and weighting of the individual classification and regression losses. A coarse grid search optimised the loss weighting by monitoring the unweighted classification and regression losses individually, followed by a fine grid search which optimised by monitoring the combined loss.

3. Results and discussion

To highlight the differences between the features extracted by the two methodologies, Fig. 7 displays the Coarse features (Fig. 7a) and the Convolutional PCs (Fig. 7b) for the first batch of the Flat Cleaning experiments. In Fig. 7a it is shown that the Energy, Sum Absolute Amplitude (SAA), Sum Root Ampltiude (SRA), Kurtosis, and standard deviation (STD) all follow similar trends. In contrast, the convolutionally extracted PCs follow different trajectories, highlighting the additional waveform information presented to the ML models through use of the Convolutional approach.

Overall, the Convolutional method was more accurate for over half of the tasks evaluated. For the fermentation datasets (Fig. 8), the Convolutional approach achieved lower accuracies than the Coarse feature method. In contrast, the Convolutional method proved more accurate for all cleaning tasks (Fig. 9.) and flour-water batter mixing (Fig. 10). However, the results were mixed for the honey-water mixing datasets (Fig. 11) with a Convolutional based approach scoring the highest for three tasks compared with five using the Coarse features method. Table 4 compares the results from this work with previous published works using these datasets. It should be noted that training, validation, and test sets, along with validation and testing procedure, differ between the previous published results and the current work. As such, the accuracies are not directly comparable. In practice, optimising for the number of PCs, employing k-fold cross validation, and possibly using past process information (in the form of feature gradients, time-lagged feature representations, or the time since the beginning of the process) would improve model accuracy on the test set data. However, this is not necessary in the current work in which the aim is to present the superior feature extraction ability of the Convolutional method compared with the Coarse features. Interestingly, for the datasets where the Convolutional method was more accurate than the Coarse method, cleaning and flour-water batter mixing, high accuracy was achieved in previous works using complex feature extraction methodologies. For example, Bowler et al. (2020) achieved 92.5% accuracy in classifying the end of flour-water batter mixing through using a CNN training on the continuous wavelet decomposition of the waveform. Escrig et al. (2020a, 2020b) used a K-best predictors method to selected the 200 most informative sample points in the waveform to predict the end point of pipe section cleaning. Furthermore, no LSTM layers or past process information (e.g. features gradients or time-lagged features) were required for these tasks. Contrastingly, for the datasets where the Coarse method was more accurate than the Convolutional method, honey-water mixing and fermentation, previous work suggests that using past process information as features was vital for high model accuracy but not complex feature extraction methodologies.

The increased accuracy of the Convolutional feature method for tasks that require a lot of waveform information in the previous works, namely; cleaning and flour-water batter mixing, shows that this method is capable of extracting more usable information from the waveform. As such, this proves that the Convolutional method is a superior feature extractor to using Coarse features. Resultantly,

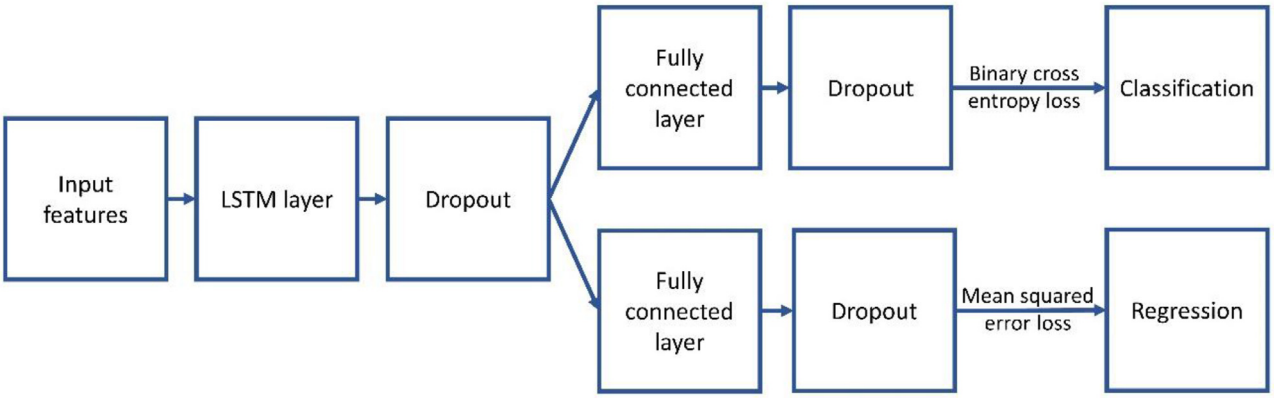


Fig. 6. The structure of the multi-task learning network evaluated using the cleaning and mixing datasets. The cleaning and mixing datasets were used as both entail classification and regression tasks.

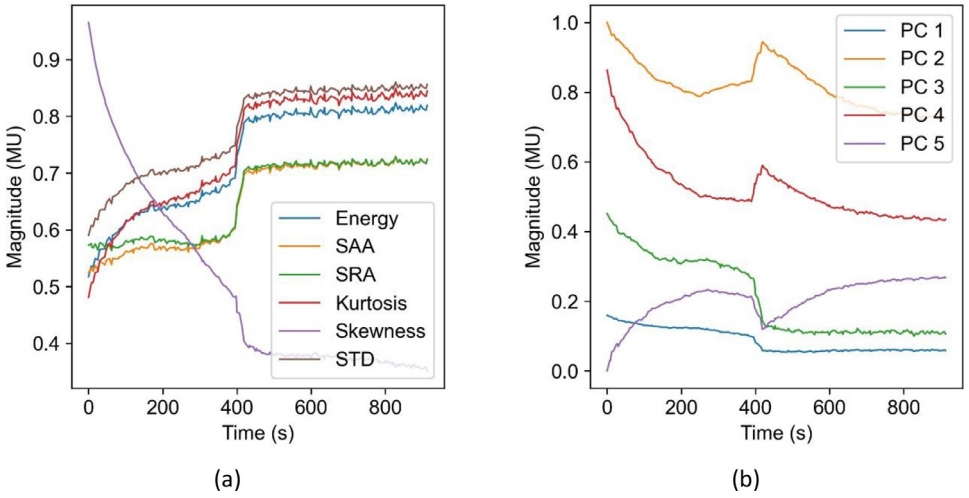


Fig. 7. A comparison between (a) the Coarse features and (b) the Convolutional extracted features for the first batch of the Flat Cleaning experiments. The end of cleaning was identified using the camera at 425 s. Note the similar process trajectories of the Energy, Sum Squared Amplitude (SAA), Sum Root Amplitude (SRA), Kurtosis, and Standard Deviation (STD). In contrast, the five convolutionally extracted principal components show differing trajectories, making additional US waveform information more accessible to the ML models.

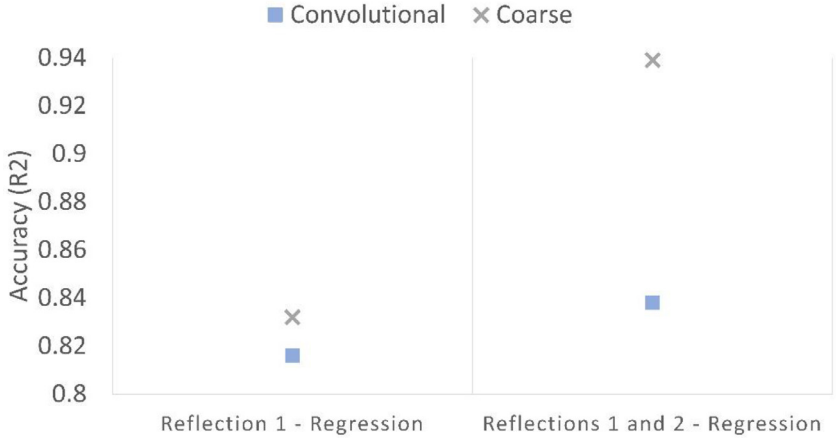


Fig. 8. The R^2 scores for the feature extraction methodologies applied to the fermentation dataset.

the lower accuracy of the fermentation and honey-water mixing results indicates that the Convolutional feature method degraded the feature trajectory learning of the LSTM layer. There are several reasons why this may be the case. Firstly, the more complicated trajectories of the PCs could have been more difficult for the LSTM layer to learn. To overcome this, the results from previous works suggest the use of feature gradients aids in LSTM layer learning of feature trajectories. Secondly, due to the increased waveform information extracted, the Convolutional method may have overfitted to the range of the training data with the testing data falling outside of the training feature ranges. This can be overcome through using a k-fold cross-validation procedure instead of the single-fold

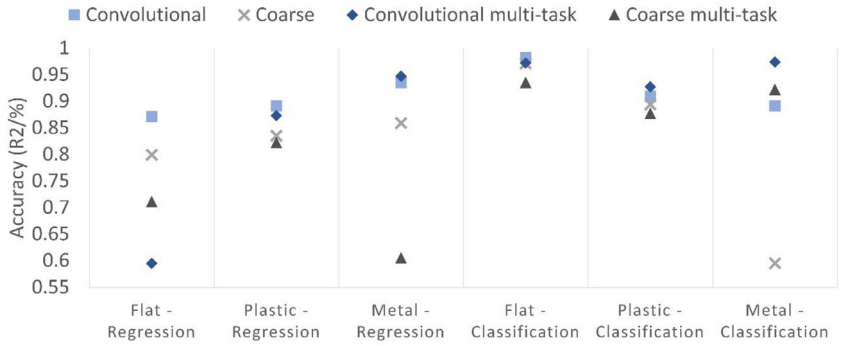


Fig. 9. The regression (R^2) and classification (% correct) accuracy for the feature extraction methodologies evaluated on the cleaning tasks. A CNN method was most accurate for every task.

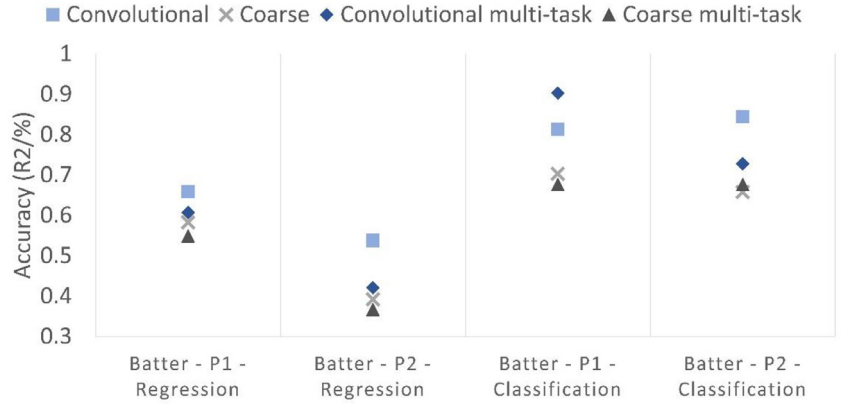


Fig. 10. The regression (R^2) and classification (% correct) accuracies for the feature extraction methodologies evaluated on the flour-water batter mixing dataset. A Convolutional method was most accurate for every task.

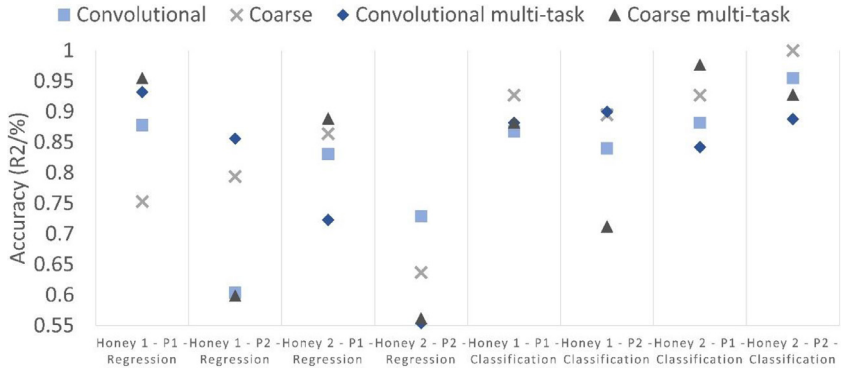


Fig. 11. The regression (R^2) and classification (% correct) accuracies for the feature extraction methodologies evaluated on the honey-water mixing datasets. P1 indicates the non-central sensors and P2 denotes the central sensors.

validation used in this study. Single-fold validation was used to reduce model development time, owing to the large number of tasks and hyperparameters evaluated. K-fold cross-validation was not required in this study, where the aim was to showcase the superior feature extraction capability of the Convolutional method, as has been presented. Thirdly, the Coarse feature method may have benefitted from the similarity in the input features. These features will most strongly follow the changes in US properties of the monitored materials, similar to the first PC extracted using the Convolutional method. Therefore, the Coarse feature method allows the LSTM layer more opportunities to learn this strong feature trend. In contrast, the Convolutional method can only learn the trajectory of the first principal component through a single path in the network. It is anticipated that, again, k-fold cross-validation would strengthen the impact of the first PC relative to the subsequent, less informative PCs.

The results for the multi-task learning neural networks were mixed. Overall, multi-task learning performed worse for 23 out of 36 tasks compared with the single task learning counterparts. The reason for this likely that the networks failed to optimise for both tasks but instead generalised across them. The hyperparameters for the task-specific branches of the multi-task neural networks were fixed as the optimal values from the respective optimised single-task networks. Optimising for these hyperparameters as well may improve multi-task learning accuracy though requires longer development time. As such, the decision to use multi-task learning should be made during the task validation stage. However, multi-task learning showed more benefits to the classification

Table 4
A comparison of the presented convolutional feature extraction method to previously published ML results obtained using the same datasets.

Previous works	Task	ML accuracy (R^2 /%)	Presented convolutional feature extraction method accuracy (R^2 /%)	Differences in previous works methodology	Conclusions
Bowler et al. (2021)	Fermentation monitoring Regression	0.952 – 1st and 2nd reflections 0.948 – 1st reflection	0.816 – 1st and 2nd reflections 0.838 – 1st reflection	Feature gradients used as features	The results indicate that the addition of time-lagged feature representations improves LSTM model training
Bowler et al. (2020)	Honey-water mixing Classification	96.3% central sensor 89.8% non-central sensor	95.5% central sensor 88.2% non-central sensor		
	Honey-water mixing Regression	0.960 central sensor 0.965 non-central sensor	0.856 central sensor 0.932 non-central sensor		
	Flour-water batter mixing Regression	0.976	0.659		
	Flour-water batter mixing Classification	92.50%	90.30%	Wavelet analysis used	The results indicate a greater number of PCs may improve model accuracy
Escrig et al. (2020a)	Cleaning Flat pipe Classification	Up to 99%	98.20%	200 waveform sample points used as features, selected	
Escrig et al. (2020b)	Cleaning Plastic and Metal pipes Classification	Up to 100%	92.7% - Plastic pipe 97.4% - Metal pipe	though K-best predictors	
Simeone et al., 2020	Cleaning Flat pipe Regression	0.955	0.871	US sensor data combined with optical sensor data	The image analysis allowed for early monitoring of the cleaning process

tasks compared with regression, providing improved accuracy for 8 out of 18 tasks. This is likely because the regression part of the network aids in identifying the approximate position of the classification decision boundary for the classification branch to optimise. This indicates that the regression results for the multi-task learning networks may be improved around the classification decision boundary but failed to learn feature trajectories far from this point. Multi-task learning showed more benefits in the regression tasks for the honey-water mixing experiments, achieving higher accuracy for half of the tasks. As the results from previous works show that learning feature trajectories is vital for these tasks, this indicates that multi-task learning may allow improved feature trend learning in the LSTM layer. This is further supported by multi-task learning proving more benefits for the Convolution method, achieving higher accuracies for 7 out of 18 tasks compared with 5 for the Coarse method. As feature trajectory learning is more difficult using the Convolutional method without feature gradients, this indicates that multi-task learning could alleviate this problem.

3.1. Advantages of the convolutional feature extraction method

The Convolutional feature extraction method presented in this work and evaluated on time-series data also has benefits for non-time series data. Firstly, it obtains informative convolutional filter weights from an easier task to be used for a more difficult desired task as either a feature extraction method or as starting points for weight fine-tuning. Data augmentation and the minimisation of the number of neurons in the fully connected layer of the auxiliary task CNN ensures useful convolutional layer feature learning. Secondly, by using the pre-trained filter weights as feature extractors rather than a starting point for fine-tuning time, model development time is saved. Thirdly, the use of PCA allows the incorporation of other features useful to process monitoring, such as the process temperature, speed of sound, standard deviation between consecutively acquired signals, feature gradients or time lagged feature representations, and other process parameters. Furthermore, the use of PCA reduces the dimensionality of the data to improve model training and amplifies the contribution the previously listed additional features.

4. Conclusion

The performance of ML models is partly dependant on the quality of features extracted from the data. This work compared two feature extraction methodologies for process monitoring using US sensor data. The Convolution feature extraction method produces more informative waveform features; however, presents a more difficult feature trajectory learning task. Multi-task learning improves process trajectory learning but regression accuracy is degraded far from the classification decision boundary. This may be overcome through more extensive hyperparameter selection though at increased model development time. Once trained, the convolutional method represents a fast and convenient way of extracting high quality US waveform features for future applications.

Declaration of Competing Interest

The authors declare no conflict of interest.

CRediT authorship contribution statement

Alexander Bowler: Conceptualization, Data curation, Formal analysis, Investigation, Methodology, Software, Validation, Visualization, Writing – original draft, Writing – review & editing. **Michael Pound:** Conceptualization, Supervision, Writing – review & editing. **Nicholas Watson:** Funding acquisition, Project administration, Resources, Supervision, Writing – review & editing.

Funding

This work was supported by the Engineering and Physical Sciences Research Council (EPSRC) standard research studentship (EP/R513283/1) and EPSRC network+ Connected Everything (EP/P001246/1).

References

Abbagoni, B.M., Yeung, H., 2016. Non-invasive classification of gas-liquid two-phase horizontal flow regimes using an ultrasonic Doppler sensor and a neural network. Meas. Sci. Technol. 27, 084002. doi:10.1088/0957-0233/27/8/084002.

Abdi, H., Williams, L.J., 2010. Principal component analysis. Wiley Interdiscip. Rev. Comput. Stat. 2, 433–459. doi:10.1002/wics.101.

Awad, T.S., Moharram, H.A., Shaltout, O.E., Asker, D., Youssef, M.M., 2012. Applications of ultrasound in analysis, processing and quality control of food: a review. Food Res. Int. 48, 410–427. doi:10.1016/j.foodres.2012.05.004.

Bowler, A.L., Bakalis, S., Watson, N.J., 2020. Monitoring mixing processes using ultrasonic sensors and machine learning. Sensors 20, 1813. doi:10.3390/s20071813.

Bowler, A.L., Escrig, J., Pound, M., Watson, N. Predicting Alcohol Concentration during Beer Fermentation Using Ultrasonic Measurements and Machine Learning. Fermentation 7, 34, 2021. 10.3390/fermentation7010034

Caesarendra, W., Tjahjowidodo, T., 2017. A Review of Feature Extraction Methods in Vibration-Based Condition Monitoring and Its Application for Degradation Trend Estimation of Low-Speed Slew Bearing. Machines 5, 21. doi:10.3390/machines5040021.

Cau, F., Fanni, A., Montisci, A., Testoni, P., Usai, M., 2005. Artificial neural networks for non-destructive evaluation with ultrasonic waves in not accessible. IEEE Ind. Applic. Soc. 1, 685–692. doi:10.1109/IAS.2005.1518382.

Chen, J., Ran, X., 2019. Deep Learning With Edge Computing: a Review. P. IEEE doi:10.1109/JPROC.2019.2921977.

De Beer, T., Burggraeve, A., Fonteyne, M., Saerens, L., Remon, J.P., Vervaet, C., 2011. Near infrared and Raman spectroscopy for the in-process monitoring of pharmaceutical production processes. Int. J. Pharm. 417, 32–47. doi:10.1016/j.ijpharm.2010.12.012.

Escrig, Escrig, .J.., Woolley, E., Rangappa, S., Simeone, A., Watson, N.J., 2019. Clean-in-place monitoring of different food fouling materials using ultrasonic measurements. Food Control 104, 358–366. doi:10.1016/j.foodcont.2019.05.013.

Escrig, J.E., Simeone, A., Woolley, E., Rangappa, S., Rady, A., Watson, N.J., 2020a. Ultrasonic measurements and machine learning for monitoring the removal of surface fouling during clean-in-place processes. Food Bioprod. Process 123, 1–13. doi:10.1016/j.fbp.2020.05.003.

Escrig, J., Woolley, E., Simeone, A., Watson, N.J., 2020b. Monitoring the cleaning of food fouling in pipes using ultrasonic measurements and machine learning. Food Control 116, 107309. doi:10.1016/j.foodcont.2020.107309.

Henning, B., Rautenberg, J., 2006. Process monitoring using ultrasonic sensor systems. Ultrasonics 44, 1395–1399. 10.1016/j.ultras.2006.05.048.

Figueiredo, M.M.F., Goncalves, J.L., Nakashima, A.M.V., Fileti, A.M.F., Carvalho, R.D.M., 2016. The use of an ultrasonic technique and neural networks for identification of the flow pattern and measurement of the gas volume fraction in multiphase flows. Exp. Therm. Fluid Sci. 70, 29–50. 10.1016/j.expthermflusci.2015.08.010.

Hochreiter, S., Schmidhuber, J., 1997. Long Short-Term Memory. Neural Comput 9, 1735–1780. doi:10.1162/neco.1997.9.8.1735.

Lecun, Y., Bengio, Y., Hinton, G., 2015. Deep learning. Nature 521, 436–444. doi:10.1038/nature14539.

Li, X., Zhao, L., Wei, L., Yang, M.-.H., Wu, F., Zhuang, Y., Ling, H., Wang, J., 2016. DeepSaliency: multi-Task Deep Neural Network Model for Salient Object Detection. IEEE T. Image. Process. 25, 3919–3930. 10.1109/TIP.2016.2579306.

Meng, M., Chua, Y.J., Wouterson, E., Ong, C.P.K., 2017. Ultrasonic signal classification and imaging system for composite materials via deep convolutional neural networks. Neurocomputing 257, 128–135. 10.1016/j.neucom.2016.11.066.

Mohd Khairi, M.T., Ibrahim, S., Md Yunus, M.A., Faramarzi, M., 2015. Contact and non-contact ultrasonic measurement in the food industry: a review. Meas. Sci. Technol. 27, 012001. doi:10.1088/0957-0233/27/1/012001.

Munir, N., Kim, H.-J., Song, S.-J., Kang, S.-S., 2018. Investigation of deep neural network with drop out for ultrasonic flaw classification in weldments. J. Mech. Sci. Technol. 32, 3073–3080. doi:10.1007/s12206-018-0610-1.

Munir, N., Kim, H.-J., Park, J., Song, S.-J., Kang, S.-S., 2019. Convolutional neural network for ultrasonic weldment flaw classification in noisy conditions. Ultrasonics 94, 74–81. doi:10.1016/j.ultras.2018.12.001.

Munir, N., Park, J., Kim, H.-J., Song, S.-J., Kang, S.-S., 2020. Performance enhancement of convolutional neural network for ultrasonic flaw classification by adopting autoencoder. NDT&E Int 111, 102218. doi:10.1016/j.ndteint.2020.102218.

Ojha, K.S., Mason, T.J., O'Donnell, C.P., Kerry, J.P., Tiwari, B.K., 2017. Ultrasound technology for food fermentation applications. Ultrason. Sonochem. 417, 32–47. doi:10.1016/j.ultsonch.2016.06.001.

Utomo, M.B., Sakai, T., Uchida, S., Maezawa, A., 2001. Simultaneous measurement of mean bubble diameter and local gas holdup using ultrasonic method with neural network. Chem. Eng. Technol. 24, 493–500. doi:10.1002/1521-4125(200105)24:5<493::AID-CEAT493>3.0.CO;2-L.

Utomo, M.B., Sakai, T., Uchida, S., 2002. Use of neural network-ultrasonic technique for measuring gas and solid hold-ups in a slurry bubble column. Chem. Eng. Technol. 25, 293–299. doi:10.1002/1521-4125(200203)25:3293::AID-CEAT2933.0.CO;2-X.

Ren, W., Jin, N., Ouyang, L., Zhai, L., Ren, Y., 2021. Gas Volume Fraction Measurement of Oil-Gas-Water Three-Phase Flows in Vertical Pipe by Combining Ultrasonic Sensor and Deep Attention Network. IEEE T. Instrum. Meas. 70, 9244102. doi:10.1109/TIM.2020.3031186.

Simeone, A., Woolley, E., Escrig, J., Watson, N.J., 2020. Intelligent industrial cleaning: a multi-sensor approach utilising machine learning-based regression. Sensors 20, 1–22. doi:10.3390/s20133642.

Sjödin, D.R., Parida, V., Leksell, M., Petrovic, A., 2018. Smart Factory Implementation and Process Innovation: a Preliminary Maturity Model for Leveraging Digitalization in Manufacturing Moving to smart factories presents specific challenges that can be addressed through a structured approach focused on people, processes, and technologies. Res. -Technol. Manag. 61, 22–31. doi:10.1080/08956308.2018.1471277.

Sun, Z., Jen, C.-.K., Yan, J., Chen, M.-.Y., 2005. Application of ultrasound and neural networks in the determination of filler dispersion during polymer extrusion processes. Polym. Eng. Sci. 45, 764–772. doi:10.1002/pen.20328.

Supardan, M.D., Maezawa, A., Uchida, S., 2003. Determination of local gas holdup and volumetric mass transfer coefficient in a bubble column by means of an ultrasonic method and neural network. Chem. Eng. Technol. 26, 1080–1083. doi:10.1002/ceat.200301752.

Virupakshappa, K., Marino, M., Oruklu, E., 2018. A Multi-Resolution Convolutional Neural Network Architecture for Ultrasonic Flaw Detection. IEEE Int. Ultra. Sym. 2018, 8579888. doi:10.1109/ULTSYM.2018.8579888.

Wallhäußer, E., Hussein, W.B., Hussein, M.A., Hinrichs, J., Becker, T., 2013. Detection of dairy fouling: combining ultrasonic measurements and classification methods. Eng. Life Sci. 13, 292–301. doi:10.1002/elsc.201200081.

Wallhäußer, E., Sayed, A., Nöbel, S., Hussein, M.A., Hinrichs, J.b,, Becker, T., 2014. Determination of cleaning end of dairy protein fouling using an online system combining ultrasonic and classification methods. Food Bioprocess Tech 7, 506–515. doi:10.1007/s11947-012-1041-0.

Wu, D., Liu, S., Zhang, L., Terpenny, J., Gao, R.X., Kurfess, T., Guzzo, J.A., 2017. A fog computing-based framework for process monitoring and prognosis in cybermanufacturing. J. Manuf. Syst. 43, 25–34. doi:10.1016/j.jmsy.2017.02.011.

Zhan, X., Jiang, S., Yang, Y., Liang, J., Shi, T., Li, X., 2015. Inline Measurement of Particle Concentrations in Multicomponent Suspensions using Ultrasonic Sensor and Least Squares Support Vector Machines. Sensors 15, 24109–24124. doi:10.3390/s150924109.

7 Transfer learning for process monitoring using reflection-mode ultrasonic sensing

Article title: Transfer learning for process monitoring using reflection-mode ultrasonic sensing

Journal: Ultrasonics

Date published: 27/07/2021

DOI: 10.1016/j.ultras.2021.106468

Authors: Bowler, A.L., Watson, N.J.

Author contributions (as published): A.L.B.: Conceptualization, Formal analysis, Funding acquisition, Methodology, Project administration, Supervision, Visualization, Writing - review & editing. N.J.W.: Data curation, Investigation, Resources, Software, Validation, Writing - review & editing.

The aim of this thesis was to develop ML methods to facilitate optimal deployment of US sensors for process monitoring applications in industrial environments. This article contributed to the thesis objectives by presenting an unlabelled domain adaptation approach to transfer a single US feature between similar mixing and similar cleaning processes. This method was compared with Transfer Component Analysis (TCA). This work and the subsequent section (Section 8, titled: “Domain Adaptation and Federated Learning for Ultrasonic Monitoring of Beer Fermentation”) comprise the portion of this thesis that explores the development of unlabelled and labelled domain adaptation methods for process monitoring using US sensors. These methods leverage previously collected datasets which would negate or reduce the burden of collecting labelled data in industrial environments.

This work investigates unlabelled domain adaptation for cleaning and mixing processes, whereas Section 8 uses labelled domain adaptation approaches for monitoring fermentation. Unlabelled domain adaptation is used to transfer ML models to a new process where no further labelled data can be collected, for example, for monitoring the internal fouling of pipe sections or monitoring mixing processes in vessels. The labelled domain adaptation methods in Section 8 are used to transfer ML models from a laboratory fermentation process to an industrial process. The domain adaptation approaches are used to minimise the data that must be collected for the industrial process. Trialling domain adaptation methods for all processes (mixing, cleaning, and fermentation) also allows these approaches to be tested across all the factor that may impact as US waveform: changing acoustic impedance or attenuation that alter the magnitude of the waveform, or variations in sound velocity that alter the displacement of the waveform in the time domain (Henning and Rautenburg, 2006).

In this work, classification ML models were trained to predict whether the mixture was non-mixed or fully mixed and whether the pipe test section is fouled or clean. Regression ML models were trained to predict the process time remaining until fully mixed or clean. For the honey-water mixing, ML models were trained on either of the two datasets (Dataset 1 or Dataset 2) and used to predict on the other dataset. This was performed for the Non-Central and Central sensors individually and then by combining data from both sensors. Therefore, an ML model is trained on a labelled mixing system and transferred to monitor a similar mixing process which has no labelled data. For the cleaning of pipe fouling, models were trained on one or several datasets and tested on another (Datasets 1-12). This is representative of training an ML model on a pipe section with labelled data available and

transferring this knowledge to an unlabelled process pipe where the pipe material, fouling material, cleaning fluid properties and US sensor could be different.

The novelty of this work was the use of unlabelled domain adaptation with US sensors for process monitoring, the use of unlabelled domain adaptation on features of US waveform opposed to the waveform itself, and, finally, the investigation of omitting the process temperature as a feature. Transfer of a single US waveform feature was found to be optimal as it was most representative of the changing material properties at the sensor measurement area. Further, training on a greater number of source datasets, or source datasets with larger variability in feature distributions, improved transfer learning.

The single feature method achieved higher prediction accuracies (96.0 % and 98.4 % accuracy and 0.947 and 0.999 R^2 values for mixing and cleaning, respectively) compared with TCA (92.6 % and 95.3 % accuracy and 0.942 and 0.966 R^2 values for mixing and cleaning, respectively). The waveform Energy was used for the single feature method as it is a measure of the total magnitude of the sound wave and therefore is representative of the acoustic impedance of the process material at the vessel wall. Other features, such as the maximum or minimum peaks, position of peaks, skewness or kurtosis, are useful to monitor changes in the waveform shape and aid identifying multiple overlapping sound waves. Therefore, the trend in these other features does not follow changes in the process material. Features similar to the waveform Energy that monitor the process material properties include the Sum Absolute Amplitude or Sum Root Amplitude, however, these give greater or lesser weighting to larger amplitudes, respectively. Therefore, the discrepancy between the Energy and either of these features would be due to the shape of the waveform rather than the changing process material. As the shape of the waveform is unlikely to follow the same trends across domains, inclusion of these features may degrade model accuracy. This effect is magnified further when using more than two domains and therefore use of a single feature is optimal. However, as described in Section 8 (3.3. Future Research Directions), this single feature method may be used to obtain predictions in the new domain, after which, these predictions can be used as a feature to be inputted into another model trained for the new domain which also uses more waveform features. In this way, the model gains the knowledge from the source domain as well as being able to use many features that describe the changing waveform.

To investigate whether additional waveform features are required to monitor the mixing and cleaning case studies, TCA was used to extract three features, or transfer components, to train the transfer learning models. TCA minimises the distance between the source and target domain feature spaces using the Maximum Mean Discrepancy and extracts transfer components that maximise variance across this shared feature space (Pan et al., 2011). A 5-level DWT decomposition was applied to the US waveforms producing six resultant waveforms (five sets of detail coefficients and one set of approximation coefficients). Six feature extraction methods were combined to these resultant waveforms along with the original US waveform (waveform energy, sum root amplitude, sum absolute amplitude, standard deviation, skewness, and kurtosis) to produce 42 features to be inputted into the TCA algorithm. Every run in the source domain dataset was used for model training and validation, and every run in the target domain dataset was used for testing. An additional model, named the Non-Transfer Learning model, was trained using only the target domain data to provide a comparative result to the transfer learning models' accuracy. This Non-Transfer Learning model provides evaluation of what a supervised learning method would achieve of the target domain if labelled data were available. It thereby provides a comparative assessment of the accuracy of the transfer learning models and is not used to draw conclusions about the relative merits of each ML methodology used in the article.

In this work, the single feature unlabelled domain adaptation method was compared to TCA where three transfer components were extracted. On reflection, for accurate comparison, this method should be compared to other unlabelled domain adaptation methods which also extract fewer than three features. For example, during model validation, the single feature method could be trialled against TCA where one or two transfer components are extracted, or other domain adaptation methods such as domain adversarial training of neural networks to produce a small number of features (Ganin et al., 2016). In this way, the maximum number of features that provide positive transfer can be identified.

A limitation of this article was the varying validation approaches utilised for each algorithm (single-fold validation with 20% of total training and validation set size for ANNs, and no validation data LSTMs). This benefits the hyperparameters selected for the ANNs as these were chosen to generalise to a validation set. In comparison, the hyperparameters selected for the LSTMs were disadvantaged owing to not being chosen to generalise to unseen data. However, despite this, LSTMs were best for 88.5 % of tasks. Therefore, the conclusions drawn from this article were not influenced by this limitation. Furthermore, the use of LSTM layers for feature trajectory learning during process monitoring were confirmed in Section 4 of this thesis.

7.1 References

Ganin, Y., Ustinova, E., Ajakan, H., Germain, P., Larochelle, H., Laviolette, F., Marchand, M., Lempitsky, V. (2016) ‘Domain-adversarial training of neural networks’ Journal of Machine Learning Research, 17, 1–35.

Henning, B., Rautenberg, J. (2006) ‘Process monitoring using ultrasonic sensor systems’ *Ultrasonics* 44, e1395–e1399. Doi:10.1016/j.ultras.2006.05.048.

Pan, S.J., Tsang, I.W., Kwok, J.T., Yang, Q. (2011) ‘Domain adaptation via transfer component analysis’ *IEEE Transactions on Neural Networks* 22(2), 199–210. Doi:10.1109/TNN.2010.2091281.

Ultrasonics 115 (2021) 106468

Contents lists available at ScienceDirect

Ultrasonics

journal homepage: www.elsevier.com/locate/ultras

Transfer learning for process monitoring using reflection-mode ultrasonic sensing

Alexander L Bowler, Nicholas J Watson *

Faculty of Engineering, University of Nottingham, University Park, Nottingham NG7 2RD, United Kingdom

ARTICLE INFO

Keywords:
Domain adaptation
Transfer learning
Ultrasonic sensors
Machine learning
Industry 4.0
Industrial Digital Technologies

ABSTRACT

The fourth industrial revolution is set to integrate entire manufacturing processes using industrial digital technologies such as the Internet of Things, Cloud Computing, and machine learning to improve process productivity, efficiency, and sustainability. Sensors collect the real-time data required to optimise manufacturing processes and are therefore a key technology in this transformation. Ultrasonic sensors have benefits of being low-cost, in-line, non-invasive, and able to operate in opaque systems. Supervised machine learning models can correlate ultrasonic sensor data to useful information about the manufacturing materials and processes. However, this requires a reference measurement of the process material to label each data point for model training. Labelled data is often difficult to obtain in factory environments, and so a method of training models without this is desirable. This work compares two domain adaptation methods to transfer models across processes, so that no labelled data is required to accurately monitor a target process. The two method compared are a Single Feature transfer learning approach and Transfer Component Analysis using three features. Ultrasonic waveforms are unique to the sensor used, attachment procedure, and contact pressure. Therefore, only a small number of transferable features are investigated. Two industrially relevant processes were used as case studies: mixing and cleaning of fouling in pipes. A reflection-mode ultrasonic sensing technique was used, which monitors the sound wave reflected from the interface between the vessel wall and process material. Overall, the Single Feature method produced the highest prediction accuracies: up to 96.0% and 98.4% to classify the completion of mixing and cleaning, respectively; and R^2 values of up to 0.947 and 0.999 to predict the time remaining until completion. These results highlight the potential of combining ultrasonic measurements with transfer learning techniques to monitor industrial processes. Although, further work is required to study various effects such as changing sensor location between source and target domains.

1. Introduction

Whilst the third industrial revolution automated individual unit operations, the fourth industrial revolution (Industry 4.0) will use Industrial Digital Technologies (IDTs) such as the Internet of Things to integrate entire manufacturing processes and Machine Learning (ML) to provide automatic decision making [39]. This has the potential to improve process productivity, raw material and energy efficiency, product quality and increase manufacturing sustainability [14]. Sensors collect the real-time data required to optimise manufacturing processes making them a key technology in this new industrial revolution. Although sensors exist for basic measurements such as temperature and pressure, there is a need for more advanced techniques that can monitor materials or processes. Active ultrasonic sensors are low-cost, small, operate non-invasively, and can characterise opaque systems. Furthermore, they are in-line, meaning they directly measure the process stream without need for manual sampling. Ultrasonic sensors have been used in process manufacturing for food material characterisation [2,31]; monitoring chemical, pharmaceutical, and biotechnology processes [17]; monitoring fermentation [34]; monitoring freezing of food materials [7]; and quality control in the dairy industry including monitoring reactions, process stream rheology, material structural changes, and component concentrations [30].

Typically, either first principle models or calibration curves are developed to determine properties from US sensor data. However, these can become complex when the sound waves are transmitted through multiple materials or there is variability in process parameters, e.g. temperature. In contrast, supervised ML models can be trained to

* Corresponding author.
E-mail addresses: alexander.bowler@nottingham.ac.uk (A.L. Bowler), nicholas.watson@nottingham.ac.uk (N.J. Watson).

https://doi.org/10.1016/j.ultras.2021.106468
Received 27 July 2020; Received in revised form 8 May 2021; Accepted 10 May 2021
Available online 18 May 2021

correlate sensor data to useful classes (classification) or values (regression) without having to define the complex underlying physical models. ML has been used with US sensors for applications such as monitoring cleaning of dairy fouling in heat exchangers [40,41] and classifying weldment flaws [33,32]. Previous work from our group has shown that ML and a reflection-mode US sensing technique can be combined to effectively monitor two important processes in manufacturing: mixing and cleaning of fouling in pipes [5,10,11]. The reflection-mode sensing technique monitors the sound wave reflected from the vessel wall and process material interface. Mixing is ubiquitous across process manufacturing, being used to combine materials, suspend solids, provide aeration, promote heat and mass transfer, and modify material structure [4]. Being able to determine when a mixing process is complete would provide the benefit of less over or under mixing of materials and therefore less off-specification product. Furthermore, this would lead to a reduction in raw material and energy use. Additionally, accurate prediction of the time remaining until mixing completion would allow for improved scheduling of batch processes leading to higher productivity. Processing equipment is usually cleaned using automated Clean-in-Place (CIP) systems. Cleaning internal surfaces of processing equipment is important to uphold product quality and optimal operating conditions. However, cleaning comes at a cost of lost production time and consumes a vast amount of water and energy [9,38]. CIP processes operate to a standard procedure which is designed to clean the materials which are most difficult to remove from equipment surfaces. This means equipment is often over-cleaned to ensure complete removal of fouling. A sensor able to detect when the cleaning process was complete would eliminate unnecessary resource use and maximise production time.

For training, supervised ML models require a reference measurement to label each sensor data point with a class or value, also termed ground truth data. For both case studies, a camera was used to determine the time for mixing or cleaning completion. This methodology is appropriate in a laboratory, but in a factory, reference measurements are seldom available or require considerable time and cost to obtain, presenting a considerable a barrier to widespread US sensor deployment at industrial scale. To overcome this, a technique is required that can train an ML model to be used on a process where no labelled data is available. In addition to transferring models from laboratory to industrial scale, transferring models for use between different US sensors is also desired. US sensors are transducers which convert electrical pulses to pressure waves, and vice versa, through piezoelectric elements [2]. Owing to differences arising during manufacture of the piezoelectric materials, US sensors of the same model can have different central resonant frequencies and bandwidth. Additionally, US sensors are typically fastened in place with the contact pressure between the sensor and vessel affecting the sound wave transfer across this material boundary. Both these factors result in differences in the received US waveform shapes and magnitudes. Therefore, each ML model is limited to that individual sensor and attachment method, even when monitoring the same process. As such, a method to transfer ML models developed from existing US sensor measurements to new sensors which monitor similar processes would prevent the need for new labelled data for each sensor deployment.

Transfer learning is an area of ML which uses data from a different domain (data distribution) or task (the prediction being made) to reduce the labelling burden of the target domain or task [35]. For example, Zhu et al. [49] recently used transfer learning by fine-tuning a pre-trained Convolutional Neural Network (CNN) to classify thyroid and breast lesions in ultrasound images, and Alguri et al. [1] used numerical simulations and dictionary learning to produce ultrasonic guided wave baselines for damage visualisations in test materials. For a similar task, an ML model trained on source domain data and used to predict on the target domain data will perform poorly if the data distributions between the two domains are different. Domain adaptation is a subcategory of transfer learning which alters how an ML model trains on source domain data so that it also predicts accurately on the target domain data for a similar task [20]. Several review articles covering aspects of domain adaptation are available to the interested reader: Patel et al. [36], Csurka [8], Wang and Deng [42], Pan and Yang [35], Weiss et al [43]. Heimann et al. [16] used instance weighting to overcome the differences in feature space density between synthetic and real data for ultrasound transducer localisation in X-ray fluoroscopy. After applying principal component analysis on features extracted from radiofrequency ultrasound signals or B-mode images together, Azizi et al. [3] used a deep belief network to minimise the divergence between the feature distributions of the two sensing modalities for an unlabelled dataset. Then a labelled dataset was passed through the pre-trained domain adaptation pipeline and a support vector machine was trained to classify the data instances. For application in foetal ultrasound imaging, Meng et al. [29] utilised mutual information minimisation to disentangle categorical features and domain features, and used feature clustering to align categorical features from both domains. For ultrasonic well logging images, [48] used an adversarial method to train an autoencoder to fool a discriminator in being able to distinguish whether the training instance originated from the source or target domains. Gao et al. [13] minimised the maximum mean discrepancy distance metric for domain adaptation between microseismic and pulse-echo data for ultrasonic logging. These works either use convolutional layers, or, in the case of Azizi et al. [3], established feature extraction methodologies. However, in this work, the differences in transducer construction and attachment, as previously outlined, means that few US waveform features will follow the same process trajectory in both the source and target domains. Therefore, this work focuses on investigating methods to extract features which transfer across domains.

This work focuses on transfer learning to an unlabelled target domain using domain adaptation of US sensor data for the two aforementioned case studies: mixing and cleaning of fouling in pipe test sections. Two domain adaptation techniques which transfer a small set of features across domains are compared: a Single Feature (SF) method and Transfer Component Analysis (TCA) using three features. The SF method uses the energy of the US waveform, a physical measurement of the acoustic impedance material being monitored. In contrast, 42 waveform features evaluating the shape of the US waveform are provided to the TCA and three transfer components are produced.

2. Methodology

2.1. Ultrasonic sensors

In this work, the US sensors were used in pulse-echo mode where they transmit a sound wave into the system and receive the returning waves. The received sound waves have reflected from material interfaces approximately perpendicular to the initial wave's direction of travel. The reflected sound wave of interest is that reflected from the interface between the vessel and the process material. The magnitude of this reflected sound wave is proportional to the difference in acoustic impedance between these two neighbouring materials [28]. This monitoring technique requires no transmission of the sound wave through the process material being characterised. This is beneficial as, in a factory setting, process streams usually contain many components such as particles, bubbles or other heterogeneities which cause scattering, reflection and attenuation of the transmitted sound wave. This makes through-transmission methods impractical without higher power, and subsequently higher cost, transducers.

2.2. Mixing case study

Honey-water blending is used as a case study to evaluate these domain adaptation techniques. Full details of the experimental methodology are provided in Bowler et al. [5]. Two transducers (5 MHz resonance, M1057, Olympus) were externally mounted to a 250 ml glass mixing vessel. An overhead stirrer was used to stir the mixture. As honey

is miscible in water, the US sensors monitor the change in component concentration at the sensor measurement area as homogeneity develops. One sensor was attached in the centre of the vessel base (Central sensor) and another was attached approximately 2 cm offset from the centre (Non-Central sensor). The experimental equipment is depicted in Fig. 1a. A US box (Lecoeur Electronique) was used to excite the transducers and digitise the received sound waves. A temperature sensor was attached to the base of the vessel and connected to a PT-104 Data Logger (Pico Technology) to monitor the temperature local to the sensors. US signals were acquired continuously from each probe for 1 s. On average, two US waveforms were recorded during this 1 s interval. The acquired waveforms were averaged to reduce the impact of signal noise. An example of the received US waveforms for a non-mixed and fully mixed system are provided in Fig. 1b. Two different volumes of pure clear honey were used for the experiments: 20 ml and 30 ml. 200 ml tap water was used for all runs. The impeller speed was set to either 200 or 250 rpm. These four parameter permutations were repeated three times whilst varying the laboratory thermostat set point, producing a set of 12 runs across a range of temperatures. The ground truth data for ML model development was obtained by filming the mixing process with a camera to determine the time when the honey had fully dissolved. This experimental procedure was followed on two different days to produce two datasets consisting of 12 runs each. Between the two sets of experiments, the sensors were removed and reattached meaning that their contact and precise location were not the same. This reattachment of the sensors produces a change in the reflected waveforms, necessitating domain adaptation to perform transfer learning across the two datasets. Mixing Dataset 1 had a temperature variation of 19.3–22.1 °C. Mixing Dataset 2 had a temperature variation of 19.8–21.2 °C (Table 1).

Table 1
A summary of the datasets for the mixing experiments, including number of runs and the temperature range each were conducted over.

Mixing dataset	Runs	Temperature range (°C)
Dataset 1	12	19.3–22.1
Dataset 2	12	19.8–21.2

Table 2
A summary of the datasets for the cleaning experiments, including the fouling material used, pipe construction, cleaning fluid temperature and number of runs.

Cleaning dataset	Fouling material	Cleaning fluid temperature	Pipe material	Pipe geometry	Runs
Dataset 1	Malt	Cold	SS340 (base)	Flat	7
Dataset 2	Malt	Hot	SS340 (base)	Flat	7
Dataset 3	Tomato	Cold	SS340 (base)	Flat	7
Dataset 4	Tomato	Hot	SS340 (base)	Flat	7
Dataset 5	Malt	Cold	PMMA	Circular	7
Dataset 6	Malt	Hot	PMMA	Circular	7
Dataset 7	Tomato	Cold	PMMA	Circular	7
Dataset 8	Tomato	Hot	PMMA	Circular	7
Dataset 9	Malt	Cold	SS316	Circular	7
Dataset 10	Malt	Hot	SS316	Circular	7
Dataset 11	Tomato	Cold	SS316	Circular	9
Dataset 12	Tomato	Hot	SS316	Circular	7

2.3. Cleaning case study

Cleaning of pipe fouling was also investigated as a case study for domain adaptation using US sensor data. The full details of the experimental methodology are provided in Escrig et al. [12] and Escrig et al. (2020). Three test sections were used: A rectangular rig with a SS340 bottom plate and clear PMMA sides, a circular pipe constructed from PMMA, and an opaque, circular pipe constructed from SS316. Two food materials (tomato paste and concentrated malt) were used to foul the test sections (Table 2). Fouling material was placed in the centre of the bottom plate for the rectangular rig and 30 mm from the exit for the pipe sections. The fouling material was then spread with a spatula to form a layer of approximately 5 mm thickness and left for 10 min to dry. Cleaning was performed by water with a fluid temperature of either 12 °C or 45 °C and a flowrate of 6 l/s. Cleaning was performed until all the fouling was removed. A minimum of 7 repeats were conducted for all combinations of test sections, fouling materials and fluid temperatures. For the flat test section, the same magnetic transducer as for the honey-water mixing experiments was attached to the base plate. For the pipe sections, different transducers (2 MHz, Yushi, 2P10N) were glued externally to the bottom of the pipes in the location the fouling material would be placed. The same US box, temperature sensor, temperature data logger and laptop were used to acquire the data. A camera was used to record images of the cleaning processes. The camera position was moved depending on whether the pipe section was clear or opaque, as depicted in Fig. 2a. US and temperature data were recorded every 4 s and camera images were recorded every 20 s during the cleaning process. The camera images were used as the ground truth data to label the recorded US data for ML model development.

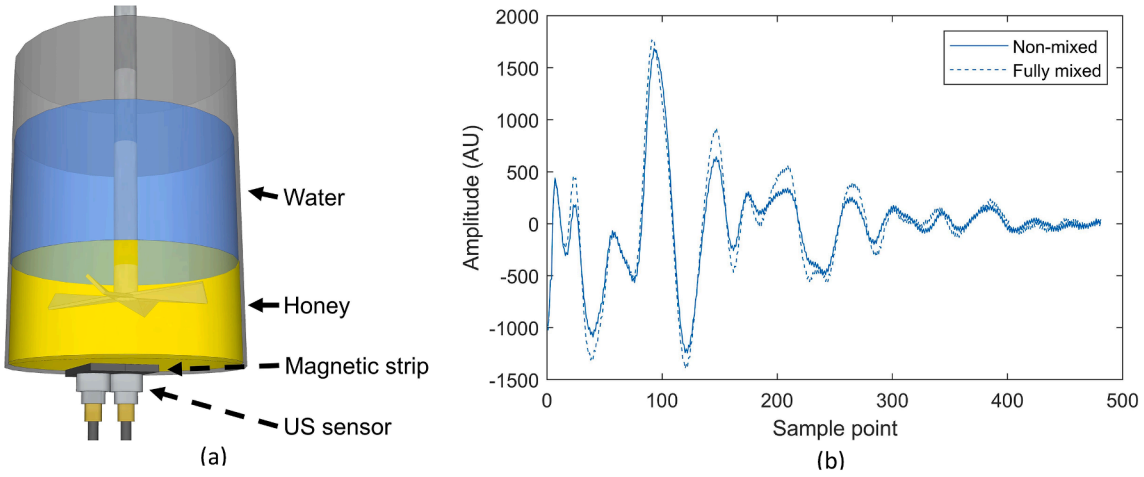


Fig. 1. (a) A diagram of the equipment for the mixing experiments; including 250 ml glass vessel, impeller, and US sensors (Adapted from Bowler et al., [5]). (b) Two received US waveforms corresponding to a non-mixed and a fully mixed system.

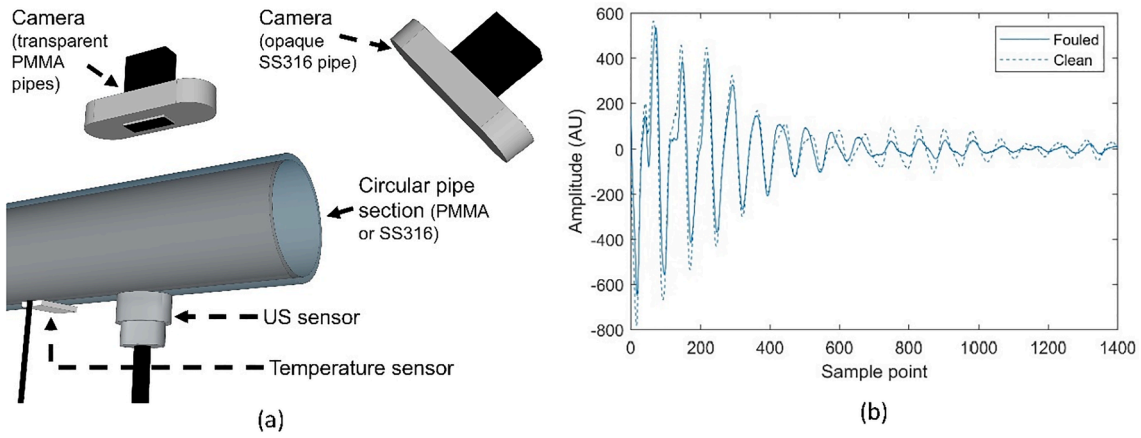


Fig. 2. (a) A diagram of the equipment for the cleaning experiments including pipe section, camera positioning, and sensor locations. (b) Two received US waveforms taken from Cleaning Dataset 9 corresponding to a fouled and clean pipe section.

2.4. Machine learning

Classification ML models were trained to predict whether the mixture was non-mixed or fully mixed and whether the pipe test section is fouled or clean. Regression ML models were trained to predict the process time remaining until fully mixed or clean. For the honey-water mixing, ML models were trained on either Dataset 1 or Dataset 2 and used to predict on the other dataset. This was performed for the Non-Central and Central sensors individually and then by combining data from both sensors. Therefore, an ML model is trained on a labelled mixing system and transferred to monitor a similar mixing process which has no labelled data. For the cleaning of pipe fouling, models were trained on one or several datasets and tested on another. This is representative of training an ML model on a pipe section with labelled data available and transferring this knowledge to an unlabelled process pipe where the pipe material, fouling material, cleaning fluid properties and US sensor could be different.

Shallow ML algorithms, as employed in this study, require features extracted from the US sensor waveform as inputs. Typical features extracted from US waveforms include the waveform shape (e.g. skewness, kurtosis, standard deviation) [6], the amplitude at every sample point in the waveform (Escrig et al., 2020) or frequency components

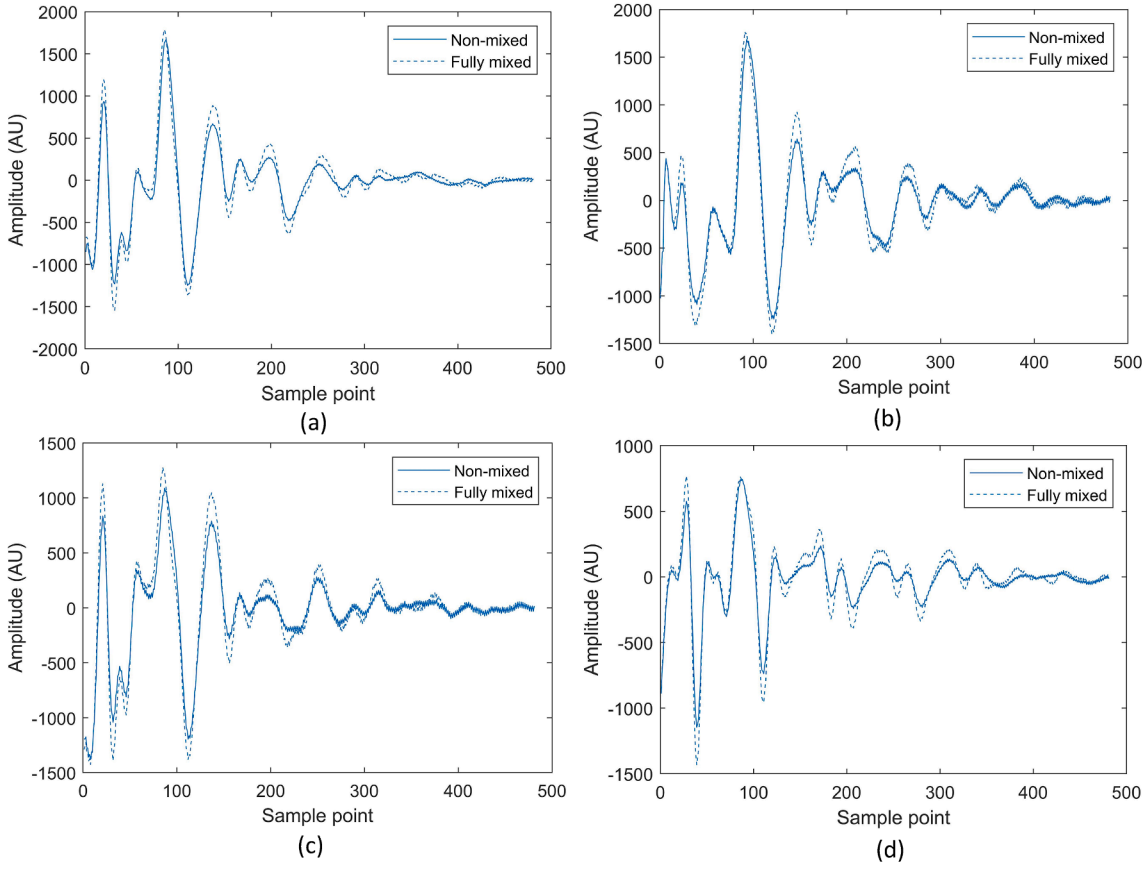


Fig. 3. US waveforms from the mixing experiments corresponding to non-mixed and fully mixed systems. (a) Dataset 1 Non-Central sensor. (b) Dataset 2 Non-Central sensor. (c) Dataset 1 Central sensor. (d) Dataset 2 Central sensor.

obtained after Fourier or Wavelet transforms [5]. However, US waveforms vary each time a sensor is attached. This effect is presented in Fig. 3, where each US waveform differs despite using the same sensors, attachment procedure, vessel and process material. Furthermore, Fig. 4 compares waveforms obtained from Cleaning Datasets 5 and 9, where different pipe construction materials and US sensors were used.

In these case studies, the US sensors are monitoring the magnitude of the sound wave reflecting at the interface between the vessel and process material. The Energy of the US waveform is therefore an effective measure of this, as it is the squared sum of the waveform amplitude at each sample point (Equation (1)). The waveform Energy has previously been used to monitor these two case studies in Bowler et al. (2020) and Escrig et al. [12]. However, the obtained US waveforms are comprised of multiple superimposed sound waves reflecting from different material interfaces. Therefore, the waveform Energy is not entirely colinear with the change in process material at the desired measurement area and additional waveform features can be used to unravel this complexity. Owing to the uniqueness of the waveforms as previously presented, these additional waveform features are unlikely to follow similar trends for different US waveforms. Therefore, the SF method only uses the Energy as a description of the waveform. To investigate whether additional waveform features are required to monitor these case studies, TCA was used to extract three features, or transfer components, to train the transfer learning models. TCA minimises the distance between the source and target domain feature spaces using the Maximum Mean Discrepancy and extracts transfer components that maximise variance across this shared feature space [37]. A total of 42 waveform features were inputted into the TCA algorithm (Sections 2.4.1 and 2.4.2). Every run in the source domain dataset was used for model training and every run in the target domain dataset was used for testing. An additional model, named the Non-Transfer Learning model, was trained using only the target domain data to provide a comparative result to the transfer learning models' accuracy. A k-fold testing procedure was used for the Non-Transfer Learning model, where k is the number of runs in the dataset. One run was held back for testing and training was carried out on the remaining runs. The run held back was changed sequentially and the average accuracy of this procedure was used to provide a measure for model generalisability. Only the waveform Energy was used as a feature in this model. An overview of this methodology is presented in Fig. 5. All data analysis and ML algorithms were completed in MATLAB R2019a.

2.4.1. Features

The waveform energy is the summed squared amplitude of every sample point in a waveform.

$$E = \sum_{i=1}^{i=SP} A_i^2 \tag{1}$$

where E is the waveform energy, SP is the total number of sample points in the waveform, and A_i is the amplitude at sample point I [46].

$$SRA = \sum_{i=1}^{i=SP} \sqrt{|A_i|} \tag{2}$$

where SRA is the sum root amplitude [46].

$$SAA = \sum_{i=1}^{i=SP} |A_i| \tag{3}$$

where SAA is the sum absolute amplitude [46].

$$\mu = \frac{\sum_{i=1}^{i=SP} A_i}{SP} \tag{4}$$

$$STD = \sqrt{\frac{1}{SP}\sum_{i=1}^{i=SP} (A_i - \mu)^2} \tag{5}$$

where μ is the mean waveform amplitude and STD is the standard deviation [46].

$$S = \frac{\sum_{i=1}^{i=SP} (A_i - \mu)^3}{SP \times STD^3} \tag{6}$$

where S is the waveform skewness [6].

$$K = \frac{\sum_{i=1}^{i=SP} (A_i - \mu)^4}{SP \times STD^4} \tag{7}$$

where K is the waveform kurtosis [46].

2.4.1.1. Feature gradient. Using the gradient of the waveform features provides a measure of the process trajectory. The difference between consecutive waveform features were calculated after applying a backwards, one-sided moving mean. A backwards, one-sized gradient uses only the past process data. The size of the moving mean was chosen as 5% of the average run time for the respective dataset. This is to ensure that the energy gradient is similar feature across the source and target domains.

$$MMV_i = \frac{1}{N}\sum_{i}^{i-N} V_i \tag{8}$$

$$G = MMV_i - MMV_{i-1} \tag{9}$$

where G is the gradient of a parameter, MMV is the moving mean value

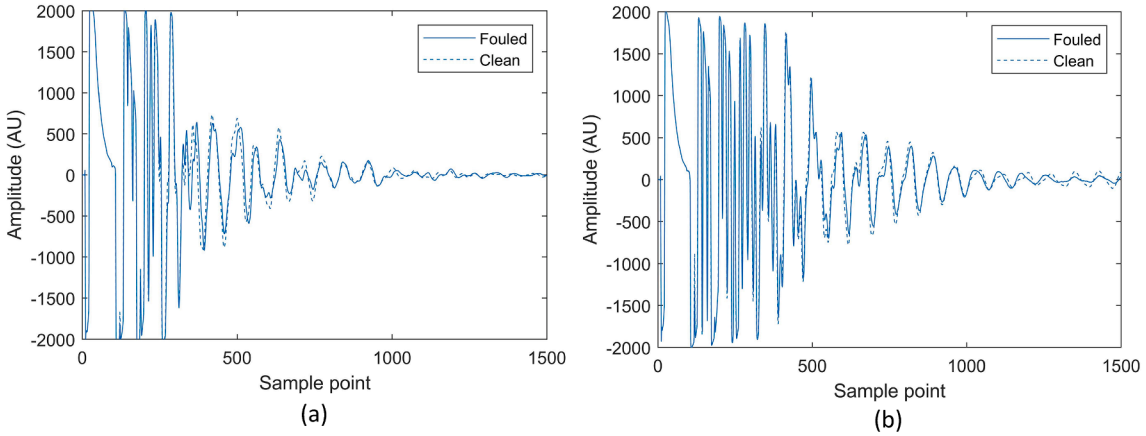


Fig. 4. US waveforms from the pipe cleaning experiments corresponding to fouled and clean pipe section. (a) Dataset 5 – circular plastic pipe section. (b) Dataset 9 – circular metal pipe section.

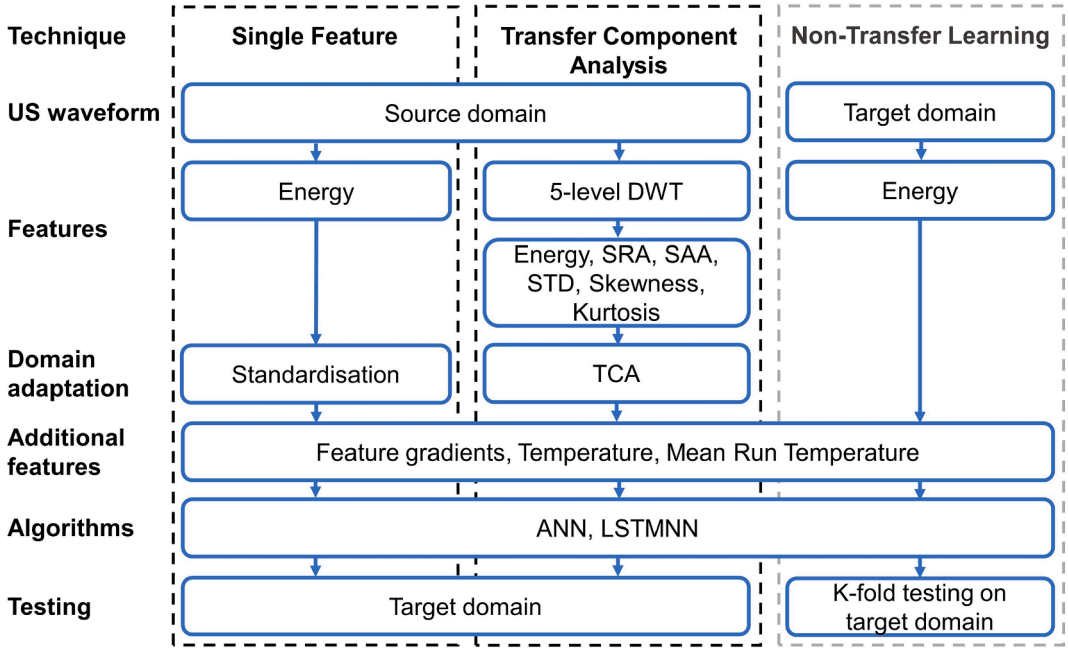


Fig. 5. A methodology flow diagram for the three models being compared. The two transfer learning models, SF and TCA, and the Non-Transfer Learning model.

of a parameter, N is the size of backwards, one-sided moving mean, and V is the original parameter value [25,26].

2.4.1.2. Temperature and mean run temperature. As the acoustic properties of materials are highly dependent on temperature [17], the local temperature measurement was also investigated as a feature. The additional Temperature feature was the measured temperature at the time each US waveform was obtained. Furthermore, the Mean Run Temperature (the average temperature for that repeat of the process) was investigated as the temperature sensors are located external to the process vessels. Therefore, any change in temperature may not be representative of temperature changes of the process material.

2.4.2. Discrete waveform analysis

The Discrete Wavelet Transform (DWT) is a method of obtaining the frequency-time information of a waveform [23]. At each decomposition, an orthogonal wavelet transform function produces a detail and approximate waveform which contain no redundant information [22]. The frequency of the analytical wavelet is successively halved for each decomposition level. The Symlet 6 wavelet was selected as the Mother wavelet owing to it being the least asymmetric, and therefore most visually similar to the expected waveforms [24], along with its previous success in analysing US waveforms [5]. 5 decomposition levels were used, and the previously described waveform features were applied to each resultant waveform producing a total of 42 features as inputs to the TCA algorithm.

2.4.3. Standardisation

For the SF transfer learning method, the features of each domain were standardised to produce feature spaces with a mean of 0 and a standard deviation of 1. This was to align and scale the feature spaces so that the ML model trained on the source domain could predict accurately on the target domain data. The process of feature standardisation is provided in equations 10–12.

$$\mu = \frac{\sum_{i=1}^{i=n} x_i}{n} \tag{10}$$

$$\sigma = \sqrt{\frac{1}{n-1}\sum_{i=1}^{n} |x_i - \mu|^2} \tag{11}$$

$$Z = \frac{x - \mu}{\sigma} \tag{12}$$

where μ is the mean of feature x, n is the number of data points for feature x, σ is the standard deviation of x, and Z is the new standardised feature.

Furthermore, for the honey-water blending experiments, prior to standardisation, the waveform energy of the first data point in each run was subtracted from all data points of that run so that they all began at a waveform energy of 0. The process material being measured at the start of each run is known to be honey as the honey settles to the bottom and the sensors are located on the vessel base. This is analogous to an industrial process having the same process material located at the sensor measurement area at the start of each run. This procedure further aligns the feature spaces despite the wide temperature range the honey-water mixing experiments were conducted over. As the laboratory set point temperature was not altered for the pipe section cleaning experiments, this additional operation was not performed. The feature standardisation method for the mixing data and the cleaning data is presented in Figs. 6 and 7, respectively.

2.4.4. Transfer component analysis

TCA attempts to extract transfer components across the source and target domains in a Reproducing Kernel Hilbert Space using the Maximum Mean Discrepancy [37]. Three dimensions, or transfer components, were selected to allow for comparison against the SF method. The TCA code provided in the MATLAB domain adaptation toolbox produced by Yan [45] was used.

2.4.5. Algorithms

2.4.5.1. Artificial neural networks. Artificial neural networks (ANNs) can create linear relationships between combinations of input variables and the activation function [19]. For this reason, despite the few input features, 5 neurons were used in the hidden layer to ensure production of a linear relationship. The "trainlm" training function was used for

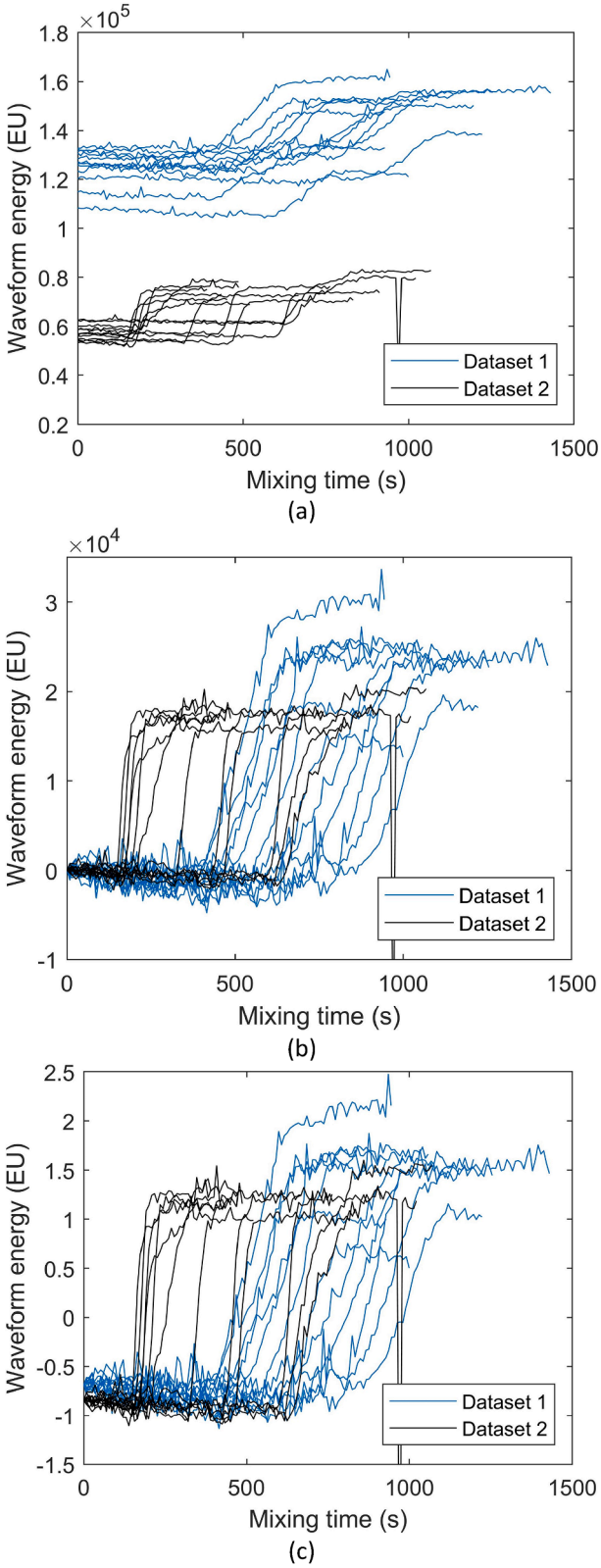


Fig. 6. The standardisation procedure for the mixing datasets. (a) All runs from Mixing Datasets 1 and 2. (b) All runs following the subtraction of the first waveform energy in each run, thereby aligning each of the first data points. (c) All runs following standardisation.

regression models and the "trainscg" training function was used for the classification models [27]. To prevent overfitting, the model training was stopped once the validation loss had increased for 6 consecutive iterations. For each prediction task, 10 neural networks were trained and tested, and the average accuracy was used. This is to account for the effects of random weight initialisation and that ANNs converge to local minima. 80% of the training data was used as a training set and the remaining 20% was used as the validation set.

2.4.5.2. Long short-term memory neural networks. To evaluate whether a more complex process trajectory memory was required rather than the gradient of the waveform energy alone, Long Short-Term Memory neural networks (LSTMNNs) were also investigated. LSTMNNs can store representations of all previous time-steps in a process though updating an internal network state using gate units [18]. No validation set was used to maximise the training data set size for the LSTMNN. The inputs were standardised and a mini-batch size of 1 was used. The training was carried out for 600 epochs to ensure fitting, using the "adam" optimisation algorithm, a learning rate of 0.01, and a gradient threshold of 1 to prevent problems of exploding gradients. Only 5 hidden units were used in the LSTM layer, as the processes did not follow a complex sequence. 5 neurons were used in the fully connected layer to ensure linear fitting of the feature combinations with the activation function.

3. Results and discussion

3.1. Honey-water mixing

For the honey-water mixing experiments, classification ML models were trained to predict whether the mixture is non-mixed or fully-mixed, and regression models to predict the time remaining until mixing completion. The models were trained on a source domain dataset (either Dataset 1 or Dataset 2) and used to predict on the other, target domain dataset.

3.1.1. Classification

Overall, transfer learning models trained for the Non-Central sensor produced poor classification accuracy (Table 3). The highest classification accuracy for the SF method was 73.9% and the highest for TCA was 74.6%. This is compared to the Non-Transfer Learning model, which produced accuracies of up to 92.2%. The cause of the poor classification accuracy for the Non-Central sensor is due to the difference in the sensor's location between Dataset 1 and Dataset 2, being closer to the vessel sides in Dataset 1. As the honey is mixed earlier at the vessel sides than in the centre of the vessel base, the waveform Energy of the Non-Central sensor in Dataset 1 begins to rise earlier with respect to the Central sensor. This is shown in Fig. 8. There is greater variability in the waveform Energy for the Non-Central sensor compared with the Central sensor due to the base of the vessel not being flat at this location, creating discrepancies in the sound wave received by the sensor [5]. The point defined as complete mixing (the time at which all honey has dissolved) is located at the centre of the vessel base and therefore non-local to the Non-Central sensor. The ML models correlate the sensor data to this non-local phenomenon. If the location of the sensor changes between the source and target domains, there is now an offset in the prediction. This demonstrates that if applying transfer learning models to unlabelled target systems which correlate sensor data to non-local phenomena, this offset in prediction must be similar across domains.

The SF method produced higher classification accuracies than TCA for all tasks using the Central sensor, indicating that the waveform Energy alone is more amenable to domain adaptation than the three transfer components. The SF method was able to produce high prediction accuracies of up to 96.0% using Dataset 1 as the source domain and Dataset 2 as the target domain. This accuracy was similar to the Non-Transfer Learning model trained on Dataset 2 which achieved 95.9%.

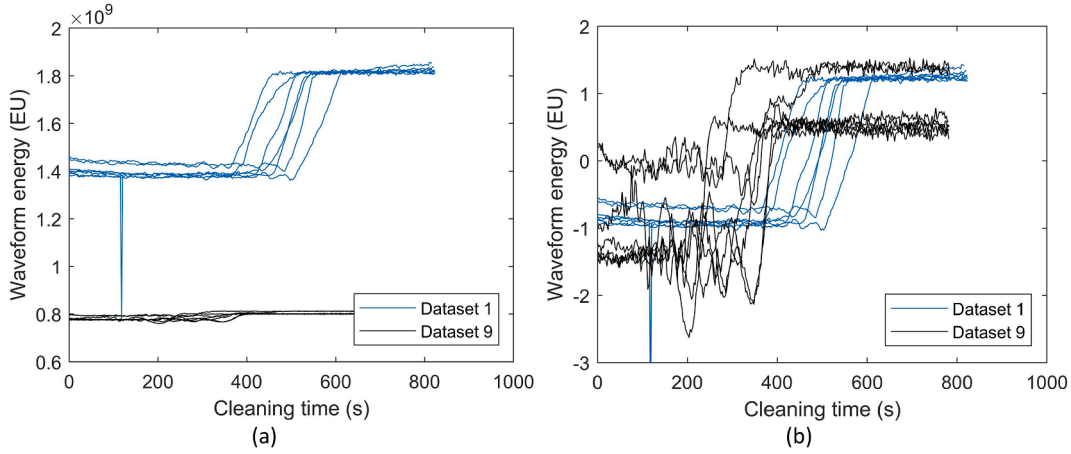


Fig. 7. The standardisation procedure for the cleaning datasets. (a) All runs from Malt Cold Flat and Malt Cold Metal datasets. (b) All runs from Malt Cold Plastic and Malt Cold Metal datasets following standardisation.

Table 3
Classification results for honey-water mixing experiments. Two of the algorithm and feature combinations which produced the highest accuracy for each model are included; one using the temperature as feature, and one without. The Additional features column denotes the features inputted into the model other than the features used for domain adaptation, e.g. the waveform Energy for the SF method, or the three transfer components used for TCA. G – Gradient of features, T – Temperature, MT – Mean run temperature.

Sensor	Source domain	Target domain	Transfer learning method	Accuracy (% correct)	Algorithm	Additional features
Non-Central	Dataset 1	Dataset 2	SF	70.8	ANN	G
				73.4	LSTM	G, MT
			TCA	74.7	ANN	–
				74.7	LSTM	G, MT
			NTL	90.3	ANN	G
				92.2	LSTM	G, T
	Dataset 2	Dataset 1	SF	72.6	ANN	G
				73.9	ANN	G, MT
			TCA	68.4	ANN	G
				70.3	ANN	G, MT
			NTL	90.1	LSTM	–
				84.9	LSTM	G, T
Central	Dataset 1	Dataset 2	SF	92.5	LSTM	G
				96.0	LSTM	G, T
			TCA	92.2	LSTM	–
				92.6	LSTM	G, MT
			NTL	94.4	LSTM	G
				95.9	LSTM	T
	Dataset 2	Dataset 1	SF	92.8	LSTM	G
				93.8	LSTM	MT
			TCA	87.6	LSTM	–
				89.9	LSTM	MT
			NTL	96.7	LSTM	–
				95.1	LSTM	G, T
Combined	Dataset 1	Dataset 2	SF	92.1	ANN	G
				92.2	ANN	G, MT
			TCA	92.1	LSTM	G
				90.4	LSTM	G, MT
			NTL	95.4	LSTM	–
				94.8	LSTM	G, MT
	Dataset 2	Dataset 1	SF	91.6	LSTM	–
				91.9	LSTM	MT
			TCA	87.3	ANN	–
				89.2	LSTM	G, MT
			NTL	95.4	ANN	G
				95.6	LSTM	G, T

The Central sensors were located at the centre of the vessel base for both datasets, and as mixing completion occurred at the sensor measurement area, there was no offset in the classification model prediction. Using Dataset 1 as the source domain produced higher classification accuracies as Dataset 1 was performed over a wider temperature range. This led to more variability in the waveform energy (as shown in Fig. 6) and hence provides a form of regularisation during model training and improved model generalisability to the target domain. This highlights that source domain datasets should be gathered over a wide process parameter range to enable the model to generalise. LSTMNNs produced the highest classification accuracies for all tasks using the Central sensor. The more complex process trajectory stored by the LSTMNNs was beneficial compared with using the waveform energy gradient with the ANNs and did not lead to overfitting.

Using both sensors produced lower classification accuracies than using the Central sensor alone due to incorporating the poorly

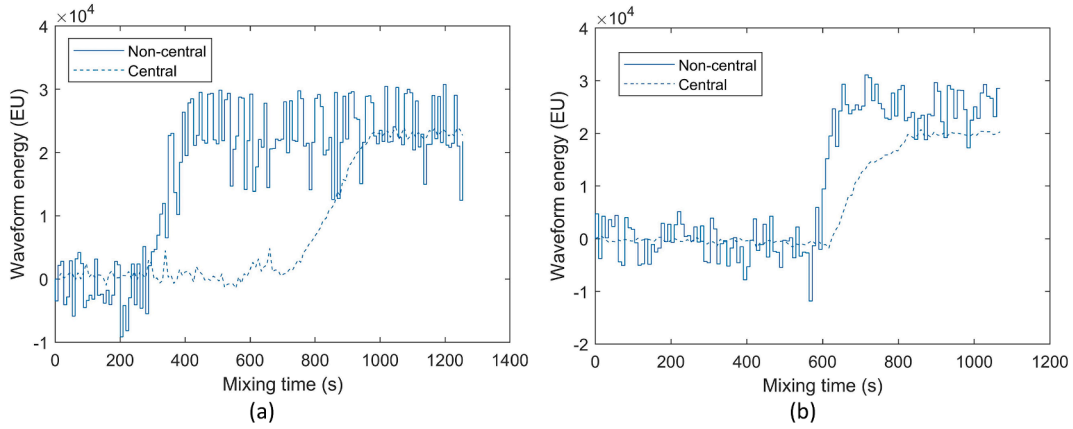


Fig. 8. The waveform Energy of the Non-Central sensor increases earlier with respect to the Central sensor during the mixing process for Dataset 2 due to the difference in sensor location. (a) Waveform Energy profiles for the Non-Central and Central sensors during Run 1 of Dataset 1. (b) Waveform energy profiles for the Non-Central and Central sensors during Run 1 of Dataset 2.

performing Non-Central sensor. Using the temperature as a feature produced higher classification accuracies for all domain adaptation tasks, excluding TCA from Dataset 1 to Dataset 2. This enhanced performance is due to the large effect of temperature on material acoustic impedance and subsequently the waveform shape and Energy. Furthermore, the models were also able to learn the relationship of higher temperature reducing the mixing time by lowering the viscosity of the honey. However, an accuracy of 92.1% using the Central sensor was achieved without incorporating the temperature using both the SF method and TCA.

3.1.2. Regression

Similar to the classification results, domain adaptation of the Non-Central sensor data produced significantly lower regression accuracies (up to 0.905) than the Non-Transfer Learning models which were trained on the target domain data (up to 0.978) (Table 4). Again, this is

Table 4
Regression results for honey-water mixing experiments. Two of the algorithm and feature combinations which produced the highest accuracy for each model are included; one using the temperature as feature, and one without. The Additional features column denotes the features inputted into the model other than the features used for domain adaptation, e.g. the waveform Energy for the SF method, or the three transfer components used for TCA. G – Gradient of features, T – Temperature, MT – Mean run temperature.

Sensor	Source domain	Target domain	Transfer learning method	Accuracy (R^2)	Algorithm	Features
Non-Central	Dataset 1	Dataset 2	SF	0.903	LSTM	–
				0.894	LSTM	G, MT
			TCA	0.846	LSTM	G
				0.902	LSTM	MT
			NTL	0.932	LSTM	G
				0.938	LSTM	T
	Dataset 2	Dataset 1	SF	0.877	LSTM	–
				0.810	LSTM	MT
			TCA	0.883	LSTM	–
				0.905	LSTM	T
			NTL	0.978	LSTM	–
				0.953	LSTM	T
Central	Dataset 1	Dataset 2	SF	0.919	ANN	G
				0.945	LSTM	G, MT
			TCA	0.942	LSTM	–
				0.941	LSTM	MT
			NTL	0.931	LSTM	–
				0.950	LSTM	MT
	Dataset 2	Dataset 1	SF	0.899	LSTM	–
				0.908	LSTM	MT
			TCA	0.798	LSTM	G
				0.878	LSTM	G, T
			NTL	0.930	LSTM	G
				0.939	LSTM	G, T
Combined	Dataset 1	Dataset 2	SF	0.942	LSTM	G
				0.947	LSTM	G, T
			TCA	0.939	LSTM	–
				0.929	LSTM	MT
			NTL	0.941	LSTM	–
				0.946	LSTM	MT
	Dataset 2	Dataset 1	SF	0.930	LSTM	–
				0.921	LSTM	T
			TCA	0.673	LSTM	G
				0.896	LSTM	MT
			NTL	0.981	LSTM	G
				0.981	LSTM	MT

attributed to the change in sensor position. As the position of the Central sensor has not changed between datasets, R^2 values of up to 0.945 were achieved using the SF method, similar to the Non-Transfer Learning models' regression accuracy of up to 0.950.

Again, using temperature as a feature aided prediction accuracy of the Central sensor, most likely because of the aforementioned effect on temperature on the mixing time. Therefore, these models were able to infer the time until mixing completion near the beginning of the process, where no change in acoustic impedance had yet been detected by the Central sensor. In contrast to the classification tasks, using both sensors together led to greater regression accuracies for the SF method. This is owed to the greater resolution of the Non-Central sensor near the beginning of the mixing process, as the honey is first removed from the vessel base in this location, and the Central sensor's greater resolution at the end, where the last of the honey is mixed [5]. As with the classification models, using Dataset 1 as the source domain and Dataset 2 as the target domain produced more accurate models for most regression tasks due to the wider temperature range in Dataset 1. Again, LSTMNN models were more accurate owing to their ability to store representations of all previous process time-steps and therefore learn more complex feature trajectories than the ANNs.

3.2. Cleaning of fouling in pipes

For the cleaning experiments, classification ML models were trained to predict whether the pipe section is fouled or clean, and regression models predict the time remaining until cleaned. The models were trained on a source domain dataset, or multiple datasets for the SF method, and used to predict on another, target domain dataset.

3.2.1. Classification

For all classification tasks, the SF method produced higher classification accuracies than TCA, again suggesting that a single feature is optimal for domain adaptation of US waveforms (Table 5). For all classification tasks, excluding Datasets 11 and 12, the SF domain adapted models were either equal to or more accurate than the Non-Transfer Learning models trained on the target domain data. Using temperature as a feature was not required for high classification accuracy, and only led to higher accuracy for the Dataset 12 as the target domain. Combining multiple source domain datasets for the SF method produced the highest classification accuracy for Datasets 5 and 11 as the target domain. This is because using multiple source domain datasets provides regularisation of the ML models by training them to generalise over multiple domains. Similar to the honey-water blending experiments, LSTMNNs were in general more accurate than ANNs due to their ability to learn complex process trajectories.

3.2.2. Regression

Similar to the classification tasks, the SF method produced higher prediction accuracies than TCA for most regression tasks (Table 6). For all target domain datasets, except for Dataset 7, the domain adaptation models produced equally high regression accuracy as the Non-Transfer Learning models which were trained on the target domain dataset. Unlike the classification tasks where using the temperature as a feature led to no improvements in prediction accuracy, incorporating the temperature into the models produced higher regression accuracies for Datasets 5, 6 and 10. This is because for most of the process there is no change in the material at the sensor measurement area and so accounting for the effects of temperature on the waveform energy would aid regression accuracy during these sections of the process. In contrast, the classification tasks are focused on the section of the process where the fouling material is being removed, resulting in large changes in the waveform Energy. Other than for Datasets 7 and 8 as the target domain, using multiple datasets as the source domain produced the highest regression accuracies for the SF method. Again, this is attributed to the models being trained to generalise across multiple datasets, increasing the likelihood of accurate prediction of the target dataset. LSTMNNs produced the highest regression accuracies for every domain adaptation task. This suggests that they were not prone to overfitting despite their ability to learn complex process trajectories.

Table 5
Classification results the cleaning of food fouling experiments. Two of the algorithm, feature, and source domain datasets combinations which produced the highest accuracy for each model are included; one using the temperature as feature, and one without. The Additional features column denotes the features inputted into the model other than the features used for domain adaptation, e.g. the waveform Energy for the SF method, or the three transfer components used for TCA. G – Gradient of features, T – Temperature, MT – Mean run temperature.

Target domain	Transfer learning method	Accuracy (% correct)	Source domain	Algorithm	Features
Dataset 5	SF	93.6	Datasets 1 & 2	LSTM	–
		93.2	Datasets 1 & 2	LSTM	G, T
	TCA	87.1	Dataset 2	LSTM	–
		86.7	Dataset 2	ANN	MT
	NTL	93.8	–	LSTM	–
		87.0	–	ANN	G, T
Dataset 6	SF	96.4	Dataset 4	LSTM	–
		95.4	Dataset 3	LSTM	G, T
	TCA	92.8	Dataset 2	LSTM	–
		93.7	Dataset 4	LSTM	T
	NTL	92.2	–	LSTM	G
		96.1	–	LSTM	G, T
Dataset 7	SF	95.4	Dataset 2	LSTM	–
	TCA	88.1	Dataset 2	LSTM	G
	NTL	91.2	–	LSTM	–
Dataset 8	SF	96.4	Dataset 3	LSTM	G
	TCA	94.1	Dataset 4	ANN	G
	NTL	95.6	–	LSTM	–
Dataset 9	SF	93.2	Dataset 1	LSTM	G
		90.0	Dataset 2	LSTM	MT
	TCA	81.0	Dataset 5	LSTM	G
		84.8	Dataset 5	LSTM	T, G
	NTL	92.2	–	LSTM	G
		91.8	–	LSTM	T
Dataset 10	SF	98.4	Dataset 3	LSTM	–
		97.5	Dataset 5	LSTM	G, T
	TCA	94.7	Dataset 4	ANN	G
		95.3	Dataset 4	LSTM	G, MT
	NTL	98.2	–	LSTM	–
		95.4	–	LSTM	MT
Dataset 11	SF	91.6	Datasets 1 & 2	LSTM	–
		86.5	Datasets 1, 2, 5 & 6	LSTM	T
	TCA	81.0	Dataset 1	ANN	–
		81.0	Dataset 2	ANN	MT
	NTL	95.9	–	LSTM	–
		95.9	–	LSTM	T
Dataset 12	SF	90.0	Dataset 7	LSTM	G
		92.4	Dataset 5	LSTM	MT
	TCA	89.9	Dataset 7	LSTM	G
		85.7	Dataset 4	LSTM	G, MT
	NTL	95.2	–	LSTM	–
		96.7	–	LSTM	G, T

3.3. Comparison with previous work

Despite using fewer ML model input features and training the models on a different data distribution to the target domain, the accuracies of the transfer learning models tested in this work are only slightly lower than our previously published results. For the honey-water mixing experiments, classification accuracies of 96.0% and regression accuracies of 0.947 are achieved using the SF method compared with 96.3% and 0.977 [5]. For the cleaning of pipe fouling, classification of accuracies of between 91.6 and 98.4 % are achieved in this work compared with previous results of 98–100% [10,11]. These results are similar to the

Table 6
Regression results for cleaning of food fouling experiments. Two of the algorithm, feature, and source domain datasets combinations which produced the highest accuracy for each model are included; one using the temperature as feature, and one without. The Additional features column denotes the features inputted into the model other than the features used for domain adaptation, e.g. the waveform Energy for the SF method, or the three transfer components used for TCA. G – Gradient of features, T – Temperature, MT – Mean run temperature.

Target domain	Transfer learning method	Accuracy (R^2)	Source domain	Algorithm	Features
Dataset 5	SF	0.894	Datasets 1 & 2	LSTM	G
		0.987	Datasets 1 & 2	LSTM	G, MT
	TCA	0.861	Dataset 1	LSTM	–
		0.820	Dataset 1	LSTM	T
	NTL	0.947	–	LSTM	–
		0.949	–	LSTM	G, T
Dataset 6	SF	0.998	Datasets 1, 2, 3 & 4	LSTM	–
		0.999	Datasets 1, 2, 3 & 4	LSTM	T
	TCA	0.870	Dataset 4	LSTM	–
		0.775	Dataset 4	LSTM	G, T
	NTL	0.997	–	LSTM	–
		0.987	–	LSTM	T
Dataset 7	SF	0.639	Dataset 2	LSTM	G
	TCA	0.747	Dataset 2	LSTM	–
	NTL	0.959	–	LSTM	G
Dataset 8	SF	0.992	Dataset 4	LSTM	–
	TCA	0.890	Dataset 3	LSTM	–
	NTL	0.983	–	LSTM	–
Dataset 9	SF	0.996	Datasets 1, 2, 5 & 6	LSTM	–
		0.988	Datasets 1, 2, 5 & 6	LSTM	MT
	TCA	0.962	Dataset 1	LSTM	–
		0.922	Dataset 1	LSTM	G, MT
	NTL	0.990	–	LSTM	G
		0.990	–	LSTM	T
Dataset 10	SF	0.947	Datasets 5, 6, 7 & 8	LSTM	G
		0.991	Datasets 1, 2, 3, 4, 5, 6, 7 & 8	LSTM	MT
	TCA	0.966	Dataset 1	LSTM	–
		0.947	Dataset 4	LSTM	G, T
	NTL	0.998	–	LSTM	–
		0.998	–	LSTM	G, T
Dataset 11	SF	0.983	Datasets 1, 2, 5 & 6	LSTM	–
		0.956	Datasets 1, 2, 5 & 6	LSTM	G, MT
	TCA	0.880	Dataset 1	LSTM	–
		0.687	Dataset 3	LSTM	T
	NTL	0.919	–	LSTM	–
		0.855	–	LSTM	G, MT
Dataset 12	SF	0.993	Datasets 5, 6, 7 & 8	LSTM	–
		0.992	Datasets 1, 2, 3 & 4	LSTM	G, MT
	TCA	0.937	Dataset 3	LSTM	–
		0.890	Dataset 4	LSTM	G, T
	NTL	0.948	–	LSTM	–
		0.902	–	LSTM	T

domain adaptation methodologies used for motor bearing fault diagnosis by vibration signal monitoring. Wen et al. [44] achieved classification accuracies averaging 99.79% on the widely-studied Case Western Reserve University dataset using a Convolutional Neural Network (CNN) based model. In comparison, Zhang et al. [47] achieved average classification accuracies of 95.5% using a CNN based domain adaptation method across different load domains and Li et al. [21] achieved accuracies >92% using a generative model. Furthermore, Guo et al. [15] achieved classification accuracies of up to 89.9% when transferring models from different machines. This similarity demonstrates the efficacy of the techniques proposed in this work to monitor processes with no labelled data available. To improve the accuracy of the trained models, a small set of labelled data in the target domain would allow for aligning not only the marginal probabilities but also the conditional probabilities. Furthermore, a small set of labelled data would allow the presented techniques to be combined with semi-supervised learning approaches to train robust ML models.

4. Conclusion

Sensors are a key technology in the fourth industrial revolution, especially for process manufacturing sectors which have greater variability in material streams and process conditions than in discrete manufacturing. However, to fully realise the potential benefits, the problem of training ML models on limited labelled sensor data must be overcome. This work has compared two domain adaptation approaches for monitoring processes using US sensors to reduce the burden of data labelling in factory environments. These were: a Single Feature method and Transfer Component Analysis using three features. US waveforms are dependent on the sensor used, attachment procedure, and contact pressure. Therefore, this work investigated transferring a small number of features across domains. It was shown that ML models using US sensor data can be trained on a similar task in a source domain and can accurately predict using sensor data from a target domain. Two case studies were investigated: honey-water mixing using datasets recorded on different days after sensor reattachment, and cleaning of fouling in pipe sections of different geometry and construction materials. Overall, the Single Feature method produced the highest prediction accuracies, indicating that using the waveform Energy alone is optimal for domain adaptation between US sensors. Classification accuracies of up to 96.0% and 98.4% were achieved for predicting the completion of mixing or cleaning, and R^2 values of up to 0.947 and 0.999 were reached to predict the processing time remaining for each process, respectively. These results were similar to comparative supervised models which did not employ transfer learning, indicating that the domain adaptation approach was successful.

Increasing the feature variability in the source domains aided prediction accuracy by providing regularisation to the ML models during training. For the honey-water mixing, using a source domain dataset obtained over a wider temperature range increased prediction accuracy. For cleaning of pipe fouling, combining multiple source domain datasets trained the model to generalise across domains and thereby improved performance on the target domain data. For the honey-water mixing experiments, the Non-Central sensor produced low accuracy predictions because the sensor position had changed between the source and target domains. When correlating sensor data to phenomena non-local to the sensor measurement area, an offset between process material changes at the sensor location and the prediction task is learned. This suggests that when using a transfer learning model to correlate sensor data to non-local phenomena, the learned offset must be ensured to be similar across domains. To monitor cleaning of fouling in pipes, it was shown that ML models could be trained using different US sensors, pipe materials, pipes geometries, fouling materials and cleaning fluid properties.

Funding

This work was supported by the Engineering and Physical Sciences Research Council (EPSRC) standard research studentship (EP/R513283/1).

CRediT authorship contribution statement

Alexander L Bowler: Conceptualization, Formal analysis, Funding acquisition, Methodology, Project administration, Supervision,

Visualization, Writing - review & editing. **Nicholas J Watson:** Data curation, Investigation, Resources, Software, Validation, Writing - review & editing.

Declaration of Competing Interest

The authors declare that they have no known competing financial interests or personal relationships that could have appeared to influence the work reported in this paper.

References

[1] K.S. Alguri, C.C. Chia, J.B. Harley, Sim-to-Real: employing ultrasonic guided wave digital surrogates and transfer learning for damage visualization, Ultrasonics 111 (2021), https://doi.org/10.1016/j.ultras.2020.106338.

[2] T.S. Awad, H.A. Moharram, O.E. Shaltout, D. Asker, M.M. Youssef, Applications of ultrasound in analysis, processing and quality control of food: a review, Food Res. Int. 48 (2) (2012) 410–427, https://doi.org/10.1016/j.foodres.2012.05.004.

[3] S. Azizi, P. Mousavi, P. Yan, A. Tahmasebi, J.T. Kwak, S. Xu, B. Turkbey, P. Choyke, P. Pinto, B. Wood, P. Abolmaesumi, Transfer learning from RF to B-mode temporal enhanced ultrasound features for prostate cancer detection, Int. J. Comput. Assist. Radiol. Surg. 12 (7) (2017) 1111–1121, https://doi.org/10.1007/s11548-017-1573-x.

[4] A.L. Bowler, S. Bakalis, N.J. Watson, A review of in-line and on-line measurement techniques to monitor industrial mixing processes, Chem. Eng. Res. Des. 153 (January) (2020) 463–495, https://doi.org/10.1016/j.cherd.2019.10.045.

[5] A.L. Bowler, S. Bakalis, N.J. Watson, Monitoring mixing processes using ultrasonic sensors and machine learning, Sensors (Switzerland) 20 (7) (2020), https://doi.org/10.3390/s20071813.

[6] W. Caesarendra, T. Tjahjowidodo, A review of feature extraction methods in vibration-based condition monitoring and its application for degradation trend estimation of low-speed slew bearing, Machines 5 (4) (2017) 21, https://doi.org/10.3390/machines5040021.

[7] X. Cheng, M. Zhang, B. Xu, B. Adhikari, J. Sun, The principles of ultrasound and its application in freezing related processes of food materials: a review, Ultrason. Sonochem. 27 (2015) 576–585, https://doi.org/10.1016/j.ultsonch.2015.04.015.

[8] G. Csurka, A comprehensive survey on domain adaptation for visual applications, Adv. Comput. Vis. Pattern Recogn. (2017) 1–35, https://doi.org/10.1007/978-3-319-58347-1_1, 9783319583464.

[9] M.H. Eide, J.P. Homleid, B. Mattsson, Life cycle assessment (LCA) of cleaning-in-place processes in dairies, LWT - Food Sci. Technol. 36 (3) (2003) 303–314, https://doi.org/10.1016/S0023-6438(02)00211-6.

[10] J. Escrig, E. Woolley, A. Simeone, N.J. Watson, Monitoring the cleaning of food fouling in pipes using ultrasonic measurements and machine learning, Food Control 116 (2020), https://doi.org/10.1016/j.foodcont.2020.107309.

[11] J.E. Escrig, A. Simeone, E. Woolley, S. Rangappa, A. Rady, N.J. Watson, Ultrasonic measurements and machine learning for monitoring the removal of surface fouling during clean-in-place processes, Food Bioprod. Process. (September) (2020) 1–13, https://doi.org/10.1016/j.fbp.2020.05.003.

[12] J.E. Escrig, E. Woolley, S. Rangappa, A. Simeone, N.J. Watson, Clean-in-place monitoring of different food fouling materials using ultrasonic measurements, Food Control 104 (October) (2019) 358–366, https://doi.org/10.1016/j.foodcont.2019.05.013.

[13] X. Gao, Y. Shi, Q. Zhu, Z. Li, H. Sun, Z. Yao, W. Zhang, Domain adaptation in intelligent ultrasonic logging tool: from microseismic to pulse-echo, IEEE Trans. Instrum. Meas. 70 (2021), https://doi.org/10.1109/TIM.2021.3050154.

[14] M. Ghobakhloo, Industry 4.0, digitization, and opportunities for sustainability, J. Cleaner Prod. 252 (2020), https://doi.org/10.1016/j.jclepro.2019.119869.

[15] L. Guo, Y. Lei, S. Xing, T. Yan, N. Li, Deep convolutional transfer learning network: a new method for intelligent fault diagnosis of machines with unlabeled data, IEEE Trans. Ind. Electron. 66 (9) (2019) 7316–7325, https://doi.org/10.1109/TIE.2018.2877090.

[16] T. Heimann, P. Mountney, M. John, R. Ionasec, Real-time ultrasound transducer localization in fluoroscopy images by transfer learning from synthetic training data, Med. Image Anal. 18 (8) (2014) 1320–1328, https://doi.org/10.1016/j.media.2014.04.007.

[17] B. Henning, J. Rautenberg, Process monitoring using ultrasonic sensor systems, Ultrasonics 44 (SUPPL) (2006) e1395–e1399, https://doi.org/10.1016/j.ultras.2006.05.048.

[18] S. Hochreiter, J. Schmidhuber, Long short-term memory, Neural Comput. 9 (8) (1997) 1735–1780, https://doi.org/10.1162/neco.1997.9.8.1735.

[19] A.K. Jain, J. Mao, K.M. Mohiuddin, Artificial neural networks: a tutorial, Computer 29 (3) (1996) 31–44, https://doi.org/10.1109/2.485891.

[20] W.M. Kouw, M. Loog, A review of domain adaptation, IEEE Trans. Pattern Anal. Mach. Intell. (2019), https://doi.org/10.1109/TPAMI.2019.2945942.

[21] X. Li, W. Zhang, Q. Ding, Cross-domain fault diagnosis of rolling element bearings using deep generative neural networks, IEEE Trans. Ind. Electron. 66 (7) (2019) 5525–5534, https://doi.org/10.1109/TIE.2018.2868023.

[22] S.G. Mallat, A theory for multiresolution signal decomposition: the wavelet representation, IEEE Trans. Pattern Anal. Mach. Intell. 11 (1989) 674–693, https://doi.org/10.1109/34.192463.

[23] S.G. Mallat, C. Mallat, IV Time meets frequency, A Wavelet Tour Signal Process. 2 (1999) 67–124.

[24] S.G. Mallat, C. Mallat, 7.2 Classes of Wavelet Bases, Elsevier Science & Technology, 1999.

[25] Mathworks, Gradient, 2002a. Accessed May 27, 2020. https://uk.mathworks.com/help/matlab/ref/gradient.html#bvhp8_i.

[26] Mathworks, Movmean, 2002b. Accessed May 27, 2020. https://uk.mathworks.com/help/matlab/ref/movmean.html#bu2yug_-1_seealso.

[27] Mathworks, Choose a Multilayer Neural Network Training Function, 2020c. Accessed May 27, 2020. https://uk.mathworks. com/help/deeplearning/ug/choose-a-multilayer-neural-network-training-function.html;jsessionid= e378b9dfbf595a83f44348fc1e7c.

[28] D.J. McClements, Advances in the application of ultrasound in food analysis and processing, Trends Food Sci. Technol. 6 (9) (1995) 293–299, https://doi.org/10.1016/S0924-2244(00)89139-6.

[29] Q. Meng, J. Matthew, V.A. Zimmer, A. Gomez, D.F.A. Lloyd, D. Rueckert, B. Kainz, Mutual information-based disentangled neural networks for classifying unseen categories in different domains: application to fetal ultrasound imaging, IEEE Trans. Med. Imaging 40 (2) (2021) 722–734, https://doi.org/10.1109/TMI.2020.3035424.

[30] V. Mohammadi, M. Ghasemi-Varnamkhasti, R. Ebrahimi, M. Abbasvali, Ultrasonic techniques for the milk production industry, Meas.: J. Int. Meas. Confed. 58 (2014) 93–102, https://doi.org/10.1016/j.measurement.2014.08.022.

[31] M.T. Mohd Khairi, S. Ibrahim, M.A. Md Yunus, M. Faramarzi, Contact and non-contact ultrasonic measurement in the food industry: a review, Meas. Sci. Technol. 27 (1) (2015), https://doi.org/10.1088/0957-0233/27/1/012001.

[32] N. Munir, H.-J. Kim, J. Park, S.-J. Song, S.-S. Kang, Convolutional neural network for ultrasonic weldment flaw classification in noisy conditions, Ultrasonics 94 (2019) 74–81, https://doi.org/10.1016/j.ultras.2018.12.001.

[33] N. Munir, H.-J. Kim, S.-J. Song, S.-S. Kang, Investigation of deep neural network with drop out for ultrasonic flaw classification in weldments, J. Mech. Sci. Technol. 32 (7) (2018) 3073–3080, https://doi.org/10.1007/s12206-018-0610-1.

[34] K.S. Ojha, T.J. Mason, C.P. O'Donnell, J.P. Kerry, B.K. Tiwari, Ultrasound technology for food fermentation applications, Ultrason. Sonochem. 34 (2017) 410–417, https://doi.org/10.1016/j.ultsonch.2016.06.001.

[35] S.J. Pan, Q. Yang, A survey on transfer learning, IEEE Trans. Knowl. Data Eng. 22 (10) (2010) 1345–1359, https://doi.org/10.1109/TKDE.2009.191.

[36] V.M. Patel, R. Gopalan, R. Li, R. Chellappa, Visual Domain Adaptation: a survey of recent advances, IEEE Signal Process Mag. 32 (3) (2015) 53–69, https://doi.org/10.1109/MSP.2014.2347059.

[37] S.J. Pan, I.W. Tsang, J.T. Kwok, Q. Yang, Domain adaptation via transfer component analysis, IEEE Trans. Neural Networks 22 (2) (2011) 199–210, https://doi.org/10.1109/TNN.2010.2091281.

[38] L. Pettigrew, V. Blomenhofer, S. Hubert, F. Groß, A. Delgado, Optimisation of water usage in a brewery clean-in-place system using reference nets, J. Cleaner Prod. 87 (1) (2015) 583–593, https://doi.org/10.1016/j.jclepro.2014.10.072.

[39] K.-D. Thoben, S.A. Wiesner, T. Wuest, "Industrie 4.0" and smart manufacturing - a review of research issues and application examples, Int. J. Autom. Technol. 11 (1) (2017) 4–16, https://doi.org/10.20965/ijat.2017.p0004.

[40] E. Wallhäußer, A. Sayed, S. Nöbel, M.A. Hussein, J. Hinrichs, T. Becker, Determination of cleaning end of dairy protein fouling using an online system combining ultrasonic and classification methods, Food Bioprocess Technol. 7 (2) (2014) 506–515, https://doi.org/10.1007/s11947-012-1041-0.

[41] E. Wallhäußer, W.B. Hussein, M.A. Hussein, J. Hinrichs, T. Becker, Detection of dairy fouling: Combining ultrasonic measurements and classification methods, Eng. Life Sci. 13 (3) (2013) 292–301, https://doi.org/10.1002/elsc.201200081.

[42] M. Wang, W. Deng, Deep visual domain adaptation: a survey, Neurocomputing 312 (2018) 135–153, https://doi.org/10.1016/j.neucom.2018.05.083.

[43] K. Weiss, T.M. Khoshgoftaar, D.D. Wand, A survey of transfer learning, J. Big Data 3 (1) (2016) 9, https://doi.org/10.1186/s40537-016-0043-6.

[44] L. Wen, X. Li, L. Gao, Y. Zhang, A new convolutional neural network-based data-driven fault diagnosis method, IEEE Trans. Ind. Electron. 65 (7) (2018) 5990–5998, https://doi.org/10.1109/TIE.2017.2774777.

[45] K. Yan, A domain adaptation toolbox, 2020. Accessed June 14, 2020. https://www.github.com/viggin/domain-adaptation-toolbox.

[46] X. Zhan, S. Jiang, Y. Yang, L. Jian, T. Shi, X. Li, Inline measurement of particle concentrations in multicomponent suspensions using ultrasonic sensor and least squares support vector machines, Sensors (Basel, Switzerland) 15 (2015) 24109–24124, https://doi.org/10.3390/s150924109.

[47] W. Zhang, C. Li, G. Peng, Y. Chen, Z. Zhang, A deep convolutional neural network with new training methods for bearing fault diagnosis under noisy environment and different working load, Mech. Syst. Sig. Process. 100 (2018) 439–453, https://doi.org/10.1016/j.ymssp.2017.06.022.

[48] W. Zhang, T. Wu, Z. Li, S. Liu, A. Qiu, Y. Li, Y. Shi, Fracture recognition in ultrasonic logging images via unsupervised segmentation network, Earth Sci. Inf. (Article in press) (2021), https://doi.org/10.1007/s12145-021-00605-6.

[49] Y.-C. Zhu, A. AlZoubi, S. Jassim, Q. Jiang, Y. Zhang, Y.-B. Wang, X.-D. Ye, H. Du, A generic deep learning framework to classify thyroid and breast lesions in ultrasound images, Ultrasonics 110 (2021), 106300, https://doi.org/10.1016/j.ultras.2020.106300.

8 Domain Adaptation and Federated Learning for Ultrasonic Monitoring of Beer Fermentation

Article title: Domain Adaptation and Federated Learning for Ultrasonic Monitoring of Beer Fermentation

Journal: Fermentation

Date published: 1/11/2021

DOI: 10.3390/fermentation7040253

Authors: Bowler, A.L., Pound, M.P., Watson, N.J.

Author contributions (as published): Conceptualization, A.L.B., M.P.P. and N.J.W.; methodology, A.L.B., M.P.P. and N.J.W.; software, A.L.B.; validation, A.L.B.; formal analysis, A.L.B.; investigation, A.L.B. and N.J.W.; resources, A.L.B. and N.J.W.; data curation, A.L.B. and N.J.W.; writing—original draft preparation, A.L.B.; writing—review and editing, A.L.B., N.J.W. and M.P.P.; visualization, A.L.B.; supervision, N.J.W. and M.P.P.; project administration, N.J.W.; funding acquisition, N.J.W. All authors have read and agreed to the published version of the manuscript.

The aim of this thesis is to develop ML methods to facilitate optimal deployment of US sensors for process monitoring applications in industrial environments. This article contributed to the thesis objectives by comparing labelled domain adaptation methods to transfer knowledge for process monitoring using US sensors between different fermentation processes. This work and the previous section (Section 7, titled: “Transfer learning for process monitoring using reflection-mode ultrasonic sensing”) comprise the portion of this thesis that explores the development of unlabelled and labelled domain adaptation methods for process monitoring using US sensors. These methods leverage previously collected datasets which would negate or reduce the burden of collecting labelled data in industrial environments.

The work uses labelled domain adaptation methods to transfer ML models from a laboratory fermentation process to an industrial process. The domain adaptation approaches are used to minimise the data that must be collected for the industrial process. Compared with Section 7, simultaneous ML model training using labelled data for both tasks is possible. On reflection, domain adversarial training of the neural networks could have also been used to extract features discriminative as to the state of the fermentation but non-discriminative between the datasets (Ganin et al., 2016). This is achieved through training the ML predictions whilst simultaneously confusing a discriminator module as to whether the input data are from the source or target domain.

For both the laboratory and industrial datasets, the frequency of US data collection was the same. Blocks of US and temperature data were collected periodically (between each block of data collected, 200 s elapsed). Each of the blocks consisted of 36 US waveforms and 36 temperature readings. The time between each US waveform acquisition was 0.55 s. The laboratory scale fermentations lasted between 4 and 7 days whilst the industrial fermentation lasted between 6 and 7 days. Therefore, no pre-processing of the time series length was needed to use both datasets in the LSTM models.

In total, 14 US waveform features were inputted into the ML models. Zero-padding was applied to the US features to make every fermentation sequence equal to the maximum sequence length of 1556 timesteps. Therefore, the final datasets dimensions were 5x14x1556 for the industrial fermentations and 13x14x1556 for the lab-scale fermentations. The LSTMs were used for a sequence to sequence tasks where the sequence of features were used to predict the sequence of output values. At each timestep the model predicted the output value at the current timestep. At the following timestep, the model obtains the new US sensor measurement features from this timestep with which to predict the output value. Therefore, only US sensor measurement features are used to make predictions and previous ground truth data or output predictions are not used. All timesteps for each dataset were used as a single sequence rather than being truncated into multiple sequences of shorter length.

During training, all laboratory-scale fermentation batches were used. The number of industrial fermentation batches included in the training and validation set was varied between 1 and 4 (Table 1). A single industrial fermentation batch was used for validation with the number of validation folds being dependent on the number of industrial fermentation batches in the training and validation set (Table 1). The domain adaptation methodologies are compared with a model trained only on the industrial scale fermentation data, i.e., without using the laboratory scale data or domain adaptation. This is named the No DA model and is used as a base-case comparison. In the article, the column headings 1-4 in Table 3 indicated the number of industrial scale fermentation batches used in the training and validation set.

Table 1. Selected parameters for the domain adaptation networks depending on number of industrial scale fermentation batches in the training set.

Parameter	**Size of Training Set**			
Number of industrial scale fermentation batches in training and validation set	1	2	3	4
Number of industrial scale fermentation batches in test set	4	3	2	1
Number of validation folds	0	2	3	4

Federated learning achieved the highest model accuracy and performed best for 14 out of 16 ML tasks compared with the models not using knowledge from the source domain. This is likely due to the order of the dataset runs the model is trained on compared with the other two methodologies evaluated. During federated learning, the local models were trained for a single epoch on the corresponding domain dataset before passing the model weights for collation. This full epoch of training allows for increased gradient descent to an optimum point compared with a procedure that alternates between source and target domain data, as used for the other domain adaptation methods. This indicates that the order of the dataset runs may impact the ability of the models to find a better local optimum and should be trialled during the validation stage.

Federated learning could be utilized by beer manufacturers in a mutually beneficial agreement to leverage data from other breweries. This can be used to improve model accuracy when monitoring beer fermentation whilst maintaining dataset privacy. Firstly, this would require identifying participating breweries. As larger beer manufacturers would likely be able to create large fermentation datasets, it would be anticipated that smaller, craft

breweries with wider product ranges would be most interested in this collaborative set-up. Due to a lack of specialist knowledge to train ML models, most likely a research group or consultancy would be used to develop the models. Local models would be trained within the data collection platforms of each brewery to avoid the transferring of private data. The model weights would then be passed to a global model for aggregation. However, this approach would also benefit from techniques that minimised the number of iterations between local and global models. A method to achieve this could be to use pre-trained models (e.g. the final models developed in this work) to provide starting network weights to produce more similar local models and facilitate aggregation of the network weights into a global model.

8.1 References

Ganin, Y., Ustinova, E., Ajakan, H., Germain, P., Larochelle, H., Laviolette, F., Marchand, M., Lempitsky, V. (2016) ‘Domain-adversarial training of neural networks’ Journal of Machine Learning Research, 17, 1–35.

fermentation

MDPI

Article

Domain Adaptation and Federated Learning for Ultrasonic Monitoring of Beer Fermentation

Alexander L. Bowler [1], Michael P. Pound [2] and Nicholas J. Watson [1,*]

[1] Food, Water, Waste Research Group, Faculty of Engineering, University of Nottingham, University Park, Nottingham NG7 2RD, UK; alexander.bowler@nottingham.ac.uk
[2] School of Computer Science, Jubilee Campus, University of Nottingham, Nottingham NG8 1BB, UK; michael.pound@nottingham.ac.uk
* Correspondence: nicholas.watson@nottingham.ac.uk

Abstract: Beer fermentation processes are traditionally monitored through sampling and off-line wort density measurements. In-line and on-line sensors would provide real-time data on the fermentation progress whilst minimising human involvement, enabling identification of lagging fermentations or prediction of ethanol production end points. Ultrasonic sensors have previously been used for in-line and on-line fermentation monitoring and are increasingly being combined with machine learning models to interpret the sensor measurements. However, fermentation processes typically last many days and so impose a significant time investment to collect data from a sufficient number of batches for machine learning model training. This expenditure of effort must be multiplied if different fermentation processes must be monitored, such as varying formulations in craft breweries. In this work, three methodologies are evaluated to use previously collected ultrasonic sensor data from laboratory scale fermentations to improve machine learning model accuracy on an industrial scale fermentation process. These methodologies include training models on both domains simultaneously, training models in a federated learning strategy to preserve data privacy, and fine-tuning the best performing models on the industrial scale data. All methodologies provided increased prediction accuracy compared with training based solely on the industrial fermentation data. The federated learning methodology performed best, achieving higher accuracy for 14 out of 16 machine learning tasks compared with the base case model.

Keywords: ultrasonic measurements; fermentation; machine learning; federated learning; domain adaptation; long short-term memory

Citation: Bowler, A.L.; Pound, M.P.; Watson, N.J. Domain Adaptation and Federated Learning for Ultrasonic Monitoring of Beer Fermentation. *Fermentation* **2021**, *7*, 253. https://doi.org/10.3390/fermentation7040253

Academic Editor: Bernard Chen

Received: 6 October 2021
Accepted: 29 October 2021
Published: 1 November 2021

1. Introduction

Beer is one of the world's oldest and most widely consumed alcoholic beverages [1]. Beer fermentation processes are conventionally monitored through sampling and off-line wort density measurements [2]. This method is typically performed every couple of hours, requires manual operation, is time-consuming, and does not produce real-time results [3]. Automatic acquisition of real-time data pertaining to the fermenting wort would enable accurate process end point determination and identification of lagging fermentations. This would provide benefits of improved product consistency, fewer lost batches, time savings, and environmental benefits of less waste and less resource and energy use [3]. This can be achieved through in-line and on-line sensing techniques, where in-line methods directly measure properties of the fermenting wort and on-line methods use bypasses to automatically collect, analyse, and return samples to the vessel [4]. Furthermore, manufacturing is undergoing the fourth industrial revolution, where industrial digital technologies such as the Internet of Things (IoT), cloud computing and Machine Learning (ML) are implemented to integrate not only entire processes but also markets and supply chains [5]. This has the potential to increase the efficiency, productivity, product quality, and flexibility of manufacturing processes [5]. In-line and on-line sensors underpin this transformation by

collecting the real-time data to provide automatic decision- making and minimise human involvement [6]. Several in-line and on-line methods to monitor alcoholic fermentation have been investigated, such as near-infrared spectroscopy [3,7], Raman spectroscopy [8,9], mid-infrared spectroscopy [10], Fourier transform infrared spectroscopy [11], MEMS resonators [12], CO_2 emission monitoring [13], and ultrasonic (US) sensors [14–18]. Typically, these techniques use calibration techniques to correlate sensor data to material composition across the full range of process conditions (e.g., temperature) [3]. Conversely, ML can be used to map sensor data directly to target variables (such as classifying the stage of the fermentation process or predicting the time remaining until significant process milestones) without requiring extensive calibration procedures. Moreover, ML is able to fit complex non-linear relationships between multiple variables, or features, extracted from sensor readings. Furthermore, validation procedures encourage the development of models which accurately predict when process parameters are outside of the range they were trained on. Ultrasonic sensors have benefits of being low-cost, are non-invasive, small in size, have low energy consumption, and are able to characterise opaque materials. ML has previously been combined with US sensors to monitor fermentation processes. Hussein et al., (2012) used the US velocity, process temperature, and nine signal features extracted from the time and frequency domains to predict wort density using an artificial neural network [14]. Bowler et al., (2021) inputted time domain signal features into Long Short-Term Memory (LSTM) neural networks to predict the volume of alcohol percentage throughout fermentation [18].

ML methods require sufficient volumes of data for model training. However, fermentation processes can last for many days, imposing a significant time investment for data collection. Therefore, industrial fermentation monitoring using sensors and ML would benefit from using knowledge gained from previously monitored fermentation processes whether conducted in a laboratory or from other breweries. This would be of particular benefit to the growing craft breweries industry, where a wider range of beers are produced at smaller volumes, necessitating ML models which can be trained on fewer fermentation batches whilst being robust across different formulations of beer [19,20]. However, US sensor readings acquired from different fermentation vessels (different domains) present different data distributions to the ML models [21]. This can be due to differing US sensor contact between the two vessels, a difference in vessel construction affecting US waveform propagation, or differing waveform frequency distributions produced by the sensors [21]. Therefore, even for a similar fermentation task, the ML model trained on the source domain data will perform poorly when asked to make a prediction based on the target domain data. Domain adaptation is a subcategory of transfer learning which alters how the ML model is trained to predict accurately across both domains [22]. Unlabeled domain adaptation techniques can be used for tasks with no reference measurement available in the target domain to correlate input features to output variables during ML model training [21]. Conversely, labelled domain adaptation can be used for tasks where a reference measurement is obtainable. Common unlabeled domain adaptation techniques include minimising the distance between features from different domains using metrics such as the Maximum Mean Discrepancy [21,23–27], adversarial methods to confuse domain membership classifiers [28–32], generative methods to transform domain features [33–36], or Adaptive Batch Normalisation, which aligns the feature distributions across the domains for each batch [37,38]. Labelled domain adaptation can be achieved through either pre-training on the source domain and fine-tuning on the target domain, retraining the last few layers of a network using the target domain data, or by training using the data from both domains simultaneously [39]. While training ML models across fermentation processes from multiple breweries, the companies may not wish to share the US sensor data which could reveal information about their product formulation or process control strategies. In this case, federated learning may be used to share network weights from local models trained on an individual brewery's data to update a common global model as opposed to transferring the acquired sensor data and thus maintain privacy [40].

In this work, US sensor data acquired from a laboratory fermentation process is used to aid ML prediction on an industrial scale fermentation task. The industrial scale fermentations were monitored at a Small and Medium-sized Enterprise (SME) company, and so the data is of limited volume. Therefore, the laboratory scale dataset is used to improve ML model accuracy on these limited number of batches. The models are trained as multi-task networks to predict four outputs: classification of whether ethanol production has started, classification of whether ethanol production has ended, the time remaining until ethanol production begins, and the time remaining until ethanol production ends. Rather than using US sensor data to predict the wort density or alcohol by volume, this methodology directly predicts the most important information required from the fermentation process: whether the fermentation is lagging and determination of the fermentation end point.

Three domain adaptation methodologies are investigated. Firstly, labelled domain adaptation is used to simultaneously train the models on data from both domains. Simultaneous training on both domains is used as opposed to pre-training on the laboratory scale data and fine-tuning on the industrial scale data or retraining the last few layers of the network which are usually used for training convolutional layers in transfer learning for image recognition tasks. This is because, unlike convolutional filters which can detect features compared to a background of neighbouring pixels, the differences in feature magnitudes and trajectories in this work mean that features extracted in the source domain would not transfer to the target domain and the network would undergo catastrophic forgetting [41]. Secondly, the networks are also trained in a federated learning strategy to evaluate the impact of privacy preservation on ML model accuracy. Lastly, fine-tuning of the best performing models which have been trained on the source and target domains simultaneously are investigated again.

2. Materials and Methods

Two sets of fermentations were monitored: one in a 30 L laboratory scale vessel at the University of Nottingham and the second in a 2000 L industrial scale fermenter at the Totally Brewed brewery in Nottingham, UK. Full experimental details for the laboratory scale fermentations are included in [18]. The laboratory scale dataset consisted of 13 fermentations and the industrial scale dataset consisted of 5 fermentations. For the laboratory scale dataset, the same type and quantity of malt (Coopers Real Ale, Adelaide, Australia), yeast (Coopers Real Ale, Adelaide, Australia), sugar (brewing sugar, the Home Brew Shop, Farnborough, UK) and water (22 L) were used for all fermentations. For the industrial scale dataset, three different beers were monitored: three fermentations consisting of Slap in the Face, one Guardian of the Forest, and one 4 Hopmen of the Apocalypse. The same US probe was used to monitor both the laboratory and industrial scale fermentation processes (Figure 1). The US probe contained a US transducer (Sonatest, 2 MHz central frequency, Milton Keynes, UK) and a temperature sensor (RTD, PT1000, RS Components, Corby, UK). The US transducer was connected to a Lecouer Electronique US Box (Chuelles, France) that provided the excitation pulse to the transducer and digitised the received US signal. The temperature sensor was connected to a Pico electronic box (PT-104 Data Logger, Pico Technology, St Neots, UK). The two electronic boxes were connected to a laptop that controlled the data acquisition. Coupling gel was applied between the US transducer and the probe material, and a spring maintained the contact pressure. For the laboratory scale fermentations, a Tilt hydrometer provided real-time density measurements as a reference measurement of the fermentation progress and to provide labelled data for ML model training. For the industrial scale fermentations, samples were removed every two hours (except during night-time) and the wort density was measured using a hydrometer. For the industrial scale fermentations only, the temperature was decreased once the desired wort density was reached. Blocks of US and temperature data were collected periodically. Each of the blocks consisted of 36 US waveforms and 36 temperature readings. The US signal consisted of 7000 sampling points at 80 MHz sampling frequency. The time between each waveform acquisition was 0.55 s. Between each block of data collected, 200 s elapsed.

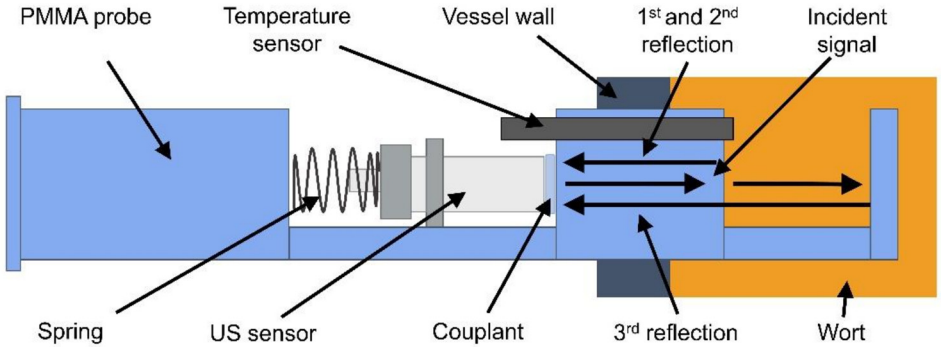


Figure 1. The probe consisting of US and temperature sensors and the paths of the received US sound wave reflections. Adapted from [18].

As depicted in Figure 1, the US transducer emitted sound waves which travelled along the PMMA probe material. At the interface between the probe material and the wort, a portion of the sound wave was reflected and the rest continued through the fermenting wort. Part of the reflected sound wave travelled through the probe-couplant boundary and was received by the transducer (the first reflection) whilst some reflected from this interface and repeated the previously described path (the second reflection). Therefore, the second reflection is a reverberation of the first reflection's path. The portion that passed through the fermenting wort was reflected at the opposite probe wall and travelled back to the transducer (the third reflection). An example of the US waveform recorded by the transducer is presented in Figure 2a. Each of the reflections in isolation are presented in Figure 2b–d. The start of the waveform (sample points <1000 in Figure 2a) was reflected back to the transducer before it contacted the probe-wort interface and therefore contains no useful information about the fermentation.

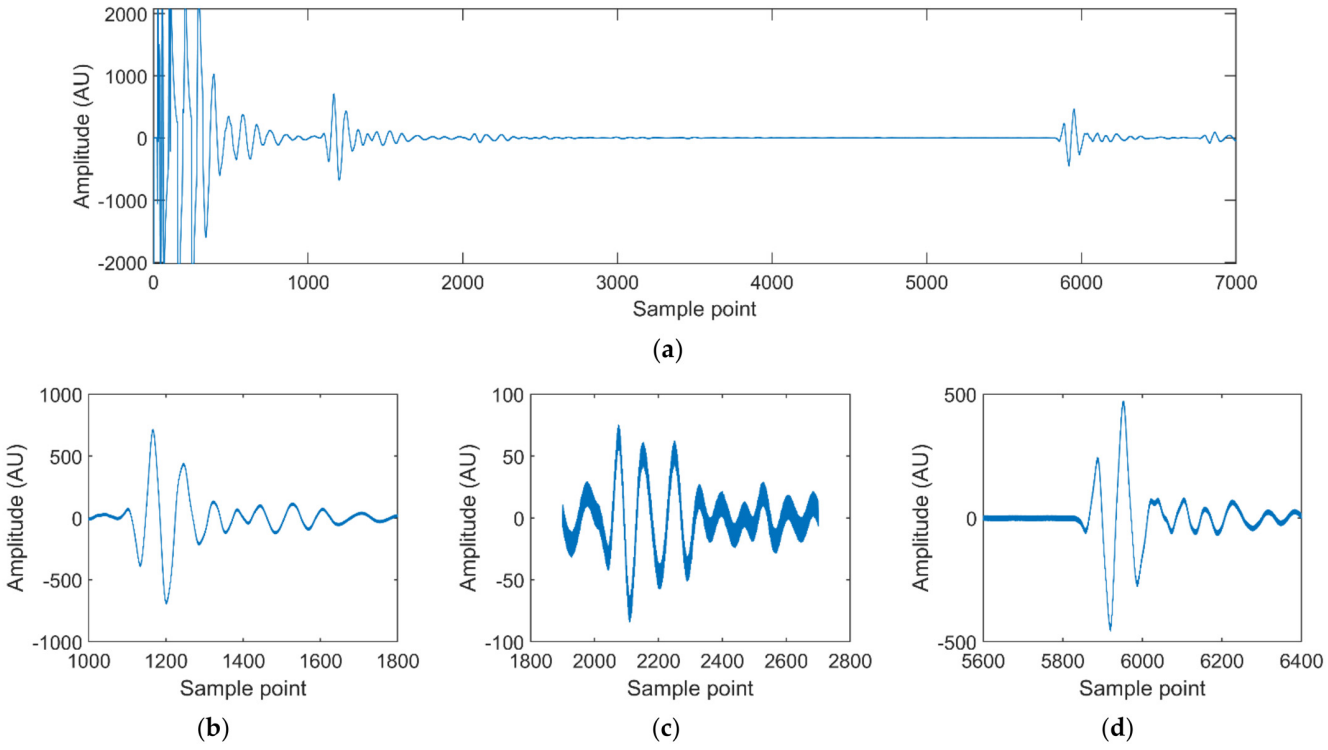


Figure 2. An example US waveform acquired: (**a**) The full waveform received; (**b**) the 1st reflection isolated; (**c**) the 2nd reflection isolated; and (**d**) the 3rd reflection isolated.

2.1. Ultrasonic Waveform Features

In total, 14 US waveform features were inputted into the ML models. Explanation of the calculation method and justification of the feature choices are provided in the following sections. In addition to the US waveform features, the process temperature was also used as an input. Although US sensors can accurately monitor fermentations without inclusion of the temperature as a feature [18], temperature sensors are already installed on most industrial vessels. As such, this data can be exploited in the ML models with no further effort in sensor installation or data collection.

2.1.1. Energy

The waveform energy is a measure of the total magnitude of the sound wave received by the transducer during an enveloped period. For the first reflection, this is a measure of the proportion of the sound wave reflected from the probe-wort interface and provides a measure of the changing wort density. Similarly, the energy of the second reflection is also dependent on the density of the fermenting wort in contact with the probe material. The energy of the third reflection is dependent on the previously discussed probe-wort boundary, the far wort-probe boundary, sound wave attenuation in the wort through which it travels, and the level of sound wave attenuation caused by CO_2 bubbles present in the wort [42].

$$\mathrm{E} = \sum_{i=start}^{i=end} {A_i}^2, \quad (1)$$

where E is the waveform energy, *Ai* is the waveform amplitude at sample point *i*, and start and end denote the range of samples points for the reflection of interest [43].

The waveform energy was the only feature selected from the oscillating part of the US waveform. Other features are commonly extracted to be used as ML model inputs, e.g., the peak-to-peak amplitude, maximum amplitude, minimum amplitude, skewness, kurtosis, and standard deviation [18,21]. However, previous work performing domain adaptation with US waveforms has shown that these additional features are unlikely to follow the same trend in both domains and their inclusion will degrade ML accuracy [21]. Therefore, only the waveform energy is used in this work as it is a measure of physical changes in the monitored wort.

2.1.2. Energy Standard Deviation

The standard deviation in the waveform energy was calculated across the 36 US waveforms obtained during each acquisition block. As CO_2 bubbles may be present in the wort through which the 3rd reflection travels, or on the probe surface affecting the 1st and 2nd reflections, the energy standard deviation monitors CO_2 formation throughout fermentation.

$$\mathrm{STD} = \sqrt{\frac{1}{W} \sum_{i=1}^{i=W} \left(E_i - \overline{E}\right)^2} \quad (2)$$

where STD is the standard deviation, *W* is the number of waveforms collected in the block, *i* is an individual waveform, and *E* is the mean waveform energy in the block.

2.1.3. Time of Flight

The time of flight was calculated using three different methods to overcome the noise and low amplitude signals present in the acquired US waveforms. Firstly, a thresholding method identified the earliest waveform sample point that rises above a predetermined value, and was calculated for all three reflections. A zero-crossing method identified the sample point where the waveform crosses zero after the threshold value had been reached, and this was also calculated for all three reflections. Finally, an auto-correlation method identified the sample point where the correlation between the first reflection and the subsequent reflections are determined to be most similar. The time of flight is a measure of the speed of sound through the materials, i.e., the probe material for the first and second

reflections (dependent on the temperature of the material) and the wort for the third reflection (dependent on wort temperature and density) [44].

2.2. Machine Learning

Multi-task deep neural networks consisting of a fully connected layer followed by an LSTM layer were used for all ML tasks. A summary of the three domain adaptation methods used is provided in Table 1. The fully connected layer enabled the creation of new features that are similar across both domains from combinations of the original inputs. The LSTM layer learns the trajectories of these modified features. The multi-task models were trained to simultaneously predict whether the production of ethanol had begun (classification), whether the production of ethanol had ended (classification), the time remaining until the start of ethanol production (regression), and the time remaining until ethanol production finishes (regression). In an industrial environment, this would provide benefits of identifying lagging fermentations by monitoring the start of ethanol production and estimating process end times by monitoring when ethanol production was complete. Multi-task learning is advantageous as it can allow for more effective process learning in the ML model when multiple metrics are desired whilst reducing the redundant information being stored [45]. Furthermore, multi-task learning is likely to reduce overfitting by preventing a single task from dominating the learning process.

LSTM layers in neural networks are able to retain information from previous timesteps in a sequence. LSTMs are a type of recurrent neural network that reduces the likelihood of vanishing or exploding gradients by using gate units. This enables their use over much longer sequences [46]. Zero-padding was applied to the US features to make every fermentation sequence equal to the maximum sequence length of 1556 timesteps. A masking layer designated that the LSTM units ignore this padding. All timesteps for each fermentation were used as a single sequence rather than being truncated into multiple sequences of shorter length. While long sequences (250–500 timesteps) are prone to producing vanishing gradients in LSTM layers when predicting a single output, this is not a concern when predicting an output at every timestep, as used in this work [47]. The input features from each dataset were independently normalised so that every feature ranged between 0 and 1 for both domains. This step aids domain adaptation capability by aligning the feature distributions from both domains, and is similar to the methodology used in [21].

A k-fold cross-validation procedure determined the optimal batch size, number of neurons in the fully connected layer, number of LSTM units, learning rate, L2 regularisation penalty, and number of epochs. As five industrial fermentation batches were monitored, the number of these fermentations used in the training set ranged from one to four, corresponding with the number of fermentations in the test set ranging from four to one (Table 2). Therefore, k was determined by the number of industrial fermentations present in the training set. For example, if only one fermentation was used in the training set, no cross-validation could be performed. However, when four fermentations were used, fourfold cross-validation was performed (Table 2).

The Adam optimisation algorithm and a gradient norm clipping value of 1 was used to reduce the likelihood of exploding gradients. The order of the training sets was shuffled after every epoch. The regression losses (mean squared error, Equation (3)) were multiplied by 0.1 to ensure their magnitudes were similar to the classification losses (binary cross-entropy, Equation (4)). This aided the network in learning both the classification and regression tasks. After cross-validation, the optimal hyperparameters which resulted in the lowest average validation error were used to train a final model using the entire training set. The networks were trained using TensorFlow 2.3.0. The coefficient of determination (R2), mean squared error (MSE), and mean absolute error (MAE) were used as performance metrics to evaluate the regression tasks during cross-validation. The accuracy, precision, and recall were used to evaluate the classification tasks during cross-validation. Evaluating multiple metrics provides a comprehensive assessment of a model's ability to fit to the

validation and test sets and facilitates improved comparison between models. In the results section, only the MAE and accuracy are discussed to aid clarity.

$$\text{BCE} = -\frac{1}{N}\sum_{i=1}^{N} y_i \cdot log\hat{y}_i + (1 - y_i) \cdot \log(1 - \hat{y}_i) \quad (3)$$

$$\text{MSE} = \frac{1}{N}\sum_{i=1}^{N} (y_i - \hat{y}_i)^2 \quad (4)$$

where BCE is the binary cross-entropy loss, MSE is the mean squared error loss, N is the number of samples, y is the target variable and $\hat{y}$ is the predicted value.

Table 1. Summary of the three domain adaptation machine learning methodologies investigated.

Method	Simultaneous Cross-Domain Training	Federated Learning	Fine-Tuning
Training datasets	Both source and target domain	Both source and target domain	Both source and target domain Followed by fine-tuning on target domain
Training strategy	Trained on both domains simultaneously	Trained on each domain sequentially	Either, depending on starting model used
Application	Transfer learning for laboratory data Transfer learning from other processes within the same company	Transfer learning between companies	Either, depending on starting model used
Advantages	More training options available as both datasets can be used simultaneously	Preserves privacy between domains	Either, depending on starting model used
Problem definition	Define N datasets $\{D_1, \ldots D_N\}$ used to train a ML model M_{DA}.	Define N data owners wishing to train a ML model M_{FED} using all their data $\{D_1, \ldots D_N\}$ without sharing the datasets and thus maintaining privacy.	Define N datasets $\{D_1, \ldots D_N\}$ used to train a ML model M_S. Define D_T as the target domain dataset (D_T included in $\{D1, \ldots DN\}$.
Algorithm	θ = model weights E = number of epochs Initialise θ_0 For i = 1 to E Iterate θ for 1 epoch using a combined dataset consisting of $D_1, \ldots D_N$. End	θ = model weights C = number of communication rounds w = weighting factor Initialise θ_0 For i = 1 to C Global model: $\theta_G = \Sigma\, w_j\theta_j$ Local models: For j = 1 to N Initialise $\theta_j = \theta_G$ Iterate θ_j for 1 epoch using D_j Return θ_j End End	θ = model weights E = number of epochs Initialise $\theta = \theta_S$ For i = 1 to E Iterate θ for 1 epoch using D_T End

In the domain adaptation case studied in this work, the source domain, D_S, and target domain, D_T, are different because the marginal probabilities of the features are different, $P_S(X) \neq P_T(X)$. Domain adaptation aims to improve model prediction accuracy on the target domain by altering how the model trains on the source domain. Three domain adaptation investigations were conducted; network training on both datasets simultaneously, network training in a federated learning set-up, and fine-tuning of the best performing previously trained networks on the target domain (industrial scale) dataset.

For the networks trained on both datasets simultaneously, the impact of dropout on the domain adaptation performance was evaluated. Dropout layers randomly remove neurons and their connections during training according to the designated probability [48]. Thus "thinned" networks are trained during each training batch encouraging more propagation paths through the network to be learned. Two dropout layers are used, one after the input layer and before the fully connected layer, and one after the fully connected layer and before the LSTM layer. The dropout layer probabilities were set to 0 or 0.5, producing four parameter combinations. Dropout was used to investigate whether it aided domain mixing in the network rather than certain neurons only learning a single domain and the remaining neurons co-adapting. There were more fermentation batches in the laboratory scale dataset compared to the industrial scale dataset. As such, to ensure both domains were learned, the frequency of the industrial dataset in the training set was increased. For example, when a single industrial fermentation batch was present in the training set, this was passed to the network 13 times during one epoch. Similarly, when four industrial fermentation batches were present, each was used three times during training for each epoch (Table 2).

Table 2. Selected parameters for the domain adaptation networks depending on number of industrial scale fermentation batches in the training set.

Parameter	Size of Training Set			
Number of industrial scale fermentation batches in training set	1	2	3	4
Number of industrial scale fermentation batches in test set	4	3	2	1
Number of validation folds	0	2	3	4
Number of industrial fermentation batch occurrences per epoch when training on both domains simultaneously	13	6	4	3
Industrial dataset weighting factor for federated learning	0.9	0.85	0.8	0.75

For the federated learning investigations, local models were trained on each dataset and a weighting factor was applied to the resulting local network weights before being summed to produce a global model. The global model weights were used as the initialisation weights for the next epoch of local network training. After training, the global model was evaluated on the test set. The weighting factors were changed depending on the number of industrial fermentation runs present in the training set. I.e., 0.9 for the industrial scale data local model and 0.1 for the laboratory scale model when a single industrial fermentation run was present in the training data, and 0.75 and 0.25 when four industrial fermentation runs were used in the training data (Table 2).

Finally, fine-tuning the best performing models on the target domain data was assessed. As the models are used to monitor the industrial scale fermentations, the final models do not need to be accurate on the source domain laboratory scale fermentations. Therefore, after initial training to transfer knowledge from the source domain, fine-tuning on the target domain can increase model accuracy of the industrial scale data. All network weights were tuned. Preliminary investigations froze the model weights for the fully connected and LSTM layers and only tuned the output layers. However, this resulted in lower accuracy models on the validation sets than when all weights could be updated.

These domain adaptation methodologies are compared with a model trained only on the industrial scale fermentation data, i.e., without using the laboratory scale data or domain adaptation. This is named the No DA model and is used as a base-case comparison.

3. Results

3.1. Ultrasonic Measurements

Figure 3a–f displays the US feature and temperature results for the industrial scale fermentations. Full discussion of the US feature and temperature results for the laboratory

dataset are included in [18]. A comparison between the two datasets is provided in the text. For the industrial scale dataset, the process temperature was decreased after the desired wort density had been reached, determined through off-line sampling and hydrometer measurements. As such, Figure 3b-f display the results until one day after the temperature was decreased so that the US feature changes during ethanol production are clearly presented. The results show that the time of flight for the third reflection decreased, corresponding to an increase in the speed of sound, during ethanol production for all fermentations (Figure 3f). This agrees with [14,15] but contradicts the results found in [16,17,49] which monitored a decreasing speed of sound throughout fermentation. The reason for this is likely because [14,15] monitored an industrial fermentation process, similar to the industrial scale dataset in this work, whereas [16,17,49] monitored a small laboratory scale process (250 cm^3). Therefore, the specific combination of water, ethanol, sugar, yeast, and CO_2 concentrations present in industrial processes may produce an increasing speed of sound during ethanol production. Overall, the energy of the first reflection increases during ethanol production (Figure 3c), as found in [18]. This indicates an increase in acoustic impedance mismatch at the probe-wort interface. As the acoustic impedance is a product of the material density and speed of sound, this shows that the decreasing wort density has a larger impact than the increasing speed of sound on the wort acoustic impedance [42]. The energy of the third reflection shows no general trend during ethanol production (Figure 3d) indicating that the reduced sound wave proportion travelling through the first buffer-wort interface is offset by the increased sound wave reflection at the far wort-buffer interface. The third reflection energy displays increased variation over the first reflection energy due to sound wave attenuation in the presence of CO_2 bubbles, similar to the results found in [17,18]. In contrast, the laboratory scale data shows no trend in the speed of sound during fermentation and the third reflection energy follows a similar profile to the first reflection [18]. This is likely due to these effects being masked due to the varying temperature during ethanol production for the laboratory scale dataset, whereas the temperature was controlled during this period for the industrial fermentations. Figure 4 displays the first reflection energy for the first five fermentations from the laboratory dataset. The differing feature magnitudes and trajectories compared with Figure 3c showcases the need for domain adaptation techniques.

3.2. Machine Learning

Figure 5a,c,e and Figure 5b,d,f display the classification accuracies for the beginning of ethanol production and end of ethanol production for the trained networks, respectively. Although the multi-task networks were also trained to predict the time remaining until (and had passed since) the start and end of ethanol production, the regression predictions are most useful close to the classification boundaries. For example, an accurate prediction of the time since ethanol production started is not needed near the end of the fermentation process, or an approximate time for when ethanol production will end would not be useful when the fermentation is lagging and never begins. Therefore, the classification results are most valuable when evaluating the utility of the trained model. Furthermore, due to the multi-task nature of the model, the accuracy of the classification results correlates with the ability to learn the regression tasks close to the classification boundaries. As such, only the classification results are included in the presented graphs. However, the regression accuracies are presented in Table 3 and discussed in the text.

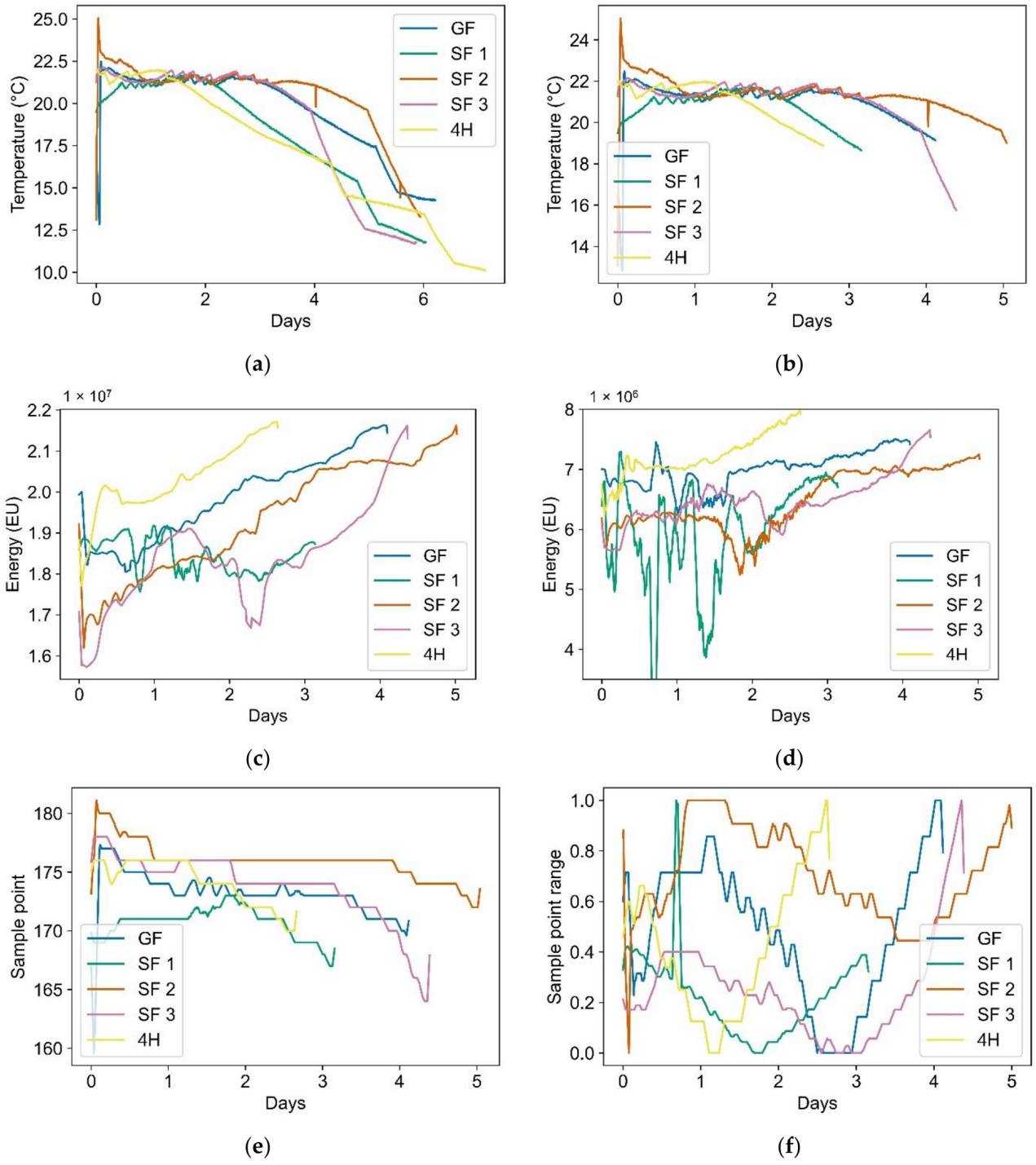


Figure 3. US feature and temperature results for the five industrial fermentation batches. (**a**) The process temperature. (**b**) The process temperature until one day post the end of ethanol production. (**c**) The first reflection energy until one day post the end of ethanol production. (**d**) The third reflection energy until one day post the end of ethanol production. (**e**) The first reflection time of flight measured using a thresholding method until one day post the end of ethanol production. (**f**) The third reflection time of flight measured using a thresholding method until one day post the end of ethanol production.

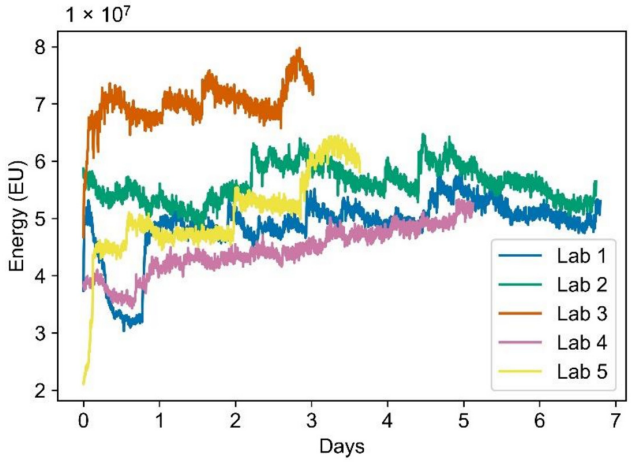


Figure 4. The first reflection energy for the first five laboratory scale fermentation batches.

Figure 5a,b display the results for the networks which were trained on the source and target domain data simultaneously. Preliminary investigations determined that the 0.5, 0.5 dropout rate models failed to train accurately for all training set sizes. Models with 0.5, 0 dropout rates produced inconsistent results, with some models accurately predicting using the test set data and some models performing worse than the model trained on only the industrial scale fermentations (No DA). However, the 0, 0 and 0, 0.5 models achieved higher accuracy than the No DA model for six out of eight classification tasks. Furthermore, the 0, 0 model achieved lower MAE for seven out of eight regression tasks compared to the No DA model. Therefore, the 0, 0 and 0, 0.5 dropout rates were used for subsequent investigations and the results of these models are presented in Figure 5a–f and Table 3. These higher accuracy results for the domain adaptation models prove that using the laboratory scale data to train the networks benefits the predictions on the industrial scale dataset.

Figure 5c,d display results for the models trained in a federated learning strategy. The two federated models are trained using the best performing dropout probabilities determined from the previous investigation and are compared with the No DA baseline results. The 0, 0 model achieved higher classification accuracies and lower MAE for six out of eight classification and regression tasks than the No DA model. When using four industrial scale fermentation batches in the training set, the 0, 0 model reached accuracies of 99.8% and 99.9% for predicting the start and end of ethanol production, respectively. Furthermore, the 0, 0.5 models achieved better results for seven out of eight of the classification and regression tasks. Overall, the federated learning models were more accurate than their corresponding non-federated training models using the same dropout probabilities, achieving higher classification accuracies on eight tasks compared to seven for the non-federated learning models. Similarly, the federated learning models achieved lower MAEs on 10 regression tasks compared with five for the non-federated learning models. This is an encouraging result as it indicates that not only can federated training provide benefits over models that train without the laboratory scale data, but that they can also perform better than conventionally trained domain adaptation networks in addition to maintaining data privacy. The reason for this may be the increased model learning afforded in the industrial scale dataset local model. During training, this model learns from an epoch full of the industrial scale training dataset compared with the non-federated model which only learns from the industrial scale target domain intermittently between source domain fermentation runs. This increased learning without switching between domains may allow the network weights to travel further towards local optima for the industrial scale dataset in each epoch. This contrasts with results presented in the wider literature, where federated learning degraded model accuracy compared with non-federated learning by 3.3% [50], 1.66% [51], and <10% [52].

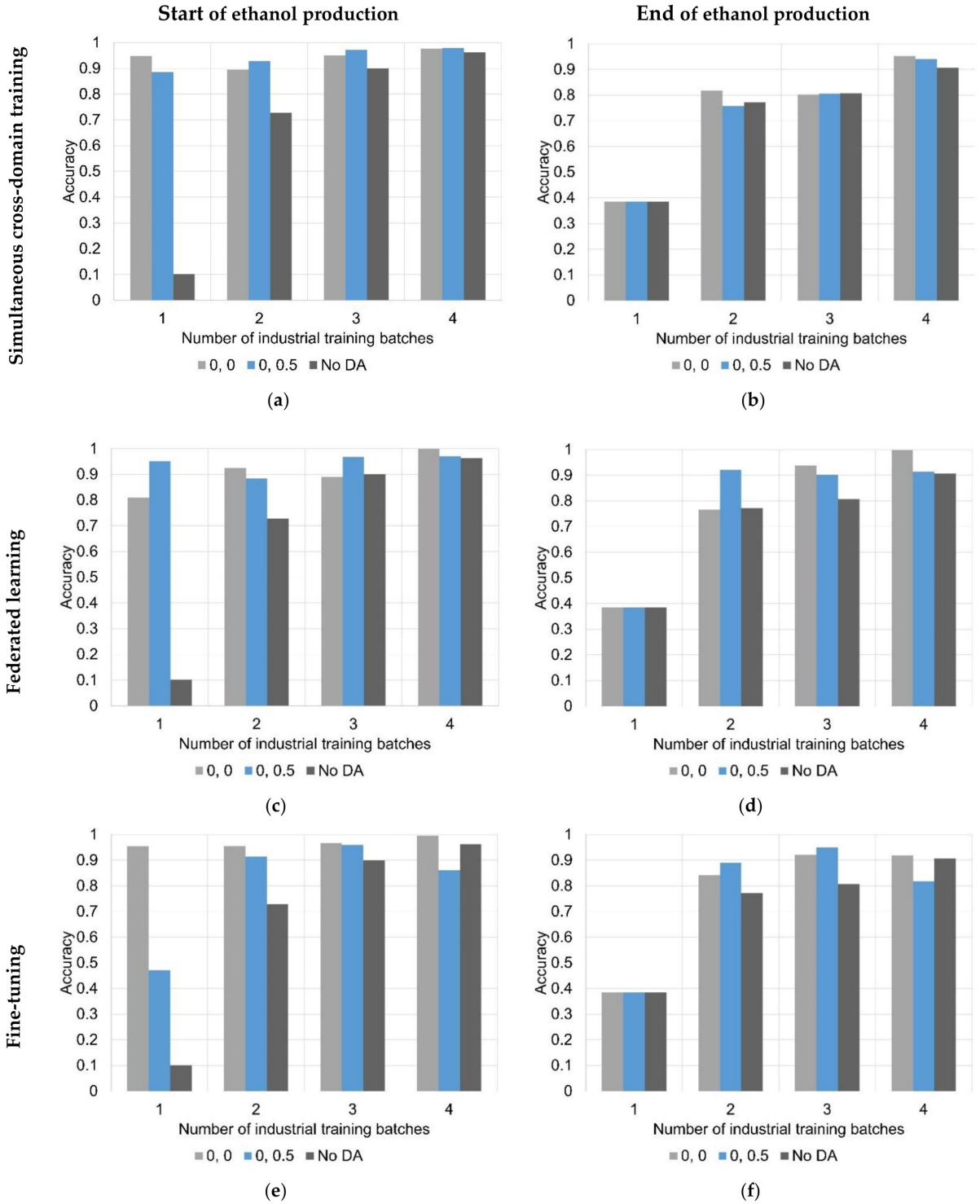


Figure 5. The classification results on the industrial scale fermentations test set. The numbers in the legend indicate the dropout layer probability for the two dropout layers. E.g., 0, 0 indicates a dropout probability of zero in both layers. (**a**) Classification results for the start of ethanol production for the networks trained on both domain datasets simultaneously. (**b**) Classification results for the end of ethanol production for the networks trained on both domain datasets simultaneously. (**c**) Classification results for the start of ethanol production for the networks trained using federated learning. (**d**) Classification results for the end of ethanol production for the networks trained using federated learning. (**e**) Classification results for the start of ethanol production for the federated training networks fine-tuned on the industrial scale dataset. (**f**) Classification results for the end of ethanol production for the federated training networks fine-tuned on the industrial scale dataset.

Figure 5e,f display the classification results for the previously discussed federated models fine-tuned on the industrial dataset. While still providing improvements over the No DA base case, achieving higher classification accuracies for 12 out of 16 tasks, their accuracy is reduced over the starting federated learning models. This is most likely due to the fine-tuning method overfitting during training. The reason for this is the large network size required to learn both domains in the starting models. For example, the No DA models had a maximum optimum number of eight neurons in the fully connected layer and four LSTM units to learn only the target domain. However, the federated learning models required a maximum of 128 neurons in the fully connected layer and eight LSTM units to fit to both dataset domains. Therefore, when fine-tuning on the industrial dataset after fitting to both domains, the model begins to overfit, especially when four industrial batches are used in the training set.

Table 3. The regression accuracies of each of the models for predicting the time remaining until the start and end of ethanol production, where MAE is the Mean Absolute Error of the prediction. The base-line model was trained using only data from the industrial fermentations. The numbers in the Model column indicate the dropout probability used in each dropout layer. E.g., 0,0 represents 0 dropout probability in both layers.

Method	Model	Start of Ethanol Production Accuracy (MAE)				End of Ethanol Production Accuracy (MAE)			
		1	2	3	4	1	2	3	4
Base-line model	No DA	2.769	1.099	0.646	0.710	2.035	1.278	1.047	0.534
Conventional domain adaptation	0, 0	1.942	0.93	0.423	0.541	1.767	0.980	0.950	0.681
	0, 0.5	3.326	1.528	0.836	0.171	7.29	2.027	1.528	0.920
Federated Learning	0, 0	2.496	0.540	0.431	0.726	3.884	1.133	0.599	0.351
	0, 0.5	2.482	0.423	0.520	0.296	3.073	1.089	0.937	0.663
Fine-tuning	0, 0	2.536	0.485	0.334	0.402	4.998	0.833	0.517	1.061
	0, 0.5	3.376	0.514	0.338	0.416	5.110	0.837	0.64	1.451

3.3. Future Research Directions

Overall, transferring knowledge from the source domain increased model accuracy when applied to the target domain data. Using more than two datasets could increase this benefit further, especially using more similar datasets, e.g., from multiple industrial fermentation processes. The two datasets used in this work had distinct differences. For example, no temperature control on the laboratory scale dataset and an increasing time of flight during fermentation for the industrial scale dataset. It is anticipated that more similar datasets would provide even greater benefits. Furthermore, other than increasing model accuracy, the domain adaptation methodology can also reduce the time for ML model development. After training across two domains, the final models could be used to predict using data from a new fermentation process without having been trained on this new domain. However, incorporation of a small number of batches from this new fermentation process would be expected to aid model accuracy.

In this work, the waveform energy was the single feature used to describe the oscillating part of the US waveform. The reason for this was that previous work demonstrated that multiple oscillating waveform features are unlikely to follow similar trends across domains and their inclusion would degrade model accuracy [21]. However, for many applications of ML and US sensors, multiple features may need to be used to accurately monitor changes in this portion of the US waveform. In this case, the methodologies presented in this work may be used to obtain predictions on the target domain data from models trained on both the source and target domains. These predictions can then be used as an additional feature in a model only trained on the target domain data. In this way, other features describing the oscillating part of the waveform can be used as no domain adaptation is required while also incorporating knowledge from the source domain.

The combination of ML and US measurements should be used in further research over calibration procedures. In this work, the speed of sound increased during fermentation, agreeing with [14,15], which were conducted at large scale, but contradicting [16,17,49], which were conducted at small scale. This indicates that there is a discrepancy in the speed of sound trend at the ethanol, sugar, yeast, and CO_2 concentrations and temperature used at small and large scales. Therefore, extensive and complicated calibration procedures would need to be used to account for these effects. In addition, ML offers several distinct advantages: it negates the need for these complex calibration procedures accounting for all the parameters previously listed; more information from the waveforms is typically used through feature extraction; more complex fitting procedures are used, allowing for increased prediction accuracy; and validation procedures encourage model accuracy even on process parameters outside the range the model was trained on.

Acceptable ML model accuracy is dependent on its desired application. In this work, the highest accuracy model (federated learning, zero dropout, four industrial training batches) achieved 99.8% and 99.9% for predicting the start and end of ethanol production, respectively. This is equivalent to the current method of determination, off-line wort density measurements using hydrometers, which are only conducted once every several hours (or even less frequently overnight) and have reduced accuracy when foam is present. However, these model accuracies were obtained using only a single test set batch and therefore a large dataset size would be needed to determine whether these accuracies were consistent.

US measurements and ML could also be used in combination with sampling methods to reduce the amount of sampling required (and therefore also reducing operator burden), provide timely results between samples (for example, overnight), and predict when fermentation stages will be reached to improve plant scheduling. In this case, ML models can be continuously updated using the labelled data from the sample measurements. If US sensors are desired to eliminate the use of sampling, higher accuracy models would be required and longer model development times would be needed. In addition, a model that stated a confidence level of its prediction would increase trust in the model by identifying when sample measurements should be used as a safeguard.

4. Conclusions

This work has used previously collected US sensor data from laboratory scale fermentations to improve ML model accuracy on an industrial scale process. Overall, all methodologies led to improvements in model accuracy over training on the target domain alone. The federated learning methodology performed best, achieving higher accuracy for 14 out of 16 machine learning tasks compared with the base case model, and achieving around 100% test set accuracy when trained on four industrial datasets and no dropout was used. Federated learning improved model accuracy over the traditional simultaneous domain training by allowing increased tuning of the network weights to converge on local target domain optima. However, fine-tuning led to a decrease in model accuracy due to overfitting of networks caused by the larger number of neurons and LSTM units needed to accurately train on both domains. The methodologies investigated not only provide increased accuracy, but also speed up model development time by reducing the number of fermentation runs required to be monitored in the target domain.

Author Contributions: Conceptualization, A.L.B., M.P.P. and N.J.W.; methodology, A.L.B., M.P.P. and N.J.W.; software, A.L.B.; validation, A.L.B.; formal analysis, A.L.B.; investigation, A.L.B. and N.J.W.; resources, A.L.B. and N.J.W.; data curation, A.L.B. and N.J.W.; writing—original draft preparation, A.L.B.; writing—review and editing, A.L.B., N.J.W. and M.P.P.; visualization, A.L.B.; supervision, N.J.W. and M.P.P.; project administration, N.J.W.; funding acquisition, N.J.W. All authors have read and agreed to the published version of the manuscript.

Funding: This work was supported by the Engineering and Physical Sciences Research Council (EPSRC) standard research studentship (EP/R513283/1) and EPSRC network+ Connected Everything (EP/P001246/1).

Data Availability Statement: The researchers at the University of Nottingham can be contacted for access to data.

Conflicts of Interest: The authors declare no conflict of interest.

References

1. Grassi, S.; Amigo, J.M.; Lyndgaard, C.B.; Foschino, R.; Casiraghi, E. Beer fermentation: Monitoring of process parameters by FT-NIR and multivariate data analysis. *Food Chem.* **2014**, *155*, 279–286. [CrossRef]
2. Jan, M.V.S.; Guarini, M.; Guesalaga, A.; Pérez-Correa, J.R.; Vargas, Y.; Perez-Correa, J. Ultrasound based measurements of sugar and ethanol concentrations in hydroalcoholic solutions. *Food Control* **2008**, *19*, 31–35. [CrossRef]
3. Vann, L.; Layfield, J.B.; Sheppard, J.D. The application of near-infrared spectroscopy in beer fermentation for online monitoring of critical process parameters and their integration into a novel feedforward control strategy. *J. Inst. Brew.* **2017**, *123*, 347–360. [CrossRef]
4. De Beer, T.; Burggraeve, A.; Fonteyne, M.; Saerens, L.; Remon, J.; Vervaet, C. Near infrared and Raman spectroscopy for the in-process monitoring of pharmaceutical production processes. *Int. J. Pharm.* **2011**, *417*, 32–47. [CrossRef] [PubMed]
5. Zhong, R.Y.; Xu, X.; Klotz, E.; Newman, S.T. Intelligent Manufacturing in the Context of Industry 4.0: A Review. *Engineering* **2017**, *3*, 616–630. [CrossRef]
6. Ghobakhloo, M. Industry 4.0, digitization, and opportunities for sustainability. *J. Clean. Prod.* **2020**, *252*, 119869. [CrossRef]
7. Corro-Herrera, V.A.; Gómez-Rodríguez, J.; Hayward-Jones, P.M.; Barradas-Dermitz, D.M.; Gschaedler-Mathis, A.C.; Aguilar-Uscanga, M.G. Real-time monitoring of ethanol production during Pichia stipitis NRRL Y-7124 alcoholic fermentation using transflection near infrared spectroscopy. *Eng. Life Sci.* **2018**, *18*, 643–653. [CrossRef]
8. Wang, Q.; Li, Z.; Ma, Z.; Liang, L. Real time monitoring of multiple components in wine fermentation using an on-line auto-calibration Raman spectroscopy. *Sens. Actuators B Chem.* **2014**, *202*, 426–432. [CrossRef]
9. Schalk, R.; Frank, R.; Rädle, M.; Methner, F.-J.; Beuermann, T.; Braun, F.; Gretz, N. Non-contact Raman spectroscopy for in-line monitoring of glucose and ethanol during yeast fermentations. *Bioprocess Biosyst. Eng.* **2017**, *40*, 1519–1527. [CrossRef]
10. Mazarevica, G.; Diewok, J.; Baena, J.R.; Rosenberg, E.; Lendl, B. On-Line Fermentation Monitoring by Mid-Infrared Spectroscopy. *Appl. Spectrosc.* **2004**, *58*, 804–810. [CrossRef]
11. Veale, E.; Irudayaraj, J.; Demirci, A. An On-Line Approach to Monitor Ethanol Fermentation Using FTIR Spectroscopy. *Biotechnol. Prog.* **2007**, *23*, 494–500. [CrossRef]
12. Toledo, J.; Ruiz-Díez, V.; Pfusterschmied, G.; Schmid, U.; Sánchez-Rojas, J. Flow-through sensor based on piezoelectric MEMS resonator for the in-line monitoring of wine fermentation. Sens. *Actuators B Chem.* **2018**, *254*, 291–298. [CrossRef]
13. Ete-Carmona, E.C.; Gallego-Martinez, J.-J.; Martin, C.; Brox, M.; Luna-Rodriguez, J.-J.; Moreno, J. A Low-Cost IoT Device to Monitor in Real-Time Wine Alcoholic Fermentation Evolution through CO2 Emissions. *IEEE Sens. J.* **2020**, *20*, 6692–6700. [CrossRef]
14. Hussein, W.B.; Hussein, M.A.; Becker, T. Robust spectral estimation for speed of sound with phase shift correction applied online in yeast fermentation processes. *Eng. Life Sci.* **2012**, *12*, 603–614. [CrossRef]
15. Hoche, S.; Krause, D.; Hussein, M.A.; Becker, T. Ultrasound-based, in-line monitoring of anaerobe yeast fermentation: Model, sensor design and process application. *Int. J. Food Sci. Technol.* **2016**, *51*, 710–719. [CrossRef]
16. Resa, P.; Elvira, L.; De Espinosa, F.M. Concentration control in alcoholic fermentation processes from ultrasonic velocity measurements. *Food Res. Int.* **2004**, *37*, 587–594. [CrossRef]
17. Resa, P.; Elvira, L.; De Espinosa, F.M.; González, R.; Barcenilla, J. On-line ultrasonic velocity monitoring of alcoholic fermentation kinetics. *Bioprocess Biosyst. Eng.* **2008**, *32*, 321–331. [CrossRef]
18. Bowler, A.; Escrig, J.; Pound, M.; Watson, N. Predicting Alcohol Concentration during Beer Fermentation Using Ultrasonic Measurements and Machine Learning. *Fermentation* **2021**, *7*, 34. [CrossRef]
19. Donadini, G.; Porretta, S. Uncovering patterns of consumers' interest for beer: A case study with craft beers. *Food Res. Int.* **2017**, *91*, 183–198. [CrossRef]
20. Gatrell, J.; Reid, N.; Steiger, T.L. Branding spaces: Place, region, sustainability and the American craft beer industry. *Appl. Geogr.* **2018**, *90*, 360–370. [CrossRef]
21. Bowler, A.L.; Watson, N.J. Transfer learning for process monitoring using reflection-mode ultrasonic sensing. *Ultrasonics* **2021**, *115*, 106468. [CrossRef] [PubMed]
22. Kouw, W.M.; Loog, M. A Review of Domain Adaptation without Target Labels. *IEEE T Pattern Anal.* **2021**, *43*, 766–785. [CrossRef] [PubMed]
23. Li, X.; Zhang, W.; Ding, Q.; Sun, J.-Q. Multi-Layer domain adaptation method for rolling bearing fault diagnosis. *Signal Process.* **2019**, *157*, 180–197. [CrossRef]
24. Li, X.; Zhang, W.; Ding, Q. A robust intelligent fault diagnosis method for rolling element bearings based on deep distance metric learning. *Neurocomputing* **2018**, *310*, 77–95. [CrossRef]
25. Guo, L.; Lei, Y.; Xing, S.; Yan, T.; Li, N. Deep Convolutional Transfer Learning Network: A New Method for Intelligent Fault Diagnosis of Machines with Unlabeled Data. *IEEE T. Ind. Electron.* **2019**, *9*, 7316–7325. [CrossRef]
26. Lu, W.; Liang, B.; Cheng, Y.; Meng, D.; Yang, J.; Zhang, T. Deep Model Based Domain Adaptation for Fault Diagnosis. *IEEE T. Ind. Electron.* **2017**, *3*, 2296–2305. [CrossRef]

27. Geng, B.; Tao, D.; Xu, C. DAML: Domain adaptation metric learning. *IEEE T. Image Process.* **2011**, *10*, 2980–2989. [CrossRef]
28. Tzeng, E.; Hoffman, J.; Saenko, K.; Darrell, T. Adversarial discriminative domain adaptation. *Proc. CVPR IEEE* **2017**, *2017*, 2962–2971. [CrossRef]
29. Zhang, W.; Ouyang, W.; Li, W.; Xu, D. Collaborative and Adversarial Network for Unsupervised Domain Adaptation. *Proc. CVPR IEEE* **2018**, *2018*, 3801–3809. [CrossRef]
30. Zhang, Y.; Qiu, Z.; Yao, T.; Liu, D.; Mei, T. Fully Convolutional Adaptation Networks for Semantic Segmentation. *Proc. CVPR IEEE* **2018**, *2018*, 6810–6818. [CrossRef]
31. Tsai, Y.-H.; Hung, W.-C.; Schulter, S.; Sohn, K.; Yang, M.-H.; Chandraker, M. Learning to Adapt Structured Output Space for Semantic Segmentation. *Proc. CVPR IEEE.* **2018**, *2018*, 7472–7481. [CrossRef]
32. Chen, W.; Wang, H.; Li, Y.; Su, H.; Wang, Z.; Tu, C.; Lischinski, D.; Cohen-Or, D.; Chen, B. Synthesizing training images for boosting human 3D pose estimation. *Proc. 3DV* **2016**, *2016*, 479–488. [CrossRef]
33. Sankaranarayanan, S.; Balaji, Y.; Castillo, C.D.; Chellappa, R. Generate to Adapt: Aligning Domains Using Generative Adversarial Networks. Proc. *CVPR IEEE* **2018**, *2018*, 8503–8512. [CrossRef]
34. Sankaranarayanan, S.; Balaji, Y.; Jain, A.; Lim, S.N.; Chellappa, R. Learning from Synthetic Data: Addressing Domain Shift for Semantic Segmentation. *Proc. CVPR IEEE* **2018**, *2018*, 3752–3761. [CrossRef]
35. Bousmalis, K.; Silberman, N.; Dohan, D.; Erhan, D.; Krishnan, D. Unsupervised pixel-level domain adaptation with generative adversarial networks. *Proc. CVPR IEEE* **2017**, *2017*, 95–104. [CrossRef]
36. Bousmalis, K.; Irpan, A.; Wohlhart, P.; Bai, Y.; Kelcey, M.; Kalakrishnan, M.; Downs, L.; Ibarz, J.; Pastor, P.; Konolige, K.; et al. Using Simulation and Domain Adaptation to Improve Efficiency of Deep Robotic Grasping. *IEEE Int. Conf. Robot.* **2018**, *2018*, 4243–4250. [CrossRef]
37. Zhang, W.; Peng, G.; Li, C.; Chen, Y.; Zhang, Z. A new deep learning model for fault diagnosis with good anti-noise and domain adaptation ability on raw vibration signals. *Sensors* **2017**, *17*, 425. [CrossRef]
38. Du, Y.; Jin, W.; Wei, W.; Hu, Y.; Geng, W. Surface EMG-based inter-session gesture recognition enhanced by deep domain adaptation. *Sensors* **2017**, *17*, 458. [CrossRef]
39. Han, Y.; Yoo, J.; Kim, H.H.; Sin, H.J.; Sung, K.; Ye, J.C. Deep learning with domain adaptation for accelerated projection-reconstruction MR. *Magn. Reson. Med.* **2018**, *80*, 1189–1205. [CrossRef]
40. Yang, Q.; Liu, Y.; Chen, T.; Tong, Y. Federated machine learning: Concept and applications. *ACM T. Intel. Syst. Tec.* **2019**, *10*, 12. [CrossRef]
41. Kirkpatrick, J.; Pascanu, R.; Rabinowitz, N.; Veness, J.; Desjardins, G.; Rusu, A.A.; Milan, K.; Quan, J.; Ramalho, T.; Grabska-Barwinska, A.; et al. Overcoming catastrophic forgetting in neural networks. *Proc. Natl. Acad. Sci. USA* **2017**, *114*, 3521–3526. [CrossRef]
42. McClements, D. Advances in the application of ultrasound in food analysis and processing. *Trends Food Sci. Technol.* **1995**, *6*, 293–299. [CrossRef]
43. Zhan, X.; Jiang, S.; Yang, Y.; Liang, J.; Shi, T.; Li, X. Inline Measurement of Particle Concentrations in Multicomponent Suspensions using Ultrasonic Sensor and Least Squares Support Vector Machines. *Sensors* **2015**, *15*, 24109–24124. [CrossRef]
44. Henning, B.; Rautenberg, J. Process monitoring using ultrasonic sensor systems. *Ultrasonics* **2006**, *44*, e1395–e1399. [CrossRef]
45. Li, X.; Zhao, L.; Wei, L.; Yang, M.-H.; Wu, F.; Zhuang, Y.; Ling, H.; Wang, J. DeepSaliency: Multi-Task Deep Neural Network Model for Salient Object Detection. *IEEE T. Image Process.* **2016**, *25*, 3919–3930. [CrossRef]
46. Hochreiter, S.; Schmidhuber, J. Long Short-Term Memory. *Neural Comput.* **1997**, *9*, 1735–1780. [CrossRef]
47. Machine Learning Mastery. Available online: https://machinelearningmastery.com/handle-long-sequences-long-short-termmemory-recurrent-neural-networks/ (accessed on 11 August 2021).
48. Srivastava, N.; Hinton, G.; Krizhevsky, A.; Sutskever, I.; Salakhutdinov, R. Dropout: A simple way to prevent neural networks from overfitting. *J. Mach. Learn. Res.* **2014**, *15*, 1929–1958.
49. Lamberti, N.; Ardia, L.; Albanese, D.; Di Matteo, M. An ultrasound technique for monitoring the alcoholic wine fermentation. *Ultrasonics* **2009**, *49*, 94–97. [CrossRef]
50. Chen, Y.-T.; Chunag, Y.-C.; Wu, A.-Y.A. Online Extreme Learning Machine Design for the Application of Federated Learning. In Proceedings of the IEEE International Conference on Artificial Intelligence Circuits and Systems (AICAS), Genova, Italy, 31 August–2 September 2020; pp. 188–192. [CrossRef]
51. Dib, M.A.D.S.; Ribeiro, B.; Prates, P. Federated Learning as a Privacy-Providing Machine Learning for Defect Predictions in Smart Manufacturing. *Smart Sustain. Manuf. Syst.* **2021**, *5*, 1–17. [CrossRef]
52. Ge, N.; Li, G.; Zhang, L.; Liu, Y. Failure prediction in production line based on federated learning: An empirical study. *J Intell Manuf.* **2021**. [CrossRef]

9 Discussion

This section justifies the methods used in this thesis to achieve the aim of developing ML methods to facilitate optimal deployment of US sensors for process monitoring applications in industrial environments. The methods discussed consist of the data collected; the training, validation, and testing procedures used; the performance metrics utilised to evaluate the ML models; and the level of accuracy required by the ML models in this thesis. Considerations for the application of US sensing and ML combinations in industrial environments is included, from which, limitations of this thesis towards achieving its aim are drawn.

9.1 Data collection

The collection and use of data underpins ML as it is used for training the models. The quantity and quality of the data seen during training, along with the data used during validation to optimise hyperparameters, determines whether the ML model is suitable for its desired purpose. The aim of this thesis was to develop ML methods to facilitate optimal deployment of US sensors for process monitoring applications in industrial environments. To achieve this purpose, the conclusions drawn from this thesis, such as the relative strengths and weaknesses of each US sensing or ML technique, must be representative of what may be expected in industrial environments. The data must therefore satisfy the following criteria:

1. The US sensor data must monitor the desired process phenomena. This allows conclusions from this thesis to be drawn about the relative merits of US sensing approaches to monitor processes. To achieve this:
    a. Data must be collected at a sufficient timescale granularity to ensure that the US sensors are able to measure the required process phenomena.
    b. It must be evidenced that the US sensors are able to provide useful information about the process.
2. The US data must be collected from a sufficient variety of processes that covers the range encountered in industrial environments. This enables the conclusions drawn from this thesis to be expected to apply to new processes encountered.

9.1.1 Criterion 1a

Table 1 compares the timescale granularity required to monitor the process phenomena investigated compared to the granularity used in this thesis. The minimum timescale granularity required has been estimated using the fastest process times monitored during the experimental runs (honey-water blending, pipe cleaning, and beer fermentation), using reference measurements (flour-water batter mixing), and comparison to previous literature (beer fermentation). Importantly, all utilised timescales were below the required sizes and thereby satisfy criterion 1a.

Table 1: A comparison between the minimum timescale granularity required to monitor the desired process phenomena compared with the granularity used in this thesis. The granularity used in this thesis is below the required minimums, thereby satisfying criterion 1a.

Process	**Phenomena to be monitored using US sensors**	**Granularity required (time per measurement)**	**Explanation**	**Granularity used in this thesis (time per measurement)**
Honey-water blending	Development of homogeneity	20 s	The shortest time for mixing completion was 200 s. Assuming	1 s

			an order of magnitude lesser timescale is needed to monitor this process, a minimum granularity of 20 s per measurement is required.	
Flour-water batter mixing	Development of gluten network	30 s	Figure 1 displays the power consumption to the impeller during flour-water batter mixing acquired every 30 s and displays sufficient granularity to determine the time for optimal mixing (the peak in the power draw).	1 s
Pipe cleaning	Removal of fouling	13.2 s	The shortest cleaning process occurred in 132 s. Assuming an order of magnitude lesser timescale is needed to follow this process, a minimum granularity of 13.2 s per measurement is required.	4 s
Beer fermentation	Decreasing wort density	30 minutes	The shortest time for ethanol production to commence was 0.21 days and the shortest time until the end of ethanol production was 1.3 days. Using, the value of 0.21 days and assuming that an order of magnitude lesser timescale is needed, a minimum granularity of 30 minutes per measurement is required.	200 s
	Production of carbon dioxide bubbles	432 s	Resa et al. (2008) monitored variability in sound wave attenuation during carbon dioxide production using a timescale granularity	

			of approximately 200 measurements per day. This equated to one measurement every 432 s.	

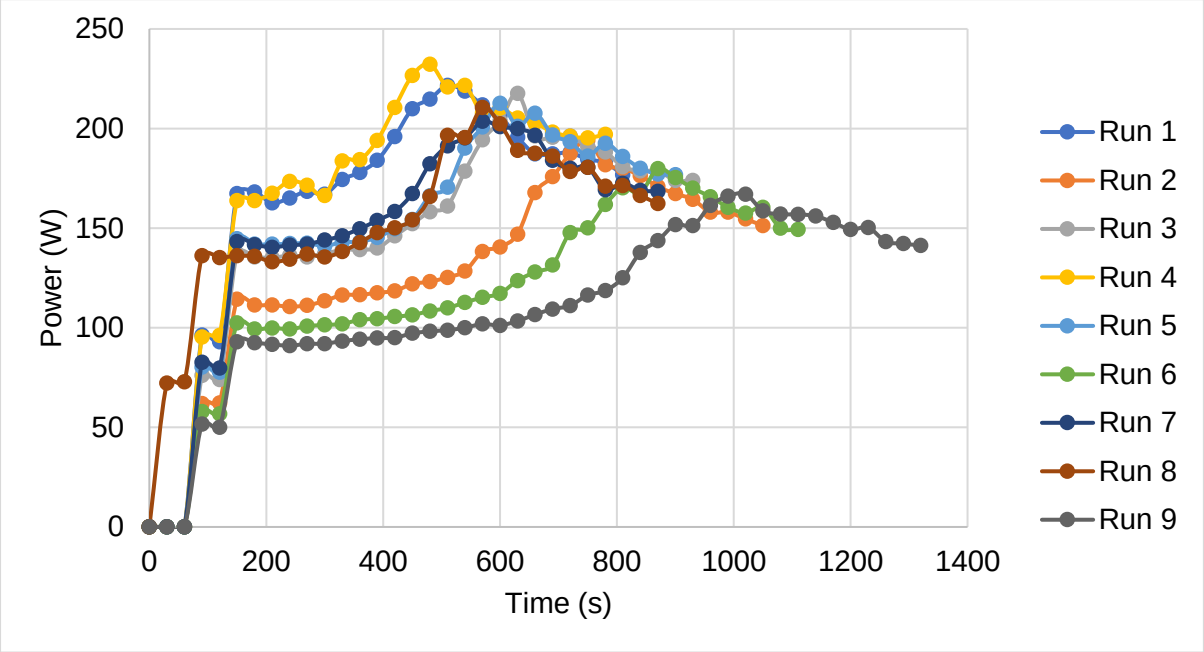
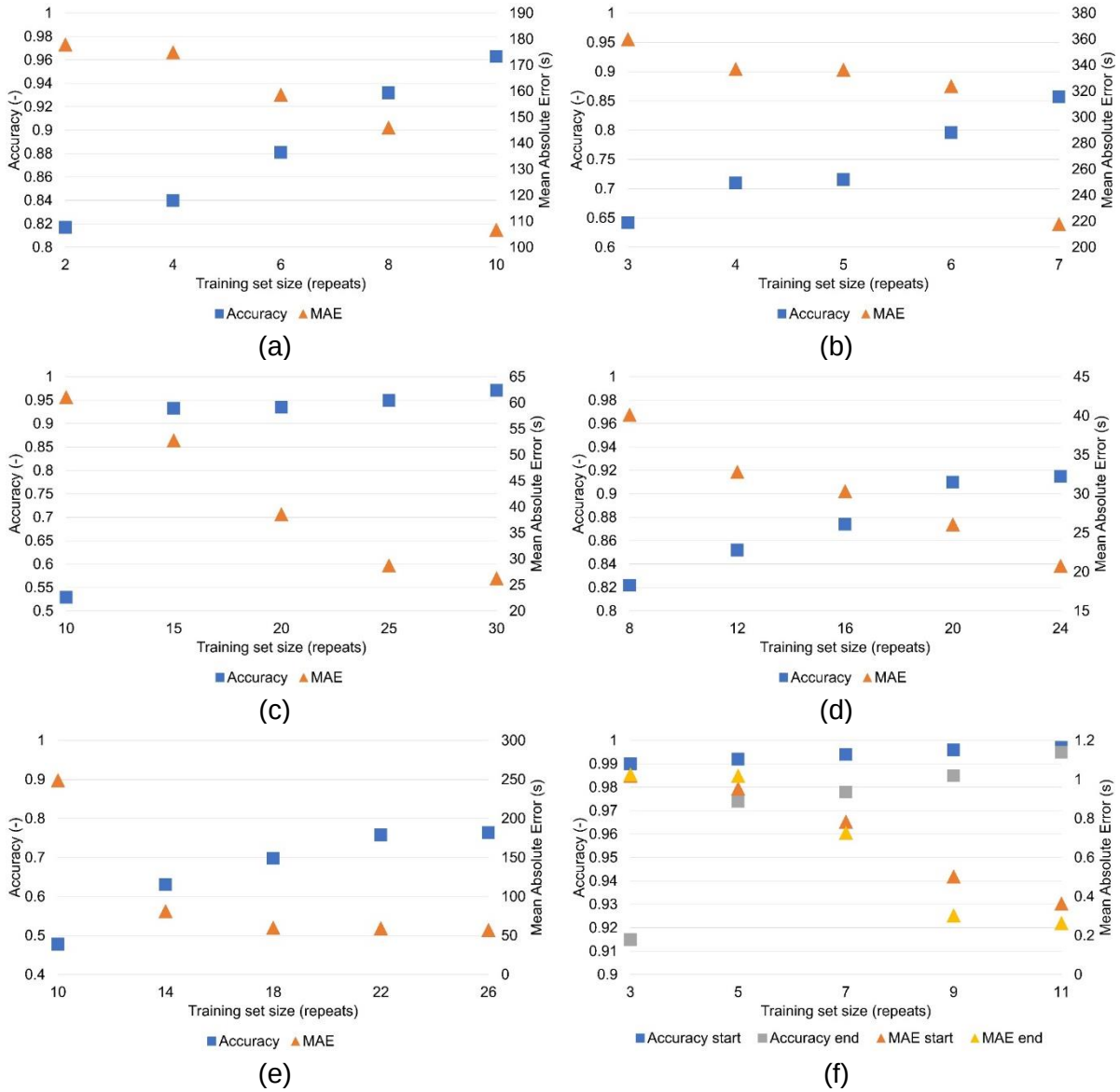


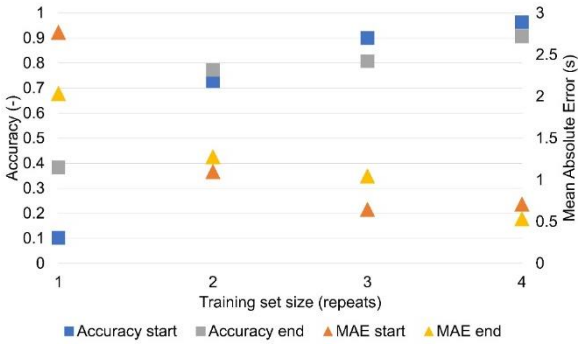
Figure 1: The power draw to the impeller measurements during the flour-water batter mixing process runs. The power draw was recorded every 30 seconds and had sufficient granularity to determine the time for optimal mixing (the peak in the power draw) (Perez Alvarado et al., 2016).

9.1.2 Criterion 1b

To evidence that the US sensors provide useful information about the processes being monitored, models were trained with varying training set sizes to evaluate whether increased accuracy was achieved with greater volumes of US sensor data. The purpose of this exercise was not to obtain as accurate predictions as possible but rather to show that with increasing volumes of US data, the ML models perform are able to better predict information relevant to each process. The models were trained using the convolutional feature extraction, multi-task networks presented in Section 6 ('Convolutional feature extraction for process monitoring using ultrasonic sensors'). Figure 2a-g displays the results of these investigations. For honey-water blending (Figure 2a), flour-water batter mixing (Figure 2b), and pipe cleaning (Figures 2c-e), the models were trained to predict whether the process was complete (classification task) and the time remaining until completion (regression task). For the laboratory (Figure 2f) and industrial (Figure 2g) beer fermentation processes, the models were trained to predict whether ethanol production had begun or finished (classification task), and the processing time remaining until these process stages were reached (regression task). The model performance for these tasks is represented by the model's percentage accuracy (classification task) and MAE (regression task). For all processes, it is shown that increasing the volume of US data used for ML models training increases prediction accuracy, thereby suggesting that the US sensor measurements provide useful process information and satisfying criterion 1b. This proves that the US sensors are able to provide useful information about each process and allows the

conclusions drawn from this thesis to be expected to apply to new process monitoring applications. However, plateaus in ML model accuracy are not achieved for any datasets, suggesting that collection of further data is required to fully represent the case studies. Although, it must be noted that these models were the final multi-task models produced in Section 6 and further optimisation for each case study and dataset size is possible. Increasing dataset size would increase the magnitude of this thesis' contribution towards achieving its aim as with greater data volumes, greater variations in the proportion of the dataset used during training may be investigated. As such an adaptable US sensing and ML pipeline for varying generalisability requirements could be developed to produce a hierarchy of techniques for manufacturers to use within their applications. This is discussed further in Section 9.6 'Limitations'.





(g)

Figure 2: a) Honey-water blending, b) flour-water batter mixing, c) flat pipe section cleaning, d) plastic pipe section cleaning, e) metal pipe section cleaning, f) laboratory beer fermentation, and g) industrial beer fermentation. The models were trained using the coarse feature extraction, multi-task networks presented in Section 6 ('Convolutional feature extraction for process monitoring using ultrasonic sensors'). For Figures 2a-e, the models were trained to predict whether the process was complete (classification task) and the time remaining until completion (regression task). For Figures 2f-g, the models were trained to predict whether ethanol production had begun or finished (classification task), and the processing time remaining until these process stages were achieved (regression task).

9.1.3 Criterion 2

US waveforms can be affected by two phenomena: Firstly, the magnitude of the waveform can be altered by changing acoustic impedance or attenuation. Secondly, variations in sound velocity alter the displacement of the waveform in the time domain (Henning and Rautenburg, 2006). In this thesis, the processes monitored were honey-water blending, flour-water batter mixing, cleaning of food fouling from pipe sections, and alcoholic beer fermentation. During blending, mixing, and cleaning, the acoustic impedance at the measurement areas change as the material composition varies. Furthermore, conducting these processes over a range of temperatures alters the sound velocity through the materials. Throughout alcoholic beer fermentation, the acoustic impedance and speed of sound of a transmitted US wave change as the density of the wort decreases (Bowler et al., 2021). Moreover, attenuation increases as CO_2 bubbles are produced. Therefore, this thesis meets criterion 2 by monitoring the full range of impacts to US waveforms during process monitoring. This means that the feature extraction methods for use in the ML pipeline or domain adaptation methods were developed using all possible impacts to US waveforms. Furthermore, the methods in this thesis can be applied to both stationary and evolving processes. Processes which evolve over time, such as those investigated in this thesis, require information from previous timesteps as inputs to the ML model. In these instances the LSTM layers used in the final ML pipeline may be used to learn process trajectories. However, if only information from the current timestep is required, then the feature extraction methodologies developed in this thesis can be applied into neural network architectures.

9.2 Training, validation, and testing procedures used

The aim of this thesis was to develop ML methods to facilitate optimal deployment of US sensors for process monitoring applications in industrial environments. To achieve this purpose, the conclusions drawn from this thesis, such as the relative strengths and weaknesses of each US sensing or ML technique, must be representative of what may be expected in industrial environments. Therefore, the training, validation, and testing procedures must be similar to those that would be used in industrial settings so that the

conclusions drawn from this thesis can be expected to transfer to industrial processes. K-fold cross validation splits the combined training and validation sets into *k* sections. Each of these sections is used as individual validation sets and an average metric for the validation accuracy is obtained (Jung and Hu, 2015). This is a robust method of validation as it repeats the procedure *k* number of times. The size of *k* typically ranges between 2 and 10 (Xu and Goodacre, 2018). However, in industrial environments, single-fold validation may be utilized to reduce ML model development time. Furthermore, drawing conclusions from test set results for models that have undergone single-fold validation biases for ML techniques with greater generalisation capabilities. As the level of generalisation required in industrial applications varies (see Section 9.5 'Considerations for industrial application'), this does not pose a limitation to the results generated. Therefore, either k-fold or single-fold validation methods are representative of industrial practices (Table 2). However, no validation set was used for the LSTMs and CNNs in Section 4 or the LSTMs in Section 7. Whilst this represents a limitation, this was not found to influence the conclusions drawn from this thesis. The reason for this is that the omission of a validation set disadvantages the hyperparameters selected for the LSTMs and CNNs compared to the other algorithms used in the articles (ANNs and SVMs in Section 4 and ANNs in Section 7) as they were not chosen to generalise to unseen data. Despite this, LSTMs and CNNs performed best in both studies. Further discussion is provided in Sections 4 and 7. The chosen test set size typically lies in the range of 10 to 50 % of the data with the validation set forming the same proportion of the total training and validation set size (Xu and Goodacre, 2018). However, in industrial applications, the desired level of model generalisability is dependent on the volume of representative data able to be collected. This is dependent on the process complexity, variability, and level of disruption to the manufacturing process caused by labelled data collection (see Section 9.5 'Considerations for industrial application'). Therefore, all validation and test set sizes used in this thesis allow the conclusions drawn from this thesis to apply to industrial processes and therefore contribute to the thesis aim of developing ML methods to facilitate optimal deployment of US sensors for process monitoring applications in industrial environments. Although, a greater contribution to this aim could obtained through development of an adaptable US sensing and ML pipeline to different generalisability requirements. For example, by comparing each US sensing method and ML technique over varying validation procedures and training, validation, and test split sizes, a hierarchy of techniques suitable for each application could be provided. This is also discussed in Section 9.6 'Limitations'.

Table 2: The validation procedures used in the sections of this thesis. For all validation procedures, the datasets were splits by number of runs so that all datapoints for a run were included in either the training, validation, or test set.

Section	Algorithm	Validation procedure
4	ANN	Single-fold (~ 20% of data)
	SVM	5-fold
	LSTM	None
	CNN	None
5	ANN and LSTM	5-fold
6	LSTM	Single-fold (~ 20% of data)
7	ANN	Single-fold (~ 20% of data)
	LSTM	None
8	LSTM	k-fold (where k ranges from 1 to 4)

9.3 Metrics used and levels of accuracy attained

In this thesis, several metrics were utilised to evaluate the predictive capabilities of the ML models including accuracy for classification tasks and R^2, MAE, and MSE for regression tasks (see Section 3.2.3.2 'Metrics'). To achieve the aim, the conclusions drawn from this thesis about the relative strengths of each US sensing or ML technique must be representative of what may be achieved in industrial applications. For the classification tasks, only the accuracy metric was used (the sum of the number of correct predictions divided by the total number of predictions and multiplied by 100 to convert into a percentage). While additional metrics (such as precision, specificity, and recall or sensitivity) may be used to obtain greater information about model performance especially on imbalanced datasets, in this thesis, the model accuracies are used to compare between ML pipelines and are not used to predict a final model's performance in application. Therefore, as accuracy provides a method of comparing between ML techniques, this fulfils the thesis aim. Similarly to the classification metric, in this thesis, the regression metrics were used to compare between US and ML techniques rather than evaluate an ML model's performance in industrial application. Therefore, any of the regression metrics may be used to compare between models and satisfy the thesis aim. However, in industrial application, the choice of evaluation metric may alter the optimal feature extraction method, algorithm, or hyperparameters required to accomplish the desired task. For example, MSE provides larger weighting to high errors and therefore a lower MSE indicates a model more able to correctly predict over the full process length. In comparison, models with low MAE may provide increased accuracy at certain process stages compared to models with low MSE scores. Furthermore, for mixing and cleaning applications, a false positive (i.e., an incorrect prediction that the pipe section is fully clean or a vessel is fully mixed) may trigger the ending of the process and thereby be less desirable than a false negative. Therefore, precision may be a better indicator of model performance than accuracy. This therefore represents a limitation of this thesis that further tasks could have been explored using the case studies to identify the relative strengths of US sensing and ML techniques for different applications. This is also discussed in Section 9.6 'Limitations'.

To draw conclusions about the strengths of each US sensing and ML technique, no criteria for minimum accuracy attained during process monitoring is required. The reason for this is that the techniques are compared to one another using the same data and therefore the relative accuracies can be expected to apply to industrial process monitoring. However, during industrial implementation, minimum accuracy is required to satisfy the manufacturers' criteria, as discussed in Section 9.5 'Considerations for industrial application'.

9.4 Chosen temperature range

For the honey-water blending datasets, the temperature range investigated spanned between 19.3 and 22.1 °C. Specifically, the temperature ranged between 19.3 to 22.1°C for one dataset compared with 19.8 to 21.2°C for the second. When developing unlabelled domain adaptation methods in Section 7, it was concluded that training ML models across a greater process parameter range (the greater range of temperatures for one of the honey-water blending datasets) or multiple datasets improved transfer learning to further datasets by enabling the models to adapt to a wider range of feature distributions. The aim of this thesis was to develop ML methods to facilitate optimal deployment of US sensors for process monitoring applications in industrial environments. Therefore, the temperature range chosen, spanning from 19.3 to 22.1 °C, is used to draw conclusions about the relative strengths of each ML method investigated rather than to reflect actual industrial temperature spectra. The choice of temperature range and granularity is therefore justified as the emphasis of this thesis is comparative evaluation of ML methods rather than achieving high

accuracy in an industrial application. The results drawn from this thesis are indicative of differing temperature ranges between the two mixing processes and provide insight into which ML methods may be most generalisable to varying temperature ranges in industrial scenarios. However, further exploration of greater parameter ranges would have increased the strength of the conclusions drawn. For example, multiple datasets could have been collected for each case study (mixing, cleaning, and fermentation) across varying process parameter ranges, other process parameter ranges could have been made to be different other than temperature (for example, material quantities or impeller speed during honey-water blending), or more than two datasets with differing parameter ranges could have been collected (e.g., a low, medium, and a larger parameter range dataset). Each of these methods could have evaluated whether the insight that training ML models across a greater process parameter range improved transfer learning to further datasets was consistent for all investigations.

9.5 Considerations for industrial application

The aim of this thesis was to develop ML methods to facilitate optimal deployment of US sensors for process monitoring applications in industrial environments. However, during application in industrial environments, the aim would be to develop ML models that are able to accurately predict information about the process being monitored from new, real-time US data. This aim presents a different set of criteria for the data collection stage compared with this thesis:

1. The collection of data should not present an unacceptable level of disruption to the manufacturer.
2. The US sensor and ML combination should achieve the level of accuracy desired by the manufacturer.

To achieve the first criteria, the sensing approach should be chosen according to the manufacturer's requirements. For example, it must be determined whether an invasive sensing solution is acceptable or whether a non-invasive sensing approach must be used. For example, in the food and drink industry, invasive probes may be discounted for some applications due to risk of contamination between product batches (Bowler et al, 2020). During data collection, normal production schedules may have to be postponed in order to sample the process or to obtain data over a wider range of process parameters than observed during normal operation. Therefore, the volume of data collected must be below the tolerable threshold to the manufacturer. This may require the use of additional techniques such as semi-supervised learning to utilise unlabelled data during normal production (Ge et al., 2017), transfer learning from similar tasks (Bowler et al., 2022b), active learning to select process parameters with high model uncertainty (Bull et al., 2019), or on-line learning to continuously update the models during normal operation (Kaushik et al., 2021) to achieve the required accuracy whilst reducing the disruption to the manufacturing process. The method of obtaining labelled data to train the ML models must also be decided. For example, sampling may require stopping the process at a greater frequency than during normal operation or an additional reference measurement sensor may need to be installed during the data collection stage.

To achieve the second criteria, firstly, it must be determined that the US sensors measure the desired process phenomena. The sensing solution must be positioned in a location that is able to monitor these phenomena. For example, to monitor pipe cleaning, the sensors should monitor the areas of most fouling to allow identification of the end of the full cleaning process (e.g. the bottom of the pipe or at corners of pipe connections). For the honey-water blending datasets, two sensors were used. One sensor was located in the centre of the

vessel base and the second was attached to the vessel base but offset from the sensor. In Section 4, the highest R^2 values (up to 0.977) were achieved by combining the inputs from both sensors. This is because the non-central sensor had better prediction ability nearer the beginning of the process as the honey was removed from this sensor measurement area first, and the central sensor had greater resolution nearer the end of the mixing process as the last of the honey was dissolved. Therefore, the location of the sensors should be chosen to obtain representative measurements of the process to be monitored. Multiple sensors may require use on large vessels where the dynamics at different locations in the vessel vary.

The sensing approach utilised must also be able to measure the desired phenomena. For example, it must be decided whether a reflection-mode sensing technique is sufficient or whether transmission of the sound wave through the process material is required. In this thesis, transmission-based techniques were used for fermentation monitoring. However, in Section 5, the non-invasive, reflection-mode method (R^2 = 0.948, MAE = 0.283 % ABV, MSE = 0.146 % ABV) achieved similar accuracy to the transmission-based technique (R^2 = 0.952, MAE = 0.265 % ABV, MSE = 0.136 % ABV). If transmission methods are required, then non-invasive sensors that transmit across the full vessel diameter (although this may lead to strong attenuation of the sound wave), invasive probe sensors (as used for fermentation monitoring in this thesis), or non-invasive sensors positioned on recycle lines or sampling points may be used. For example, Hussein et al. (2012) implemented a single non-invasive US sensor on a circulation line for in-line monitoring of a 60 litre fermentation process. The selection of the timescale granularity for data acquisition as well as determining whether the US sensors provide useful information about the process can be conducted using the same approaches taken in Sections 9.1.1 (Criterion 1a) and 9.1.2 (Criterion 1b).

During data collection, similar plots to Figures 2a-g can be constructed to identify a plateau in model accuracy to determine when a sufficient volume of data has been obtained to represent the process. It should be noted that this plateau does not identify that the desired level of accuracy is achievable, only that further data collection would not lead to an increase in model accuracy. The choice of datapoints for use in the training, validation, and tests sets must also be selected to obtain representative accuracy estimations for the ML models during production. For example, validation set datapoints may be selected to be outside the input feature range of the training set datapoints, and the test set datapoints to be located outside of both the training and validation sets. In this way, hyperparameters that provide the best extrapolation of the model predictions are chosen in the validation stage and the accuracies obtained on the test set are conservative estimates based on datapoints from outside the range that the models were trained on. This produces models that are robust to process deviations which may be caused by varying input streams, alterations to operating conditions, or seasonal variations (Fisher et al., 2020). Furthermore, large validation and test sets (e.g. 25 % training, 25 % validation, and 50 % test splits) could be employed to evaluate the model's ability to generalise. Further ML techniques can be employed to increase model robustness to changing feature distributions such as real-time domain adaptation strategies including feature alignment, prediction alignment, and feature removal (Bowler et al., 2023a).

During production, outlier detection methods (e.g., isolation forests (scikit-learn, 2023)) may be used to identify datapoints located outside the feature range encountered during model training. Furthermore, autoencoders may be used to identify datapoints that are located within the model training data range but whose feature combinations signal the datapoint as an anomaly (Ha et al., 2022) and uncertainty quantification techniques (such as model ensembles (Bowler et al., 2023b)) can be used to identify datapoints with low model confidence. Each of these methods may be used to firstly identify datapoints where the

model prediction should be questioned and secondly where further data collection efforts should be allocated.

It is also important to understand what the manufacturer defines as accurate and how this informs the ML approach. For example, during beer fermentation, wort density is usually measured once or twice per day (Controllo e Misura, 2021). However, the wort density may be used for many purposes such as monitoring the rate of fermentation, the beer alcohol content, and the start or end of ethanol production (further discussion is provided in Section 5). Therefore, to monitor the start or end of ethanol production, increased sampling rates may be required to obtain enhanced data labelling around these process stages. The choice of metric is an important consideration for determining accuracy. As discussed previously, for mixing and cleaning applications, a false positive (i.e., an incorrect prediction that the pipe section is fully clean or a vessel is fully mixed) may trigger the ending of the process and thereby be less desirable than a false negative. Therefore, precision may be a better indicator of model performance than accuracy. Furthermore, if using neural networks, the choice of metric during the gradient descent operation determines the model produced. For example, using MSE for regression tasks penalises larger errors leading to models that more accurately monitoring the full process length. Whereas, using MAE may produce models that are more accurate at process stages that provide greater US waveform information.

9.6 Limitations

As the aim of this thesis was to develop ML methods to facilitate optimal deployment of US sensors for process monitoring applications in industrial environments, there is no minimum data collection volume required to achieve this (aside from proving that useful information is provided by the sensors as highlighted by Section 9.1.2 ‘Criterion 1b’). However, increasing dataset size would increase the magnitude of this thesis’ contribution towards achieving this aim. With greater data volumes, greater variations in the proportion of the dataset used during training may be investigated. This is evidenced by the results presented in Figures 2a-g where plateaus in ML model accuracy are not achieved for any datasets, suggesting that collection of further data is required to fully represent the case studies. However, it must be noted that these models were the final multi-task models produced in Section 6 and further optimisation for each case study and dataset size is possible. Collection of further data would enable investigation into the relative strengths of US sensing and ML approaches for a wider range of training dataset sizes. This would provide a greater contribution to facilitating the implementation of US sensing and ML combinations in industrial environments by providing adaptable recommendations to the size of dataset available. For example, a larger dataset volume which more completely represents the monitored process’ feature range and distribution may enable the use of more complex algorithms such as CNNs and deep neural networks with LSTM layers. However, a smaller, less representative dataset may require simpler ML techniques such as ANNs with feature gradients as inputs to reduce overfitting to the training data. Furthermore, transmission-based US sensing approaches obtain more US information compared with reflection-mode techniques. Therefore, transmission-based methods may provide increased accuracy with larger dataset sizes whereas reflection-mode methods may provide better generalisation due to the fewer features extracted at lower dataset sizes. The current conclusions drawn from this thesis are useful starting pipelines to trial but require evaluation for each new process encountered.

This adaptable pipeline can also be extended to varying validation procedures (e.g. k-fold or single-fold), sizes of validation sets, or to validation datapoints located outside of the training dataset range (thereby evaluating model robustness to process deviations). Similarly, this could produce a hierarchy of techniques to utilise depending on the level of generalisability required by the model. To further improve the contribution to the thesis aim, the US sensing

and ML combinations could be applied to a greater number of tasks using the same case studies as presented. For example, ML techniques could be compared for each of the possible desired tasks for fermentation monitoring such as predicting fermentation rate, alcohol concentration, and the start or end of ethanol production. For the cleaning and mixing datasets, precision could be used as the performance metric to minimise the false positive rate triggering the early ending of the processes. This could be used to develop an adaptable pipeline to greater range of ML tasks. These limitations are also discussed in Section 10.1 ‘Limitations of thesis’ after the results from the thesis have been presented.

As discussed in Section 9.4 ‘Chosen temperature range’, when developing unlabelled domain adaptation methods in Section 7, it was concluded that training ML models across a greater process parameter range (a greater range of temperatures; 19.3 to 22.1°C compared with 19.8 to 21.2°C for one of the honey-water blending datasets) or multiple datasets improved transfer learning to further datasets by enabling the models to adapt to a wider range of feature distributions. Further exploration of greater parameter ranges would have increased the strength of the conclusions drawn. For example, multiple datasets could have been collected for each case study (mixing, cleaning, and fermentation) across varying process parameter ranges, other process parameter ranges could have been made to be different other than temperature (for example, material quantities or impeller speed during honey-water blending), or more than two datasets with differing parameter ranges could have been collected (e.g., a low, medium, and a larger parameter range dataset). Each of these methods could have evaluated whether the insight that training ML models across a greater process parameter range improved transfer learning to further datasets was consistent for all investigations.

9.7 References

Bowler, A.L., Bakalis, S., Watson, N.J. (2020) ‘A review of in-line and on-line measurement techniques to monitor industrial mixing processes’ *Chemical Engineering Research and Design* 153, 463-495. Doi: 10.1016/j.cherd.2019.10.045.

Bowler, A.L., Escrig, J., Pound, M., Watson, N. (2021) ‘Predicting alcohol concentration during beer fermentation using ultrasonic measurements and machine learning’ *Fermentation* 7, 34. Doi: 10.3390/fermentation7010034.

Bowler, A.L., Pound, M.P., Watson, N.J. (2022a) ‘A review of ultrasonic sensing and machine learning methods to monitor industrial processes’ *Ultrasonics* 124. Doi:10.1016/j.ultras.2022.106776.

Bowler, A.L., Pound, M.P., Watson, N.J. (2022b) ‘Domain adaptation and federated learning for ultrasonic monitoring of beer fermentation’ *Fermentation* 7, 253. Doi:10.3390/fermentation7040253.

Bowler, A., Ozturk, S., di Bari, V., Glover, Z.J., Watson, N.J. (2023) ‘Machine learning and domain adaptation to monitor yoghurt fermentation using ultrasonic measurements’ *Food Control* 147, 109622. Doi:10.1016/j.foodcont.2023.109622.

Bowler, A.L., Ozturk, S., Rady, A., Watson, N. (2022) ‘Domain Adaptation for In-Line Allergen Classification of Agri-Food Powders Using Near-Infrared Spectroscopy’ *Sensors* 22, 7239. Doi: 10.3390/s22197239.

Bull, L.A., Rogers, T.J., Wickramarachchi, C., Cross, E.J., Worden, K., Dervilis, N. (2019) ‘Probabilistic active learning: an online framework for structural health monitoring’ *Mechanical Systems and Signal Processing* 134, 106294. Doi:10.1016/j.ymssp.2019.106294.

Controllo e Misura (2021) The Brewing Process Is under Control. Available online: https://www.publiteconline.it/controlloemisura/2021/09/10/the-brewing-process-is-under-control/ (Accessed 18th April 2023)

Fisher, O.J., Watson, N.J., Escrig, J.E., Witt, R., Porcu, L., Bacon, D., Rigley, M., Gomes, R.L. (2020) 'Considerations, challenges and opportunities when developing data-driven models for process manufacturing systems' *Computers and Chemical Engineering* 140, 106881. Doi:10.1016/j.compchemeng.2020.106881.

Ge, Z., Song, Z., Ding, S.X., Huang, B. (2017) 'Data mining and analytics in the process industry: the role of machine learning' *IEEE Access* 5, 20590–20616. Doi: 10.1109/ACCESS.2017.2756872.

Ha, J.M., Seung, H.M., Choi, W. (2022) 'Autoencoder-based detection of near-surface defects in ultrasonic testing' *Ultrasonics* 119, 106637. Doi: 10.1016/j.ultras.2021.106637.

Henning, B., Rautenberg, J. (2006) 'Process monitoring using ultrasonic sensor systems' *Ultrasonics* 44, e1395–e1399. Doi:10.1016/j.ultras.2006.05.048.

Hussein, W.B., Hussein, M.A, Becker, T. (2012) 'Robust spectral estimation for speed of sound with phase shift correction applied online in yeast fermentation processes' *Engineering in Life Sciences* 12(6), 603–614. Doi: 10.1002/elsc.201100183.

Jung, Y., Hu, J. (2015) 'A K-fold averaging cross-validation procedure' *Journal of Nonparametric Statistics* 27(2), 167-179. Doi:10.1080/10485252.2015.1010532.

Kaushik, P., Gain, A., Kortylewski, A., Yuille, A.L. (2021) 'Understanding Catastrophic Forgetting and Remembering in Continual Learning with Optimal Relevance Mapping' *CoRR* abs/2102.11343.

Perez Alvarado, F.A., Hussein, M.A., Becker, T. (2016) 'A Vision System for Surface Homogeneity Analysis of Dough Based on the Grey Level Co-occurrence Matrix (GLCM) for Optimum Kneading Time Prediction' *Journal of Food Process Engineering* 39, 166–177. Doi: 10.1111/jfpe.12209.

Resa, P., Elvira, L., De Espinosa, F.M., González, R., Barcenilla, J. (2008) 'On-line ultrasonic velocity monitoring of alcoholic fermentation kinetics' *Bioprocess and Biosystems Engineering* 32, 321–331. Doi: 10.1007/s00449-008-0251-3.

Scikit-learn (2023) 2.7. Novelty and Outlier Detection. Available online: https://scikit-learn.org/stable/modules/outlier_detection.html#novelty-detection-with-local-outlier-factor (accessed 14 April 2023).

10 Conclusions and recommendations

The aim of this thesis was to develop ML methods to facilitate optimal deployment of US sensors for process monitoring applications in industrial environments. Table 1 is included below as a reminder of the sections in this thesis. The objectives were:

1. **To collect US sensor data for process monitoring applications that enable the thesis conclusions to be expected to extend to industrial environments.** Three types of processes were monitored: material mixing, cleaning of pipe sections, and alcoholic fermentation covering the full range of impacts to US waveforms during process monitoring (see Section 3.1 Experimental datasets). Furthermore, the data was collected at sufficient timescale granularity and the US sensor data was shown to contain useful information about the processes (see Section 9 Discussion).

2. **To evaluate different US sensing techniques to determine their benefits and limitations for industrial process monitoring.** Two US sensing techniques were investigated: a non-invasive, reflection-mode technique that can be externally retrofitted to existing processing equipment, and an invasive US probe with reflector plate.

3. **To evaluate different feature extraction, feature selection, algorithm types and hyperparameter values to determine the optimal ML pipeline for process monitoring using US measurements.** This reduces time for ML model development in industrial environments by suggesting ML pipelines that achieve the highest accuracy for previous process monitoring tasks.

4. **To develop unlabelled domain adaptation methods to utilise previously collected datasets and negate the data labelling burden for sensor deployment.** These methods can be used to transfer ML models between processes without requiring labelled data (the outputs required during ML model training and often acquired by using reference measurements) and therefore negate disruption to a manufacturing process during the data collection stage.

5. **To develop labelled domain adaptation methods to utilise previously collected datasets, reduce the data labelling burden for sensor deployment, and improve ML model accuracy on target processes.** These methods can be used to reduce disruption to a manufacturing process during the data collection stage.

Table 1: A summary of the articles in each section of this thesis.

Section	Title of article	Novelty
4	Monitoring Mixing Processes Using Ultrasonic Sensors and Machine Learning	ML was combined with non-invasive, reflection-mode US sensing to monitor mixing processes. A range of ML algorithms were trialled including LSTMs and CNNs.
5	Predicting Alcohol Concentration during Beer Fermentation Using Ultrasonic Measurements and Machine Learning	ML was combined with non-invasive, reflection-mode US sensing to monitor fermentation processes. Omission of the temperature as a feature was evaluated.
6	Convolutional feature extraction for process monitoring using ultrasonic sensors	A convolutional feature extraction method is presented. Multi-task learning and data augmentation were applied to US sensor data.

7	Transfer learning for process monitoring using reflection-mode ultrasonic sensing	Unlabelled domain adaptation was applied to US sensor data for monitoring mixing and cleaning processes. Omission of the temperature as a feature was evaluated.
8	Domain Adaptation and Federated Learning for Ultrasonic Monitoring of Beer Fermentation	Labelled domain adaptation was applied to US sensor data for fermentation monitoring. Federated learning and multi-task learning were combined with US sensor data.

The first and second objectives were fulfilled through the work in Sections 4, 5, 6, and 7. The potential of an industrially relevant, reflection-mode, non-invasive US sensing technique was demonstrated by the work undertaken in Sections 4, 6, and 7 for mixing and cleaning processes. Furthermore, by using only the first waveform reflection and omitting the process temperature as an ML model feature, Section 5 demonstrated the potential of the sensing technique to monitor beer fermentation with comparative accuracy as transmission-based methods. This non-invasive method can be easily retrofitted onto existing production equipment. The non-invasive method (R^2 = 0.948, MAE = 0.283, MSE = 0.146) achieved similar accuracy to the transmission-based technique (R^2 = 0.952, MAE = 0.265, MSE = 0.136). Although the non-invasive US sensing approach was compared to a transmission-based approach for a single case study (laboratory beer fermentation), only Meng et al. (2012) has previously used a non-invasive, reflection-mode US sensing technique to monitor fermentation. They used a single sensor to monitor yogurt fermentation by monitoring the change in acoustic impedance at the wall-yogurt interface. All other previous applications of US sensors to monitor fermentation have utilised transmission-based approaches, highlighting the reduction in process information available to reflection-mode sensing (Bowler et al., 2022). Ultimately, the consideration of whether to use invasive or non-invasive, reflection-mode or transmission sensing approaches must be decided based on the criteria presented in Section 9.5 ‘Considerations for industrial application’: 1.) The collection of data should not present and unacceptable level of disruption to the manufacturer, and 2.) the US sensor and ML combination should achieve the level of accuracy desired by the manufacturer. It must be determined whether an invasive sensing solution is acceptable to the manufacturer or whether a non-invasive sensing approach must be used. If a reflection-mode sensing solution does not achieve the level of accuracy required by the manufacturer, then transmission-based approaches can be trialled.

The third objective was fulfilled by developing an optimal ML pipeline for process monitoring using US sensors (Figure 1). A pretrained CNN is used as a feature extractor. PCA is used to extract a small set of orthogonal features from the CNN outputs to be combined with additional features such as the time of flight or deviations between consecutively acquired waveforms. These features are then used as inputs in a deep neural network with LSTM layers. Multi-task learning is recommended for use if multiple predictions are available. The development of this pipeline is discussed in the following paragraphs.

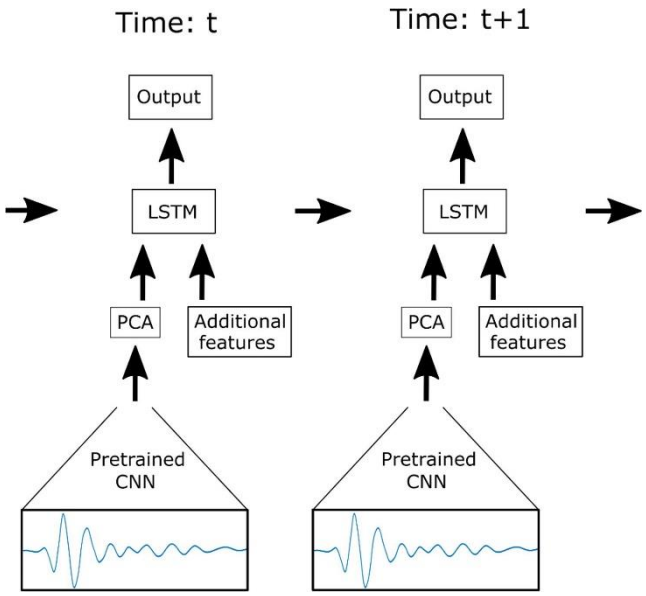


Figure 1: The optimised ML pipeline for process monitoring using US measurements (Bowler et al., 2022).

The convolutional feature extraction method presented in Section 6 was found to extract more informative waveform features from the US waveforms studied compared with traditional coarse features, achieving higher model accuracy for datasets requiring substantial waveform information and for 65% of tasks overall. The convolutional feature extraction method achieved higher prediction accuracies for tasks that required complex feature extraction methodologies in other sections in this thesis, such as the flour-water batter mixing (increasing R^2 by 0.1 ± 0.046 for regression tasks and increasing accuracy by 13.9 ± 8.8 % for classification tasks) and cleaning datasets (increasing R^2 by 0.11 ± 0.23 for regression tasks and increasing accuracy by 15.4 ± 14.3 % for classification tasks). However, the convolutional feature extraction method requires choice of the number of PCs to extract. The author recommends that the number of PCs required for this method be determined through cross-validation. These results agree with theory as CNNs are able to monitor changes to the waveform directly, rather than indirectly through extraction of coarse features such as the waveform Energy. Furthermore, CNNs are robust to the spatial variance of features unlike ANNs that use every waveform sample point amplitude as a feature (Lecun et al., 2015). The position of features may vary due to a changing speed of sound through the process material. From these results, the author recommends that CNNs should be used for feature extraction from US waveforms. This recommendation enables faster development of US and ML combinations by providing a stipulation that the feature extraction power of CNNs be used within the ML pipeline. However, CNNs often need a large volume of data for training, with even small CNNs needing at least approximately 1000 datapoints to train on (LUNA, 2016). In this thesis, the developed ML pipeline uses a CNN pre-trained on an auxiliary task to classify to which dataset previously collected US waveform originated from and was therefore trained on over 30,000 US waveforms. In this way, the CNN could learn to identify features of US waveforms and be used as a feature extractor for new tasks. Therefore, a prerequisite for utilising the developed pipeline is a large volume of US data either collected from the current process to be monitored or from previously investigated processes. If neither of the requirements are available, the coarse feature extraction methodology may be utilised, as recommended in Bowler et al. (2022).

The use of the process temperature as a feature was not necessary in order to obtain comparatively high ML model accuracy, as evidenced by its omission from all models in

Sections 4 (achieving up to 96.3 % classification and 0.977 R^2 regression accuracies for honey-water blending) and 6 (achieving up to 98.2 % classification and 0.947 R^2 regression accuracies for pipe section cleaning) compared with Section 7 (up to 96.0% and 98.4% to classify the completion of mixing and cleaning, respectively; and R^2 values of up to 0.947 and 0.999 to predict the time remaining until completion). Notably, in Section 5, the inclusion of the process temperature degraded ML model accuracy when using only the first waveform reflection during beer fermentation monitoring (increasing the mean squared error from 0.146 to 0.345 % alcohol by volume). This was most likely due to the high variability (20 – 30°C presented in Figure 3d, Section 5) of this feature creating a more difficult optimisation problem for the model. However, in subsequent works (Sections 6 and 8) an additional fully connected layer was used in the deep neural networks to reduce the burden on the LSTM layer to learn the feature trajectories and also fit to the target variables. Therefore, this new network structure is expected to handle the extra complexity of the process temperature and thereby achieve an increase in model accuracy, although, this was not confirmed in this thesis. In Section 7, using the temperature as a feature produced higher classification accuracies for all but one task for the honey-water blending datasets (increasing accuracy by 1.8 ± 1.8 %). Furthermore, incorporation of the process temperature as a feature increased the regression accuracy for the honey-water blending (increasing R^2 values by 0.078 ± 0.145) as the temperature affected the waveform features and the trajectory of the process. Overall, if available, using the process temperature as a feature is anticipated to simplify the optimisation problem and increase model accuracy. However, if not available, comparative model accuracy can still be expected. Therefore, the absence of available temperature measurement in industrial applications is not expected to hinder US sensor and ML combination deployment based on the results from these datasets. This is evidenced by the results from Section 5 where models omitting the process temperature as a feature achieving increased accuracy (previously stated in the paragraph) and Section 7 where comparable accuracy was attained when omitting temperature as a feature (Table 2).

Table 2: A comparison between the maximum accuracies achieved for the ML models in Section 7 when the process temperature was either included or excluded as an input feature.

Use of temperature as a feature	**Honey-water blending**		**Pipe cleaning**	
	Classification (% accuracy)	Regression (R^2)	Classification (% accuracy)	Regression (R^2)
Included	92.8	0.942	98.4	0.998
Excluded	96	0.947	97.5	0.999

In Section 4, for classification (96.3 % achieved using LSTM layers and feature gradients) and regression tasks (0.977 R^2 achieved using time domain input CNNs which use the previous 10 s of acquired waveforms) using the honey-water blending dataset, and regression tasks using the flour-water batter mixing dataset (0.976 R^2 achieved using time domain input CNNs), using information from previous time steps was critical for model accuracy. However, for classification tasks using the flour-water batter mixing dataset, this was not required (92.5% accuracy using CWT-input CNNs). Despite this, a single ML pipeline approach is desired for deployment with process monitoring applications. A single, universally applied ML pipeline would greatly reduce the time for model development, despite, potentially, not being the optimal model for each use case. Therefore, it is recommended that information from previous time steps be utilised for all applications where processes evolve over time, even if this creates more difficult optimisation problems by requiring the use of LSTM layers in deep neural networks. In Section 4, LSTM layers and

CNNs using up to 10 s of previously acquired data were more accurate than using fixed feature gradient lengths in ANNs (results listed above). Furthermore, in Sections 5 (LSTMs achieved R^2 = 0.952, MAE = 0.265, and MSE = 0.136 compared with R^2 = 0.398, MAE = 1.01, and MSE = 1.942 for ANNs) and 7 (where LSTMs were best for 88.5 % of tasks), neural networks with LSTM layers were more accurate for all tasks than ANNs using feature gradients. Therefore, the use of LSTM layers in deep neural networks are recommended to learn feature trajectories through flexible use of past process information. It is recommended that the impact of using of feature gradients as inputs on model accuracy should also be evaluated during the validation stage to reduce the burden on the LSTM layers to learn the feature trajectories. However, LSTMs require large training set sizes, are prone to overfitting, and require long times for training (Sugandhi, 2023). Therefore, in applications where limited data is available or there is large variability in the process and thereby input features to the ML model, the use of feature gradients as input features into ANNs may be used.

Multi-task learning has the potential to increase model accuracy through global learning of feature trajectories important to the process being monitored, allowing for more stable model training by optimising for two combined losses, and reducing overfitting by preventing a single task from dominating the learning process. In Section 6, multi-task learning improved feature trajectory learning by providing increased accuracy on 8 out of 18 regression tasks but led to reduced model accuracy on data points far from the classification decision boundaries. However, to reduce model development time, the multi-task learning network hyperparameters were not fully optimised. Therefore, the accuracy of the multi-task models could be further improved despite showing improvements in accuracy for several tasks. Although multi-task learning was not shown to consistently provide increased accuracy over single-task learning in Section 6, it should be used when possible due to the desire to deploy a single ML pipeline approach for process monitoring applications.

The fourth objective is fulfilled by the work undertaken in Section 7. Section 7 concluded that only a single feature, in this case the waveform Energy, should be used in domain adaptation tasks to describe the oscillating part of the US waveform. The single feature method achieved higher prediction accuracies (96.0 % and 98.4 % accuracy and 0.947 and 0.999 R^2 values for mixing and cleaning, respectively) compared with TCA (92.6 % and 95.3 % accuracy and 0.942 and 0.966 R^2 values for mixing and cleaning, respectively). The waveform Energy was used as it is a measure of the total magnitude of the sound wave and therefore is representative of the acoustic impedance of the process material at the vessel wall. Other features, such as the maximum or minimum peaks, position of peaks, skewness or kurtosis, are useful to monitor changes in the waveform shape and aid identifying multiple overlapping sound waves. Therefore, the trend in these other features does not follow changes in the process material. Features similar to the waveform Energy that monitor the process material properties include the Sum Absolute Amplitude or Sum Root Amplitude, however, these give greater or lesser weighting to larger amplitudes, respectively. Therefore, the discrepancy between the Energy and either of these features would be due to the shape of the waveform rather than the changing process material. As the shape of the waveform is unlikely to follow the same trends across domains, inclusion of these features may degrade model accuracy. This effect is magnified further when using more than two domains. However, as described in Section 8 (3.3. Future Research Directions), this single feature method may be used to obtain predictions in the new domain, after which, these predictions can be used as a feature to be inputted into another model trained for the new domain which also uses more waveform features. In this way, the model gains the knowledge from the source domain as well as being able to use many features that describe the changing waveform.

The fourth and fifth objective are fulfilled by the works carried out in Section 7 and 8. Using multiple source datasets in domain adaptation tasks provided increased model accuracy for the pipe cleaning datasets in Section 7 for six out of eight regression tasks. This was attributed to the model being trained to generalise across two datasets making it easier to adapt to the third, target, dataset. In addition, for the honey-water blending datasets, models trained on source honey-water blending datasets across a greater parameter range, in this case a greater range of temperatures (19.3 to 22.1°C compared with 19.8 to 21.2°C presented in Figures 1b and 1c, Section 3), improved model accuracy in Section 7 (improving classification accuracy by 2.3 ± 4.1 % and R^2 values by 0.12 ± 0.15). This allowed the LSTM layers to learn feature trajectories across a wider range of feature distributions, thereby making them more adaptable to the different feature distributions of the target domain. In Section 8, three domain adaptation methodologies were evaluated. Federated learning achieved the highest model accuracy and performed best for 14 out of 16 ML tasks compared with the models not using knowledge from the source domain. This is likely due to the order of the dataset runs the model is trained on compared with the other two methodologies evaluated. During federated learning, the local models were trained for a single epoch on the corresponding domain dataset before passing the model weights for collation. This full epoch of training allows for increased gradient descent to an optimum point compared with a procedure that alternates between source and target domain data, as used for the other domain adaptation methods. This indicates that the order of the dataset runs may impact the ability of the models to find a better local optimum and should be trialled during the validation stage.

10.1 Limitations of thesis

The aim of this thesis was to develop ML methods to facilitate optimal deployment of US sensors for process monitoring applications in industrial environments. This thesis developed an ML pipeline for process monitoring using US sensors consisting of convolutional feature extraction, LSTM layers, and multi-task learning. Furthermore, labelled and unlabelled domain adaptation methods were developed to transfer knowledge from previously collected datasets. However, this thesis did not achieve several considerations for deployment of US sensing and ML combinations in industrial environments as outlined in Section 9 ‘Discussion’. These were: 1.) the creation of an adaptable ML pipeline to different dataset sizes, generalisability requirements, or ML tasks 2.) semi-supervised, active, or on-line learning approaches to reduce the data collection burden during industrial implementation.

As outlined in Section 9.5 ‘Considerations for industrial application’, to achieve the first consideration listed above (the creation of adaptable ML pipelines) US sensing and ML approaches may be evaluated over a range of dataset sizes, using datapoints that are outliers the training dataset range, on process deviations, or using large validation and test set sizes. Furthermore, a greater range of ML tasks could be applied to the same case studies used in this thesis such as expanding multi-task learning for fermentation (simultaneous prediction of the fermentation rate, alcohol concentration, and the start and end of ethanol production) monitoring or varying evaluation metrics (as discussed in Section 9.6 ‘Limitations’). A hierarchy of each US sensing (reflection-mode or transmission-based) and ML method (e.g. CNNs or LSTM layers) may be developed for each task or generalisability requirement. In this way, manufacturers would have a framework for the different approaches to trial based on the volume of data able to be collected for the desired process or the process monitoring task. To contribute to the second requirement, active learning methods are also discussed in Section 9.5 ‘Considerations for industrial application’ including the use of outlier detection methods, autoencoders for anomaly detection, and

uncertainty quantification during production to identify datapoints that require labelling. Semi-supervised, active and on-line learning are discussed in Section 11 ‘Directions for future work’.

At present, the recommendations developed in this thesis would still need to be evaluated by a manufacturer to determine whether the use of CNNs, LSTMs layers, or transmission-based techniques would provide the level of model accuracy and generalisability required, or whether the coarse feature extraction method, ANNs with feature gradients, or reflection-mode sensing should be utilised.

In Section 7, it was concluded that training ML models across a greater process parameter range (a greater range of temperatures; 19.3 to 22.1°C compared with 19.8 to 21.2°C for one of the honey-water blending datasets) or multiple datasets improved transfer learning to further datasets by enabling the models to adapt to a wider range of feature distributions. Further exploration of greater parameter ranges would have increased the strength of the conclusions drawn. For example, multiple datasets could have been collected for each case study (mixing, cleaning, and fermentation) across varying process parameter ranges, other process parameter ranges could have been made to be different other than temperature (for example, material quantities or impeller speed during honey-water blending), or more than two datasets with differing parameter ranges could have been collected (e.g., a low, medium, and a larger parameter range dataset). Each of these methods could have evaluated whether the insight that training ML models across a greater process parameter range improved transfer learning to further datasets was consistent for all investigations.

10.2 References

Bowler, A.L., Pound, M.P., Watson, N.J. (2022) ‘A review of ultrasonic sensing and machine learning methods to monitor industrial processes’ *Ultrasonics* 124. Doi:10.1016/j.ultras.2022.106776.

Lecun, Y., Bengio, Y., Hinton, G. (2015) ‘Deep learning’ *Nature* 521, 436–444. Doi:10.1038/nature14539.

LUNA (2016) ‘Lung Nodule Analysis’. Available online: https://luna16.grand-challenge.org/Data/ (accessed 14 April 2023).

Meng, R., Zhou, J., Ye, X., Liu, D. (2012) ‘On-line monitoring of yogurt fermentation using acoustic impedance method’ *Applied Mechanics and Materials* 101-102, 737-742. Doi:10.4028/www.scientific.net/AMM.101-102.737.

Sugandhi, A. (2023) ‘What is Long Short Term Memory (LSTM) - Complete Guide’ Available online: https://www.knowledgehut.com/blog/web-development/long-short-term-memory. (accessed 14 April 2023).

11 Directions for future work

The aim of this thesis was to develop ML methods to facilitate optimal deployment of US sensors for process monitoring applications in industrial environments.

Objectives:

1. To collect US sensor data for process monitoring applications that enable the thesis conclusions to be expected to extend to industrial environments.
2. To evaluate different US sensing techniques to determine their benefits and limitations for industrial process monitoring.
3. To evaluate different feature extraction, feature selection, algorithm types and hyperparameter values to determine the optimal ML pipeline for process monitoring using US measurements.
4. To develop unlabelled domain adaptation methods to utilise previously collected datasets and negate the data labelling burden for sensor deployment.
5. To develop labelled domain adaptation methods to utilise previously collected datasets, reduce the data labelling burden for sensor deployment, and improve ML model accuracy on target processes.

11.1 Objectives 1, 2, and 3

As discussed in Section 10.1 ‘Limitations of thesis’, this thesis did not produce an adaptable ML pipeline to different dataset sizes, generalisability requirements, or ML tasks for use in industrial environments. In an industrial setting, the volume of data obtained for ML model training may be limited (see Section 9.5 ‘Considerations for industrial application’). Therefore, to achieve the level of accuracy desired by the manufacturer, each US sensing and ML application may require different generalisation performance to unseen data. The collection of further data for all case studies used in this thesis (see Section 9.6 ‘Limitations of the data collected in this thesis’) would enable ML approaches to be evaluated over a range of dataset sizes and generalisation requirements, on datapoints that are outliers the training dataset range, on process deviations from the training data, or using large validation and test set sizes (Objective 1). Furthermore, evaluation of the US sensing and ML combinations over a wider range of tasks (such as expanding multi-task learning for fermentation to simultaneously predict the fermentation rate, alcohol concentration, and the start and end of ethanol production, or by varying the evaluation metrics used for each ML task as discussed in Section 9.6 ‘Limitations’) would enable production of a adaptable ML pipeline to differing process monitoring tasks. A hierarchy of each US sensing (reflection-mode or transmission-based, Objective 2) and ML method (e.g. CNNs or LSTM layers, Objective 3) may be developed for each task or generalisability requirement. In this way, manufacturers would have a framework for the different approaches to trial based on the desired task or volume of data able to be collected.

Furthermore, this could be extended to minimisation of the time for development of ML models in industrial applications. Another constraint during industrial implementation of US sensor and ML combination will be the time required for development prior to deployment. This will be due to the time required to find optimal model hyperparameters and the computational time to train the models. There could be several methods to reduce the severity of this problem. For example, increasing model generalisability (by using out of training data distribution validation points or large validation set sizes) would obtain less complex model architecture (as less complex model architecture is comprised of fewer model parameters giving less chance of overfitting) which will also be less computationally expensive to train. Models can also be trained for fewer epochs with a higher learning rate

(whilst using learning rate decay to prevent missing optima) which will reduce computational time whilst also limiting hyperparameter ranges. Coarse-fine grid searches can also be used to reduce the time required during cross-validation. Finally, techniques that are parallelisable during training may also be of benefit, such as the ensemble methods. These methods may be evaluated and a hierarchy of ML methods to trial when model development time in industrial application is a constraint may be created.

Improve the ML pipeline further (Objective 3), other methods for dimensionality reduction following the convolutional feature extraction method presented in Section 6 can be investigated. Two examples include using autoencoders for non-linear feature extraction and feature importances determined using random forest algorithms. As PCA (used in Section 6) extracts linear dependencies, autoencoders are able to retain more waveform information but at the cost of longer training times, the requirement of hyperparameter selection, a difficult network optimisation problem, and possible coadaptation of the latent space nodes leading to redundant information being stored. Random forest algorithms can be used to determine feature importance by summing the error or impurity at each decision node in the tree structure (Mathworks, 2021). Using random forests would provide the advantages of little hyperparameter selection, fast training time, and retaining all the information from the original features, however, at the expense of potentially identifying several similar features as being the most important to the ML task.

11.2 Objectives 4 and 5

As described in Section 8 (3.3. Future Research Directions) the waveform Energy and domain adaptation can be used to produce target variable predictions for the new domain. These predictions may be used as a feature in a model along with additional features from the waveform in the new domain. This would enable the benefits of transfer learning from previous datasets as well as more accurate fitting by using all available target domain waveform features. Furthermore, federated learning was shown to improve domain adaptation capability in Section 8 which was attributed to aiding gradient descent by training on a full epoch from each dataset sequentially. During federated learning, the local models were trained for a single epoch on the corresponding domain dataset before passing the model weights for collation. This full epoch of training allows for increased gradient descent to an optimum point compared with a procedure that alternates between source and target domain data, as used for the other domain adaptation methods. This indicates that the order of the datasets during domain adaptation model training should be investigated to aid network optimisation. However, this would add an additional step in the ML pipeline.

Other domain adaptation methodologies may also be trialled. Self-training may be used to train an ML model (or ensemble of models) in the source domain and predict the target variables of data from the target domain (this may also be following some transformation of either feature space). Then, data points that the model confidently predicts the target variable (through either choosing the highest model outputs for classification or the narrowest distribution of ensemble voting) can be pseudo-labelled and added to the training set for model retraining. This iterative procedure can enable the original model to adapt to the new domain (Mishra and Woltering, 2021; Lin et al., 2019). TrAdaBoost is a method that uses labelled source domain data along with a small amount of labelled target domain data. A model is trained using these combined datasets and then the weightings of source datapoints which are incorrectly classified are decreased due to them having low similarity to the target domain, and the weightings of target domain datapoints which are incorrectly classified are increased to produce a model better fitted to the new domain. The iterative process produces an ensemble of models that can be used to predict the target variables of the remaining unlabelled target domain data (Yu et al., 2021; Li et al., 2021). Autoencoders

may also be investigated to extract waveform feature combinations which are similar across domains to identify additional features that may be used rather than only the waveform energy during domain adaptation. Furthermore, autoencoders could be used to find a transformation between US waveforms from different domains, an idea similar to near-infrared spectroscopy calibration transfer by measuring standard samples across multiple spectrometers (Andries, Kalivas, and Gurung, 2019). For example, a standard material such as water at standard temperatures can be measured using US sensors on both pieces of process equipment. Autoencoders can then learn this transformation by inputting the source domain waveform and training the reconstruct the target domain waveform. This transformation may then be applied to all source domain data and subsequently used to predict the target variables of the target domain data.

As discussed in Section 10.1 ‘Limitations of thesis’, this thesis did not develop semi-supervised, active, or on-line learning approaches to reduce the data collection burden during industrial implementation. The ability to obtain labelled data is likely to be a constraint during industrial deployment of US and ML sensing techniques. Further work should consider other in-line or on-line sensors that could be used as reference measurements during the monitored process, whether the process must be ended at different stages and off-line sensors used to obtain reference measurements, if semi-supervised learning must be used to pseudo-label the unlabelled data, if US measurements could be used to infer and label process states, and, lastly, if sampling and off-line reference measurements are available. For example, for pipe cleaning processes, it is unlikely that an in-line or on-line reference measurement could be used, or that sampling could be conducted during the process. Therefore, a plateau in the waveform Energy could identify the potential end of cleaning and the process stopped and visual inspection of the pipe section to verify. In industrial applications, the wort density during fermentation is typically measured once or twice per day (Controllo e Misura, 2021). To label the remaining US measurements, techniques such as interpolation or semi-supervised learning may be used. Active learning, where analysis of model prediction confidence identifies datapoints that should be labelled, could be used to identify points in the process where the labelled data should be obtained and semi-supervised learning could be used to pseudo-label the remaining data (Zhou, Chen, and Wang, 2013). During industrial deployment, models will also need to be continuously trained (on-line learning) to adapt to changing process specifications, environmental conditions, or rare process states. If new labelled data can be collected, this new data must be combined with the previously collected data to prevent catastrophic forgetting (Kirkpatrick et al., 2016). However, to reduce computational time, an ensemble approach that uses simple neural networks trained on all of the available data or complex models trained on different subsections of data could be used to reduce overall training time. The self-training methods discussed in Section 10.4 (Other domain adaptation methodologies) can also be used to adapt models to changing environments, even where no new labelled data is available.

As discussed in Section 10.1 ‘Limitations of thesis’, further exploration of greater parameter ranges could be undertaken. For example, multiple datasets could be collected for each case study (mixing, cleaning, and fermentation) across varying process parameter ranges, other process parameter ranges could be made to be different other than temperature (for example, material quantities or impeller speed during honey-water blending), or more than two datasets with differing parameter ranges could be collected (e.g., a low, medium, and a larger parameter range dataset). Each of these methods could evaluated whether the insight that training ML models across a greater process parameter range improves transfer learning to further datasets was consistent for further investigations.

11.3 ML model trust

A further criteria for industrial implementation of US sensing and ML combinations could be to achieve the level of model trust desired by the manufacturer. ML models that output a confidence level associated with each prediction would increase operator trust. This may be achieved through an ensemble of models which outputs a prediction distribution although at higher computational cost. In the case of neural networks, differences in models can be achieved by using random weight initialisation (Gawlikowski et al., 2022). An ensemble of neural networks also overcomes the problem of adversarial examples. Adversarial examples are sets of features that fool a neural network into making an inaccurate prediction despite possibly being imperceptible to a human operator (Carlini and Wagner, 2017). While an adversarial example may confuse a single neural network, an ensemble of neural networks would be robust to these inputs (Pang et al., 2019). An ensemble of neural networks could also identify when input features are outside of the training data distribution. In this case, a wider range of predictions would be made and the confidence score will decrease highlighting that the models are uncertain about the prediction being made (Lakshminarayanan, Pritzel, and Blundell, 2017). This would indicate that the model predictions should not be trusted in this instance. For an added layer of protection, input features outside of the training distribution can also be identified using anomaly detection methods such as isolation forests (Scikit-learn, 2021). Identification that a new instance is from outside the training distribution can indicate to an operator that extra caution should be taken in using the ML model prediction or possible feature distributions on which the model should be retrained. Model generalisability may also be desired in industrial applications at the expense of accuracy. This is to ensure that ML models are robust to outliers, process parameter drift, or input features that were not captured by the training set distribution. One method of increasing model generalisability to out of distribution samples is to select the validation data as the datapoints furthest from the centre of the distribution. This could again be achieved through using anomaly detection methods such as isolation forests. Secondly, large validation sets could be used (for example, 2-fold cross-validation) to select hyperparameter that generalise from a small training set to a large validation set.

11.4 References

Andries, E., Kalivas, J.H., Gurung, A., (2019) ‘Sample and feature augmentation strategies for calibration updating’ *Journal of Chemometrics* 33(1), e308. Doi: 10.1002/cem.3080.

Carlini, N. and Wagner, D., (2017) ‘Towards Evaluating the Robustness of Neural Networks’ *Proceedings - IEEE Symposium on Security and Privacy* 2017, 39-57. Doi: 10.1109/SP.2017.49.

Controllo e Misura (2021) The Brewing Process Is under Control. Available online: https://www.publiteconline.it/controlloemisura/2021/09/10/the-brewing-process-is-under-control/ (Accessed 18th April 2023)

Gawlikowski, J., Tassi, C.R.N., Ali, M., Lee, J., Humt, M., Feng, J., Kruspe, A., Triebel, R., Jung, P., Roscher, R. et al. (2021) ‘A survey of uncertainty in deep neural networks’. *arXiv* arXiv:2107.03342.

Kirkpatrick, J., Pascanu, R., Rabinowitz, N., Veness, J., Desjardins, G., Rusu, A.A., Milan, K., Quan, J., Ramalho, T., Grabska-Barwinska, A., Hassabis, D., Clopath, C., (2017) ‘Overcoming catastrophic forgetting in neural networks’ *Proceedings of the National Academy of Sciences of the United States of America* 114(13), 3521-3526. Doi: 10.1073/pnas.1611835114.

Lakshminarayanan, B., Pritzel, A., Blundell, C. (2017) ‘Simple and scalable predictive uncertainty estimation using deep ensembles’ *Advances in Neural Information Processing Systems* 2017, 6403-6414.

Li, X., Li, Z., Yang, X., He, Y., (2021) ‘Boosting the generalization ability of Vis-NIR-spectroscopy-based regression models through dimension reduction and transfer learning’ *Computers and Electronics in Agriculture* 186, 106157. Doi: 10.1016/j.compag.2021.106157.

Lin, X., Zhuang, Y., Dan, T., Guanglin, L., Xiaodong, Y., Jie, S. and Xuwen, L., (2019) ‘The model updating based on near infrared spectroscopy for the sex identification of silkworm pupae from different varieties by a semi-supervised learning with pre-labeling method’ *Spectroscopy Letters*, 52(10), pp. 642-652. Doi: 10.1080/00387010.2019.1681463.

Mathworks (2021). Available online: https://uk.mathworks.com/help/stats/compactclassificationensemble.predictorimportance.html (accessed 19 February 2022).

Mishra, P. and Woltering, E., (2021) ‘Handling batch-to-batch variability in portable spectroscopy of fresh fruit with minimal parameter adjustment’ *Analytica Chimica Acta*, 1177. Doi: 10.1016/j.aca.2021.338771.

Pang, T., Xu, K., Du, C., Chen, N., Zhu, J. (2019) ‘Improving adversarial robustness via promoting ensemble diversity’ *36th International Conference on Machine Learning, ICML 2019* 2019, 8759-8771.

Scikit-learn (2021). Available online: https://scikit-learn.org/stable/auto_examples/miscellaneous/plot_anomaly_comparison.html#sphx-glr-auto-examples-miscellaneous-plot-anomaly-comparison-py (accessed 21 February 2022).

Yu, Y., Huang, J., Liu, S., Zhu, J., Liang, S., (2021) ‘Cross target attributes and sample types quantitative analysis modeling of near-infrared spectroscopy based on instance transfer learning’ *Measurement: Journal of the International Measurement Confederation* 77, 109340. Doi: 10.1016/j.measurement.2021.109340.

Zhou, S., Chen, Q. and Wang, X., (2013) ‘Active deep learning method for semi-supervised sentiment classification’ *Neurocomputing*, 120, 536-546. Doi: 10.1016/j.neucom.2013.04.017.

Glossary

Artificial neural networks: Artificial neural networks (ANNs) have the ability to create new features in their hidden layers from combinations of input features to perform the ML task. ANNs are formed of weight and bias terms that connect the model inputs to the outputs that are iteratively updated during the training procedure.

Convolutional neural networks: Convolutional neural networks (CNNs) have convolutional layers as well as fully connected layers. The convolutional layers consist of filters that perform cross-correlation on the input data. This enables CNNs to automatically learn spatially invariant features from the input data.

Domain adaptation: Domain adaptation is a subcategory of transfer learning where a model is trained on a source domain and transferred to a target domain where the data distributions may differ. Domain adaptations methodologies are therefore used to increase the accuracy of the model trained on the source domain to predict on the target domain. Unlabelled domain adaptation consists of domain adaptation methods where no labelled data is available in the target domain. In contrast, labelled domain adaptation methods use some labelled data from the target domain to aid transfer of the model.

Federated learning: Federated learning is a ML technique where multiple datasets can be used to collaboratively train a model without sharing the raw data with one another. Each dataset is used to train a local model and the model weights are aggregated to update a global model.

In-line and on-line sensing: In-line specifies techniques that directly measure the process stream and on-line techniques utilise automatic sampling methods.

Long-short term memory units: Long short-term memory neural networks (LSTMs) are able to learn sequences of time series data. LSTMs are a development of recurrent neural networks that reduce the likelihood of exploding or vanishing gradients and thereby enable the learning of long-term dependencies. LSTMs store representations of sequences by using gate units to update their internal network state. At each time step, LSTMs use the input features at the current time step as well as information passed from the previous time steps to make their prediction. Therefore, they have the capability to learn feature trajectories during processes that evolve over time.

Machine learning: Machine Learning (ML) is the use of computer algorithms to learn patterns in data to perform a task such as making predictions or decisions. The correlations in the data that the ML models learn during training have not been explicitly programmed by human operators. There are four main types of ML: supervised, unsupervised, semi-supervised, and reinforcement learning, of which supervised and unsupervised learning are used in this thesis.

Multi-task learning: Multi-task learning is an ML approach where a single model is trained on multiple related tasks simultaneously, sharing some or all of the model parameters between tasks to improve performance on all tasks.

Principal Component Analysis: Principal Component Analysis (PCA) is an unsupervised ML method that linearly transforms input variables into new, uncorrelated features called principal components (PCs).

Supervised machine learning: Supervised ML maps inputs to outputs during training with the aim being to create a model that accurately predicts the outputs of data that was not previously used during training.

Support vector machines: For classification tasks, support vector machines (SVMs) find a hyperplane that separates two classes of data by maximizing its distance from the closest data points from each category. In regression tasks, support vector regressors fit lines to continuous data by only accounting for the error from data points outside a set distance from the fitted line. SVMs generalise well to new data and, as they are effective with high dimensional feature spaces, make use of the kernel trick for non-linear fitting.

Transfer Component Analysis: TCA minimises the distance between source and target domain feature spaces by using the Maximum Mean Discrepancy (MMD) and extracts transfer components that maximise the variance across this shared feature space. The MMD is a measure of the distance between feature distribution embeddings in a reproducing kernel Hilbert space.

Transfer learning: Transfer learning is a machine learning technique where a model trained on one task is used to improve performance on a different but related task. The most common example is using pre-trained CNNs trained on large datasets to apply to new tasks.

Ultrasonic sensors: Low power (intensities below 1 Wcm2), high frequency (greater than 100 kHz) ultrasonic (US) sensors monitor the interaction of materials with mechanical sound waves. Ultrasonic sensors have advantages of being low cost, small in size, able to monitor opaque materials, low in power consumption, able to operate non-invasively, non-destructive, real-time, in-line, and do not cause changes to the structure of the material through which they pass.

Wavelet Transform: Continuous Wavelet Transform (CWT) uses a continuous range of frequencies to decompose the US signal whereas the Discrete Wavelet Transform (DWT) and Wavelet Packet Transform (WPT) use discrete frequencies at each decomposition. The WPT performs successive decompositions on each branch of the original signal whereas the DWT only applies successive decompositions to the lower frequency signal content.