Transfer Learning for Process Monitoring using Reflection-Mode Ultrasonic Sensing

Authors

Alexander L Bowler [a], Nicholas J Watson [a] *

[a] Faculty of Engineering, University of Nottingham, University Park, Nottingham, NG7 2RD, United Kingdom; alexander.bowler@nottingham.ac.uk (A.L.B.);

* Corresponding author: nicholas.watson@nottingham.ac.uk

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.

Keywords

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

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 (Thoben, et al. 2017). This has the potential to improve process productivity, raw material and energy efficiency, product quality and increase manufacturing sustainability (Ghobakhloo 2020). 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 (Awad, et al. 2012, Mohd Khairi, et al. 2015); monitoring chemical, pharmaceutical, and biotechnology processes (Henning and Rautenberg 2006); monitoring fermentation (Ojha, et al. 2017); monitoring freezing of food materials (Cheng, et al. 2015); and quality control in the dairy industry including monitoring reactions, process stream rheology, material structural changes, and component concentrations (Mohammadi, et al. 2014).

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 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 (Wallhäußer, et al. 2014, Wallhäußer, et al. 2013) and classifying weldment flaws (Munir, et al. 2018, Munir, et al. 2019). 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 (Bowler, et al. 2020b, Escrig, et al. 2020a, Escrig, et al. 2020b). 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 (Bowler, et al. 2020a). 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 (Eide, et al. 2003, Pettigrew, et al. 2015). 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 (Awad, et al. 2012). 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 (Pan and Yang 2010). For example, Zhu et al. (2021) 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. (2021) 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 (Kouw and Loog 2019). Several review articles covering aspects of domain adaptation are available to the interested reader: Patel et al. (2015), Csurka (2017), Wang and Deng (2018), Pan and Yang (2010), Weiss et al (2016). Heimann et al. (2014) 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. (2017) 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. (2021) 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, Zhang et al. 2021 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. (2021) 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. (2017), 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 (McClements 1995). 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. (2020b). 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 Figure 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 Figure 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 °C to 22.1 °C. Mixing Dataset 2 had a temperature variation of 19.8 °C to 21.2 °C.

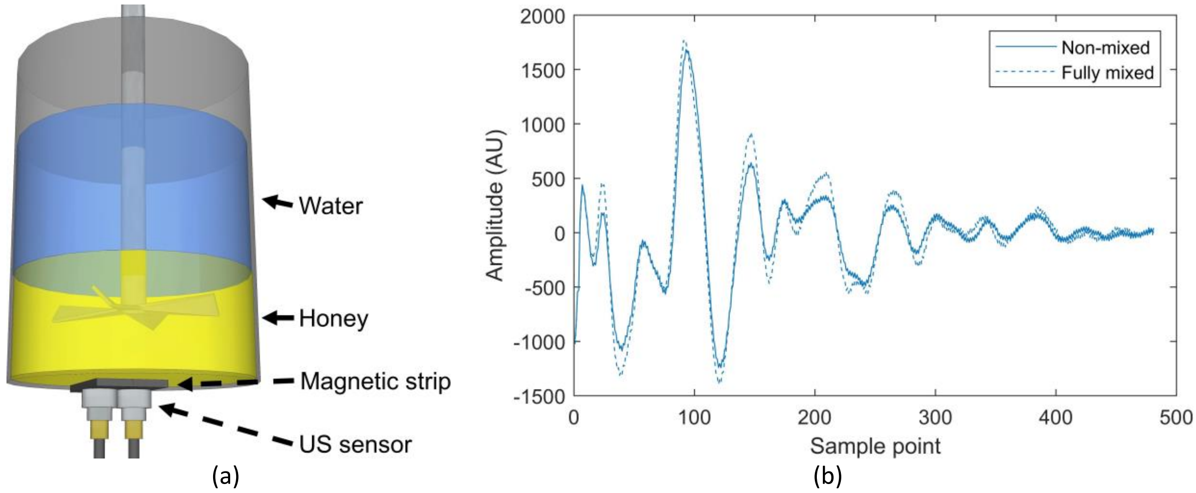


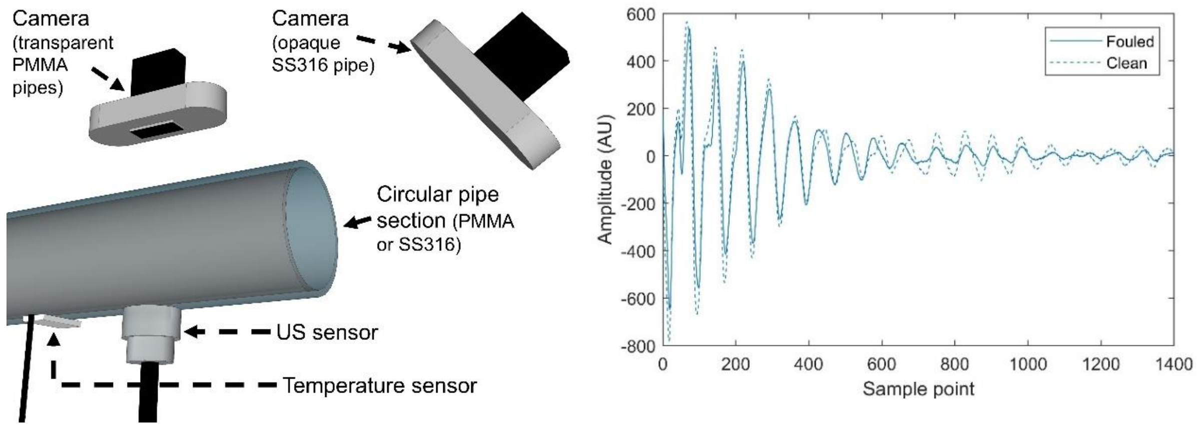
Figure 1. (a) A diagram of the equipment for the mixing experiments; including 250 ml glass vessel, impeller, and US sensors (Adapted from Bowler, et al. (2020b)). (b) Two received US waveforms corresponding to a non-mixed and a fully mixed system.

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

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. (2019) 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. 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 minutes 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 Figure 2a. US and temperature data were recorded every 4 seconds and camera images were recorded every 20 seconds during the cleaning process. The camera images were used as the ground truth data to label the recorded US data for ML model development.



(a) (b)

Figure 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.

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

2.4 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) (Caesarendra and Tjahjowidodo 2017), the amplitude at every sample point in the waveform (Escrig, et al. 2020) or frequency components obtained after Fourier or Wavelet transforms (Bowler, et al. 2020b). However, US waveforms vary each time a sensor is attached. This effect is presented in Figure 3, where each US waveform differs despite using the same sensors, attachment procedure, vessel and process material. Furthermore, Figure 4 compares waveforms obtained from Cleaning Datasets 5 and 9, where different pipe construction materials and US sensors were used.

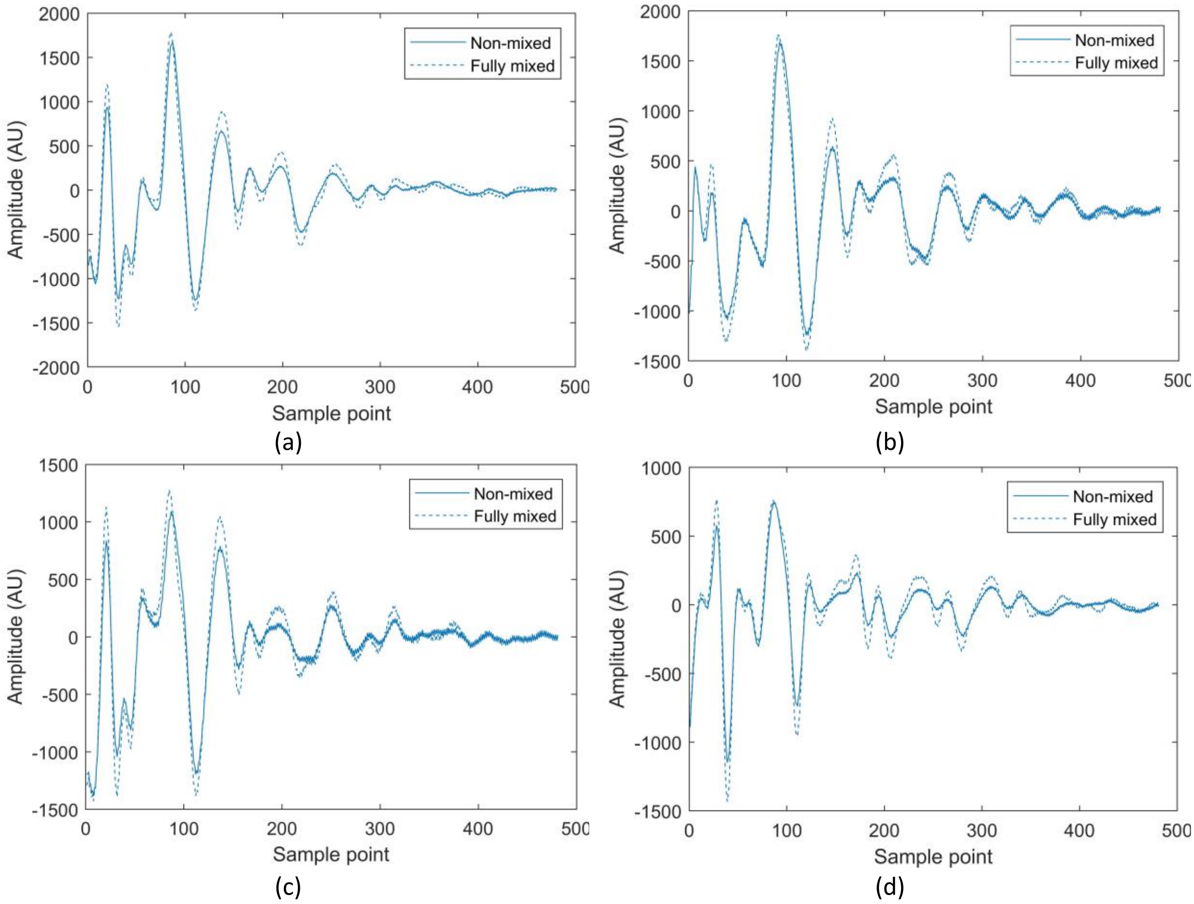


Figure 3. US waveforms from the mixing experiments corresponding to non-mixed and fully mixed systems. (a) Dataset 1 Non-Central sensor. (b) Dataset 2 Non-Central sensor. (c) Dataset 1 Central sensor. (d) Dataset 2 Central sensor.

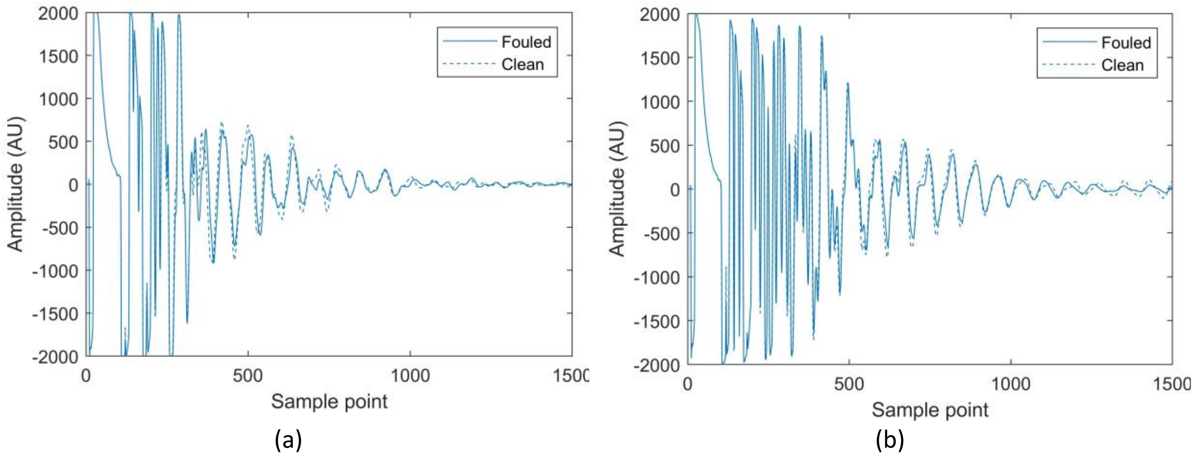


Figure 4. US waveforms from the pipe cleaning experiments corresponding to fouled and clean pipe section. (a) Dataset 5 – circular plastic pipe section. (b) Dataset 9 – circular metal pipe section.

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. (2019). 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 (Pan, et al. 2011). 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 Figure 5. All data analysis and ML algorithms were completed in MATLAB R2019a.

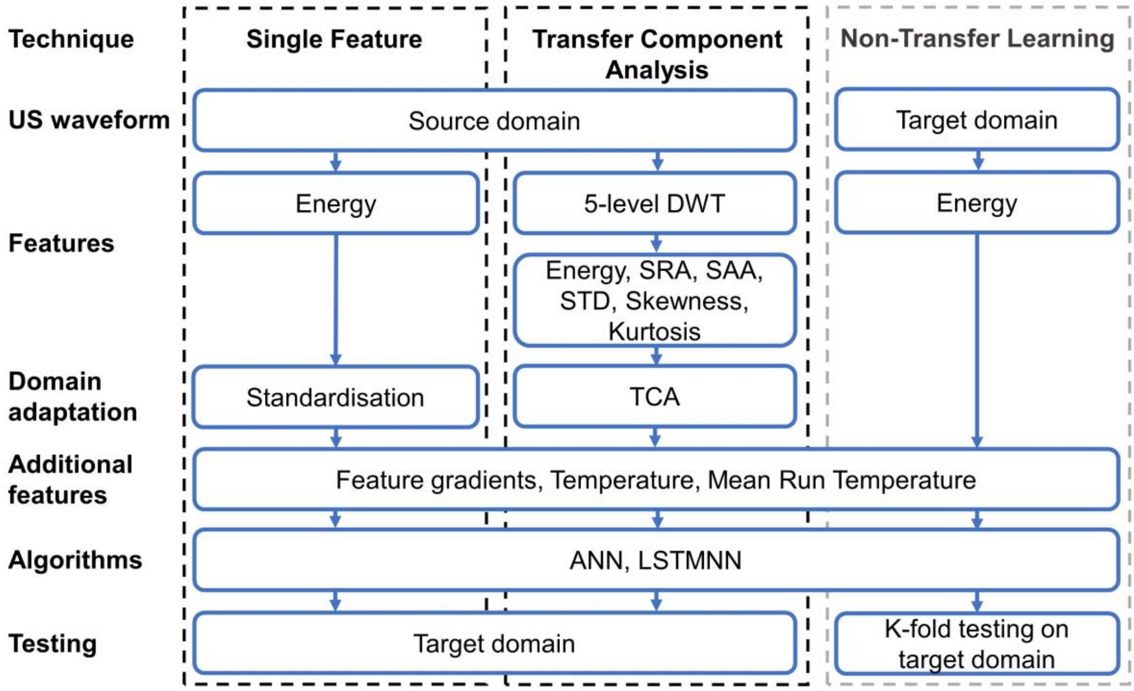


Figure 5. A methodology flow diagram for the three models being compared. The two transfer learning models, SF and TCA, and the Non-Transfer Learning model.

2.4.1 Features

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

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

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* (Zhan, et al. 2015).

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

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

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

Where *SAA* is the sum absolute amplitude (Zhan, et al. 2015).

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

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

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

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 of a parameter, *N* is the size of backwards, one-sided moving mean, and *V* is the original parameter value (Mathworks 2020a, Mathworks 2020b).

2.4.1.2 Temperature and Mean Run Temperature

As the acoustic properties of materials are highly dependent on temperature (Henning and Rautenberg 2006), 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 (Mallat and Mallat 1999a). At each decomposition, an orthogonal wavelet transform function produces a detail and approximate waveform which contain no redundant information (Mallat 1989). 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 (Mallat and Mallat 1999b), along with its previous success in analysing US waveforms (Bowler, et al. 2020b). 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.

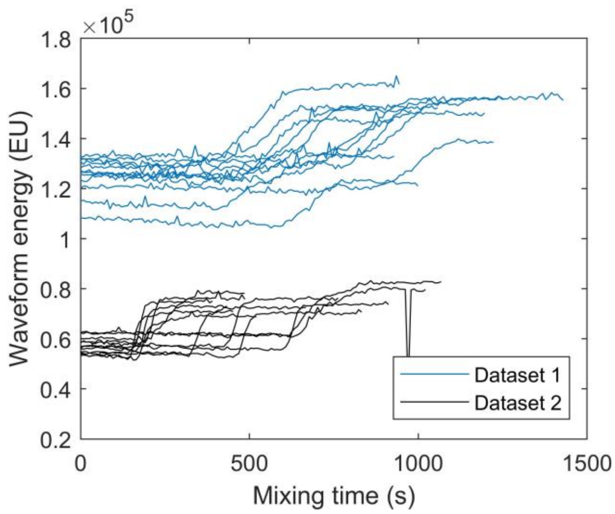
$$\mu = \frac{\sum_{i=1}^{i=n} x_i}{n} \tag{10}$$

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

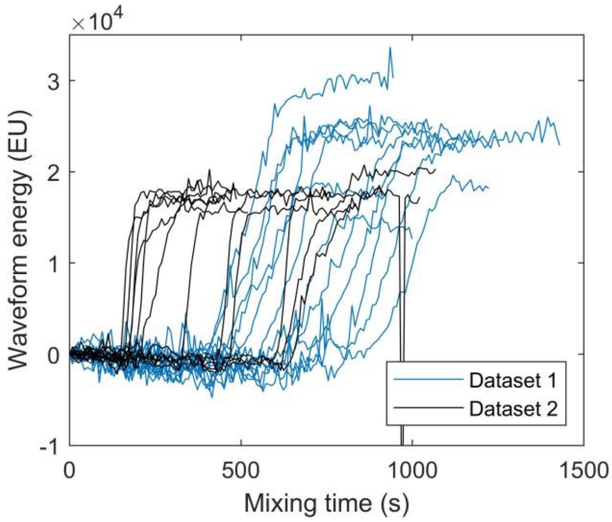
$$Z = \frac{x - \mu}{\sigma} \tag{12}$$

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.

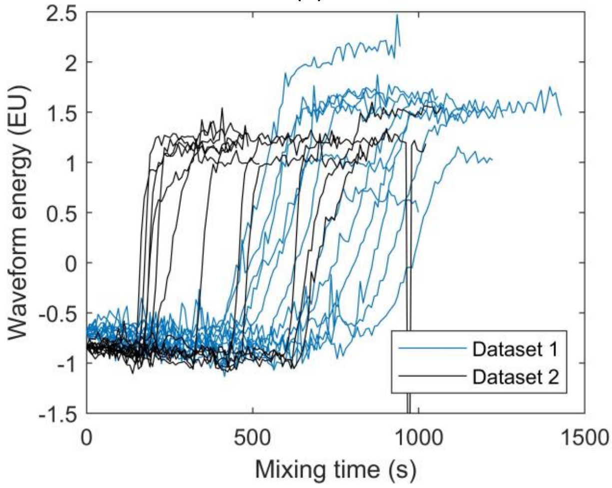
Furthermore, for the honey-water blending experiments, prior to standardisation, the waveform energy of the first data point in each run was subtracted from all data points of that run so that they all began at a waveform energy of 0. The process material being measured at the start of each run is known 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 Figures 6 and 7, respectively.



(a)

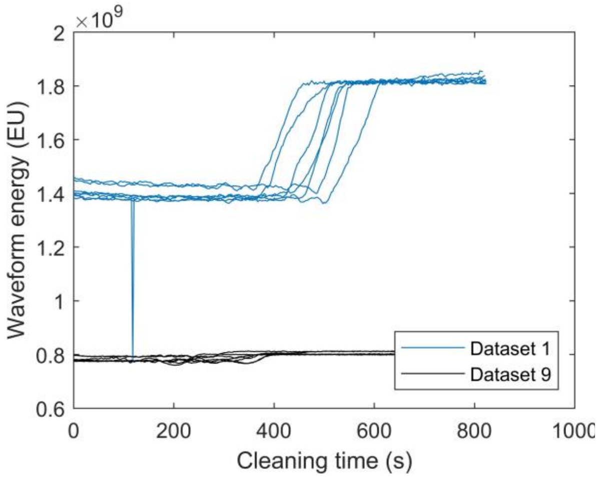


(b)

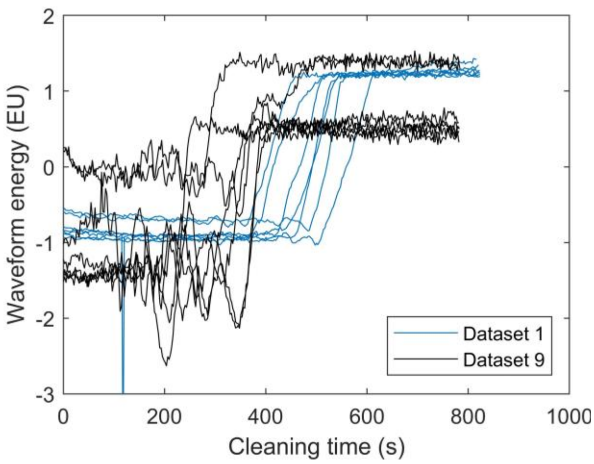


(c)

Figure 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.



(a)



(b)

Figure 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.

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 (Pan, et al. 2011). 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 (2020) 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 (Jain, et al. 1996). 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 regression models and the “trainscg” training function was used for the classification models (Mathworks 2020c). 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 (Hochreiter and Schmidhuber 1997). 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.

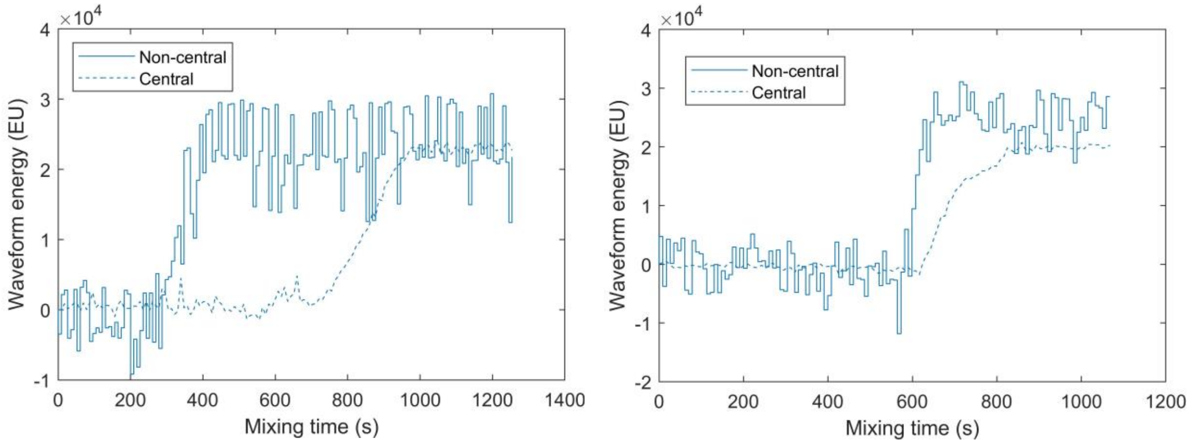
3.1.1 Classification

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 Figure 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

(Bowler, et al. 2020b). 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 %. 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 Figure 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 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.



(a) (b)

Figure 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.

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

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 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 (Bowler, et al. 2020b). 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.

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

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.

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

3.2.2 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 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

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 (Bowler, et al. 2020b). For the cleaning of pipe fouling, classification of accuracies of between 91.6-98.4 % are achieved in this work compared with previous results of 98-100 % (Escrig, et al. 2020a, Escrig, et al. 2020b). These results are similar to the domain adaptation methodologies used for motor bearing fault diagnosis by vibration signal monitoring. Wen et al. (2018) 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. (2018) achieved average classification accuracies of 95.5 % using a CNN based domain adaptation method across different load domains and Li et al. (2019) achieved accuracies >92 % using a generative model. Furthermore, Guo et al. (2019) 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.

Author Contributions: Conceptualization, A. Bowler, N. Watson; Data curation, A. Bowler; Formal analysis, A. Bowler; Funding acquisition, N. Watson; Investigation, A. Bowler; Methodology, A. Bowler; Project administration, A. Bowler, N. Watson; Resources, N. Watson; Software A. Bowler; Supervision, N. Watson; Validation, A. Bowler; Visualization, A. Bowler; Roles/Writing - original draft, A. Bowler; Writing - review & editing, A. Bowler, N. Watson.

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

5 References

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

Awad, T.S., H.A. Moharram, O.E. Shaltout, D. Asker, and M.M. Youssef. 2012. “Applications of ultrasound in analysis, processing and quality control of food: A review.” *Food Research International* 48 (2): 410-427. doi:10.1016/j.foodres.2012.05.004.

Azizi, S., Mousavi, P., Yan, P., Tahmasebi, A., Kwak, J.T., Xu, S., Turkbey, B., Choyke, P., Pinto, P., Wood, B., Abolmaesumi, P. 2017. “Transfer learning from RF to B-mode temporal enhanced ultrasound features for prostate cancer detection.” *International Journal of Computer Assisted Radiology and Surgery* 12 (7): 1111-1121. doi:10.1007/s11548-017-1573-x.

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

Bowler, A.L., S. Bakalis, and N.J. Watson. 2020b. “Monitoring mixing processes using ultrasonic sensors and machine learning.” *Sensors (Switzerland)* 20 (7). doi:10.3390/s20071813.

Caesarendra, W., and T. Tjahjowidodo. 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.

Cheng, X., M. Zhang, B. Xu, B. Adhikari, and J. Sun. 2015. “The principles of ultrasound and its application in freezing related processes of food materials: A review.” *Ultrasonics Sonochemistry* 27: 576-585. doi:10.1016/j.ultsonch.2015.04.015.

Csurka, G. 2017. “A comprehensive survey on domain adaptation for visual applications.” *Advances in Computer Vision and Pattern Recognition.* (9783319583464): 1-35. doi:10.1007/978-3-319-58347-1_1.

Eide, M.H., J.P. Homleid, and B. Mattsson. 2003. “Life cycle assessment (LCA) of cleaning-in-place processes in dairies.” *LWT - Food Science and Technology* 36 (3): 303-314. doi:10.1016/S0023-6438(02)00211-6.

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

Escrig, J.E., A. Simeone, E. Woolley, S. Rangappa, A. Rady, and N.J. Watson. 2020b. "Ultrasonic measurements and machine learning for monitoring the removal of surface fouling during clean-in-place processes." *Food and Bioproducts Processing* September: 1-13. doi:10.1016/j.fbp.2020.05.003.

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

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 doi:10.1109/TIM.2021.3050154.

Ghobakhloo, M. 2020. “Industry 4.0, digitization, and opportunities for sustainability.” *Journal of Cleaner Production* 252. doi:10.1016/j.jclepro.2019.119869.

Guo, L., Y. Lei, S. Xing, T. Yan, and N. Li. 2019. “Deep Convolutional Transfer Learning Network: A New Method for Intelligent Fault Diagnosis of Machines with Unlabeled Data.” *IEEE Transactions on Industrial Electronics* 66 (9): 7316-7325. doi:10.1109/TIE.2018.2877090.

Heimann, T. Mountney, P., John, M., Ionasec, R. 2014. “Real-time ultrasound transducer localization in fluoroscopy images by transfer learning from synthetic training data.” *Medical Image Analysis* 18 (8): 1320-1328. doi:10.1016/j.media.2014.04.007

Henning, B., and J. Rautenberg. 2006. “Process monitoring using ultrasonic sensor systems.” *Ultrasonics* 44 (SUPPL): e1395-e1399. doi:10.1016/j.ultras.2006.05.048.

Hochreiter, S., and J. Schmidhuber. 1997. “Long Short-Term Memory.” *Neural Computation* 9 (8): 1735-1780. doi:10.1162/neco.1997.9.8.1735.

Jain, A.K., J. Mao, and K.M. Mohiuddin. 1996. “Artificial Neural Networks: A Tutorial.” *Computer* 29 (3): 31-44. doi:10.1109/2.485891.

Kouw, W.M., and M. Loog. 2019. “A review of domain adaptation.” *IEEE transactions on pattern analysis and machine intelligence.* doi:10.1109/TPAMI.2019.2945942.

Li, X., W. Zhang, and Q. Ding. 2019. “Cross-domain fault diagnosis of rolling element bearings using deep generative neural networks.” *IEEE Transactions on Industrial Electronics* 66 (7): 5525-5534. doi:10.1109/TIE.2018.2868023.

Mallat, S.G. 1989. “A Theory for Multiresolution Signal Decomposition: The Wavelet Representation.” *IEEE Trans. Pattern Anal. Mach. Intell.* 11: 674-693. doi:10.1109/34.192463.

Mallat, S.G., and C. Mallat. 1999a. “IV Time meets frequency.” *A Wavelet Tour of Signal Processing* 2: 67-124.

Mallat, S.G., and C. Mallat. 1999b. *7.2 CLASSES OF WAVELET BASES.* Elsevier Science & Technology.

Mathworks. 2020a. *Gradient.* Accessed May 27, 2020. https://uk.mathworks.com/help/matlab/ref/gradient.html#bvhp8_i.

Mathworks. 2020b. *Movmean.* Accessed May 27, 2020. https://uk.mathworks.com/help/matlab/ref/movmean.html#bu2yug_-1_seealso.

Mathworks. 2020c. *Choose a Multilayer Neural Network Training Function.* Accessed May 27, 2020. https://uk.mathworks. com/help/deeplearning/ug/choose-a-multilayer-neural-network-training-function.html;jsessionid= e378b9dfbf595a83f44348fc1e7c.

McClements, D.J. 1995. "Advances in the application of ultrasound in food analysis and processing." *Trends in Food Science and Technology* 6 (9): 293-299. doi:10.1016/S0924-2244(00)89139-6.

Meng, Q., Matthew, J., Zimmer, V.A., Gomez, A., Lloyd, D.F.A., Rueckert, D., Kainz, B. 2021. "Mutual Information-Based Disentangled Neural Networks for Classifying Unseen Categories in Different Domains: Application to Fetal Ultrasound Imaging." *IEEE Transactions on Medical Imaging* 40 (2): 722-734. doi:10.1109/TMI.2020.3035424.

Mohammadi, V., M. Ghasemi-Varnamkhasti, R. Ebrahimi, and M. Abbasvali. 2014. "Ultrasonic techniques for the milk production industry." *Measurement: Journal of the International Measurement Confederation* 58 (2014): 93-102. doi:10.1016/j.measurement.2014.08.022.

Mohd Khairi, M.T., S. Ibrahim, M.A. Md Yunus, and M. Faramarzi. 2015. "Contact and non-contact ultrasonic measurement in the food industry: A review." *Measurement Science and Technology* 27 (1). doi:10.1088/0957-0233/27/1/012001.

Munir, N., H.-J. Kim, J. Park, S.-J. Song, and S.-S. Kang. 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., H.-J. Kim, S.-J. Song, and S.-S. Kang. 2018. "Investigation of deep neural network with drop out for ultrasonic flaw classification in weldments." *Journal of Mechanical Science and Technology* 32 (7): 3073-3080. doi:10.1007/s12206-018-0610-1.

Ojha, K.S., T.J. Mason, C.P. O'Donnell, J.P. Kerry, and B.K. Tiwari. 2017. "Ultrasound technology for food fermentation applications." *Ultrasonics Sonochemistry* 34: 410-417. doi:10.1016/j.ultsonch.2016.06.001.

Pan, S.J., and Q. Yang. 2010. "A survey on transfer learning." *IEEE Transactions on Knowledge and Data Engineering* 22 (10): 1345-1359. doi:10.1109/TKDE.2009.191.

Patel, V.M., Gopalan, R., Li, R., Chellappa, R. 2015. "Visual Domain Adaptation: A survey of recent advances." *IEEE Signal Processing Magazine* 32 (3): 53-69. doi:10.1109/MSP.2014.2347059.

Pan, S.J., I.W. Tsang, J.T. Kwok, and Q. Yang. 2011. "Domain adaptation via transfer component analysis." *IEEE Transactions on Neural Networks* 22 (2): 199-210. doi:10.1109/TNN.2010.2091281.

Pettigrew, L., V. Blomenhofer, S. Hubert, F. Groß, and A. Delgado. 2015. "Optimisation of water usage in a brewery clean-in-place system using reference nets." *Journal of Cleaner Production* 87 (1): 583-593. doi:10.1016/j.jclepro.2014.10.072.

Thoben, K.-D., S. A. Wiesner, and T. Wuest. 2017. ““Industrie 4.0” and smart manufacturing - a review of research issues and application examples.” *International Journal of Automation Technology* 11 (1): 4-16. doi:10.20965/ijat.2017.p0004.

Wallhäußer, E., A. Sayed, S. Nöbel, M.A. Hussein, J. Hinrichs, and T. Becker. 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.

Wallhäußer, E., W.B. Hussein, M.A. Hussein, J. Hinrichs, and T. Becker. 2013. “Detection of dairy fouling: Combining ultrasonic measurements and classification methods.” *Engineering Life Sciences* 13 (3): 292-301. doi:10.1002/elsc.201200081.

Wang, M., Deng, W. 2018. “Deep visual domain adaptation: A survey.” *Neurocomputing* 312: 135-153. doi:10.1016/j.neucom.2018.05.083.

Weiss, K., Khoshgoftaar, T.M., Wand, D.D. 2016. “A survey of transfer learning.” *Journal of Big Data* 3 (1): 9. doi:10.1186/s40537-016-0043-6

Wen, L., X. Li, L. Gao, and Y. Zhang. 2018. “A New Convolutional Neural Network-Based Data-Driven Fault Diagnosis Method.” *IEEE Transactions on Industrial Electronics* 65 (7): 5990-5998. doi:10.1109/TIE.2017.2774777.

Yan, K. 2020. *A domain adaptation toolbox.* Accessed June 14, 2020. https://www.github.com/viggin/domain-adaptation-toolbox.

Zhan, X., S. Jiang, Y. Yang, L. Jian, T. Shi, and X. Li. 2015. “Inline Measurement of Particle Concentrations in Multicomponent Suspensions using Ultrasonic Sensor and Least Squares Support Vector Machines.” *Sensors (Basel, Switzerland)* 15: 24109-24124. doi:10.3390/s150924109.

Zhang, W., C. Li, G. Peng, Y. Chen, and Z. Zhang. 2018. “A deep convolutional neural network with new training methods for bearing fault diagnosis under noisy environment and different working load.” *Mechanical Systems and Signal Processing* 100: 439-453. doi:10.1016/j.ymssp.2017.06.022.

Zhang, W., Wu, T., Li, Z., Liu, S., Qiu, A., Li, Y., Shi, Y. 2021 “Fracture recognition in ultrasonic logging images via unsupervised segmentation network.” *Earth Science Informatics* (Article in press). doi:10.1007/s12145-021-00605-6.

Zhu, Y.-C., A. AlZoubi, S. Jassim, Q. Jiang, Y. Zhang, Y.-B. Wang, X.-D. Ye, and H. Du. 2021. “A generic deep learning framework to classify thyroid and breast lesions in ultrasound images.” *Ultrasonics* 110: 106300. doi:10.1016/j.ultras.2020.106300