A Comparative Study of Machine Learning Approaches for Anomaly Detection in Industrial Screw Driving Data

Nikolai West

RIF Institute for Research and Transfer e.V.
Work and Production Systems
Dortmund, Germany

Institute of Production System
Technical University Dortmund
Dortmund, Germany

nikolai.west@tu-dortmund.de

Jochen Deuse

Institute of Production Systems
Technical University Dortmund
Dortmund, Germany

Centre for Advanced Manufacturing
University of Technology Sydney
Sydney, Australia

jochen.deuse@ips.tu-dortmund.de

Abstract

This paper investigates the application of Machine Learning (ML) approaches for anomaly detection in time series data from screw driving operations, a pivotal process in manufacturing. Leveraging a novel, open-access real-world dataset, we explore the efficacy of several unsupervised and supervised ML models. Among unsupervised models, DBSCAN demonstrates the best performance with an accuracy of 96.68% and a Macro F1 score of 90.70%. Within the supervised models, the Random Forest classifier excels, achieving an accuracy of 99.02% and a Macro F1 score of 98.36%. These results not only underscore the potential of ML in boosting manufacturing quality and efficiency, but also highlight the challenges in their practical deployment. This research encourages further investigation and refinement of ML techniques for industrial anomaly detection, thereby contributing to the advancement of resilient, efficient, and sustainable manufacturing processes. The entire analysis, comprising the complete dataset as well as the Python-based scripts are made publicly available via a dedicated repository. This commitment to open science aims to support the practical application and future adaptation of our work to support business decisions in quality management and the manufacturing industry.

Keywords: Anomaly detection, screw driving operations, tightening process, supervised learning, unsupervised learning.

1. Introduction

Anomaly detection in manufacturing operations is a cornerstone for maintaining process quality and efficiency (Schlegl et al., 2022; Stojanovic et al., 2016). The advent of *Machine Learning* (ML) in recent decades has transformed this area, paving the way for new methods and innovative strategies (Glaser et al., 2022). ML has opened up a wide field of possibilities for improved monitoring of screw driving data. As a subset of *Artificial Intelligence*, ML uses algorithms that allow computers to learn from data and make decisions or predictions without specific programming. Anomaly detection, especially unsupervised learning, can identify complex failure patterns and thus represents a significant improvement to existing quality assurance processes that often require prior knowledge of the nature of the anomalies (Chandola et al., 2009).

The aim of this paper is to expand the scope of business decisions in quality management by introducing a data-driven approach to detecting complex error scenarios. This entails the exploration of anomaly detection methods based on unsupervised ML techniques, which will be tested and validated in the context of a screw driving case study. In achieving this, the paper makes several important contributions:

- Offering a comprehensive summary of the contributions made by related research in the field of anomaly detection, thereby enabling readers to understand the landscape of existing solutions and their implications more clearly (see **Sec. 3**).
- Providing a new real-world data set of screw driving operations, thereby enriching the resources available for studying these processes (see **Sec. 5**).
- Conducting an extensive study of popular models using the provided dataset, thereby, contributing to the understanding of their practical applications and limitations (see **Sec. 6**).

Following this introduction, we will explore the fundamentals of tightening processes as well as anomaly

URI: https://hdl.handle.net/10125/106504
978-0-9981331-7-1

detection. Then, we provide a review of related works, an introduction to the methods as well as the case study, and finally a discussion of our results and their implications, followed by a brief conclusion.

2. Fundamentals

2.1. Tightening process

Various methods exist within the manufacturing industry to oversee and manage the screw tightening process. These methods differ in their supervisory techniques and the parameters they use for control. For instance, *torque-controlled* methods monitor the applied torque, ceasing the screwing process once a preset torque is attained. This approach is cost-effective and efficient for a majority of applications (J. H. Bickford & Nassar, 1998). In contrast, *angle-controlled* methods govern the screwing process via the rotation angle. After reaching a specific torque, the screw undergoes further rotation over a set angle, providing superior control and repeatability, particularly in safety-critical connections (J. Bickford, 1995). Other techniques include pulse-controlled, yield point-controlled, torsion-free tightening, or hybrid methods that combine torque and angle control. The use case in this paper mainly utilizes the torque-controlled method.

Upon examining the tightening process, the role of collected data in assuring its overall quality becomes evident. Handheld screwdrivers or automated stations gather thousands of *angle-torque pairs*, essential for enabling operators to observe the tightening curve and deduce process correctness. This data collection serves a dual purpose; beyond offering immediate feedback, it aids in developing a *defect catalog*. Such a catalog proves invaluable in understanding potential tightening process issues and mitigating them, thus fostering continual enhancements in quality. The amassed data points allow for meticulous inspection of each unit, aiding in isolating those needing further examination. This approach prevents the transition of defective units to subsequent assembly stations, thereby reinforcing the quality control within the manufacturing process.

Maintaining a high degree of process quality during tightening is crucial for any manufacturing process, emphasizing the role of data analyses. They can provide a visual representation of the complex interplay between torque and angle. In the simplest form, such visualizations allow process experts to monitor each process, with regard to a collected defect catalog. For a higher degree of automation, statistical tools for quality monitoring, such as specific quality control limits, are used more frequently (Schlegl et al., 2021). While these approaches are widely used and easy to apply, they also exhibit limitations. Stochastic and technological noise within the collected data may introduce anomalies, potentially resulting in incorrect process interpretations. Processes falsely labeled as erroneous are not as damaging in this regard as observations falsely labeled as correct. Nevertheless, they are accompanied by increased expenses for manual inspections or rework.

This emphasizes the need for more sophisticated analytics, capable of effectively identifying and addressing these issues. Such analytics could transform the tightening process and quality control. The following section reviews anomaly detection methods and their capability to improve quality control.

2.1. Anomaly detection

In practice, an *anomaly* is characterized as a pattern that deviates from expected behavior (Chandola et al., 2009; Pang et al., 2021). A practical approach to identify these patterns involves establishing a range of values indicative of normal behavior. However, several multi-faceted influencing factors make this straightforward approach more challenging. These hurdles can encompass the multivariate character of time-series-based use cases, an ambiguous boundary separating normal and anomalous behavior, the potential for normal behavior to shift over time, the presence of noise or other interferences in the data, and the scarcity of labeled data to define and validate decision boundaries.

As stated before, traditional inspection methods usually require substantial human effort to detect potential defects or deviations. As such, these manual inspections are generally regarded as time-consuming, exhausting, and prone to errors (Cao et al., 2019). Given the repetitive nature of these tasks, operators might experience fatigue or miss subtle differences, resulting in inaccurate findings. This could have significant ramifications in industries like manufacturing, where quality control is essential. Furthermore, traditional methods often fall short in adapting to the dynamic and complex nature of manufacturing processes. This is especially true concerning screw tightening, where numerous factors such as torque, angle, and position can influence assembly quality. Thus, ensuring consistent and precise inspections in such complex conditions becomes challenging, particularly for human operators. As a result, contemporary strategies are increasingly adopting automated and algorithmic approaches from the anomaly detection domain (see **Sec. 3**). These approaches leverage the potential of advanced analytics and ML to minimize the need for human inspections, offering a more accurate, consistent, and efficient way to detect anomalies in manufacturing processes.

Deep Learning and ML, particularly *Deep Neural Network* (DNN), are especially relevant in this context. A DNN can learn intricate patterns and create a model capable of distinguishing between defective and non-defective instances through training with known examples of normal and abnormal cases. This concept, known as supervised learning, has yielded notable outcomes in industrial settings. However, supervised learning methods have a significant limitation in that they require labeled instances of damage patterns during training to accurately classify them during subsequent operational use. Given the high standardization level of industrial processes, instances of relevant damage patterns are seldom available. This implies that deviations from normal conditions occur infrequently, rendering the collection of a sufficient number of labeled examples that accurately reflect representative error types nearly impossible. Moreover, generating and labeling anomalous samples synthetically for model training can be expensive and time-consuming.

3. Related work

The following review of related work presents recent research and applications of ML approaches for anomaly detection in screw driving operations. The studies encompass a range of approaches including unsupervised and supervised models. The overarching objective is to identify reoccurring models as well as the structure of the explored screw driving scenarios. **Table 1** provides an overview of the works discussed.

Cheng et al. (2019) refer to their work as the first known approach of unsupervised learning in tightening data. In it, they investigate the use of *Hidden Markov Models* (HMM) to identify erroneous patterns in 1,013 samples of four different screw types. HMMs are statistical models capable of representing stochastic processes with hidden states, and they can be applied for anomaly detection, where unusual sequences of observed outcomes, deviating from the normal patterns inferred by trained HMMs, are identified as potential anomalies (Dorj & Altangerel, 2013). In their data, with 75% faults deliberately caused by various types of misalignment errors, the authors manage to determine the classes of screw runs with an accuracy of over 97%.

In the same year, *Cao et al.* (2019) demonstrated a successful application of *Long Short-Term Memory Networks* (LSTM), a specialized type of *Recurrent Neural Networks* (RNN), to detect anomalies in screw driving data, achieving an accuracy of 93% within a dataset of 2,000 observations with four distinct classes. LSTM help to alleviate the vanishing gradient problem inherent in traditional RNN, thus enhancing their ability to learn from, and remember, information over long sequences of data (Hochreiter & Schmidhuber, 1997). The supervised approach significantly outperforms the two benchmarks, both also supervised, *Support Vector Machine* and *Random Forest*, and emphasizes the good suitability of LSTM for use cases with time series data.

The work of *Li et al.* (2020), in which *Synthetic Minority Over-sampling Technique* (SMOTE) is used in combination with *Density-Based Spatial Clustering of Applications with Noise* (DBSCAN) for anomaly detection in screw data, introduces a hybrid or semi-supervised approach. SMOTE is a supervised learning technique since it requires class labels to generate synthetic samples for minority classes (Chawla et al., 2002), while DBSCAN is a prominent, unsupervised method used for clustering (Ester et al., 1996). In combination, the methods are often applied to use cases with unevenly distributed classes (Sanguanmak & Hanskunatai, 2016), a scenario that is typical for screw driving use cases, since faults usually occur several

Table 1: Overview of related work on supervised and unsupervised machine learning for anomaly detection in tightening data

Source	Approach	Proposed or deployed methods *(Methods selected for benchmarking)*	Observations?		Open-Access?	
			#OK	#NOK	Data	Code
Cheng et al. (2019)	Unsupervised	Hidden Markov Model *(none)*	253	760	No	No
Cao et al. (2019)	Supervised	LSTM-RNN *(Random Forest, SVM)*	664	1,336	No	No
Li et al. (2020)	Hybrid	SMOTE with DBSCAN *(none)*	98,693	1,703	No	No
Ribeiro et al. (2021)	Unsupervised	LOF, iForest, AE *(Random Forest)*	6,088	74	No	No
Schlegl et al. (2021)	Unsupervised	Shape-based AE *(traditional AE)*	9,950	50	No	No
West et al. (2021)	Supervised	kNN, Naïve Bayes, DT, MLP, Random Forest, AdaBoost *(six DL-classifier)*	14,083 100	60 100	No Yes	Yes
Ribeiro et al. (2022)	Unsupervised	LOF, iForest, AE *(Random Forest)*	67,337	220	No	No
Leporowski et al. (2022)	Supervised	ResNet, Temporal Attention-augmented Bilinear Network *(none)*	1,420 1,574	625 288	Yes Yes	Yes
Sakamoto et al. (2023)	Unsupervised	iForest *(none)*	300	100	No	No
West et al. (2023)	Unsupervised	k-Means clustering with DTW *(none)*	50,000	96	No	Yes

orders of magnitude less frequently than correct runs do. With this hybrid approach, the authors manage to achieve an accuracy of more than 99% in a scenario with a very large data set with over one hundred thousand screw runs and four variants of different screw types.

Instead of just one model, *Ribeiro et al.* (2021) test and evaluate three unsupervised approaches: *Local Outlier Factor* (LOF), *Isolation Forest* (iForest), and a *Deep Learning Autoencoder* (AE). LOF is an unsupervised anomaly detection method that calculates the local density deviation of a given sample with respect to its neighbors, identifying those that have significantly lower density as outliers (Breunig et al., 2000). iForest is an anomaly detection algorithm that isolates observations by randomly selecting a feature and then randomly selecting a split value between the maximum and minimum values of the selected feature, with anomalies being identified as those observations that require fewer random partitions to be isolated (Liu et al., 2008). AE are a type of neural network used for learning efficient codings of input data, typically used for dimensionality reduction or anomaly detection, by training the network to reconstruct its inputs, with the hidden layers encoding a compressed representation of the input (Zhou & Paffenroth, 2017). In their scenario with 6,162 observations, iForest performed best with a 99% accuracy, while AE with about 96% and LOF with at least 80% also produced acceptable results. At the same time, a *Random Forest* classifier, chosen as a supervised benchmark, also achieved a result of 99%. Only about a year later, the authors repeated the analysis for a related scenario with significantly more observations (Ribeiro et al., 2022). With 67,337 screw runs, ten times as many observations were available. Despite this, the authors managed to obtain results of over 99% using iForest and AE, thus again demonstrating the suitability of their approaches.
Schlegl et al. (2021) implement a custom implementation of a Deep Learning model with an AE-logic specialized for mapping characteristic fault shapes in screw driving data. The proposed model, composed of two sub-networks, learns interpretable representations of normal process behaviors from manufacturing sequences through a custom convolution operation and unique loss function, and then employs these learned representations within a convolutional-RNN-AE structure to perform anomaly detection, using the inverse reconstruction error as a measure of normality. Besides an increase in interpretability, the model achieves a better detection rate compared to a conventional AE approach.

West et al. (2021) present an approach that aims at efficient feature extraction through statistical representations of time series data. In addition to using a variety of traditional classification models, such as *k-Nearest Neighbors* (kNN), *Gausian Naive Bayes*, *Decision Trees* (DT), *Multi-layer Perceptrons* (MLP), *Random Forest* and *Adaptive Boosting* (AdaBoost), the paper stands out for being the first to publish the code to their analysis. Furthermore, in addition to the application for an unpublished screw driving scenario from the automotive industry, they demonstrate the suitability of the developed approach for an open-source dataset with 200 observations. Unfortunately, this was no screw driving scenario, but a case study with time series observations of human motion data.

Leporowski et al. (2022) (2022) employ two models, called *Residual Neural Networks* (ResNet) and *Temporal Attention-augmented Bilinear Network*, to classify screw driving data. Of particular note is that the authors used their own dataset with 2,045 observations, called AURSAD, which they previously made publicly available (Leporowski et al., 2021). Furthermore, they apply the approach to another open-source screw driving dataset, called *The Manipulation Lab Screwdriving Dataset* (Aronson et al., 2017) and make the generated analysis publicly available. Thus, making a valuable contribution to the deployment of ML-based approaches for anomaly detection and paving the way for future approaches.

Sakamoto et al. (2023) present an approach that also utilizes iForest, but unlike previous methods, does not primarily consider torque-angle angle pairs, but detects faulty screw runs based on data from the AC servo system during tightening. In the scenario, the errors are detected with almost complete accuracy and false positive and false negative rates are reported as mostly zero. Their work demonstrates that torque-angle values are not the only basis to successfully detect anomalies in screw driving operations.

West et al. (2023) implement an unsupervised approach that relies on the *K-Means* clustering method and uses *Dynamic Time Warping* (DTW) as a similarity measure to compare different screw runs. As a result, the faulty screw runs stand out as separate cluster, clearly distinguishable by their respective error shape. The approach succeeds in predicting the class of screw runs with up to 89% accuracy, not requiring prior knowledge about class distributions.

The review shows the continued relevance and steady success in the application of ML methods for anomaly detection in tightening data. High accuracy rates in anomaly detection have been achieved using a range of methods, both with unsupervised and supervised approaches. It should be noted, that the reported metrics of the results are highly dependent on the respective use case and were provided in this review primarily to quantify the achievement of the approaches for anomaly detection. Additionally, we want to stress that this is a selection of the most recent work. In the past, work such as *Matsuno et al.* (2013) or *Diez-Olivan et al. (2017)*, presented promising work for anomaly detection in screw driving as well.

4. Methods

This paper explores four supervised and four unsupervised anomaly detection methods, spanning a diverse set of approaches, such as density-based, tree-based, encoder or Deep Learning. We selected the respective models based on the related contributions, summarized in **Table 1**. Our goal is not to develop a novel method but to explore the applicability of established models in detecting anomalies within screw driving data using the data descripted in **Sec. 5**. At the same time, it is not our goal to provide an in-depth look at how the selected models function, so as not to exceed the intended scope of this paper. For this, we refer to the related works, which provide an in-depth description of the models.

Next, we introduce the four unsupervised methods, later deployed in our case study in **Sec. 6.**

- **Autoencoder**. Designed for unsupervised ML, AE are a powerful tool for anomaly detection in screw driving torque and angle data. It achieves this by encoding high-dimensional input data into a lower-dimensional representation, before reconstructing it (Zhou & Paffenroth, 2017). Discrepancies between the original and reconstructed data, which likely correlate with unusual screw driving runs that deviate from standard torque-angle patterns, represent potential anomalies indicative of anomal instances.
- **DBSCAN**. Utilizing the DBSCAN algorithm for unsupervised anomaly detection within screw driving data enables efficient distinction between dense, typical data clusters and sparse, irregular ones (Ester et al., 1996). This algorithm spatially characterizes torque-angle data points, identifying those anomalous screw runs that do not belong to denser regions, generally indicative of anomalies.
- **Isolation Forest**. iForest excels at detecting anomalies in torque and angle data, operating without any prior knowledge of good or bad screw runs. It isolates anomalies based on their rarity and uniqueness, quickly identifying and isolating screw driving runs that exhibit anomalous torque or angle readings (Liu et al., 2008).
- **Local Outlier Factor**. Utilizing the LOF algorithm for unsupervised anomaly detection within screw driving data enables the detection of anomaly instances, identified as those deviating significantly from the density of their neighboring runs (Breunig et al., 2000). By designating lower density instances as outliers, the LOF method effectively differentiates between standard and unusual screw driving data, facilitating the detection of potential process anomalies.

While unsupervised methods have the inherent advantage of not requiring labels from screw driving data, we also leverage four supervised learning techniques as benchmarks. Their application aims to demonstrate the capabilities of supervised approaches.

- **Random Forest.** For supervised time series clustering with screw driving data, a Random Forest classifier, an ensemble learning method utilizing multiple decision trees, provides a robust tool that curtails the risk of overfitting prevalent in individual trees (Breiman, 2001). Its integration of randomness in feature selection contributes to a sturdy and versatile method for identifying complex, nonlinear relationships typical in time series data, also found in torque and angle data.
- **Long Short-Term Memory.** As a type of RNN, LSTMs are particularly suited for analyzing sequential data like screw driving torque and angle measurements over time. Their unique memory cells can capture long-term dependencies, enabling the LSTM to learn from extended sequences commonly observed in time series data (Hochreiter & Schmidhuber, 1997). This feature allows the LSTM to detect patterns over time, aiding the accurate clustering of high-dimensional data.
- **Convolutional Neural Network.** Within the scope of time series clustering, a CNN can effectively extract relevant features from screw driving torque and angle measurements. The convolutional layers in the network perform a series of local operations across the time sequence, identifying localized temporal patterns within the data, providing a powerful and flexible tool for classifying sequential observations.
- **Encoder**. An Encoder classifier can translate high-dimensional time series data, such as screw driving torque and angle measurements, into a lower-dimensional, more digestible representation. This transformation of complex sequential data reveals underlying structures and dependencies, potentially improving the efficiency of supervised clustering tasks within the high-dimensional time series domain.

The next section will provide detailed information on the use case and the dataset utilized in our study.

5. Use case

The data set this study is based on was generated using an automatic screwing station. The station was used in serial production to connect two housing halves of a motor control unit of an electronic vehicle from the consumer sector. Since its end of serial production, the station serves as demonstrator for different research applications. On the station, one spindle performs two identical tightening operations for each component. *Delta PT 40x12* screws are used, specifically designed for superior performance in thermoplastics. Accordingly, the housing halves are also made of a thermoplastic material and the targeted tightening torque is comparatively low at 1.4 Nm.

The data set contains measurements of 5,000 runs. Of these, there are 4,089 normal and 911 anormal runs. We generated the data with the assembly of 100 work pieces of the same type, each consisting of an upper and lower part. Each work piece was tightened 25 times, resulting in multiple cuttings of the thread in the same work piece. With two connections per component, this results in the total number of 5,000 screw runs. Unlike some related work (e.g. Cheng et al., 2019; Leporowski et al., 2022; Sakamoto et al., 2023), no other measures were taken to artificially generate defective components. Only material wear due to the repeated threading of the screws led to faulty observations. To illustrate this effect, **Figure 1** shows an example of 25 screw runs of the same work piece, with the color gradient showing the number of the respective tightening cycles. The example shows that the torque that has to be applied in the tightening phase decreases with the increasing number of tightening cycles, and at the same time the maximum achievable angle of rotation appears to be slightly decreasing.

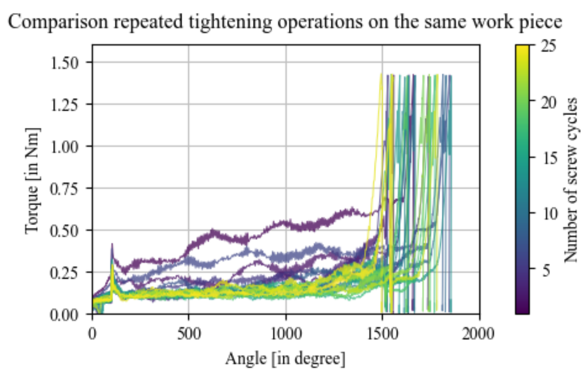

Figure 1: Exemplary representation of 25 screw driving operations for the same work piece, colored according to the respective number of subsequent operations

We provide binary class labels, *okay* ("OK") and *not okay* ("NOK"), for every screw run, recorded by the process monitoring of the station's control unit. For OK, a connection must achieve a maximum torque of between 1.2 and 1.6 Nm, where 1.4 Nm was the experimentally determined target torque in serial production. In addition, the tightening phases of the screw runs has to meet the torsion angle-based specifications of the screw station's program, else a NOK is assigned. **Figure** 2 shows the effect of the cycle number to the relative distribution of the labels.

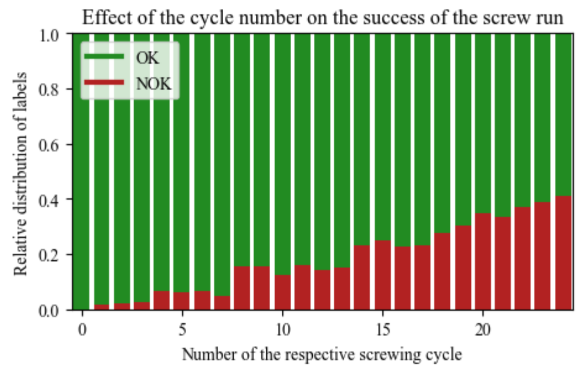

Figure 2: Visualization of the relation between the cycle number of the screw run and the label of the observation

The figure shows that 1.5% of defects occur from the second screw connection onward, with the relative ratio of OK to NOK rising almost constantly with the increasing number of the respective screwing cycle. As expected, the highest percentage of anomalies occurs for the last screw run, i.e. the twenty-fifth, with 41% of all recorded screw runs. In the analysis that follows in **Sec. 6**, we treat the 5,000 observations as individual screw runs and do not include the number of the screwing cycle while training the models.

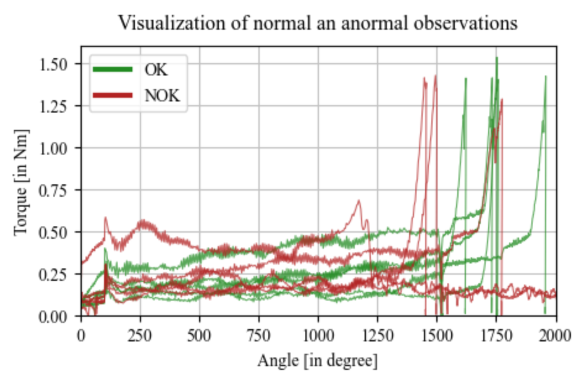

Figure 3: Exemplary display of five normal and five anormal screw runs

Figure 3 shows ten randomly selected screw runs *(five OK and five NOK)* to provide an insight into the structure of the normal and abnormal nature of the time series. The OK observations are shown in green and the NOK observations in red. The illustration shows the manifold forms of the different time series.

In **Table 2,** we provide a summary of the data that is available in the project repository (i.e. West, 2023). For each screw run i, the DMC code of the work piece *("id code")*, a timestamp at the start of the process *("date")* and the label of the screw run *("result")*, i.e. OK or NOK, are recorded. In addition, four tightening steps were performed for each run, which have individual names *("name")*, a continuous angular velocity *("speed")* and an indication of the success of the respective step *("result")*. Several hundreds of values k are recorded as a time series for each step. The current angle in degree *("angle vales")*, the torque in Nm *("torque values")* as well as the gradient *("gradient values")* and the time in seconds *("time values")* are determined. This is a selection of recorded and stored values. The total number of recorded values is several hundred and contains more information on the screw program. Each run is saved as a JSON file and can be viewed in the project repository (see *data/*).

For an in-depth look at the model's parameterization, we again refer to the full published code for this analysis (West, 2023, see *models/*).

To summarize, this use case is an application with real-world tightening data, in which no artificial errors were induced. To the best of our knowledge, this is only the third open-source dataset for anomaly detection in tightening data, besides the two datasets AURSAD and TMLSD discussed in **Sec. 3**. Other special features of this study are the application in the screw driving for a plastics work piece as well as the investigation the effects of re-tightening work pieces.

Table 2: Description of some selected variables from the recorded screw driving data (available in the project repo)

Variable		Description
One value for each recorded screw run (i)		
id code		Unique id of the work piece
date		Timestamp of the screw run
result		Binary label of the screw run
Four values per screw run (one for each screw step):		
name		Defined name of the tightening step
speed		Target speed [in degree per minute]
result		Binary label of the screw step
Hundreds of values per individual screw step (k):		
angle	$\theta_{i,k}$	Value for the angle [in degrees]
torque	$\tau_{i,k}$	Value for the torque [in Nm]
gradient	$g_{i,k}$	Value for the gradient
time	$t_{i,k}$	Value for the time [in seconds]

6. Results

6.1. Modeling and evaluation

In this section, we apply the models introduced in **Sec. 4** to check whether the classes of observations are recognizable with ML methods. For each model, we perform a *tenfold cross-validation* of the screw driving data to obtain comparable results. Ten-fold cross-validation is a technique where a dataset is divided into ten subsets. A model is trained ten times, each time using nine subsets for training and the remaining subset for testing, which helps in obtaining a more robust estimate of the model's performance by averaging the results from each of the ten iterations. In **Sec. 6**, we therefore additionally provide the variance of the ten determined modeling metrics.

For comparison, we calculate the *Accuracy*, *Precision*, *Recall* and the *Macro F1 Score*. Accuracy is the ratio of correctly predicted labels to the total number of runs, providing a simple, straightforward measure of a model's performance. Conversely, the Macro F1 Score is the harmonic mean of *Precision* and *Recall*, computed independently for each class and then averaged. This metric is particularly insightful in scenarios with unbalanced classes, as it gives equal weight to each class performance, regardless of its size. In the use case with about 81.78% OK and 18.22% NOK observations, a decent ML model has to achieve an Accuracy of more than 81.78% and a Macro F1 Score of more than 50.00%.

Since not all models can handle time series of different lengths, we had to limit the length of all series to 750 after a preliminary study. Shorter series were extended accordingly using *zero padding*. While the majority of screw runs had to be padded, 140 series were *shortened* due to their length exceeding 750. In addition, we limited the analysis to the torque values, since the angle was specified as a constant by the process control for each screw driving step.

6.2. Unsupervised models

We evaluate four unsupervised models, using the screw driving data from our use case introduced in **Sec. 5**, and summarize their averaged performance across a ten-fold cross-validation in **Table 3**. Again, both the models (see *models/*) and results (see *results/*) are in the provided project repository (West, 2023).

The **Autoencoder** delivered an average accuracy of 89.30% ± 0.05%, exceeding the discussed baseline accuracy of 81.78%. In addition, its Macro average F1 Score of 83.76% ± 0.07% indicates an adequate performance in discerning both OK and NOK classes.

Table 3: Averaged results of the anomaly detection with regard to the type of machine learning models

Type	Model	Accuracy	Precision	Recall	Macro avg. F1
Unsupervised	Autoencoder	89.30% ± 0.05%	66.71% ± 0.24%	84.00% ± 0.14%	83.76% ± 0.07%
	DBSCAN	96.68% ± 0.08%	75.52% ± 0.73%	99.26% ± 0.04%	90.70% ± 0.12%
	Isolation Forest	69.78% ± 0.02%	12.50% ± 0.06%	11.00% ± 0.04%	46.73% ± 0.01%
	LOF	73.82% ± 0.04%	30.10% ± 0.27%	33.05% ± 0.27%	57.64% ± 0.09%
Supervised	CNN	97.52% ± 0.01%	92.57% ± 0.15%	94.31% ± 0.11%	95.92% ± 0.02%
	Encoder	98.98% ± 0.00%	98.78% ± 0.01%	95.65% ± 0.05%	98.27% ± 0.00%
	LSTM	87.96% ± 6.99%	36.72% ± 22.58%	38.16% ± 24.30%	63.34% ± 6.96%
	Rand. Forest	99,02% ± 0.00%	99.04% ± 0.01%	95.67% ± 0.05%	98.36% ± 0.00%

The **DBSCAN** model surpassed the Autoencoder, demonstrating an accuracy of 96.68% ± 0.08% and a Macro average F1 score of 90.70% ± 0.12%. Furthermore, the high values of Precision with 75.52% ± 0.73% and Recall with 99.26% ± 0.04 clearly show the superior performance of DBSCAN compared to the three other unsupervised methods. However, the slightly lower Precision implies that the model achieves a higher rate of False Positives, whereas the rate of False Negatives is very low, expressed by the great Recall. The spatial distribution of torque-angle data points appears to provide meaningful information for the differentiation between regular and anomalous screw runs, as evidenced by the DBSCAN results.

In contrast, the **Isolation Forest** and **Local Outlier Factor** models were less successful. The iForest attained an accuracy of merely 69.78% ± 0.02%, and its Macro average F1 score of 46.73% ± 0.01% indicates potential for improvement. This result suggests that this model's foundational approach of isolating anomalies based on their distinctiveness and rarity may not be entirely suitable for this particular application. Similarly, the LOF model, which assigns lower density instances as outliers, achieved an accuracy of 73.82% ± 0.04%. The Macro's average F1 score of 57.64% ± 0.09% was slightly better than iForest's, but still significantly below average compared to Autoencoder and DBSCAN.

6.2. Supervised models

Similarly, we evaluated the four supervised models from **Sec. 3** to provide a comparison, with their averaged performance across ten-fold cross-validation also presented in **Table 3** in the previous section.

The **Convolutional Neural Network** performs well, delivering an accuracy of 97.52% ± 0.01% and a Macro average F1 score of 95.92% ± 0.02%. This result indicates that the CNN's ability to extract localized temporal patterns from screw driving torque and angle measurements is highly effective.

The **Encoder** model reached the highest accuracy among the considered models, with an Accuracy score of 98.98% ± 0.00%, and a Macro average F1 score of 98.27% ± 0.00%. With 98.78% ± 0.01% and 95.65% ± 0.05% respectively, Precision and Accuracy again clearly show the success of the model predictions. We emphasize the high ratio of observations correctly identified as erroneous, expressed by the Precision. This superior performance underscores the Encoder's capability to transform high-dimensional time series data into lower-dimensional representations, which is particularly useful in detecting anomalies in screw driving data, as evident in this use case's results.

Among all supervised models, the **LSTM** classifier presented the widest range of performances, averaging an accuracy of 87.96% ± 6.99% and a Macro average F1 score of 63.34% ± 6.96%. This variability mainly stems from an uneven performance across cross-validation folds. The variations are also reflected in the high variance of the values of Precision and Accuracy, which are at 36.72% ± 22.58% and at 38.16% ± 24.30%, respectively. Notably, in four folds, the model displayed high efficacy, with accuracies and Macro F1 scores exceeding 95% and 92%, respectively, due to the accurate classification of True Positives and True Negatives. Conversely, in the six remaining folds, the model's inability to detect any True Positives critically affected its Precision, Recall, and Marco F1 Score, hence reducing overall Accuracy. These findings underscore the need to consider data characteristics and distribution when deploying ML models.

The **Random Forest** model achieved an excellent accuracy of 99.02% ± 0.00%, surpassing the Encoder, and a slightly superior Macro average F1 of 98.36% ± 0.00%. A consideration of Precision and Recall, which are 99.04% ± 0.01% and 95.67% ± 0.05% respectively, shows that the Random Forest is superior to DBSCAN in this respect as well. The significantly higher precision shows the ability to avoid False Positives. This robust performance demonstrates the effectiveness of this model in identifying complex, nonlinear relationships in the time series data. The use of multiple decision trees in a Random Forest model may have contributed to this superior performance.

7. Conclusion

In this paper, we undertook the task of exploring anomaly detection within industrial manufacturing processes, specifically focusing on screw driving data. Our aim was to explore the potential of unsupervised (ML) techniques in identifying anomalous patterns within time series data, a data-driven approach that enhances traditional quality assurance measures.

We introduced a real-world dataset of screw driving operations, enriching the resources available for studying these processes and enabling exploration of the utility and limitations of various ML models. Our study demonstrated the efficacy of CNN, AE, and Random Forest, with Random Forest achieving the highest accuracy and Macro average F1 scores. Nevertheless, the LSTM model displayed high variability in performance, underlying the importance of data distribution and model parameter tuning.

We have shown that unsupervised ML techniques can efficiently detect both known and unexplored anomalies in screw driving data, contributing to improving manufacturing process quality and efficiency. The implications of these findings extend beyond the scope of quality management: Enhancing the reliability of screw runs can lower production costs, improve customer satisfaction, and contribute to more sustainable manufacturing practices.

Since the scope of this work did not allow us to examine the eight modeling processes in depth, we intend to do so for selected models in future work. In addition, we consider publishing a complementary publication for the data set (akin to Leporowski et al., 2021), in which we provide detailed explanations of the data collection set-up as well as the nature of the collected data. We are also considering providing a simpler extract-transform loading process for the dataset, instead of simply publishing the raw data, to facilitate future use in subsequent scientific work.

Summarizing, despite the promising results, our research highlighted that the practical implementation of ML models in manufacturing anomaly detection is not without challenges. For example, the LSTM results demonstrated the role that data characteristics and distribution may play in modeling, underscoring the need for careful data preprocessing, parameter tuning during practical applications as well as the need for a holistic approach to anomaly detection.

Acknowledgement

The work on this paper was supported by the *German Ministry of Education and Research* (BMBF) as part of program "Strengthening the data skills of the next generation of scientists" as well as by the program "NextGenerationEU" of the *European Union*.

References

Aronson, R. M., Bhatia, Jia, Z., & Guillame-Bert, M. (2017). Data-driven classification of screwdriving operations. In D. Kulić, Y. Nakamura, O. Khatib, & G. Venture (Eds.), 2016 International Symposium on Experimental Robotics (pp. 244–253). Springer International Publishing. https://doi.org/10.1007/978-3-319-50115-4_22

Bickford, J. (1995). An introduction to the design and behavior of bolted joints: Revised and expanded (3rd ed.). Routledge. https://doi.org/10.1201/9781315137353

Bickford, J. H., & Nassar, S. (1998). Handbook of Bolts and Bolted Joints. Marcel Dekker, Inc.

Breiman, L. (2001). Random Forest. Machine Learning, 45(1), 5–32. https://doi.org/10.1023/A:1010933404324

Breunig, M. M., Kriegel, H.-P., Ng, R. T., & Sander, J. (2000). LOF: Identifying density-based local outliers. Proceedings of the International Conference on Management of Data (ACM SIGMOD), 29(2), 93–104. https://doi.org/10.1145/335191.335388

Cao, X., Liu, J., Meng, F., Yan, B., Zheng, H., & Su, H. (2019). Anomaly detection for screw tightening timing data with LSTM recurrent neural network. Proceedings of the IEEE International Conference on Mobile Ad-Hoc and Sensor Networks (MSN), 15(1), 348–352. https://doi.org/10.1109/MSN48538.2019.00072

Chandola, V., Banerjee, A., & Kumar, V. (2009). Anomaly detection: A survey. ACM Computing Surveys, 41(3), 1–58. https://doi.org/10.1145/1541880.1541882

Chawla, N. V., Bowyer, K. W., Hall, L. O., & Kegelmeyer, W. P. (2002). SMOTE: Synthetic minority over-sampling technique. Journal of Artificial Intelligence Research, 16(1), 321–357. https://doi.org/10.1613/jair.953

Cheng, X., Jia, Z., & Mason, M. T. (2019). Data-efficient process monitoring and failure detection for robust robotic screwdriving. Proceedings of the IEEE International Conference on Automation Science and Engineering, 15(1), 1705–1711. https://doi.org/10.1109/COASE.2019.8842854

Diez-Olivan, A., Penalva, M., Veiga, F., Deitert, L., Sanz, R., & Sierra, B. (2017). Kernel density-based pattern classification in blind fasteners installation. In F. J. Martínez de Pisón, R. Urraca, H. Quintián, & E. Corchado (Eds.), Hybrid Artificial Intelligent Systems (Vol. 10334, pp. 195–206). Springer Int. Publishing. https://doi.org/10.1007/978-3-319-59650-1_17

Dorj, E., & Altangerel, E. (2013). Anomaly detection approach using Hidden Markov Model. Proceedings of the IEEE International Forum on Strategic Technology

(IFOST), 8(1), 141–144. https://doi.org/10.1109/IFOST.2013.6616874

Ester, M., Kriegel, H.-P., Sander, J., & Xu, X. (1996). A density-based algorithm for discovering clusters in large spatial databases with noise. Proceedings of the International Conference on Knowledge Discovery and Data Mining (KDD), 2(1), 226–231.

Glaser, A. E., Harrison, J. P., & David, J. (2022). Anomaly detection methods to improve supply chain data quality and operations. SMU Data Science Review, 6(1), 1–24.

Hochreiter, S., & Schmidhuber, J. (1997). Long short-term memory. Neural Computation, 9(8), 1735–1780. https://doi.org/10.1162/neco.1997.9.8.1735

Jia, Z., Bhatia, A., Aronson, R. M., Bourne, D., & Mason, M. T. (2019). A survey of automated threaded fastening. IEEE Transactions on Automation Science and Engineering, 16(1), 298–310. https://doi.org/10.1109/TASE.2018.2835382

Leporowski, B., Tola, D., Hansen, C., & Iosifidis, A. (2021). AURSAD: Universal robot screwdriving anomaly detection dataset, 1–27.

Leporowski, B., Tola, D., Hansen, C., & Iosifidis, A. (2022). Detecting Faults During Automatic Screwdriving: A Dataset and Use Case of Anomaly Detection for Automatic Screwdriving. In A.-L. Andersen, R. Andersen, T. D. Brunoe, M. S. S. Larsen, K. Nielsen, A. Napoleone, & S. Kjeldgaard (Eds.), Lecture Notes in Mechanical Engineering Ser. Towards Sustainable Customization: Proceedings of the 8th Changeable, Agile, Reconfigurable and Virtual Production Conference (CARV2021) (pp. 224–232). Springer International Publishing AG. https://doi.org/10.1007/978-3-030-90700-6_25

Li, X., Wu, Y., & Jia, Q. (2020). Anomaly detection of bolt tightening process based on improved SMOTE. Proceedings of the International Conference on Robotics and Artificial Intelligence (ICRAI), 6(1), 9–14. https://doi.org/10.1145/3449301.3449304

Liu, F. T., Ting, K. M., & Zhou, Z.-H. (2008). Isolation Forest. Proceedings of the IEEE International Conference on Data Mining, 8(1), 413–422. https://doi.org/10.1109/ICDM.2008.17

Matsuno, T., Huang, J., & Fukuda, T. (2013). Fault detection algorithm for external thread fastening by robotic manipulator using linear support vector machine classifier. Proceedings of the IEEE International Conference on Robotics and Automation (ICRA), 30(1), 3443–3450. https://doi.org/10.1109/ICRA.2013.6631058

Pang, G., Shen, C., Cao, L., & van den Hengel, A. (2021). Deep learning for anomaly detection: A review. ACM Computing Surveys, 54(2), 1–38. https://doi.org/10.1145/3439950

Powell, T. C. (1995). Total quality management as competitive advantage: A review and empirical study. Strategic Management Journal, 16(1), 15–37. https://doi.org/10.1002/smj.4250160105

Ribeiro, D., Matos, L. M., Cortez, P., Moreira, G., & Pilastri, A. (2021). A comparison of anomaly detection methods for industrial screw tightening. In O. Gervasi, B. Murgante, S. Misra, C. Garau, I. Blečić, D. Taniar, B. O. Apduhan, A. M. A. Rocha, E. Tarantino, & C. M. Torre (Eds.), Computational Science and Its Applications (ICCSA 2021) (Vol. 12950, pp. 485–500). Springer International Publishing.

Ribeiro, D., Matos, L. M., Moreira, G., Pilastri, A., & Cortez, P. (2022). Isolation forests and deep autoencoders for industrial screw tightening anomaly detection. Computers, 11(4), 54. https://doi.org/10.3390/computers11040054

Sakamoto, Y., Nakamura, Y., & Sugioka, M. (2023). A case study of real-time screw tightening anomaly detection by machine learning using real-time processable features. Omron Technics, 55.001EN(1), 1–11.

Sanguanmak, Y., & Hanskunatai, A. (2016). DBSM: The combination of DBSCAN and SMOTE for imbalanced data classification. Proceedings of the IEEE International Joint Conference on Computer Science and Software Engineering (JCSSE), 13(1), 1–5. https://doi.org/10.1109/JCSSE.2016.7748928

Schlegl, T., Schlegl, S., Tomaselli, D., West, N., & Deuse, J. (2022). Adaptive similarity search for the retrieval of rare events from large time series databases. Advanced Engineering Informatics, 52(101629), 1–12. https://doi.org/10.1016/j.aei.2022.101629

Schlegl, T., Schlegl, S., West, N., & Deuse, J. (2021). Scalable anomaly detection in manufacturing systems using an interpretable deep learning approach. Proceedings of the CIRP Conference on Manufacturing Systems, 54(1), 1547-1552. https://doi.org/10.1016/j.procir.2021.11.261

Stojanovic, L., Dinic, M., Stojanovic, N., & Stojanovic, A. (2016). Big-data-driven anomaly detection in Industry 4.0: An approach and a case study. Proceedings of the IEEE International Conference on Big Data(1), 1647–1652. https://doi.org/10.1109/BigData.2016.7840777

West, N. (2023). Project repository: github.com/nikolaiwest/2023-anomaly-detection-hicss.git.

West, N., Schlegl, T., & Deuse, J. (2021). Feature extraction for time series classification using univariate descriptive statistics and dynamic time warping in a manufacturing environment. Proceedings of the IEEE International Conference on Big Data, Artificial Intelligence and Internet of Things Engineering, 2(1), 762–768. https://doi.org/10.1109/ICBAIE52039.2021.9389954

West, N., Schlegl, T., & Deuse, J. (2023). Unsupervised anomaly detection in unbalanced time series data from screw driving processes using k-means clustering: (Preprint: Accepted for publication). Proceedings of the CIRP Conference on Manufacturing Systems, 56(1), 1–6.

Zhou, C., & Paffenroth, R. C. (2017). Anomaly detection with robust deep autoencoders. Proceedings of the International Conference on Knowledge Discovery and Data Mining (KDD), 23(1), 665–674. https://doi.org/10.1145/3097983.3098052