

Active Learning Framework for Time-Series Classification of Vibration and Industrial Process Data

Sergio Martin-del-Campo^{1,2}, William Lindskog³, Davide Badalotti⁴, Oskar Liew¹, and Arash Toyser¹

¹ *Research and Development, Viking Analytics, Gothenburg, 41109, Sweden*
sergio.mdc@vikinganalytics.se
oskar.liew@vikinganalytics.se
arash.toyser@vikinganalytics.se

² *EISLAB, Luleå University of Technology, Luleå, 97187, Sweden*

³ *Center for Mathematical Sciences, Lund University, Lund, 22100, Sweden*
williamlindskog@protonmail.com

⁴ *Department of Informatics, System and Communication, University of Milano-Bicocca, Milano, 20126, Italy*
d.badalotti@campus.unimib.it

ABSTRACT

Recent technical developments have facilitated the collection and storage of large amounts of time series data for many condition monitoring and maintenance processes. However, most of this data is unlabeled, and producing high-quality labeled data is expensive, time-consuming, and a lot of times inaccurate given the uncertainty surrounding the labeling process and annotators. Active Learning (AL) has emerged as an approach that enables cost and time reductions of the labeling process. Here, we present an active learning framework to be used in the classification of time series from industrial process data, which can be vibration waveforms or control process data. Previous work has focused on active learning for image classification problems. Alternatively, when active learning has focused on time series classification problems, it has not dealt with the cold start problem, which consists of a complete absence of labels at the beginning of the training process. The active learning framework proposed incorporates a pre-clustering step to create an initial labeled dataset. Furthermore, we incorporate two strategies for the generation of features to be used in the AL framework, which are time series imaging and automatic feature generation. We study the learning curves of the different feature extraction techniques and evaluate them in two case studies. The first case is based on vibration data from a ball bearing experiment with faults seeded in the bearings. The second case is based on a produc-

tion dataset from an industrial control process. We discover that with only having to label up to 10% of the unlabeled instances, after they had been properly queried, it is possible to achieve accuracy over 90%. An active learning framework offers a real possibility to achieve high accuracy while reducing the amount of work that needs to be incorporated into the labeling process.

1. INTRODUCTION

The condition monitoring, maintenance, and health management fields make ample use of time series for diagnostics, monitoring, and prognostics tasks. These time series originate from electrical sensors that sample individual measurements (e.g. temperature or pressure) or waveforms, which are full-time series by themselves, such as vibration. The demand for raw time series waveforms is increasing given the interest to build data-based models around them. Unfortunately, most of this data, which is becoming widely available, comes without labels. Labels are an important component to build these data-based models and describe the conditions to be identified by such models. However, these labels are a scarce resource. This is especially true for machine fault detection issues where collecting large amounts of data is easier than obtaining the corresponding labels, especially when the faults evolve naturally over time (Zhang et al., 2020). Furthermore, producing high-quality labeled data is expensive, time-consuming, and a lot of times inaccurate given the uncertainty surrounding the labeling process and the annotators.

These conditions, where the amount of unlabeled data is rich

Sergio Martin-del-Campo et al. This is an open-access article distributed under the terms of the Creative Commons Attribution 3.0 United States License, which permits unrestricted use, distribution, and reproduction in any medium, provided the original author and source are credited.

but labeled data is poor, are suitable for a semi-supervised approach (Zhu, 2005). Active Learning is a machine learning approach where an external annotator is queried for labels to be given to the unlabeled data. The aim is to improve the model accuracy by asking for labels from the annotator until reaching a desired accuracy or model convergence. Active learning algorithms are iterative and on each iteration, the annotator, who is also referred to as the *oracle*, must label a number of samples. The benefit of pursuing an active learning approach is that machine learning models are able to achieve larger accuracy while having to use a smaller training size (Settles, 2011). These approaches had produced significant results in the fields of medical natural language processing (Rosales et al., 2007), image classification for remote sensing (Haut et al., 2018) and object detection for autonomous driving (Hausmann et al., 2020).

The use of active learning for time-series problems is still limited given the high dimension of the data and concerns on how to present such time-series to the annotator. This limitation is further constrained by the limited availability of fault data, a common occurrence on condition monitoring or maintenance applications, given that most of the machines operate in healthy operating conditions. Furthermore, there are multiple challenges that need to be considered when using an active learning framework in practical scenarios (Settles, 2009). One of these challenges is addressed by Shekhar et al. (2021), who considers the scenario where the learner abstains from providing a label under certain scenarios at a penalty cost. Agarwal et al. (2021) considers the challenge where there is uncertainty around the label provided by the oracle. An active learning challenge, which is highly relevant to time-series problems is how to evaluate the informativeness metrics of the time-series instances. Peng et al. (2017) addresses this challenge by proposing a set of informativeness metrics to be used in multi-class time-series classification problems. Finally, an active learning challenge that is significant to the maintenance and condition monitoring fields is the so-called *cold-start* problem. Active learning needs an initial labeled set of samples before asking the oracle to label additional samples. The cold-start problem is the scenario where there are no labels at all when starting to train the model. Lughofer (2012) proposes a solution to this problem, which is applicable to image classification problems.

In this paper, we present an active learning framework for time-series classification of vibration and industrial-processes signals. The framework incorporates a pre-clustering step that considers the scenario where no labels are available at the beginning of the training stage. Our interest here focuses on feature extraction alternatives that can facilitate the presentation of time-series signals to the annotator for labeling. We aim to prepare the unlabeled time series through an automated data pre-processing step that enables subject-matter experts to incorporate their expertise without knowl-

edge of particular machine learning tools. We consider two alternatives in this pre-processing step: time series imaging and automated feature engineering. Z. Wang & Oates (2015) present multiple techniques for the encoding of time series as images. We are interested in the use of images because it enables us to emphasize or capture local patterns that can be spread over time, which might assist the annotator on his work (Rodriguez-Garcia et al., 2021). Automated feature engineering enables to automatically calculate a large number of time series characteristics that the annotator can use as a starting point. The libraries considered for automatic feature generation are Time Series Feature Extraction Library (TSFEL) (Barandas et al., 2020) and Time Series Feature Extraction based on Scalable Hypothesis tests (TSFRESH) (Christ et al., 2018). Using time series imaging and automated feature engineering in tandem with the active learning framework can enable us to skip the feature selection step given that we can use the resulting images and features directly for query and classification. These features are used in the active learning framework which intends to find which instances should be labeled next based on the selection of the instance that is most useful. Thus, we adopt the uncertainty sampling active learning strategy, which we use together with the metrics proposed by Peng et al. (2017). This entire framework is evaluated on vibration data from a rolling element bearing experiment and time series data originated from a production facility. The work presented here is novel because it describes an end-to-end active learning framework focused on the typical data conditions of the condition monitoring and maintenance fields where no labeled data is more common. It further incorporates an automated pre-processing stage that converts the time series signals into images or a set of characteristics that can facilitate the work of annotators. We observe that a small number of annotations is required before reaching accuracy over 90%. These results indicate that our proposed active learning framework is useful for the classification of time series signals.

2. ACTIVE LEARNING FRAMEWORK

Active learning is an area of machine learning where the algorithm is allowed to choose the data from which it learns. This is a desirable property given that typical supervised learning approaches require a large number of labeled instances. However, these labeled instances are not always available, and generating them is difficult or expensive. On active learning methods an annotator, also known as *oracle* provides labels to a series of queries. The aim is to achieve greater accuracy with as few labels as possible Settles (2009).

We present our framework by describing the techniques used in the feature extraction stage, followed by the pre-clustering step that addresses the cold-start problem, and a description of the active learning strategy used in our work. Figure 1 shows a diagram of the framework, which presents the stages

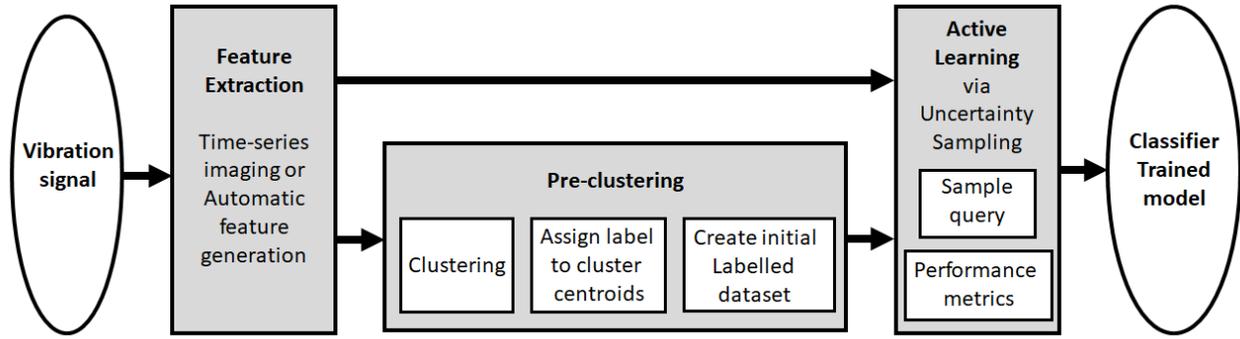


Figure 1. Diagram of our proposed Active Learning framework.

that form it and how each stage is related. The inputs are the raw time series signals, which can be a vibration signal or any other time series signal. Each block describes how their output contributes to the following block. The output is a trained classification model that is capable to handle to provide a label to new unseen data.

2.1. Feature extraction

Recent advancements within the Internet-of-Things have enabled to collect large amounts of data in the condition monitoring and maintenance fields. Typically, the collected data is time series data and requires specialists working to identify suitable features to be used by the machine learning models. The feature engineering work required to identify these features, which can vary from application to application, is extensive, particularly for supervised models and it remains as a research opportunity for unsupervised models. In our active learning framework, we consider two alternatives for the pre-processing of the data. These alternatives are time series imaging and automated feature engineering.

In a time series imaging approach, the time series data inputs are converted into images, which may be easier to label by an oracle given that it can present a larger amount of information in a compressed manner. The time series data can be a vibration waveform or any sensor data stream, and the resulting image is presented to the oracle and can be used by a classifier. The imaging techniques consider in this work are Z. Wang & Oates (2015):

- Gramian Angular Fields - time series are rescaled and represented in polar coordinates by encoding the magnitude as the angular cosine and the timestamp as the radius. Then, by exploiting this transformation we can use the trigonometric sum/difference between points to establish the temporal correlation resulting in the Gramian Angular Summation Field (GASF) and the Gramian Angular Difference Field (GADF).
- Markov Transition Fields (MTF) - time series are encoded as a Markov Chain transition matrix, where the

magnitudes are discretized in a fixed number of bins of uniform width. Temporal information is encoded transition probabilities between points at specific intervals.

- Recurrence Plots (RP) - time series are encoded as a matrix of recurrences, where the time series is split in a number of sub-sequences and it is kept track of the number of times points returns to previously visited states.

We define the initial formulations of these algorithms as the original formulation. However, these original formulations can be invariant to some signal transformations, particularly in large datasets. Thus, we also adopt the imaging encoding formulations proposed by Rodriguez-Garcia et al. (2021), which we refer to as the modified formulation. Under this modified formulation, GASF/GADF uses a dataset-wide rescaling step instead of instance-wide rescaling. Meanwhile, in MTF, the width of the bins is proportional to the quantiles of distribution centered in the mean values instead of uniform width. Likewise, the mean value is considered in the estimation of the transitions for the RP encoding, which is not considered under the original formulation.

In the automated feature engineering approach, we use the automatic feature extraction frameworks: Time Series Feature Extraction based on Scalable Hypothesis tests (TSFRESH) (Christ et al., 2018) and Time Series Feature Extraction Library (TSFEL) (Barandas et al., 2020). TSFRESH can produce up to 111 statistical features, while TSFEL can produce up to 60 statistical features. Some of these features can result in NaNs or infinite values, which are replaced by the average and extreme values respectively. Also, if all the values for a calculated feature result in infinite, the entire feature is set to zero. In addition, we implement the pattern discovery strategy described by Peng et al. (2017) on their active learning framework for time series classification, known as ACTS. We use this implementation as a benchmark to our proposed framework. ACTS is based on shapelet discovery, which identifies discriminative patterns in the time series. The splits generated during pattern discovery are used to estimate the data entropy of the pattern, which in turn is used

to identify the optimal split and pattern. The selected patterns together with the time series and labels are used to construct a probabilistic model where the Euclidean norm is used to measure the distance between the pattern and the time series. Further details of the ACTS strategy are available in Peng et al. (2017).

2.2. Cold-start problem

Active learning algorithms have multiple label gathering scenarios that define the query process. The most common scenario, which considers the real-world problem of large amounts of unlabeled data, is known as pool-based active learning Settles (2009). Under this scenario, a small set of labeled data is required before starting to select new samples to label from the pool of unlabeled data. The cold-start problem originates from this scenario where there are no labels in the beginning to train the initial classification model. Therefore, we solve this problem by introducing a pre-clustering step. During pre-clustering, we apply an unsupervised clustering algorithm to the unlabeled data and select the points closest to the centroids of each cluster as the initial instances of the labeled dataset. This procedure is based on the work by Souza et al. (2017). The steps in this procedure consist of:

1. Define unsupervised clustering algorithm, such as k-means, and apply it to the unlabeled data.
2. Select the number of points closest to the centroids of each cluster.
3. Assign a label to the selected points.
4. Labeled points are the instances that form the initial labeled dataset.

The number of clusters to be identified by the clustering algorithm is defined by the number of categories upon which the data is expected to be classified. Alternatively, the elbow method (Marutho et al., 2018) can be used to determine the optimal number of clusters when this information is not known. In addition, the number of points to be selected out of each cluster is defined by a pre-defined threshold, which can be in the order of 1-2%, that describes the size of the initial dataset with respect to the size of the unlabeled dataset.

2.3. Active learning

An active learning framework enables data-based models to perform better with less training given that they can choose the data from which it can learn. The basic procedure of the active learning framework consists of the following steps:

- Step 1** Algorithm begins with a small labeled training dataset and a large set of unlabeled data.
- Step 2** A classifier algorithm is trained using the labeled dataset.
- Step 3** An active learning technique queries for a data point from the unlabeled dataset.

Step 4 The oracle gives a label to the previously selected data point and the labeled and unlabeled datasets are updated.

Step 5 Step 2 to Step 4 are repeated iteratively until reaching a stop condition.

The unlabeled dataset is the vibration or time series signals of the condition monitoring or maintenance problem at hand. The small labeled training dataset mentioned in Step 1 is the product of the pre-clustering step of our framework described in Section 2.2. The classifier algorithm in Step 2 that is used as part of this proposed framework is the Support Vector Machine algorithm (SVM). SVM is a non-probabilistic binary linear classifier that identifies a hyperplane separating the two categories and intends to maximize the width of the gap between these two categories (L. Wang, 2005). In the case of multi-class classification problems, the one-versus-all method is adopted.

Sample query is Step 3 in the active learning framework. In this step, the data sample from the unlabeled dataset to be presented to the oracle for labeling is selected. The selection of the data sample requires evaluating the usefulness of the unlabeled dataset to select the sample that might contribute the most to the improvement of the model. There are multiple query strategies and Kumar & Gupta (2020) presents a review of all of them. The query strategy selected in this work is uncertainty sampling given its simplicity and straightforward understanding. Under this strategy, the oracle queries the instances upon which it is least certain on how to label. In particular, the framework utilizes the method of classification uncertainty. Under this method, the uncertainty for each data point x is defined as

$$U(x) = 1 - P(\hat{x}|x), \quad (1)$$

where \hat{x} is the most likely prediction for the specified instance. This can be interpreted as the classification uncertainty been the uncertainty of \hat{x} having the prediction given to data point x .

Figure 2 presents an example of the labeling interface used in Step 4, where the oracle is queried for the label of an unlabeled time series. In this example, a raw vibration signal from a rolling element bearing is shown. Alternatively, the interface can show the spectra of the raw signal or the converted time series image. The annotator needs to provide a label on the location of the fault via pressing the buttons to the right of the interface. Please notice that in this example, just a small subset of raw time series values are shown. The active learning procedure ends at Step 5, once the model has reached a stop condition. This stop condition can be defined by reaching a desired accuracy level in the classifier model or having carried out a pre-defined number of queries.

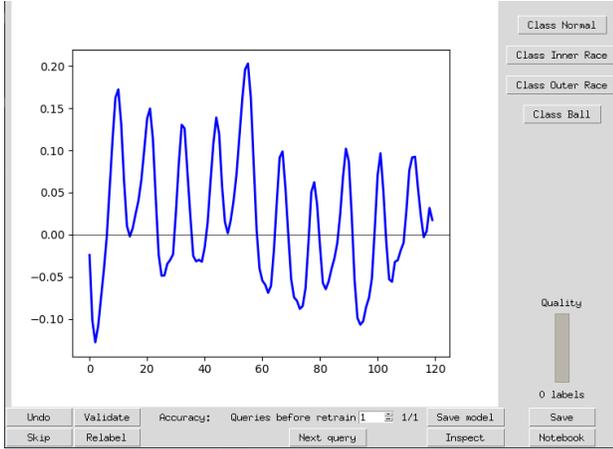


Figure 2. Example of the active learning user interface employed to query the oracle. The interface shows an example of the ball bearing dataset along with the four label alternatives.

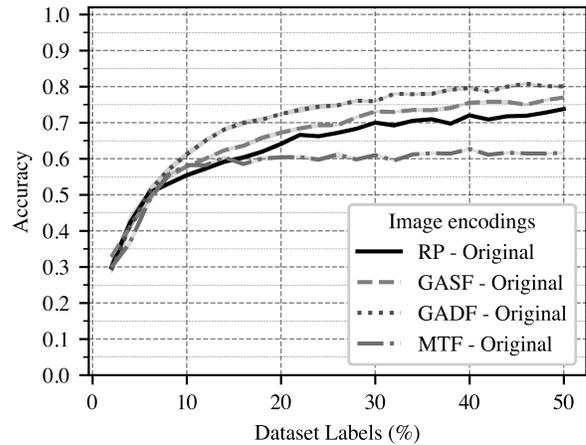
3. CASE STUDIES

We are interested in the performance of the time series classification model with respect to the number of queries carried out by the active learning framework. We consider two case studies to evaluate this performance. We analyze vibration signals from the bearing data center at Case Western Reserve University (Loparo, 2003) and time series data from a control problem belonging to one of our customers, henceforth as the production dataset.

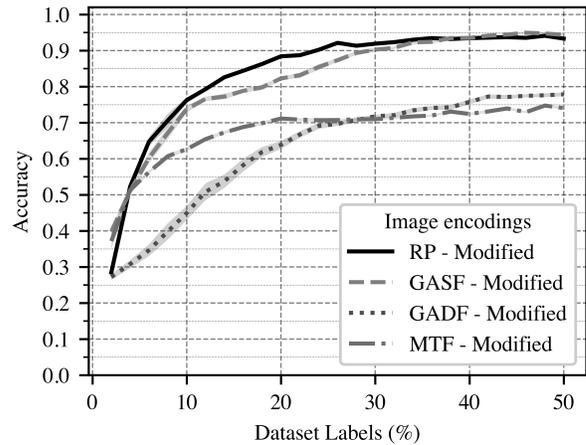
The vibration signals of the ball bearing dataset are generated by a rotating machine, which consists of an electric motor, a torque transducer, a dynamometer, and a ball bearing supporting the motor shaft. Vibration data is recorded via an accelerometer located at the drive end of the motor with a sampling rate of 12 kHz. The machine operates with a varying load between 0 HP and 3 HP, resulting in varying motor speed between 1800 and 1730 rpm. Faults are manually introduced at the inner raceway, outer raceway, and balls. Thus, there are four possible labels that the oracle can assign to each instance corresponding to the three fault location cases and a case without no faults. Meanwhile, the production dataset originates from data provided by one of the customers of Viking Analytics. The time series data belongs to a control system, where samples were taken with a sampling rate of 25 Hz. This is a binary classification problem where the annotator has to classify the presence of a bump in the data.

The initialization of the active learning algorithm requires an initial labeled dataset, which typically would be the output of the pre-clustering step. The labels are assigned to each cluster by the oracle after the total number of clusters has been defined. In these case studies, ten instances form the initial labeled dataset where we ensure the presence of at least one instance from each of the categories for both datasets. This small dataset of ten labeled instances is known as the seed,

while all other instances remain unlabeled. In the case of features originated from the imaging techniques, the images are flattened into a single vector for their use in the classifier algorithm. The selected image resolution is 64×64 pixels and each pixel represents a single feature. In the case of the automated feature engineering framework, each framework produces a set of features. As described in Section 2.3, we utilize SVM as the classifier algorithm with four categories for the bearing dataset and two categories for the production dataset. The stop condition for the active learning algorithm will be a specific number of queries to evaluate the performance at that point.



(a) Original



(b) Modified

Figure 3. Comparison of learning curves for imaging techniques in the ball bearing dataset. Gray areas represent a confidence interval of one standard deviation.

3.1. Ball bearing dataset

The vibration signals of the ball bearing dataset are long sequence time series sequences recorded for each load and fault

condition. The normal case vibration signals are approximately 480k samples long and the vibration signals for induced damages on the bearing are approximately 120k samples long. Therefore, these long sequences were split into minibatches of 120 samples long for experimental purposes. Each minibatch represents a time series instance and these instances are the ones used in the active learning framework. In practical terms, the size of the batch can be large enough to contain enough health state information of the bearing.

Active learning frameworks are normally evaluated via learning curves. A learning curve evaluates a metric of interest, such as accuracy, as a function of the number of new instances queries (Settles, 2009). Thus, accuracy represents the fraction of accurate predictions of a trained model that used the given number of instances queries for training. Figure 3 shows the learning curves for the scenario where an imaging technique was used to generate the features. In this scenario, the size of the unlabeled set is 2000 instances with equal parts of each label case. Figure 3a shows the original imaging implementation, while Figure 3b the imaging implementation with the modifications proposed by Rodriguez-Garcia et al. (2021). On both implementations, the performance of the GADF technique remains similar, although it reaches a steady accuracy level at a faster pace with the original implementation. Furthermore, this technique has the widest confidence interval among all the techniques, which is shown by the gray area. The performance of all other imaging techniques improves with the modified implementation. However, the spread in performance across techniques is wider in the modified implementation than in the original implementation, especially at lower percentages of labeled instances. It should be noted that the recurrence plot (RP) technique has the most significant improvement by reaching the highest level of accuracy at the fastest pace.

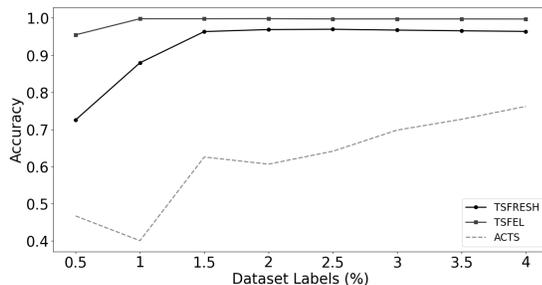
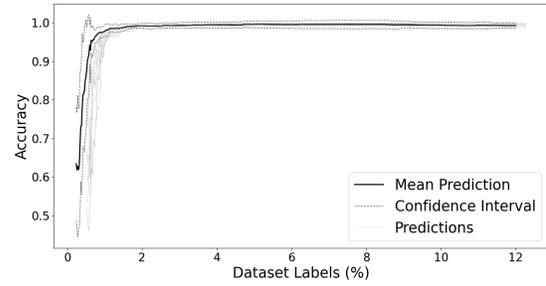
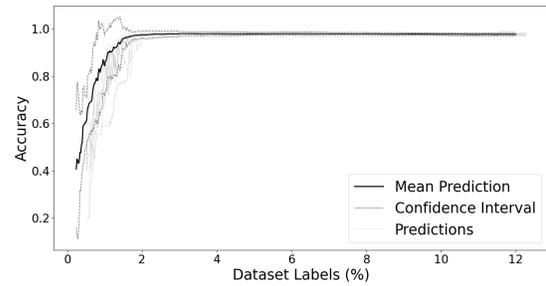


Figure 4. Learning curves for the automatic feature extraction frameworks and ACTS in the ball bearing dataset.

Figure 4 shows the learning curve for the automatic feature extraction frameworks and ACTS. Meanwhile, the confidence intervals are shown in Figure 5. It should be noted that the size of the unlabeled set is 5000 instances with equal parts of each label case. The motivation for a larger unlabeled is the ease to generate the features and the lower computational cost



(a) TSFEL



(b) TSFRESH

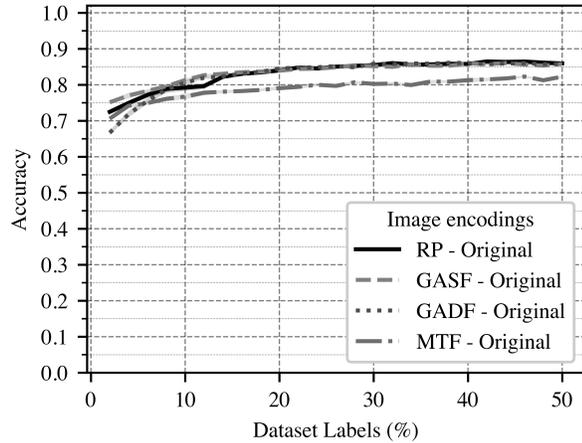
Figure 5. Comparison of the confidence intervals of the learning curves for the automatic feature extraction frameworks in the ball bearing dataset.

than in the imaging techniques. The confidence interval is the averaged result of ten time series predictions. It can be seen in Figure 4 that TSFEL has a higher accuracy level at a faster rate than TSFRESH and ACTS. This result is reinforced in Figure 5a, where even the confidence intervals described by TSFEL reach a faster convergence than TSFRESH. It should be noted that TSFEL has a lower computational cost given that it estimates a lower number of statistical features than TSFRESH.

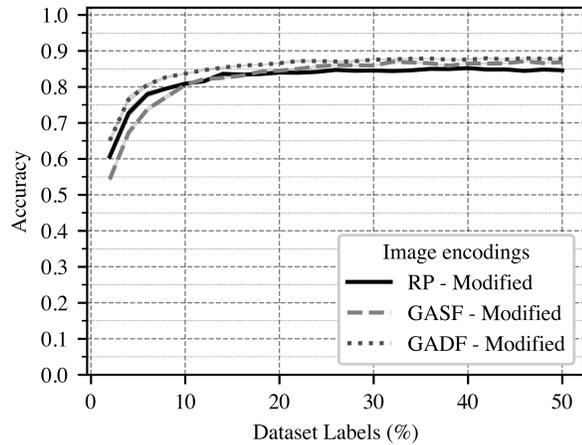
3.2. Production dataset

The time series data in the production dataset is formed by instances of different lengths resulting in the need for data padding to ensure equal data lengths for ease of calculations. Padding was carried out by extending or clipping the beginning of the time series to the initial stationary value of the series. An earlier labeling effort of the 2000 available instances was the ground truth for this dataset.

The performance of the active learning framework in this dataset is also evaluated using learning curves. Figure 6 shows the learning curves for the imaging scenario where the production time series data was converted into images. Figure 6a presents the original imaging implementation, while Figure 6b presents the modified implementation with the update proposed by Rodriguez-Garcia et al. (2021). It should be noted that it was not possible to evaluate the production



(a) Original



(b) Modified

Figure 6. Comparison of learning curves for imaging techniques in the production dataset. Grey areas represent a confidence interval of one standard deviation.

dataset using the modified MTF technique because in multiple instances the returned images were black. This situation might be partially due to the quasi-stationary behavior of the time series resulting in difficulty to partition the magnitude values on bins proportional to the quantile distribution. On this dataset, the accuracy of all imaging techniques on both implementations remains fairly similar and high. However, the original implementations tend to reach a higher accuracy at smaller sizes of the queried dataset. It is worth noting that the performance of the RP technique remains similar between the original implementation and the modified implementation, which contrasts with the results shown in Figure 3. At the same time, GASF and GADF have improved performance in the modified implementation with respect to the original implementation.

Figure 7 shows the learning for the automatic feature extrac-

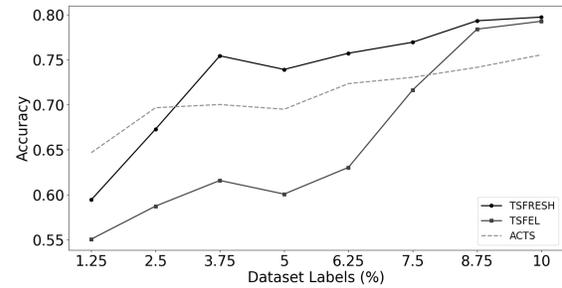


Figure 7. Learning curves for the automatic feature extraction frameworks and ACTS in the production dataset.

tion frameworks and ACTS, while the confidence intervals are shown in Figure 8. Similar to the ball bearing dataset, the confidence intervals are averaged over ten time series predictions. It can be seen in Figure 7 that TSFRESH has a higher accuracy at most of the dataset queries percentages than TSFEL and ACTS. Similarly, the confidence levels for accuracy of TSFRESH shown in Figure 8b are higher than TSFEL. However, both automatic feature extraction frameworks require a similar amount of queries of approximately 12.5% before they reach convergence. This result differs significantly with respect to the ball bearing dataset, which required a lower number of queries, at approximately 2% before reaching convergence despite having a larger number of categories.

4. DISCUSSION

We present an active learning framework that focuses on time series data, such as vibration or industrial process signals, belonging to condition monitoring and maintenance applications. In particular, we investigate the use of time series imaging and automatic feature extraction to generate features to be presented to an annotator for labeling and subsequently to be used by a classifier algorithm. The aim is to facilitate the annotation of time series data by having created an interface that enables to assign labels to the time series shown to the annotator. Furthermore, in our framework, we had considered the situation where there are no labels present at the beginning of the active learning algorithm and we solve that challenge by incorporating a pre-clustering step that intends to identify some samples belonging to a predefined number of expected labels. We find that the framework proposed by this work is capable of achieving high accuracy by just having to label 10% of the overall unlabeled dataset. These results are observed across two case studies, which included vibration data and industrial data. However, each dataset has a different performance depending on the feature extraction technique used. The ball bearing dataset achieved high accuracy at a faster rate using automatic feature extraction techniques, while the production dataset achieved high accuracy with the time series imaging techniques. These results moti-

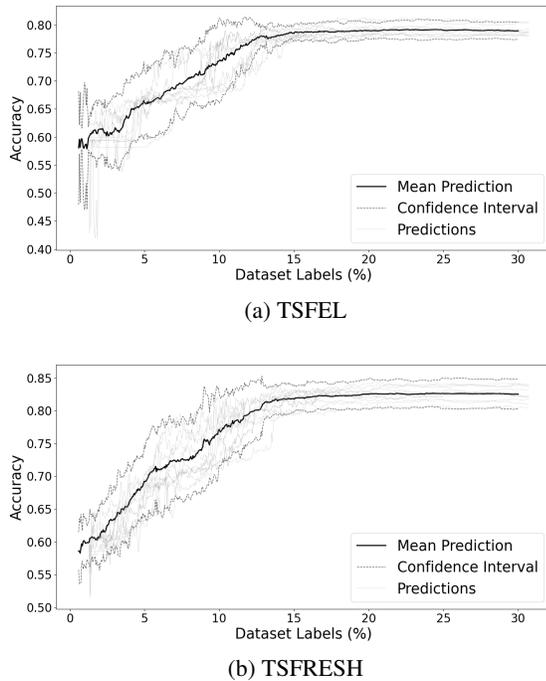


Figure 8. Comparison of the confidence intervals of the learning curves for the automatic feature extraction frameworks in the production dataset.

vate further investigations of the framework in a wider set of time series data to identify if vibration and process data have more suitable feature extraction techniques only to them. Furthermore, an investigation of additional active learning query strategies would be beneficial to understand the limitations of the framework. Active learning frameworks are a useful approach to deal with the large amounts of unlabeled data that is constantly produced by IoT systems. Further work is required to integrate these approaches into an efficient production pipeline that can manage the unlabeled data as it becomes available and users just need to respond to uncertain conditions that might require further investigation. Active learning can help in reducing the amount of work and expenses related to the labeling process of data while helping to maintain a level of consistency on the labeling procedure across users.

ACKNOWLEDGMENT

The authors would like to express their gratitude towards Mohsen Nosratinia, Fredrik Wartenberg, and Rajet Krishnan for the support and feedback provided while carrying out this work. We also appreciate the conversations and exchanges with Deepesh Agarwal, Pravesh Srivastava, Balanatarajan Subramaniam, and Babji Srinivasan.

REFERENCES

- Agarwal, D., Srivastava, P., Martin-del Campo, S., Nataraajan, B., & Srinivasan, B. (2021). Addressing uncertainties within active learning for industrial IoT. In *IEEE 7th world forum on internet of things*.
- Barandas, M., Folgado, D., Fernandes, L., Santos, S., Abreu, M., Bota, P., ... Gamboa, H. (2020). TSFEL: Time series feature extraction library. *SoftwareX*, 11.
- Christ, M., Braun, N., Neuffer, J., & Kempa-Liehr, A. W. (2018). Time series feature extraction on basis of scalable hypothesis tests (tsfresh – a python package). *Neurocomputing*, 307, 72-77.
- Hausmann, E., Fenzi, M., Chitta, K., Ivanecky, J., Xu, H., Roy, D., ... Alvarez, J. M. (2020). Scalable active learning for object detection. In *2020 IEEE intelligent vehicles symposium (iv)* (p. 1430-1435).
- Haut, J. M., Paoletti, M. E., Plaza, J., Li, J., & Plaza, A. (2018). Active learning with convolutional neural networks for hyperspectral image classification using a new bayesian approach. *IEEE Transactions on Geoscience and Remote Sensing*, 56(11), 6440-6461.
- Kumar, P., & Gupta, A. (2020). Active learning query strategies for classification, regression, and clustering: a survey. *Journal of Computer Science and Technology*, 35(4), 913-945.
- Loparo, K. (2003). *Bearing vibration data set*. Case Western Reserve University. (<http://csegroups.case.edu/bearingdatacenter/>)
- Lughofer, E. (2012). Hybrid active learning for reducing the annotation effort of operators in classification systems. *Pattern Recognition*, 45(2), 884-896.
- Marutho, D., Handaka, S. H., Wijaya, E., et al. (2018). The determination of cluster number at k-mean using elbow method and purity evaluation on headline news. In *2018 international seminar on application for technology of information and communication* (pp. 533-538).
- Peng, F., Luo, Q., & Ni, L. M. (2017). ACTS: An active learning method for time series classification. In *IEEE 33rd international conference on data engineering* (p. 175-178).
- Rodriguez-Garcia, G., Michau, G., Ducoffe, M., Sen-Gupta, J., & Fink, O. (2021). Temporal signals to images: Monitoring the condition of industrial assets with deep learning image processing algorithms. *Proceedings of the Institution of Mechanical Engineers, Part O: Journal of Risk and Reliability*.
- Rosales, R., Krishnamurthy, P., & Rao, R. B. (2007). Semi-supervised active learning for modeling medical concepts from free text. In *Sixth international conference on machine learning and applications* (p. 530-536).

- Settles, B. (2009). Active learning literature survey. *Technical Report 1648 - University of Wisconsin-Madison Department of Computer Sciences*.
- Settles, B. (2011, 16 May). From theories to queries: Active learning in practice. In I. Guyon, G. Cawley, G. Dror, V. Lemaire, & A. Statnikov (Eds.), *Active learning and experimental design workshop in conjunction with aistats 2010* (Vol. 16, pp. 1–18). Sardinia, Italy: JMLR Workshop and Conference Proceedings.
- Shekhar, S., Ghavamzadeh, M., & Javidi, T. (2021). Active learning for classification with abstention. *IEEE Journal on Selected Areas in Information Theory*, 2(2), 705-719.
- Souza, V., Rossi, R. G., Batista, G. E., & Rezende, S. O. (2017). Unsupervised active learning techniques for labeling training sets: An experimental evaluation on sequential data. *Intelligent Data Analysis*, 21(5), 1061–1095.
- Wang, L. (2005). *Support vector machines: theory and applications* (Vol. 177). Springer Science & Business Media.
- Wang, Z., & Oates, T. (2015). Imaging time-series to improve classification and imputation. In *Twenty-fourth international joint conference on artificial intelligence*.
- Zhang, S., Zhang, S., Wang, B., & Habetler, T. G. (2020). Deep learning algorithms for bearing fault diagnostics—a comprehensive review. *IEEE Access*, 8, 29857-29881.
- Zhu, X. J. (2005). Semi-supervised learning literature survey. *Technical Report 1530 - University of Wisconsin-Madison Department of Computer Sciences*.

BIOGRAPHIES



Sergio Martin-del-Campo received the B.Sc. degree in Mechatronic Engineering from Monterrey Institute of Technology and Higher Education (ITESM), Guadalajara, Mexico in 2007, the M.Sc. degree in Space Science and Technology from Luleå University of Technology (LTU), Sweden and Julius Maximilians University of Würzburg, Germany in 2012, and the Ph.D. degree in Industrial Electronics from LTU in 2017. He completed a post-doc in 2020 in the Embedded Intelligent Systems LAB (EISLAB) at LTU.

Since 2020, he is a Data Scientist at Viking Analytics focused on analytics and machine learning approaches to condition

monitoring, maintenance and manufacturing. His research interests include anomaly detection and time series analysis.



William Lindskog received his M.Sc. degree in Industrial Engineering and Management, majoring in Data Science, from Lund University, Faculty of Engineering, Lund, Sweden and ETH Zürich, Zürich, Switzerland in 2021. He wrote his master thesis about Time Series Active Learning using Automated Feature Extraction with Viking Analytics. He is an incoming Industrial PhD student at DENSO Automotive Deutschland GmbH and Technische Universität München, München, Germany where he will pursue his Ph.D. in Connected Vehicles.



Davide Badalotti received the B.Sc. degree in Physics from Università degli Studi di Milano (UNIMI), Milano, Italy in 2019. He is currently pursuing his M.Sc. degree in Data Science at Università degli Studi di Milano Bicocca (UNIMIB), Milano, Italy. His main interests include quantum machine learning and time series analysis.



Arash Toyser received the B.Sc. degree in Electrical Engineering from Sharif University of Technology, Tehran, Iran in 2002, the M.Sc. degree in Wireless Communication from Royal Institute of Technology (KTH), Stockholm, Sweden in 2004, and the Ph.D. degree in Communication from Chalmers University of Technology, Gothenburg, Sweden in 2010. Since 2018, he is a Data Scientist at Viking Analytics focused on machine learning solutions for Industry 4.0 applications. His research interests include semi-supervised and active learning.



Oskar Liew received the B.Sc. degree in Engineering Physics in 2018 and the M.Sc. degree in Engineering Mathematics and Computational Science in 2020 from Chalmers University of Technology, Gothenburg, Sweden. On his thesis, he investigated integrative data analysis methods and explainable machine learning. He has been working as a Data Scientist and Software Developer at Viking Analytics since 2020 with focus on topics such as active learning and machine learning model deployment. His current research interests include human-in-the-loop machine learning and explainable AI.