IEEE Access
Multidisciplinary ⋮ Rapid Review ⋮ Open Access Journal

Received June 2, 2021, accepted June 12, 2021, date of publication June 22, 2021, date of current version June 30, 2021.

Digital Object Identifier 10.1109/ACCESS.2021.3091476

A TOPSIS-Assisted Feature Selection Scheme and SOM-Based Anomaly Detection for Milling Tools Under Different Operating Conditions

MARYAM ASSAFO[1] AND PETER LANGENDÖRFER[1,2]
[1]Chair of Wireless Systems, BTU Cottbus–Senftenberg, 03046 Cottbus, Germany
[2]IHP–Leibniz-Institut für innovative Mikroelektronik, 15236 Frankfurt (Oder), Germany

Corresponding author: Maryam Assafo (maryam.assafo@b-tu.de)

This work was supported by the Federal Ministry of Education and Research of Germany (BMBF) through the iCampus Cottbus Project under Grant 16ES1128K.

ABSTRACT Anomaly detection modeled as a one-class classification is an essential task for tool condition monitoring (TCM) when only the normal data are available. To confront with the real-world settings, it is crucial to take the different operating conditions, e.g., rotation speed, into account when approaching TCM solutions. This work mainly addresses issues related to multi-operating-condition TCM models, namely the varying discriminability of sensory features with different operating conditions; the overlap between normal and anomalous data; and the complex structure of input data. A feature selection scheme is proposed in which the Technique for Order Preference by Similarity to Ideal Solution (TOPSIS) is presented as a tool to aid the multi-objective selection of sensory features. In addition, four anomaly detection approaches based on Self-Organizing Map (SOM) are studied. To examine the stability of the four approaches, they are applied on different single-operating-condition models. Further, to examine their robustness when dealing with complex data structures, they are applied on multi-operating-condition models. The experimental results using the NASA Milling Data Set showed that all the studied anomaly detection approaches achieved a higher assessment accuracy with our feature selection scheme as compared to the Principal Component Analysis (PCA), Laplacian Score (LS), and extended LS in which we added a final step to the original LS method in order to eliminate redundant features.

INDEX TERMS Anomaly detection, condition monitoring, feature selection, one-class classification, predictive maintenance, self-organizing feature maps, TOPSIS.

I. INTRODUCTION

Milling is one of the popular operations in the manufacturing industry. It can be used to manufacture different geometric components, e.g., flat surfaces, grooves, threads, etc. The health state of the milling tool has a significant influence on the quality of milling processes as well as the availability of the production line [1]. The tool wear is a natural result of the thermal-force coupling effect that the tool encounters during machining [2]. A worn tool adversely affects the surface finish and dimension accuracy of the finished products, leading to reworking or scrapping the workpiece [3], which in turn causes time and financial losses [4]. Moreover, the tool wear could eventually lead to the tool breakage which is considered the major cause of the unscheduled machine downtime. Handling the tool wear problem using the schedule-based maintenance strategy, i.e., preventive maintenance, could cause the tool to be replaced even when it is still in a good condition, leading to an inefficient use of the available resources and an increase in the overall costs. On the other hand, predictive maintenance is based on monitoring the actual condition of the system, and making the maintenance plan accordingly. Hence, tool condition monitoring (TCM) allows for replacing the tool in a timely manner, and thus, achieving the joint objective of maximizing the tool usage and avoiding the consequences of using a worn tool. It is estimated that an effective TCM system could reduce costs by 10-40% [1].

The proliferation of sensors and the advancements in the machine learning area makes it possible to develop automated TCM systems, which in turn enables the realization of smart manufacturing. Automated TCM systems involve deploying

The associate editor coordinating the review of this manuscript and approving it for publication was Ming Luo.

sensors to measure variables that are sensitive to the tool condition. When the traditional machine learning methods are implemented for TCM, features are extracted first from the raw sensor signals, and then a few representative features are selected to train the machine learning model. On the other hand, most of the approaches based on Deep Learning (DL) do not involve feature extraction/selection processes as the deep model can extract features automatically from the raw data [5].

The TCM task is handled in some researches, e.g., [6], [7], as a multi-class classification problem where different classes corresponding to different tool conditions are used to label the run-to-failure data, then the labeled data are used to train the supervised learning model. A relevant issue to the multi-class classification is when the training normal samples are far more than the abnormal ones. To this end, some classification approaches were shown to be resistant to the imbalance class problem, e.g., the Self Organizing Map (SOM)-based classifier employed in [8], and the hybrid approach proposed in [9] that combines the extended Kalman filter and cost-sensitive extreme learning machine. Intuitively, the availability of labeled run-to-failure data allows for learning the different data patterns associated with the normal and worn tools [4]. However, considering the fact that performing machining using a worn tool will generate poor quality products and that it is time consuming to acquire labeled data on the shop floor, using a worn tool merely for the sake of collecting measurements could lead to a considerable waste of resources and time. Therefore, relying only on the normal data can represent an efficient and easily applicable implementation of the predictive maintenance [4], [10].

Anomaly detection is one of the essential elements of predictive maintenance. It is usually modeled as a one-class classification problem which is particularly useful when only the normal behavior of the system is known, and no knowledge of the abnormal behavior is available. Data-driven reconstruction methods can be employed to build a model representing normal data, e.g., SOM [11], Generative Probabilistic Adversarial Autoencoder (GPAA) [12], Generative Adversarial Network (GAN) [4], archetypal analysis [13]. These methods are based on projecting the training data onto another data space with the aim of minimizing the reconstruction errors [13]. An unknown sample is then assessed as normal or anomalous based on a health indicator that is derived from the built model [12], [13]. Data-driven statistical methods can also be implemented, e.g., the authors in [14] combined Re-weighted Minimum Covariance Determinant (RMCD) estimator and Hidden Markov Model (HMM) to perform a two-step anomaly detection. Furthermore, anomaly detection can be realized through the Discord Detection (DS) technique. For instance, the authors in [15] proposed a novel computational approach to identify discords in multivariate time-series data. In addition to detecting the abnormal behavior of a system, anomaly detection can also be implemented with other predictive maintenance tasks, e.g., to improve the estimation accuracy of the remaining useful life [16].

Regarding the existing researches that addressed the anomaly detection task, little attention has been paid to handle the problem under different operating conditions of the milling tool, e.g., rotation speed, depth of cut, etc. Considering the real-world settings, the same tool is subject to different operating conditions during its lifetime [17]. A built TCM model will not perform as expected when applied to data taken under different operating conditions from those of the training data, as it is experimentally shown in [8]. Hence, the learning model that fits data under different operating conditions will show robustness on the shop floor where different, non-stationary operating conditions exist [6], [18]. Even though the operating conditions cannot be indicative of the system condition, their significant role in the predictive maintenance stems from the fact that they influence the system degradation rate [17], [19] as well as the variables that are used to monitor the system health [17], [20]. Three main issues arise when building a multi-operating-condition model, i.e., a model that is trained to represent data from different operating conditions.

1) The ability to distinguish between normal and anomalous data based on a specific sensory feature varies with different operating conditions, as experimentally shown in [8].
2) The likelihood of overlapping between normal and anomalous data increases when dealing with multi-operating-condition data. Since the sensor signals carry the combined effect of the system health and the operating conditions, similar sensor patterns could be generated under different health conditions, which makes the anomaly detection even more challenging [20].
3) The data structure under different operating conditions is highly complex. This is due to the fact that the data variation is not only driven by the system dynamics caused by the tool wear progression, but also by the different operating conditions. This could lead some anomaly detection approaches to fail to achieve a satisfactory diagnosis accuracy.

Handling the aforementioned issues using only the normal data makes it even more challenging. In [4] and [12], a multi-operating-condition model is built and tested for the purpose of anomaly detection. However, they are both based on DL approaches that can automatically perform the feature extraction/selection processes, which does not allow the researchers to know the specific sensory features that led to the results achieved. On the other hand, most of the existing unsupervised feature selection approaches, e.g., [21], consider only the case when the data are taken from one set of operating conditions.

This paper aims to present a rigorous, interpretable framework for the anomaly detection for milling tools under different operating conditions. SOM is chosen as the data-driven learning approach. SOM is an unsupervised neural network

that allows for mapping high-dimensional input data onto a low-dimensional grid of neurons. It does not have high demands in terms of the required memory and computations [22], which facilitates its implementation on an edge device, and achieving a real-time tool monitoring. Furthermore, it does not need large training data to generalize [23], and it has the merits of interpretability and understandability [24]. SOM has been employed in various predictive maintenance applications, including anomaly detection [11], fault classification [25], and remaining useful life estimation [26].

The contribution of this paper can be summarized as follows.

1) A feature selection scheme is proposed to tackle the first two aforementioned issues associated with the multi-operating-condition models. The scheme aims to select the sensory features whose discrimination ability is independent of the operating conditions being used, while reducing the overlap between normal and anomalous data of different operating conditions. In this scheme, a discrimination metric that exists in the literature is used in two forms that are tailored for the problem at hand, and the Technique for Order Preference by Similarity to Ideal Solution (TOPSIS) is employed to rank the sensory features. TOPSIS has been utilized in multitude of areas, including business, energy management, chemical engineering, manufacturing [27], mobile communications [28]. Its manufacturing-related applications include the selection of the best lubricant, tool insert, and machining parameters, among others [29]. To the best of our knowledge, it has not been used yet for the selection of sensory features in the area of predictive maintenance. We show in this paper that TOPSIS could be used as a tool to aid the multi-objective feature selection.
2) An investigation and analytical study of four different SOM-based anomaly detection approaches are conducted in this paper using the NASA Milling Data Set [30]. The powerful ability of SOM to represent data of complex structures is already well-known. However, the built-upon-SOM anomaly detection approaches are yet to be investigated under different operating conditions. In order to examine the stability and robustness of the studied approaches, they are applied on different single-operating-condition models as well as multi-operating-condition models.

The remainder of the paper is organized as follows. Related works are reviewed in Section II. The main workflow and assumptions are described in Section III. Our feature selection scheme is presented in Section IV. The background about SOM as well as the studied SOM-based anomaly detection approaches are presented in Section V. Section VI encompasses the implementation details, including the experimental data, the implemented scenarios, the conducted feature extraction/selection processes, as well as the different results and discussion. Finally, the paper is concluded and the future directions are presented in Section VII.

II. RELATED WORKS

Performing diagnosis tasks under different operating conditions is considered in some researches, yet it is rarely investigated and reviewed thoroughly. Due to the lack of related works, and in order to present the most relevant insights to the problem at hand, the literature review in this section is not only confined to the anomaly detection for milling tools. Concerning the researches that considered different operating conditions in the diagnosis tasks, three main approaches can be distinguished in the literature.

1) The first approach is based on building separate single-operating-condition models, i.e., each model fits data under only one set of operating conditions. This approach is used in some works, e.g., [18], [31], [32], to examine the generalization ability of a specific diagnosis approach when applied to different data sets belonging to different sets of operating conditions. It is also adopted in some researches, e.g., [33], to avoid the complexity of setting a decision boundary that is associated with the multi-operating-condition models.
2) The second approach is based on building a multi-operating-condition model, i.e., mixed data from different operating conditions are used to train and test the model. The robustness of a specific model can be reflected in its ability to achieve a satisfactory system diagnosis under different operating conditions [6], [18]. Research examples that fall under this category include [4], [6], [7], [18].
3) The third approach is based on building several multi-operating-condition models so that each model fits only data with specific sets of operating conditions. To realize this, clustering of the operating conditions is performed first, and then a general model is built for the data corresponding to each cluster. This approach is used in [34] and [20] to perform an anomaly detection for aircraft engines and wind turbines, respectively. In [34], the effect of the operating conditions is removed first from the original data, and then a SOM model is trained by the residuals. In [20], an anomaly detection model based on isolation forest is built for each cluster of the operating conditions. However, since the formed clusters of operating conditions might need to be validated by an expert [20], the third approach is not considered any further in this paper.

The first approach that involves building separate single-operating-condition models poses no special challenges, at least from the individual model point of view. On the other hand, several challenges arise when building a multi-operating-condition model as discussed in Section I. The effect of incorporating multiple operating conditions on the assessment accuracy of the model can be found in the researches that applied both the first and second aforementioned approaches, as in [18] and [12]. The authors in [18] used a Sparse Auto Encoder (SAE) for feature fusion, and a Deep Belief Network (DBN) for classifying bearing faults. Three separate models are built for 3 different bearing speeds,

and one general model is built for the mixed data. The accuracy of the general model was about 6% less than the average accuracy of the separate models. In [12], GPAA was proposed for anomaly detection tasks. The accuracy of GPAA model under 2 sets of operating conditions was about 5% less than that of the single-operating-condition model.

Most of the existing multi-operating-condition TCM models are performed through a supervised multi-class classification. In [7], the Principal Component Analysis (PCA) is used for feature reduction, and a support vector machine-based classifier is employed to diagnose the tool under 4 operating conditions. In [6], a correlation-based feature selection is performed, and different machine learning approaches and ensemble methods are tested to classify the tool state under 8 operating conditions. In both [7] and [6], machining parameters, e.g., speed rotation, are used along with the sensory features to feed the learning model. The approach presented in [19] is based on removing the component related to the operating conditions from the original spindle current signal, and then feeding the residuals to the Deep Convolutional Neural Network (DCNN) that in turn classifies different tool states. In [35], four model coefficients of the cutting force signal are shown to be independent of the operating conditions while being correlated with the tool wear. However, both the approaches in [19] and [35] are based on particular sensor signals, which limits their applicability with other sensor types.

The authors in [13] developed an anomaly detection mechanism through a non-convex archetypal analysis. The suitability of this approach for non-convex data makes it promising for the data taken under different operating conditions.

Regarding the anomaly detection-based TCM, most of the researches consider only one set of operating conditions, such as [10] in which DCNN is employed. On the other hand, the authors in [4] applied a DL anomaly detection model based on GAN to detect anomalous tool states under 9 different operating conditions. However, most of the DL-based approaches rely on the inherent capability of the deep networks to automatically perform the tasks of feature extraction/selection, which does not allow for knowing the particular sensory features that contributed to obtaining the results achieved. Moreover, DL methods usually need large training data, long training time, and high computational demands, which could limit their applicability in the industrial applications.

As mentioned in Section I, SOM is selected as the learning approach in this paper. Several SOM-based anomaly detection approaches exist in the literature, e.g., [11], [34], [36]. They mainly differ in the health indicator, distance measure, and threshold-setting technique. Examining the ability of TCM approaches to maintain a consistent performance under different operating conditions is essential to ensure a reliable TCM in the real-world manufacturing settings. In [37], a comparative evaluation of different SOM-based anomaly detection approaches was performed. However, the data sets that were used are non-time series data. Moreover, the effect of the different operating conditions are not considered in the evaluation. Therefore, the literature still lacks an investigation and an in-depth study of different SOM-based anomaly detection approaches under different operating conditions. To fill this gap, we perform a comprehensive comparison of different SOM-based approaches, with the aim of examining their stability when applied to different single-operating-condition models, as well as their robustness when applied to multi-operating-condition models. Since the effectiveness of anomaly detection is not only related to the applied approach but also to the feature samples that are used to train SOM, it is of great importance to select the most significant sensory features with regard to the task at hand. Most of the feature selection schemes proposed in the literature are designed based on data from only one set of operating conditions, e.g., [21]. Furthermore, the most commonly used feature reduction/selection techniques, e.g., PCA [7], Laplacian Score (LS) [38], are based on underlying assumptions regarding what would lead to good results, e.g., the criterion for PCA is the data variance, and it is the locality preserving power for LS, which might not be consistent with multi-operating-condition data. In Section IV, a feature selection scheme which utilizes multi-operating-condition normal data is proposed.

III. THE MAIN WORKFLOW AND SYSTEM DESCRIPTION

As mentioned previously, the main focus of this work is on addressing the issues related to building a multi-operating-condition SOM model that can be employed for anomaly detection for milling tools based on a one-class classification. In this work, the tool is regarded as normal when its health condition still allows for carrying out accurate machining and generating good quality products. Otherwise, it is considered worn and should be replaced.

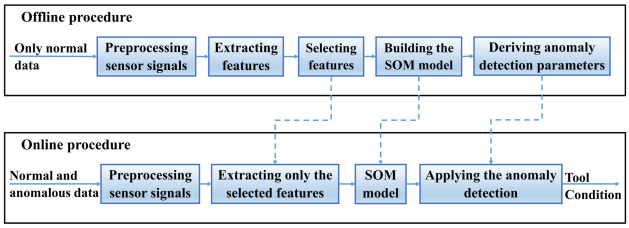


FIGURE 1. The main workflow of the offline and online procedures performed in this paper for TCM; dashed lines represent the outcome of the offline procedure that will be used in the online one.

The main workflow of the TCM performed in this paper is divided into offline and online procedures, as depicted in Fig. 1. The offline procedure encompasses the different processes and decisions that are needed to build the SOM model as well as to derive the required parameters for the anomaly detection. Only the normal data corresponding to the normal tool are used in this procedure. First, features are extracted from the preprocessed raw sensor signals.

Second, the most significant features are determined based on our proposed feature selection scheme that will be presented in Section IV. The selected features are then used to build the SOM model, and finally anomaly detection-related parameters are derived from the trained SOM. The offline procedure outcome, represented by the selected features, the trained SOM, and the parameters for the anomaly detection approach, will be used in the online procedure where data from both the normal and worn tool are presented, and an assessment of the tool condition is made at last.

Since we aim to build a multi-operating-condition SOM model, normal data observed under different operating conditions are needed to train the model. In the next paragraph, we clarify the major assumptions and notions needed for our subsequent analysis.

Let $\{case^i\}, i = 1, \ldots, L$, be a set of experimental cases where each case represents observations taken during the milling process under a specific set of operating conditions, e.g., rotation speed, depth of cut, etc. Each observation comprises a time stamp and the sensor signals recorded during the observation period. For each case, the observations are stored in time order as the milling machine is running, and cover the portion of the tool lifetime when the tool is still normal, i.e., starting from when the tool is new up to a certain point of its lifetime. Hence, the observations under each case do not constitute run-to-failure data but rather degradation data. Furthermore, the available normal data are not labeled with respect to the exact tool state. It should be noted that the number of observations might not be the same for all the cases, as fewer observations are expected for the cases in which the tool wear progresses faster. Let $\{F_f\}, f = 1, \ldots, N$, be a set of the extracted features, e.g., mean, variance, etc., from each of the recorded sensor signals.

IV. OUR FEATURE SELECTION SCHEME FOR MULTI-OPERATING-CONDITION MODELS

Feature selection aims to reduce the dimension of the feature sample that will be an input to the learning model by selecting the most significant features among the extracted ones, while excluding irrelevant and redundant features. This is not only crucial to reduce the computation and storage demands, but also to not degrade the learning model performance [39].

The aim of our proposed scheme is to select the features with the highest discrimination ability under different operating conditions while eliminating redundant features. The scheme can be described briefly in the following steps:

1) Calculating the discriminability indicators of all the extracted sensory features.
2) Using TOPSIS to rank the features based on their calculated discriminability indicators.
3) Eliminating the redundant features based on the inter-feature Pearson's correlation.
4) Selecting the top n features of the remaining ranked features.

A. CALCULATING THE DISCRIMINABILITY INDICATORS

The significance of each extracted feature is examined on the basis of its discrimination ability with respect to different patterns of data. The Statistical Overlap Factor (SOF) presented in [40] is employed for this purpose. SOF is used to measure the ability of a given feature to discriminate between two different classes. For this, labeled data are needed to identify the two classes. SOF is given as in (1) [40].

$$\mathrm{SOF} = \left| \frac{\mu_1 - \mu_2}{(\sigma_1 + \sigma_2)/2} \right| \tag{1}$$

where μ_1 and σ_1 are respectively the mean and standard deviation of the feature samples corresponding to the first class, μ_2 and σ_2 are respectively the mean and standard deviation of the feature samples corresponding to the second class. However, since the focus of this work is on dealing with unlabeled normal data, we will make a coarse division of the normal data under each case, $case^i$, so that SOF can be implemented.

Recall that the observations representing the normal data are stored in time order, the observations under each case can be divided into two groups— the first group contains the first 70% of the total observations, whereas the second one includes the remaining 30% of the observations. It can be considered that the tool will be noticeably more degraded over the second group as compared to the first one. As such, the first group is more representative of the healthy state of the tool, whereas the second group is more representative of the degraded state. Such a division could help to exploit the known normal data to make a rough estimation of the feature behavior over the unknown anomalous data. Therefore, the SOF metric will be used in this paper within the same class, i.e., normal state, but with the two aforementioned data groups representing the healthy and degraded states.

Since the model will be trained by data from different operating conditions, it is important to measure the discriminability of the feature not only within the data of the same set of operating conditions but also with the different ones. As such, two SOF-based discriminability indicators of the feature f are used in our scheme proposed in this section, namely intra-case SOF and inter-case SOF.

- Intra-case SOF, denoted by $SOF_f^i(intra)$: It measures the ability of the feature f to separate between the healthy and degraded states under the intended case, $case^i$.
- Inter-case SOF, denoted by $SOF_f^i(inter)$: It measures the ability of the feature f to separate between the healthy state under the intended case, $case^i$, and the degraded states under the other cases, $\{case^j\}, j = 1, \ldots, K$; $case^j \in \{case^i\}$; $j \neq i$, whose operating conditions are different from those of the intended case, $case^i$.

SOF (intra-case) and SOF (inter-case) under the $case^i$ are calculated by (2) and (3), respectively.

$$SOF_f^i(intra) = \left| \frac{\mu_{f1}^i - \mu_{f2}^i}{\left(\sigma_{f1}^i + \sigma_{f2}^i\right)/2} \right| \tag{2}$$

$$SOF_f^i(inter) = \frac{1}{K}\sum_{j=1}^{K}\left|\frac{\mu_{f1}^i - \mu_{f2}^j}{\left(\sigma_{f1}^i + \sigma_{f2}^j\right)/2}\right|; \quad j \neq i \tag{3}$$

where μ_{f1}^i and σ_{f1}^i are respectively the mean and standard deviation of feature f's samples corresponding to the healthy state under the $case^i$, whereas μ_{f2}^i and σ_{f2}^i are respectively the mean and standard deviation of feature f's samples corresponding to the degraded state under the $case^i$. Similarly, μ_{f2}^j and σ_{f2}^j are respectively the mean and standard deviation of feature f's samples corresponding to the degraded state under the $case^j$. The equation (3) represents the average of the individual SOF values calculated for the intended case, $case^i$, and all the cases, $\{case^j\}$, that have different operating conditions than those of the intended case, $case^i$. Fig. 2 illustrates the concept of intra-case SOF and inter-case SOF for feature f under $case^i$.

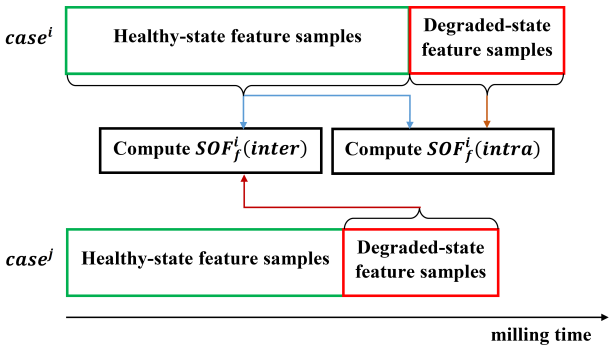


FIGURE 2. The concept of intra-case SOF and inter-case SOF used in our proposed feature selection scheme to assess the discriminability of the sensory feature f with the multi-operating-condition data; the operating conditions of $case^i$ and $case^j$ are different.

Given that the significance of a specific feature varies with different operating conditions [8] and the aim here is to use the same learning model for the different operating conditions incorporated in building the model, it is critical to take all these operating conditions into account when selecting the features. To realize this, for each feature f, the value of each of the two discriminability indicators will be calculated first for the individual cases, and then an average over the total cases will be calculated. Hence, the final discriminability indicators for the feature f are represented by (4) and (5).

$$SOF_f(intra) = \frac{1}{L}\sum_{i=1}^{L} SOF_f^i(intra) \tag{4}$$

$$SOF_f(inter) = \frac{1}{L}\sum_{i=1}^{L} SOF_f^i(inter) \tag{5}$$

where L is the number of all the cases represented by the TCM model, as mentioned in Section III.

B. RANKING SENSORY FEATURES BASED ON TOPSIS

TOPSIS is one of the multi-criteria decision-making methods. It is characterized by being simple, logical and computationally efficient, and is considered one of the best ranking methods [41]. It is based on an organized procedure that ranks the alternatives with respect to some attributes (criteria) in such a way that the best alternative would have the smallest distance from the ideal solution and the largest distance from the non-ideal solution [42].

In the following paragraphs, the ranking procedure of TOPSIS will be reviewed first, and then our idea to apply TOPSIS for ranking sensory features is presented.

Let $\{A_f\}, f = 1, \ldots, N$, represents a set of alternatives. Each alternative is characterized by a set of attributes, $\{C_c\}, c = 1, \ldots, M$. The extent to which the attribute C_c contributes to ranking the alternatives is determined by its weight w_c, where the attributes' weights are fed to TOPSIS algorithm and $\sum_c w_c = 1$. The value x_{fc} represents the attribute C_c corresponding to the alternative A_f. The decision matrix D contains all the values of x_{fc}, as in (6).

$$D = \begin{bmatrix} x_{11} & \ldots & x_{1M} \\ \vdots & \ddots & \vdots \\ x_{N1} & \ldots & x_{NM} \end{bmatrix} \tag{6}$$

The ranking procedure of TOPSIS involves the following steps [42]:

Step 1. Normalizing the elements of the decision matrix D with respect to each attribute C_c. This step accounts for the different units and scales among the different attributes.

$$r_{fc} = \frac{x_{fc}}{\sqrt{\sum_{f=1}^{N} x_{fc}^2}}; \quad f = 1, \ldots, N; \; c = 1, \ldots, M \tag{7}$$

Step 2. Constructing the weighted normalized matrix by multiplying each normalized value r_{fc} with its corresponding weight w_c.

$$v_{fc} = r_{fc} \times w_c; \quad f = 1, \ldots, N; \; c = 1, \ldots, M \tag{8}$$

Step 3. Determining the positive ideal solution (S^+) and the negative ideal solution (S^-) with respect to all the attributes.

$$\begin{aligned} S^+ &= \{v_1^+, \ldots, v_M^+\} \\ &= \left\{\left(\max_f v_{fc} | c \in C'\right), \left(\min_f v_{fc} | c \in C''\right)\right\} \end{aligned} \tag{9}$$

$$\begin{aligned} S^- &= \{v_1^-, \ldots, v_M^-\} \\ &= \left\{\left(\min_f v_{fc} | c \in C'\right), \left(\max_f v_{fc} | c \in C''\right)\right\} \end{aligned} \tag{10}$$

where C' represents the set of benefit attributes (the larger the better), and C'' represents the cost attributes (the smaller the better).

Step 4. Calculating the distances d_f^+ and d_f^- that separate each alternative from S^+ and S^-, respectively.

$$d_f^+ = \sqrt{\sum_{c=1}^{M}\left(v_{fc} - v_c^+\right)^2}; \quad f = 1, \ldots, N \tag{11}$$

$$d_f^- = \sqrt{\sum_{c=1}^{M}\left(v_{fc} - v_c^-\right)^2}; \quad f = 1, \ldots, N \tag{12}$$

Step 5. Calculating the relative closeness (Q_f^+) of each alternative to the ideal solution. This value represents the score of the alternative A_f.

$$Q_f^+ = \frac{d_f^-}{d_f^- + d_f^+}; \quad f = 1, \ldots, N \tag{13}$$

Step 6. Ranking the alternatives in descending order based on Q_f^+. The best alternative would be the one with the highest Q_f^+.

The attributes' weights play a significant role in ranking the alternatives. Weight determination can be broadly classified into two categories: subjective weighting and objective weighting. In the former, the importance of the attributes are determined by the decision maker, e.g., expert, based on some factors, such as knowledge, preference, perception of the problem, design constraints [43], etc. Whereas in the latter, the weights are determined based on the information provided by the decision matrix. Entropy method, which assigns higher weights to the attributes that carry more diverse information among the alternatives [43], is one of the most widely adopted objective weighting methods.

In our TOPSIS-assisted feature selection scheme, each sensory feature represents an alternative A_f. The attributes that characterize each feature are $SOF_f(intra)$ and $SOF_f(inter)$ which are calculated using (4) and (5), respectively. Hence, N represents the total number of the extracted sensory features, and the number of the attributes, i.e., M in (6)–(12), is equal to 2. It should be noted that both of the utilized attributes are benefit attributes. As for determining the attributes' weights, the objective weighting was not employed in this paper due to the fact that, in the considered system, only normal data are available in the offline procedure in which the feature selection scheme is performed (see Fig. 1). Since the feature attributes were computed based only on the normal data but the ultimate goal is to distinguish between normal and anomalous data, the attributes in the decision matrix do not fully reflect the feature behavior over all the states of interest (normal and worn tool states). Therefore, performing a fine tuning of the weights based on these values might lead to selecting suboptimal features when it comes to the online procedure in which both normal and anomalous data are presented. In light of that, a compromise decision would be to give an equal importance to the two attributes since, for multi-operating-condition models, it is crucial for the selected features to have good discriminability properties regarding data of the same-and-different operating conditions, i.e., $SOF_f(intra)$ and $SOF_f(inter)$. Thus, $w = [0.5, 0.5]$ is considered in this work. Unlike the typical TOPSIS-based decision making applications in which only the best alternative is selected, e.g., selecting the best material, or in which the rank of top alternatives is meaningful for the further decisions, e.g., ranking companies, our TOPSIS-assisted feature selection scheme selects the top n ranked features after eliminating redundant features, as it will be shown in the next step of our scheme. Moreover, the rank order of the eventually selected features carries no significance when it comes to feeding them to the learning model.

C. ELIMINATING THE REDUNDANT FEATURES

Unlike the features that carry no relevant information to the task at hand, the redundant features might be informative. However, they carry no unique information, and hence they are not considered indispensable [39]. It is a common practice in the literature to utilize the inter-feature Pearson's correlation to examine the redundancy among the extracted features [44]–[47].

To remove the redundant features, we adopted the procedure presented in [44] in which the redundant feature elimination is carried out iteratively. The main principle of this procedure is that each higher ranked feature will eliminate the lower ranked features with which it has a correlation exceeding a specific threshold. First, the highest ranked feature starts eliminating the strongly correlated features with it. Next, the highest ranked feature among the remaining features is determined, and the same elimination process is performed. The same procedure is iterated until no feature is remaining.

Since multi-operating-condition models are considered in our work, the inter-feature correlation is examined over all the cases. The inter-feature correlation between the features F_a and F_b under $case^i$ is denoted by R_{ab}^i, and is calculated through the Pearson's correlation coefficient as in (14) [48].

$$R_{ab}^i = \frac{\sum_s \left(y_s^i - \bar{y^i}\right)\left(z_s^i - \bar{z^i}\right)}{\sqrt{\sum_s \left(y_s^i - \bar{y^i}\right)^2}\sqrt{\sum_s \left(z_s^i - \bar{z^i}\right)^2}} \tag{14}$$

where y_s^i and z_s^i are the feature samples corresponding to the features F_a and F_b, respectively. $\bar{y^i}$ and $\bar{z^i}$ are the mean values of feature samples for F_a and F_b, respectively. R_{ab}^i is within the range of [-1,1]. The correlation direction (positive or negative) is not important for examining redundancy among features, but rather, the correlation strength expressed by the absolute value whose range is [0,1] [47]. The overall correlation strength between the features F_a and F_b is denoted by R_{ab} and is calculated as in (15).

$$R_{ab} = \frac{1}{L}\sum_{i=1}^{L}\left|R_{ab}^i\right| \tag{15}$$

where L is the number of the total cases. The greater R_{ab} the stronger the correlation between the two examined features. Similarly to [46], feature elimination occurs when R_{ab} exceeds 0.8.

After eliminating the redundant features, the top n features among the remaining ranked features are selected. n can be regarded as an optimization parameter since the optimum number of the selected features is related to the number of training samples as well as the overall complexity of the task performed by the learning model. Fig. 3 shows a flowchart

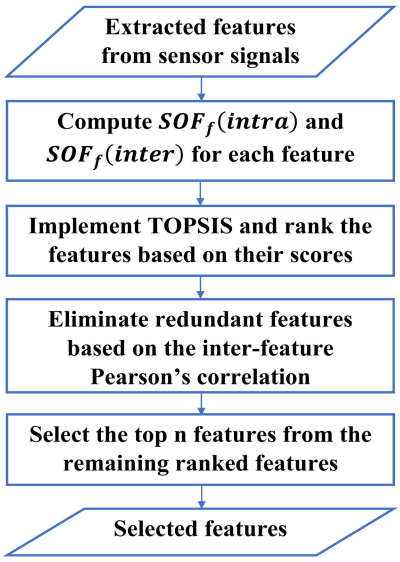


FIGURE 3. The flowchart of our feature selection scheme for the multi-operating-condition models.

illustrating our feature selection scheme proposed in this paper.

V. SOM-BASED ANOMALY DETECTION

A. BACKGROUND ABOUT SOM

SOM, also known as Kohonen maps, is an unsupervised learning neural network that maps high-dimensional input data onto a lower-dimensional, usually a 2-dimensional, grid of neurons. This mapping has the special property of preserving the topology of the input space, meaning that the input vectors that are closely spaced in the input space will be mapped onto closely spaced neurons in the grid [49]. This organized 2D-representation is particularly useful for analyzing and visualizing high-dimensional data. The topology-preserving aspect is what distinguishes SOM from other clustering techniques, e.g., k-means [22].

Each neuron is characterized by its position in the grid as well as its weight vector in the input space. The former is used to determine the spatial relationship between the different map neurons, and the latter is used to measure the distance, usually the Euclidean distance, between the input vector and the neuron. The neuron grid can be rectangular, hexagonal, etc. [49].

There are two main training algorithms for SOM, namely the sequential learning algorithm and the batch learning algorithm. In the former, one input vector is presented to the SOM at a time, whereas in the latter, all the input vectors are presented to the SOM at once [50].

For the sequential learning, when an input vector, $x(t)$, is presented to SOM, the Euclidean distances between $x(t)$ and the weight vectors of all the map neurons are calculated. The neuron that has the smallest distance from the input vector is called the winner or the Best Matching Unit (BMU), and its distance to the input vector is called the Minimum Quantization Error (MQE) [36]. Hence, if c is the index of the winner, it is determined as in (16) [50].

$$c = \arg\min_i \{||x(t) - m_i(t)||\} \tag{16}$$

where $m_i(t)$ is the current weight vector of the neuron i. $||.||$ is the Euclidean distance. The BMU and its neuron neighbors in the grid will update their weight vectors so that they get closer to the input vector in the input space, as in (17) [51].

$$m_i(t+1) = m_i(t) + \alpha(t)h_{ci}(t)[x(t) - m_i(t)] \tag{17}$$

where $m_i(t+1)$ is the updated weight vector of the neuron i, $\alpha(t)$ is the learning rate at the time t, $h_{ci}(t)$ is the neighborhood function centered at the BMU, neuron c, and is commonly chosen as in (18) [51].

$$h_{ci}(t) = \exp\left(-\frac{\text{sqrtdis}(c,i)}{2\sigma_c^2(t)}\right) \tag{18}$$

where sqrtdis(c, i) is the square of the distance between the neuron i and neuron c in the grid, $\sigma_c(t)$ represents the neighborhood radius associated with the winning neuron c at the time t. Both $\alpha(t)$ and $\sigma_c(t)$ are monotonically decreasing functions with respect to t [51].

The overall training comprises two learning phases: the initial phase and fine tuning phase. The topology-preserving mapping is achieved roughly in the initial phase where relatively large values of both the learning rate and neighborhood function are used. During the fine tuning phase, small values of learning rate and neighborhood function are used to enable the neurons to converge based on their corresponding input vectors, and thus the distribution of the input vectors is also learned in this phase [21].

The training in the batch learning algorithm is also iterative. However, all the training samples are fed to the SOM at once. The batch learning is advisable for practical applications as it converges faster and does not involve the learning rate parameter [50]. The BMUs corresponding to the input vectors are identified as in (16). The weight vectors are updated as in (19).

$$m_i(t+1) = \frac{\sum_{j=1}^{T} h_{ci}(t)x_j}{\sum_{j=1}^{T} h_{ci}(t)} \tag{19}$$

where T is the number of the training samples, and $h_{ci}(t)$ is the value of the neighborhood function for the neuron i when the input vector is x_j [51].

The number of SOM neurons has a significant impact on the representation quality and implementation efficiency. Too few neurons might cause different input patterns to be represented by the same neuron. Too many neurons, on the other hand, might cause similar input patterns to get separated on the map [23], and it also increases the computational burden [51]. It is a common practice in the literature to set the grid size based on (20) and (21) to avoid the two aforementioned scenarios [51].

$$U \approx 5\sqrt{T} \tag{20}$$

$$\frac{u_1}{u_2} \approx \sqrt{\frac{e_1}{e_2}} \tag{21}$$

where U is the number of grid neurons, and T is the number of training samples. u_1 and u_2 are respectively the numbers of rows and columns of the grid. e_1 and e_2 are respectively the largest and second largest eigenvalues of the training data's covariance matrix.

B. THE STUDIED ANOMALY DETECTION APPROACHES

SOM-based anomaly detection approaches usually involve three major steps. First, the map neurons are trained by the normal data. Second, the values of the employed health indicator are calculated for the training data. Third, a threshold is drawn from the calculated health indicator values. The health indicator serves as a similarity measure between the test samples and the normal data represented by the built model. Therefore, the test samples that generate health indicator values exceeding the threshold are regarded as anomalous.

Four SOM-based anomaly detection approaches are investigated in this paper, namely the traditional approach which is employed in many researches, e.g., [11], [52]; the approach proposed in [34]; the approach proposed in [36]; and a hybrid approach of the last two approaches. These approaches differ mainly in the employed health indicator and the threshold-setting technique.

In the traditional approach, the health indicator is the MQE. After training SOM, the threshold is set to be a quantile of 0.99 of all the MQEs corresponding to the training samples. If the MQE corresponding to a test sample exceeds the defined threshold, then the sample is assessed as anomalous. One reason behind setting the quantile to the value of 0.99, and not 1, is to consider the fact that the real-world training samples are affected by some factors, e.g., noise. Hence, setting the threshold to the maximum MQE generated by the training samples might adversely affect the anomaly detection rate [11].

The approach presented in [34] is the same as the traditional one, except that the threshold applied to the MQE values is set locally per neuron instead of globally. The threshold associated with a particular neuron is set to be the quantile of 0.99 of the MQEs corresponding to only the training samples for which that neuron was the BMU. Therefore, when a test sample is to be assessed, the MQE will be compared with the particular threshold associated with its BMU.

The approach in [36] differs from the traditional approach in the health indicator. First, for a given input sample, the K nearest neighbors (KNNs) among the map neurons are determined based on their corresponding quantization errors generated for that input sample. Next, the health indicator is calculated as the distance between the input sample and the centroid of the KNNs. Similarly to the traditional approach, the threshold is set globally, i.e., a quantile of 0.99 of the total generated health indicators by the training samples is set as the anomaly threshold. The aim of this approach is to tackle the false anomaly detection caused by the noisy training samples. Previous solutions suggested taking the health indicator as the average of all the quantization errors corresponding to all the map neurons. However, such solutions are not suitable when the map neurons are non-convex. Hence, the approach proposed in [36] aimed to present a compromise solution through considering only the KNNs to the input sample when calculating the health indicator. In this paper, the value of K is set to 3, as in [36].

In order to make the study of the different approaches more comprehensive, we also consider a hybrid approach of the two last above approaches. Thus, the health indicator is the same as in [36], i.e., the distance of the input sample from the centroid of the KNNs, whereas the anomaly threshold for that health indicator is set locally per neuron as in [34].

TABLE 1. The studied SOM-based anomaly detection approaches.

Approach's name[a]	Health indicator of the input sample	Th[b] type	Reference
MQE-G	MQE	Global	[11], [52]
MQE-L	MQE	Local	[34]
KNN-G	Distance from the KNN centroid	Global	[36]
KNN-L	Distance from the KNN centroid	Local	[34], [36]

[a]as used in this paper. [b]Threshold.

For simplicity, in the subsequent sections of this paper, the four approaches are respectively referred to as MQE-G, MQE-L, KNN-G, and KNN-L, where G and L refer to the global and local thresholds, respectively. It should be noted that for MQE-L; KNN-G; and KNN-L, after training SOM, only the neurons that became BMU at least once are considered in the processes involved in the corresponding approach. Table 1 summarizes the most important aspects of the four aforementioned approaches.

VI. IMPLEMENTATION DETAILS

A. EXPERIMENTAL DATA

The publicly available NASA Milling Data Set [30] is used in this paper. It represents several runs from a milling machine under various operating conditions. The operating conditions are chosen in accordance with industrial applicability and manufacturer's recommendations. They encompass two values of the depth of cut: 0.75mm and 1.5mm, two values of the feed rates: 0.25mm/rev and 0.5mm/rev, and two workpiece materials: cast iron and steel. A cutting speed of 826 rev/min is used for all the experiments. The combination of these operating conditions results in 8 different cases. However, the experiments with the same operating conditions are repeated a second time using another set of inserts, leading to a total of 16 cases in this data set. Each case includes many observations taken from many runs of the milling machine. Table 2 lists the different operating conditions and the number of runs for each of the 16 cases. The case indexed '6' is not considered in this paper as it only has one observation. All the remaining cases include run-to-failure data, where the milling process starts with a new tool and continues until the tool reaches its limit, and sometimes beyond. The observations under each case are provided with the machining time, the run

number, and 6 sensor signals with 9000 samples each. In addition, the majority of observations are provided with the flank wear (VB) that were measured with the help of a microscope after the corresponding runs. As in [53], we calculated the missing flank wear values through interpolation using the available values corresponding to the intended case. Recall that our work only deals with unlabeled data with respect to the exact tool state (See Section III), the flank wear values are only used in this paper to split the run-to-failure data under each case of the NASA Milling Data Set into normal and anomalous data. The tool wear can be divided into three stages [32]: the initial worn state, the normal worn state, and the severe worn state. In the last stage, the tool wear exceeds 0.3mm. Hence, in this paper, the tool is considered worn when its flank wear value exceeds 0.3mm. Otherwise, it is considered normal. As such, the total NASA Milling Data Set is divided into normal data and anomalous data.

TABLE 2. The main parameters of the NASA milling data set.

Case index	Depth of cut (mm)	Feed rate (mm/rev)	Workpiece material	Number of runs
1	1.5	0.5	Cast iron	17
2	0.75	0.5	Cast iron	14
3	0.75	0.25	Cast iron	14
4	1.5	0.25	Cast iron	7
5	1.5	0.5	Steel	6
6	1.5	0.25	Steel	1
7	0.75	0.25	Steel	8
8	0.75	0.5	Steel	6
9	1.5	0.5	Cast iron	9
10	1.5	0.25	Cast iron	10
11	0.75	0.25	Cast iron	23
12	0.75	0.5	Cast iron	15
13	0.75	0.25	Steel	15
14	0.75	0.5	Steel	9
15	1.5	0.25	Steel	7
16	1.5	0.5	Steel	6

The sensor signals include the AC spindle motor current, DC spindle motor current, table Acoustic Emission (AE), spindle AE, table vibration, and spindle vibration. Fig. 4 shows an example of the six sensor signals for one run. Excluding the DC spindle motor current, we consider only five sensor signals in this paper. In addition, out of the 9000 samples for each sensor signal, we only consider 4096 samples in the middle. This is to exclude the samples corresponding to the entry cut and exit cut, as in [3]. More details about this data set can be found in [30].

B. IMPLEMENTED SCENARIOS

In order to perform a comprehensive evaluation of the different anomaly detection approaches, we consider two extreme scenarios with respect to the diversity of the operating condition sets included in the individual models. The models in the first scenario include only one set of the operating conditions, whereas all the 8 operating conditions are included in the second-scenario models. More specifically,

1) In the first scenario, single-operating-condition models are built. The data under the cases of identical operating conditions, e.g., the cases {1,9} (see Table 2), are combined to constitute the data set for the corresponding model. Since the NASA Milling Data Set encompasses 8 different sets of operating conditions, 8 single-operating-condition models are built and tested for this scenario. The main aim here is to examine the stability of the approaches over the aforementioned 8 models.
2) In the second scenario, multi-operating-condition models are built and tested, with each of them encompassing 8 sets of operating conditions. Hence, the whole NASA Milling Data Set, except the data of case 6, is used for this scenario. For all the built models, we will study the effect of incorporating the numerical machining parameters, i.e., feed rate and depth of cut, to the selected sensory feature set. Moreover, our proposed feature selection scheme will be compared with PCA and LS in this scenario.

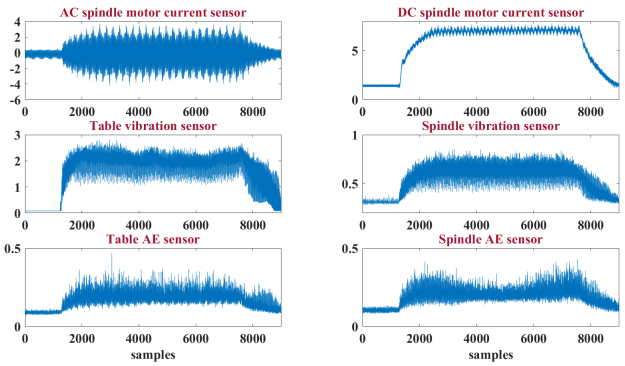


FIGURE 4. Sensor signals for one run of the NASA milling data set.

The hexagonal shape is chosen for the two-dimensional SOM grid and the batch algorithm is applied for learning, as it is recommended in [50]. For each validation run, the training and test samples are normalized based on the corresponding training samples to fall within the range [0,1]. The SOM grid size is also configured based on the training samples using (20) and (21). After the completion of each training run of SOM, the four anomaly detection approaches are applied in order to derive the corresponding anomaly indicators and thresholds. Finally, the corresponding test samples are fed to SOM and assessed as normal or anomalous.

For all the SOM models, the training data contain only normal data, whereas the test data include both normal and anomalous data. A 10-fold cross validation is implemented for each model. Similarly to [6], this procedure is repeated 10 times using 10 different partition seeds, leading to 100 validation runs for each evaluated SOM model. Hence, each result obtained in this work is the average of its corresponding values over the 100 runs.

The different anomaly detection approaches are evaluated on the basis of the following performance indicators: accuracy [45], anomaly detection rate (recall) [34], false alarm rate [34], and F1-score [45] which are expressed in (22)–(25), respectively. The precision indicator that is

needed for calculating F1-score is expressed in (26) [45].

$$\text{Accuracy} = \frac{\text{TP}+\text{TN}}{\text{TP}+\text{TN}+\text{FP}+\text{FN}} \tag{22}$$

$$\text{Anomaly Detection Rate (Recall)} = \frac{\text{TP}}{\text{TP}+\text{FN}} \tag{23}$$

$$\text{False Alarm Rate} = \frac{\text{FP}}{\text{TP}+\text{FP}} \tag{24}$$

$$\text{F1-score} = 2 \times \frac{\text{Recall} \times \text{Precision}}{\text{Recall}+\text{Precision}} \tag{25}$$

$$\text{Precision} = \frac{\text{TP}}{\text{TP}+\text{FP}} \tag{26}$$

where TP (True Positive) and TN (True Negative) represent the test samples that are correctly assessed as anomalous and normal, respectively. FP (False Positive) and FN (False Negative) represent the test samples that are wrongly assessed as anomalous and normal, respectively.

C. THE CONDUCTED FEATURE EXTRACTION/SELECTION PROCESSES

The sensor signals recorded during each milling run are divided into non-overlapping segments of 1024 samples. Therefore, each observation will consist of 4 segments; without including the entry and exit cut. The sensory features are then extracted from each segment. There exist a variety of feature extraction methods in the literature, among which wavelet transform emerged as one of the most powerful feature extraction methods for the predictive maintenance. This is due to the fact that the wavelets are localized in both the time and frequency domains, which allows for extracting distinctive features from the non-stationary sensor signals. In this paper, three different signal analysis methods are performed for feature extraction, namely:

1) Time-domain statistical analysis in which the following seven features are extracted from each segment: mean, variance, skewness, kurtosis, impulse factor, crest factor, and Root Mean Square (RMS). The equations for these features can be found in [7].
2) Multi-resolution analysis using a non-decimated discrete wavelet transform. It is a common practice in the literature to extract statistical features from the discrete wavelet transform coefficients [32]. In our work, a 6-level decomposition of each segment is performed using each of the following mother wavelet functions: db1, db2, and db3. Each decomposition results in 6 details, D1, D2, …, D6; and one approximation, A6. Each of the resultant approximations and details is used to extract the following four statistical features: mean, variance, skewness, and kurtosis.
3) Time-frequency analysis using continuous wavelet transform. The bump wavelet is used for this transform. The resultant scalogram for each segment is used to compute the mean peak frequency. First, the peak frequency is determined for each time instance, where the peak frequency is the frequency with the maximum power. Next, the mean peak frequency is calculated as the average of the peak frequencies determined at all the time instances.

Table 3 shows the number of extracted features per sensor for each of the feature extraction methods. Since the same features are extracted from the five sensors, a total of 460 (92 × 5) sensory features are extracted.

TABLE 3. Number of extracted features per sensor.

Feature extraction method	Number of features (per sensor)
Time-domain statistical analysis	7
Discrete wavelet transform (db1/db2/db3)	28/28/28
Continuous wavelet transform	1
Sum of the extracted features per sensor	92

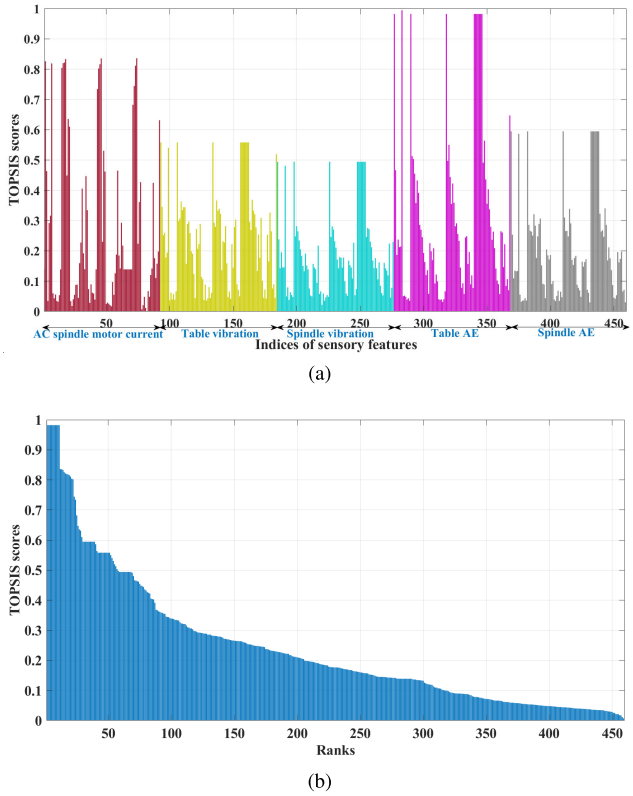


FIGURE 5. TOPSIS scores of all the extracted sensory features. (a) Feature scores for the individual sensors. (b) Ranked scores.

The extracted features undergo the processes involved in our feature selection scheme (see Section IV). The two criteria of each feature are calculated based on (4) and (5). The TOPSIS is then implemented to rank the features. It is worth inspecting the scores of the total features extracted from all the sensors before eliminating any feature or sensor. Fig. 5 shows the TOPSIS scores of the total features extracted from the five sensors. As it is depicted in Fig. 5a, among all the sensors, the table AE sensor and the AC spindle motor current sensor have the highest scored features. In addition, for all the sensors, the low scored features tend to be more

than the high scored features. This is to be expected since it would be difficult to maintain a good discrimination ability over the 8 operating conditions considered in the evaluation criteria of the features. Fig. 5b illustrates the ranked TOPSIS scores. It can be noticed that some features have identical scores. Such features are most likely to be strongly correlated. This explains the importance of the third step in our feature selection scheme in which redundant features are eliminated.

Deploying five sensors on the machine center could not be practical in real-world implementations, especially when the sensors are mounted near the machining area. Since finding the optimal multi-sensor configuration is out of the scope of this paper, we will only select two sensors to enable the sensor fusion. The AE table sensor and the AC spindle motor current sensor are selected since, among all the other sensors, they have the features with the highest TOPSIS scores. Hence, these two sensors are the only sensors that will be considered in the rest of the paper.

Concerning the feature selection in the first scenario that corresponds to the single-operating-condition models, the feature selection will be performed separately for each model. It should be noted that the second criterion in our proposed scheme, i.e., the criterion related to the inter-case SOF, is not relevant in this scenario since the data of a specific model share the same operating conditions. This is equivalent to setting the criteria weights to $w = [1, 0]$ (see Section IV-B). When calculating the first criterion through (4), only the cases that share the intended set of operating conditions are included. As for the feature selection in the second scenario that corresponds to the multi-operating-condition models, our proposed feature selection scheme is implemented exactly as described in Section IV. In both scenarios, the number of the selected sensory features for each model is set to 5. Table 4 lists the selected five features based on our proposed feature selection scheme for the multi-operating-condition models.

TABLE 4. The features selected by our scheme for the multi-operating-condition models.

Feature rank[a]	Sensor signal	Selected feature
1	Table AE	RMS
2	AC spindle motor current	Variance of D4 (db3)
3	Table AE	Mean of the peak frequency
4	AC spindle motor current	Mean of the peak frequency
5	AC spindle motor current	Variance of A6 (db1)

[a]After eliminating redundant features.

D. EXPERIMENTAL RESULTS AND DISCUSSION

This section starts with the results obtained in the first implementation scenario, and then proceeds with the second scenario where the multi-operating-condition models are applied and our proposed feature selection scheme is compared with PCA and LS.

1) THE FIRST-SCENARIO RESULTS AND DISCUSSION

As mentioned previously, 8 single-operating-condition SOM models are built for the first scenario. Table 5 lists all the performance indicators of the studied anomaly detection approaches for these models, whereas Fig. 6 only shows the accuracies of the approaches. As illustrated in Fig. 6, the approaches with global thresholds, MQE-G and KNN-G, show a less stable accuracy compared to MQE-L and KNN-L. In particular, the accuracy of KNN-G fluctuates considerably over the individual models. It reached as low as 13.3% (for the cases{7,13}) and as high as about 97% (for the three case sets{4,10}, {5,16}, {15}). Clearly, obtaining these two extreme accuracies from the same approach emphasizes the significance of evaluating the TCM models on different operating conditions. The reason for this fluctuation could be attributed to the fact that the health indicator used in KNN-G is based on the 3 nearest neurons to the input sample, which makes the performance of this approach highly sensitive to the training data structure which is in turn reflected by the map neuron distribution in the feature space. If the 3 neurons are close enough to each other, their centroid would still lie in the feature space representing the normal data. Otherwise, the centroid could fall in an abnormal area. The more scattered the clusters of the training samples are, the more sever this effect would be, which in turn leads to a higher global threshold for the health indicator, and thus, a higher probability that an anomalous sample would be wrongly assessed as normal. This can be verified in Table 5 where the anomaly detection rate achieved by KNN-G is considerably lower than those achieved by the other three approaches for the same aforementioned cases. As for KNN-L, since the thresholds are set separately for the map neurons, the influence of the data structure will be limited to the training samples lying in the vicinity of the intended neuron. This explains why the performance of KNN-L is substantially more stable than the KNN-G even though they both use the same health indicator. The influence of the data structure is reduced even more for MQE-L as it is not only based on local thresholds but also the health indicator, i.e., MQE, is based only on one neuron, meaning that the only relevant training samples are those mapped to the intended neuron.

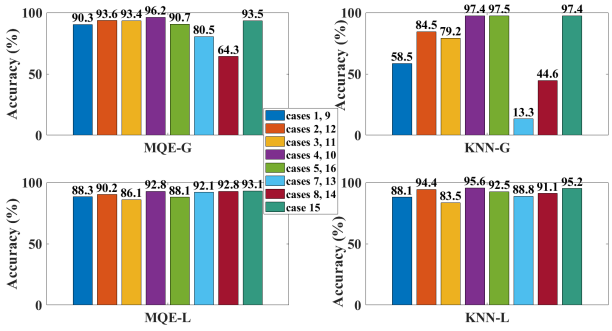


FIGURE 6. The accuracies of the studied SOM-based anomaly detection approaches for the single-operating-condition models.

The improved accuracy that can be attained when using local thresholds instead of global ones can be vividly seen for the case sets {7,13},{8,14}. The accuracies achieved by

TABLE 5. Performance indicators of the studied SOM-based approaches for the single-operating-condition models.

Anomaly detection approach	Index of the experimental cases															
	1 & 9	2 & 12	3 & 11	4 & 10	5 & 16	7 & 13	8 & 14	15	1 & 9	2 & 12	3 & 11	4 & 10	5 & 16	7 & 13	8 & 14	15
	Accuracy (%)								F1-score (%)							
MQE-G	**90.3**	93.6	**93.4**	96.2	90.7	80.5	64.3	93.5	**94.4**	96.4	**96.1**	98.0	95.0	87.5	75.3	96.6
MQE-L	88.3	90.2	86.1	92.8	88.1	**92.1**	**92.8**	93.1	93.7	94.7	92.3	96.2	93.6	**95.9**	**96.2**	96.4
KNN-G	58.5	84.5	79.2	**97.4**	**97.5**	13.3	44.6	**97.4**	68.1	90.3	85.9	**98.6**	**98.6**	12.9	57.0	**98.6**
KNN-L	88.1	**94.4**	83.5	95.6	92.5	88.8	91.1	95.2	93.3	**96.9**	90.0	97.7	95.9	93.9	95.1	97.5
	Anomaly detection rate (%)								False alarm rate (%)							
MQE-G	93.6	96.2	96.1	**100**	**100**	83.9	62.3	**100**	4.5	3.3	3.8	3.9	9.4	**5.1**	1.5	6.5
MQE-L	**98.3**	**99.8**	**98.2**	**100**	**100**	**98.9**	**99.5**	**100**	10.5	9.8	12.9	7.2	12.0	7.0	6.8	6.9
KNN-G	53.3	83.1	75.6	98.3	**100**	07.0	40.1	**100**	**0.5**	**0.7**	**0.3**	**1.1**	**2.7**	5.2	**0.8**	**2.7**
KNN-L	93.5	99.7	88.2	**100**	**100**	93.5	94.4	**100**	6.8	5.7	8.1	4.5	7.7	5.5	4.1	4.9

Bold values indicate the best values with respect to the corresponding performance indicator

MQE-L for these case sets are 92.1% and 92.8% respectively, which correspond to an outperformance of 11.6% and 28.5% over MQE-G which achieved 80.5% and 64.3%. Similarly, the accuracies achieved by KNN-L for these case sets are 88.8% and 91.1% respectively, which are by 75.5% and 46.5% higher than those achieved with KNN-G (the accuracies for KNN-G are 13.3% and 44.6%). The difference in the anomaly detection rate is the reason behind these considerable differences in the accuracy between the aforementioned counterpart approaches, as it can be seen in Table 5.

On the other hand, for some cases, e.g., {5,16}, the approaches with global thresholds achieved a higher accuracy than the approaches with local thresholds, primarily due to an outperformance in the false alarm rate indicator. These cases are likely to correspond to when the distribution of the SOM neurons in the feature space is relatively even, i.e., their density in the different areas are relatively similar. In such cases, the availability of more data samples when the threshold is set globally compared to when it is set locally helps to achieve a higher precision in anomaly detection. However, when a better accuracy is achieved by the global thresholds, the maximum difference compared to that achieved by the local thresholds reached 7.3% for the MQE approaches (for the cases {3,11}), and 5% for the KNN approaches (for the cases {5,16}). These differences can be neglected when compared to the cases in which the local thresholds outperformed the global ones with differences of up to 28.5% for the MQE-based approaches and up to 75.5% for the KNN-based approaches, as mentioned previously.

It can be seen that for the case sets {4,10}, {5,16}, and {15}, the accuracies achieved by KNN-G and KNN-L outperform those achieved by MQE-G and MQE-L, respectively. In general, if the three neighbors in the KNN approaches occupy a relatively compact area in the feature space, incorporating the 3 neighbors in calculating the health indicator would reduce the effect of noise, and thus, would lead to a higher precision in detecting anomalies compared to the MQE approaches whose health indicator is based only on the distance to the BMU. This can be verified in Table 5 where the anomaly detection rates are nearly identical for all the approaches, while the false alarm rates of the KNN approaches are noticeably lower than those of the MQE approaches.

TABLE 6. Statistical summary of the accuracies achieved by the approaches for the single-operating-condition models.

Approach	Maximum (%)	Minimum (%)	Range (%)	Average (%)	Standard deviation (%)
MQE-G	96.2	64.3	31.9	87.8	10.6
MQE-L	93.1	86.1	06.9	90.4	02.7
KNN-G	97.5	13.3	84.3	71.5	30.5
KNN-L	95.6	83.5	12.2	91.1	04.2

Table 6 lists a statistical summary of the accuracies achieved by the studied anomaly detection approaches in the first scenario. The conclusions of the first-scenario evaluation can be summarized as follows.

- MQE-L and KNN-L achieved a more stable performance compared to MQE-G and KNN-G, owing to being less affected by the overall structure of the training data. Moreover, they were able to yield an F1-score of over 90% for all the individual experimental cases, as shown in Table 5. In addition, their average accuracy over the eight first-scenario models are over 90% (90.4% for MQE-L, and 91.1% for KNN-L), as shown in Table 6.
- The stability of the different approaches with respect to the different operating conditions can be evaluated on the basis of the range and standard deviation of the accuracy values achieved over the different single-operating-condition models (see Table 6); the more stable the approach is, the smaller the accuracy range and standard deviation will be. As such, the approaches can be ranked with respect to their performance stability as follows (from the best to the worst): MQE-L, KNN-L, MQE-G, and KNN-G. We consider the stability a key factor to evaluate the approaches, since it reflects the ability of the approach to yield a consistent TCM regardless of which operating conditions the data come from.

2) THE SECOND-SCENARIO RESULTS AND DISCUSSION

As mentioned previously, multi-operating-condition models are built in this scenario. We start with the results obtained

TABLE 7. Performance indicators of the studied SOM-based approaches for the multi-operating-condition models with our proposed feature selection scheme.

Anomaly detection approach	Without adding DoC and feed rate to the feature set				With adding DoC and feed rate to the feature set			
	Accuracy (%)	Anomaly detection rate (%)	False alarm rate (%)	F1-score (%)	Accuracy (%)	Anomaly detection rate (%)	False alarm rate (%)	F1-score (%)
MQE-G	68.8	65.9	1.2	78.6	79.9	78.2	0.8	87.3
MQE-L	**89.4**	**94.0**	5.9	**94.0**	**94.4**	**98.2**	4.4	**96.9**
KNN-G	33.4	25.6	**0.8**	40.6	62.1	57.8	**0.4**	72.9
KNN- L	73.1	73.0	4.0	82.9	89.8	91.1	2.6	94.1

Bold values indicate the best values with respect to the corresponding performance indicator

with our feature selection scheme proposed in this paper, and then a comparison with PCA and LS will be performed.

a: RESULTS WITH OUR PROPOSED SCHEME

First, only the five sensory features selected by our proposed scheme (see Table 4) are fed to the SOM models. The average accuracy achieved by a specific approach in the first scenario will be used as a baseline for comparison with the accuracy achieved in the second scenario. Such a comparison helps to evaluate the robustness of the individual approaches when dealing with complex data structures. Next, the numerical machining parameters in the NASA Milling Data Set, i.e., feed rate and Depth of Cut (DoC), are incorporated into the selected sensory features, constituting a 7-feature input sample to the SOM models.

Table 7 shows the different performance indicators of the anomaly detection approaches. Starting with the case without including the machining parameters, it can be seen that the accuracy achieved by MQE-G in this scenario (68.8%) is considerably lower than the average accuracy achieved in the first scenario (87.8%), i.e., there is a degradation of 19%. Similarly, the accuracy of KNN-L (73.1%) is lower than the average accuracy achieved in the first scenario (91.1%), i.e., a 18% degradation. As for KNN-G, an accuracy of 33.4% is achieved against 71.5% in the first scenario (a 38.1% degradation). On the other hand, MQE-L achieved an accuracy of 89.4% which is nearly the same as the average accuracy achieved in the first scenario (90.4%), i.e., the degradation is only 1%. Among the four approaches, MQE-L achieved the highest accuracy for the multi-operating-condition model.

Since the SOM neurons'weights follow the density distribution of the training samples, the patterns that appear more frequently in the input space will be represented by a greater number of map neurons compared to the input patterns that appear less frequently. Hence, for all the approaches, lower values of the health indicators are expected in the areas where the neurons are denser, and vice versa. Such irregularities affect the tool assessment when the thresholds are set globally. This explains why the accuracy of MQE-G and KNN-G is more prone to deterioration in general, and in particular, when the training data belong to different operating conditions. It can be vividly seen that MQE-L is the only approach that succeeded in maintaining nearly the same level of accuracy as in the first scenario, which reflects the robustness of this approach. These results reinforce that MQE-L is the least vulnerable approach to the influence of the overall structure of the training data, as it was also seen in the first scenario. This aspect is particularly useful to achieve a reliable TCM.

As it is illustrated in Table 7, adding the machining parameters to the input feature sample led to a noticeable performance improvement for all the four approaches. This is to be expected since adding these values would contribute to reducing the overlapping between the normal and anomalous multi-operating-condition data. It can be seen that this addition has simultaneously enhanced both the anomaly detection rate and false alarm rate for all the approaches. An accuracy of 94.4% was achieved by MQE-L.

b: RESULTS OF COMPARING OUR SCHEME WITH PCA AND LS

As mentioned previously, our proposed feature selection scheme will be compared with both PCA and LS in the second scenario.

PCA is one of the most commonly used feature reduction methods [7], [39], [48], [54]. It is based on generating a new low-dimensional reference system that contains the directions with the largest variance of the original data. This system is formed by the so-called principal components, where each principal component is a linear combination of the original variables [39]. Using PCA in conjunction with SOM was already used in many researches, e.g., [54]. In our implementation, the first five principal components are selected, hence the 184 features extracted from the two selected sensors will be projected onto these components, and the resultant scores will constitute a 5-feature input sample to SOM.

Laplacian score (LS) for feature selection was proposed in [38]. It is one of the most popular unsupervised feature selection methods, and has been adopted in a multitude of research works, e.g., [55], [56]. LS reflects the ability of a feature to preserve locality. The basic idea behind it is that closely spaced data samples are likely to represent the same state. The LS algorithm can be summarized as follows. First, the local data structure is modeled through a nearest neighbor graph. Each node in this graph represents one data sample constituted by the values of all the candidate features corresponding to that sample. Two nodes are considered connected if one of them is among the K nearest neighbors of the other. Moreover, the inter-node connections are weighted

based on a similarity measure. Next, LS will be computed for each individual feature, where LS reflects the extent to which each feature is able to preserve the constructed graph, and hence the features can be ranked based on their LS. More details about this method and the equations involved can be found in [38]. In our implementation, the number of K nearest neighbors in the nearest neighbor graph is taken to be 5 as in [38]. It should be noted that LS method does not take into account redundancy among features. In order to make a more relevant comparison between our scheme and LS method, we extended the LS method by adding a final step in order to account for redundant features. In the extended LS, after ranking the features based on their importance as specified by the Laplacian scores, the ranked features will undergo the procedure we implemented in our scheme to eliminate redundant features (see Section IV-C), and then the top 5 features will be selected from the remaining features. Table 8 and Table 9 list the 5 most important features selected based on LS and extended LS, respectively. Comparing the two tables, it can be seen that there exist two redundant features, namely the features ranked second and fifth, among the 5 features selected based on LS.

TABLE 8. The features selected by LS method for the multi-operating-condition models.

Feature rank	Sensor signal	Selected feature
1	AC spindle motor current	Kurtosis of D5 (db2)
2	AC spindle motor current	Kurtosis of D5 (db1)
3	Table AE	Variance of A6 (db1)
4	Table AE	Mean of the peak frequency
5	Table AE	Variance of A6 (db2)

TABLE 9. The features selected by the extended LS method for the multi-operating-condition models.

Feature rank[a]	Sensor signal	Selected feature
1	AC spindle motor current	Kurtosis of D5 (db2)
2	Table AE	Variance of A6 (db1)
3	Table AE	Mean of the peak frequency
4	AC spindle motor current	Variance of D2 (db2)
5	AC spindle motor current	Kurtosis of D1 (db2)

[a]After eliminating redundant features.

Fig. 7 shows the accuracies of the different approaches in the second scenario for our proposed feature selection scheme, PCA, LS, and extended LS (with and without adding the machining parameters). As it is observed for all the experiments in this figure, our proposed feature selection scheme outperformed the other methods. This is attributed to the fact that these methods are based on underlying assumptions that are not consistent with the multi-operating-condition data whose variation stems from both the operating conditions and the tool wear progression. More specifically, PCA is based on maximizing the data variance. However, the directions of the maximum variance of the data could not be linked to the tool state. Hence, it is possible to generate similar PCA scores

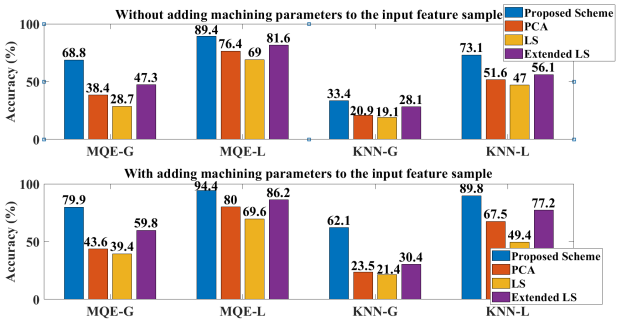


FIGURE 7. The accuracies of the studied SOM-based anomaly detection approaches for the multi-operating-condition models with our proposed feature selection scheme, PCA, LS, and extended LS (with/without adding machining parameters to the input feature sample).

by data belonging to different tool states, especially with the increased overlapping between normal and anomalous data under the existence of different operating conditions. In addition, LS-based methods select the features with the highest locality preserving power, assuming that similar data points belong to the same state. However, as mentioned previously, with the combined effect of both the tool degradation and operating conditions, similar data patterns could be generated even under different tool states. On the other hand, our proposed scheme selects the features with the highest discriminability concerning the data of the same/different operating conditions. As for the difference between LS and extended LS, it can be noticed that the extended LS considerably outperformed the LS. This emphasizes the importance of the added step that aims to eliminate redundant features. Regarding when no machining parameters are added to the input feature sample, the maximum accuracy gain achieved by our scheme over PCA, LS, and extended LS reached 30.4% (for MQE-G), 40.1% (for MQE-G), and 21.5% (for MQE-G), respectively. With the addition of machining parameters, the accuracy gain over PCA, LS, and extended LS reached 38.6% (for KNN-G), 40.7% (for KNN-G), and 31.7% (for KNN-G), respectively.

The second-scenario evaluation can be summarized as follows.

- With our proposed feature selection scheme and 5-feature input sample, the MQE-L was the only approach that succeeded in maintaining the same average accuracy it achieved over the single-operating-condition models (about 90%). Contrarily, the performance of all the other approaches considerably suffered with the multi-operating-condition data. Based on the difference between the accuracy achieved in the second scenario and the average first-scenario accuracy, the approaches can be ranked as follows (from the best to the worst): MQE-L, KNN-L, MQE-G, and KNN-G.
- For all the four approaches, our proposed feature selection scheme outperformed PCA, LS, and extended LS.
- Adding the machining parameters to the input feature sample could substantially improve the assessment accuracy achieved by the multi-operating-condition

models, which can be attributed to the decreased overlapping between the normal and anomalous data.

It should be noted that the performance of our proposed feature selection scheme can be enhanced with selecting more than 5 sensory features. However, keeping the number of the selected features constant for both the single-operating-condition models and multi-operating-condition models allowed us to link the results obtained in the latter models to the robustness of our scheme and the anomaly detection approaches. In addition, the performance of all the studied approaches can be optimized by calibrating the corresponding quantile values. However, using the quantile of 0.99 would reveal the inherent performance tendency of the different approaches. In addition, it represents the most commonly used quantile value in the literature.

VII. CONCLUSION AND FUTURE WORK

This work addressed the anomaly detection, modeled as a one-class classification, of milling tools under different operating conditions. A feature selection scheme for multi-operating-condition models was proposed in this paper. In this scheme, the SOF metric was used in two forms to reflect the discriminability of a given feature regarding the different data patterns belonging to the same/different operating conditions. In addition, TOPSIS was used to rank the features, and redundant features were eliminated. SOM was employed to model the normal data, and four SOM-based anomaly detection approaches are studied. The NASA Milling Data Set was used for the evaluation experiments. Two extreme evaluation scenarios with respect to the diversity of the operating conditions were performed. First, the anomaly detection approaches were applied on 8 single-operating-condition models, with the aim of examining their stability over these different models. Second, they are applied on multi-operating-condition models (each one incorporating 8 different sets of operating conditions), with the aim of examining their robustness when dealing with complex data structures.

Feature extraction from 5 sensor signals was first performed based on the time-domain statistical analysis as well as the continuous/discrete wavelet transforms. TOPSIS scores of the total features were utilized as a basis to choose only the two sensors that have the highest ranked features. The selected sensors are: the table AE sensor and AC spindle current sensor.

The overall results showed that the studied SOM-based anomaly detection approaches can be ranked in terms of the stability and robustness as follows (from the best to the worst): MQE-L, KNN-L, MQE-G, and KNN-G. This ranking indicates that the threshold type (global or local) is the most significant factor that affects the stability and robustness of the SOM-based approaches studied in this paper. The global thresholds render the approach more susceptible to the influence of the overall data structure. In addition, incorporating more than one neuron in calculating the health indicator, as in the KNN-based approach, has also shown to increase the sensitivity of the approach to the data structure. Further, the accuracy fluctuation of some approaches over the different models emphasizes that achieving a high diagnosis accuracy is not the only important indicator to be sought after when approaching TCM solutions. The stability and robustness of the TCM models are just as important to ensure a reliable TCM in the real-world settings.

Our proposed feature selection scheme was compared with PCA, LS, extended LS (in which we added a step to the original LS to eliminate redundant features). The results showed that higher accuracies were achieved by all the approaches with our scheme, with differences of up to 38.6%, 40.7% and 31.7% as compared to PCA, LS, and extended LS, respectively. These results show the feasibility of our feature selection scheme for multi-operating-condition models used for anomaly detection, and highlight the fact that the existing feature reduction/selection techniques might not always be powerful in regards to the complex industrial tasks, owing to their dependency on specific criteria that might not be consistent with the task at hand. For such tasks, utilizing many existing and/or task-tailored criteria to evaluate the sensory features might be essential; not only to seek the features that allow for a high diagnosis accuracy, but also to gain more insights about the different sensors based on their signals, as it was performed in this paper. To this end, TOPSIS can play a vital role in ranking the features through a simple yet powerful procedure.

For the future work, the stability and robustness of other SOM-based TCM solutions that rely on novel health indicators, e.g., the health indicator proposed in [26], will also be investigated under different operating conditions. In addition, applying the non-convex archetypal analysis proposed in [13] on top of a SOM model is also intriguing. Further, we consider applying TOPSIS-assisted feature selection on labeled run-to-failure data for multi-class classification tasks, and investigating subjective and objective weight determination methods for TOPSIS with these tasks.

ACKNOWLEDGMENT

Responsible for the content of this publication are the authors.

REFERENCES

[1] Y. Zhou and W. Xue, "Review of tool condition monitoring methods in milling processes," *Int. J. Adv. Manuf. Technol.*, vol. 96, nos. 5–8, pp. 2509–2523, May 2018.
[2] K. Zhu, G. Li, and Y. Zhang, "Big data oriented smart tool condition monitoring system," *IEEE Trans. Ind. Informat.*, vol. 16, no. 6, pp. 4007–4016, Jun. 2020.
[3] S. Das, R. Hall, S. Herzog, G. Harrison, M. Bodkin, and L. Martin, "Essential steps in prognostic health management," in *Proc. IEEE Conf. Prognostics Health Manage.*, Denver, CO, USA, Jun. 2011, pp. 1–9.
[4] C. Cooper, J. Zhang, R. X. Gao, P. Wang, and I. Ragai, "Anomaly detection in milling tools using acoustic signals and generative adversarial networks," *Procedia Manuf.*, vol. 48, pp. 372–378, Jan. 2020.
[5] G. Serin, B. Sener, A. M. Ozbayoglu, and H. O. Unver, "Review of tool condition monitoring in machining and opportunities for deep learning," *Int. J. Adv. Manuf. Technol.*, vol. 109, pp. 1–22, Jul. 2020.
[6] S. Binsaeid, S. Asfour, S. Cho, and A. Onar, "Machine ensemble approach for simultaneous detection of transient and gradual abnormalities in end milling using multisensor fusion," *J. Mater. Process. Technol.*, vol. 209, no. 10, pp. 4728–4738, Jun. 2009.

[7] W. J. Lee, G. P. Mendis, and J. W. Sutherland, "Development of an intelligent tool condition monitoring system to identify manufacturing tradeoffs and optimal machining conditions," *Procedia Manuf.*, vol. 33, pp. 256–263, Jan. 2019.

[8] L. C. Brito, M. B. da Silva, and M. A. V. Duarte, "Identification of cutting tool wear condition in turning using self-organizing map trained with imbalanced data," *J. Intell. Manuf.*, vol. 32, no. 1, pp. 127–140, Jan. 2021.

[9] K. Yan, Z. Ji, H. Lu, J. Huang, W. Shen, and Y. Xue, "Fast and accurate classification of time series data using extended ELM: Application in fault diagnosis of air handling units," *IEEE Trans. Syst., Man, Cybern., Syst.*, vol. 49, no. 7, pp. 1349–1356, Jul. 2019.

[10] G. Li, Y. Fu, D. Chen, L. Shi, and J. Zhou, "Deep anomaly detection for CNC machine cutting tool using spindle current signals," *Sensors*, vol. 20, no. 17, p. 4896, Aug. 2020.

[11] A. von Birgelen, D. Buratti, J. Mager, and O. Niggemann, "Self-organizing maps for anomaly localization and predictive maintenance in cyber-physical production systems," *Procedia CIRP*, vol. 72, pp. 480–485, Jan. 2018.

[12] J. Wu, Z. Zhao, C. Sun, R. Yan, and X. Chen, "Fault-attention generative probabilistic adversarial autoencoder for machine anomaly detection," *IEEE Trans. Ind. Informat.*, vol. 16, no. 12, pp. 7479–7488, Dec. 2020.

[13] P. Li and O. Niggemann, "A nonconvex archetypal analysis for one-class classification based anomaly detection in cyber-physical systems," *IEEE Trans. Ind. Informat.*, vol. 17, no. 9, pp. 6429–6437, Sep. 2021.

[14] K. Sarda, A. Acernese, V. Nole, L. Manfredi, L. Greco, L. Glielmo, and C. D. Vecchio, "A multi-step anomaly detection strategy based on robust distances for the steel industry," *IEEE Access*, vol. 9, pp. 53827–53837, 2021.

[15] M. Hu, X. Feng, Z. Ji, K. Yan, and S. Zhou, "A novel computational approach for discord search with local recurrence rates in multivariate time series," *Inf. Sci.*, vol. 477, pp. 220–233, Mar. 2019.

[16] G. Aydemir and B. Acar, "Anomaly monitoring improves remaining useful life estimation of industrial machinery," *J. Manuf. Syst.*, vol. 56, pp. 463–469, Jul. 2020.

[17] T. Liu and K. Zhu, "A switching hidden semi-Markov model for degradation process and its application to time-varying tool wear monitoring," *IEEE Trans. Ind. Informat.*, vol. 17, no. 4, pp. 2621–2631, Apr. 2021.

[18] Z. Chen and W. Li, "Multisensor feature fusion for bearing fault diagnosis using sparse autoencoder and deep belief network," *IEEE Trans. Instrum. Meas.*, vol. 66, no. 7, pp. 1693–1702, Jul. 2017.

[19] K. Song, M. Wang, L. Liu, C. Wang, T. Zan, and B. Yang, "Intelligent recognition of milling cutter wear state with cutting parameter independence based on deep learning of spindle current clutter signal," *Int. J. Adv. Manuf. Technol.*, vol. 109, nos. 3–4, pp. 929–942, Jul. 2020.

[20] H. Chen, H. Ma, X. Chu, and D. Xue, "Anomaly detection and critical attributes identification for products with multiple operating conditions based on isolation forest," *Adv. Eng. Informat.*, vol. 46, Oct. 2020, Art. no. 101139.

[21] J. Yu, "A hybrid feature selection scheme and self-organizing map model for machine health assessment," *Appl. Soft Comput.*, vol. 11, no. 5, pp. 4041–4054, Jul. 2011.

[22] J. Vesanto and E. Alhoniemi, "Clustering of the self-organizing map," *IEEE Trans. Neural Netw.*, vol. 11, no. 3, pp. 586–600, May 2000.

[23] M. Rigamonti, P. Baraldi, E. Zio, A. Alessi, D. Astigarraga, and A. Galarza, "Identification of the degradation state for condition-based maintenance of insulated gate bipolar transistors: A self-organizing map approach," *Microelectron. Rel.*, vol. 60, pp. 48–61, May 2016.

[24] C. S. Wickramasinghe, K. Amarasinghe, and M. Manic, "Deep self-organizing maps for unsupervised image classification," *IEEE Trans. Ind. Informat.*, vol. 15, no. 11, pp. 5837–5845, Nov. 2019.

[25] S. Schwartz, J. J. Montero Jimenez, M. Salaün, and R. Vingerhoeds, "A fault mode identification methodology based on self-organizing map," *Neural Comput. Appl.*, vol. 32, no. 17, pp. 13405–13423, Sep. 2020.

[26] A. Rai and J.-M. Kim, "A novel health indicator based on the Lyapunov exponent, a probabilistic self-organizing map, and the Gini-Simpson index for calculating the RUL of bearings," *Measurement*, vol. 164, Nov. 2020, Art. no. 108002.

[27] M. Behzadian, S. K. Otaghsara, M. Yazdani, and J. Ignatius, "A state-of the-art survey of TOPSIS applications," *Expert Syst. Appl.*, vol. 39, no. 17, pp. 13051–13069, Dec. 2012.

[28] M. A. Senouci, M. S. Mushtaq, S. Hoceini, and A. Mellouk, "TOPSIS-based dynamic approach for mobile network interface selection," *Comput. Netw.*, vol. 107, pp. 304–314, Oct. 2016.

[29] A. Shukla, P. Agarwal, R. S. Rana, and R. Purohit, "Applications of TOPSIS algorithm on various manufacturing processes: A review," *Mater. Today, Proc.*, vol. 4, no. 4, pp. 5320–5329, 2017.

[30] A. Agogino and K. Goebel. (2007). *Milling Data set*. BEST lab, UC Berkeley. NASA Ames Prognostics Data Repository. Moffett Field, CA, USA. [Online]. Available: https://ti.arc.nasa.gov/project/prognostic-data-repository

[31] H. Tang, Z. Liao, P. Chen, D. Zuo, and S. Yi, "A novel convolutional neural network for low-speed structural fault diagnosis under different operating condition and its understanding via visualization," *IEEE Trans. Instrum. Meas.*, vol. 70, pp. 1–11, Aug. 2021, Art. no. 3501611, doi: 10.1109/TIM.2020.3016752.

[32] D. Gao, Z. Liao, Z. Lv, and Y. Lu, "Multi-scale statistical signal processing of cutting force in cutting tool condition monitoring," *Int. J. Adv. Manuf. Technol.*, vol. 80, nos. 9–12, pp. 1843–1853, Oct. 2015.

[33] V. Bălănică, L. Liao, H. Claussen, and J. Rosca, "A multi-model approach for anomaly detection and diagnosis using vibration signals," in *Proc. IEEE Conf. Prognostics Health Manage. (PHM)*, Gaithersburg, MD, USA, Jun. 2013, pp. 1–7.

[34] A. Bellas, C. Bouveyron, M. Cottrell, and J. Lacaille, "Anomaly detection based on confidence intervals using SOM with an application to health monitoring," in *Advances in Self-Organizing Maps and Learning Vector Quantization*. Cham, Switzerland: Springer, 2014, pp. 145–155, doi: 10.1007/978-3-319-07695-9_14.

[35] M. Nouri, B. K. Fussell, B. L. Ziniti, and E. Linder, "Real-time tool wear monitoring in milling using a cutting condition independent method," *Int. J. Mach. Tools Manuf.*, vol. 89, pp. 1–13, Feb. 2015.

[36] J. Tian, M. H. Azarian, and M. Pecht, "Anomaly detection using self-organizing maps-based K-nearest neighbor algorithm," in *Proc. Eur. Conf. Prognostics Health Manage. Soc.*, 2014, pp. 1–9.

[37] B. Guo, L. Song, T. Zheng, H. Liang, and H. Wang, "A comparative evaluation of SOM-based anomaly detection methods for multivariate data," in *Proc. Prognostics Syst. Health Manage. Conf. (PHM-Qingdao)*, Qingdao, China, Oct. 2019, pp. 1–6.

[38] X. He, D. Cai, and P. Niyogi, "Laplacian score for feature selection," in *Proc. Adv. Neural Inf. Process. Syst.*, vol. 18, 2005, pp. 507–514.

[39] G. Z. Yang and G. Yang, *Body Sensor Networks*, vol. 1. London, U.K.: Springer, 2006.

[40] C. Scheffer and P. S. Heyns, "An industrial tool wear monitoring system for interrupted turning," *Mech. Syst. Signal Process.*, vol. 18, no. 5, pp. 1219–1242, Sep. 2004.

[41] C. S. Dhanalakshmi, P. Madhu, A. Karthick, M. Mathew, and R. V. Kumar, "A comprehensive MCDM-based approach using TOPSIS and EDAS as an auxiliary tool for pyrolysis material selection and its application," *Biomass Convers. Biorefinery*, vol. 186, pp. 1–16, Sep. 2020.

[42] L. Anojkumar, M. Ilangkumaran, and V. Sasirekha, "Comparative analysis of MCDM methods for pipe material selection in sugar industry," *Expert Syst. Appl.*, vol. 41, no. 6, pp. 2964–2980, May 2014.

[43] A. Jahan, F. Mustapha, S. M. Sapuan, M. Y. Ismail, and M. Bahraminasab, "A framework for weighting of criteria in ranking stage of material selection process," *Int. J. Adv. Manuf. Technol.*, vol. 58, nos. 1–4, pp. 411–420, Jan. 2012.

[44] K. Jemielniak, J. Kossakowska, and T. Urbański, "Application of wavelet transform of acoustic emission and cutting force signals for tool condition monitoring in rough turning of inconel 625," *Proc. Inst. Mech. Eng., B, J. Eng. Manuf.*, vol. 225, no. 1, pp. 123–129, Jan. 2011.

[45] Á. L. P. Gómez, L. F. Maimó, A. H. Celdrán, and F. J. G. Clemente, "MADICS: A methodology for anomaly detection in industrial control systems," *Symmetry*, vol. 12, no. 10, p. 1583, Sep. 2020.

[46] F. G. Zöllner, K. E. Emblem, and L. R. Schad, "SVM-based glioma grading: Optimization by feature reduction analysis," *Zeitschrift für Medizinische Physik*, vol. 22, no. 3, pp. 205–214, Sep. 2012.

[47] A. J. Ferreira and M. A. T. Figueiredo, "An unsupervised approach to feature discretization and selection," *Pattern Recognit.*, vol. 45, no. 9, pp. 3048–3060, Sep. 2012.

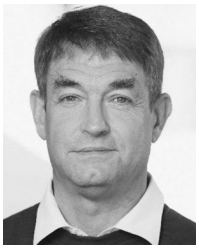
[48] A. Caggiano, "Tool wear prediction in Ti-6Al-4 V machining through multiple sensor monitoring and PCA features pattern recognition," *Sensors*, vol. 18, no. 3, p. 823, Mar. 2018.

[49] T. Kohonen, "The self-organizing map," *Proc. IEEE*, vol. 78, no. 9, pp. 1464–1480, Sep. 1990.

[50] T. Kohonen, "Essentials of the self-organizing map," *Neural Netw.*, vol. 37, pp. 52–65, Jan. 2013.

[51] H. L. García and I. M. González, "Self-organizing map and clustering for wastewater treatment monitoring," *Eng. Appl. Artif. Intell.*, vol. 17, no. 3, pp. 215–225, Apr. 2004.
[52] M. Rosengarten and S. Ramachandran, "SOM-based anomaly detection and localization for space subsystems," in *Advances in Self-Organizing Maps, Learning Vector Quantization, Clustering and Data Visualization*. Cham, Switzerland: Springer, 2019, pp. 57–69, doi: 10.1007/978-3-030-19642-4_6.
[53] J. Coble and J. W. Hines, "Incorporating prior belief in the general path model: A comparison of information sources," *Nucl. Eng. Technol.*, vol. 46, no. 6, pp. 773–782, Dec. 2014.
[54] L. Zheng, H. Yuan, X. Wang, and H. Yin, "Fault diagnosis of transformer based on principal component analysis and self-organizing map neural network," in *Proc. IEEE Int. Conf. High Voltage Eng. Appl. (ICHVE)*, Chengdu, China, Sep. 2016, pp. 1–4.
[55] Y. Zhang, Y. Lv, and M. Ge, "Complementary ensemble adaptive local iterative filtering and its application to rolling bearing fault diagnosis," *IEEE Access*, vol. 9, pp. 47275–47293, 2021.
[56] C. Liang and C. Chen, "Generalized composite multiscale diversity entropy and its application for fault diagnosis of rolling bearing in automotive production line," *IEEE Access*, vol. 9, pp. 84545–84558, 2021, doi: 10.1109/ACCESS.2021.3063322.

MARYAM ASSAFO received the B.Sc. and M.Sc. degrees in communications engineering from the University of Aleppo, Aleppo, Syria, in 2013 and 2018, respectively. She is currently pursuing the Ph.D. degree in computer science with the Brandenburg University of Technology Cottbus-Senftenberg, Germany. From 2014 to 2016, she was a University Lab Instructor. From 2014 to 2019, she was a Teacher with an industrial-and-technical secondary school. She is currently a Research Assistant with the Chair of Wireless Systems, Brandenburg University of Technology Cottbus-Senftenberg. Her current research interests include sensor fusion, predictive maintenance, data mining, and artificial intelligence.



PETER LANGENDÖRFER received the Diploma and Ph.D. degrees in computer science. Since 2000, he has been with the IHP, Frankfurt (Oder), where he is currently leading the Wireless Systems Department. From 2012 to 2020, he was leading the Chair for Security in Pervasive Systems, Technical University of Cottbus-Senftenberg. Since 2020, he owns the Chair Wireless Systems, Technical University of Cottbus-Senftenberg. He has published more than 150 refereed technical articles, filed 17 patents of which ten have been granted already. His research interests include security for resource constraint devices, low power protocols, and efficient implementations of a means and resilience. He worked as a Guest Editor for many renowned journals, such as *Wireless Communications and Mobile Computing* (Wiley) and *ACM Transactions on Internet Technology*.