International Journal of Advanced Mechanics and Control Systems

Volume 2 , Issue 2 2026

A deep learning method based on SAE and LSTM for mechanical equipment anomaly detection

Balbayev , Marco.Department of Mechanics,Italy

Research Article

Date of Submission: 16-02-2026 Date of Acceptance: 10-03-2026 Date of Publication: 11-04-2026

Abstract : Mechanical system anomalies may result in equipment failure that has major effects on the environment, economy, and safety. Anomaly detection for mechanical equipment often demands a great deal of domain expertise since many mechanical devices operate in harsh settings that make them susceptible to many forms of failures. However, a typical problem in many real-world applications is that one may not be able to get empirical information about anomalies or that the historical data is entirely unlabeled, rendering traditional fault detection techniques inapplicable. This paper fills the gap by proposing a novel deep learning-based method for anomaly detection in mechanical equipment. It does this by combining two types of deep learning architectures, long short term memory (LSTM) neural networks and stacked autoencoders (SAE), to identify anomaly conditions in an entirely unsupervised manner. When the historical data is unlabeled and there is no empirical information regarding anomalies, the suggested technique focuses on anomaly identification using numerous feature sequences. The effectiveness and stability of the suggested method are shown in an experiment for anomaly identification in rotary equipment using wavelet packet decomposition (WPD) and data-driven models. The two phases of the process are LSTM-based anomaly detection and SAE-based multiple feature sequence representation. Fivefold cross-validation was used throughout the experiment to verify the stability and effectiveness of the suggested method. The findings demonstrate that the suggested strategy may identify anomalous working conditions with 99% accuracy in an entirely unsupervised learning environment and provide a different way to use and incorporate characteristics for anomaly identification without the need for empirical information.

Keywords :Identification of anomalies Mechanical apparatus SAE LSTM

Introduction

Mechanical equipment is a huge part of manufacturing applications and encompasses a very wide variety of industrial equipment. Mechanical machinery, which is essential equipment in many manufacturing industries, often works in harsh environments, making it susceptible to many forms of malfunctions. These flaws might lead to equipment failure or deteriorate certain machinery functions, such as manufacturing quality, operation safety, and geriatric location [1]. Combined faults would practically make fault categorization more complicated and challenging [2]. In this situation, businesses often choose to detect anomalous conditions first, after which further diagnostic or prognosis might be carried out.

Anomaly detection in mechanical equipment is now a difficult problem due to the complexity of the industrial applications [3, 4]. According to studies, 70–90% of accidents in industrial settings are caused by human operators [5]. Because of this, very complicated computer-based anomaly detection systems are essential for enhancing the precision and dependability of anomaly detection and averting unforeseen mishaps. Furthermore, a manufacturing process's capacity to operate as and when needed often depends on mechanical equipment [6]. Such equipment failure may have detrimental effects on the environment, economy, and safety [7]. In mechanical equipment, the goal of maintenance is often to minimize equipment breakdowns and save maintenance costs by reducing the amount of unnecessary maintenance [8].

However, owing to time-consuming disassembly, large machine size, or environmental constraints, the subsystems in mechanical machinery, such as bearings and gear transmission systems, are often difficult to reach or physically check the failures directly [9]. In the realm of mechanical maintenance, early anomaly identification in mechanical equipment is thus a constant concern [10]. Finding

patterns in data that deviate from predicted behaviors is often the study aim. In recent years, several clever methods for detecting anomalies in mechanical equipment have been put forward and studied. By analyzing the temperature distribution between the goal and relatively stable circumstances, López-Pérez and Antonino-Daviu [11] used infrared thermography to identify failure states in induction motors. Lin et al. [12] extracted failure patterns from nonlinear data in rotary machines to propose an anomaly detection method based on crossover characteristics. Griffin et al. [13] shown how intelligent control techniques can reliably and robustly identify various circumstances by proposing a system based on neural networks and decision trees to detect anomalies for several machining operations. A stacked denoising autoencoder was developed by Lu et al. [14] to identify anomalies and the state of rotating machinery components. A modified kernel-based anomaly detection technique was presented by Aydin et al. [15] for contactless catenary system state monitoring. A data-driven approach based on deep belief networks was presented by Li et al. [16] to forecast when anomalies would occur for maintenance scheduling in machining centers. Deep learning techniques have been used in that study to forecast possible problems. Peña et al. [17] created a rule-based method to leverage data mining to identify anomalies in energy efficiency in smart buildings. For anomaly detection in machine tools, Zhou et al. [18] suggested a customizable approach that might provide explicit knowledge representation with formal semantics and effective knowledge usage. A kernel-based support vector machine was used by Diez-Olivan et al. [19] to identify anomalous circumstances in sensor data monitoring. That paper's suggested method is comprehensible and offers a real-time condition monitoring-based maintenance optimization tool. Each of these suggested approaches has attained specific goals in relevant fields and shown their viability in many instances. Nevertheless, the majority of their approaches unavoidably need for labels or information under steady conditions that might allow diagnostic models to train under supervision. A typical problem in many real-world applications is that the historical data is entirely unlabeled or the anomaly information is unavailable [20]. Unsupervised learning techniques have also been extensively studied and effectively used in several domains to address this problem. For early defect identification, Amruthnath and Gupta [21] used a number of unsupervised learning techniques, such as fuzzy C-means clustering and K-means. The authors of that study [21] employ principal component analysis to make the environment simpler and the nbClust package and elbow approach to determine the number of clusters for unsupervised clustering beforehand. A self-organizing maps-inspired method for unsupervised anomaly detection and location was proposed by Von Birgelen et al. [22]. In this method, quantization error first provides the relationship between severity and degradation, and deviations from the diagnostic model can then be used to assess the severity degree. A self-developing fuzzy-rule-based classifier was suggested by Costa et al. [23]. In that study, the classifier could automatically produce hypothetical class labels based on the differences in input data after the classification criteria were set. A technique for fault identification and detection in power plant coal mills based on dynamic residual analysis and developing fuzzy models is presented by Serdio et al. [24]. This approach does not need the previous availability of fault patterns or annotated samples. In order to identify anomalies in an unsupervised learning environment and extract important failure signs ahead of time, the authors of that work separated the offline training stage from the online detection stages. To optimize the number of clusters [21], distinguish degradation between normal condition and anomaly [22], establish initial rules for classification [23], or extract indicators from identified models [24], the majority of those unsupervised approaches still inevitably require some degree of empirical knowledge in corresponding domains. This research suggests a new deep learning-based approach for anomaly identification in mechanical equipment using long short term memory (LSTM) neural networks and stacked autoencoders (SAE) to close this gap. In particular, when data are gathered without labels, the suggested technique offers an alternate way to use and integrate characteristics for anomaly detection in place of empirical information like the number of clusters and calibrated deterioration.

We conducted an experiment to identify anomalous conditions for a rotating machine in order to test our suggested strategy. The findings showed that in a fully unsupervised learning environment, the suggested method could identify anomalous conditions with 99% accuracy. The remainder of this work is structured as follows. The experiment's setup and data gathering method are described in Section 2. SAE-based representation learning for various feature sequences is described in Section 3. In Section 4, a unique LSTM-based method for identifying anomalies in time series is proposed, along with a discussion and numerical results. The last portion of this study provides a summary of the conclusions.

Set up and data collection

In the experiment, we used a Bently Nevada Rotor Kit RK3 to simulate the working condition of rotating equipment. Three acceleration meters of Kistler 8702B100 were mounted in three directions at the top of the bearing house to collect vibration signals in both normal and anomaly conditions. The test rig is shown in Figure 1. The collection frequency of the acceleration meters is 4096 Hz with the maximum revolving speed at 4000 rpm. Rub generator and mass adjustable load could be modulated to inject failures. The vibrationmonitoring refers to a zero position of the test rig. Data recorded at the zero position would be recorded as the standard values in normal working condition.

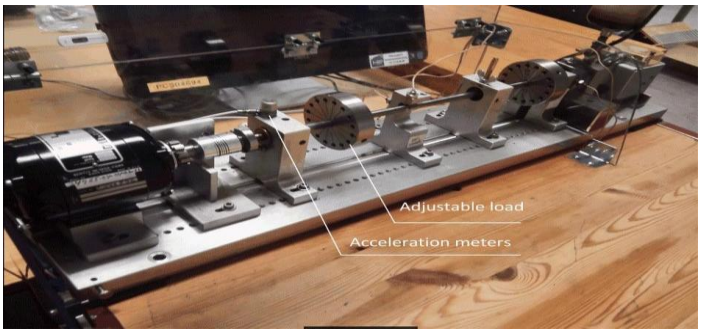


Fig. 1. Bently Nevada rotor kit

In order to mimic load imbalance, weights were added to the mass adjustable load throughout the experiment. Proximity sensors and a portable tachometer will be used to detect the vibration signals using accelerometers at various rotational speeds. Additionally, the vibration signals under typical operating conditions will be captured and combined with unlabeled failure data. The label will only be used to verify and assess how well the suggested method works. In order to characterize the target equipment's operating status in both the time and frequency domains, we used Wavelet Packet Decomposition (WPD) in this study to extract wavelet coefficient-based and energy-based feature sequences from vibration signals [25]. WPD is essentially a wavelet transform that may offer a multi-level time-frequency decomposition of signals by parsing the discrete-time signal through more filters than the discrete wavelet transform [26]. The Daubechies 4 (DB4) wavelet function has been chosen for this paper due to its good performance in estimating the local properties of signals, such as breakdown points [27], and its ability to extract a set of conventional and energy-based features from signals [28]. Different wavelet function types can result in different time-frequency structures. Determining the vibration level acceptability criterion using the acquired multiple features sequence is a frequent challenge when assessing vibration data from mechanical equipment after features are extracted from vibration signals. Using WPD to extract failure information from a series of several characteristics is likewise difficult. The next sections will introduce and use LSTM-based anomaly detection and SAE-based representation learning to the analysis of vibration data in order to address this problem.

SAE-based representation learning for multiple features sequence.As mentioned above, in many practical applications, one may face the dilemma that the history data is collected and recoded unlabeled, let alone classified. To solve this challenge, we propose a SAE-LSTM anomaly detection method in this paper to identify the anomaly condition in an unsupervised learning environment.
When the history data is collected without labels (These labels usually can be used to represent the working condition in a supervised learning manner), an alternative method is to track the changes in multiple features sequence with time-series to identify the anomaly. As shown in Figure 2, features in time domain, frequency domain, or time-frequency domain were first extracted from vibration signals. The original data includes 33 features, which are extracted through WPD and Fourier Transform from vibration signals, in both time and frequency domains. Features 1-6 represent the peak and the second peak of vibration signals at X, Y, and Z directions in frequency domain after Fourier Transform. Features 7-18 denote the standard deviation noises of wavelet coefficients at level 1-4 in direction X, Y, and Z. Features 19-33 express wavelet packet energy features. To be more specific, Features 19-21 are the percentages of energy corresponding to the approximation in three directions and Features 22-33 denote the percentages of energy corresponding to the details at level 1-4 and three directions.
After unity-based normalization, all the extracted features would be transformed into the same scale. To prevent the inputs from explosion, SAE-based representation learning will further be leveraged to reduce the number of features extracted from raw data, and reconstruct the multiple features sequence.

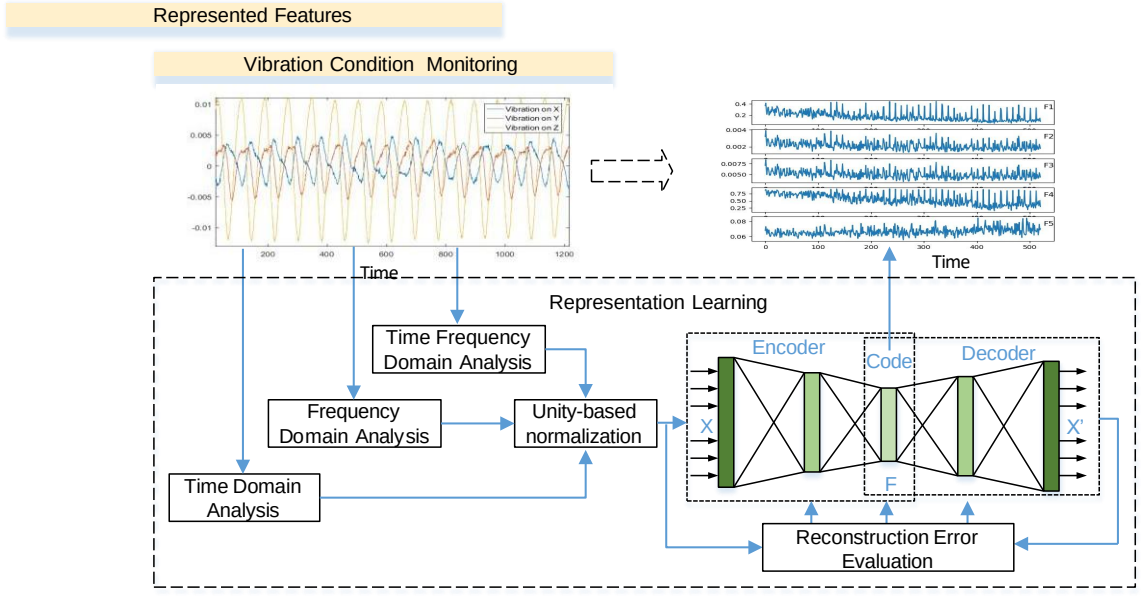


Fig. 2. Process of SAE-based representation learning for multiple features sequence

Feature normalization

During the experiment, vibration signals in both normal and anomalous conditions are collected through three accelerometers (Kistler 8702B100) at the sampling frequency of 4096 Hz. Each sampling unit has been divided into 10 parts with the same length for time series-based detection. Features of each part are first extracted through WPD and Fourier Transform, and will further be represented by autoencoders to reduce the dimension of features sequence. To adjust values measured on different scales to a notionally common scale, unity-based normalization [29] has been applied to normalize the inputs F' for SAE-based representation learning, as shown in Equation 1.

$$F'_{ij} = \frac{i}{}$$

$$F_{ij}-F^{min}$$

$$ij \qquad F^{max}-F^{min} \qquad (1)$$

$$i \qquad i$$

Where F_{ij} denotes the i^{th} feature in j^{th} samples, F^{min} and F^{max} represent the minimum and maximum values of the i^{th} feature in database, respectively. Figure 3 shows part of energy-based features after normalization, which will be used as inputs for SAE-based dimension reduction (Anomaly sampling units are collected when failures are injected, but the labels will be used in validation only). Visually, after normalization, data collected in anomaly still keep certain divergence from normal condition, though we cannot catch the rules directly in this step. It should be noticed that, during the training process, it is supposed

that we only have the data in normal condition. Data in anomaly is only collected to test and validate the proposed method.

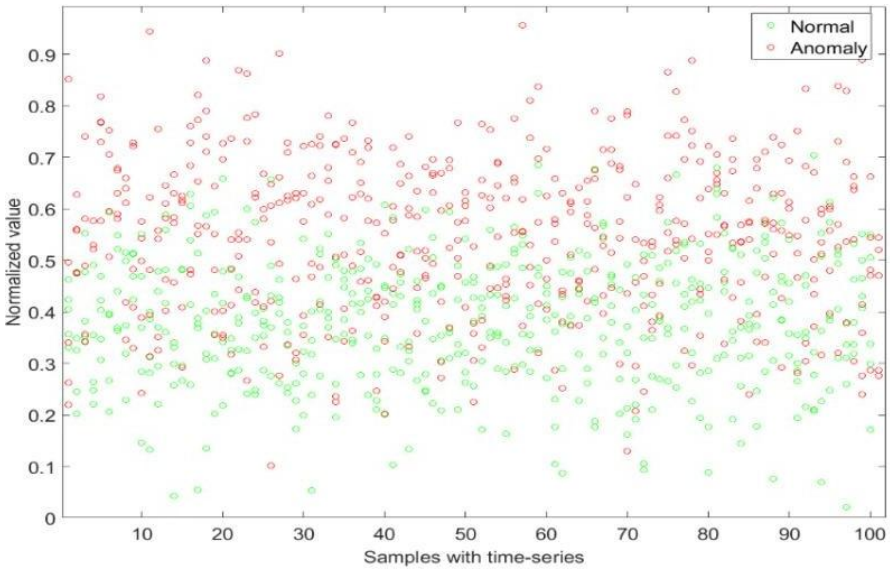


Fig. 3. Part of energy-based features after normalization

Feature representation with SAE

After feature normalization, sparse autoencoders have been leveraged to construct the deep neural network for representation learning. SAE is first proposed in 2007 [30,31]. It is a special type of deep neural networks created through stacking multiple autoencoder layers. The architecture of the deep neural network is pre-trained through single autoencoder layer by layer [32]. The output of SAE is the data input itself, which is leveraged for learning efficient encoding or dimensionality reduction for a set of data. More specifically, it is a nonlinear feature extraction method involving no class labels. Hence, it is generative. When an autoencoder uses three or more layers in the neural network, and the number of hidden layers is greater than one, the autoencoder is considered to be deep [33].

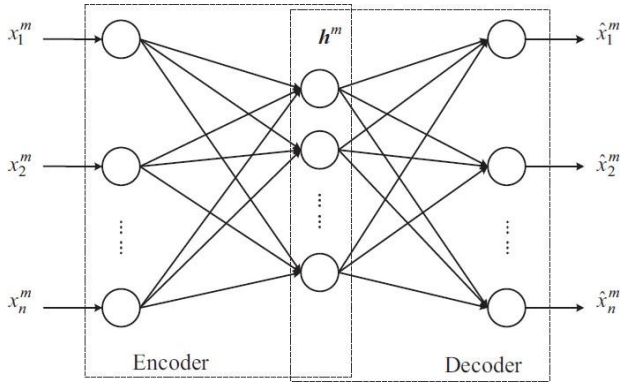


Fig. 4. Architecture of an autoencoder

As depicted in Figure 4, the input layer and hidden layer construct the encoder network, which transforms the input data from a high-dimensional space into codes as a low-dimensional space and the decoder network, which consists of the hidden layers and output layer, reconstructs the inputs from the corresponding codes. The encoder network is explicitly defined as an encoding function denoted by f_θ, which is also called as the
encoder [32]. For each input signal $\boldsymbol{x}^m$ from a dataset $\{\boldsymbol{x}^m\}_{m=1}^M$ (M is the number of training samples), we
label h^m as the obtained encode vector:

$$\boldsymbol{h}^m = f_{\boldsymbol{\theta}}(\boldsymbol{x}^m) \quad (2)$$

The decoder network is defined as a reconstruction function denoted by $g_{\theta'}$, namely the decoder. It
maps $\boldsymbol{h}^m$ from the low-dimensional space back into the high-dimensional space, producing a reconstruction as Equation 3:

$$\hat{\boldsymbol{x}}^m = g_{\boldsymbol{\theta}'}(\boldsymbol{h}^m) \quad (3)$$

The parameter sets of the encoder and decoder are learned simultaneously on the task of reconstructing as well as possible the original input, attempting to incur the lowest possible reconstruction error $L(\boldsymbol{x}, \hat{\boldsymbol{x}})$ over the M training samples. $L(\boldsymbol{x}, \hat{\boldsymbol{x}})$is a loss function that measures the discrepancy between $\boldsymbol{x}$ and $\hat{\boldsymbol{x}}$ [32].
In summary, the autoencoder training aims to find the parameter sets $\boldsymbol{\theta}$ and $\boldsymbol{\theta}'$ minimizing reconstruction error, which can be depicted as Equation (4):

$$\varphi_{AE}(\boldsymbol{\theta}, \boldsymbol{\theta}') = \frac{1}{M}\sum_{m=1}^{M} L(\boldsymbol{x}^m, g_{\boldsymbol{\theta}'}(f_{\boldsymbol{\theta}}(\boldsymbol{x}^m))) \quad (4)$$

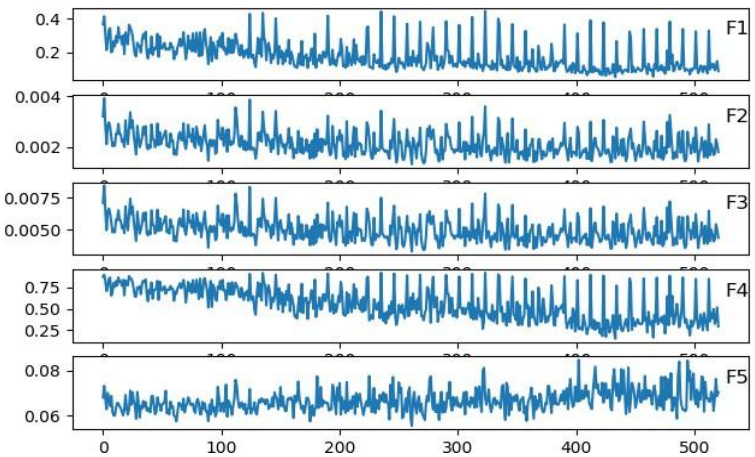
A deep neural network could be constructed by stacking multiple autoencoder layers with a final classification or regression layer on top. Stacking multiple autoencoder layers together allows the network to learn higher order features, where each successive layer represents additional complexity within the input data [34]. Since each hidden layer in SAE is pre-trained through learning multiple nonlinear transformation of the inputs indecently, SAE has the ability to capture the main variations, discover the discriminative information, and represent the features from the raw data [35]. For predictive maintenance, representations of working condition with lower-dimension can improve performance in many situations such as fault classification and detection, especially when the input data is industrial big and row data. With the code vector of the previous trained autoencoder as input for training the next autoencoder, SAE could recognize the characteristics and effectively discover the discriminative information of these signals [36], and subsequently represent mechanical health conditions in a feasible and representational manner. Some practical applications have also shown that SAE has the ability to automatically mine the important information from the frequency spectra according to the diagnosis issues [14]. For this reason, we applied SAE as a representation learning model to reduce the dimension of original features sequence. The SAE constructed during the test has three hidden layers trained through L2 regularization in an unsupervised learning manner with the hidden layers of size 20, 10, and 5, respectively, and a linear transfer function for the decoder. The L2 weight regularization, sparsity regularization, and sparsity proportion have been set to 0.001, 4, and 0.05, respectively. After representation learning, the original 33 features are transformed into a multiple features sequence with time series, as shown in Figure 5.

Fig. 5. Features after representation learning

LSTM-based anomaly identification with time series

Architecture of LSTM

To solve the problem of gradient vanishing in deep learning, a common method is to employ special architectures unaffected by gradient. LSTM is the most typical model in this type of deep learning architectures, which could avoid the fundamental problem of gradients vanishing through special architectures [37]. LSTM neural network is a type of recurrent neural network (RNN) proposed in 1997 to address the problem of insufficient, decaying error backflow in RNN training [38]. The basic idea of LSTM is simple: In a LSTM neural network, memory cells are employed as independent activation functions and identity functions with fixed weights, which are connected to themselves. Due to fixed weight, errors back-propagated through a memory cell cannot vanish or explode but stay as they are [39]. The weight matrixes in conventional RNNs are also trained via backpropagation through time series like the training process of normal neural network. Therefore, the gradients



vanishing problem also happens in RNN while the complexity of the network increases, which means

traditional RNN do not have the ability to discover information or capture dependencies hidden in long-term time series. In this background, LSTM was proposed to prevent back-propagated errors from gradients vanishing or exploding in RNN to deal with issues about long-term dependencies. The core idea behind the LSTM architecture is a memory cell, which can maintain its state over time, and non-linear gating units regulating the information flow into and out of the cell [40]. Compared with traditional RNN, LSTM neural network leverages memory cells with forget gates instead of traditional neurons to establish connections between inputs and outputs [41]. These adopted forget gates can effectively control the utilization of information in the cell states, and enable LSTM the capability to capture nonlinear dynamics in time series sensory data and learn effective representation of machine [42].

As shown in Figure 6, LSTM applies four special and interacting neural network layers, layer α, β, γ , o , instead of a single layer as in a standard RNN [43]. The first layer α is a sigmoid layer also called as forget
gate layer, which returns a value between 0 and 1 in the previous cell state C_{t-1} , while 0 means no information pass and 1 means all information pass. The equation of the first layer can be denoted as Equation (5).

$$\alpha_t = \sigma(W_\alpha \cdot [h_{t-1}, x_t] + b_\alpha) \qquad (5)$$

In Equation (5) – (10), σ is the sigmoid function, W_α is the weight of layer α, [ ] denotes the concatenate operation, x_t is the input x and time t, h_t is the output with respect to x_t, W_α , W_β , W_γ , $W_o$are the weights and b_α , b_β , b_γ, b_o are the biases of the layer α, β, γ, o, respectively. The second layer β is called as input gate layer, which is applied to decide which value shall be updated, denoted as Equation (6)

$\beta_t = \sigma(W_\beta \cdot [h_{t-1}, x_t] + b_\beta)$ (6) Next, a tanh layer γ updates the values to be stored using:

$\gamma_t = tanh(W_\gamma \cdot [h_{t-1}, x_t] + b_\gamma)$ (7) Where $tanh$ is the hyperbolic tangent function.

Then, we can update the previous state C_{t-1} to the current state C_t by Equation (8)

$C_t = \alpha_t \cdot C_{t-1} + \beta_t \cdot \gamma_t$(8)

The final layer is also a sigmoid function layer, which determines what parts of the cell state will be the output, as denoted by Equation (9)

$o_t = \sigma(W_o \cdot [h_{t-1}, x_t] + b_o)$ (9)

Then, the cell state go through tanh function and form the final output as Equation (10)

$h_t = o_t tanh(C_t)$ (10)

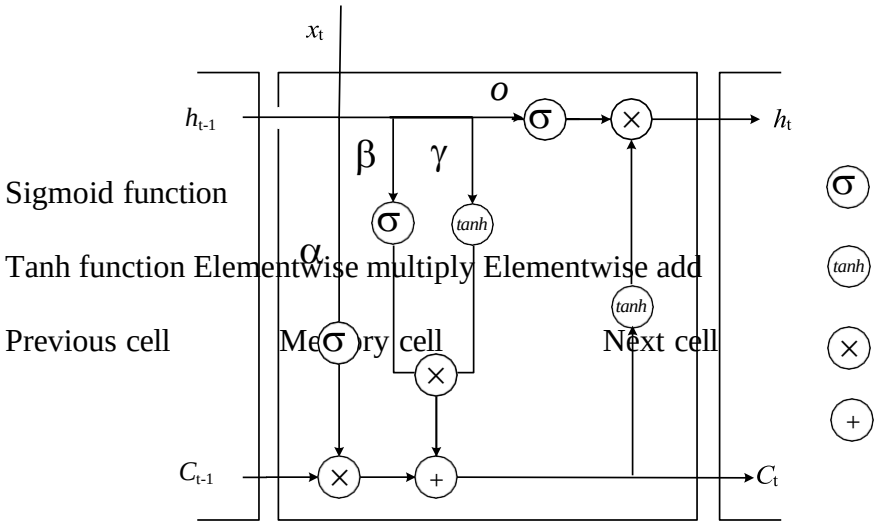


Fig. 6. Memory cell in LSTM

Through stacking multiple memory cells on top of each other, deep RNN can be created with the output sequence of one layer that forms the input sequence for the next, which then enables LSTM to discover information from a dynamically changing contextual window over the input sequence history, rather than a static one as in the fixed-sized window applied in feed-forward neural networks [44].

Application of LSTM in anomaly identification

Because of the recurrent connections in the network, which enable the network to retain memories of previous information, an RNN is a suitable candidate to address problems with high temporal reliance [45]. However, the gradient vanishing issue prevents typical RNN from learning long-term temporal dependencies, as was covered in the previous section. By using the unique memory cells in the design, LSTM may resolve this basic issue [46]. LSTM is an excellent tool for simulating time series because it may retain some information from prior inputs in the output by stacking memory cells, which are conveyed via cell state [47]. For this reason, we would provide LSTM as the time series prediction model for the suggested anomaly detection method. The LSTM network used in our experiment was built in a Python environment using the Keras deep learning toolkit, which is based on the Google TensorFlow library. Before feature extraction, each sample unit of raw vibration signals is split into ten pieces throughout the experiment. As a result, the LSTM model is built to forecast the tenth part using the preceding nine parts. There are five aspects in each phase. Throughout the experiment, the LSTM neural network uses the multiple feature sequence—obtained via unity-based normalization and SAE-based representation learning—as inputs. The procedure of our suggested SAE-LSTM method for anomaly identification is shown in Figure 7. The number of stages N in this instance is up to 10. Initially, the raw vibration signals were split into ten segments. The features at the first nine stages of each feature sequence will be utilized as inputs to map the features at the tenth phase of the training process using SAE-based representation learning.

In order to forecast the features at the tenth step and to represent the preceding nine steps in a multiple feature sequence, the applied LSTM neural network is built using nine LSTM memory cells. To ascertain if the equipment functions normally, the difference between the expected and actual values of the attributes at the tenth stage will be used.

Fig. 7. Process of SAE-LSTM approach for anomaly detection

During the experiment, a selection of 500 samples is applied to train the LSTM neural network with fivefold cross-validation to validate the proposed approach. Figure 8 illustrates the numerical result of fivefold cross-validation, including the mean square errors of all the features and their average values.

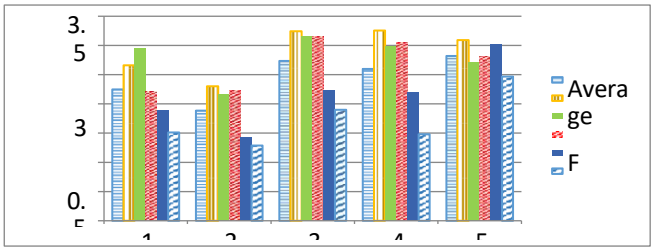


Fig. 8. Numerical result of 5-fold cross-validation

The suggested SAE-LSTM technique may predict numerous feature sequences, according to the numerical result of 5-fold cross validation. Every feature used in the cross-validation process has mean square errors less than 3.5%. Furthermore, the diversity of the same characteristic is less than 1%, confirming the stability of the suggested approach. We will use the second cross-validation run, which shares the highest performance with the lowest training error, to confirm the anomaly detection performance. The construction error with training epochs throughout the training process is shown in Figure 9, where the mean construction error began to converge at around the 300th epoch with very little volatility. The LSTM neural network's average training errors vary from -0.4 to 0.3 across all five feature sequences. The performance of the suggested technique has to be further evaluated since the goal is to differentiate between anomalous and regular working conditions rather than immediately forecasting the multiple feature sequence.

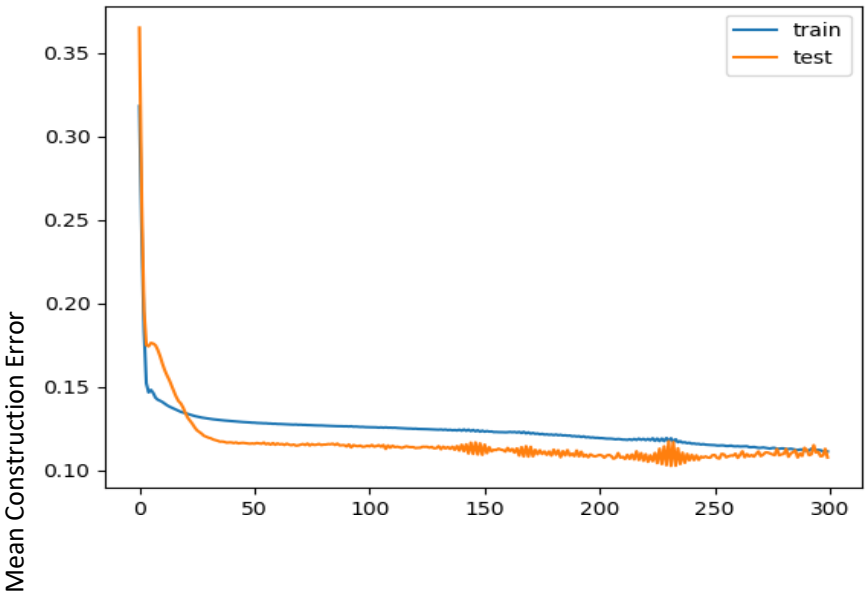


Fig. 9. Construction errors with training

Validation

To validate the performance of proposed SAE-LSTM approach for anomaly detection, a selection of 200 samples, constructed of 150 samples in normal condition and 50 samples in anomaly, is leveraged for testing. Figure 10 shows the testing result of anomaly detection through SAE-LSTM. During the experiment, we applied the largest mean square error in each feature obtained through fivefold cross-validation [48] as the criterion to detect anomaly in the equipment. Since the sensitivity of each feature

to anomaly condition is highly subjective in nature, the criterion applied during the test is based on the overall performance in all features, which means only when all the prediction errors in five features are beyond the average values, the condition would be considered as anomaly. Table 1 lists the overall performance of proposed SAE-LSTM anomaly detection approach and the result of each single feature sequence, respectively.

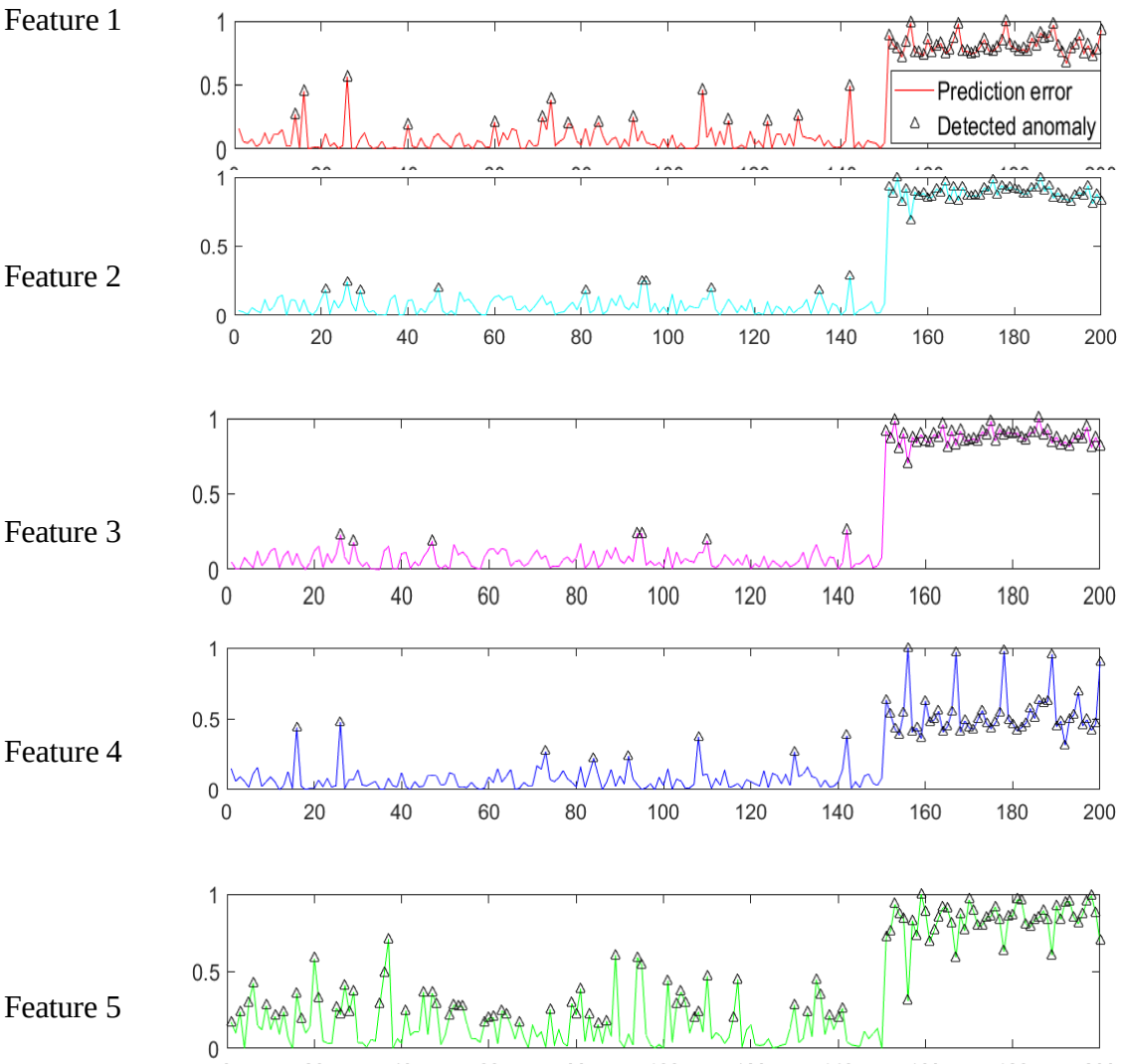


Fig. 10. Result of anomaly detection through SAE-LSTM Table 1 Performance of SAE-LSTM for anomaly detection

Threshold (%)		Number of detected anomaly	Accuracy (%)
Feature 1	18.0418	65	92.5
Feature 2	17.7565	60	95

Feature 3	17.7676	57	96.5
Feature 4	17.3675	58	96
Feature 5	16.7939	109	71.5
F1&F2&F3&F4&F5		52	99

Discussion

The numerical results of each single feature sequence are provided in Table 1 combined with the final result. As demonstrated in Figure 10, there are some samples in normal circumstances misinterpreted by each single feature sequence independently. In addition, it is also evident that not all the features taught using representation learning are acceptable for anomaly detection (e.g., the accuracy of Feature 5 is only 71.5%). We consider the reason is that fault identification is a very subjective problem in nature, and the features after merging were extracted by SAE automatically with the distribution of original data, which makes the extracted features may be partly irrelevant or insensitive to the anomaly condition in this experiment. In this situation, we suppose that Feature 5 was educated to represent the original features, which are somewhat unrelated to the goal, by SAE during the representation learning. However, we also want to highlight that it is also the superiority of deep learning methods since the features are extracted by machine automatically with certain rules hidden behind the distribution of original data instead of human experts with considerable empirical knowledge in the domain (the knowledge that may never be obtained in this case). It might also be a path for future study to further figure out and demonstrate the principles or rules about how to pick the number of learning characteristics and parameters to enhance the performance of representation learning without ground truth.
After all the 5 features are combined as the criteria to evaluate the working condition, which means anomaly samples shall be confirmed only after being identified by all the features, the overall detection accuracy could achieve 99%, which validate the performance of proposed SAE-LSTM method for anomaly detection. In this study, the data-driven model was trained and validated in a totally unsupervised learning environment, which implies the suggested SAE-LSTM technique may optimally identify abnormal working situation when the data is gathered without labels. In practical applications, if some of the data is gathered with labels, it may assist to optimize the detection criteria and further increase the detection accuracy, which would be a path for future study.

In our prior work [49], we also performed some research based on the same circumstances using various supervised approaches for fault classification and degradation assessment in a supervised learning environment, in which all the data-driven models are trained with labels or ground truth from the very beginning. The samples used in that research have been divided into 4 groups, normal condition and three types of injected failures, and the correct classification rate of applied back-propagation neural network, support vector machine, K-nearest neighbor classification, deep belief neural network, and fully connected deep neural network are 99.77%, 99.85%, 99.85%, 99.8%, and 99.87%, respectively. As the next study of that article, data gathered in one kind of the failures are combined with the normal condition without any labels to testify the practicality of the suggested unsurprised anomaly detection technique in this work. As noted previously, after integrating all the collected characteristics as criterion, the proper identification rate is 99%. From the standpoint of Taskonomy [50], the aim of machine learning in this study has been condensed into anomaly detection from categorization.

However, the task's complexity might also be seen as significantly raised since labels were absent throughout the training phase. As a result, this paper's decreased detection accuracy is acceptable. Furthermore, the problem of analyzing and using industrial data without labels is becoming more and more important in the contemporary manufacturing business as automation and sensor technologies advance. A workable solution to the problem may be the suggested deep learning-based unsupervised anomaly detection technique. Future research can also concentrate on identifying other anomalies from unlabeled data.

Conclusion

A unique SAE-LSTM method for anomaly detection in mechanical equipment was presented in this study. The two phases of the suggested approach are LSTM-based anomaly detection and SAE-based multiple features sequence representation. An experiment for anomaly identification in rotary equipment using wavelet packet decomposition (WPD) and data-driven models is carried out in order to verify and evaluate the practical performance of the suggested technique. Fivefold cross-validation findings throughout the experiment show how stable and effective the suggested method is. The findings further reveal that the suggested SAE-LSTM technique may properly identify anomalous working condition in a totally unsupervised learning environment via various characteristics sequence when the history data is unlabeled and the empirical information about anomaly is lacking. Instead of using empirical knowledge, the suggested technique can provide a different way to leverage and integrate characteristics for defect diagnostics.

References

1. He, Z., Lin, J., Lei, Y., and Zuo, M. J. (2013). An overview of empirical mode decomposition in rotating machinery failure diagnostics. Signal processing and mechanical systems, 35(1), 108-126.
2. Hernandez-Vargas, M., Garcia-Perez, A., and Cabal-Yepez, E. (2014). Multiple coupled defects in induction motors may be detected in real time using SVD. 40(7), 2193-2203, Computers & Electrical Engineering.
3. El Kadiri, S., Grabot, B., Thoben, K.-D., Hribernik, K., von Cieminski, G., Emmanouilidis, C., et al. (2016). current developments in business information systems ICT technology. 14–33 in Computers in Industry, 79 (Supplement C). The doi is
4. Prior toCup, R.-E., Sayed-Mouchaweh, M., Costa, B. S. J., and Angelov, P. (2015). An introduction to nature-inspired optimum control of industrial process applications and problem diagnostics. 75–94 in Computers in Industry, 74 (Supplement C). The doi is
5. Wang, P., & Guo, C. (2013). Based on the coal mine’s basic safety management system of safety accident cause analysis. American Journal of Environment, Energy and Power Research, 1(3), 62-68.
6. Ayele, Y. Z., & Barabadi, A. (2016) 'Risk based inspection of offshore topsides static mechanical equipment in Arctic circumstances' 2016 IEEE International Conference on Industrial Engineering and Engineering Management (IEEM). 4-7 Dec. 2016. pp. 501-506.
7. Gao, Z., Cecati, C., & Ding, S. X. (2015). A Survey of Fault Diagnosis and Fault-Tolerant Techniques—Part I: Fault Diagnosis With Model-Based and Signal-Based Approaches. IEEE Transactions on Industrial Electronics, 62(6), 3757-3767..
8. Klingert, F., Roeder, G., Schellenberger, M., Bauer, A., Frey, L., Brueggemann, M., et al. (2017) 'Condition-based maintenance of mechanical setup in aluminum wire bonding equipment by data mining' Information and Communication Technology, Electronics and Microelectronics (MIPRO), 2017 40th International Convention on. IEEE, pp. 72-77.
9. Yang, Y., Dong, X., Peng, Z., Zhang, W., & Meng, G. (2015). Vibration signal analysis employing parameterized time–frequency approach for features extraction of varying-speed rotary equipment. Journal of Sound and Vibration, 335, 350-366.
10. Bangalore, P., & Tjernberg, L. B. (2015). An Artificial Neural Network Approach for Early Fault Detection of Gearbox Bearings. IEEE Transactions on Smart Grid, 6(2), 980-.

11. López-Pérez, D., & Antonino-Daviu, J. (2017). Application of Infrared Thermography to Failure Detection in Industrial Induction Motors: Case Stories. 53(3), 1901-1908; IEEE Transactions on Industry
12. Chen, Q., and Lin, J. (2014). Using crossover properties of nonlinear data, a new feature extraction technique is used to rotary equipment failure diagnostics. Signal processing and mechanical systems, 48(1), 174-187.
13. Griffin, J. M., Doberti, A. J., Hernández, V., Miranda, N. A., & Vélez, M. A. (2017). Intelligent classification from an anomaly viewpoint in part 1: multiple categorization of the force and acceleration data retrieved from numerous machine operations. The International Journal of 93(1), 811-823, Advanced Manufacturing Technology.
14. Lu, C., Qin, W.-L., Wang, Z.-Y., & Ma, J. (2017). Health state identification based on a stacked denoising autoencoder for rotary machinery component fault diagnostics. 130, 377-388; Signal Processing.
15. Aydin, I., Karakose, M., & Akin, E. (2015). Anomaly detection utilizing a modified kernel-based tracking in the pantograph–catenary system.
Expert Systems with Applications, 42(2), 938-948.
16. Li, Z., Wang, Y., & Wang, K. (2017). A data-driven solution based on deep belief networks for backlash error prediction in machining centers. of Intelligent Manufacturing.
17. Peña, M., Biscarri, F., Guerrero, J. I., Monedero, I., & León, C. (2016). Rule-based method to discover energy efficiency anomalies in smart buildings, a data mining technique. 56, 242-255; Expert Systems with Applications.
. 18. Zhou, Q., Yan, P., Liu, H., Xin, Y., & Chen, Y. (2018). Studying a flexible approach to machine tool failure diagnostics and its use. 95(1), 937-960, International Journal of Advanced Manufacturing Technology.
19. Diez-Olivan, A., Pagan, J. A., Khoa, N. L. D., Sanz, R., & Sierra, B. (2018). Kernel-based support vector machines for automated health status evaluation while monitoring sensor data. 95(1), 327-340, International Journal of Advanced Manufacturing Technology.
20. Landry, M., Leonard, F., Landry, C., Beauchemin, R., Turcotte, O., & Brikci, F. (2008). An Improved Vibration Analysis Algorithm as a Diagnostic Tool for Detecting Mechanical Anomalies on Power Circuit Breakers. IEEE Transactions on Power Delivery, 23(4), 1986-1994.
21. "A research study on unsupervised machine learning algorithms for early fault detection in predictive maintenance" by Amruthnath, N., and Gupta, T. (2018) was presented at the 5th International Conference on Industrial Engineering and Applications (ICIEA). IEEE, 355–361.
22. von Birgelen, A., Niggemann, O., Buratti, D., and Mager, J. (2018). Self-organizing maps for predictive maintenance and anomaly localization in cyber-physical production systems. CIRP Procedia, 72, 480-485.
23. Costa, B. S. J., Angelov, P. P., & Guedes, L. A. (2015). Recursive density estimation and a self-evolving cloud-based classifier are used for fully unsupervised fault detection and identification. 150, 289-303; Neurocomputing.
24. Serdio, F., Lughofer, E., Pichler, K., Buchegger, T., & Efendic, H. (2014). Soft computing-based residual-based fault detection for rolling mill condition monitoring. Information Sciences, 259, 304-320.
25. Hu, Q., He, Z., Zhang, Z., & Zi, Y. (2007). doi:https://doi.org/10.1016/j.ins.2013.06.045. rotating machinery fault diagnostics using SVMs ensemble and an enhanced wavelet package transform. Signal processing and mechanical systems, 21(2), 688-705.
26. Zhang, Y., Liu, B., Ji, X., & Huang, D. J. N. P. L. (2017 EEG Signal Classification Using Wavelet Packet Decomposition and the Autoregressive Model (journal paper). 45(2), 365-378.
27. Borges, D. b. L., and Ferreira, C. B. R. (2003). Wavelet transform decomposition analysis of mammography categorization. Letters on Pattern Recognition, 24(7), 973-982.
28. Murugappan, M., Sazali, Y., and Ramachandran, N. (2010). employing the discrete wavelet

transform to classify human emotions from EEG. Biomedical Science and Engineering Journal, 3(04), 390.
29. Majdi, A., Gholipour, Y., and Rastbood, A. (2017). FEM-based artificial neural network for segmental tunnel lining structural force prediction. International Journal of Geo-Engineering and Mining, 51(1), 71-78.
30. "Greedy layer-wise training of deep networks" (Bengio, Y., Lamblin, P., Popovici, D., & Larochelle, H. 2007) Neural information processing systems. pages 153–160.