



A COMPARATIVE STUDY OF DIVERSE AUTOENCODER MODELS IN LOCAL GEAR PITTING FAULT DIAGNOSIS

^{1,*}Mustafa YURTSEVER , ²Rafet Can ÜMÜTLÜ , ³Hasan ÖZTÜRK 

¹İzmir Democracy University, Faculty of Economics and Administrative Sciences, Department of Management Information Systems, İzmir, TÜRKİYE

²TOFAŞ Turkish Automobile Factory Joint-Stock Company, Research & Development Center, Propulsion Systems, İzmir, TÜRKİYE

³Dokuz Eylül University, Faculty of Engineering, Department of Mechanical Engineering, İzmir, TÜRKİYE
¹mustafa.yurtsever@idu.edu.tr, ²rafet.canumutlu@tofas.com.tr, ³hasan.ozturk@deu.edu.tr

Highlights

- Gearbox pitting faults occur frequently in high-torque industrial systems.
- Vibration analysis is a common method for diagnosing gear pitting faults.
- Deep learning is increasingly used for gear pitting fault classification.
- Autoencoder models like CAE, SAE, and VAE extract gear pitting features.
- Sparse Autoencoder is efficient for diagnosing gear pitting using raw data.



A COMPARATIVE STUDY OF DIVERSE AUTOENCODER MODELS IN LOCAL GEAR PITTING FAULT DIAGNOSIS

^{1,*}Mustafa YURTSEVER , ²Rafet Can ÜMÜTLÜ , ³Hasan ÖZTÜRK 

¹İzmir Democracy University, Faculty of Economics and Administrative Sciences, Department of Management Information Systems, İzmir, TÜRKİYE

²TOFAŞ Turkish Automobile Factory Joint-Stock Company, Research & Development Center, Propulsion Systems, İzmir, TÜRKİYE

³Dokuz Eylül University, Faculty of Engineering, Department of Mechanical Engineering, İzmir, TÜRKİYE
[1mustafa.yurtsever@idu.edu.tr](mailto:mustafa.yurtsever@idu.edu.tr), [2rafet.canumutlu@tofas.com.tr](mailto:rafet.canumutlu@tofas.com.tr), [3hasan.ozturk@deu.edu.tr](mailto:hasan.ozturk@deu.edu.tr)

(Received: 21.10.2024; Accepted in Revised Form: 20.12.2024)

ABSTRACT: Gearbox, which is one of the most important and frequently used components among mechanical power transmission systems, has often been observed to occur in gear surface pitting faults in industrial applications that require high torque. For the diagnosis of gear pitting faults, vibration analysis is one of the commonly utilized techniques. Recently, there has been an increasing interest in applying deep learning approaches for classification and learning feature representations. Deep learning provides an excellent opportunity to integrate vibration signals for gear pitting fault diagnosis. Therefore, in this study, autoencoder models Contractive Autoencoder (CAE), Sparse Autoencoder (SAE) and Variational Autoencoder (VAE) are used to extract deep feature representations of gear pitting data. Without using any additional feature extraction techniques, in this study uses the raw vibrational data directly to identify the local gear pitting faults. Experimental results have shown that Sparse Autoencoder is a viable and efficient feature extraction method and provides a new research method for gear pit fault diagnosis.

Keywords: Contractive Autoencoder, Sparse Autoencoder, Variational Autoencoder, Local Gear Pitting, Fault Diagnosis

1. INTRODUCTION

In terms of cost and downtime, mechanical system maintenance is critical for the industry, since unanticipated faults can result in significant losses of production, energy, and quality, as well as damage to machine parts, unscheduled downtimes, and higher expenses. The main advantage of a monitoring technique is that it makes it possible to detect escalating problems in the machine before they become serious. As a result, vibration analysis is a practical and frequently chosen technique for evaluating the conditions of a machine while it is in operation.

For industrial applications, mechanical power transmission systems in machine components are critical. Among mechanical power transmission systems, the gearbox is one of the most often utilized components. Gear systems are used to transfer rotation or power from one shaft to another at a predetermined rate. In the absence of any flaws in the gear system, these criteria may be accomplished effectively. Manufacturing flaws, beginning torques, overload, gear misalignment, lubrication debris, bearing failures, resonant vibrations, and other factors can all increase the magnitude of forces on a gearbox's tooth surface [1]. As a result of all these factors, some teeth may be subjected to a greater and more variable load than their capacity. Surface faults will develop after a specific number of working cycles if the surface tension of the teeth exceeds the fatigue limit of the gear material. Wear, cracks, pitting and tooth breakage are common surface problems in gears [2]. It has been frequently observed that pitting faults occur on the gear surface, especially in industrial applications that require high torque.

In predictive maintenance studies, data is collected by various methods and used in analysis. Infrared thermography [3], lubricant analysis [4], vibration analysis [5,6], acoustic emission [7,8], motor current signature analysis [9], and other condition monitoring techniques may all be used to identify faults in

*Corresponding Author: Mustafa YURTSEVER, mustafa.yurtsever@idu.edu.tr

industrial systems. Among these techniques, vibration analysis is the most widely used for gearbox condition monitoring. The reason for this is that the failures on the teeth spoil the gear teeth surface and the unwanted forces that occur during meshing from these damaged areas cause vibration [10]. Based on this information, by measuring and analyzing the vibration of a gearbox, it is possible to determine the type of fault, which gear pair it is on in the gearbox, and the severity of the defect [11].

Extreme noise significantly impacts the vibration data collected from gearboxes, making it challenging to accurately interpret the signals. As a result, conventional vibration analysis methods, such as time domain and frequency domain analysis, become less effective in detecting faults within the gearbox. The presence of high levels of noise obscures the patterns that these methods rely on, complicating the identification of potential issues [12,13]. To handle fault detection difficulties, artificial neural networks (ANN) [6,14], support vector machines (SVM) [15], genetic algorithms (GA) [16], signal decomposition approaches [17], and support vector data descriptions [18] have all been developed recently. According to studies, using these approaches, it is feasible to interpret the machine's status. As a result, these categorization algorithms have become increasingly important in recent years.

In recent years, researchers have a high interest in deep learning in different fields. Deep learning uses multi-layer structures to extract hidden information from a dataset for classification or other purposes [19]. Deep learning excels at dealing with big data compared to traditional feature extraction methods. Signals collected in diagnostic problems are often affected by various forms of noise. The success of deep learning methods in noise removal and robust feature extraction increases the studies.

Chen, et al. proposed a fused-stacked AEs for planetary gear fault diagnosis method based on autoencoder methods [20]. The method is capable of learning features from raw data. Qu, et al. proposed a method for detecting gear pitting failures that incorporates dictionary learning in sparse coding into an autoencoder (AE) network [21]. This method is regarded as an adaptive feature extraction method for detecting machinery faults. Li, et al. presented a novel method for gear pitting fault diagnosis using stacked autoencoder and Gauss-Binary restricted Boltzmann machine (GBRBM) [22]. The GBRBM layer is used to process continuous time domain vibration signals. Li, et al. proposed a method for integrating vibration and acoustic emission signals using convolutional neural network (CNN) and gated recurrent unit (GRU) networks [23]. The results obtained using the presented method are more accurate than those obtained using the CNN or GRU network alone. The deep sparse autoencoder method is highly effective in diagnosing various pitting conditions in gear systems. It can also clearly illustrate trends in the severity of gear wear failures, offering valuable insights into the progression of damage. This method enables the detection and differentiation of fault levels with greater precision compared to traditional approaches [24]. Li, et al. presented a method for gear pitting fault diagnosis that integrated sparse autoencoder and deep belief network [25]. In the study the normalized frequency domain signals were used as the input.

Unsupervised learning algorithm, which is represented by deep learning method, especially AE, has a different structure from traditional neural networks. Deep learning can discover the hidden correlation between the data passed into the model. It can extract compressed features of input data without human intervention. Another diagnostic issue is the utilization of nonlinear and high-dimensional data. While having more data appears to be a benefit in terms of obtaining more information, it is challenging to put to good use. Dimensional reduction to create a smaller representation of higher-dimensional data which makes it easier to identify and store high-dimensional data is one technique of overcoming this difficulty. The autoencoder allows for greater data representation as well as reduced data [26].

The majority of the pitting diagnostic approaches listed above entail a feature extraction and feature selection process. For many years, vibration signals have been a key tool in diagnosing pitting faults in mechanical systems. Numerous signal processing and analysis methods have been developed and applied to extract valuable information from vibration data, allowing for the detection and monitoring of pitting-related issues in machinery like gearboxes and rotating equipment. These methods play an essential role in identifying fault patterns and assessing the condition of mechanical components. Recent advances in deep learning offer opportunities to use vibration signals raw. It eliminates the need for additional operations such as converting the time domain signals to the frequency domain.

Due to its higher performance when dealing with complicated, large-scale data, deep learning is a good choice for diagnostics. This paper proposes a method for diagnosis of local gear pitting based on stacked autoencoder models. In the proposed method, the input set created with various AE models. Three types of stacked network architectures are compared for gear pitting fault diagnosis. Such comparisons are rare in literature and can be extremely valuable to researchers and industry professionals. These comparisons can help us better understand the performance, efficiency and learning capacity of different stacked network structures, which can contribute to better design of future studies and applications.

2. MATERIAL AND METHODS

2.1. AutoEncoder

An autoencoder is a type of neural network designed to replicate the input data at the output layer. An autoencoder is a specialized neural network architecture used primarily for unsupervised learning. Its main purpose is to copy the input data and reproduce it at the output layer. The network consists of two parts: an encoder that compresses the input into a lower-dimensional representation, and a decoder that reconstructs the original input from this compressed form. By forcing the model to learn an efficient encoding, it can capture key features and patterns within the data, often useful for tasks like dimensionality reduction, feature extraction, and anomaly detection. At this point, the data output by the hidden layer units is the low-dimensional representation of the original data, containing all of the data's information. The autoencoder structure is shown in Figure 1.

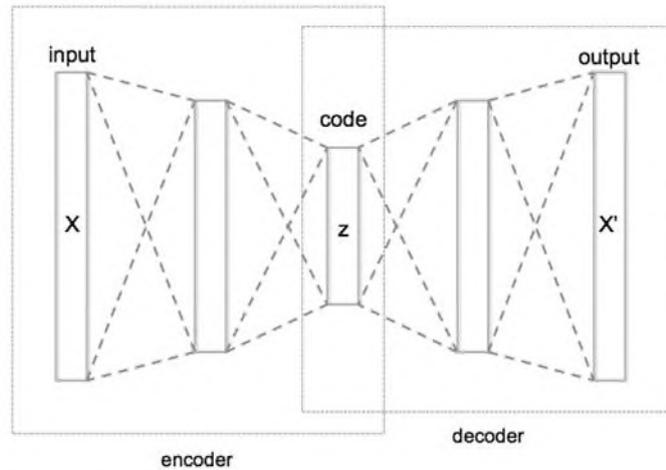


Figure 1. The Structure of an Autoencoder

The encoder network performs feature transformation on the input data. High-dimensional data is first compressed to extract its core features, then the decoder network uses this compressed data to generate an output similar to the original input. Assuming that there are n unlabeled training samples $x = \{x_1, x_2, x_3, \dots, x_n\}$. The encoder and decoder processes of AE are as follows:

$$h = f(Wx + b(1)) \quad (1)$$

$$x' = f(Wh + b(2)) \quad (2)$$

A conventional neural network with an activation function serves as the representation of the encoding network, in Eq. 1. H represents the latent dimension in Eq. 1. W is the weight matrices, $b(1)$ and $b(2)$ are the bias vectors, and f is the activation function. The decoding network can be represented similarly. It is expressed by varying weights, biases, and activation functions in Eq. 2.

2.2. Contractive AutoEncoder

CAE is a version of the classic AE that improves the resilience of the training set against noisy perturbations [27]. As a regularization term, a Frobenius norm of Jacobian is used. By adding a penalty to the Frobenius norm of the Jacobian matrix of the hidden features with respect to the inputs, the Contractive Autoencoder (CAE) encourages the model to learn more robust and stable representations. This regularization constrains how much the hidden features can change in response to small input variations, which helps the model capture the essential underlying factors in the data. As a result, CAE becomes a powerful tool for extracting meaningful, robust hidden features, making it particularly useful in tasks that require stable feature extraction under noisy or complex conditions [28].

CAE penalizes the sensitivity to input, which is measured by, in order to learn invariant and robust feature representation. A CAE is derived by augmenting the autoencoder's reconstruction loss with a regularization term. This term is computed as the Frobenius norm of the Jacobian matrix that captures the relationship between hidden layer activations and input variations, which is represented as

$$\|J_f(x)\|_F^2 = \sum_{i=1}^k \sum_{j=1}^e \left(\frac{\partial h_j(x)}{\partial x_i} \right)^2 \quad (3)$$

The total loss function can be stated numerically as:

$$L(x, x') + \lambda \|J_f(x)\|_F^2 \quad (4)$$

where λ is the penalty term parameter that regulates the penalty term's and reconstruction fault's relative relevance. The disturbance of the input in all directions can be decreased by adding the penalty term to the loss function, achieving the goal of anti-noise and therefore strengthening the robustness of the recovered features.

2.3. Sparse Autoencoder

A sparse autoencoder is a type of artificial neural network commonly used in the field of deep learning. Its primary objective is to represent and extract features from input data. An autoencoder typically consists of two main components: an encoder and a decoder. In the case of a sparse autoencoder, the concept of sparsity is introduced during the learning process, ensuring that the majority of activations in the model are close to zero. This encourages the model to learn more general and meaningful features, making it more resistant to noise. Sparsity aims to improve the feature extraction process, leading to more effective and learnable representations. One of the traditional methods, SAE penalizes the hidden unit biases in order to learn relatively sparse features [29]. The SAE emerges from augmenting the basic autoencoder's cost function with a sparsity penalty term. This modification enables the network to discover more abstract and meaningful compressed representations than what a standard autoencoder can achieve[30].

$$J_{\text{sparse}}(W, b) = J(W, b) + \beta \sum_{j=1}^{S_2} KL(p || \hat{p}_j) \quad (5)$$

The sparsity penalty term's weight is set by β . It is necessary to determine W and b ideal parameters during the coding procedure. Since W and b are directly related to the sparse cost function presented in Eq. (5), these two parameters can be obtained by solving the sparse cost function by minimizing $J_{\text{sparse}}(W, b)$.

2.4. Variational Autoencoder

As a powerful generative modeling tool in machine learning and deep learning, a VAE can learn the underlying distribution of data and generate new samples similar to the training data. VAEs are designed to learn a probabilistic representation of input data, particularly in the context of unsupervised learning. VAEs differentiate themselves from classical autoencoders by implementing a stochastic encoding process, where the latent space is characterized by probability distributions rather than deterministic encodings [31], often assuming a Gaussian distribution. This enables the generation of diverse and realistic samples during the decoding process. VAEs incorporate a loss function that not only measures the reconstruction error but also includes a term for the divergence between the learned latent distribution and a predefined prior distribution, usually a standard normal distribution. This dual-loss structure encourages the model to learn a compact and smooth representation of the data, making VAEs suitable for tasks like image generation, data synthesis, and latent space interpolation.

Real samples are encoded to latent vectors using the variational inference theory. As long as the vectors follow the Gaussian distribution, they effectively preserve the deep properties of the original data and transfer to the training dataset's distribution. The data that the decoder has deciphered will then make more sense and resemble the original data somewhat closely [32]. The most significant difference between a Variational Autoencoder (VAE) and a standard Autoencoder (AE) lies in the presence of a sampling layer. In a standard AE, there is no such sampling mechanism; the latent space is directly encoded from the input. In contrast, VAEs introduce a probabilistic approach by including a sampling layer, where latent variables are sampled from a distribution, typically a Gaussian. This enables VAEs to generate new data that closely resembles the original input. To achieve this, the VAE optimizes the marginal likelihood of the data, as demonstrated by the maximization of the marginal distribution, such as in Eq. 6. This probabilistic framework allows VAEs to generate more diverse and realistic data compared to standard AEs.

$$p_{\theta}(x) = \int p_{\theta}\left(\frac{x}{z}\right) p_{\theta}(z) dz \tag{6}$$

where $p_{\theta}(z)$ denotes the previous distribution of hidden variables z , and $p_{\theta}(x|z)$ reflects the reconstruction of the original data x through a hidden variable z . The variational autoencoder structure is shown in Figure 2.

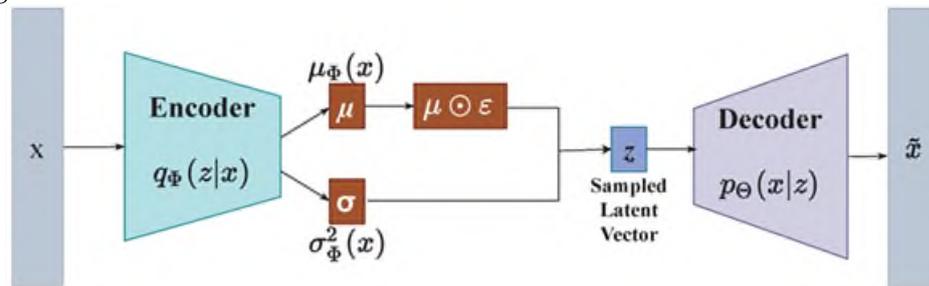


Figure 2. Variational autoencoder [33]

2.5. The General Procedure of the Proposed Method

A stacked autoencoder is constructed by connecting multiple autoencoders in sequence, one after the other. In this structure, the output of one encoder layer becomes the input for the next encoder layer. This process enables the network to learn progressively more abstract and higher-level feature representations of the input data at each layer. As each successive layer compresses the information further, stacked autoencoders are particularly useful for tasks that require deep feature extraction or hierarchical data representation. To acquire access to the optimum connection weights and bias values of the complete stacked auto-encoder network, the greedy layer-wise pre-training approach is utilized. The back-

propagation algorithm is mostly used in the network training section to alter the network's topology and hyper parameters. The optimization of weights and biases using the backpropagation algorithm can be significantly enhanced by employing a greedy layer-wise pre-training approach. This method involves initializing the network in a layer-by-layer manner, where each layer is trained independently before integrating it into the full network. Typically, unsupervised learning techniques like autoencoders are utilized during the pre-training phase to capture the hierarchical features in the data. Once pre-training is complete, the network undergoes fine-tuning using back propagation, which adjusts the weights and biases to minimize the overall loss function. This approach mitigates issues like vanishing gradients in deep networks, accelerates convergence, and improves generalization by providing a better starting point for the optimization process. The hyperparameters, such as the number of layers and neurons in each layer, were determined through a combination of domain knowledge and experimental validation. Initially, a range of configurations was explored based on commonly used heuristic values (e.g., powers of 2, such as 256 and 128 neurons). These configurations were evaluated using a separate validation dataset to identify the architecture that yielded the best performance. This iterative process ensured that the selected hyperparameters effectively balanced model complexity and generalization capability.

The purpose of the method used in the study is to classify the local gear pitting faults. The accelerometer sensor is utilized to gather the raw vibration signals of the gear in the initial data acquisition step. The complete set of collected vibration data is systematically divided into three subsets: the largest portion (60%) is allocated for training, while the remaining data is equally split between validation and testing (20% each). The collected data has been normalized for better training in the deep learning method. Test data is used to validate the model. The deep learning architecture has the capacity to turn several hidden layers into complicated nonlinear transformations. The effective deep features can be learned directly from the raw vibration signal using the layer-by-layer learning approach for deep features. The extracted deep features of each pit gear condition in each concealed layer are becoming more distinguishable. The gear pitting data set, which has 335 inputs, will be used in this instance. The encoder has 256 hidden neurons in its hidden layer, followed by 128 neurons in its center hidden layer. Because the model is symmetrical, the decoder also has an output layer with 335 neurons and a hidden layer with 256 neurons. As the number of hidden layers increases, stacked autoencoders are constructed. It implies that more than one input value will be compressed. We want to improve feature extraction performance by adjusting the stacked autoencoder's input value and hyperparameters. The deep model setup is finished after adding the last softmax layer. The structure of the proposed model is shown in Figure 3.

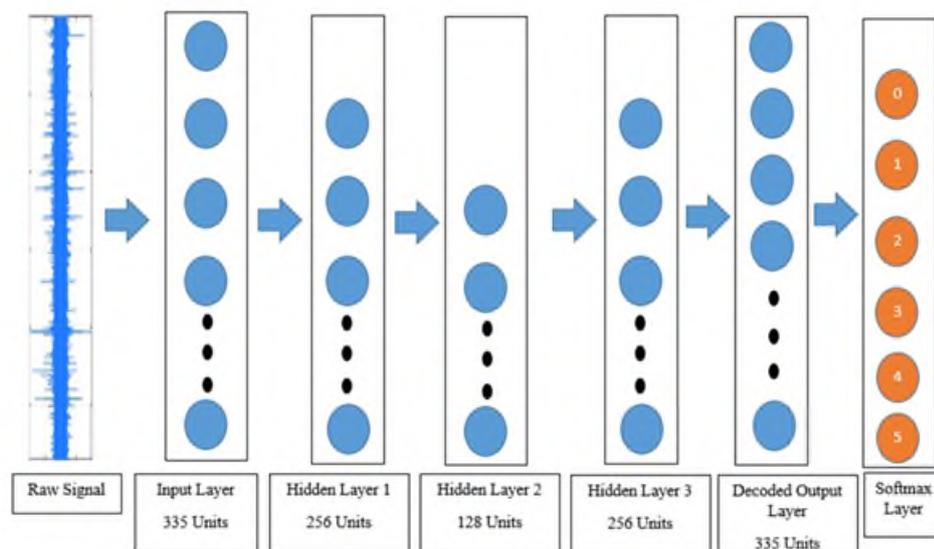


Figure 3. The Structure of Proposed Model

2.5. Experimental Setup

A two-stage industrial helical gearbox is employed in the experiment as illustrated in Figure 4. The whole set of helical gears is built from induction case-hardened steel. To avoid the negative impacts of misalignment, the alternating-current (AC) motor and direct-current (DC) generator are coupled via belt pulley contact mechanisms. The first-stage pinion gear has 29 teeth that mesh with a 40-tooth wheel. The second-stage pinion gear, which is driven directly by a 40-tooth wheel, has 13 teeth that mesh with a 33-tooth wheel. The 2.2 kW load capacity of the DC was substantially lower than the almost 8.1 kW load capacity of the gearbox employed, which was fed by a generator whose output was used to supply an adjustable resistor bank. For this reason, the face width of the pinion test gear was reduced from 12 mm to 4 mm, allowing it to be tested under reasonably high load conditions. This adjustment was made to ensure that the gear could endure the applied load without exceeding practical testing limits, enabling more accurate analysis of its performance under stress [34].

The vibration signals generated by the gears were measured using two [PCB 352A76] accelerometers, which are designed for vibration measurements within a frequency range of 5–16,000 Hz. These accelerometers were mounted mutually perpendicular to each other on the input shaft bearing housings to minimize the effects of the transmission path, as illustrated in Figure 4. A PCB 480C02 signal conditioner was employed to amplify the outputs from the accelerometers [34].

The position of the input shaft was monitored using a 5V DC ME4-S12L-PA inductive sensor, which generates a single pulse per revolution. All signals from the accelerometers and the positioning sensor were sampled at an appropriate rate and recorded using a National Instruments (NI DAQ Card 6036E) data acquisition system, with data collection and processing conducted through LabVIEW 7.0 software [34].

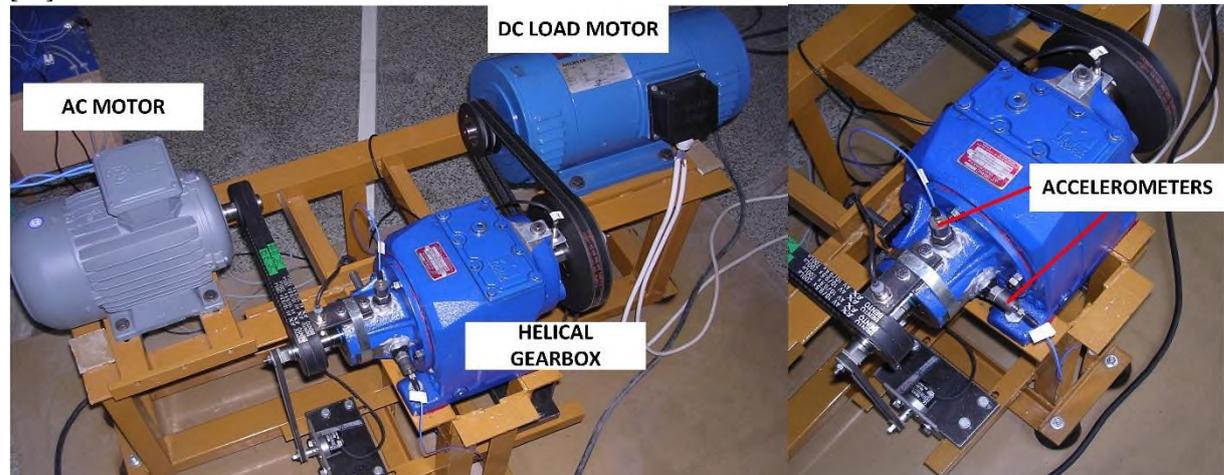


Figure 4. Experimental Setup With Helical Gearbox [34]

The operating settings of the experiments are given in Table 1. These settings were maintained throughout each test for all scenarios.

Table 1. Characteristics of the helical gearbox [34]

	First stage	Second stage
Number of teeth	29/40	13/33
Normal module (mm)	1.25	2.5
Pressure angle (o)	20	20
Profile shift (pinion/wheel)	+0.325/+0.259	+0.437/+0.340
Helix angle (o)	30	15

As a consequence of shock or cyclic load fluctuation, the specified gear load may be exceeded, and certain gear teeth may be subjected to a load greater than the gear's capacity. Pitting faults on the tooth

surfaces that are subjected to a greater load may develop over time in such instances. In the first stage, an electro-erosion machine was used to create a simulated surface defect resembling pitting on the 29th teeth of the pinion gear. These pitting faults occur as an initial fault and then recur on the teeth, increasing the severity of the failure. As illustrated in Figure 5b, a circular hole with a diameter and depth of roughly 0.7 mm and 0.1 mm was seeded into a single tooth surface. This gear tooth, called as the center tooth, was positioned so that it meshed with the pinion at about 270° pinion rotations. To represent the progression of the fault, the number of pits on the main tooth surface is raised to three, and its two neighbor teeth are pitted with one pit, as illustrated in Figure 5c. As illustrated in Figure 5d, the number of pits on main teeth is raised to five, and two pits are placed on both of the other two neighbor tooth surfaces in the third stage of fault simulation. The severity of the faults is doubled in the fourth step by increasing the number of pits on these teeth, as seen in Figure 5e. The number of pits was increased during the last stage of pitting development, and the surface of the center tooth is entirely covered by severe pitting markings, as illustrated in Figure 5f.

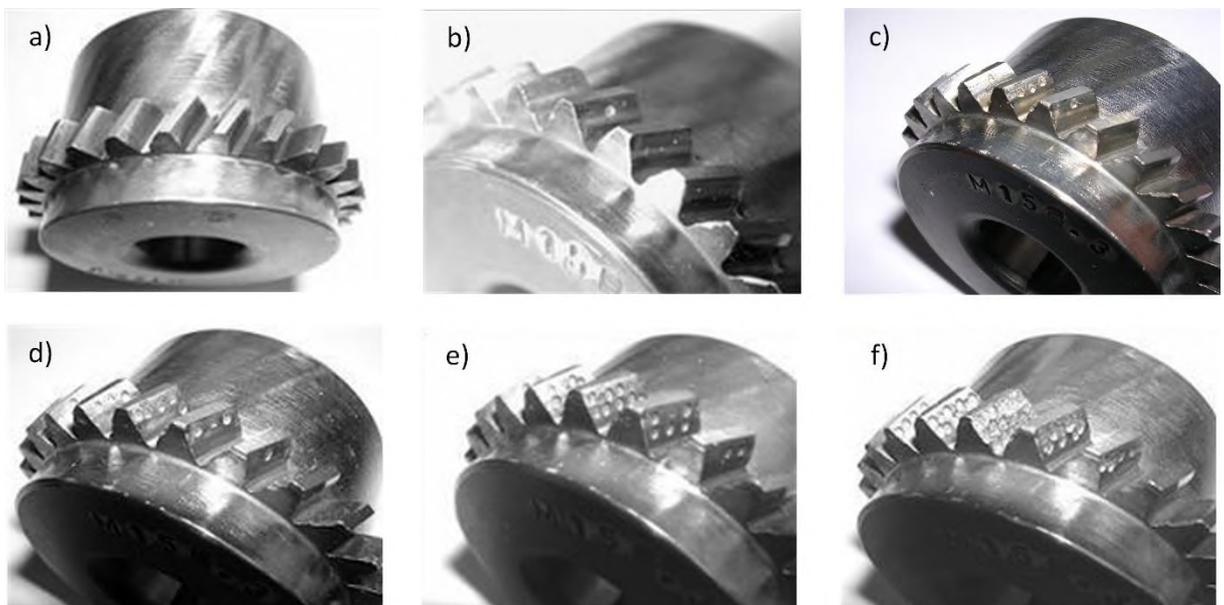


Figure 5. Fault stages used in the experiment: (a) healthy, (b) first fault, (c) second fault, (d) third fault, (e) fourth fault, and (f) fifth fault [34]

2.6. Analysis of Vibration Data in Time and Frequency Domain

The raw data were subjected to synchronous time averaging to highlight repeated characteristics and remove unnecessary noise. For every three turns of the input pinion gear, the averaged vibration signals are received. Figure 6 shows averaged signal of time domain gearbox vibration of vertical direction. It is seen that the gear vibration figures for healthy and each faulty condition are more or less similar to each other and no signs of fault progression are observed until the last fault stage. The frequency spectrum of each of the average helical gear vibrations is shown in Figure 7. The test pinion's speed is set at 2678 rpm during each testing, resulting in a fundamental tooth meshing frequency of 1294 Hz for the first stage and 420.7 Hz for the second stage. The computer-based recording and sampling of the vibration and location signals used a 15 kHz sample rate.

The spectra show a strong peak at 403 Hz, 1299 Hz and 2597 Hz, which correspond to the first gear mesh frequency for gearbox's second stage, first and second gear mesh frequencies for the gearbox's first stage. Except for Fault 5, the frequency spectra are quite similar. As in the time domain, the distortion in the last fault can be clearly observed in the frequency domain figures. Especially when the dominant frequencies and the amplitudes of these frequencies are considered, Fault 1 and Fault 2 are the most similar situations. As the fault progresses, the dominance of 5176 Hz, the fourth gear matching frequency of the

first stage, decreases. This is expected behavior due to high-frequency vibrations related to the growth of the fault.

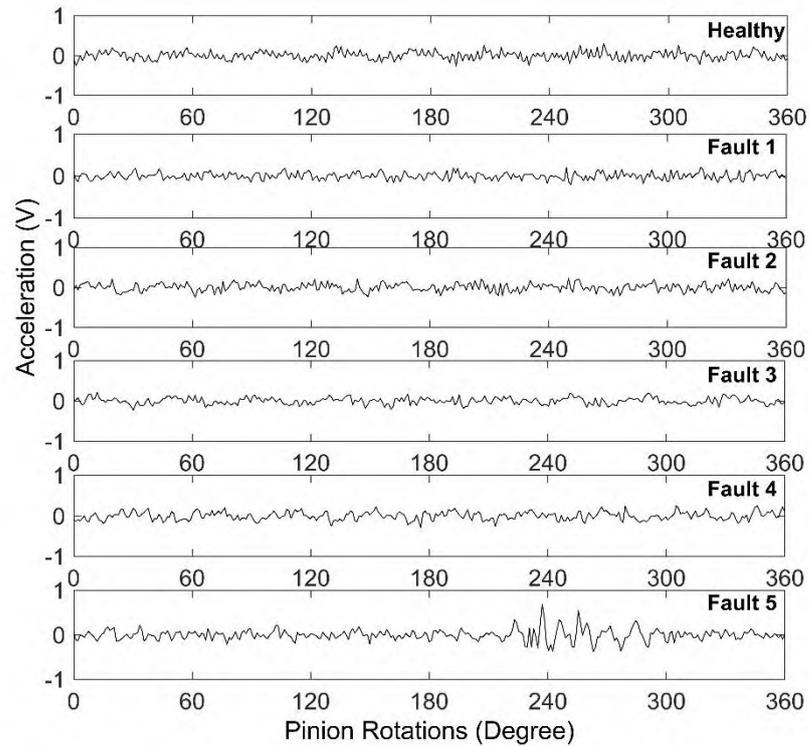


Figure 6. Averaged gear vibrations for one pinion rotation in time domain during the development of pitting fault.

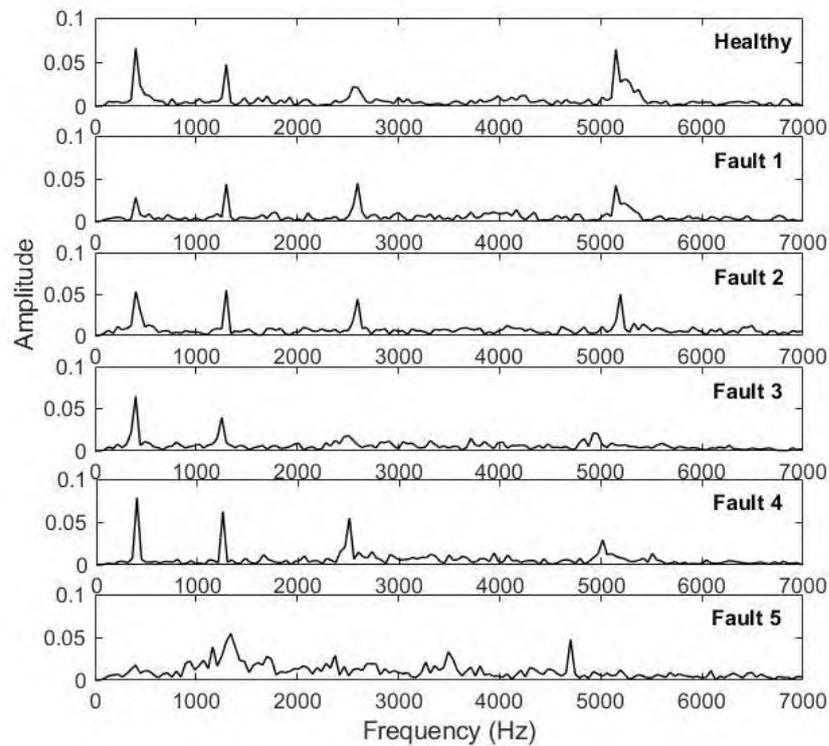


Figure 7. Averaged gear vibrations frequency domain during the development of pitting fault.

3. RESULTS AND DISCUSSION

A Local gear pitting dataset with five subsets was obtained for five different operating situations in this experiment. The total number of samples is 2670, with 445 samples for each condition. The number of input units in a deep neural network is governed by the sample length. As a result, sample length is a critical metric. On the one hand, a deep neural network with a small number of input units may not be capable enough. A deep neural network with a large number of input units, on the other hand, is likely to do better in diagnostics. However, it may result in a larger computing cost.

The vibration data, called sample consists of 335 sampling points because this data length is equal to one revolution of the drive pinion gear. The time synchronous average (TSA) method is applied to the raw data and each sample consists of three turns of the pinion gear. Indications of gear failure are caused by the vibrations of the gear surfaces that come into contact with each other, along with the rotation of the gears. Therefore, 335 vibration components related to faults tend to be periodic. Moreover, gear vibrations also contain many other undesirable components. The TSA method helps to remove unwanted components called noise from vibration data and makes the vibration signals caused by the rotation of the gearbox more pronounced. The model created with the training data was run with the test data to verify the effectiveness of the suggested method.

In this research, the raw vibration data to diagnose the local gear pitting faults without any signal preprocessing or manual feature extraction has been used. Raw signal data is reconstructed with CAE, SAE, VAE and used for fault classification. To begin, the sample length determines the input unit number. Second, the number of hidden layers and units in the proposed deep model should be large enough to support feature learning. Finally, the following hidden layer's unit number is lower than the preceding layer's, allowing feature learning to be viewed as a data compression process. The deep structure of the proposed approach is determined as 335-256-128-256-335 in the model. All of the tests were run on a Windows 10 computer with an Intel Core i5 processor and 8GB of RAM. Python3.6 is the compiler and language that was used to create and test these algorithms. The optimization parameters of the model are

given in Table 2.

Table 2. Optimization parameters

Data set	Pitting Fault Data
Input Units	335
Output Units	6
Epoch	100
Batch size	32

The model was tested using 534 samples. Table 3 shows the diagnostic recognition rates of test samples from various gear states. The test samples of each gear pitting condition have a high diagnostic recognition rate, as can be observed. Table 3 illustrates how the SAE model can accurately diagnose the various gear pitting defects, with an accuracy rate of more than 97%. The test set and the training set exhibit comparable accuracy levels, and no over-fitting phenomenon is evident.

Table 3. Diagnosis results

Autoencoder Model	Training Set	Testing Set
CAE	0,9101	0,9101
SAE	0,9757	0,9756
VAE	0,9082	0,9082

The confusion matrix shows each class's performance more clearly. This helps to evaluate the performance and verify the efficiency of the classifier. The multi-class confusion matrix displays all of the conditions' classification results in detail, including classification accuracy and misclassification fault. The confusion matrix's ordinate axis represents the actual classification label, while the horizontal axis predicts the classification label. As shown in Figure 8, the identification rate for healthy gear conditions, as well as for Fault 3 and Fault 5, is 100%. Additionally, the recognition rate for other gear pitting conditions is also very high, exceeding 98%. This demonstrates the effectiveness of the diagnostic method in accurately detecting and classifying various gear pitting faults. Fault 2 is misjudged as the gear of the Fault 1 and Fault 1 is misjudged as the gear of the Fault 2. This could be as a result of identical initial vibration signals caused by very close gear pitting. In Li's work [22], a combination of Sparse Autoencoder (SAE) and Gauss-Binary Restricted Boltzmann Machine (GBRBM) was used for fault diagnosis, effectively leveraging SAE's capabilities for feature compression and GBRBM's ability to handle continuous input data. In our study, SAE, along with CAE and VAE, was independently used to analyze raw vibration data without employing additional feature extraction methods, demonstrating the stand-alone capability of these autoencoders in learning deep feature representations directly from raw vibration signals.

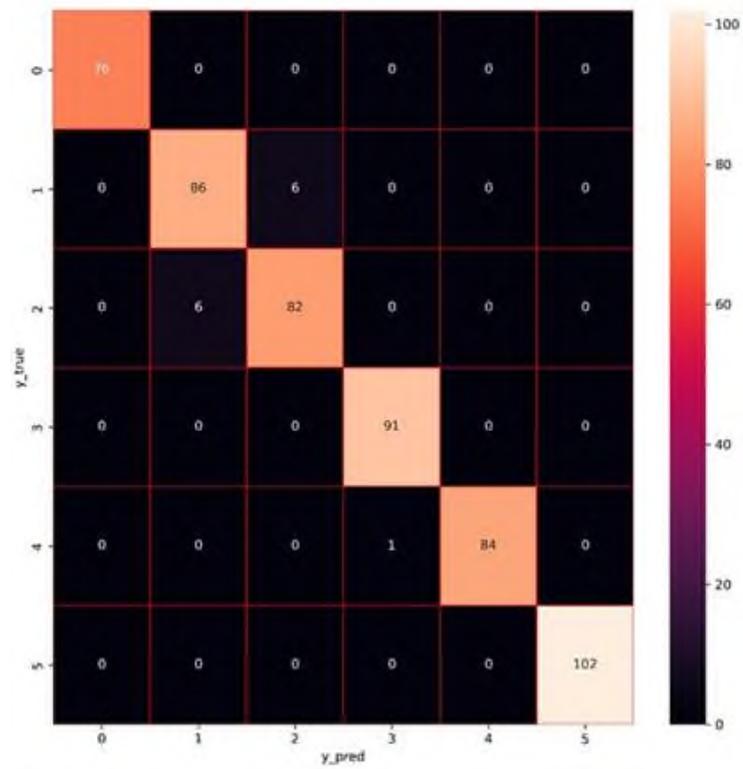


Figure 8. Multi-class confusion matrix of the SAE

As can be seen from Figure 9, the identification rate of condition healthy, Fault 3,4 and 5 are 100%. The recognition rate of the Fault 2 is also above 98%. Fault 1 is 0.50% misjudged as the gear of the Fault 2. Here, it seems that the CAE model fails to distinguish 2 faults with similar characteristics.

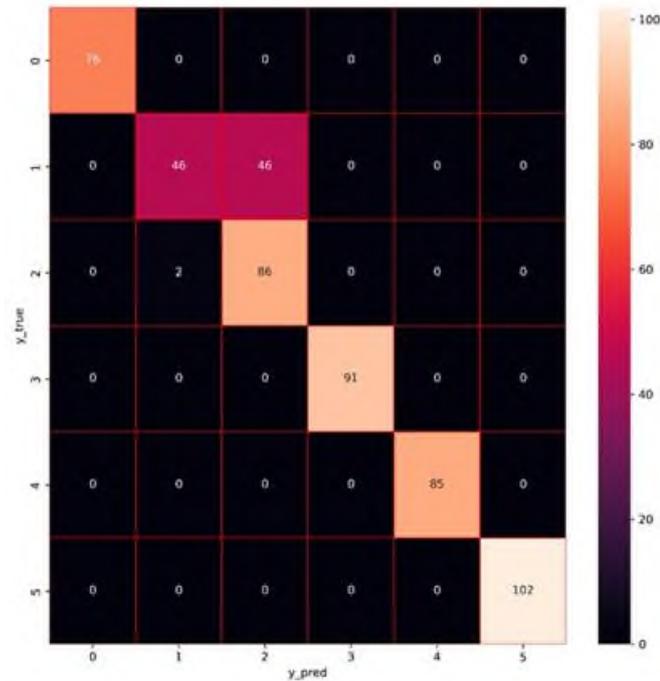


Figure 9. Multi-class confusion matrix of the CAE

As illustrated in Figure 10, the identification rate for condition Fault 5 is 100%. This indicates that the diagnostic method used was highly effective in accurately detecting and classifying this specific fault

without any errors. The recognition rate of faults 1,2 and 4 are close to 90%. Fault 4 is misjudged as the gear of the Fault 3 and Fault 3 is misjudged as the gear of the Fault 4. Also Fault 1 is misjudged as the gear of the Fault 2 and healthy. Fault 2 is misjudged as the gear of the Fault 1 and healthy. We can say that the variational autoencoder is not very successful in distinguishing faults with similar characteristics.

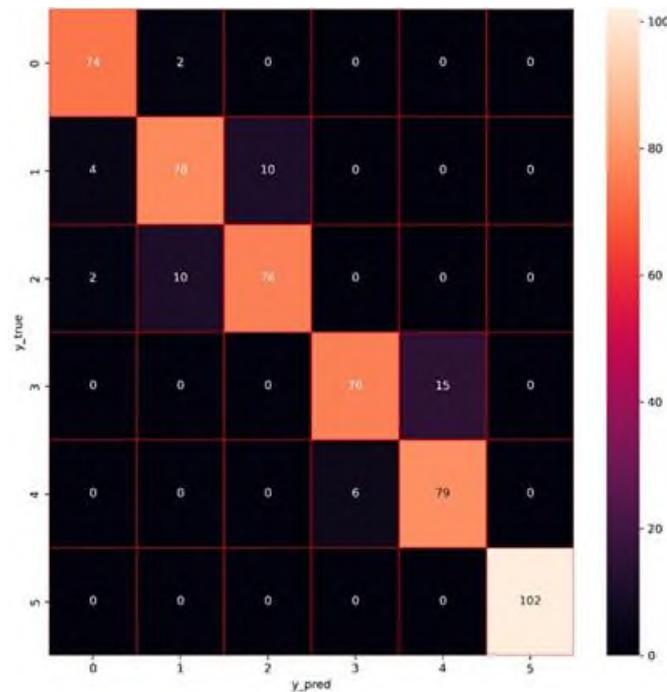


Figure 10. Multi-class confusion matrix of the VAE

To validate the feature extraction capability of the presented method, t-SNE was used to visualize the extracted features. t-SNE is a nonlinear dimensionality reduction technique that maps high-dimensional, interrelated data into a lower-dimensional space. By doing so, it preserves the local structure and relationships between the data points, allowing complex patterns in high-dimensional datasets to be visualized in 2 or 3 dimensions. The resulting reduced data provides an intuitive representation of the original data's structure, making it particularly useful for visualizing and interpreting clusters and patterns that may not be easily discernible in higher dimensions. Six different pitting faults can be clearly grouped, as shown in Figure 11. In this graph, we can see faults with similar characteristics more clearly. It can be seen that Fault 1 and Fault 2, which are the most similar to each other, overlap. We have seen in the confusion matrices above that these faults are the ones where AE models make the most mistakes.

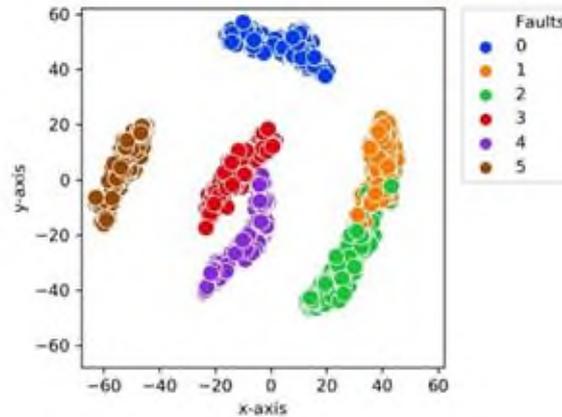


Figure 11. Two-dimensional features of gear pitting conditions

4. CONCLUSIONS

In many industrial machines, gears are among the most important parts. Pitting is one of the most frequent gear problems and is typically hard to find. When the local gear pitting defect is not early detected, it can cause serious machine failures. This article compares the performance of different autoencoder models for the local pitting fault diagnosis method. The progressive learning process in deep learning architectures allows for the direct extraction of meaningful deep features from raw vibration signals. As more hidden layers are added, the ability to differentiate between the features corresponding to various gear states improves substantially. This makes deep learning a robust and efficient approach for deep feature extraction and gear fault diagnosis. AE can be seen as a way of transforming representation.

Through experiments, the methods' efficacy was evaluated. The gear pitting defect detection rate yielded findings that were higher than 90%. The SAE approach has the highest accuracy when compared to other auto encoder techniques, and it can satisfy the classification and detection needs for gears with various pitting conditions. Comparison between auto encoder models showed that the SAE model has better performance than the CAE and VAE models in automatically extracting adaptive features. This article provides automatic feature extraction, but network parameters such as the number of hidden layers, the number of neurons in each layer, and the learning rate depend on the human experience. Extending the developed approach to the identification of other mechanical component failures, such as bearing and motor failures, will be a focus of future research.

This study focused on the comparison of three autoencoder models (SAE, CAE, and VAE) for diagnosing gear pitting faults using raw vibration data. While these methods demonstrated strong performance, the application of traditional machine learning algorithms such as Random Forest or XGBoost was not explored. Given the well-separated nature of the dataset as seen in Figure 11, these methods could also achieve high accuracy. Future studies should include such algorithms to benchmark their performance against deep learning approaches, providing a broader perspective on fault diagnosis methodologies.

Declaration of Ethical Standards

The authors of this article state that the materials and methods employed in the study do not necessitate approval from an ethics committee or any specific legal permissions.

Credit Authorship Contribution Statement

1st Author, 2nd Author, and 3rd Author designed the study. 2nd Author and 3rd Author performed

the experiments and collected data. 1st Author derived the models and analyzed the data. 1st Author and 2nd Author wrote the manuscript in consultation with 3rd Author

Declaration of Competing Interest

The authors declare that they have no known competing financial interests or personal relationships that could have appeared to influence the work reported in this paper.

Data Availability

Data sharing not applicable

REFERENCES

- [1] H. Öztürk, M. Sabuncu, and I. Yesilyurt, "Early detection of pitting damage in gears using mean frequency of scalogram," *J. Vib. Control*, vol. 14, 2008, pp. 469–484.
- [2] M. Amarnath and S. K. Lee, "Assessment of surface contact fatigue failure in a spur geared system based on the tribological and vibration parameter analysis," *Measurement*, vol. 76, 2015, pp. 32–44.
- [3] S. Bagavathiappan, B. B. Lahiri, T. Saravanan, et al., "Infrared thermography for condition monitoring – a review," *Infrared Phys. Technol.*, vol. 60, 2013, pp. 35–55.
- [4] Z. Peng, N. J. Kessissoglou, and M. Cox, "A study of the effect of contaminant particles in lubricants using wear debris and vibration condition monitoring techniques," *Wear*, vol. 258, 2005, pp. 1651–1662.
- [5] Y. E. Karabacak and N. G. Özmen, "Common spatial pattern-based feature extraction and worm gear fault detection through vibration and acoustic measurements," *Measurement*, vol. 187, 2022.
- [6] B. Hizarci, R. C. Ümütlü, H. Ozturk, and Z. Kiral, "Vibration region analysis for condition monitoring of gearboxes using image processing and neural networks," *Exp. Tech.*, vol. 43, no. 6, 2019, pp. 739–755.
- [7] T. Toutountzakis, K. T. Chee, and M. David, "Application of acoustic emission to seeded gear fault detection," *NDT E Int.*, vol. 38, no. 1, 2005, pp. 27–36.
- [8] K. Worden and J. M. Dulieu-Barton, "An overview of intelligent fault detection in systems and structures," *Struct. Health Monit.*, vol. 3, no. 1, 2004, pp. 85–98.
- [9] M. H. Benbouzid, "A review of induction motors signature analysis as a medium for faults detection," *IEEE Trans. Ind. Electron.*, vol. 47, no. 5, 2000, pp. 984–993.
- [10] S. Ebersbach and Z. Peng, "Expert system development for vibration analysis in machine condition monitoring," *Expert Syst. Appl.*, vol. 34, 2008, pp. 291–299.
- [11] F. P. G. Marquez, A. M. Tobias, J. M. P. Perez, and M. Papaalias, "Condition monitoring of wind turbines: Techniques and methods," *Renew. Energy*, vol. 46, 2012, pp. 169–178.
- [12] J. Zarei, M. A. Tajeddini, and H. R. Karimi, "Vibration analysis for bearing fault detection and classification using an intelligent filter," *Mechatronics*, vol. 24, 2014, pp. 151–157.
- [13] A. K. S. Jardine, D. Lin, and D. Banjevic, "A review on machinery diagnostics and prognostics implementing condition-based maintenance," *Mech. Syst. Signal Process.*, vol. 20, 2006, pp. 1483–1510.
- [14] R. C. Ümütlü, B. Hizarci, H. Ozturk, and Z. Kiral, "Classification of pitting fault levels in a worm gearbox using vibration visualization and ANN," *Sādhanā*, vol. 45, no. 1, 2020, pp. 1–13.
- [15] B. Samanta, "Gear fault detection using artificial neural networks and support vector machines with genetic algorithms," *Mech. Syst. Signal Process.*, vol. 18, 2004, pp. 625–644.
- [16] B. Samanta, K. R. Al-Balushi, and S. A. Al-Araimi, "Artificial neural networks and support vector machines with genetic algorithm for bearing fault detection," *Eng. Appl. Artif. Intell.*, vol. 16, no. 7, 2003, pp. 657–665.
- [17] W. Huang, H. Sun, Y. Liu, and W. Wang, "Feature extraction for rolling element bearing faults

- using resonance sparse signal decomposition," *Exp. Tech.*, vol. 41, 2017, pp. 251–265.
- [18] B. Zhang, H. Wang, Y. Tang, et al., "Residual useful life prediction for slewing bearing based on similarity under different working conditions," *Exp. Tech.*, vol. 42, 2018, pp. 279–289.
- [19] G. E. Hinton and R. R. Salakhutdinov, "Reducing the dimensionality of data with neural networks," *Science*, vol. 313, no. 5786, 2006, pp. 504–507.
- [20] X. Chen, A. Ji, and G. Cheng, "A novel deep feature learning method based on the fused-stacked AEs for planetary gear fault diagnosis," *Energies*, vol. 12, no. 23, 2019.
- [21] Y. Qu, M. He, J. Deutsch, and D. He, "Detection of pitting in gears using a deep sparse autoencoder," *Appl. Sci.*, vol. 7, no. 5, 2017.
- [22] J. Li, X. Li, D. He, and Y. Qu, "A novel method for early gear pitting fault diagnosis using stacked SAE and GBRBM," *Sensors*, vol. 19, no. 4, 2019.
- [23] X. Li, J. Li, Y. Qu, and D. He, "Gear pitting fault diagnosis using integrated CNN and GRU network with both vibration and acoustic emission signals," *Appl. Sci.*, vol. 9, no. 4, 2019.
- [24] X. Li, J. Li, Y. Qu, and D. He, "Semi-supervised gear fault diagnosis using raw vibration signal based on deep learning," *Chin. J. Aeronaut.*, vol. 33, no. 2, 2020, pp. 418–426.
- [25] J. Li, X. Li, D. He, and Y. Qu, "Unsupervised rotating machinery fault diagnosis method based on integrated SAE–DBN and a binary processor," *J. Intell. Manuf.*, vol. 31, 2020, pp. 1899–1916.
- [26] S. Khan and T. Yairi, "A review on the application of deep learning in system health management," *Mech. Syst. Signal Process.*, vol. 107, 2018, pp. 241–265.
- [27] S. Rifai, G. Mesnil, P. Vincent, et al., "Higher order contractive auto-encoder," in *Proc. Joint Eur. Conf. Mach. Learn. Knowl. Discov. Databases*, Berlin, Heidelberg, Sep. 2011, pp. 645–660, Springer.
- [28] C. Shen, Y. Qi, J. Wang, et al., "An automatic and robust features learning method for rotating machinery fault diagnosis based on contractive autoencoder," *Eng. Appl. Artif. Intell.*, vol. 76, 2018, pp. 170–184.
- [29] L. Wen, L. Gao, and X. Li, "A new deep transfer learning based on sparse auto-encoder for fault diagnosis," *IEEE Trans. Syst. Man Cybern. Syst.*, vol. 49, no. 1, 2017, pp. 136–144.
- [30] Y. Qi, C. Shen, D. Wang, J. Shi, X. Jiang, and Z. Zhu, "Stacked sparse autoencoder-based deep network for fault diagnosis of rotating machinery," *IEEE Access*, vol. 5, 2017, pp. 15066–15079.
- [31] A. He and X. Jin, "Deep variational autoencoder classifier for intelligent fault diagnosis adaptive to unseen fault categories," *IEEE Trans. Rel.*, vol. 70, no. 4, 2021, pp. 1581–1595.
- [32] S. Liu, H. Jiang, Z. Wu, and X. Li, "Rolling bearing fault diagnosis using variational autoencoding generative adversarial networks with deep regret analysis," *Measurement*, vol. 168, 2021, p. 108371.
- [33] D. Addo, S. Zhou, J. K. Jackson, G. U. Nneji, H. N. Monday, K. Sarpong, and C. A. Owusu-Agyei, "Evae-net: An ensemble variational autoencoder deep learning network for COVID-19 classification based on chest X-ray images," *Diagnostics*, vol. 12, no. 11, 2022, p. 2569.
- [34] H.Öztürk. Gearbox health monitoring and fault detection using vibration analysis. PhD Thesis, Graduate School of Natural and Applied Sciences of Dokuz Eylül University, Türkiye (2006)