Gear Pitting Fault Detection: Leveraging Anomaly Detection Methods

Ozan Can Alper[1], Hatice Doğan[2], and Hasan Öztürk[3]

[1]The Graduate School of Natural and Applied Sciences, Dokuz Eylül University, Izmir, Turkey
alper.ozancan@ogr.deu.edu.tr
[2]Department of Electrical and Electronics Engineering, Dokuz Eylül University, Izmir, Turkey
hatice.dogan@deu.edu.tr
[3]Department of Mechanical Engineering, Dokuz Eylül University, Izmir, Turkey
hasan.ozturk@deu.edu.tr

Abstract

Monitoring and maintaining the health of gears is crucial for the efficient and safe operation of mechanical systems. Due to harsh operating conditions, gear failures such as wear, pitting, and breakage are common. This study investigates the effectiveness of unsupervised and semi-supervised deep anomaly detection methods for identifying distributed pitting defects in gears using vibration data. In the experimental setup, gear faults of varying severity were created, and vibration data from helical gears were recorded for each level of fault severity. Autoencoders (AE), Variational Autoencoders (VAE), and Deviation Networks (DevNet) have been utilized to detect faulty gears. This study presents the performance of these techniques in predictive maintenance based on the availability of fault data.

1. Introduction

Gears are circular mechanical devices that transmit torque and speed at a desired rate and play a vital role in a wide range of industrial applications. Since gearboxes operate in harsh conditions, different defects may occur in gears including but not limited to wear and tear, pitting, fracture, surface fatigue, and tooth breakage [1]. These defects can negatively impact the performance, efficiency, and reliability of gears and the machinery in which they are used. Therefore, the early detection of defects is critical to maintaining operational reliability, preventing catastrophic failures, optimizing performance, extending equipment life, and improving safety in industrial environments. For the purpose of detecting anomalies and potential issues within gear systems, a diverse array of signals and data sources are employed, with vibration signals being the most popular among them [2, 3, 4, 5]. Changes in vibration patterns, such as increased amplitude or frequency, can indicate problems with gears, so time domain [6], frequency domain [7, 8, 9] and time-frequency domain methods [8, 10, 11] are used to analyze the vibration signal.

Supervised machine learning and data-driven methods are often used to detect gear faults through vibration analysis. However, during the training phase of these models, it is not always possible to simulate all possible types of faults or fault scenarios that may occur in real industrial systems. This limitation is due to the diversity and complexity of potential faults and the difficulty of reproducing all possible fault conditions in a controlled environment. Anomaly detection methods, which mostly use only normal data to train the model, may be more suitable for detecting faults in gears.

Anomaly detection is a data analysis technique used to identify patterns or data points within a dataset that deviate significantly from expected or normal behavior. In this context, an anomaly is a data point or observation that does not fit the typical pattern or distribution of the majority of the data [12]. Anomaly detection can be used for various applications such as credit card fraud detection [13, 14], cyber-security intrusion detection [15], healthcare abnormal behavior detection [16], traffic scene event detection [17, 18], etc.

Various anomaly detection methods have been employed in the literature for the detection of gearbox failures. A deep learning technique combining Long Short-Term Memory (LSTM) and Support Vector Machine (SVM) techniques is proposed in [19] to separate anomalous data from normal vibration signals obtained during an endurance test of a reduction gearbox. A method using adaptive thresholding and twin support vector machines to detect anomalies in wind turbine gearboxes is proposed in [20]. In [21], a method that is the combination of variational autoencoder (VAE) and LSTM network for anomaly and trend detection in industrial robot gear condition monitoring is proposed and in [22], the periodicity-enhanced robust principal component analysis (PRPCA) approach is used to detect the anomalies in the encoder data.

In this paper, unsupervised and semi-supervised anomaly detection methods are used for the detection of distributed pitting faults in gears, which is a challenging problem [11]. Distributed pitting faults refer to small pits or craters that occur on all the tooth surfaces, rather than in a limited area of the gear. Determining if a fault exists or emerging in distributed pitting might be challenging because the gear vibration accelerations of the different fault severities are similar to each other and the healthy gear acceleration [11]. For the detection of the distributed pitting fault, the unsupervised Autoencoders (AE), Variational Auto Encoders (VAE), and semi-supervised Deviation Networks (DevNet), which use very few anomaly data in the training set, are used. To compare the performances of the methods in detecting faults of varying severity, different levels of fault severity were simulated by introducing different numbers of pits into the tooth surface of helical gears within the test rig. Vibration data were collected for each fault level.

2. Experimental Setup

A fault monitoring test rig, as shown in Figure 1, has been set up, consisting of a two-stage industry-type helical gearbox, 2.2 kW DC load motor, and 2.2 kW AC drive motor, which are connected by belt pulley mechanisms to eliminate the undesirable effects such as of AC - DC motors and misalignment. The input

shaft position was measured using a 5V DC ME4-S12L-PA type inductive sensor, which generates a single pulse for each rotation. Additionally, a speed controller for an AC drive motor was employed to enable the gearbox to run between 0 and 3000 rpm. Table 1 lists the technical details for the gearbox's first and second stages.

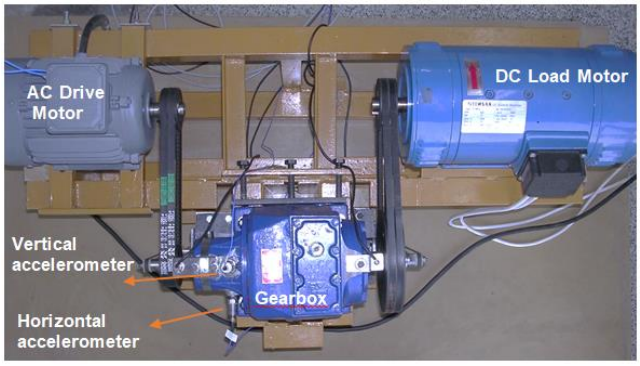

Fig. 1. Experimental setup [23]

Table 1. Specifications of the two-stage gearbox

	First Stage	Second Stage
Number of teeth	29/40	13/33
Normal module (mm)	1.25	2.5
Pressure angle (°)	20	20
Helix angle (°)	30	15

The vibration signals produced by the gears were collected using two PCB 352A76 type accelerometers, which operated between 5 and, 16000 Hz. On the housings for the input shaft bearings, these accelerometers were positioned perpendicular to one another. Using an NI (National Instrument) data acquisition system, LabVIEW 7.0 software, and 15 kHz sampling, the raw vibration data from accelerometers was recorded on a computer. If the gears are angularly misaligned, the surface contact stress may not be uniform across the face width of the mating teeth. In such cases, the likelihood of potential future pitting is greatest on those tooth surfaces that experience contact stresses in excess of permissible limits.

Initially, a circular pit, measuring approximately 0.7 mm in diameter and 0.1 mm in depth, was created on all gear surfaces using an electro-erosion machine to simulate the slightest fault (F-1). The number of pits was then increased by one to represent the progression of distributed pitting faults thought to be caused by the presence of angular misalignment (F-2, F-3). Finally, all the surfaces of the gear were covered with pits to simulate the most severe fault (F-4). The images of the faulty gears are shown in Figure 2.

Raw vibration data was collected continuously over the course of 1337 pinion revolutions. Figure 3 shows an example of a vibration signal from each class in both the time and frequency domains. The gearbox was disassembled and reassembled each time to simulate a distributed pitting defect on the pinion gear's teeth surfaces. This situation, or manufacturing faults, could have caused a modulation that manifests itself as repetitive fluctuations for each pinion rotation.

Fig. 2. Gear with pitting faults of different severities

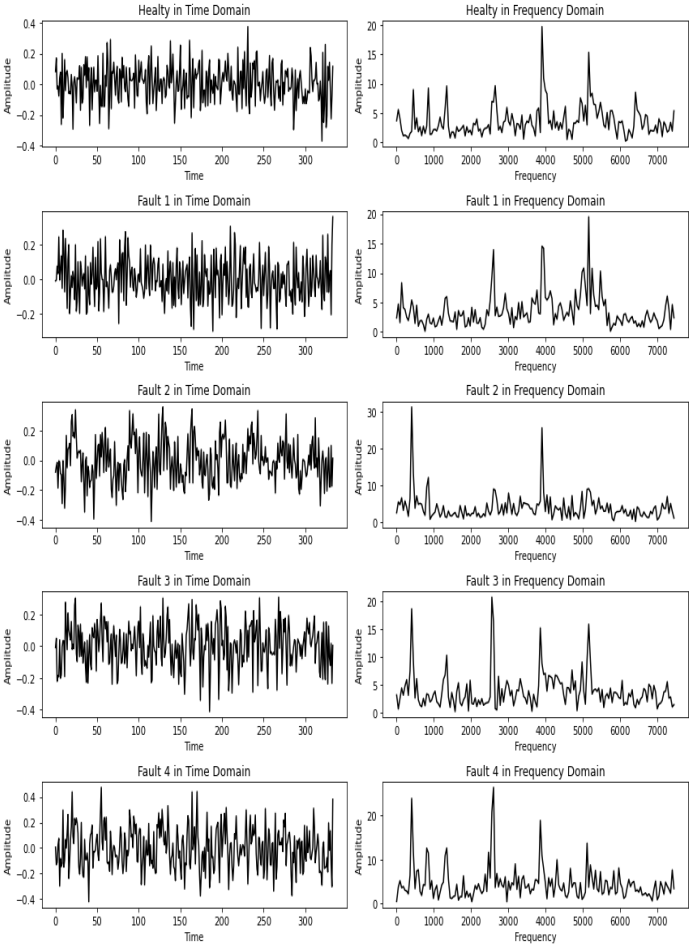

Fig. 3. Raw acceleration signals and their corresponding spectra

3. Methods

To detect distributed pitting faults, unsupervised methods such as autoencoders (AE) and variational AE can be used, as well as the semi-supervised approach known as DevNet.

3.1 Autoencoders (AE)

Autoencoders, initially introduced by Hinton and the PDP group in the 1980s [24], are used to compress high-dimensional input data into a lower-dimensional latent space while preserving key features [25]. Deep autoencoders are the extension of traditional autoencoders, designed to capture more complex and hierarchical patterns in data. The encoder part of the autoencoder compresses the input into a latent space representation, while the decoder part expands this low-dimensional latent space representation to obtain the output that is equal to the input data. The main purpose is to minimize the reconstruction error between input and output data. In gear fault detection, the autoencoder is trained only with vibration signals of healthy gears. If the input vibration signal contains patterns that deviate from the learned healthy behavior, the reconstruction error will be large and the autoencoder will identify these signals as anomalies that may indicate the presence of a gear fault.

3.2 Variational Autoencoders (VAE)

A particular kind of autoencoder known as a variational autoencoder (VAE) adds a probabilistic component to the encoding procedure. The input that is passed through the encoder in VAEs is encoded as a probability distribution, typically a Gaussian distribution. VAEs produce the mean and variance values that characterize the probability distribution as the encoder output rather than a fixed code, as is the case with conventional autoencoders [26]. A random sample is taken from the distribution created by these mean and variance values in order to create a code. The VAE can generate new data by sampling from this distribution because this sampling process enables the VAE to learn a distribution over the latent space. VAE, whose main structure is shown in Figure 4, captures the statistical structure of the data in the latent space.

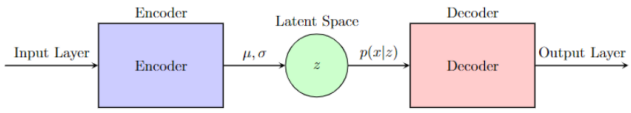

Fig. 4. Variational Autoencoder Architecture [27]

3.3 Deviation Networks (DevNet)

A deviation network is a semisupervised anomaly detection framework described in [28] that gives directly the anomaly score as an output using a few labeled anomaly data along with normal data. In contrast to other deep anomaly detection techniques that use data reconstruction to learn new representations, DevNet optimizes the anomaly scores directly, rather than optimizing feature representations, because it is built to learn the anomaly scores. The DevNet framework comprises three distinct sub-blocks. The first block is a neural network structure responsible for modeling the scalar anomaly function, denoted as ϕ, and generating an anomaly score as output for a given input. The second block is responsible for determining the mean, μ_R, and associated standard deviation, σ_R, of a given set of normal data. The final block takes inputs from the first and second blocks, namely ϕ, μ_R, and σ_R, and constructs the deviation loss function to guide the optimization process. During optimization, the objective is to drive anomaly scores towards a reference value, μ_R, for normal data inputs, while aiming to obtain significantly deviated anomaly scores from the upper tail.

4. Results

4.1 Data Preprocessing

The dataset consists of five classes, one of which is healthy. Each class in the dataset consists of a 3x450000 matrix, where the matrix rows correspond to data collected from the horizontal acceleration sensor, vertical acceleration sensor, and encoder output data. Utilizing encoder data, the horizontal and vertical sensor data were segmented into discrete windows per rotational cycle, yielding a total of 1337 different datasets per class each containing 334 data points. A filtering procedure was applied to limit the frequency components between 1000 Hz and 5400 Hz to reduce noise-induced distortion [29]. In order to benefit from the exceptional capabilities of deep learning methods in image analysis, horizontal and vertical acceleration data, each consisting of 334 data points, were normalized and written into a 26x26 matrix to create images [30, 31]. The last 8 pixels were padded with zeros to reach the final image size of 26x26, which is the size of the square matrix that can be created from 668 data points. As a result of this procedure, each data class contains 1337 images. Representative images for each class are shown in Figure 5.

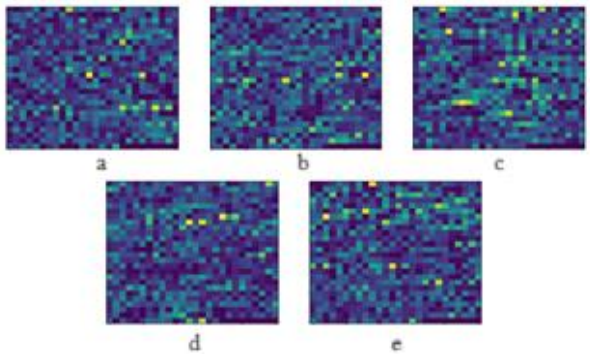

Fig. 5. a) Healthy b) Fault-1 c) Fault-2 d) Fault-3 e) Fault-4

4.2 Application

The architecture of AE, including two blocks as encoder and decoder, is as follows:

The encoder takes an input of shape (26, 26, 1), which represents a 26x26 image. It consists of two convolutional layers (Conv2D) with 32 and 64 filters, respectively. The output of the convolutional layers is flattened to a vector of shape (10816). It then passes through two dense (fully connected) layers with 1024 and 128 units, respectively. The final layer (latent space) has 50 units.

The decoder takes an input of shape (50), which is the output of the encoder. It consists of a dense layer with 10816 units, which reshapes the data to (13, 13, 64). Then, it uses three transpose convolutional layers (Conv2DTranspose) to upsample the data, gradually increasing the spatial dimensions. The final layer produces an output of shape (26, 26, 1), which aims to reconstruct the original 26x26 image.

The autoencoder was trained using only 70% of the data from the healthy class. The reconstruction error values were calculated on a test set containing the remaining 30% of the healthy class data and data from the faulty classes. The created histogram

according to the reconstruction values of the test data is shown in Figure 6. Input data selected from the test set was classified as "Healthy" if reconstruction_error ≤ 459 and "Faulty" if reconstruction_error > 459.

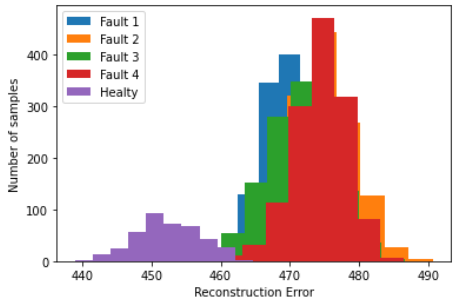

Fig. 6. Histogram of Reconstruction Errors using AE

VAE is designed to have the same structure as AE. The main difference is that the latent space in VAE consists of two vectors of 50 dimensions, representing mean and variance.

The same training and test procedure as for AE was applied, and the histogram of the test data according to the reconstruction values is shown in Figure 7. Input data selected from the test set were classified as "Healthy" if reconstruction_error ≤ 460 and "Faulty" if reconstruction_error > 460.

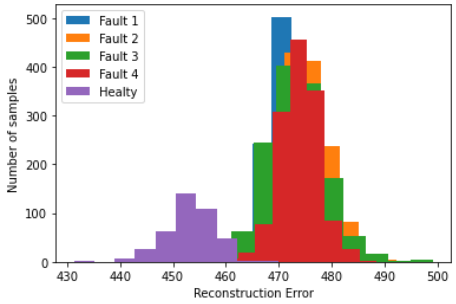

Fig. 7. Histogram of Reconstruction Errors using VAE

DevNet requires significantly fewer labeled anomalies for the training stage. To determine the optimal number of anomaly samples for DevNet, the model was trained and tested with different numbers of anomaly samples. The results revealed that 20 samples is the optimal number. Therefore, DevNet was trained by taking 70% of the Healthy class and 5 samples of each fault type (approximately 0.37% of each class), and then tested with the remaining data, the results are shown in Table 2.

Table 2. Accuracies of AE, VAE, and DevNet

	AE Acc. (%)	VAE Acc. (%)	DevNet Acc. (%)
H	91.27	94.26	100
F-1	99.70	99.85	100
F-2	99.85	99.93	99.77
F-3	99.40	99.63	100
F-4	99.63	99.70	100
TA	99.06	99.39	99.95

The columns of Table 2 show the performance of the methods, and the rows show the percentage of correct classification of each class. The bottom row shows the total accuracy (TA) of the models in healthy/faulty discrimination. When the performances are compared, it is seen that the performance of VAE is better than AE, while the performance of DevNet is the best, as expected, since it is a semi-supervised method.

In reality, situations may arise where data is not available for every possible fault type, or new fault types may appear unexpectedly. This raises the question: "Can one fault type effectively represent other fault types? To answer this question, the DevNet model is trained on a dataset containing 20 examples of a single error type and 70% of the Healthy class, rather than a dataset containing 5 examples of each error type. The test set contains all the remaining faulty data and 30% of the healthy class. The results of this approach are shown in Table 3.

Table 3. Accuracy of DevNet trained with only one fault type

Acc. (%)	H	F-1	F-2	F-3	F-4	Total Accuracy
F-1	100	100	79.13	88.78	89.23	90.00
F-2	99.75	65.37	99.92	98.95	99.18	91.45
F-3	99.75	74.50	93.49	99.70	99.78	92.40
F-4	100	75.77	73.22	97.91	99.77	87.55

The rows in Table 3 show the performance of DevNet in detecting Healthy and different fault types when only instances from a single fault class were used in the training set. For instance, in the first row, only 20 instances of the F-1 class were employed to train DevNet. In the test set, while data belonging to the Healthy class and F-1 fault were identified with 100% accuracy, the detection accuracy for F-2, F-3, and F-4 faults stood at 79.13%, 88.18%, and 89.23%, respectively. The discrimination accuracy between healthy and faulty is 90%.

Upon closer examination of the table, it becomes evident that the system excels in detecting the fault type utilized in its training, exhibiting a notably high success rate. Furthermore, it is notable that the system's performance is considerably better when trained with moderate faults (F-2 and F-3) compared to training with mild and extreme faults (F-1 and F-4). On the other hand, when the system is trained with severe faults (F-3 and F-4), it tends to incline towards misclassifying mild faults (F-1 and F-2) as healthy.

Consequently, the outcomes of this study lead to the recommendation of training the DevNet architecture with moderate faults. However, even with this adjustment, the DevNet model does not attain the same level of performance as the AE and VAE models when trained with only a single type of fault.

5. Conclusions

This study delved into the effectiveness of unsupervised and semi-supervised deep anomaly detection methods for identifying distributed pitting defects in gears using vibration data. Various deep learning techniques, including Autoencoders, Variational Autoencoders, and Deviation Networks, were employed to detect faulty gears. The findings of this study highlight the performance of these techniques in predictive maintenance based on the availability of fault data.

To create the dataset, gear failures of different severities were generated in the designed experimental setup. The vibration data of the helical gears were recorded for each different fault severity. To take advantage of the exceptional capabilities of deep learning methods in image analysis, one-dimensional vibration signals recorded in two axes were converted into images by writing them into a matrix. These images were used as input. While AE and VAE were trained with only healthy data, DevNet was trained with only 20 anomaly instances in addition to healthy data. The results indicate that VAE model demonstrated better performance than AE model, whereas DevNet outperformed AE models as expected when trained with a dataset covering all possible faults due to its semi-supervised structure. However, the DevNet model did not reach the same level of performance as the AE and VAE models when trained with only one type of fault.

6. References

[1] Kumar, Anil, et al. "Latest developments in gear defect diagnosis and prognosis: A review." *Measurement* 158 (2020): 107735.

[2] Zhang, Shouhua, et al. "State of the art on vibration signal processing towards data-driven gear fault diagnosis." *IET Collaborative Intelligent Manufacturing* 4.4 (2022): 249-266.

[3] Miltenović, Aleksandar, et al. "Detection and Monitoring of Pitting Progression on Gear Tooth Flank Using Deep Learning." *Applied Sciences* 12.11 (2022): 5327.

[4] Wei, Yu, et al. "A review of early fault diagnosis approaches and their applications in rotating machinery." *Entropy* 21.4 (2019): 409.

[5] Tama, Bayu Adhi, et al. "Recent advances in the application of deep learning for fault diagnosis of rotating machinery using vibration signals." *Artificial Intelligence Review* 56.5 (2023): 4667-4709.

[6] Praveenkumar, T., et al. "Pattern recognition based on-line vibration monitoring system for fault diagnosis of automobile gearbox." *Measurement* 114 (2018): 233-242.

[7] Kar, Chinmaya, and A. R. Mohanty. "Vibration and current transient monitoring for gearbox fault detection using multiresolution Fourier transform." *Journal of Sound and Vibration* 311.1-2 (2008): 109-132.

[8] Fan, Xianfeng, and Ming J. Zuo. "Gearbox fault detection using Hilbert and wavelet packet transform." *Mechanical Systems and Signal Processing* 20.4 (2006): 966-982.

[9] Liu, Bao, S. Riemenschneider, and Y. Xu. "Gearbox fault diagnosis using empirical mode decomposition and Hilbert spectrum." *Mechanical Systems and Signal Processing* 20.3 (2006): 718-734.

[10] Wang, W. J., and P. D. McFadden. "Application of wavelets to gearbox vibration signals for fault detection." *Journal of sound and vibration* 192.5 (1996): 927-939.

[11] Öztürk, Hasan, İsa Yeşilyurt, and Mustafa Sabuncu. "Detection and advancement monitoring of distributed pitting failure in gears." *Journal of Nondestructive Evaluation* 29 (2010): 63-73.

[12] Chandola, Varun, Arindam Banerjee, and Vipin Kumar. "Anomaly detection: A survey." *ACM computing surveys (CSUR)* 41.3 (2009): 1-58.

[13] Aleskerov, Emin, Bernd Freisleben, and Bharat Rao. "Cardwatch: A neural network based database mining system for credit card fraud detection." *Proceedings of the IEEE/IAFE 1997 computational intelligence for financial engineering (CIFEr)*. IEEE, 1997.

[14] Ahmed, Mohiuddin, Abdun Naser Mahmood, and Md Rafiqul Islam. "A survey of anomaly detection techniques in financial domain." *Future Generation Computer Systems* 55 (2016): 278-288.

[15] Kumar, Vipin. "Parallel and distributed computing for cybersecurity." *IEEE Distributed Systems Online* 6.10 (2005).

[16] Liao, Zhifang, et al. "A visual analytics approach for detecting and understanding anomalous resident behaviors in smart healthcare." *Applied Sciences* 7.3 (2017): 254.

[17] Riveiro, Maria, Mikael Lebram, and Marcus Elmer. "Anomaly detection for road traffic: A visual analytics framework." *IEEE Transactions on intelligent transportation systems* 18.8 (2017): 2260-2270.

[18] Yuan, Yuan, Dong Wang, and Qi Wang. "Anomaly detection in traffic scenes via spatial-aware motion reconstruction." *IEEE Transactions on Intelligent Transportation Systems* 18.5 (2016): 1198-1209.

[19] Vos, Kilian, et al. "Vibration-based anomaly detection using LSTM/SVM approaches." *Mechanical Systems and Signal Processing* 169 (2022): 108752.

[20] Dhiman, Harsh S., et al. "Wind turbine gearbox anomaly detection based on adaptive threshold and twin support vector machines." IEEE Transactions on Energy Conversion 36.4 (2021): 3462-3469.

[21] Nentwich, Corbinian, and Gunther Reinhart. "A combined anomaly and trend detection system for industrial robot gear condition monitoring." Applied Sciences 11.21 (2021): 10403.

[22] Liang, Kaixuan, et al. "An encoder information-based anomaly detection method for planetary gearbox diagnosis." Measurement Science and Technology 31.4 (2020): 045015.

[23] Öztürk, Hasan. "Gearbox health monitoring and fault detection using vibration analysis." (2006).

[24] Hinton, Geoffrey E. "Learning translation invariant recognition in a massively parallel networks." *International conference on parallel architectures and languages Europe*. Berlin, Heidelberg: Springer Berlin Heidelberg, 1987.

[25] Baldi, Pierre. "Autoencoders, unsupervised learning, and deep architectures." *Proceedings of ICML workshop on unsupervised and transfer learning*. JMLR Workshop and Conference Proceedings, 2012.

[26] Kingma, Diederik P., and Max Welling. "Auto-encoding variational bayes." *arXiv preprint arXiv:1312.6114* (2013).

[27] Addo, Daniel, et al. "Evae-net: An ensemble variational autoencoder deep learning network for covid-19 classification based on chest x-ray images." *Diagnostics* 12.11 (2022): 2569.

[28] Pang, Guansong, Chunhua Shen, and Anton van den Hengel. "Deep anomaly detection with deviation networks." *Proceedings of the 25th ACM SIGKDD international conference on knowledge discovery & data mining*. 2019.

[29] Hizarci, B., et al. "Vibration region analysis for condition monitoring of gearboxes using image processing and neural networks." *Experimental Techniques* 43 (2019): 739-755.

[30] LeCun, Yann, Yoshua Bengio, and Geoffrey Hinton. "Deep learning." *nature* 521.7553 (2015): 436-444.

[31] Goodfellow, Ian, et al. "Generative adversarial nets." *Advances in neural information processing systems* 27 (2014).