UDC 004.93

BEARING FAULT DETECTION BY USING AUTOENCODER CONVOLUTIONAL NEURAL NETWORK

Kysarin M. K. – Student of the Faculty of Information Technologies, State University of Trade and Economics, Kyiv, Ukraine.

ABSTRACT

Context. Bearings are an important part for the functioning of various means of transportation. They have the property of wear and failure, which requires high-quality and timely detection of faults. Failures are not always easy to detect, so the use of traditional detection methods may not be effective enough. The use of machine learning methods well-suited to the task can effectively solve the problem of detecting bearing faults. The object of study is the process of non-destructive diagnosis of bearings. The subject of study is methods of selecting hyperparameters and other optimization for building a diagnostic model based on a neural network according to observations.

Objective. The goal of the work is to create a model based on a neural network for detecting bearing faults based on the ZSL.

Method. A proposed filter smooths the data, preserving key characteristics such as peaks and slopes, and eliminates noise without significantly distorting the signal. A normalization method vibration data is proposed, which consists of centering the data and distributing the amplitude within optimal limits, contributing to the correct processing of this data by the model architecture. A model based on a neural network is proposed to detect bearing faults by data processing and subsequent binary classification of their vibrations. The proposed model works by compressing the vibration data into a latent representation and its subsequent recovery, calculating the error between the recovered and original data, and determining the difference between the errors of healthy and faulty bearing vibration data. The Zero-Shot Learning machine learning method involves training, validating the model only on healthy vibration data, and testing the model only on faulty vibration data. Due to the proposed machine learning method, the model based on a neural network is able to detect faulty bearings present in the investigated fault class and theoretically new fault classes, that is, the model can detect different classes of data that it did not see during training. The architecture of the model is built on the convolutional and max-pooling layers of the encoder, and the reverse convolutional layers for the decoder. The best hyperparameters of the model are selected using a special method.

Results. Using the Pytorch library, a model capable of binary classification of healthy and faulty bearings was obtained through training, validation, and testing in the Kaggle software environment.

Conclusions. Testing of the constructed model architecture confirmed the model's ability to classify healthy and fault bearings binaryly, allowing it to be recommended for use in practice to detect bearing faults. Prospects for further research may include testing the model through integration into predictive maintenance systems for timely fault detection.

KEYWORDS: bearing fault, autoencoder, convolutional neural network, zero-shot learning, binary classification.

ABBREVIATIONS

AE is an autoencoder;

AE-CNN is an autoencoder convolutional neural network;

CWRU is a Case Western Reserve University;

CNN is a convolutional neural network;

DAE is a deep autoencoder;

DNN is a deep neural network;

FFT is a fast Fourier transform;

GAE is a graph autoencoder;

GAF is a Gramian angular field algorithm;

GCN is a model that specializes in learning the node characteristics of graph data;

GPMS is a Green Power Monitoring Systems company;

LSTM is a long short-term memory;

MMN is a min-max normalization data method;

MSE is a mean squared error;

PCA is a principal component analysis method;

SGF is a Savitzky-Golay filter;

SSAE is a stacked sparse autoencoder;

STFT is a technology that has been developed to overcome the limitations of FFT;

TPE is a Tree-structured Parzen Estimator method;

VAE is a variational autoencoder;

XG is a Xavier Glotor initialization;

ZSL is a zero-shot learning machine learning scenario.

NOMENCLATURE

α is a hyperparameter of the learning rate of the AE-CNN model;

a is a minimum data value for the MMN;

b is a maximum data value for the MMN;

b_k are the coefficients of the polynomials for the SGF;

B_1 is a exponential smoothing coefficient of the first momentum for the Adam optimization algorithm;

B_2 is a exponential smoothing factor for the second momentum for the Adam optimization algorithm;

c_b is a probability distribution for bad combinations of hyperparameters of the AE-CNN model;

c_g is a probability distribution for good combinations of hyperparameters of the AE-CNN model;

c_{in} is a number of filters at the input to the convolutional layer;

c_{out} is a number of filters at the output of the convolutional layer;

d is a data provided by the AE-CNN model;

ε is a small value in the Adam optimization algorithm to prevent division by zero;

E is a MSE;

DOI 10.15588/1607-3274-2025-2-10

E_{test} is a set of test error values;

E_{train} is a set of training error values;

E_{val} is a set of validation error values;

f is a samples of bad vibration data with faults;

g is a samples of good vibration data;

γ is a threshold separating good and bad values of E;

h is a set of hyperparameters of the AE-CNN model;

k is a shift index of the smoothing window for the SGF;

k_c is a size of the one-dimensional convolutional layer filter;

l is a shift parameter for the AE-CNN convolutional one-dimensional layer;

L_{train} is a training loss value;

L_{val} is a validation loss value;

m is a width of the window for the SGF;

$M()$ is an AE-CNN model structure;

n is a sample of vibration data;

n_{in} is a number of input connections (neurons) included in the layer;

n_{out} is a number of output connections (neurons) coming out of the layer;

N is a number of samples of vibration data;

p is a max-pooling filter size;

$P()$ is a probability distribution;

t is a number of time points in the vibration data sample;

U is an uniform distribution;

W is a set of controlled (adjusted) weights of the AE-CNN model;

x is an one point in the original sample;

$\hat{x}$ is an one point in a recovered sample;

x_{index} is an index of one point in the original sample;

X_{test} is a test samples of bad vibration data with faults;

X_{train} is a training samples of good vibration data;

X_{val} is a validation samples of good vibration data;

y is an optimization iteration number, the current step of the Adam optimization algorithm;

z is a MSE threshold.

INTRODUCTION

The study deals with the development of a special neural network to detect faults in bearings that affect their reliability and safety. The problem is that bearings can wear or break over time, and these failures are often accompanied by vibrations that are not always easy to detect. Depending on different situations, different methods of detecting bearing faults may be effective [16]. Detection of such faults using traditional methods can be difficult and not always effective [1], [15], especially if the faults are still at an early stage. Therefore, it is necessary to develop an automated system capable of analyzing the vibration signals and determining whether the bearing is good or bad.

One of the powerful options for detecting the fault features is AE-CNN, which can be trained only on healthy vibration samples, ensuring their generalization and extracting knowledge from data, without losing the ability to further classify faulty data samples.

The object of study is the process of non-destructive diagnosis of bearings.

Vibration data are used to diagnose the condition of these bearings and identify possible faults.

The process of pre-processing the vibration data to be suitable for AE-CNN is very careful. This is caused by the fact that the initial vibration data is very noisy due to the influence of other sound factors from the environment. Therefore, to realize the classification ability of AE-CNN, a filter and normalization must be applied to the vibration data. The size and quality of the training sample used can significantly affect the training and accuracy of the AE-CNN model. Therefore, reducing the size of the samples, and ensuring the preservation of its main properties, is necessary to improve the quality of AE-CNN and the speed of its construction.

The subject of study is methods of selecting hyperparameters and other optimization for building a diagnostic model based on a neural network according to observations.

The purpose of the work is to create a model based on a neural network for detecting bearing faults based on the ZSL.

1 PROBLEM STATEMENT

Suppose $g_N(t)$ and $f_N(t)$ are given. Training, validation, and test samples are formed from data prepared for model processing: X_{train}, X_{val}, X_{test}.

For given, respectively, good and bad bearing classes {0,1}, the task of detecting the difference between bearing classes through the AE-CNN model can be represented as (if $M(h, W, d)$ with $d \in \{X_{train}, X_{val}, X_{test}\}$ then class($M(h, W, d) \in \{0,1\}$)) $\rightarrow$ opt. h of the best model $M()$ is determined by a special selection method, and W is adjusted through the optimizer of the learning process.

2 REVIEW OF THE LITERATURE

Aiming to address the issue of strong background noise in bearing failures and the lack of evident fault features, this paper [2] proposes a fault diagnosis method that combines Savitzky-Golay Gram angle field feature enhancement with ResNet18. The acquired signal is segmented, and the segmented signal is subjected to Butterworth high-pass filtering to extract the high-frequency component of the signal containing fault information.

The optimized SGF and adaptive spectrum editing are proposed to detect the fault feature of the rolling element bearing under low-speed and variable-speed conditions [3].

This manuscript elaborates on the development of a VAE-CNN model, designed for the fault diagnosis of rolling bearings. By amalgamating the CNN model's superior capacity for representing vibration signal data with the VAE model's robustness to data noise, the proposed VAE-CNN model excels in scenarios where only a limited amount of observational data is available at the initial stages of rolling bearing operation. The VAE-

DOI 10.15588/1607-3274-2025-2-10

CNN model achieves over 90% accuracy in diagnosing different fault types at various speeds [4].

This study [5] proposes a GAE-based approach for fusing node features in an Euclidean dataset. The primary advantage of the proposed approach lies in its adaptive ability to capture the complex structure of the dataset and fuse node features using a GAE, effectively extracting the latent features. Normalization is a vital pre-processing step as it addresses the issue of varying magnitudes among different base detectors, which makes direct comparison and combination challenging. To enhance the convergence speed of the algorithm, in the proposed approach normalized the output of the base detectors.

This approach [6] uses both an unsupervised and a supervised model. First, the current signal in the time domain is segmented with respect to the fundamental frequency, and then a DAE is trained with the normal state data to estimate the approximation of the system function. The residual signal is then calculated from the difference between the raw and estimated signals produced by the AE for all conditions, which helps to extract discriminative features from the current signal data without any labels. Finally, a two-layer CNN is built with the residual signals to identify bearing faults. The experiments were performed 100 times by randomly selecting the training/testing dataset, and the result showed good stable convergence with high accuracy.

This study [7] proposes a novelty detection and fault diagnosis method for bearing fault recognition and diagnosis based on a hybrid DAE. The method uses one model to accomplish two tasks, detecting new faults and classifying known faults. To address the challenge of requiring a large number of training samples in existing deep learning methods, this study combines the unsupervised training characteristics of AEs with the powerful feature extraction ability of CNNs, adopting a semi-supervised training method that can learn fault features from both labeled and unlabeled samples, reducing the required sample size for training.

The proposed method [8] incorporates a GAF-based image generation technique from a 1-dimensional current signal for a 2-layer CNN model to construct a data-driven intelligent fault-diagnosis approach for bearings. As the current signal is affected by surrounding noises, it becomes very difficult to extract the fault signatures manually. When the data are converted into the polar coordinates for image transformation, the different bearing-condition data create distinctive patterns, which helps the CNN model to easily extract the necessary high-level features. In all considering operating conditions, this proposed GAF and 2-dimensional CNN-based approach can attain good accuracy.

This study [9] is based on deep learning methods for bearing fault diagnosis. Firstly, a CNN model is designed for end-to-end bearing fault diagnosis. Then, considering the presence of strong noise in actual working conditions, a bearing fault diagnosis model based on AE-CNN is proposed to achieve bearing fault diagnosis under noisy conditions. The experiment results on the CWRU demonstrate the effectiveness of the proposed model. The method proposed in this study can be used for fault diagnosis of bearings under noise conditions and has engineering practical value.

This article [10] highlights the advantages of deep learning models, particularly AE and CNNs, for fault diagnosis in bearing systems. Unlike traditional machine learning algorithms, which require extensive manual feature design, AE and CNN can automatically extract high-level features from large-scale, unlabeled data. AE, though simple, benefits from enhancements such as stacking layers and adding noise. CNN excels in feature extraction due to its unique architecture and translation invariance.

This study [11] indicates that experimental results show that VAE is a more competent and promising dimensionality reduction tool than PCA.

This work presents SSAE-DNN, which in combination with a complex envelope spectrum for inputs performs fault diagnosis of rotary machines when there are fluctuations of the shaft speed. In the proposed scheme, vibration signals related to different health conditions of a motor bearing are preprocessed using the complex envelope signal. In the proposed method, information obtained by the stacked autoencoders from the defect frequency, as well as its principle harmonics present in the complex envelope spectrum for a given fault, makes it possible to classify faults with varying speeds. The efficiency of the proposed scheme was validated using rotating machine bearing data for four different shaft speeds. The minimum average classification accuracy for every experiment was 90%, which demonstrates that the proposed scheme can also classify faults when fluctuations of the shaft speed exist.

3 MATERIALS AND METHODS

Let's divide all vibration data into good and bad, respectively, as follows: $g_N(t)$ and $f_N(t)$. To reduce the noise in the data arising from environmental factors, the SGF is applied to the data, which is defined by the following formulas:

$$g'_N(t) = \sum_{k=-\frac{m}{2}}^{\frac{m}{2}} b_k g_N(t+k) ,$$

$$f'_N(t) = \sum_{k=-\frac{m}{2}}^{\frac{m}{2}} b_k f_N(t+k) ,$$

where

$$k = \frac{m-1}{2} . \qquad (1)$$

The filtered data must have one dimension because if the data is submitted to AE-CNN at different scales of variation, it can complicate the data processing of the

DOI 10.15588/1607-3274-2025-2-10

model and significantly degrade the classification results. Taking into account this feature, MMN is used, which is determined by the formulas:

$$s_g = \frac{g'_N(t) - \min(g'_N(t))}{\max(g'_N(t)) - \min(g'_N(t))}(b-a)+a ,$$

$$s_f = \frac{f'_N(t) - \min(f'_N(t))}{\max(f'_N(t)) - \min(f'_N(t))}(b-a)+a .$$

where b and a are determined according to research requirements.

Normalized data ready for submission to AE-CNN is divided into training, validation, and test samples: X_{train}, $X_{val}=\{s_g\}$; $X_{test}=\{s_f\}$.

Following the ZSL principle, the training and validation data consists only of the working vibration samples, and the test data consists only of the faulty vibration samples.

Given the periodicity of the vibration data, the healthy data samples have stable healthy features along their entire length, as well as the fault features in the corresponding faulty samples are observed consistently along their entire length. Taking these data properties into account, the data is reduced to an optimal size to optimize the data processing speed of the AE-CNN model.

The architecture of the AE-CNN model is divided into an encoder, a latent space, and a decoder.

The encoder architecture is proposed in four stages. Each stage has a convolutional one-dimensional layer and a maximum pooling layer. A convolutional one-dimensional layer performs a convolution operation where filters are moved over d and compute scalar products.

The data in each stage of the encoder is processed according to the formulas:

$$c = \max_{i=0,\dots,p-1}(Wd+l) ,$$

where

$$d = \{n_s \mid s \in [x_{index}, x_{index}+k_c-1]\} .$$

l is initially initialized to zero. This is a standard approach, as the offset value can be adjusted quickly during training without much damage to the result, so zero initialization works well.

W in AE-CNN models from the beginning of processing are initialized through XG according to the defined formula:

$$W = U(-\sqrt{\frac{6}{n_{in}+n_{out}}}, \sqrt{\frac{6}{n_{in}+n_{out}}}) ,$$

where

$$n_{in} = c_{in}k_c ,$$

$$n_{out} = c_{out}k_c .$$

To reduce the dimensionality of the data to the state of the latent representation, a linear layer is used, which has one matrix of weights, which determines the connections between input and output neurons. Each output neuron is connected to all input neurons.

The decoder uses similar formulas as the encoder, but in reverse order, since the main purpose of the decoder is to recover or reconstruct the input data from the latent representation.

To optimize W and α of the model, the Adam optimization algorithm is used, which is determined by the following formulas:

$$W_{y+1} = W_y - \alpha \frac{\hat{m}_y}{\sqrt{\hat{v}_y}+\varepsilon} ,$$

$$\hat{m}_y = \frac{m_y}{1-\beta_1^y} ,$$

$$\hat{v}_y = \frac{v_y}{1-\beta_2^y} ,$$

$$m_y = \beta_1 m_{y-1} + (1-\beta_1)\nabla W_y ,$$

$$v_y = \beta_2 v_{y-1} + (1-\beta_2)(\nabla W_y)^2 ,$$

$$\nabla W_y = \frac{\partial E}{\partial W_y} .$$

The initial values of m_{y-1} and v_{y-1} are initialized with zeros.

The research uses the method of TPE hyperparameter selection, which works according to the following formulas:

$$h_{new} = \arg\max \frac{c_g(h)}{c_b(h)} ,$$

$$c_g(h) = P(h \mid E \le \gamma) ,$$

$$c_b(h) = P(h \mid E > \gamma) .$$

This method divides h into good and bad groups using a certain γ, and builds models to estimate the probability of each group. This makes it possible to find new h that have a higher probability of belonging to good values, thus optimizing the choice of h. γ is usually chosen as a certain quantile.

In this case, E is the last validation error at the end of the current trial h through TPE.

The AE-CNN model goes through the processes of training, validation, and updating W and is tested on X_{test}.

DOI 10.15588/1607-3274-2025-2-10

At all stages, E determined by the following formulas is used:

$$e_{train} = \frac{\sum_{n=1}^{t}(x_{train} - \hat{x}_{train})^2}{t},$$

$$e_{val} = \frac{\sum_{n=1}^{t}(x_{val} - \hat{x}_{val})^2}{t},$$

$$e_{test} = \frac{\sum_{n=1}^{t}(x_{test} - \hat{x}_{test})^2}{t}.$$

Let the sets of errors be defined as E_{train}, E_{val}, $E_{test}=\{e_1, e_2, …, e_N\}$ which is calculated on the last epoch of the final trained AE-CNN model. To obtain the difference between the signs of good and bad bearings, the difference between E_{val} and E_{test} was used. z is determined by the formula:

$$z = \max(E_{val}).$$

All samples with $E>z$ are considered faulty.

In the final step, E_{test} is compared with z, and if $e_{test}>z$, then the corresponding sample is considered faulty.

4 EXPERIMENTS

The data for the study supplied by GMPS owner Eric Bechhoefer has 1158 samples of vibration data, of which 865 are good and 293 are bad. Each sample has 93752 values recorded over 5 minutes.

Experimentally, m was selected with a value of 24, b_k with a value of 13, and k was calculated with a value of 11 according to formula (1).

Based on the periodicity in the samples, in each SGF-treated faulty sample, pulses with increased amplitude of the fault-determining signal are observed. The research recorded that the first pulse appears on average at 513 values of the faulty sample, the period during which each subsequent pulse is recorded is 935 values.

In the process of normalization of vibration data according to formula (2), b with a value of 0.3 and a with a value of –0.3 were used, these values were selected taking into account that after normalization of healthy samples and normalization of faulty samples based on the calculated metrics of healthy samples, the maximum module values of faulty samples will not exceed the range from –1 to 1.

Based on the signs of periodicity of both healthy and faulty vibration samples, it was decided to reduce the samples to the first 1500 values to optimize the duration of data processing in the AE-CNN model.

Empirical studies show that the best results are obtained if the authors use 20–30% of the data for testing, and the remaining 70–80% ofThe data for training [12].

Adhering to the ZSL principle, good samples in the amount of 605 and 260, 70% and 30% of g_N, respectively, were selected for the training and validation processes, and all 293 faulty samples were selected for the testing process.

In the best encoder architecture selected through TPE, the parameters c_{in}=1, c_{out}=13 were used on the first one-dimensional convolutional layer; on the second layer c_{in}=13, c_{out}=22; on the third layer c_{in}=22, c_{out}=57; on the fourth layer c_{in}=57, c_{out}=94. Each one-dimensional convolutional layer used k_c=3, p=2, and a padding parameter equal to 1 to preserve dimensionality and include zero values at the edges, allowing the filters to better treat edge values that would otherwise be treated less than other values in the center. Output padding with one extra value, without significantly affecting learning, was added to control the decoder layers' output sizes.

A linear layer is used to provide an initial size for the data supplied to the decoder. An unflatten layer is used to ensure the corresponding data size expected by the reverse convolutional layers of the decoder.

The best reverse convolutional layers of the decoder selected through TPE have the parameters of the first layer c_{in}=94, c_{out}=57; on the second layer c_{in}=57, c_{out}=22; on the third layer c_{in}=22, c_{out}=13; on the fourth layer c_{in}=13, c_{out}=1. 18 epochs selected through TPE were used for training and validation processes. The best α of the AE-CNN model determined through TPE was set to 0.0004. It should be noted that TPE determines the initial possible α, after which the Adam optimizer adjusts the determined α. The batch size for simultaneous sample processing by the AE-CNN model was determined to be 16.

5 RESULTS

The results of the AE-CNN model training and validation processes are shown in Table 1. The table shows L_{train} and L_{val} for 18 epochs. z and mean of E_{test} values are also represented.

Table 1 shows the satisfactory reduction of L_{train} and L_{val} over epochs. It can be seen that the mean $E_{test}>z$, so the method of detecting faults by comparing E of the AE-CNN model of the reproduced samples with z works in practice.

Table 1 – Results of training and validation of the AE-CNN model

Epochs	L_{train}	L_{val}
1	0.0097	0.0092
2	0.0091	0.0091
3	0.0091	0.0090
4	0.0089	0.0086
5	0.0078	0.0076
6	0.0064	0.0064
7	0.0052	0.0056
8	0.0045	0.0052
9	0.0041	0.0050
10	0.0039	0.0048
11	0.0037	0.0048
12	0.0036	0.0047
13	0.0035	0.0047
14	0.0034	0.0046
15	0.0033	0.0046
16	0.0033	0.0046
17	0.0032	0.0046
18	0.0032	0.0046
z	0.00467195	
Mean of E_{test}	0.007458317	

DOI 10.15588/1607-3274-2025-2-10

L_{train} and L_{val} are calculated at the end of each epoch and are dynamic during all epochs due to changes in the performance of the AE-CNN model.

Table 2 shows the mean values of E_{val} and E_{test} calculated from groups of 16 samples with the corresponding number to batch size showing a clear difference between them. The last row of Table 2 represents the mean values for the E_{val} and E_{test} columns.

Table 2 – The results of calculating the mean values of E_{val} and E_{test} for each batch group

Batch group	E_{val}	E_{test}
1	0.0045924	0.00741719
2	0.00449713	0.00747784
3	0.00457025	0.00725762
4	0.00458623	0.00775483
5	0.00447544	0.00727971
6	0.00461615	0.00741939
7	0.00446275	0.00729698
8	0.00461744	0.00755094
9	0.00463127	0.00701465
10	0.00448145	0.00788275
11	0.00452703	0.00733029
12	0.00454546	0.00676255
13	0.0046516	0.0077562
14	0.00459267	0.00735691
15	0.00456514	0.00795846
16	0.00467195	0.00786724
17	0.00454972	0.00778848
18		0.00705316
19		0.00748284
Mean of E	0.0045667106	0.0074583173

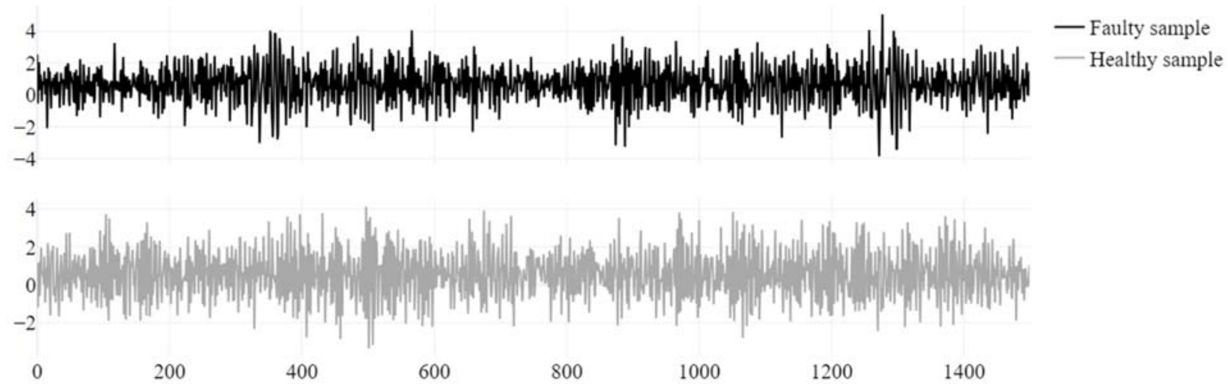

Figure 1 – The example of two faulty and healthy samples without SGF-MMN treatment

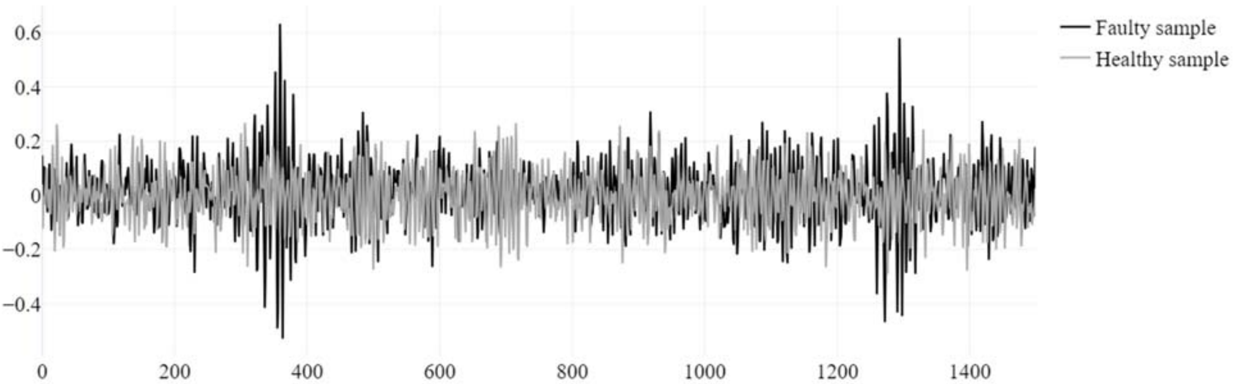

Figure 2 – The example of two SGF-MMN-treated faulty and healthy samples

DOI 10.15588/1607-3274-2025-2-10

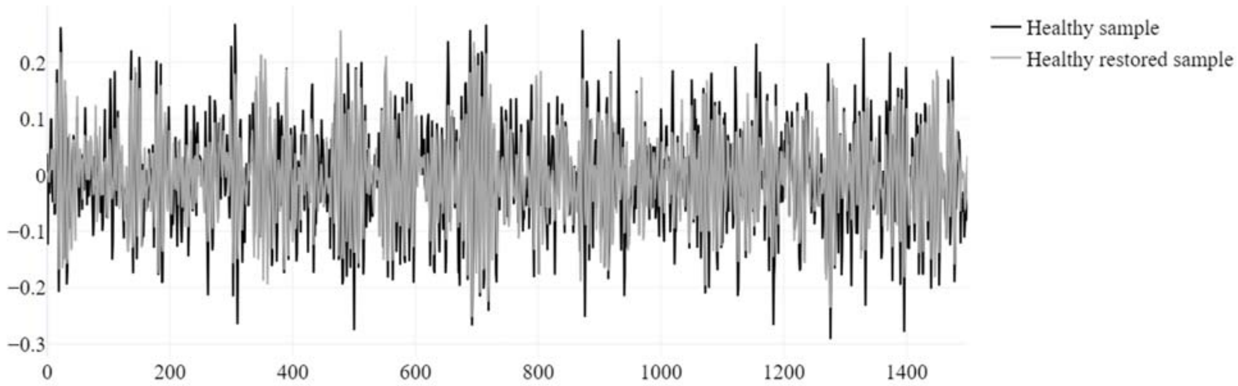

Figure 3 – The example of a healthy SGF-MMN-treated sample and its recovery from the AE-CNN model

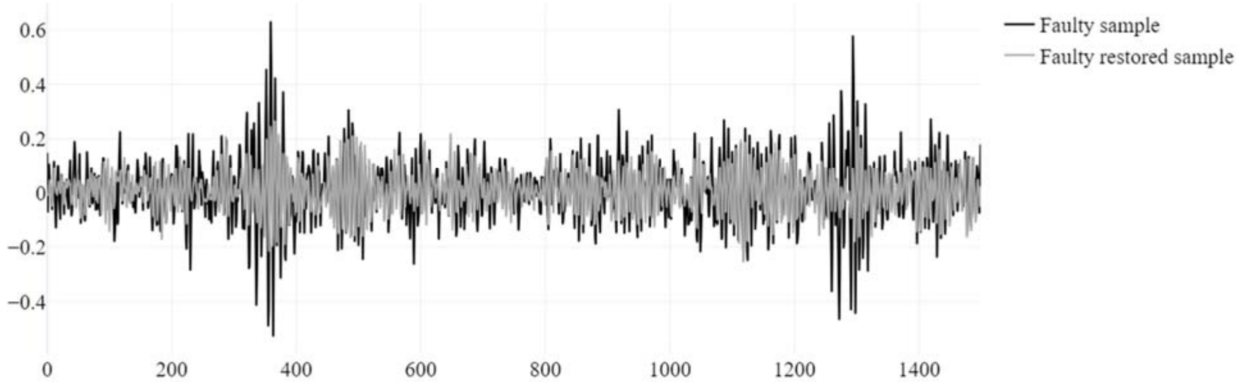

Figure 4 – The example of a faulty SGF-MMN-treated sample and its recovery from the AE-CNN model

Fig. 1 shows that the faulty and healthy samples without proper data pre-processing have blurred class characteristics, which can complicate the perception of the AE-CNN model and worsen its classification ability.

Fig. 2 shows that the SGF-MMN-treated vibrations of healthy and faulty samples mainly differ only in the characteristic pulses that are present in the faulty sample and absent in the healthy sample.

Fig. 3 shows that the AE-CNN model restores the healthy sample with satisfactory accuracy, highlighting the main details of the healthy sample, which confirms the qualitative performance of the AE-CNN model.

Fig. 4 shows that the AE-CNN model recovers a faulty sample without characteristic pulses due to the fact that the AE-CNN model learned to recover only healthy data features, which is why there is a clear difference between E_{val} and E_{test} in Table 2.

6 DISCUSSION

The dimensionality of the vibration samples was reduced to 1500 values, but still, the AE-CNN model showed good performance results in sample recovery and classification of faulty samples through the comparison of E_{val} and E_{test}. It is assumed that if the AE-CNN model will handle full-size samples, with proper tuning of $M()$ and corresponding h, the AE-CNN model will do well in the classification task. The difference when processing full-size samples will be that the difference between E_{val} and E_{test} will be much larger because the reduced faulty samples have a much smaller number of characteristic pulses than the full-size faulty samples.

It is worth noting that if the bearing faults have small amplitude and low intensity of the characteristic signals [5], that is, if the vibrations of the faulty samples are very similar to the vibrations of the healthy samples and have barely noticeable signs of the fault, the AE-CNN model constructed in this study is likely to be much more difficult to track the characteristics of faulty samples and distinguish them from healthy samples by a slight difference in class features, in this case, the E comparison method may be less effective.

The proposed combinatorial method in the study [5] solves the given problem. The method transforms the original dataset into a graph dataset and uses the feature aggregation module to aggregate the features of neighboring nodes.

Considering that during the SGF processing of the data provided for this study, only the features of the classes shown in Fig. 2 were detected, the AE-CNN model was built for the task of binary classification based on these features of the classes of healthy and faulty samples. It is also worth noting that the SGF parameters selected for processing the data provided in this study are appropriate only for this data and may vary when properly applied to other data.

DOI 10.15588/1607-3274-2025-2-10

In the study [7], the hyperparameters of the DAE hybrid network are mainly set based on common experience and are constantly adjusted by trial and error to achieve higher diagnostic accuracy. In this study, this feature was taken into account and the TPE method was used to select h in the AE-CNN model. Also, the method proposed in the study [7] can only detect new faults and cannot distinguish between different new faults, it is not able to represent and distinguish faults in more detail. In this research, the AE-CNN model can detect new faults if they have specific characteristics that are significantly different from the healthy characteristics of the samples. Still, it can also not distinguish between classes of possible faulty samples.

The method proposed in the study [9] is also based on the AE-CNN model and can be used to diagnose bearing faults under noise conditions. The AE-CNN model in this study uses data processed by SGF and MMN, however, in the case of using noisy data processed through MMN only, it is not known what the classification ability of the AE-CNN model will be, but assumed that it will be lower due to the presence of noise in the data, which can distort the characteristics of classes important for classification.

Research [10] mentions the integrated use of different deep learning models, such as LSTM network and CNN, the advantages of each model can be complementary. It is assumed that adding layers of LSTM to the AE-CNN model, given its ability to remember well the short-term and long-term features of the data, can indeed if correctly implemented, help to better capture the dependencies between successive pulses and long-term trends in vibration data, especially if the AE-CNN model processes samples of the full-size of 93752 values.

In the paper [13], a GCN-based LSTM autoencoder with a self-attention model for bearing fault diagnosis was proposed and evaluated using multivariate time series data. The proposed model was found to increase the accuracy of fault diagnosis by combining the GCN layer and the LSTM layer to extract important features from the frequency domain. In the data pre-processing step, data including various fault states and steady states were standardized, while features in the frequency domain were extracted through STFT conversion. A competent implementation of STFT can likely help improve the efficiency of AE-CNN in detecting faulty samples by taking into account frequency components and observing the change in frequency over time.

CONCLUSIONS

The task of detecting bearing faults when applying the machine learning method based on the ZSL principle has been solved.

The scientific novelty of the obtained results is that, for the first time, a machine learning method with the selection of hyperparameters was proposed for building the AE-CNN model based on the best-selected hyperparameters. The hyperparameter selection method divides the combinations of hyperparameters into good and bad using a certain threshold value of the objective function and builds models to estimate the probability of each group. This makes it possible to find new combinations of hyperparameters that have a higher probability of being good, thereby optimizing the choice of hyperparameters.

The practical significance of the obtained results is that, following the ZSL principle, a model based on a neural network was built that detects bearing faults and successfully performs binary classification of healthy and faulty samples of vibration data. The results of the experiment make it possible to recommend the proposed data pre-processing methods and the built model for practical application, as well as to determine the effective conditions for applying the data pre-processing methods and the built model based on a neural network.

Prospects for further research consist of testing the built model based on a neural network on other vibration data of bearings and its implementation in practical operations to detect bearing faults.

ACKNOWLEDGMENTS

Acknowledgments go to GPMS owner Eric Bechhoefer for providing the data.

REFERENCES

1. Mao W., He J., Li Y. et al. Bearing fault diagnosis with auto-encoder extreme learning machine: a comparative study [Electronic resource], *Journal of Mechanical Engineering Science,* 2017, Vol. 231, № 8, pp. 1560–1578. Mode of access: https://journals.sagepub.com/doi/10.1177/0954406216675896
2. Huang Z., Song X., Liao Z. et al. Bearing fault feature enhancement and diagnosis based on Savitzky-Golay filtering Gramian angular field [Electronic resource], *IEEE Access*, 2024, Vol. 12, pp. 87991–88005. Mode of access: https://ieeexplore.ieee.org/document/10570176
3. Chen X., Guo Y., Na J. Encoder signal-based optimized Savitzky-Golay and adaptive spectrum editing for feature extraction of rolling element bearing under low-speed and variable-speed conditions [Electronic resource], *ISA Transactions,* 2024, Vol. 154, pp. 371–388. Mode of access: https://www.sciencedirect.com/science/article/abs/pii/S0019057824003677?via%3Dihub
4. Wang Y., Li D., Li L. et al. A novel deep learning framework for rolling bearing fault diagnosis enhancement using VAE-augmented CNN model [Electronic resource], *Heliyon,* 2024, Vol. 10, № 15. Mode of access: https://doi.org/10.1016/j.heliyon.2024.e35407
5. Wang M., Yu J., Leng H.et al. Bearing fault detection by using graph autoencoder and ensemble learning [Electronic resource], *Scientific Reports*, 2024, Vol. 14, № 1, pp. 1–15. Mode of access: https://www.nature.com/articles/s41598-024-55620-6#citeas
6. Toma R. N., Piltan F., Kim J.-M. A deep autoencoder-based convolution neural network framework for bearing fault classification in induction motors [Electronic resource], *Sensors,* 2021, Vol. 21, № 24. Mode of access: https://www.mdpi.com/1424-8220/21/24/8453
7. Zhao Y., Hao H., Chen Y. et al. Novelty detection and fault diagnosis method for bearing faults based on the hybrid deep autoencoder network [Electronic resource],

DOI 10.15588/1607-3274-2025-2-10

Electronics, 2023,Vol. 12, № 13. Mode of access: https://www.mdpi.com/2079–9292/12/13/2826
8. Toma R. N., Piltan F., Im K. et al. A bearing fault classification framework based on image encoding techniques and a convolutional neural network under different operating conditions [Electronic resource], *Sensors,* 2022, Vol. 22, № 13. Mode of access: https://www.mdpi.com/1424-8220/22/13/4881
9. Yuan F., Xun P., Wang W. et al. Bearing fault diagnosis based on auto-encoder combined with CNN [Electronic resource], *Fuzzy Systems and Data Mining IX : Proceedings of the Fuzzy Systems and Data Mining Conference, 18–20 November 2023 : proceedings*. Xiamen, IOS Press, 2023, Vol. 354, pp. 334–344. Mode of access: https://ebooks.iospress.nl/doi/10.3233/FAIA231039
10. Rao C. A survey of autoencoder and convolutional neural network based methods for fault diagnosis [Electronic resource], *Advances in Engineering Technology Research,* 2024, Vol. 11, № 1, pp. 516–528. Mode of access: https://madison–proceedings.com/index.php/aetr/article/view/2529
11. Zhang S., Zhang S., Wang B.et al. Deep learning algorithms for bearing fault diagnostics – a comprehensive review [Electronic resource], *IEEE Access*, 2020, Vol. 8, pp. 29857–29881. Mode of access: https://ieeexplore.ieee.org/document/8988271
12. Gholamy A., V. Kreinovich, O. Kosheleva Why 70/30 or 80/20 relation between training and testing sets: a pedagogical explanation [Electronic resource], Mode of access: https://api.semanticscholar.org/CorpusID:7467506
13. Lee D. H. Choo, J. Jeong GCN-based LSTM Autoencoder with self-attention for bearing fault diagnosis [Electronic resource], *Sensors,* 2024, Vol. 24, № 15. Mode of access: https://www.mdpi.com/1424-8220/24/15/4855
14. Sohaib M., Kim J.-M. Reliable fault diagnosis of rotary machine bearings using a stacked sparse autoencoder-based deep neural network [Electronic resource], *Shock and Vibration*, 2018, Vol. 2018, № 1. Mode of access: https://onlinelibrary.wiley.com/doi/10.1155/2018/2919637
15. Gu Y., Cao J., Song X.et al. A denoising autoencoder-based bearing fault diagnosis system for time-domain vibration signals [Electronic resource], *Wireless Communications and Mobile Computing,* 2021, Vol. 2021, № 1. Mode of access: https://onlinelibrary.wiley.com/doi/10.1155/2021/9790053
16. Althubaiti A., Elasha F., Teixeira J. A. Fault diagnosis and health management of bearings in rotating equipment based on vibration analysis – a review [Electronic resource], *Journal of Vibroengineering*, 2021, Vol. 24, № 1, pp. 46–74. Mode of access: https://www.extrica.com/article/22100

Received 30.01.2025.
Accepted 10.04.2025.

УДК 004.93

ВИЯВЛЕННЯ НЕСПРАВНОСТІ ПІДШИПНИКА ЗА ДОПОМОГОЮ ЗГОРТКОВОЇ НЕЙРОННОЇ МЕРЕЖІ АВТОКОДУВАЛЬНИКА

Кисарін М. К. – студент факультету інформаційних технологій Державного торговельно-економічного університету, Київ, Україна.

АНОТАЦІЯ

Актуальність. Підшипники є важливою частиною для функціонування різних засобів пересування. Вони мають властивість зношуватися і виходити з ладу, що вимагає якісного і своєчасного виявлення несправностей. Збої не завжди легко виявити, тому використання традиційних методів виявлення може бути недостатньо ефективним. Використання методів машинного навчання, які добре підходять для завдання, може ефективно вирішити проблему виявлення несправностей підшипників. Об'єктом дослідження є процес неруйнівної діагностики підшипників. Предметом дослідження є методи підбору гіперпараметрів та іншої оптимізації для побудови діагностичної моделі на основі нейронної мережі за даними спостережень.

Мета роботи – створення моделі на основі нейронної мережі для виявлення несправностей підшипників на основі ZSL.

Метод. Запропонований фільтр згладжує дані, зберігаючи ключові характеристики, такі як піки та нахили, і усуває шум без істотного спотворення сигналу. Запропоновано метод нормалізації вібраційних даних, який полягає в центруванні даних і розподілі амплітуди в оптимальних межах, що сприяє коректній обробці цих даних архітектурою моделі. Запропоновано модель на основі нейронної мережі для виявлення несправностей підшипників шляхом обробки даних і подальшої двійкової класифікації їх коливань. Запропонована модель працює шляхом стиснення даних про вібрацію в приховане представлення та їх подальшого відновлення, обчислення похибки між відновленими та вихідними даними та визначення різниці між похибками даних про вібрацію справного та несправного підшипників. Метод машинного навчання Zero-Shot Learning передбачає навчання, перевірку моделі лише на справних даних про вібрацію та тестування моделі лише на несправних даних про вібрацію. Завдяки запропонованому методу машинного навчання модель на основі нейронної мережі здатна виявляти несправні підшипники, наявні в досліджуваному класі несправностей і теоретично нові класи несправностей, тобто модель може виявляти різні класи даних, які вона не бачила під час навчання. Архітектура моделі побудована на згорткових рівнях і рівнях максимального об'єднання кодера, а також на зворотних згорткових рівнях для декодера. Спеціальним методом вибираються найкращі гіперпараметри моделі.

Результати. Використовуючи бібліотеку PyTorch, було отримано модель, здатну до бінарної класифікації справних і несправних підшипників, шляхом навчання, валідації та тестування в програмному середовищі Kaggle.

Висновки. Тестування побудованої архітектури моделі підтвердило здатність моделі класифікувати справні та несправні підшипники двійково, що дозволяє рекомендувати її для використання на практиці для виявлення несправностей підшипників. Перспективи подальших досліджень можуть включати тестування моделі шляхом інтеграції в системи прогнозного обслуговування для своєчасного виявлення несправностей.

КЛЮЧОВІ СЛОВА: несправність підшипника, автокодувальник, згорткова нейронна мережа, навчання з нуля, бінарна класифікація.

DOI 10.15588/1607-3274-2025-2-10

ЛІТЕРАТУРА

1. Bearing fault diagnosis with auto-encoder extreme learning machine: a comparative study [Electronic resource] / [W. Mao, J. He, Y. Li et al.] // Journal of Mechanical Engineering Science. – 2017. – Vol. 231, № 8. – P. 1560–1578. – Mode of access: https://journals.sagepub.com/doi/10.1177/0954406216675896
2. Bearing fault feature enhancement and diagnosis based on Savitzky-Golay filtering Gramian angular field [Electronic resource] / [Z. Huang, X. Song, Z. Liao et al.] // IEEE Access. – 2024. – Vol. 12. – P. 87991–88005. – Mode of access: https://ieeexplore.ieee.org/document/10570176
3. Chen X. Encoder signal-based optimized Savitzky-Golay and adaptive spectrum editing for feature extraction of rolling element bearing under low-speed and variable-speed conditions [Electronic resource] / X. Chen, Y. Guo, J. Na // ISA Transactions. – 2024. – Vol. 154. – P. 371–388. – Mode of access: https://www.sciencedirect.com/science/article/abs/pii/S0019057824003677?via%3Dihub
4. A novel deep learning framework for rolling bearing fault diagnosis enhancement using VAE-augmented CNN model [Electronic resource] / [Y. Wang, D. Li, L. Li et al.] // Heliyon. – 2024. – Vol. 10, № 15. – Mode of access: https://doi.org/10.1016/j.heliyon.2024.e35407
5. Wang M. Bearing fault detection by using graph autoencoder and ensemble learning [Electronic resource] / [M. Wang, J. Yu, H. Leng et al.] // Scientific Reports. – 2024. – Vol. 14, № 1. – P. 1–15. – Mode of access: https://www.nature.com/articles/s41598-024-55620-6#citeas
6. Toma R. N. A deep autoencoder-based convolution neural network framework for bearing fault classification in induction motors [Electronic resource] / R. N. Toma, F. Piltan, J.-M. Kim // Sensors. – 2021. – Vol. 21, № 24. – Mode of access: https://www.mdpi.com/1424-8220/21/24/8453
7. Novelty detection and fault diagnosis method for bearing faults based on the hybrid deep autoencoder network [Electronic resource] / [Y. Zhao, H. Hao, Y. Chen et al.] // Electronics. – 2023. – Vol. 12, № 13. – Mode of access: https://www.mdpi.com/2079-9292/12/13/2826
8. A bearing fault classification framework based on image encoding techniques and a convolutional neural network under different operating conditions [Electronic resource] / [R. N. Toma, F. Piltan, K. Im et al.] // Sensors. – 2022. – Vol. 22, № 13. – Mode of access: https://www.mdpi.com/1424-8220/22/13/4881
9. Bearing fault diagnosis based on auto-encoder combined with CNN [Electronic resource] / [F. Yuan, P. Xun, W. Wang et al.] Fuzzy Systems and Data Mining IX : Proceedings of the Fuzzy Systems and Data Mining Conference, 18–20 November 2023 : proceedings. – Xiamen : IOS Press, 2023. – Vol. 354. – P. 334–344. – Mode of access: https://ebooks.iospress.nl/doi/10.3233/FAIA231039
10. Rao C. A survey of autoencoder and convolutional neural network based methods for fault diagnosis [Electronic resource] / C. Rao // Advances in Engineering Technology Research. – 2024. – Vol. 11, № 1. – P. 516–528. – Mode of access: https://madison-proceedings.com/index.php/aetr/article/view/2529
11. Zhang S. Deep learning algorithms for bearing fault diagnostics – a comprehensive review [Electronic resource] / [S. Zhang, S. Zhang, B. Wang et al.] // IEEE Access. – 2020. – Vol. 8. – P. 29857–29881. – Mode of access: https://ieeexplore.ieee.org/document/8988271
12. Gholamy A. Why 70/30 or 80/20 relation between training and testing sets: a pedagogical explanation [Electronic resource] / A. Gholamy, V. Kreinovich, O. Kosheleva // Mode of access: https://api.semanticscholar.org/CorpusID:7467506
13. Lee D. GCN-based LSTM Autoencoder with self-attention for bearing fault diagnosis [Electronic resource] / D. Lee, H. Choo, J. Jeong // Sensors. – 2024. – Vol. 24, № 15. – Mode of access: https://www.mdpi.com/1424-8220/24/15/4855
14. Sohaib M. Reliable fault diagnosis of rotary machine bearings using a stacked sparse autoencoder-based deep neural network [Electronic resource] / M. Sohaib, J.-M. Kim // Shock and Vibration. – 2018. – Vol. 2018, № 1. – Mode of access: https://onlinelibrary.wiley.com/doi/10.1155/2018/2919637
15. Gu Y. A denoising autoencoder-based bearing fault diagnosis system for time-domain vibration signals [Electronic resource] / [Y. Gu, J. Cao, X. Song et al.] // Wireless Communications and Mobile Computing. – 2021. – Vol. 2021, № 1. – Mode of access: https://onlinelibrary.wiley.com/doi/10.1155/2021/9790053
16. Althubaiti A. Fault diagnosis and health management of bearings in rotating equipment based on vibration analysis – a review [Electronic resource] / A. Althubaiti, F. Elasha, J. A. Teixeira // Journal of Vibroengineering. – 2021. – Vol. 24, № 1. – P. 46–74. – Mode of access: https://www.extrica.com/article/22100

DOI 10.15588/1607-3274-2025-2-10