



(RESEARCH ARTICLE)



CNC machine gearbox fault detection with convolutional neural network

Konstantin Gubernatorov * and Sergei Gavrilenkov

Department of Robotics and Complex Automation, Bauman Moscow State Technical University (BMSTU), Moscow, Russia.

World Journal of Advanced Engineering Technology and Sciences, 2024, 11(02), 379–388

Publication history: Received on 24 February 2024; revised on 06 April 2024; accepted on 09 April 2024

Article DOI: <https://doi.org/10.30574/wjaets.2024.11.2.0044>

Abstract

This study focuses on the development of a deep learning-based approach of gearbox monitoring and fault detection. The project aims to create a solution for early detection of defects in dynamic equipment based on data from vibration sensor by building a binary classifier with convolutional neural network implemented. The gearboxes condition of which is being assessed is stored in three similar computer numerically controlled (CNC) milling machines. Data is collected during 15 milling operations of different duration and with different tool's speed and feed. Vibration is measured by an accelerometer stored on the body of each gearbox. Convolutional neural network takes vibration spectra as inputs and whether fault is detected makes a prediction of a gearbox condition. To make the whole solution autonomous and be able to embed it into manufacture the project is integrated into a server with an edge-to-cloud architecture. As an end product deep learning fault classifier stored on a server is to detect possible gearbox faults, draw conclusions on condition of dynamic equipment and automate the process of fault detection.

Keywords: Gearbox; CNC machine; Vibration spectrum; Machine learning; Deep learning; Convolutional neural network

1. Introduction

The requirements for precision, reliability and safety of modern numerically controlled machines are very high. Failure or damage to transmission parts often causes a chain reaction leading to a severe accident, which significantly increases the economic cost of operating the equipment. Gear failure accounts for most of the mechanical failures. Therefore, it is very important to accurately determine the condition of the gearbox, as well as diagnose and predict gearbox malfunctions.

Currently, fault diagnosis methods are mainly divided into three types – model-based fault diagnosis methods, signal processing-based fault diagnosis methods, and data-based fault diagnosis methods [3]. Data-based fault diagnosis methods can be divided into two types, namely: traditional machine learning fault diagnosis methods and deep learning fault diagnosis methods.

Model-based fault diagnosis methods use the correlation between transmission fault characteristics and the physical model, analyze the fault mechanism to build and optimize the model, and implement real-time fault diagnosis and prediction. However, in practice, it is difficult to establish an accurate transmission model, which significantly limits the use of model diagnostic methods.

Fault diagnosis methods based on signal processing determine effective diagnostic indicators by analyzing the correlation between signals and faults [12]. Fault diagnosis is achieved by constructing fault signs using dimensional and dimensionless signal indicators. However, the operating conditions of the gearbox are complex and changeable, and

* Corresponding author: Gubernatorov Konstantin

the selected features are difficult to use in different conditions. Thus, analyzing the generality of error data in massive data is an effective means of error diagnosis.

In recent years, due to the significant increase in training resources and the rapid development of computing power, data-based fault diagnosis methods have gradually attracted more and more attention. The development of machine learning algorithms opens up a new way to diagnose transmission failures. In accordance with the signal processing technology, a spectrum of signs is analyzed and constructed that can effectively express a malfunction. Then, a machine learning algorithm is used for intelligent fault diagnosis.

However, in traditional machine learning algorithms, the selection and extraction of fault signs are still based on manual control, which introduces uncertainty into fault diagnosis and does not allow achieving the goal of real intelligent diagnostics. The deep learning method with powerful function learning ability can realize automatic feature extraction and fault classification, so it is widely used in the field of fault diagnosis.

The input data of the deep learning diagnostic model includes two types of fault samples: accelerometer readings in the form of acceleration projections along three axes and a two-dimensional vibration spectrum based on accelerometer data [6]. The first one directly extracts fault characteristics from one-dimensional vibration signals for diagnosis, and the second combines signal processing technology to convert vibration signals into two-dimensional images. Many studies have used signal preprocessing technology to improve sample quality during the conversion process. A fault feature extraction method that introduces fault image samples into a deep learning model is a necessary choice for accurate fault identification.

Aim and objectives of the research

The aim of this research is to develop a method for early detection of defects in dynamic equipment based on data from vibration sensors by constructing a binary classifier based on machine learning methods.

The study aims to achieve the following objectives:

- To analyze approaches to monitoring the condition of dynamic equipment using machine learning methods.
- To analyze the operating data of a CNC milling machine gearbox in two states.
- To select a rational convolutional neural network model.
- To test the built classifier.
- To draw conclusions on the result of convolutional neural network performance.

2. Literature Review

2.1. Gearbox Fault Detection with Vibration Analysis

The article [1] provides a brief overview of modern vibration-based methods used for condition monitoring in gear transmission systems. The authors of the article draw the following conclusions:

- Transmission vibration signals are usually intermittent and noisy. The time domain averaging method successfully removes noise from the signal and captures the dynamics of one signal period.
- Methods for analyzing vibration signals in the time domain in the form of waveform generation, indices and the overall vibration level do not provide any diagnostic information, but may have limited use in detecting faults in simple auxiliary components that are critical for safety.
- In the frequency domain, the fast Fourier transform (FFT) was able to display pulses at fault characteristic frequencies and their multiples, but other peaks are also visible — this occurs due to the effect of signal modulation. It is difficult to determine fault categories using this method.
- During the bandpass analysis of the vibration signals of the gear, it was found that this method is applicable to identify signs for fault diagnosis. It was concluded that the root mean square (RMS) value of the filtered signal frequency in three frequency ranges can be a valuable characteristic for the development of an intelligent system.
- Synchronous signal averaging potentially greatly simplifies shaft and gear fault diagnosis by providing significant attenuation of non-synchronous vibrations and signals for which ideal filtering can be used. It is necessary to carry out further development on the introduction of methods of synchronous averaging and analysis of the results.

- An expert system based on artificial neural network (ANN) and fuzzy logic can be designed to reliably classify faults using extracted features from a vibration signal.
- The results also show that signal generation in the case of multiple faults on the contact surfaces of the gear is useful only for determining the correct or faulty condition, but is not able to identify fault categories.

2.2. Deep Neural Networks for Intelligent Diagnosis of Rotating Machinery

ANNs are one of the most commonly used classifiers in the intelligent fault diagnosis methods, which generally include two main steps, i.e. fault feature extraction using signal processing techniques and fault classification using ANN classifiers [9], [10]. Feature extraction involves mapping of measured signals onto representative features characterizing the health conditions of machinery. And fault classification is to distinguish the health conditions based on the extracted features. Thanks to the representative features from the measured signals and adaptive learning capability of ANNs, the ANN-based methods are supposed to displace diagnosticians for making decisions and work well in intelligent fault diagnosis. The ANN-based methods reported in literature, however, have two obvious deficiencies. Firstly, the features input into classifiers are extracted and selected by diagnosticians from the measured signals, largely depending on prior knowledge about signal processing techniques and diagnostic expertise. In addition, the features are selected according to a specific diagnosis issue and probably unsuitable for other issues. Thus, it is necessary to adaptively mine the characteristics hidden in the measured signals to reflect the different health conditions of machinery, instead of extracting and selecting features manually. Secondly, the ANNs commonly adopted in intelligent fault diagnosis of rotating machinery have shallow architectures, which means that only one hidden layer is included in an ANN architecture. Such simple architectures limit the capacity of ANNs to learn the complex non-linear relationships in fault diagnosis issues. Thus, it is necessary to establish a deep architecture network for distinguishing the health conditions of machinery.

Based on DNNs trained through deep learning, the article [5] proposes a novel intelligent diagnosis method to overcome the two deficiencies of the ANN-based methods in fault diagnosis of rotating machinery. In this method, DNNs are utilized to implement both fault feature extraction and intelligent diagnosis. The DNNs are first pre-trained by an unsupervised layer-by-layer learning and then fine-tuned with a supervised algorithm, where the unsupervised process helps the fault characteristic mining and the supervised process contributes to construct the discriminative fault characteristics for classification [5], [11]. The merits of the proposed method are summarized as follows. It is able to adaptively mine fault characteristics from the measured signals for various diagnosis issues. The method is good at establishing the non-linear mapping relationship between the different health conditions of machinery and the corresponding measured signals. Therefore, the proposed method is expected to obtain higher diagnosis accuracy compared with the methods based on shallow ANNs.

2.3. Convolutional Neural Network for the Gearbox Condition Monitoring

Table 1 Average testing accuracy and standard deviation of tested methods

Methods	Learning features		Manual features				
	Raw data	time	Frequency data	Time-frequency data	Time features	Frequency features	Wavelet features
CNN	82.30 ± 1.07		98.67 ± 0.52	95.54 ± 1.02	65.18 ± 1.62	84.61 ± 0.49	84.11 ± 0.66
FNN	23.57 ± 4.56		54.34 ± 4.21	51.62 ± 4.28	65.40 ± 3.17	84.83 ± 0.65	84.23 ± 0.76
SVM	23.87 ± 2.40		76.21 ± 1.51	71.61 ± 1.63	63.37 ± 1.30	84.84 ± 0.42	84.73 ± 0.47
RF	42.56 ± 1.46		92.34 ± 1.18	81.31 ± 1.5	64.24 ± 1.06	88.26 ± 0.94	83.16 ± 1.26

Convolutional neural network (CNN) as one of the main types of deep neural networks (DNN) [4], has been applied with great success to learn features from raw data and has become the dominant approach for almost all recognition and detection tasks in image and speech analysis. However, very few investigations have been conducted on the application of CNN in feature learning and fault diagnosis for a planetary gearbox or a gearbox with combined gear-bearing-shaft faults. At the same time, most of the studies of DNN based feature learning only focus on one type of raw data. The study of the different performance of feature learning from various types of data is still few. In the article [7] CNN is applied to learn features from raw vibration data in time domain, raw frequency spectrum of the data and their combination, and diagnose the health conditions of gearboxes. Manual features and three common intelligent methods, including

fully-connected neural network (FNN), support vector machine (SVM) and random forest (RF), are used as comparisons, as can be seen in Table 1.

3. Material and methods

3.1. Experimental Setup

To keep the research as close as possible to the industrial scenario, the data is collected from different 4-axis horizontal CNC machining centers during production [8]. The machines are processing aluminum workpieces as depicted in Figure 1. For the data acquisition, an indirect method by collecting accelerometer data from Bosch CISS sensors mounted to the rear end of the spindle housing is used. Other approaches opt for mounting the sensors in the machining area. This rear area remains unaffected by extreme machining environment, coolant or material chips and is available for retrofitting new sensors to brownfield machines. The sensor maintains a constant distance to the tool center point and the three axes of the accelerometer are in alignment with the linear motion axis of the machine. The sensor coordinate system is indicated in Figure 1.

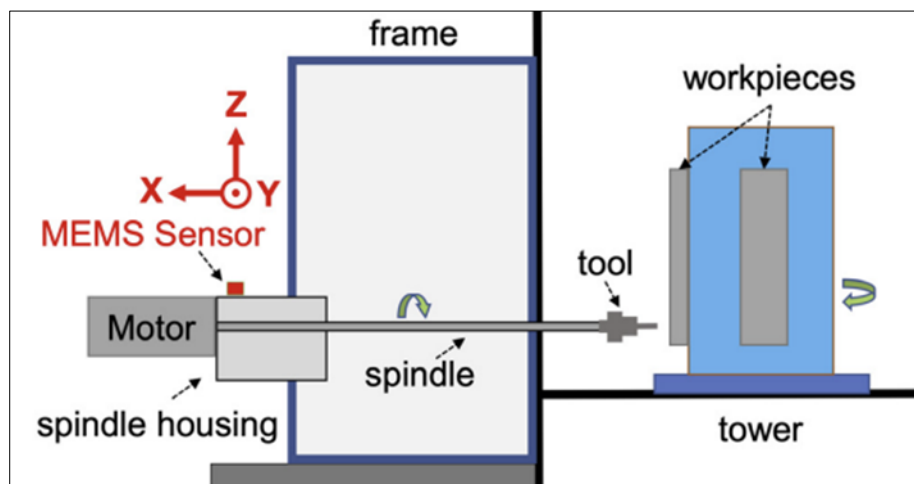


Figure 1 Schematic sketch of the experimental setup: 4-axis machining center with mounted sensor

Using the low-cost tri-axial CISS sensor, acceleration data is collected with a sampling rate of 2 kHz. Most relevant frequencies to monitor the machining processes are low integer multiples of the spindle speed. For tool operations, these frequencies will be in the range of 75 Hz to 1 kHz. According to the Nyquist-Shannon theorem, a minimum sampling rate of 2 kHz is sufficient to detect machine anomalies. Sampling with this rate along the 3-axes produces an amount of 4.14 GB per day. Such volumes of data cannot be fully stored and processed in on-premise solutions. It demands a smart data mining system to collect, store, annotate, process and learn from the gathered data.

3.2. Cloud Server

To have reliable annotation, continuous data collection and simultaneous machine learning (ML) evaluation, an internet of things (IoT) architecture is required which enables:

- central aggregation of selected anomalies and processes across different machining centers and locations,
- local storage and processing of raw sensor data including event annotation by product experts,
- aggregation of annotated data in a central database,
- centralized training of ML models, and
- management and deployment of models and modules from the cloud to the edge device.

The data collection system presented in this work is characterized in an edge-to-cloud architecture. The main goal of this architecture is the simplification of data annotation, the use of expert knowledge in the shop floor, and the centralized storage of annotated data in the cloud. Through an anomaly detector module, potential events and anomalies are pre-selected for annotation.

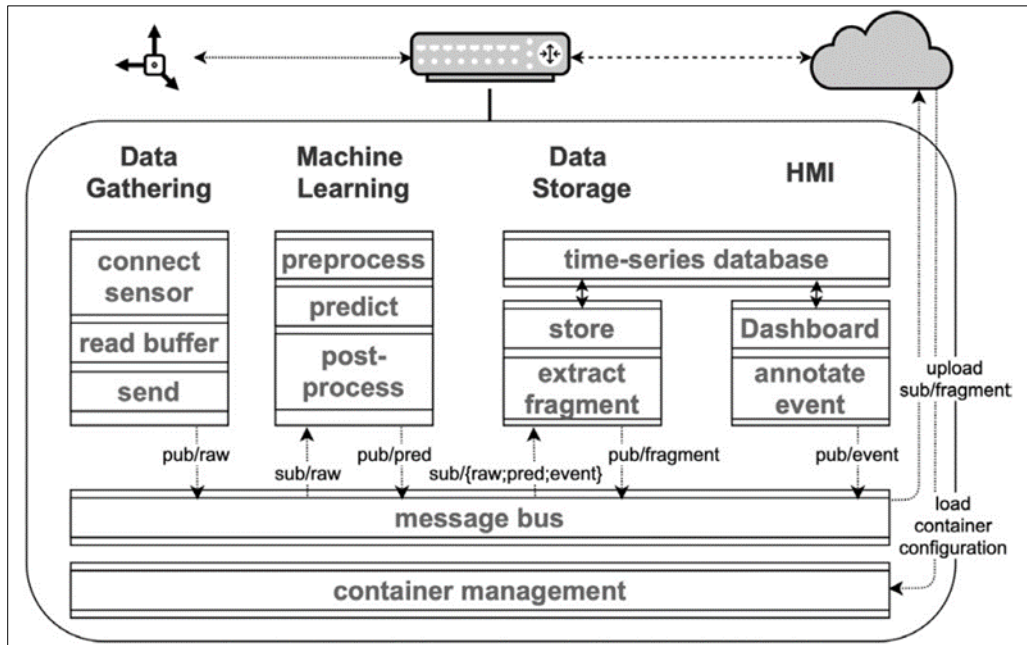


Figure 2 Concept and interaction of containers in the edge stack

The edge stack represented in Figure 2 describes the modules running in the production line on site. The modules are managed from the cloud side by an orchestration client running on the edge device. A messaging bus using the Message Queuing Telemetry Transport (MQTT) protocol provides a standardized interface for local inter application communication. The data gathering and annotation system involves multiple modules. Firstly, a data gathering module establishes a connection to the accelerometer sensor and triggers the read. The data stream is afterwards published on the message bus. Secondly, the data stream is subscribed by a ML module, which with predictions on the stream, supports the quality check process by pre-selecting the correct time frame for anomalies. This allows time-delayed annotations to be entered by the end-of-line quality check, while retaining the majority of data only in the edge time-series database. Ultimately, a dashboard allows the visualization of the ML pseudo-labels and manual annotation via the user interface. Once an event is validated by the experts, the corresponding data segment gets acquired and queried for upload to the cloud. The major benefit of the architecture is the collaboration of data science and domain expertise. It allows additionally in-place distribution of updated ML modules, which support and improve data annotation.

3.3. Initial Data

The data is collected in a production plant from 3 different CNC machines (M01, M02 and M03) on a regular basis during the time interval of October 2018 to August 2021 [8]. The time frame is tagged as "Month Year" and represents the 6-month interval before the label. For example, "Aug 2019" would refer to the period between February 2019 and August 2019. The machine performs a sequence of several operations using different tools on aluminum parts to work the specified design. It is important to mention that the machines produce different parts and the process flow changes over time. To study the drift between machines and over time, the dataset is built with 15 different tool operations that run on all 3 machines at different time frames. Table 2 gives an overview on the characteristics of the different operations.

During machining, the different process operations are conducted in high-speed, requiring a frequent mounting and unmounting of tools on the spindle chuck. These factors lead occasionally to process failures mainly caused by tool misalignment, chip clamping, chip in chuck, tool breakage. To reach the optimal product quality, after each batch an expert on the shop floor controls the resulting workpiece in a gauging station and annotate the process health. Nevertheless, labeling during production is still very challenging. Due to the manual drudgery gauging, some processes are wrongly labelled and precise annotations are missing. The published dataset focuses on the quality process failures: the OK class refers to a healthy process and NOK refers to a faulty process. Figure 3 shows an unbalance rate of 816:35 between the OK/NOK in the dataset. In the real production, the number of OK samples are significantly higher. To provide an exemplary dataset, a reasonable number of OK processes were selected from the different time periods, which reduces the class imbalance.

Table 2 Tools operations collected from M01, M02 and M03

Tool operation	Description	speed [Hz]	feed [mm*s ⁻¹]	duration [s]
OP00	Step Drill	250	100	132
OP01	Step Drill	250	100	29
OP02	Drill	200	50	42
OP03	Step Drill	250	330	77
OP04	Step Drill	250	100	64
OP05	Step Drill	200	50	18
OP06	Step Drill	250	50	91
OP07	Step Drill	200	50	24
OP08	Step Drill	250	50	37
OP09	Straight Flute	250	50	102
OP10	Step Drill	250	50	45
OP11	Step Drill	250	50	59
OP12	Step Drill	250	50	46
OP13	T-Slot Cutter	75	25	32
OP14	Step Drill	250	100	34

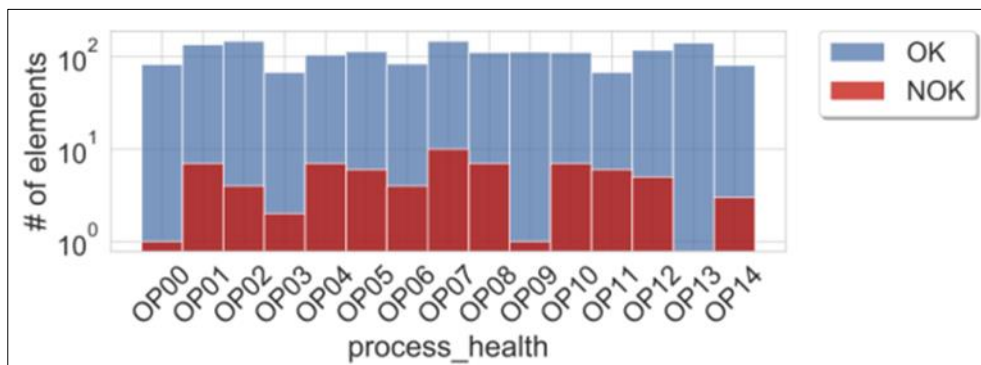


Figure 3 Class distribution per process operation

3.4. Building a Convolutional Neural Network

To further verify the operability of the algorithms, the input data is divided into a test sample (necessary to determine the quality) and a training sample (on which the model is trained). The general algorithm for a classifier working process is shown in Figure 4. Tiny VGG architecture, as can be seen in Figure 5, was chosen to implement.

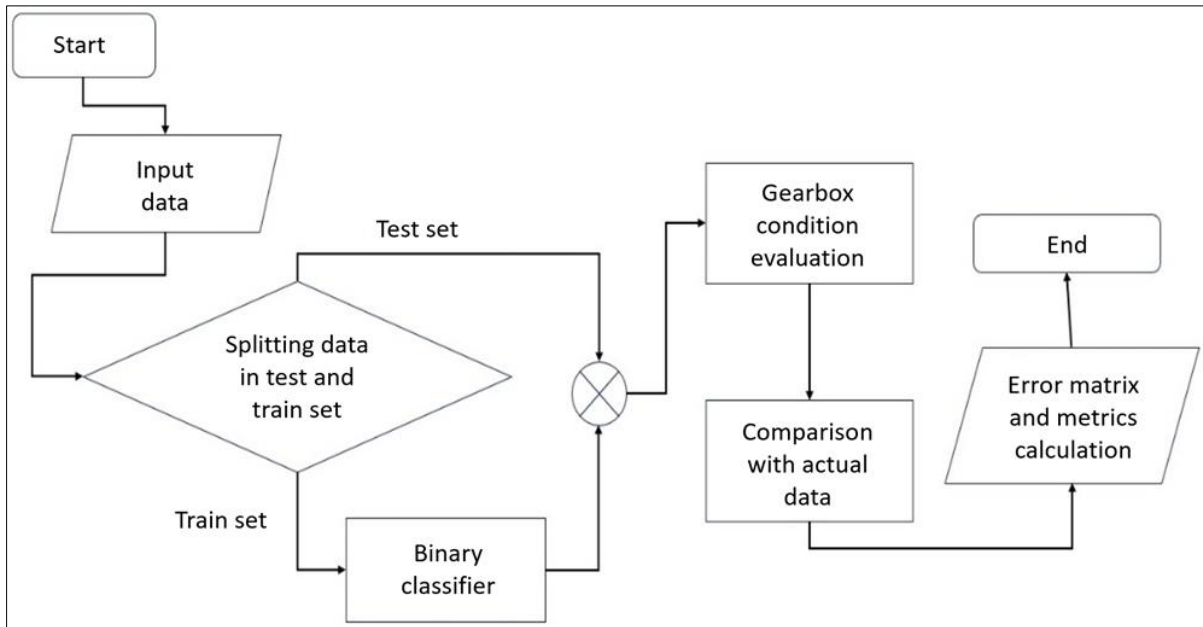


Figure 4 Classifier work algorithm

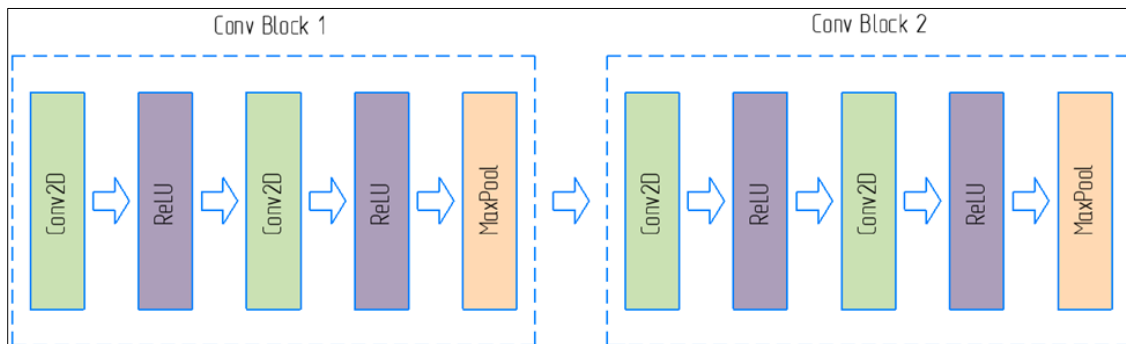


Figure 5 Tiny VGG architecture

3.5. Tiny VGG Layers

3.5.1. Convolutional Layer

The convolutional layers are the foundation of CNN [2], as they contain the learned kernels (weights), which extract features that distinguish different images from one another — this is what we want for classification! As you interact with the convolutional layer, you will notice links between the previous layers and the convolutional layers. Each link represents a unique kernel, which is used for the convolution operation to produce the current convolutional neuron’s output or activation map.

The convolutional neuron performs an elementwise dot product with a unique kernel and the output of the previous layer’s corresponding neuron. This will yield as many intermediate results as there are unique kernels. The convolutional neuron is the result of all of the intermediate results summed together with the learned bias.

3.5.2. Activation Layer

ReLU function, as shown in Figure 6, applies much-needed non-linearity into the model. Non-linearity is necessary to produce non-linear decision boundaries, so that the output cannot be written as a linear combination of the inputs. If a non-linear activation function was not present, deep CNN architectures would devolve into a single, equivalent convolutional layer, which would not perform nearly as well. The ReLU activation function is specifically used as a non-

linear activation function, as opposed to other non-linear functions such as Sigmoid because it has been empirically observed that CNNs using ReLU are faster to train than their counterparts.

$$\text{ReLU}(x) = \max(0, x) \dots \dots \dots (1)$$

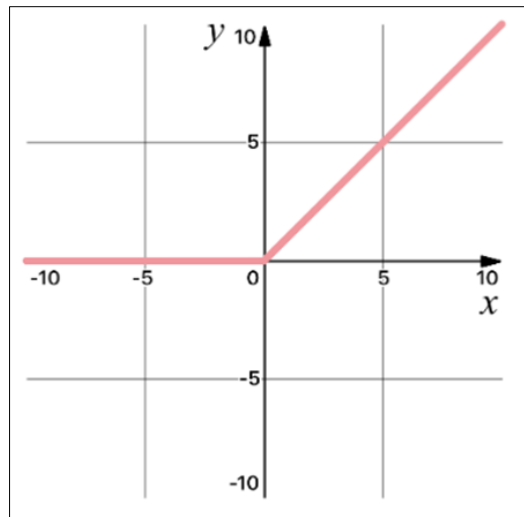


Figure 6 ReLU activation function graphed, which disregards all negative data

3.5.3. Pooling Layer

There are many types of pooling layers in different CNN architectures, but they all have the purpose of gradually decreasing the spatial extent of the network, which reduces the parameters and overall computation of the network. The type of pooling used in the Tiny VGG architecture is Max-Pooling.

The Max-Pooling operation requires selecting a kernel size and a stride length during architecture design. Once selected, the operation slides the kernel with the specified stride over the input while only selecting the largest value at each kernel slice from the input to yield a value for the output.

In the Tiny VGG architecture above, the pooling layers use a 2x2 kernel and a stride of 2. This operation with these specifications results in the discarding of 75% of activations. By discarding so many values, Tiny VGG is more computationally efficient and avoids overfitting.

4. Result and Discussion

The main visualization of the result of the work of the CNN is the error matrix, presented in Figure 7 – a matrix on the main diagonal of which there are the number of correctly classified objects, and outside of it – falsely classified.

	Fault present	Fault absent
Fault present	758	59
Fault absent	108	155

Figure 7 Error matrix of CNN classifier

Metrics are calculated based on the error matrix of the binary classifier, as seen in Table 3:

Accuracy – proportion of correctly classified to the number of all signs:

$$A_i = \frac{\Sigma TP + TN}{\Sigma TP + FN + TN + FP} \dots\dots\dots (2)$$

Loss – estimation of the proximity of true and predicted values:

$$H(p, q) = - \Sigma_x p(x) \log(q(x)) \dots\dots\dots (3)$$

Precision – proportion of correct classifications to all classifications of the trait:

$$P_i = \frac{\Sigma TP}{\Sigma TP + FP} \dots\dots\dots (4)$$

Recall – proportion of correctly defined objects to all objects of this class:

$$R_i = \frac{\Sigma TP}{\Sigma TP + FN} \dots\dots\dots (5)$$

F-score – average harmonic of accuracy and completeness:

$$F1 = 2 \times \frac{P_i \times R_i}{P_i + R_i} \dots\dots\dots (6)$$

Table 3 CNN classifier performance metrics

Architecture	Accuracy	Loss	Precision	Recall	F1-score
Tiny VGG	0.843	0.072	0.846	0.843	0.845

5. Conclusion

In conclusion, a technique for early detection of gearbox defects based on vibration sensor data using a deep learning model was proposed. The study of existing approaches and methods for the implementation of monitoring, data collection and analysis using classical machine learning and deep learning methods has been carried out.

A method for implementing a binary classifier for determining the presence of a defect was proposed and one of the existing CNN architectures was built.

The methodology for implementing the approach to early detection of defects has been tested on a set of historical data. The classification quality was determined using an error matrix and basic metrics: accuracy, loss, precision, completeness and F1-score.

Furthermore, this project can be used in an existing manufacturing process to fully automate the process of monitoring the condition and fault detection of dynamic equipment.

Compliance with ethical standards

Acknowledgments

The author would like to thank all parties who have contributed to this research.

Disclosure of conflict of interest

No conflict of interest is to be disclosed.

References

[1] Aherwar A. (2012). An Investigation on Gearbox Fault Detection Using Vibration Analysis Techniques: A review. Australian Journal of Mechanical Engineering, vol. 10, no. 2, pp. 169-183.

- [2] Azamfar M, Singh J, Bravo-Imaz I and Lee J. (2020). Multisensor data fusion for gearbox fault diagnosis using 2-D convolutional neural network and motor current signature analysis. *Mechanical Systems and Signal Processing*, vol. 144.
- [3] Barbieri N, Barbieri G, Bruno Matos M, Barbieri L and Fonseca de Lima K. (2019). Analysis of automotive gearbox faults using vibration signal. *Mechanical Systems and Signal Processing*, vol. 129, pp. 148-163.
- [4] Chen Z, Li C and René-Vinicio S. (2015). Gearbox Fault Identification and Classification with Convolutional Neural Networks. *Shock and Vibration*, vol. 2015.
- [5] Feng J, Yaguo L, Jing L, Xin Z and Na L. (2016). Deep neural networks: A promising tool for fault characteristic mining and intelligent diagnosis of rotating machinery with massive data. *Mechanical Systems and Signal Processing*, vol. 72–73, pp. 303-315.
- [6] Lei Y, Ming Zuo J, He Z and Zi Y. (2010). A multidimensional hybrid intelligent method for gear fault diagnosis. *Expert Systems with Applications*, vol. 37, no.2, pp. 1419-1430.
- [7] Luyang J, Ming Z, Pin L, and Xiaoqiang X. (2017). A convolutional neural network based feature learning and fault diagnosis method for the condition monitoring of gearbox. *Measurement*, vol. 111, pp. 1-10.
- [8] Mohamed-Ali T, Michael F and Klaus D. (2022). Smart Data Collection System for Brownfield CNC Milling Machines: A New Benchmark Dataset for Data-Driven Machine Monitoring. *Procedia CIRP*, vol. 107, pp. 131-136.
- [9] Samanta B, Al-Balushi KR and Al-Araimi SA. (2003) Artificial neural networks and support vector machines with genetic algorithm for bearing fault detection. *Engineering Applications of Artificial Intelligence*, vol. 16, no. 7-8, pp. 657-665.
- [10] Yang Z, Hoi WI and Zhong J. (2011). Gearbox fault diagnosis based on artificial neural network and genetic algorithms. *International Conference on System Science and Engineering*, pp. 37-42.
- [11] Zhang W, Li C, Peng G, Chen Y and Zhang Z. (2018). A deep convolutional neural network with new training methods for bearing fault diagnosis under noisy environment and different working load.
- [12] *Mechanical Systems and Signal Processing*, vol. 100, pp. 439-453.