Kufa Journal of Engineering
Vol. 16, No. 3, July 2025, P.P. 217-234
Article history: Received 5 December 2024, last revised 17 May 2025, accepted 21 January 2025

CLASSIFICATION OF FAULT SIGNALS BASED ON DCT AND DEEP LEARNING

Haitham Al Akabi[1] and Tawfiq Al-assadi[2]

[1] Master's Degree Student in College of Information Technology, University of Babylon, Email: haithammohammeda.sw@student.uobabylon.edu.iq

[2] Assistant Prof. at College of Information Technology, University of Babylon, Iraq. Email: Tawfiqasadi@itnet.uobabylon.edu.iq

https://doi.org/10.30572/2018/KJE/160413

ABSTRACT

Industrial sound analysis plays a key role in machine monitoring systems, as experienced machine operators can hear and detect many rotary machine faults. However, due to factories having complicated noise environments, they are challenging to detect. Detecting rotary motor faults is important to reduce repair costs, avoid unexpected machine downtime, and stop production. One of the primary approaches for fault diagnosis involves deep learning (DL) techniques that classify the fault sound produced by motors. This research proposes a method to detect and classify faults using a Convolution Neural Network (CNN) based on converting sound to a gray image. First, converts the 1-D time domain vibration sound into a 2-D gray image depending on Discrete Cosine Transform (DCT) to extract feature patterns by dividing the 1-D matrix of coefficients resulting from the DCT after normalization into equal parts and storing them in a column followed by a column in a 2-D matrix whose dimensions have been determined. Second, these gray images are utilized as input for CNN, which performs multi-classification in a supervised learning framework. (IDMT-ISA Electric Engine) dataset is used to demonstrate the effectiveness of the suggested approach. The results show high classification Accuracy 99.7%, Precision 99.6%, Recall 99.6%, and F1-Score 100% with effectiveness in diagnosing faults under various noise conditions associated with sound recording. This study could have an impact in the field of sounds analysis.

KEYWORDS

DCT, DL, features extract, gray image, CNN, induction motor.

1. INTRODUCTION

Induction motors play an important role in industry due to their low cost, robustness, and ease of maintenance. With the rapid development of technology, Induction motors have become more accurate and automatic, Recently, these electric motors have been introduced as an alternative to diesel engines in electric or hybrid cars. However, failures in these machines may lead to unexpected stoppage of the manufacturing process. Early detection of these failures and their reliable diagnosis is a major problem in modern industries (Nguyen et al., 2013). Much research has been conducted to develop DL models, proving their ability and efficiency in various fields, especially in diagnosing machine faults in industry (Hoang and Kang, 2019). DL is distinguished from traditional neural networks in that it works with multiple layers, which enables it to perform complex calculations and is suitable for various disciplines (Mohammed, Kareem and Mohammed, 2022). CNN networks emerged as part of DL, which can understand and classify many categories of large data (Alaa, Hussein and Al-Libawy, 2024). In recent years, different techniques have appeared to obtain the signal, including vibration, sound, temperature, etc. This signal is processed to extract features. Wang et al. Showed how to transform a one-dimension vibration signal into a two-dimension grayscale picture, and they were successful in classifying the data using CNN (Wang, Mao and Li, 2021). He et al. Introduced a conversion of the time domain signal into a grayscale image by applying the Gabor filter and Wasserstein Generative Adversarial Network (WGAN)(He, Lv and Chen, 2023). Information from vibration signals that are difficult to perceive in the time domain may be revealed using the frequency domain analysis approach. In Fourier analysis, time domain vibration signals are transformed to the frequency domain using the Discrete Fourier Transform (DFT), Fast Fourier Transform (FFT), and Short Time Fourier Transform (STFT) approaches (Zhang, Wang and Wang, 2012) (Betta et al., 2002) (Manhertz and Bereczky, 2021). In these experiments, the raw vibration signal is pre-processed to extract features and use them as (CNN) inputs, but there are some serious limitations in these studies including

1- Extracting features requires human expertise.

2- Difficulty in identifying faults by using time domain or frequency domain or both.

3- Deep learning techniques were originally designed on two-dimensional inputs such as images. Deep learning techniques prove powerful and effective for fault diagnosis such as CNN (Han, Oh and Jeong, 2021), DNN (Zhou et al., 2021), and AutoEncoder (Hao Yijia et al., 2020). For their ability to overcome signal distortion conditions and complex operating environments with high classification accuracy. Zhou et al. Suggested using a network with Long-Short-Time-Memory (LSTM) to extract the time series. However, noise may affect the diagnosis accuracy, and more CNN layers lead to redundant parameters and increased training time (Zhao, Sun and Jin, 2018). To address these problems, Li et al. Proposed a machine attention technique and achieved excellent fault diagnosis results with good training speed and high accuracy (Li, Zhang and Ding, 2019).

In this research, sounds acquired by recording motor sounds are used, which are less affected by noise due to their high power spectrum, instead of using current and voltage signals which may contain unnecessary signals, such as characteristic harmonics (Tran et al., 2013).The sound is preprocessed and converted into a gray image based on DCT transform to extract features and detect the main elements clearly, instead of adopting the transform raw signal into signal 1-D time domain, 1-D frequency domain, or 1-D time frequency domain and using it straightforwardly as inputs to CNN network due to its poor obfuscation ability (Lei et al., 2016). This research aims to convert the fault sounds generated by the motors into a gray image to extract features, eliminate irrelevant features, and develop a CNN network to identify the type of fault in the rotating motors. The contribution of this study is described as follows:

1- The power of our study compared to traditional methods is that it involves effectively extracting features despite the noise associated with the recordings.

2- A unique way to create 2-D gray images that makes it easier for neural networks to recognize patterns more effectively.

3- Developing a CNN architecture designed for fault diagnosis in audio signal images.

4- Providing an effective approach to using deep learning techniques in audio analysis.

The rest of this article is structured as follows: Section 2 describes related works. Section 3 explains the suggested strategy. Section 4 describes the experiment outcomes. Section 5 concludes with a summary of the findings.

2. RELATED WORKS

In general, to diagnose faults in induction motors, data that records the vibration or sounds of the motors is needed. Methods based on analyzing the vibration signal include analysis of the time domain, frequency domain, or time-frequency domain, which undergoes processing and transformation operations to extract features and use them as input to a Convolutional Neural Network for classification. The literature will be reviewed, focusing on the methods of processing the acquired signal and the networks used to classify the error signal. The most common way to obtain the signal from induction motors is through sensors connected to the motor to draw the vibration signal. Mukherjee et al. Developed a system to transform the time-frequency domain vibration signal to a spectral image by applying STFT, it breaks a signal into smaller segments by applying the Fourier transform mathematical to each segment and, then uses a CNN to classify multiple faults. Disadvantages in this study where the sensors are placed have a big impact on the model accuracy, for example, compared to data received from a nearby location, the classification accuracy decreases more dramatically when the sensor data is obtained from a place that is physically further away from the training sensor. This indicates a limitation in the model adaptability to variations in sensor placement (Mukherjee and Tallur, 2021). Misra et al. Suggested three techniques for feature extraction: the first is based on time analysis, the second uses the FFT to transform time analysis into frequency analysis, and the third is converting the time domain to the time-frequency domain using the STFT. Relying on the time domain only may lose important frequency information, while the frequency domain assumes that the signal is stationary, and this does not happen in reality for induction motors, which leads to the loss of information related to time. STFT representation of the time-frequency domain provides a better representation of the non-stationary signal, but it has some limitations including choosing the window size to compare between time and frequency accuracy, a larger window may provide better frequency accuracy with lower time accuracy, and vice versa, which may hinder good identification of a fault. The researcher used the Random Forest (RF) algorithm to classify errors for both the time and frequency domain vibration signals and used the CNN network to classify the time-frequency domain vibration signal (Misra et al., 2022).

Ahmed et al. Discussed converting a 1-D vibration signal into a color image (RGB), transforming it first into a grayscale picture, then into a binary image, and finally into a color image (RGB) by the process of extracting the Region of Interest (ROI). The limitation of this study requires a high-quality input signal without noise, in addition to the fact that the signal undergoes more than one conversion, which may cause a loss of the original information of the signal, and training a network CNN requires high computational resources, which may make its implementation this method in hardware environments with limited capabilities more difficult (Ahmed and Nandi, 2021). He et al. Suggested splitting a 1-D vibration signal into sub-windows to turn it into a 2-D grayscale picture without the need for any techniques, and storing them in the 2-D binary matrix in the form of a row-by-row for each window or column after column, and then using CNN to classify bearing faults. In this method, data augmentation was adopted using a network (GAN), and it is noticeable that there is a difference between the results of diagnosing the original data and the results of diagnosing the generated data (He, Lv and Chen, 2023). Kang et al. Created a method that uses a feature extract based on Singular Value Decomposition (SVD) to identify motor defects. Demonstrates two methods, in time domain analysis, the first uses Short-Time Energy (STE) plus (SVD) To extract features, and the second method is to transform the signal from the time analysis to the frequency analysis using the Discrete Cosine Transform DCT plus SVD. The classification is affected by the performance of the sensors used to collect data, as a difference in size may occur with incorrect timing (Kang and Kim, 2013). Liu et al. Suggested using the sound coming from rotating motors instead of the vibration signal to avoid noise while recording the signal and converting these sounds into a grayscale image. Convert the sounds issued by the motors into a time frequency analysis image by applying STFT and then converting this spectral image into a grayscale image and using the Autoencoder Network to classify faults. The method's effectiveness depends on high-quality sound signals, as noise can impair feature extraction and classification.in addition Implementing complex deep learning models like stacked Sparse AutoEncoders (SAE) requires significant resources and expertise, posing challenges for practitioners (Liu, Li and Ma, 2016). Ahmad et al. Presented a new approach using vibration images for fault diagnosis, which includes a processing step using Discrete Cosine Stockwell Transform (DCST) to create consistent patterns from the vibration signal image, and then converting the coefficient matrix into 2-D grayscale picture in an innovative way. The researcher used a (LeNet5) network with a Stochastic Gradient Desent (SGD) optimizer, and for classification, (SoftMax) was used, which ensures that the network learns to accurately classify bearing fault cases. This approach depends largely on the quality of the vibration images, and any quality defect leads to a wrong diagnosis. (Ahmad, Hasan and Kim, 2022). Freire Moraes et al. Suggested an approach to detect faults in diesel engines by using STFT and Continuous Wavelet Transform (CWT) techniques and generating images that are fed into CNN. Despite the high accuracy of 96.5% in classifying eight engine faults, the model performance deteriorated when tested on audio signals containing 40% noise, where the accuracy reached 70%. In addition, the CWT transformation requires more time and complex computational resources, which makes its practical application difficult (Freire Moraes, Ribeiro Junior and Gomes, 2024).

However, the large challenge to adopting sounds to identify motor faults the effectiveness heavily relies on the quality of the sound signals recorded. Background noise or interference during recording can adversely affect the accuracy of the fault diagnosis. If the sound signals are not clear, the feature extraction process may yield suboptimal results, Additionally, to increase the model performance and strength for error identification, fault sound data should be diversified.

3. PROPOSED APPROACH

As is known, CNN has been widely used in image classification. This section proposes a deep learning model for image classification to diagnose faults. First, the motor sounds are converted into gray images using DCT transform and a deep convolutional model that learns abstract features from the transformed gray images. Finally, faults are classified based on the type of fault. Fig. 1 represents the block diagram of the proposed approach. The proposed method will be presented in two phases: The first phase represents pre-processing sounds and converts sound to a gray image, and the second phase presents the Convolution Neuron Network CNN for classification of motor faults.

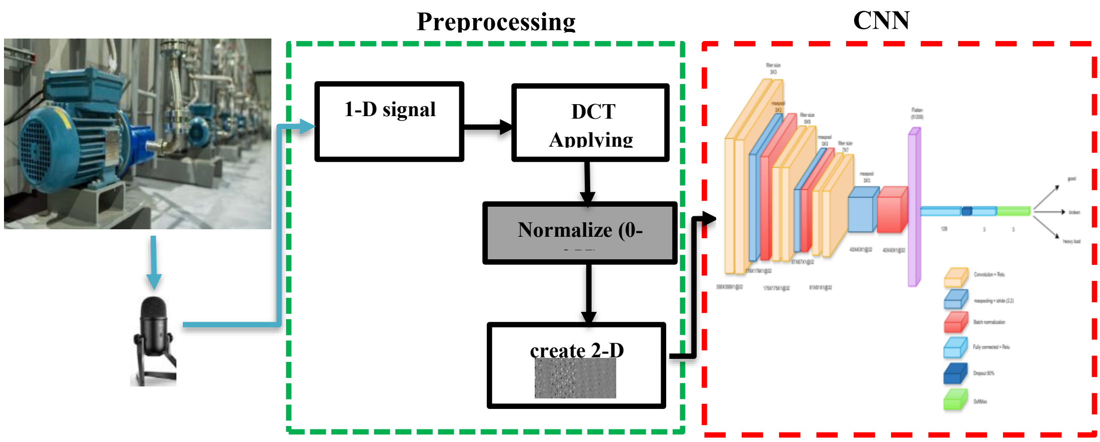

Fig. 1. Block diagram of the proposed method.

3.1. First phase: Preprocessing (Sound To 2-D Gray Image Conversion)

Converting 1-D vibration signals or sound into 2-D visual representations may be done in a variety of ways. As explained in Section II, these methods range from straightforward ones that only reshape the 1-D vibration signal into a 2-D image to transformation-based methods such as FFT, STFT, WT, and DFT that first convert the 1-D time analysis vibration signal to frequency or time-frequency analysis before producing the 2-D image representation.

Therefore, in this research, converting sound to a gray image will be discussed in four steps as follows:

3.1.1. Transform Sound To 1-D Time Domain Signal

The sounds are converted to the 1-D time domain signal, by reading the digital values of the signal frequency concerning time and storing them in a 1-D array in preparation for processing these values. The Scipy and Librosa libraries in Python are used to read and store the values. Fig. 2 shows a colored image of the waveform of the audio signal converted to the time domain.

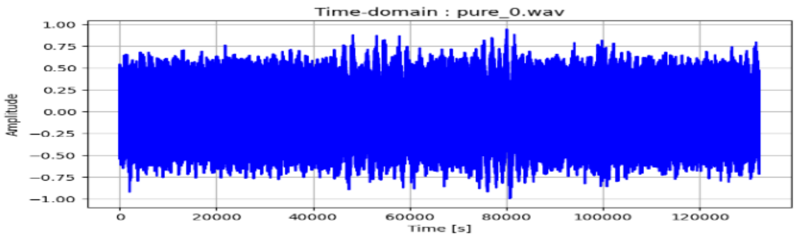

Fig. 2. 1-D Time domain signal

3.1.2. DCT Applying

DCT has shown a high ability to reduce high-frequency effects and block low-frequencies (Jaber, 2011). DCT is applied to each block of eight pixels of the 1-D time domain signal, The resulting coefficients represent the 1-D frequency domain. Fig. 3 illustrates the 1-D frequency domain signal after the application of DCT. This conversion greatly helps in separating important information (such as high or low frequencies) from unimportant information such as noise. The DCT coefficients are calculated using Eq.1 (Lam and Goodman, 2000):

$$c(u) \quad = \alpha(u) \sum_{k=0}^{N-1} f(x) cos \, [\frac{\pi(2x+1)u}{2N}] \tag{1}$$

where u = 0, 1, 2, …, N – 1, N=8, and $\alpha(u) = \begin{cases} \sqrt{\frac{1}{N}}, & for\ u = 0 \\ \sqrt{\frac{2}{N}}, & for\ u \neq 0 \end{cases}$

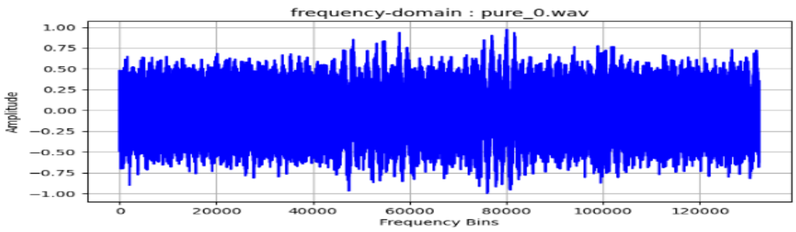

Fig. 3. 1-D Frequency domain signal

3.1.3. Data Normalization (0-255)

Min-Max normalization is a normalization technique that creates balanced value comparisons between data before and after the procedure by executing linear modifications on the original data (Saranya and Manikandan, 2013). The following formula may be used for this procedure. Eq.2 (Ribaric and Fratric, 2011):

$$Xnew = \frac{x-min(x)}{max(x)-min(x)} \tag{2}$$

Xnew: new value, x: old value in the array, max(X): max value in the array, and min(X): min value in the array.

Our approach must normalize every pixel in the range gray image, which is 0 to 255, so the equation will be multiplied by 255, in addition to most pixels produced from DCT transform are real numbers, so converted to integers using the (round) function. The final form of the equation normalization is as follows by Eq.3 (Tang et al., 2021) (Di, Magistrale and Gaizo, 2023):

$$Xnew = \text{round}\ \{\frac{x-min(x)}{max(x)-min(x)} \ \times 255\} \tag{3}$$

This equation is applied to all 1-D frequency domain signal values to shrink values between 0 and 255.

3.1.4. Construct a 2-D Gray Image

In this process, the 1-D frequency domain signal is divided into M equal segments, when the size of each segment $M=\sqrt{N}$, N indicates the 1-D frequency domain signal length, M×M equal to the length of the 1-D frequency domain (N), the number of segments is N/M, and X refers to a pixel. As shown in Fig. 4.

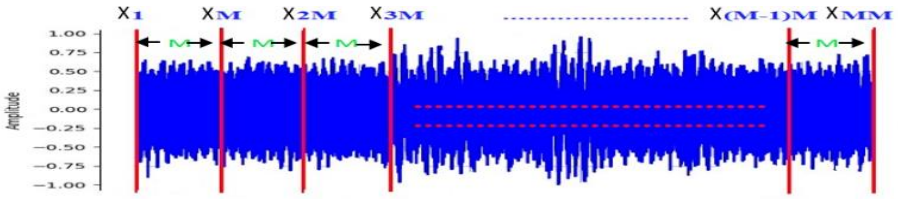

Fig. 4. Segmentation the 1-D frequency domain signal

After segmenting the signal, we got a definite number of segments which along with the segment size helped to define the rows and columns of the 2-D gray matrix. The number of rows in a gray matrix is equal to the size of each segment, and the number of columns in the matrix is equal to the number of segments. for instance, if our segment size is M, the number of the segments should be M, because the number of segments = N/M, then our size of the desired gray matrix is MxM. Finally, each segment is put column by column in the gray matrix, shown in Fig. 5 (Islam, Uddin and Kim, 2018) (Uddin et al., 2014).

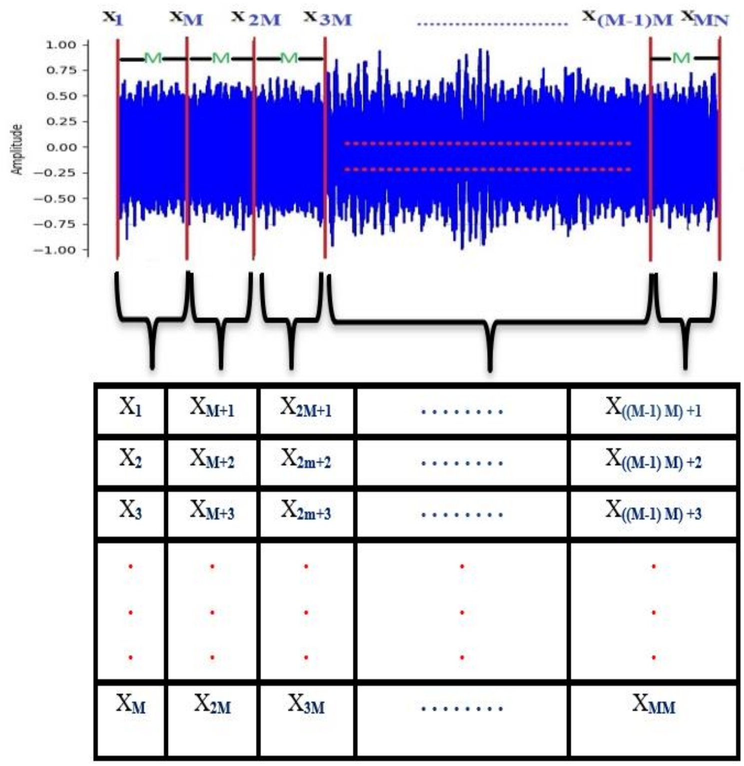

Fig. 5. Construct 2-D gray image.

Therefore, the result images from motor sound conversion into gray images are shown in Fig.6.

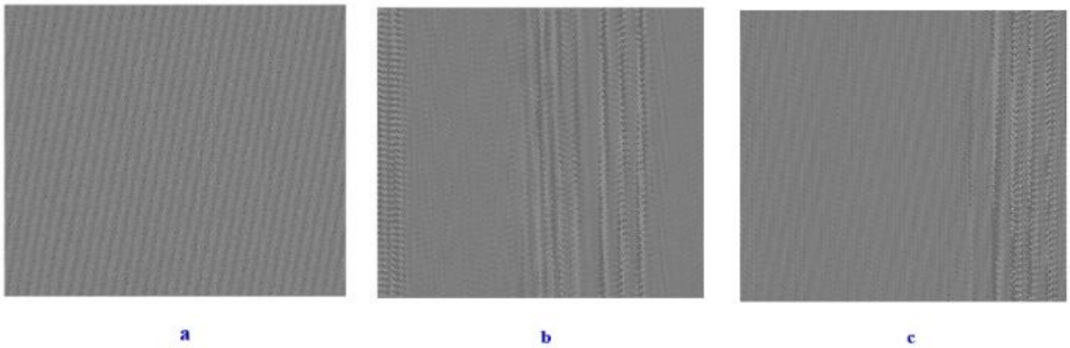

Fig. 6. Converted image from fault sound: (a)engine1_ good, (b) engine2_ broken, (c) engine3_heavy load

3.2. Second Phase: CNN for Diagnosis of Motor State

Feature extraction and classification are the two fundamental functions of Convolutional Neural Networks, which are mostly utilized for object identification and classification. The convolution and pooling layers use the proper activation function to extract features, while the fully connected layers use the SoftMax function to predict the output in the classification test. The convolutional computations are represented mathematically in Eq.4 (Ma et al., 2019).

$$hj = f\left(\sum_{i=1}^{n} \mathrm{xi} * w\{ij\} + bj\right) \tag{4}$$

where hj: indicates the output feature map, n: refers to filter size, xi: represents the output feature map for the previous convolutional layer, *: indicates the convolution operation, w{ij}: indicates the convolution filter, bj: denotes bais, and f: is the activation function.

Although CNNs have unique advantages in classification and transfer learning, they have some limitations, some of which we faced as challenges in this research:

1- CNNs require large and multiple data for training, the small size of the dataset and the lack of data diversity expose the model to the problem of overfitting and becoming unable to generalize well (unseen data).

2- To achieve high accuracy and improve the network, it is necessary to increase the depth of the network and change its structure, which results in complex computational operations and expensive resources.

3- Choosing the appropriate filter size for each layer, as choosing a fixed one for all layers fails to learn new features. However, different techniques are discovered to handle the overfitting problem, such as:

3.2.1. Dropout

It is an algorithm that depends on randomly canceling some neurons during the training task. It was presented as a research paper by Hinton et al. (2012) (Hinton et al., 2012). Neural feedback networks are used, and dropout is carried out by percentage, such as 50%, 30%, or 20%.

Eliminating the overfitting issue is one of the primary reasons for dropout. The dropout process is shown in Fig. 7.

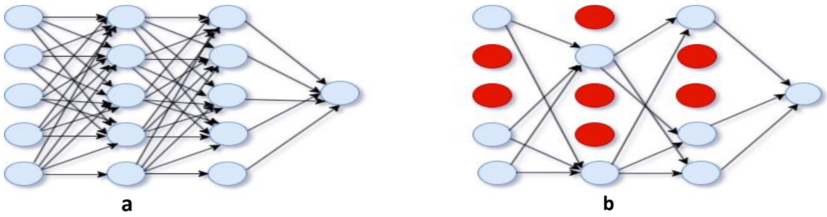

Fig. 7. Dropout operation: (a) Standard neuron network, (b) Applying dropout.

3.2.2. Batch Normalization

A technique used to normalize activation in the feature extraction part of DL networks to speed up training, solve the local minima problem, and fast convergence. It is usually added as a layer alongside the convolution and pooling layer to make the mean output close to 0 and the standard deviation close to 1 (Ioffe, 2017). In our research, batch normalization was used to overcome the previously mentioned problems, especially overfitting, and applying a dropout percentage of 50%. Fig. 8 illustrates the general structure of the proposed CNN network.

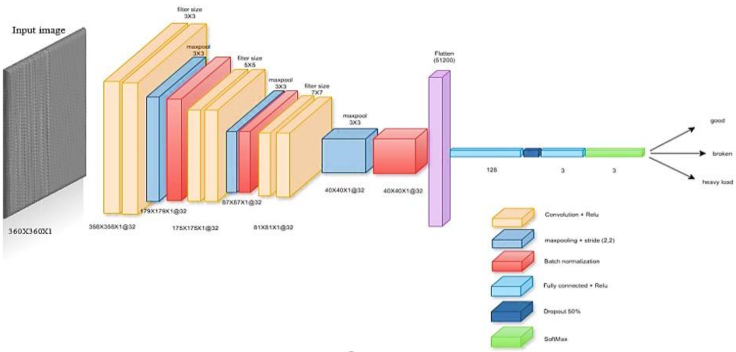

Fig. 8. CNN structure for model classification.

In this proposed network, three different sizes of filters were used to extract features better, The size of each filter is calculated according to the following Eq.5 (Xu et al., 2022).

$$S=2n+1 \tag{5}$$

where the convolution number is denoted by n. For example, the filter size in the first convolutional layer will be 2*1+1=3, etc. The sizes of the filters used in convolution sequence (3,3), (5,5), and (7,7) with stride (2,2), with a learning rate of 0.0001, the Adam optimizer was used to optimize the hybrid parameters, and using loss function Sparse Categorical Cross entropy, batch size=16, epochs= 20. Table 1 shows the layers and the parameters for each layer. Therefore, the Model performance monitor to avoid overfitting by using early stopping with patience=2 and validation model by splitting the training dataset into 20% validation and 80% training. Trained the model on a GPU (Graphic Process Unit) within the Colab environment to achieve high speed during data training.

Table 1. Model parameter.

Layer Type	Kernal Size	Stride	Channel	Output
Con1	(3,3)	(1,1)	32	(358,358,1)
MaxPooling1	(3,3)	(2,2)	32	(179,179,1)
BatchNormalization1	(1,1)	(1,1)	32	(179,179,1)
Con2	(5,5)	(1,1)	32	(175,175,1)
MaxPooling2	(5,5)	(2,2)	32	(87,87,1)
BatchNormalization2	(1,1)	(1,1)	32	(87,87,1)
Con3	(7,7)	(1,1)	32	(81,81,1)
MaxPooling3	(7,7)	(2,2)	32	(40,40,1)
Batch Normalization 3	(1,1)	(1,1)	32	(40,40,1)
Flatten	51200	-	1	51200
FC1	128	-	1	128
FC2	3	-	1	3
Softmax	3	-	1	3

4. EXPERIMENTS AND RESULTS

In this section, the model was trained and tested on a set of images resulting from converting the sounds of rotating motors to a database IDMT-ISA Electric Engine. with 16GB of RAM and a powerful image processing unit, the model was trained using the Google Colab environment and the GPU that Colab supplied. In addition to the test data sounds being recorded without noise, the model demonstrated high accuracy when tested on unseen test data and high efficiency when trained on data with noise accompanying the sound recording.

4.1. (IDMT-ISA ELECTRIC ENGINE) Dataset Description

The IDMT-ISA-ELECTRIC-ENGINE dataset, recorded at the Fraunhofer Institute for Digital Media Technology (IDMT) in March 2017 in Germany, Contains sound files of three similar electric motor engines (2ACT Motor Brushless DC 42BLF01, 4000 RPM, 24VDC). Resulting in three operational states: "engine1_good", "engine2_broken" and "engine3_heavyload" each file features only one active motor, ensuring that each engine can only be in one of the three states. Recording time between (31-34) minutes for the training process for each engine state and between (17-19) minutes for the testing process for each engine state, All recordings are cut into 3-second segments to facilitate evaluation over short periods. The dataset used in this research in the training folder consists of six different types of noises accompanying audio recording, as follows in Table 2 (Grollmisch, Abeßer and Liebetrau, 2019). However, the testing folder consists of recordings without background noise (pure).

Table 2. Noises Accompanying Audio Recording.

Noise	Description
atmo_high	Recording audio with factory environment noise(high)
atmo_low	Recording audio with factory environment noise(low)
atmo_medium	Recording audio with factory environment noise(medium)
stresstest	Record audio with stress change
talking	People talking during recording sounds
white_noise	Using speakers outside of the casing

4.2. Preprocessing Stage Analysis

After converting the sounds to the time domain applying DCT-based pre-processing and generating a two-dimensional grayscale image (360, 360,1), It is observed that the patterns obtained for each operating condition are nearly identical, for instance, the “engine1_good” operating condition all the image patterns are nearly similar to one another, as well as to the other operating conditions, despite the presence of noise accompanying the recording of these motor sounds. Therefore, we conclude that the pre-processing step is effective and successful in producing similar patterns for every operating condition under different conditions. It is essential to note that CNN does not scale the input picture sizes to retain significant portions of the images that arise from a little alteration in the machine noises.

4.3. Evaluation Metrics

To evaluate the performance of the proposed model and provide a robust analysis for performance, three measures were used, namely (Accuracy, Precision, Recall, and F1-Score). These measures are calculated using the following Eq.6,7,8,9 (Sivalingan, 2024):

$$\text{Accuracy} = \frac{TP(True\ Positive)+TN(True\ Negative)}{TP(True\ Positive)+TN(True\ Negative)+FP(False\ Positive)+FN(False\ Negaitve)} \tag{6}$$

$$\text{Precision} = \frac{TP(True\ Positive)}{TP(True\ Positive)+FP(False\ Positive)} \tag{7}$$

$$\text{Recall} = \frac{TP(True\ Positive)}{TP(True\ Positive)+FN(False\ Negative)} \tag{8}$$

$$\text{F1 Score} = \frac{2*Precision*Recall}{Precision+Recall} \tag{9}$$

4.4. CNN Performance and Result Analysis

During the model training, the training took 100 minutes and stopped until epochs 9. It did not finish all the specified epochs due to the lack of improvement in the validation accuracy used to monitor model performance to avoid overfitting. The training accuracy reached 97.1%. Fig.9 illustrates the fast convergence of the loss function during the training phase. The model achieved a high classification accuracy on the test data close to 99.7%, demonstrating that the suggested CNN can successfully finish the fault diagnosis classification task. Fig. 10 illustrates the confusion matrix.

However, the nature of the sound imposes certain features. For example, our method succeeded efficiently with the sounds of motors, but when tried on other sounds, its necessary other extract features.

However, Table 3 illustrates the results of classification using different metrics.

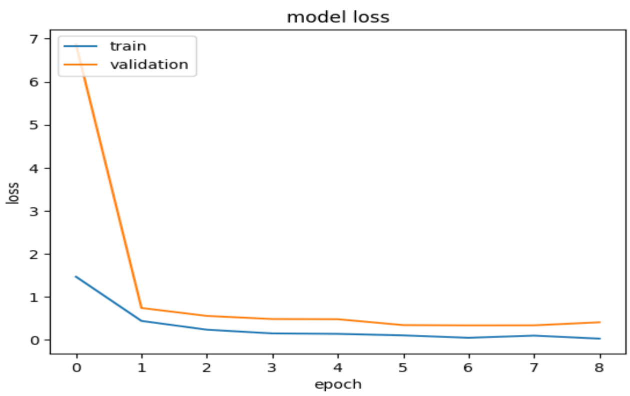

Fig. 9. Model loss function in the training.

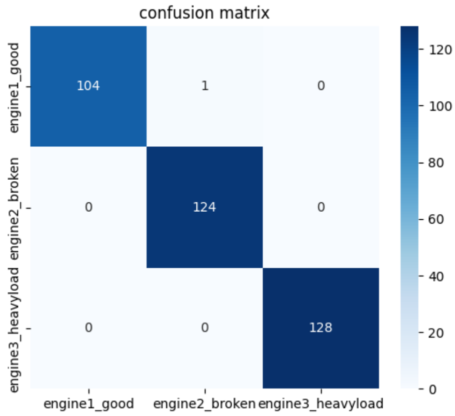

Fig. 10. Confusion matrix for testing data.

Table 3: The Performance Evaluation Classification by Different Metrics Metrics.

Class	Precision	Recall	F1 Score	support
engine1 good	1.00	0.99	1.00	105
engine2 broken	0.99	1.00	1.00	124
engine3 heavy load	1.00	1.00	1.00	128
accuracy			1.00	375
macro avg	1.00	1.00	1.00	375
weighted avg	1.00	1.00	1.00	375

Additionally, Table 4 further shows that the suggested strategy achieves faster convergence and higher accuracy when compared to the other methods.

Table 4. Comparison of our methodology with previous studies.

Reference	Dataset	Preprocessing	Network	Accuracy
(Grollmisch, Abeßer and Liebetrau, 2019)	IDMT_ISA_ ELECTRIC_ ENGINE dataset	Transform sound to Spectrogram mage by applying STFT.	DNN	98.8%
(Jiang, Lan and Li, 2020)	IDMT_ISA_ ELECTRIC_ ENGINE dataset & DCASE 2020 Task 2 Dataset	Transform sound to log Mel Spectrogram image.	Autoencoder	75.99%
(Strantzalis *et al.*, 2022)	IDMT_ISA_ ELECTRIC_ ENGINE dataset	Transform sound to Spectrogram image by applying FFT.	TensorFlow lite CNN model and generated c-model STM tools	100%
		Transform sound to Spectrogram image by applying FFT.	TensorFlow lite CNN model and generated edge impulse	97.8%
This Study	IDMT_ISA_ ELECTRIC_ ENGINE dataset	Transform sound to the gray image by applying DCT.	CNN	99.7%

5. CONCLUSION

This study suggests an approach to diagnose the condition of rotating motors based on machine sounds. The primary contribution of this article is preprocessing the sounds based on DCT transform with 2-D grayscale images and proposing a CNN network capable of fault diagnosis without overfitting due to a lack of dataset and network complexity using Adam optimizer and Batch Normalization. Comparing the suggested technique to conventional approaches, experimental findings show that it is accurate and converges quickly under various noise situations, due to the use of DCT which helped in extracting features that affected the classification process. The model achieved an accuracy of 99.7%. However, this method has some limitations, including that the proposed approach belongs to supervised learning, in addition to the limited size of the dataset utilized in this study. Future work will include sound that contains more than one motor working at the same time and try to isolate each sound and classify the faults.

ACKNOWLEDGMENT

The author expresses gratitude to the reviewers and editors for their work, My great gratitude to my supervisor, Dr. Prof. Tawfiq Al-assadi. No government, private, or nonprofit organization provided money for this study.

6. REFERENCES

Ahmad, Z., Hasan, M.J. and Kim, J.M. (2022) ‘Transfer Learning with 2D Vibration Images for Fault Diagnosis of Bearings Under Variable Speed’, in proceedings of 21st International conference on intelligent systems design and applications (ISDA 2021). Springer Science and Business Media Deutschland GmbH, pp. 154–164. Available at: https://doi.org/10.1007/978-3-030-96308-8_14.

Ahmed, H.O.A. and Nandi, A.K. (2021) ‘Connected Components-based Colour Image Representations of Vibrations for a Two-stage Fault Diagnosis of Roller Bearings Using Convolutional Neural Networks’, Chinese Journal of Mechanical Engineering (English Edition), 34(1), pp. 1–21. Available at: https://doi.org/10.1186/s10033-021-00553-8.

Alaa, R., Hussein, E.A. and Al-Libawy, H. (2024) ‘OBJECT DETECTION ALGORITHMS IMPLEMENTATION ON EMBEDDED DEVICES: CHALLENGES AND SUGGESTED SOLUTIONS’, Kufa Journal of Engineering, 15(3), pp. 148–169. Available at: https://doi.org/10.30572/2018/KJE/150309.

Betta, G. et al. (2002) ‘A DSP-based FFT-analyzer for the fault diagnosis of rotating machine based on vibration analysis’, IEEE Transactions on Instrumentation and Measurement, 51(6), pp. 1316–1322. Available at: https://doi.org/10.1109/TIM.2002.807987.

Di, T., Magistrale, L. and Gaizo, D. Del (2023) Enhancing Drone Spectra Classifi-cation: A Study on Data-Adaptive Pre-processing and Efficient Hard-ware Deployment, IEEE (Institute of Electrical and Electronics Engineers). Politecnico.

Freire Moraes, G.H., Ribeiro Junior, R.F. and Gomes, G.F. (2024) ‘Fault Classification in Diesel Engines Based on Time-Domain Responses through Signal Processing and Convolutional Neural Network’, Vibration, 7(4), pp. 863–893. Available at: https://doi.org/10.3390/vibration7040046.

Grollmisch, S., Abeßer, J. and Liebetrau, J. (2019) ‘Sounding Industry: Challenges and Datasets for Industrial Sound Analysis’, 2019 27th European Signal Processing Conference (EUSIPCO), pp. 1–5. Available at: https://doi.org/978-9-0827-9703-9/19/$31.00.

Han, S., Oh, S. and Jeong, J. (2021) ‘Bearing Fault Diagnosis Based on Multiscale Convolutional Neural Network Using Data Augmentation’, Journal of Sensors, 2021. Available at: https://doi.org/10.1155/2021/6699637.

Hao Yijia et al. (2020) ‘Multi-Scale CNN based on Attention Mechanism for Rolling Bearing Fault Diagnosis’, in the 9th Asia-Pacific International Symposium on Advanced Reliability and Maintenance Modeling (APARM). IEEE, pp. 1–5. Available at: https://doi.org/978-1-7281-7102-9/20/$31.00©2020IEEE.

He, J., Lv, Z. and Chen, X. (2023) ‘Rolling bearing fault diagnosis method based on 2D grayscale images and Wasserstein Generative Adversarial Nets under unbalanced sample condition’, Complex Engineering Systems, 3(3), pp. 1–15. Available at: https://doi.org/10.20517/ces.2023.20.

Hinton, G.E. et al. (2012) ‘Improving neural networks by preventing co-adaptation of feature detectors’, Neural Information Processing Systems (NeurIPS), pp. 1–5. Available at: http://www.utstat.toronto.edu/.

Hoang, D.T. and Kang, H.J. (2019) ‘Rolling element bearing fault diagnosis using convolutional neural network and vibration image’, Cognitive Systems Research, 53, pp. 1–9. Available at: https://doi.org/10.1016/j.cogsys.2018.03.002.

Ioffe, S. (2017) 'Batch Renormalization: Towards Reducing Minibatch Dependence in Batch-Normalized Models', in 31st Conference on Neural Information Processing Systems (NIPS 2017), Long Beach, CA, USA, pp. 1–9.

Islam, R., Uddin, J. and Kim, J.-M. (2018) 'Texture analysis based feature extraction using Gabor filter and SVD for reliable fault diagnosis of an induction motor Texture analysis based feature extraction using Gabor filter and SVD 21', Int. J. Information Technology and Management, 17(2), pp. 20–32.

Jaber, A. (2011) ' LOW COMPLEXITY EMBEDDED IMAGE COMPRESSION ALGORITHM USING SUBBAND CODING OF THE DCT COEFFICIENTS', Kufa Journal of Engineering, 3(2), pp. 1–16.

Jiang, H., Lan, B. and Li, H. (2020) 'ABNORMAL SOUND DETECTION SYSTEM BASED ON AUTOENCODER Technical Report', Detection and Classification of Acoustic Scenes and Events, pp. 1–4.

Kang, M. and Kim, J.M. (2013) 'Singular value decomposition based feature extraction approaches for classifying faults of induction motors', Mechanical Systems and Signal Processing, 41(1–2), pp. 348–356. Available at: https://doi.org/10.1016/j.ymssp.2013.08.002.

Lam, E.Y. and Goodman, J.W. (2000) 'A mathematical analysis of the DCT coefficient distributions for images', IEEE Transactions on Image Processing, 9(10), pp. 1661–1666. Available at: https://doi.org/10.1109/83.869177.

Lei, Y. et al. (2016) 'An Intelligent Fault Diagnosis Method Using Unsupervised Feature Learning Towards Mechanical Big Data', IEEE Transactions on Industrial Electronics, 63(5), pp. 3137–3147. Available at: https://doi.org/10.1109/TIE.2016.2519325.

Li, X., Zhang, W. and Ding, Q. (2019) 'Understanding and improving deep learning-based rolling bearing fault diagnosis with attention mechanism', Signal Processing, 161, pp. 136–154. Available at: https://doi.org/10.1016/j.sigpro.2019.03.019.

Liu, H., Li, L. and Ma, J. (2016) 'Rolling Bearing Fault Diagnosis Based on STFT-Deep Learning and Sound Signals', Shock and Vibration, 2016, pp. 1–13. Available at: https://doi.org/10.1155/2016/6127479.

Ma, P. et al. (2019) 'A novel bearing fault diagnosis method based on 2D image representation and transfer learning-convolutional neural network', Measurement Science and Technology, 30(5), pp. 1–23. Available at: https://doi.org/10.1088/1361-6501/ab0793.

Manhertz, G. and Bereczky, A. (2021) 'STFT spectrogram based hybrid evaluation method for rotating machine transient vibration analysis', Mechanical Systems and Signal Processing, 154, pp. 1–16. Available at: https://doi.org/10.1016/j.ymssp.2020.107583.

Misra, S. et al. (2022) 'Fault Detection in Induction Motor Using Time Domain and Spectral Imaging-Based Transfer Learning Approach on Vibration Data', Sensors, 22(21), pp. 1–16. Available at: https://doi.org/10.3390/s22218210.

Mohammed, H.A., Kareem, S.W. and Mohammed, A.S. (2022) 'A COMPARATIVE EVALUATION OF DEEP LEARNING METHODS IN DIGITAL IMAGE CLASSIFICATION', Kufa Journal of Engineering, 13(4), pp. 53–69. Available at: https://doi.org/10.30572/2018/KJE/130405.

Mukherjee, I. and Tallur, S. (2021) 'Light-Weight CNN Enabled Edge-Based Framework for Machine Health Diagnosis', IEEE Access, 9, pp. 84375–84386. Available at: https://doi.org/10.1109/ACCESS.2021.3088237.

Nguyen, D. et al. (2013) 'Highly reliable state monitoring system for induction motors using dominant features in a two-dimension vibration signal', New Review of Hypermedia and Multimedia, 19(3–4), pp. 248–258. Available at: https://doi.org/10.1080/13614568.2013.832407.

Ribaric, S. and Fratric, I. (2011) 'Experimental Evaluation of Matching-Score Normalization Techniques on Different Multimodal Biometric Systems', IEEE International Conference on Biometrics: Theory, Applications and Systems (BTAS), pp. 1–4.

Saranya, C. and Manikandan, G. (2013) 'A Study on Normalization Techniques for Privacy Preserving Data Mining', International Journal of Engineering and Technology (IJET), 5, pp. 2701–2704.

Sivalingan, H. (2024) 'CLOUD-SMART SURVEILLANCE: ENHANCING ANOMALY DETECTION IN VIDEO STREAMS WITH DF-CONVLSTM-BASED VAE-GAN', Kufa Journal of Engineering, 15(4), pp. 125–140. Available at: https://doi.org/10.30572/2018/KJE/150409.

Strantzalis, K. et al. (2022) 'Operational State Recognition of a DC Motor Using Edge Artificial Intelligence', Sensors, 22(24). Available at: https://doi.org/10.3390/s22249658.

Tang, H. et al. (2021) 'A novel intelligent fault diagnosis method for rolling bearings based on wasserstein generative adversarial network and convolutional neural network under unbalanced dataset', Sensors, 21(20), pp. 1–24. Available at: https://doi.org/10.3390/s21206754.

Tran, V.T. et al. (2013) 'An application to transient current signal based induction motor fault diagnosis of Fourier-Bessel expansion and simplified fuzzy ARTMAP', Expert Systems with Applications, 40(13), pp. 5372–5384. Available at: https://doi.org/10.1016/j.eswa.2013.03.040.

Uddin, J. et al. (2014) 'Reliable fault classification of induction motors using texture feature extraction and a multiclass support vector machine', Mathematical Problems in Engineering, 2014, pp. 1–10. Available at: https://doi.org/10.1155/2014/814593.

Wang, X., Mao, D. and Li, X. (2021) 'Bearing fault diagnosis based on vibro-acoustic data fusion and 1D-CNN network', Measurement, 173, pp. 1–13. Available at: https://doi.org/10.1016/j.measurement.2020.108518.

Xu, M. et al. (2022) 'Bearing-Fault Diagnosis with Signal-to-RGB Image Mapping and Multichannel Multiscale Convolutional Neural Network', Entropy, 24(11), pp. 1–20. Available at: https://doi.org/10.3390/e24111569.

Zhang, Z., Wang, Y. and Wang, K. (2012) 'Fault diagnosis and prognosis using wavelet packet decomposition, Fourier transform and artificial neural network', Journal of Intelligent Manufacturing, 24(6), pp. 1–15. Available at: https://doi.org/10.1007/s10845-012-0657-2.

Zhao, H., Sun, S. and Jin, B. (2018) 'Sequential Fault Diagnosis Based on LSTM Neural Network', IEEE Access, 6, pp. 12929–12939. Available at: https://doi.org/10.1109/ACCESS.2018.2794765.

Zhou, F. et al. (2021) 'A sparse denoising deep neural network for improving fault diagnosis performance', Signal, Image and Video Processing, 15(8), pp. 1889–1898. Available at: https://doi.org/10.1007/s11760-021-01939-w.