SPECTA Journal of Technology Vol 10, No 2 August, 2026 pg: 283 – 293
DOI: 10.35718/specta.v10i2.1344

SPECTA Journal of Technology
E-ISSN : 2622-9099
P-ISSN : 2549-2713
Homepage jurnal: https://journal.itk.ac.id/index.php/sjt

Effect of MFCC Feature Dimensionality on Deep Learning-Based Fault Classification of Industrial Machinery Using Acoustic Signals

Thorikul Huda[1*], Min-Fu Hsieh[2], Faaris Mujaahid[3], Muhammad Rhido Dewanto[4]
[1*2]Department of Electrical Engineering, College of Electrical Engineering and Computer Science, National Cheng Kung University (NCKU), No. 1, University Road, East District, Tainan City 70101, Taiwan
[3]Department of Electrical Engineering, Universitas Muhammadiyah Yogyakarta (UMY), Building F3, 2nd Floor, Jl. Brawijaya, Geblagan, Tamantirto, Kasihan, Bantul, Yogyakarta 55183, Indonesia
[14]Department of Electrical Engineering, Informatics, and Business, Faculty of Science and Information Technology, Institut Teknologi Kalimantan (ITK), Jl. Soekarno Hatta KM 15, Karang Joang, Balikpapan 76127, Indonesia

*Corresponding email: thorikulhuda@gmail.com

Received: 24/March/2025 Revised: 29/August/2026
Accepted: 30/August/2026 Published: 31/August/2026

To cite this article:
Huda, T., Hsieh, M. F., Mujaahid, F & Dewanto, M. R (2026). Effect of MFCC Feature Dimensionality on Deep Learning - Based Fault Classification of Industrial Machinery Using Acoustic Signals. *SPECTA Journal of Technology,* 10(2), 283-293. 10.35718/specta.v10i2.1344

Abstract

This study investigates the effect of MFCC feature dimensionality on deep learning-based induction motor fault classification using acoustic signals. Four conditions were classified: normal, unbalance, horizontal misalignment, and vertical misalignment. LSTM models using 13, 26, and 39 MFCC features achieved validation accuracies of 97.4%, 96.12%, and 96.3%, respectively. The LSTM was also compared with CNN, RNN, and FNN models. The CNN achieved the highest average training accuracy of approximately 97%, while the LSTM achieved the best testing accuracy of approximately 94% using 39 MFCC features. Overall, the LSTM and CNN outperformed the RNN and FNN. The results indicate that increasing MFCC dimensionality did not improve classification performance, as the 13-feature configuration achieved the highest validation accuracy with the lowest input complexity. Future research may investigate adaptive feature selection to improve computational efficiency and classification performance.

Keywords: MFCC, Deep Learning, Electric Motor, Fault Classification.

Abstrak

Penelitian ini mengkaji pengaruh dimensi fitur MFCC terhadap klasifikasi kerusakan motor induksi berbasis deep learning menggunakan sinyal akustik. Empat kondisi yang diklasifikasikan meliputi kondisi normal, ketidakseimbangan, misalignment horizontal, dan misalignment vertikal. Model LSTM dengan 13, 26, dan 39 fitur MFCC masing-masing menghasilkan akurasi validasi sebesar 97,4%, 96,12%, dan 96,3%. Kinerja LSTM juga dibandingkan dengan model CNN, RNN, dan FNN. CNN memperoleh rata-rata akurasi pelatihan tertinggi sekitar 97%, sedangkan LSTM menghasilkan akurasi pengujian terbaik sekitar 94% dengan 39 fitur MFCC. Secara umum, LSTM dan CNN menunjukkan kinerja lebih baik dibandingkan RNN dan FNN. Hasil penelitian menunjukkan bahwa peningkatan jumlah fitur MFCC tidak meningkatkan kinerja klasifikasi karena konfigurasi 13 fitur menghasilkan akurasi validasi tertinggi dengan kompleksitas input paling rendah. Penelitian selanjutnya dapat mengembangkan pemilihan fitur adaptif untuk meningkatkan efisiensi komputasi dan kinerja klasifikasi.

Kata Kunci: MFCC, Deep Learning, Motor Listrik, Klasifikasi Kerusakan.

1. Introduction

Fault diagnosis of electrical rotating machines has long been recognized as a critical area of research due to its impact on operational reliability, safety, and maintenance efficiency in industrial environments. Unanticipated failures in induction motors can lead to significant downtime, economic loss, and potential safety hazards (Ahmed Murtaza et al., 2024; Glowacz, 2018). As such, robust condition monitoring techniques are essential for early detection and classification of faults, thereby ensuring the continuous and optimal performance of industrial systems.

Acoustic signal analysis has emerged as a promising non-invasive method for diagnosing mechanical and electrical faults in induction motors (Germen et al., 2014) (Glowacz & Glowacz, 2017). Among the various signal processing techniques, the Mel-Frequency Cepstral Coefficient (MFCC) has gained popularity due to its ability to capture relevant audio features that reflect the operating state of machinery (Abdul & Al-Talabani, 2022; Fang et al., 2001; Mishra et al., 2023). However, the effectiveness of MFCC-based feature extraction depends significantly on the number of coefficients used, which directly affects the accuracy and computational complexity of the classification process.

Therefore, some work focuses on applying machine learning techniques to improve the fault diagnosis performance recently, in motor fault diagnosis the implemented system was based on acoustic signals of three-phase induction motors was performed using 3 classifiers LDA (Linear Discriminant Analysis), NBC (Naive Bayes Classifier), CT (Classification Tree) (Glowacz et al., 2021). Using GoogleNet model result shows that GoogleNet achieves 98.23% accuracy in the fault classification for denoised signals, while only 89.66% accuracy for the original signals (Yang et al., 2022). These GRU-based non-linear predictive denoising autoencoders (GRU-NP-DAEs) are trained with strong generalization ability for each different fault pattern (Liu et al., 2018).

The objective of this study is to investigate the influence of the number of MFCC features on the performance of deep learning models for fault classification of three-phase induction motors using acoustic signals. In particular, the proposed approach focuses on distinguishing between normal operation, imbalance, and two types of misalignment faults (horizontal and vertical). We employ Long Short-Term Memory (LSTM) networks as the primary classification model and benchmark its performance against three others widely used deep learning architectures: Feed-Forward Neural Network (FNN), Convolutional Neural Network (CNN), and Recurrent Neural Network (RNN).

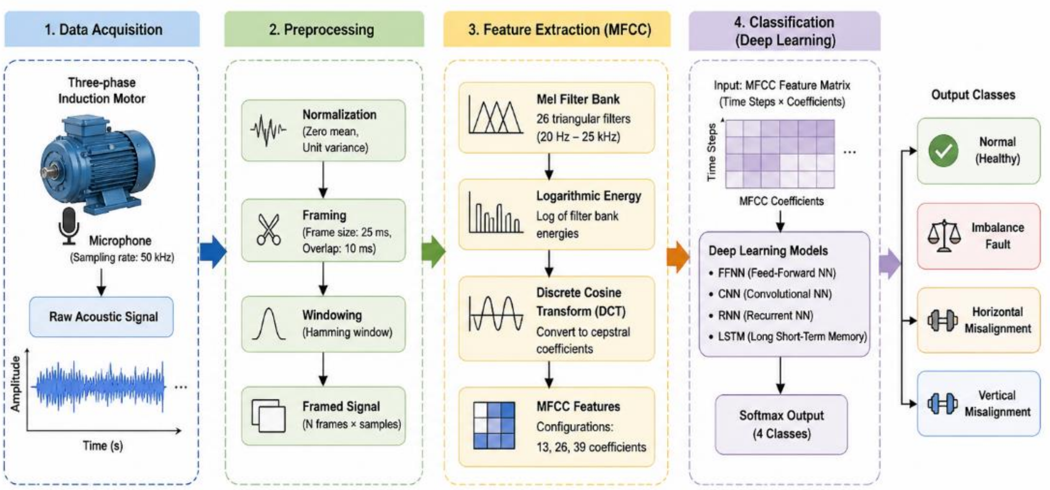


Figure 1. Overall pipeline of the proposed acoustics signals based fault classification system

MaFaulDa was used to develop the experiment in this study, the database includes multiple types of faults with different levels of severity and rotational speeds. It is composed of multivariate time-series signals acquired by sensors from Machinery Fault Simulator (MFS) test rig that simulates the machine under normal operation and fault conditions (Khan et al., 2021; Nath et al., 2021). Figure 1 presents the overall pipeline of the proposed acoustic signal-based fault classification system. The process starts with data acquisition from a three-phase induction motor using a microphone to capture raw acoustic signals. Preprocessing includes normalization, framing, and windowing to prepare the signals for feature extraction. MFCC features are then computed using a Mel filter bank, logarithmic energy, and discrete cosine transform, producing cepstral coefficients. These features are subsequently fed into deep learning classifiers, including FNN, CNN, RNN, and LSTM, to predict machine conditions as normal, imbalance, horizontal misalignment, or vertical misalignment. The diagram illustrates the end-to-end workflow from signal acquisition to fault classification.

2.1. Motor Fault Simulator

Machinery Fault Simulator (MFS) test rig that simulates the machine under normal operation and five fault conditions, namely, mass imbalance, horizontal and vertical misalignment, and overhang and under-hang bearing faults at different rotation speeds, Figure 2 shows the data acquisition setup of the MaFaulDa system (SpectraQuest Inc., Machinery Fault Simulator, 2021).

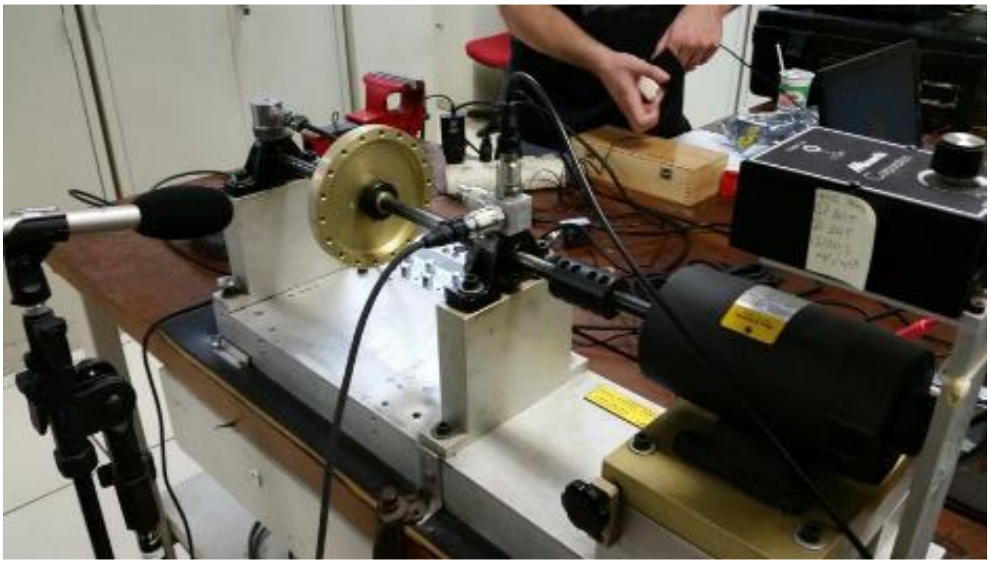


Figure 2. The acquisition device of MaFaulDa (Wang & Cao, 2023)

1. Imbalance is the state of a rotating component when its mass distribution is uneven in relation to the axis of rotation. This results in unbalanced centrifugal forces and moments, which cause dynamic reactions in bearings, resulting in vibration and noise, thus leading to bearing damage. A symptom of imbalance is an increase in the amplitude of the rotational frequency that occurs in the diagnostic signal.
2. The misalignment in the induction motor drive concerns the connection between the motor and the loading machine

2.2. Extract Audible Noise Feature Using MFCC

In this study, the MFCC extraction process is configured with the following parameters: the audio signal is segmented into frames of 25 ms with an overlap of 10 ms. A Hamming window is applied to each frame to reduce spectral leakage. The Mel filter bank consists of 26 triangular filters distributed across the frequency range. The sampling frequency of the signal is 50 kHz, consistent with the MaFaulDa dataset specification. After applying the Mel filter bank, the logarithmic energy is computed and transformed using Discrete Cosine Transform (DCT) to obtain the cepstral coefficients. Three configurations of MFCC features are evaluated in this study: 13, 26, and 39 coefficients.

MFCCs are the most widely used perceptual features because of their success in embodying the speech amplitude spectrum concisely. These features are obtained after applying a series of functions on the audio signal . It starts with applying a filter that amplifies the higher frequencies in the signal, which elevates the energy at these parts. After amplification, the entire signal is fragmented into small frames ranging typically between 20ms and 40ms. The next process in the line is applying the Hamming window function which eliminates the edge effects and assimilates all the frequency lines that are placed close together.

The output of the Hamming window is expressed in eq. (1).

$$y(n) = x(n) \times w(n) \tag{1}$$

$$w(n) = 0.54 - 0.46 \times \left(\cos\frac{2\pi n}{N-1}\right), 0 \leq n \leq N-1 \tag{2}$$

In the eq. (2), *N* represents the number of samples in each frame and in eq. (1), *y(n)* is the output signal. The input signal is represented by *x(n)* and *w(n)* represents the Hamming window operation. A time signal is composed of its constituent frequencies and a FT on the time signals gives us the corresponding complex-valued function in the frequency domain. This transformation is represented by eq. (3).

$$F(\xi) = \int_{-\infty}^{\infty} f(x) e^{-2\pi i x \xi} dx \tag{3}$$

$$M(f) = 1125 \times \ln\left(1 + \frac{f}{700}\right) \tag{4}$$

The Mel scale helps to incorporate this behavior into the extracted features so that it is analogous to how humans process audio. The Mel transform is given by eq. (4). In the above equation, f is the frequency expressed in hertz. To represent the power spectrum on the Mel scale, filter banks are applied to it.

2.3. Long Short-Term Memory (LSTM)

Long Short-term Memory (LSTM) is proposed by Hochreiter and Schmid Huber (Hochreiter & Schmidhuber, 1997)to solve the gradient vanishing problem in Recurrent Neural Network (RNN). In addition to the hidden state vector h_t, LSTM maintains a memory cell c_t encoding memory of observed information up to the time step t. The behavior of the memory cell is determined by three gates: input gate it , output gate i_t and forget gate h_t. The updating equations are given as follows equation 5 - 7:

$$i_t = \sigma(w_i[h_{t-1}, x_t] + b_i) \tag{5}$$

$$f_t = \sigma\left(w_f[h_{t-1}, x_t] + b_f\right) \tag{6}$$

$$o_t = \sigma(w_o[h_{t-1}, x_t] + b_o) \tag{7}$$

3. Deep Learning for Fault Classification

3.1. System Overview

The overall workflow of the proposed fault classification system is illustrated as a sequential pipeline consisting of four main stages: data acquisition, preprocessing, feature extraction, and classification. First, acoustic signals are obtained from the MaFaulDa dataset using a microphone sensor. The raw audio signals are then preprocessed through normalization and segmentation into short frames. Next, Mel-Frequency Cepstral Coefficients (MFCC) are extracted to represent the acoustic characteristics of the motor conditions. In this study, three different feature configurations are evaluated, namely 13, 26, and 39 MFCC coefficients. Finally, the extracted features are used as input to deep learning models (FFNN, CNN, RNN, and LSTM) for classification into four classes: normal, imbalance, horizontal misalignment, and vertical misalignment. This structured pipeline ensures a clear relationship between each processing stage and improves reproducibility of the proposed method.

3.2. Noise Dataset

In this paper, datasets provided by Machinery Fault Database (MaFaulDa) Website are employed to verify the proposed method (*MAFAULDA:MachineryFaultDatabase [Online]*, 2021). The website provides a huge database composed of 1951 multivariate time-series acquired by sensors on a SpectraQuest's Machinery Fault Simulator (MFS) Alignment-Balance-Vibration (ABVT).

Table 1 summarizes the different simulated conditions available in the MaFaulDa dataset. The complete dataset comprises six machine states: normal operation, unbalance, horizontal misalignment, vertical misalignment, inner bearing fault, and outer bearing fault. However, this study considers only four

SPECTA Journal of Technology Vol 10, No 2 August, 2026 pg: 283 – 293
DOI: 10.35718/specta.v10i2.1344

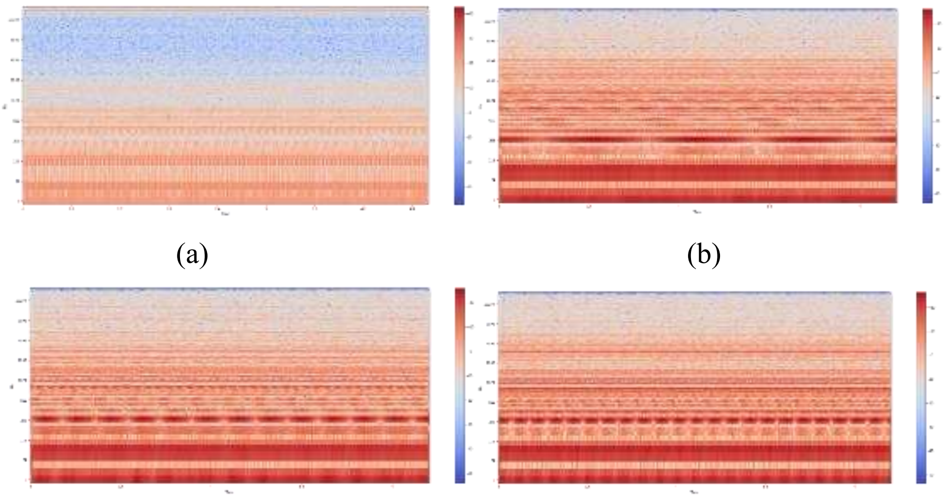
conditions: normal operation, unbalance, horizontal misalignment, and vertical misalignment. The test setup utilized three Industrial IMI Sensors model 601A01 accelerometers positioned in the radial, axial, and tangential directions; one IMI Sensors model 604B31 triaxial accelerometer; a Monarch Instrument MT-190 analog tachometer; a Shure SM81 microphone; and National Instruments NI 9234 four-channel analog acquisition modules. Each sequence was recorded for 5 seconds at a sampling rate of 50 kHz, resulting in 250,000 samples per sequence. The table presents the severity levels, number of measurements, and operating speeds for each condition considered in this study.

Table 1. Summary of different conditions.

Conditions	Severity Levels	Number of Measurements	Speed of Operation
Normal	No Fault	49	737 to 3886 rpm
Unbalance	10 g, 15 g 20 g, 25 g, 30 g, 35 g	333	737 to 3886 rpm
Horizontal Misalignment	0.5 mm, 1.0 mm, 1.5 mm, 2.0 mm	197	737 to 3886 rpm
Vertical Misalignment	0.51 mm, 0.63 mm, 1.27 mm, 1,4 mm, 1.78mm, 1,9 mm.	301	737 to 3886 rpm

For imbalance condition load values within the range from 6 g to 35 g. For each load value below 30 g, the rotation frequency assumed in the same 49 values employed in the normal operation case. For loads equal to or above 30 g, however, the resulting vibration makes it impracticable for the system to achieve rotation frequencies above 3300 rpm, limiting the number of distinct rotation frequencies.

Figure 3 illustrates the MFCC patterns of audible noise data under different conditions. Subfigure (a) shows the normal condition without distortion, while (b) represents amplitude imbalance. Subfigures (c) and (d) depict vertical and horizontal misalignment, respectively, indicating temporal and frequency shifts in the signal. This visual analysis highlights how different types of distortions affect MFCC representations, which is critical for subsequent speech or noise classification tasks.



(a) (b) (c) (d)

Figure 3. Different MFCC pattern audible noise data (a) Normal Condition, (b) Imbalance, (c) Vertical Misalignment, (d) Horizontal Misalignment.

3.3. Deep Learning Configuration

This research focuses on the use of Deep Learning models for classification tasks, specifically FNN, LSTM, RNN and CNN. These models are favored for their ability to analyze time-series or sequential data and provide accurate predictions even with raw input data. The models classify outcomes into multiple classes thanks to the various layers included in their design. The choice of model structure, including the number of layers and units, depends on the nature and complexity of the dataset. In some cases, the model may be more complex if the input data is nonlinear and intricate. The details of the model structures used in the research are provided in table 2.

Table 2. Comparison of Deep Learning Architectures Used in This Study

Model	Layers Type	Number of Layers	Units/Filters	Activation Function
FNN	Dense	3	64, 128, 128	ReLU
	Output	1	4	Softmax
LSTM	LSTM	2	64, 128	ReLU
	Dense	1	128	ReLU
	Output	1	4	Softmax
RNN	SimpleRNN	2	64, 128	ReLU
	Dense	1	128	ReLU
	Output	1	4	Softmax
CNN	Conv2D	2	64, 128	-
	MaxPooling2D	2	3x3	-
	Dense	1	128	ReLU
	Output	1	4	Softmax

Table 2 presents a unified comparison of the deep learning architectures used in this study, including FNN, LSTM, RNN, and CNN. The table highlights the structural configuration of each model in terms of layer types, number of layers, number of units or filters, and activation functions. By organizing the architectures into a single table, it becomes easier to compare their complexity and design characteristics. All models use a Softmax activation function in the output layer to perform multi-class classification across four fault conditions. This unified representation ensures clarity and supports a fair comparison of model performance, as all architectures are trained using the same MFCC feature inputs. During the training process, the hyperparameters of deep learning models varied, and their values which resulted in optimal performance are given as number in table 3.

Table 3. Hyperparameters setting.

Hyperparameter	Unit
Batch size	128
Number of epochs	100
Learning rate	1×10^{-4}
Dataset split ratio	75:15:10

4. Result

4.1. LSTM Result

Figure 3 shows a training comparison of LSTM model performance using three different configurations based on the number of Mel-Frequency Cepstral Coefficients (MFCC) data inputs: 13, 26, and 39 MFCC features. The graphs display the model's accuracy and loss across training and testing phases, providing insights into how varying the number of input features impacts the model's learning effectiveness and generalization capabilities.

The accuracy graphs exhibit a rapid ascent in model performance during the initial epochs across all three configurations, indicating efficient initial learning regardless of the number of features. All three curves converge to a similar performance level, with slight variations in the plateau phase. This suggests that all three models can capture sufficient information to achieve high performance, although the models with more features (26 and 39) may provide a slight edge in achieving peak accuracy slightly earlier. This could imply that additional features provide more useful information that aids in quicker convergence.

The testing graphs display in figure 4 a sharp decline in the initial training phase for all feature configurations, with all three curves stabilizing at a low level of loss thereafter. The closeness of the training and testing loss curves indicates good generalization across all feature sets. Notably, the models do not show significant divergence in loss between the training and test datasets, suggesting none of the configurations is prone to overfitting. However, the 39 MFCC feature model exhibits slightly lower and more stable loss values, possibly indicating more robust learning from a richer feature set.

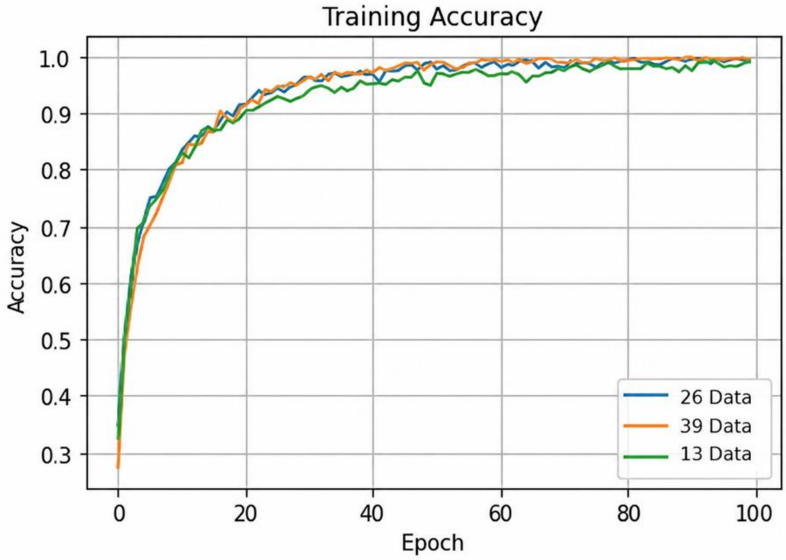


Figure 4. Training accuracy graphic comparison between 13, 26, and 39 input number

In a more detailed view of the loss behavior, especially in the latter epoch, we observe minor fluctuations which are common in deep learning models as they adapt to the inherent noise and complexity in the data. These fluctuations are minimal and do not significantly impact the overall stability, reinforcing the effectiveness of all three configurations.

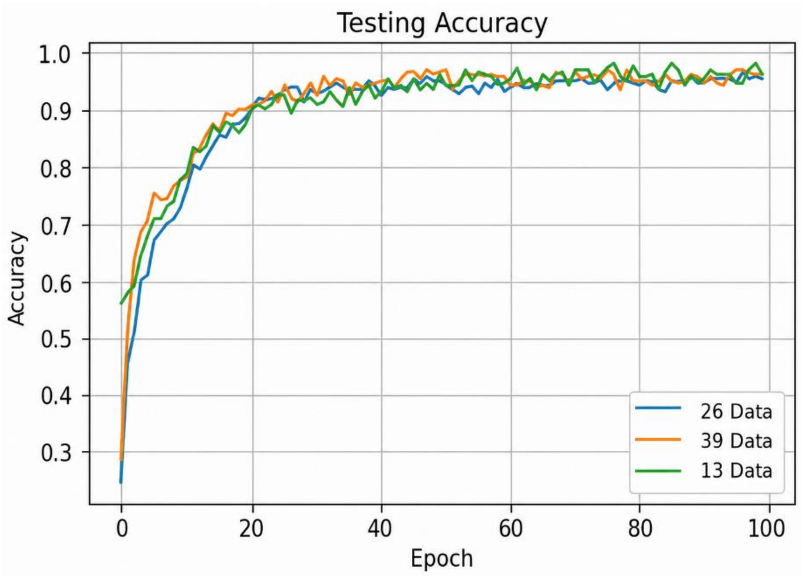


Figure 5. Testing accuracy graphic comparison between 13, 26, and 39 input number

The comparison across different numbers of MFCC features illustrates that while all models perform competently, models with a higher number of features (26 and 39) might provide marginal benefits in terms of early learning speed and stability in loss metrics. However, the performance differences are relatively small, indicating that even the 13 MFCC feature model is sufficiently capturing the necessary information for accurate fault diagnosis and classification. This analysis can guide decisions on model complexity versus performance trade-offs, suggesting that increasing the number of features can enhance model performance but up to a certain point beyond which the gains may diminish.

4.3. Comparison LSTM result with other types of NN algorithms

This result uses 39 MFCC data, the accuracy score is a metric that measures the model's ability to predict the correct output. The scores range from 0 to 1, with 1 being the best possible score. The training data show in figure 6 the CNN model has the highest accuracy, with an average score of 0.97. It consistently performs well across all trials, with only a few dips in accuracy below 0.97.

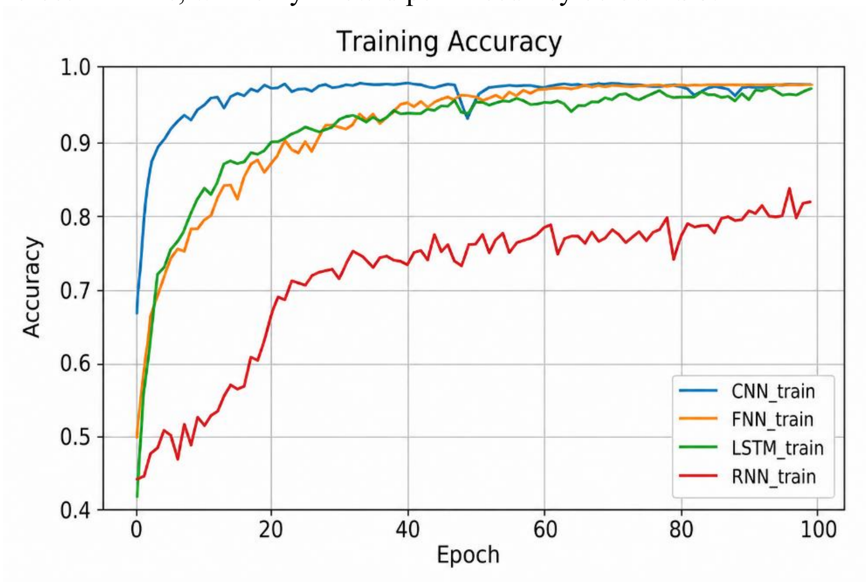


Figure 6. Training accuracy graphic comparison between the LSTM, CNN, FNN, and RNN

On the other hand, the LSTM model has an average accuracy of 0.924, with some trials showing accuracy scores above 0.95, while others are below 0.9. The FNN and RNN models have lower accuracy scores, with an average accuracy of 0.74 and 0.46, respectively. The results suggest that the CNN model is the best performing network for noise classification tasks. The LSTM model performs well but not as well as the CNN model. The FNN and RNN models have lower accuracy scores and thus are not the best options for this task.

The accuracy testing data shows that in figure 7 LSTM and CNN have generally higher accuracy compared to RNN and FNN. The highest accuracy achieved in this data set was by LSTM with 0.943 and the lowest accuracy was by CNN with 0.299. It is important to note that this data only represents the validation accuracy, and that the models may perform differently on a different set of data. It is also important to consider the variability of accuracy between the different models. LSTM and RNN models tend to have more consistent accuracy over the iterations, with a standard deviation of 0.028 and 0.042, respectively. In comparison, the standard deviation for RNN and FNN were 0.077 and 0.055. This suggests that LSTM and CNN models have a more stable performance, while CNN and FNN may have more unpredictable accuracy.

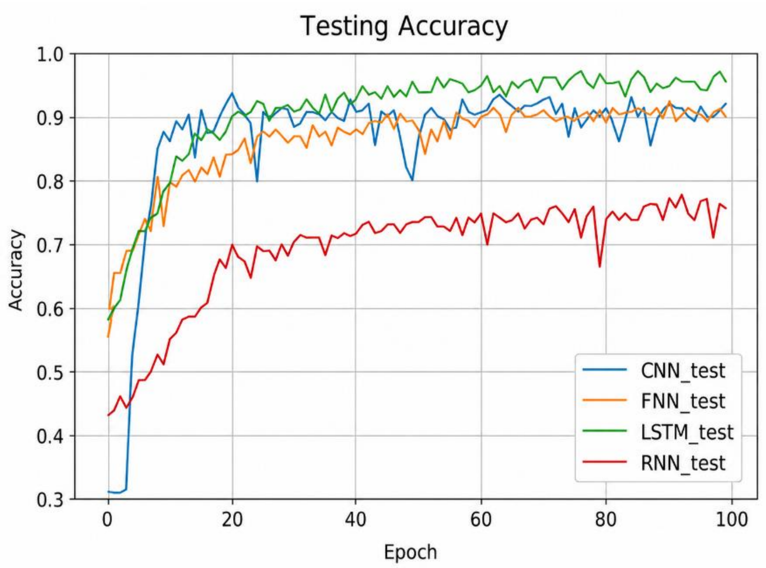


Figure 7. Validation accuracy graphic comparison between the LSTM, CNN, FNN, and RNN

Overall, the CNN architecture seems to perform the best in terms of minimizing the loss values. The average loss for CNN is lower than the other three architectures throughout the training process. The LSTM architecture also shows a good performance, with loss values lower than the FNN and RNN, although slightly higher than the CNN. The loss value for LSTM decreases smoothly throughout the training process. FNN architecture has the highest average loss throughout the training process, indicating that it is less effective in accurately classifying the noises compared to the other three architectures. The loss value for the FNN decreased gradually but not as fast as the other three. The RNN architecture has performance between the CNN and LSTM, with loss values lower than the FNN but higher than the CNN. The loss value for the RNN decreased throughout the training process, but with some fluctuations. The results suggest that the CNN and LSTM architectures are better suited for noise classification compared to the FNN and RNN.

5. Conclusion

The methods and results presented in this paper demonstrate that varying the MFCC data (13, 26, and 39 coefficients) on the LSTM model does not significantly impact accuracy. 97.4% for 13 MFCC data, 96.12%. for 26 MFCC data, and 96.3% for 39 MFCC data. For a more efficient model, using a smaller dataset is preferable. During the training process, the CNN model achieved the highest accuracy, averaging 97%. However, in the testing phase, the LSTM model showed superior performance, achieving 94% accuracy with 39 MFCC coefficients. Both LSTM and CNN models generally outperformed RNN and FNN in terms of validation accuracy. The 13-feature configuration achieved the highest validation accuracy while using the smallest input dimension, indicating that increasing the number of features did not improve classification performance in this experiment. It is important to note that these results reflect only one aspect of model performance. Further analysis is necessary to determine which model is best suited for specific tasks. Future work may explore adaptive feature selection or feature learning approaches to further optimize the trade-off between computational efficiency and classification performance.

Acknowledgment

This research was supported by the Institute of Research and Community Service (LPPM) at the Kalimantan Institute of Technology (ITK). We would like to express our sincere gratitude for their valuable funding, guidance, and support, which made this work possible.

References

MAFAULDA:MachineryFaultDatabase [Online]. (2021, February 3). Http://Www02.Smt.Ufrj.Br/~{}offshore/Mfs/Page_01.Html.

Abdul, Z. K., & Al-Talabani, A. K. (2022). Mel Frequency Cepstral Coefficient and its Applications: A Review. In *IEEE Access* (Vol. 10, pp. 122136–122158). Institute of Electrical and Electronics Engineers Inc. https://doi.org/10.1109/ACCESS.2022.3223444

Ahmed Murtaza, A., Saher, A., Hamza Zafar, M., Kumayl Raza Moosavi, S., Faisal Aftab, M., & Sanfilippo, F. (2024). Paradigm shift for predictive maintenance and condition monitoring from Industry 4.0 to Industry 5.0: A systematic review, challenges and case study. In *Results in Engineering* (Vol. 24). Elsevier B.V. https://doi.org/10.1016/j.rineng.2024.102935

Fang, Z., Guoliang, Z., & Zhanjiang -, S. (2001). Comparison of Different Implementations of MFCC. In *J. Comput. Sci. A Technol* (Vol. 16, Issue 6).

Germen, E., Başaran, M., & Fidan, M. (2014). Sound based induction motor fault diagnosis using Kohonen self-organizing map. *Mechanical Systems and Signal Processing*, *46*(1), 45–58. https://doi.org/10.1016/j.ymssp.2013.12.002

Glowacz, A. (2018). Acoustic-based fault diagnosis of commutator motor. *Electronics (Switzerland)*, *7*(11). https://doi.org/10.3390/electronics7110299

Glowacz, A., & Glowacz, Z. (2017). Diagnosis of stator faults of the single-phase induction motor using acoustic signals. *Applied Acoustics*, *117*, 20–27. https://doi.org/10.1016/j.apacoust.2016.10.012

Glowacz, A., Tadeusiewicz, R., Legutko, S., Caesarendra, W., Irfan, M., Liu, H., Brumercik, F., Gutten, M., Sulowicz, M., Antonino Daviu, J. A., Sarkodie-Gyan, T., Fracz, P., Kumar, A., & Xiang, J. (2021). Fault diagnosis of angle grinders and electric impact drills using acoustic signals. *Applied Acoustics*, *179*. https://doi.org/10.1016/j.apacoust.2021.108070

Khan, A., Hwang, H., & Kim, H. S. (2021). Synthetic data augmentation and deep learning for the fault diagnosis of rotating machines. *Mathematics*, *9*(18). https://doi.org/10.3390/math9182336

Liu, H., Zhou, J., Zheng, Y., Jiang, W., & Zhang, Y. (2018). Fault diagnosis of rolling bearings with recurrent neural network-based autoencoders. *ISA Transactions*, *77*, 167–178. https://doi.org/10.1016/j.isatra.2018.04.005

Mishra, S. P., Warule, P., & Deb, S. (2023). Speech emotion recognition using MFCC-based entropy feature. *Signal, Image and Video Processing*. https://doi.org/10.1007/s11760-023-02716-7

Nath, A. G., Sharma, A., Udmale, S. S., & Singh, S. K. (2021). An Early Classification Approach for Improving Structural Rotor Fault Diagnosis. *IEEE Transactions on Instrumentation and Measurement*, *70*. https://doi.org/10.1109/TIM.2020.3043959

SpectraQuest Inc.,: Machinery Fault Simulator. (2021, February 5). Https://Spectraquest.Com/Machinery-Fault-Simulator/Details/Mfs/.

Wang, Y., & Cao, G. (2023). A multiscale convolution neural network for bearing fault diagnosis based on frequency division denoising under complex noise conditions. *Complex and Intelligent Systems*, *9*(4), 4263–4285. https://doi.org/10.1007/s40747-022-00925-0

Yang, G., Wei, Y., & Li, H. (2022). Acoustic Diagnosis of Rolling Bearings Fault of CR400 EMU Traction Motor Based on XWT and GoogleNet. *Shock and Vibration*, *2022*. https://doi.org/10.1155/2022/2360067

Hochreiter, S., & Schmidhuber, J. (1997). Long short-term memory. *Neural Computation, 9*(8), 1735–1780. https://doi.org/10.1162/neco.1997.9.8.1735