Audio Enhancement for Gamelan Instrument Recognition using Spectral Subtraction

Viga Laksa Hardjanto
Department of Computer Science and Electronics, Faculty of Mathematics and Natural Sciences, Universitas Gadjah Mada, Yogyakarta, Indonesia
vigalaksa00@mail.ugm.ac.id

Wahyono
Department of Computer Science and Electronics, Faculty of Mathematics and Natural Sciences, Universitas Gadjah Mada, Yogyakarta, Indonesia
wahyo@ugm.ac.id (corresponding author)

Received: 10 January 2025 | Revised: 13 February 2025 and 21 February 2025 | Accepted: 27 February 2025

Licensed under a CC-BY 4.0 license | Copyright (c) by the authors | DOI: https://doi.org/10.48084/etasr.10181

ABSTRACT

Artificial intelligence has made significant progress in processing audio, text, and images, but noise remains a major challenge, especially in real-world audio data. This research presents a novel approach to improve audio classification by integrating noise reduction techniques with machine learning models. Focusing on the bonang barung, a traditional Javanese gamelan instrument, the study uses Mel Frequency Cepstral Coefficients (MFCC) and Mel spectrograms to identify the most effective features for classification, and the Multi-Layer Perceptron (MLP) model for the classification task. In addition, the spectral subtraction method is used to reduce noise, which resulted in significant improvements in audio quality, although some artifacts remain. The main contribution of this study is the integration of noise reduction with the MLP model to improve the classification performance. The MLP model successfully classified various bonang barung playing techniques, achieving a classification accuracy of 90% after noise reduction compared to 87.22% with noise, highlighting the importance of preprocessing steps, such as noise reduction. It is also demonstrated that MLP models can be a viable alternative to more complex deep learning models, such as CNN and RNN, for audio classification tasks. Overall, this research provides new insights into the role of noise reduction in audio analysis and offers potential advances in the field of audio classification.

Keywords-pattern recognition; audio enhancement; MFCC; multi-layer perceptron

I. INTRODUCTION

The term "karawitan", which comes from the word "rawit" meaning complicated, is inseparable from traditional Javanese music performance, especially gamelan. This musical ensemble consists of 23 instruments, each with its own playing techniques. For example, the bonang barung instrument has five different playing techniques: gĕmbyang, mipil lamba, mipil rangkĕp, mbalung, and nduduk gĕmbyang [1]. It is difficult for people to distinguish these playing techniques in one hearing, so this study aims to classify the playing techniques of bonang barung instruments based on audio recordings. The background noise significantly reduces the quality and clarity of the information of an audio file, so the file must go through various stages before it can be classified [2]. However, the noise reduction method must also be adapted to the type of noise, whether stationary or non-stationary. Stationary noise tends to be stable and it is expected to produce better audio quality if the spectral subtraction method is used in its processing [3]. On the other hand, for non-stationary noise, spectral subtraction methods are generally less suitable and produce a lot of residual noise [4]. However, according to the authors in [5], this can be overcome by implementing a combination method between spectral subtraction and wavelet transform, although it requires a relatively complex computational process. In this research, the types of noise used include gray noise, instrument noise, and speech noise, as these different noise conditions can significantly affect the quality of the audio signal and the performance of the model. Briefly, gray noise is audio noise with a uniform frequency spectrum, instrument noise comes from other musical instruments played together, and speech noise is noise from human voices. The noise will have a significant effect on feature extraction from audio data, such as in speech recognition [6].

The classification process involves the use of neural networks, which are able to recognize patterns in the classification process during training. In this research, the classifier presented in [7] is employed and is trained using the Mel-Frequency Cepstral Coefficients (MFCC) and Mel spectrogram feature extraction processes. Thus, two-dimensional features are obtained for each audio file. Both types of features are considered to be completely accurate in representing audio waves [8, 9]. This study presents the classification of bonang barung playing techniques based on audio data recorded by researchers. Audio preprocessing is used to reduce the noise before processing the audio data in the classifier [10]. The spectral subtraction method was chosen because it is considered relevant to the type of stationary noise data in the recording of the bonang barung instrument.

II. METHODOLOGY

This research was conducted through several stages, such as literature review, data collection, preprocessing, feature extraction, model building, testing, and evaluation. The flow chart of the research process is shown in Figure 1. In order to have knowledge of the latest methods and developments in the field of audio classification, a literature review of previous studies was conducted. Then, audio data were collected by recording on the corresponding object, bonang barung, and preprocessing was performed using spectral subtraction. The preprocessing is a key focus of this research because, based on [7], future research opportunities focus on considering noise reduction effects to improve the performance of classification models. Next, relevant features in each audio file are extracted to provide the main characteristics of the classification process. In the next stage, the classification model is developed using a Multi-Layer Perceptron (MLP) and adjusted in the parameter tuning stage to optimize the performance and increase the classification accuracy. In the testing phase, the model is tested based on its effectiveness. Finally, an evaluation is performed to assess the performance of the developed audio classification model.

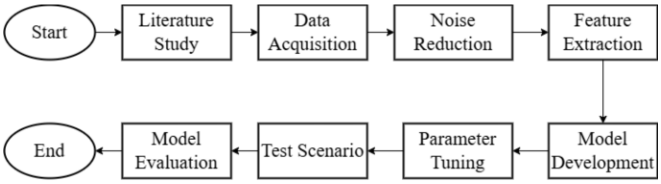

Fig. 1. Flow chart of the research process.

A. Algorithm Design

The data were collected by recording the bonang barung of the gamelan instrument and saving it in .wav file format. Each playing technique (gĕmbyang, mipil lamba, mipil rangkĕp, mbalung, and nduduk gĕmbyang) is represented by 60 audio files of 30 s each. To analyze the audio, features were extracted using MFCC [11] and Mel spectrogram, resulting in a two-dimensional array stored in JSON format for input to the MLP model. Since noise in the raw audio files can affect classification accuracy, a noise reduction process was applied to create two different datasets: one with noise preserved and one with noise reduction. Both datasets then underwent audio segment trimming before Mel spectrogram and MFCC features were extracted. Finally, these extracted features were fed into the MLP model and the results were evaluated to compare the classification accuracy between the noise-maintained and noise-reduced datasets. The entire process is illustrated in Figure 2.

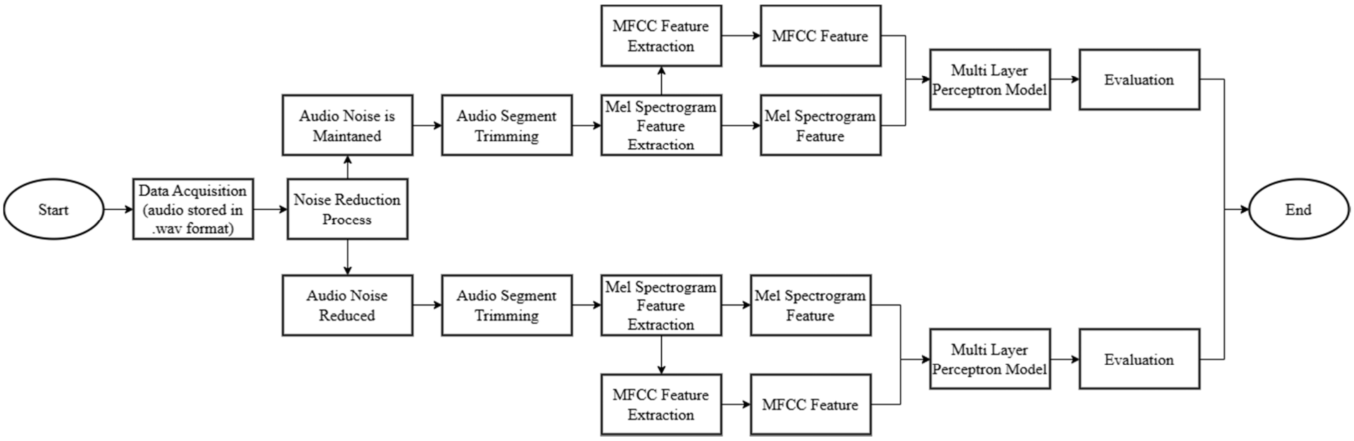

Fig. 2. Proposed process phases.

B. Data Acquisition

The data acquisition process began with the recording of the slendro-tuned gamelan bonang barung instruments. The recordings were made independently by the researcher, who is a gamelan player in Yogyakarta. The recording process utilized Adobe Audition software, and the .wav format was chosen for its superior quality and lower compression artifacts compared to .mp3. A dynamic microphone and an audio interface with a sample rate of 48,000 Hz were used. Each playing technique was recorded at a consistent tempo of 70 bpm to maintain the uniformity of the instrument's playing techniques. To increase the diversity of the dataset for classification purposes, the recorded audio files were segmented into multiple smaller clips, allowing for a greater variety of patterns in the recognition system. Since the researcher is both the performer and the creator of the recordings, this dataset qualifies as a primary dataset. No external sources or commercial music have been incorporated, ensuring that there are no copyright issues. The researcher retains full ownership and consent over the recorded music, allowing its use in this study. The dataset has been publicly archived on Zenodo to support further research and reproducibility [12].

C. Noise Reduction Strategy

This process aims to improve audio quality by reducing noise in the audio files using a noise profile reference and spectral subtraction. Assuming that $y(t)$ is the noisy sound, $s(t)$ is the pure sound signal, and $n(t)$ is the audio noise signal, (1) gives the relationship between them:

$$y(t) = s(t) + n(t) \quad (1)$$

The noise removal process starts with analyzing the audio signal in the frequency domain to identify the frequency components of the noise signals. The next step is to determine the noise spectrum, which is the portion of the audio signal that contains only contains only noise, by recording under stationary conditions or with a stationary device. This noise spectrum is then subtracted from the original audio spectrum to reduce the noise. The result is an audio signal with less noise. The final step is to reconstruct the modified audio signal back into the time domain. By subtracting the noise spectrum from the original signal spectrum, the energy corresponding to the noise frequency is reduced. The estimated signal value after spectral subtraction is calculated using the following formula:

$$\left|\hat{S}(w)\right| = \ [|Y(w)|^2 - |N(w)|^2\,]^{\frac{1}{2}} \quad (2)$$

where $\hat{S}(w)$ represents the signal after spectral subtraction, $Y(w)$ is the noisy signal in the frequency domain, and $N(w)$ represents the audio noise signal in the frequency domain.

D. Feature Extraction

The feature extraction process used in this study is MFCC and Mel spectrogram. Referring to the standard dataset for audio classification problems, GTZAN has 30 s on each audio track [13-15]. To expand the dataset, each track is divided into segments of 3 s, 5 s, and 10 s to increase the amount of data available for analysis. The total number is calculated based on the number of tracks, segments, and classes; for example, if there are 60 tracks, 10 segments, and five classes, the total data is 3000.

E. Multi-Layer Perceptron

In this research, MLP is used for the classification task. The architecture of the MLP model is inspired by the single-layer perceptron model, a neural network that has only input and output layers. Mathematically, the output of a single-layer perceptron can be calculated as follows:

$$y = f(w_1x_1 + w_2x_2 + \cdots + w_nx_n + b) \quad (3)$$

where y is the output value of the single-layer perceptron, x_n is the input feature, w_n is the weight corresponding to the input feature, b represents the bias, and f is the activation function. The MLP model was chosen for this study because of its simplicity, effectiveness, and suitability for the two-dimensional input generated by MFCC and Mel spectrograms. The input size of 216×13 for a 5 s audio segment is manageable for MLP, making it a computationally efficient choice compared to more complex models like Convolutional Neural Networks (CNNs) and Recurrent Neural Networks (RNNs). While CNNs excel at handling spatial data and RNNs are designed for sequential data, MLPs provide a balanced trade-off, offering good performance in classification tasks without requiring extensive computational resources. The MLP model used in this study consists of an input layer, three hidden layers, and an output layer, with ReLU and softmax activation functions, and a dropout layer to prevent overfitting. Optimized with the Adam algorithm and sparse categorical cross-entropy loss, the MLP model demonstrates competitive performance in audio classification tasks, making it an ideal choice for this research, where both accuracy and computational efficiency are important. The architecture includes an input layer, three hidden layers, and an output layer, as shown in Figure 3. The output layer consists of five classes: class 1 to class 5, representing the five categories of bonang barung instrument playing, which are gĕmbyang, mbalung, mipil lamba, mipil rangkĕp, and nduduk gĕmbyang.

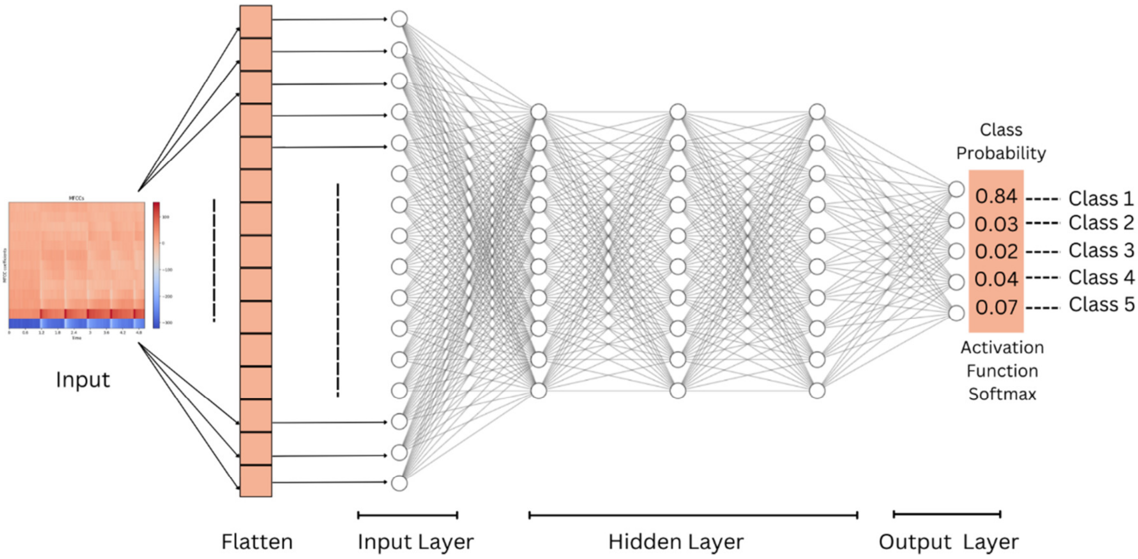

Fig. 3. MLP architecture.

F. Evaluation

The model is tested on noisy audio and audio enhanced with spectral subtraction techniques, using MFCC features and Mel spectrograms to identify the most appropriate features for classifying gamelan bonang barung instruments. The performance of the model is evaluated using metrics such as confusion matrix, accuracy, precision, recall, and F1-score, which assess how well the model classifies the data [16] and highlight its strengths and weaknesses. The confusion matrix provides detailed insight into the distribution of data between correct and incorrect classes [17].

III. RESULTS AND DISCUSSION

A. Noise Reduction Results with Spectral Subtraction

The effectiveness of the spectral subtraction method in reducing noise in audio data was analyzed through spectrogram visualization of the first 10 s sample of the gĕmbyang bonang barung playing technique. In Figure 4, the intensity difference of the purple color between the first (before) and second (after) spectrograms shows significant noise reduction, especially for static or stationary noise. However, the appearance of artifacts in the form of small boxes at the bottom of the spectrogram indicates that this method still has limitations in properly removing noise.

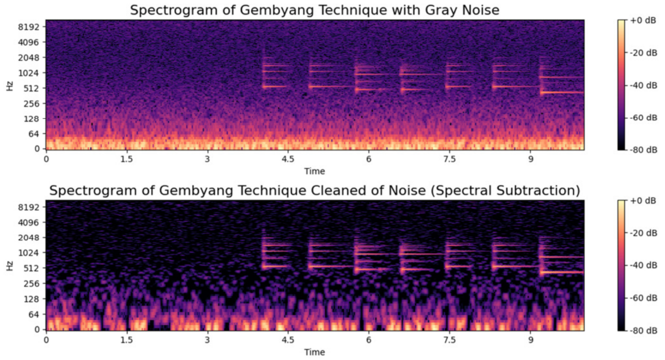

Fig. 4. Comparison of spectral subtraction in audio with gray noise types.

B. Baseline Model

The baseline MLP model was tested on a 3 s audio segment with a learning rate of 0.0001, 30% dropout, and 50 epochs on a data dimension of 130 frames × 13 MFCC coefficients (2,100 samples), resulting in a test data accuracy of 75.9%. The model achieved 78.4% precision, 76% recall, and 75.8% F1-score, leaving room for performance improvement through further parameter tuning. Testing the initial model with a 3 s audio segment that had been cleaned of gray noise showed a significant performance improvement, with a test data accuracy of 81.8%, as well as an increase in precision (83.7%) and recall (81.8%), providing early evidence that noise cleaning has a positive impact on classification performance.

C. Feature Extraction Method Selection

The choice of feature extraction methods, such as MFCC and Mel spectrogram, affects the performance of audio classification models, especially in the context of noise effects such as gray, instrument, and speech. Increasing the number of epochs to 100 to improve model training showed positive results, especially for MFCC features, improving accuracy from 80.89% to 84.33% after noise removal. On the other hand, the Mel spectrogram did not show a consistent improvement, with accuracy decreasing from 82.77% to 68.22% after noise removal. In addition, cleaning resulted in low accuracy between 19.77% and 20.44% for instrument noise types, indicating that MFCC is more effective in handling noise than Mel spectrogram. Based on the comparison of the results, MFCC performs better for audio classification problems, and the spectral subtraction method works better for static noise types such as gray noise.

D. Learning Rate Selection

Further tests were conducted with learning rate values of 0.01, 0.001, 0.005, and 0.0001 to evaluate their impact on model training using a 3 s audio segment containing gray noise and MFCC features. The best results were obtained with a learning rate of 0.0001, resulting in a test accuracy of 80.89%. Higher learning rates, such as 0.01, 0.001, and 0.005, yielded lower results. After noise removal, a learning rate of 0.0001 again showed the best performance with an accuracy of 84.33%. A similar test using spectrograms also showed that a learning rate of 0.0001 produced optimal performance with an accuracy of 82.78%. However, after noise removal, the performance drops to 68.22% accuracy, although it is still better than other values.

E. Dropout Selection

A dropout mechanism with probabilities of 0.3 and 0.5 was applied to prevent overfitting, and tests were conducted using 3 s audio segments, MFCC and Mel spectrogram features, with a learning rate of 0.0001 and 100 epochs, all with gray noise. The results showed that with MFCC features, a dropout probability of 0.3 achieved an accuracy of 80.89%, whereas a dropout probability of 0.5 achieved an accuracy of only 28.78%. After applying spectral subtraction for noise reduction, the 0.3 dropout accuracy increased to 84.33%, whereas the 0.5 dropout accuracy remained at 56.22%. Using the Mel spectrogram features without noise reduction, the 0.3 dropout accuracy was 82.78%. However, the 0.5 dropouts caused a drop in performance similar to the MFCC results, with a test accuracy of 20.44% after noise reduction. The 0.3 dropouts consistently outperformed the 0.5 dropouts, both with and without noise reduction. Table I details the differences between the two dropout probabilities and the effect of noise reduction.

TABLE I. DROPOUT TESTING

Dropout	Features	Noise reduction	Test accuracy (%)
0.3	MFCC	None	80.89
0.5	MFCC	None	28.78
0.3	MFCC	Yes	84.33
0.5	MFCC	Yes	56.22
0.3	Mel spectrogram	None	82.78
0.5	Mel spectrogram	None	20.44
0.3	Mel spectrogram	Yes	68.22
0.5	Mel spectrogram	Yes	39.89

F. Number of Epochs

Epoch testing on the MLP model, using 10, 50, 80, and 100 epochs, aims to find the optimal number of iterations for efficient learning. The results using MFCC features show that increasing the number of epochs significantly improves the classification performance, with 100 epochs achieving 80.89% accuracy for noisy data and 84.33% for cleaned data. 80 epochs proved to be sufficient, as no significant improvements were seen beyond that. Testing with the Mel spectrogram features initially showed low accuracy, with cleaned data performing slightly better. However, as epochs increased, accuracy improved more for noisy data and precision improved for cleaned data, indicating that Mel spectrograms are less effective for noisy data.

G. Effect of the Sample Length

The length of audio segments affects the size of the input matrix and model performance. Testing 3 s, 5 s, and 10 s segments with MFCC and Mel spectrogram features showed that segment length has a significant impact on performance. The 5 s segment with MFCC achieved the highest accuracy, reaching 87.22% for noisy audio and 90% for cleaned audio. The 3 s segment performed well, whereas the 10 s segment showed a decrease in accuracy, especially for noisy data. Mel spectrogram features performed best with the 3 s segment. Overall, the 3 s and 5 s segments were the most effective, indicating an optimal segment length for accuracy. The results of test sample length on model performance are shown in Table II.

TABLE II. SAMPLE LENGTH DIFFERENCE TEST

Segment (s)	Test accuracy (%)	Noise type	Features
3	79.22	Gray noise	MFCC
3	82.22	Gray cleaned	MFCC
5	87.22	Gray noise	MFCC
5	90.00	Gray cleaned	MFCC
10	65.19	Gray noise	MFCC
10	87.41	Gray cleaned	MFCC
3	79.44	Gray noise	Mel spectrogram
3	66.56	Gray cleaned	Mel spectrogram
5	20.00	Gray noise	Mel spectrogram
5	42.78	Gray cleaned	Mel spectrogram
10	20.00	Gray noise	Mel spectrogram
10	20.00	Gray cleaned	Mel spectrogram

H. Effect of Spectral Subtraction on Noise Type

Further experiments using MFCC and Mel spectrogram features with 3 s and 5 s audio segments showed different results depending on the type of noise. Using MFCC features, for gray noise, the 3 s segment increased the accuracy from 79.22% to 82.22% after the noise removal, whereas the 5 s segment increased the accuracy from 87.22% to 90% after noise removal. For the instrument noise type, the 3 s segment decreased the accuracy from 67.22% to 61.89% after noise removal, whereas the 5 s segment slightly increased the accuracy from 66.11% before noise removal to 67.33% after noise removal. This suggests that the 5 s segment provides more consistent results. The test results for the speech noise types are shown in Table III.

TABLE III. EFFECT OF SPECTRAL SUBTRACTION ON SPEECH NOISE

Noise type	Test accuracy (%)	Segment (s)	Features
Speech noise	58,33	3	MFCC
Speech cleaned	45,67	3	MFCC
Speech noise	65,37	5	MFCC
Speech cleaned	60,93	5	MFCC

In addition, when the model was tested using the Mel spectrogram features on a 3 s segment with gray noise, the model achieved 79.44% accuracy. Furthermore, when the noise was removed, the accuracy decreased to 66.56%. On the other hand, in the 5 s segment, the performance of the model decreased drastically, with an accuracy of only 20% before noise removal, increasing to 42.78% after noise removal. The results of comparing the effect of gray noise removal with Mel spectrogram as the feature extraction method are shown in Table IV. In the instrument noise test, a 3 s segment resulted in a test accuracy of 19.89%. After noise reduction, the accuracy increased to 20.56%. For the 5 s segment, the accuracy was 20.19% before and 20% after noise reduction, showing better performance than the 3 s segment. For speech noise, the 3 s segment had an accuracy of 20.33%, which increased to 20.78% after noise reduction. For the 5 s segment, the accuracy was 19.81% before noise reduction and increased slightly to 21.48% after noise reduction.

TABLE IV. EFFECT OF SPECTRAL SUBTRACTION ON GRAY NOISE

Noise Type	Test accuracy (%)	Segment (s)	Features
Gray noise	79.44	3	Mel spectrogram
Gray cleaned	66.56	3	Mel spectrogram
Gray noise	20	5	Mel spectrogram
Gray cleaned	42.78	5	Mel spectrogram

I. Comparison with Bandpass Filter and Wiener Filter Methods

The proposed method was compared with the bandpass filter and Wiener filter noise reduction methods. The bandpass filter, which removes unwanted frequencies by combining low-pass and high-pass filters, was found to be less effective with low accuracy. The Wiener filter performed better, although accuracy remained relatively low, especially for instrument and speech noise. When combined with the MLP model and the MFCC feature extraction process, the spectral subtraction method was the most effective for audio classification, achieving up to 90% accuracy, especially for signals cleaned of gray noise. Improvements were also seen for instrument and speech noise, but not as much as for gray noise. Overall, spectral subtraction with MLP gave the best results, whereas bandpass and Wiener filters showed lower performance for most noise types. Detailed accuracy comparisons are shown in Figure 5.

J. Comparison with CNN and RNN Models

The proposed MLP method was compared with CNN and RNN models, using the optimal parameters of MLP: 5 s audio segments, MFCC features, and 80 epochs. MLP achieved

87.22% accuracy on gray noise data, which improved to 90% after noise removal. Figure 6 shows MLP's evaluation using the confusion matrix, whereas Figure 7 shows the training accuracy and error improvement, indicating no overfitting. The classification results in these evaluations correspond to the five output classes previously explained in the MLP architecture section, which represent different styles of bonang barung instrumental performance.

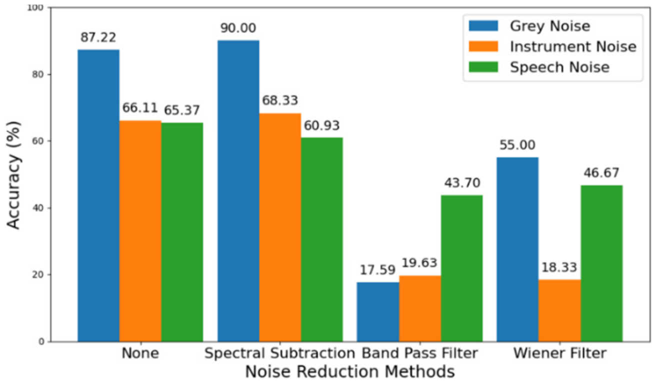

Fig. 5. Comparison of the accuracy of noise removal methods for gray, instrument, and speech noise types.

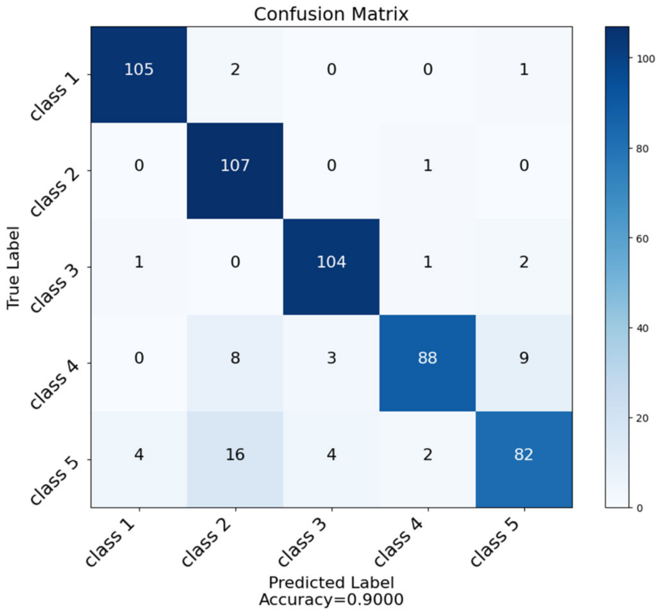

Fig. 6. Confusion matrix of the MLP model without gray noise.

Overall, CNN performed better than MLP despite the slower execution time. The RNN model also showed consistent accuracy. In the context of audio classification, deep learning models (CNN and RNN) excel at handling spatial and sequential data, with RNN suitable for applications requiring high accuracy, CNN suitable for applications requiring a balance accuracy and time, and MLP suitable for applications requiring more efficient use of computational resources. The comparison results are shown in Table V.

TABLE V. ACCURACY COMPARISON OF MLP, CNN, AND RNN MODELS

Noise Type	Accuracy (%)		
	MLP	**CNN**	**RNN**
Gray noise	87.22	96.11	87.96
Gray cleaned	90.00	95.93	91.85
Instrument noise	66.11	81.67	84.63
Instrument cleaned	68.33	86.30	81.67
Speech noise	65.37	73.52	77.22
Speech cleaned	60.93	76.11	70.74

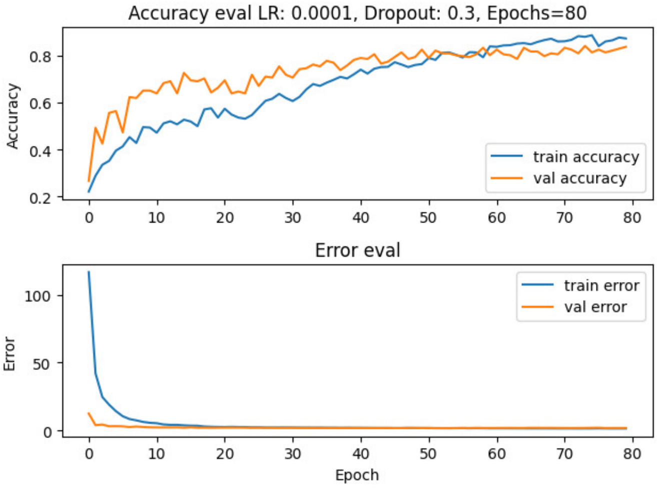

Fig. 7. MLP model training graph with data cleaned of gray noise.

IV. CONCLUSION

This study demonstrates that the Multi-Layer Perceptron (MLP) model is effective in classifying bonang barung music techniques such as gĕmbyang, mipil lamba, mipil rangkĕp, mbalung, and nduduk gĕmbyang. The study also highlights the important role of noise reduction, with the spectral subtraction method improving audio quality, although it may introduce some artifacts. After noise removal, Mel-Frequency Cepstral Coefficients (MFCC) features performed better than Mel spectrograms features, and 5 s audio segments produced the best results. With the optimal MLP settings, the accuracy of the model increased from 87.22% with noise to 90% after noise removal. Compared to similar studies [7], which achieved 87% accuracy using an Artificial Neural Network (ANN) without noise reduction, this research underscores the significant impact of noise reduction in improving classification accuracy. While deep learning models such as CNN and RNN are better at handling complex data, the MLP model offers a strong alternative with lower computational requirements. However, this study has several limitations. First, the spectral subtraction method is more effective for stationary noise, so the use of other noise reduction techniques could help to address non-stationary noise. Second, real-time audio recognition was not investigated, which could be an important area for future work. Finally, other feature extraction methods besides MFCC and Mel spectrograms could be considered to further improve the performance of the system. For future research, this approach could be applied to other musical instruments or genres. Improving noise reduction methods and combining them with other techniques could further increase the accuracy and robustness of audio classification systems.

ACKNOWLEDGMENT

The publication of this work was partially supported by the Department of Computer Science and Electronics, Universitas Gadjah Mada under the Publication Funding Year 2025.

REFERENCES

[1] Soeroso, *Bagaimana bermain gamelan*. Jakarta, Indonesia: Balai Pustaka, 1982.

[2] J. Umamaheswari and A. Akila, "Improving Speech Recognition Performance using Spectral Subtraction with Artificial Neural Network," *International Journal of Advanced Studies of Scientific Research*, vol. 3, no. 11, pp. 214–219, 2018.

[3] J. S. Ashwin and N. 92, Jan Manoharan, "Audio Denoising Based on Short Time Fourier Transform," *Indonesian Journal of Electrical Engineering and Computer Science*, vol. 9, no. 1, pp. 89–. 2018, https://doi.org/10.11591/ijeecs.v9.i1.pp89-92.

[4] S. K. Shridhar, L. Doddimani, A. Hirekoppa, K. Kodliwad, and A. Viraktamath, "Speech Enhancement using Spectral Subtraction," *International Journal of Engineering Research*, vol. 10, no. 7, pp. 744–748, Jul. 2021.

[5] Y. Yang, P. Liu, H. Zhou, and Y. Tian, "A Speech Enhancement Algorithm combining Spectral Subtraction and Wavelet Transform," in *2021 IEEE 4th International Conference on Automation, Electronics and Electrical Engineering*, Shenyang, China, 2021, pp. 268–273, https://doi.org/10.1109/AUTEEE52864.2021.9668622.

[6] A. A. Alasadi, T. H. Aldhayni, R. R. Deshmukh, A. H. Alahmadi, and A. S. Alshebami, "Efficient Feature Extraction Algorithms to Develop an Arabic Speech Recognition System," *Engineering, Technology & Applied Science Research*, vol. 10, no. 2, pp. 5547–5553, Apr. 2020, https://doi.org/10.48084/etasr.3465.

[7] K. S. Harshavardhan and Mahesh, "Urban sound classification using ANN," in *2022 International Interdisciplinary Humanitarian Conference for Sustainability*, Bengaluru, India, 2022, pp. 1475–1480, https://doi.org/10.1109/IIHC55949.2022.10060146.

[8] M. Shah, N. Pujara, K. Mangaroliya, L. Gohil, T. Vyas, and S. Degadwala, "Music Genre Classification using Deep Learning," in *2022 6th International Conference on Computing Methodologies and Communication*, Erode, India, 2022, pp. 974–978, https://doi.org/10.1109/ICCMC53470.2022.9753953.

[9] R. Shah, P. Shah, C. Joshi, R. Jain, and R. Nikam, "Heartbeat Prediction using Mel Spectrogram and MFCC Value," in *2023 IEEE IAS Global Conference on Emerging Technologies*, London, United Kingdom, 2023, pp. 1–5, https://doi.org/10.1109/GlobConET56651.2023.10150129.

[10] X. Zhou, K. Hu, and Z. Guan, "Environmental sound classification of western black-crowned gibbon habitat based on spectral subtraction and VGG16," in *2022 IEEE 5th Advanced Information Management, Communicates, Electronic and Automation Control Conference*, Chongqing, China, 2022, pp. 578–582, https://doi.org/10.1109/IMCEC55388.2022.10019981.

[11] H. A. Owida, A. Al-Ghraibah, and M. Altayeb, "Classification of Chest X-Ray Images using Wavelet and MFCC Features and Support Vector Machine Classifier," *Engineering, Technology & Applied Science Research*, vol. 11, no. 4, pp. 7296–7301, Aug. 2021, https://doi.org/10.48084/etasr.4123.

[12] V. L. Hardjanto, "Bonang Barung Instrument." Zenodo, Feb. 17, 2025, https://doi.org/10.5281/zenodo.14880567.

[13] M. S. Rao, O. Pavan Kalyan, N. N. Kumar, Md. Tasleem Tabassum, and B. Srihari, "Automatic Music Genre Classification Based on Linguistic Frequencies Using Machine Learning," in *2021 International Conference on Recent Advances in Mathematics and Informatics*, Tebessa, Algeria, 2021, pp. 1–5, https://doi.org/10.1109/ICRAMI52622.2021.9585937.

[14] Y.-H. Cheng, P.-C. Chang, and C.-N. Kuo, "Convolutional Neural Networks Approach for Music Genre Classification," in *2020 International Symposium on Computer, Consumer and Control*, Taichung City, Taiwan, 2020, pp. 399–403, https://doi.org/10.1109/IS3C50286.2020.00109.

[15] J. K. Bhatia, R. D. Singh, and S. Kumar, "Music Genre Classification," in *2021 5th International Conference on Information Systems and Computer Networks*, Mathura, India, 2021, pp. 1–4, https://doi.org/10.1109/ISCON52037.2021.9702303.

[16] M. Rahmandani, H. A. Nugroho, and N. A. Setiawan, "Cardiac Sound Classification Using Mel-Frequency Cepstral Coefficients (MFCC) and Artificial Neural Network (ANN)," in *2018 3rd International Conference on Information Technology, Information System and Electrical Engineering*, Yogyakarta, Indonesia, 2018, pp. 22–26, https://doi.org/10.1109/ICITISEE.2018.8721007.

[17] X. Mu and C.-H. Min, "MFCC as Features for Speaker Classification using Machine Learning," in *2023 IEEE World AI IoT Congress*, Seattle, WA, USA, 2023, pp. 0566–0570, https://doi.org/10.1109/AIIoT58121.2023.10174566.