Journal of Artificial Intelligence and Engineering Applications

Website: https://ioinformatic.org/

15th June 2025. Vol. 4. No. 3; e-ISSN: 2808-4519

Enhancing Mujawwad Qur'anic Recitation Rhythm Classification Using Optimized LSTM Algorithm

Japar Sidik[1]*, Ade Irma Purnamasari[2], Agus Bahtiar[3], Edi Tohidi[4]

[1,2,3,4]STMIK IKMI Cirebon, Cirebon, Indonesia
japarssidik820@gmail.com[1]*

Abstract

This research develops an automated classification system for Qur'anic recitation rhythms using the Long Short-Term Memory (LSTM) deep learning approach. The study aims to enhance rhythm identification accuracy by applying hyperparameter optimization techniques. Audio data was collected from mujawwad recitation recordings at Al-Falah 2 Nagreg Islamic Boarding School. Mel-frequency Cepstral Coefficients (MFCC) was extracted as acoustic features, and a systematic GridSearch approach was used to optimize the LSTM model. The proposed model achieved 88.07% classification accuracy, significantly outperforming the Naïve Bayes Classifier (38.97%). Confusion matrix analysis revealed superior performance in classifying complex rhythms, particularly bayati (95%), jiharkah (92%), and rast (90%) styles. This research demonstrates the potential of deep learning in understanding intricate Qur'anic recitation patterns and provides a foundation for developing advanced learning and assessment tools.

Keywords: *Deep Learning; LSTM; Rhythm Classification; Mujawwad; MFCC*

1. Introduction

The recitation of the Holy Qur'an, known as *tilawah*, represents a profound spiritual and artistic practice in Islamic tradition, as documented in various studies on the phonetic and prosodic dimensions of Qur'anic recitation. Beyond mere textual reading, *tilawah* encompasses a sophisticated art form that requires mastery of *Tajweed* rules, precise pronunciation, and intricate melodic styles. In the Indonesian context, the mujawwad recitation style stands out for its rich musical complexity, featuring diverse rhythmic patterns such as *bayati, shoba, hijaz, nihawand, rast, sikah,* and *jiharkah* [1], [2].

Traditionally, identifying and distinguishing these nuanced recitation rhythms has been the exclusive domain of expert reciters (*qari'*) who possess deep cultural and musical knowledge. This expertise typically requires years of dedicated training under master instructors, making comprehensive understanding of these rhythmic styles challenging for novice learners [3]. The intricacies of each rhythm involve subtle variations in melodic structure, tempo, and emotional expression that demand sophisticated auditory discernment.

Technological advancements in machine learning and artificial intelligence have revolutionized the approach to analyzing and classifying complex musical patterns, offering unprecedented opportunities to decode intricate cultural and artistic expressions through computational methods. These innovations provide researchers with powerful tools to transform subjective musical interpretation into objective, data-driven analysis. Specifically, deep learning techniques like Long Short-Term Memory (LSTM) neural networks demonstrate remarkable potential in capturing temporal dependencies within sequential audio data [4], [5], [6]. Utilizing sophisticated feature extraction techniques like Mel-frequency Cepstral Coefficients (MFCC), a widely recognized method in audio processing known for its effective representation of sound's spectral properties [7], [8], [9], these algorithms have the potential to accurately decode the complex characteristics of various Qur'anic recitation styles with remarkable precision [10], [11].

The objective of this study is to establish an automated classification system for mujawwad recitation rhythms, thereby connecting traditional musical knowledge with contemporary computational methods. Through the implementation of deep learning techniques, we intend to develop a resource that aids learners in comprehending various recitation styles while also offering an objective framework for the preservation of musical and cultural heritage.

2. Research Method

This research implements a systematic approach to develop an automatic classification system for Qur'anic recitation rhythms in mujawwad style. The methodology follows a structured process that begins with data collection and culminates in model evaluation, as illustrated in Figure 1.

Fig. 1: Research Method

The research methodology comprises four main stages: data collection, data preprocessing, model development, and model evaluation. Each stage is carefully designed to address the challenges associated with audio data processing and deep learning model optimization for Qur'anic recitation analysis. Data preprocessing, including audio standardization and signal processing, is crucial for preparing high-quality audio data [4], [12].

2.1. Data Collection

Data used in this research was obtained through direct recording of mujawwad recitations performed by students and teachers at Al-Falah 2 Nagreg Islamic Boarding School [13]. The recording process utilized the Sony IC Recorder ICD-PX470 device, which supports high-quality audio capture with specifications including LPCM 44.1 kHz/16-bit frequency range (50 Hz - 20,000 Hz) and WAV file format with PCM_16 encoding. Additional data was collected from voice note recordings shared via WhatsApp to ensure diversity in the dataset. The initial dataset consisted of 24 audio files representing seven different rhythmic patterns (*bayati, shoba, hijaz, nihawand, rast, sikah,* and *jiharkah*), with varying distribution across categories. This approach ensured that the data accurately reflected the characteristics of mujawwad recitation rhythms [1], [3].

2.2. Data Preprocessing

2.2.1. Audio Standardization

All audio files were standardized to ensure consistency in format and quality. Files from different sources (voice notes and direct recordings) were converted to WAV format with PCM_16 encoding, 44.1 kHz sample rate, and mono channel configuration. This standardization was implemented using Python libraries such as librosa [14] and soundfile.

2.2.2. Audio Signal Processing

The standardized audio files underwent several processing steps:

1. Resampling to 22050 Hz to reduce computational complexity while preserving essential frequency information.
2. Audio trimming to remove silent sections at the beginning and end of recordings.
3. Segmentation using a sliding window approach with 50% overlap (window size of 2048 samples and hop length of 1024 samples).

2.2.3. Feature Extraction

MFCC was extracted as the primary acoustic feature for rhythm classification. MFCC was chosen due to its effectiveness in capturing the timbral characteristics of audio signals and its proven performance in speech and music classification tasks. For each audio segment, 40 MFCC coefficients were extracted, creating a rich representation of the recitation's acoustic properties.

2.2.4. Data Transformation

The extracted features underwent further modifications to prepare them for model training. This process involved label encoding, which converted categorical rhythm labels into numerical representations, along with data reshaping to meet the input requirements of the LSTM model. Additionally, the dataset was partitioned into training and testing subsets, consisting of 80% and 20% of the data, respectively. Consequently, these pre-processing steps transformed the original 24 audio files into 7,543 labelled segments, with the distribution across rhythm categories outlined in Table 1.

Table 1: Data Distribution Before and After Pre-processing

Label	Before	After
Bayati	4	1090
Hijaz	3	931
Jiharkah	3	890
Nihawand	4	898
Rast	4	1582
Shoba	3	989
Sikah	4	1163
Total	**24**	**7543**

2.3. Model Development

2.3.1. LSTM Architecture Design

The foundational architecture of the LSTM was developed with an emphasis on the sequential characteristics inherent in audio data, as well as the intricate nature of rhythmic patterns. This model included an input layer tailored to align with the dimensions of the pre-processed MFCC features, followed by two LSTM layers that incorporated dropout regularization to mitigate the risk of overfitting. Additionally, the architecture featured a dense output layer that employed softmax activation, facilitating multi-class classification. This design effectively addresses the challenges posed by audio data, ensuring robust performance in recognizing and categorizing complex rhythmic structures.

2.3.2. Hyperparameter Optimization

To optimize the model's performance, a detailed hyperparameter search was undertaken using GridSearch to find the most suitable configuration [15]. The parameters analyzed included the number of LSTM units, which ranged from 128 to 1024, the dropout rate from 0.1 to 0.5, the learning rate between 0.0001 and 0.01, and batch sizes of 32 or 64. This hyperparameter tuning was carried out with Keras Tuner [16], emphasizing validation accuracy as the primary metric for optimization. The search process included 20 trials, with several runs for each trial to ensure the findings were robust.

2.3.4. Model Training

The training of the model was conducted utilizing the Adam optimizer alongside a categorical cross-entropy loss function. To mitigate the risk of overfitting, early stopping was employed, and model checkpoints were established to preserve the most effective configuration. Additionally, class weights were computed and incorporated during the training process to tackle the dataset's imbalance. This study was conducted using a local computational environment with the following hardware specifications in table 2.

Table 2: Hardware Specification

Component	Specification
Processor	Intel Core I7-6700HQ
RAM	12 GB
Integrated GPU	Intel HD Graphics 530
Dedicated GPU	Nvidia GTX 950M
VRAM	4 GB
OS	Windows 11 (WSL Ubuntu)

The hardware configuration provided the necessary computational resources for training and evaluating the deep learning model for Quranic recitation rhythm classification. This setup enabled efficient processing of audio features and model optimization, allowing researchers to perform complex machine learning tasks without relying on cloud-based computational platforms. The local computational environment was carefully selected to balance processing power, memory capacity, and graphics performance, ensuring robust implementation of the LSTM-based classification approach for Quranic recitation rhythms.

2.4. Model Evaluation

Model evaluation is a crucial step in developing machine learning systems, where the model's performance is assessed using test data. This process ensures that the model not only learns effectively from the training dataset but also makes accurate predictions on new, unseen data. In classifying Quranic recitation rhythms, evaluation employs quantitative metrics such as accuracy (1), precision (2), and recall (3), which are derived from the confusion matrix. True Positives (TP) represent correctly identified positive instances, while True Negatives (TN) indicate accurately identified negatives. False Positives (FP) are positive instances misclassified, and False Negatives (FN) are negative instances incorrectly identified. These metrics enable a comprehensive analysis of the classification model's performance, offering insights into its ability to distinguish between different rhythm types. By utilizing various metrics, researchers can gain a nuanced understanding of the model's strengths and weaknesses in diverse classification scenarios. All of the evaluation metrics are computed with equations 1 through 3 [3], [17].

$$\text{Accuracy} = \frac{TP + TN}{TP + FP + TN + FN} \tag{1}$$

$$\text{Precision} = \frac{TP}{TP + FP} \tag{2}$$

$$\text{Recall} = \frac{TP}{TP + FN} \tag{3}$$

3. Result and Discussion

3.1. Data Collection and Pre-processing

The study employed audio recordings of Quranic recitation (*tilawah*) in the mujawwad style, sourced from both students and instructors at Al-Falah 2 Nagreg Islamic Boarding School. The original dataset consisted of 24 audio files that encompassed seven distinct rhythm types: *bayati, shoba, hijaz, nihawand, rast, sikah,* and *jiharkah*. During the pre-processing phase, the audio files were standardized to the WAV format, adhering to specific parameters: WAV format, PCM_16 encoding, a sample rate of 44.1 kHz, and mono channel configuration. By implementing audio segmentation with a 50% overlap and extracting MFCC, the dataset was significantly increased from 24 to 7,543 samples. Following pre-processing, the distribution of rhythm labels revealed that *rast* (1,582 samples), *sikah* (1,163 samples), and *bayati* (1,090 samples) were the most prevalent categories. This effective data augmentation strategy successfully mitigated the initial limitation of a small audio sample size, thereby enhancing the dataset's suitability for deep learning applications.

3.2. Model Development and Hyperparameter Optimization

The LSTM model represents an advanced approach to classifying Qur'anic recitation rhythms, specifically tailored to capture the complex temporal nuances found in musical performances. With a carefully designed architecture that includes two LSTM layers and dropout regularization, this model overcomes the traditional limitations of neural networks in processing intricate audio signals. The GridSearch methodology enabled a comprehensive exploration of the parameter space, leading to the identification of a configuration that accurately recognizes subtle rhythmic variations. This optimization not only improved the model's technical performance but also highlighted the potential of deep learning in understanding complex cultural musical expressions. The model's enhanced ability to distinguish between different tilawah styles demonstrates the significant promise of computational techniques in analyzing intricate artistic and cultural phenomena. By transforming subjective musical interpretations into objective, data-driven insights, this research effectively bridges technological advancements with musical heritage, as illustrated by the optimal hyperparameter settings in table 3.

Table 3: Best Parameter Configuration

Parameter	Value
LSTM Layer	2
Units per layer	128
Dropout	0.2
Optimizer	AdamW
Learning Rate	0.001
Batch Size	32
Loss Function	Categorical Cross-Entropy

3.3. Model Evaluation

The performance comparison of the three models in Table 4 shows that Hyperparameter LSTM excels with 88% accuracy, followed by Base LSTM 81%, while NBC only reaches 39%. The low performance of NBC indicates its difficulty in learning complex temporal patterns from MFCC features.

Table 4: Model Performance Comparison

Model	Accuracy	Precision	Recall
NBC	39%	39%	39%
Base LSTM	81%	81%	81%
Hyperparameter LSTM (proposed model)	**88%**	**88%**	**88%**

The analysis derived from the confusion matrix presented in Figure 2 offers detailed insights into the performance of various classification models. The Naïve Bayes Classifier (NBC) exhibited the least effectiveness, facing considerable difficulties in differentiating among various rhythm types. Its classification accuracy peaked at 48% for the bayati rhythm, while it plummeted to 32% for the sikah rhythm, highlighting the model's struggle to discern the subtle distinctions inherent in rhythm patterns. In contrast, the Base LSTM model showed a significant enhancement in classification performance. It achieved the highest precision (84.71%) and recall (93.91%) for the bayati rhythm, indicating a strong capability to identify this specific rhythm. Additionally, the jiharkah and rast rhythms also displayed commendable classification results, with precision values ranging from 83% to 89% and recall rates between 81% and 83%. However, the model encountered more challenges with the nihawand and sikah rhythms, where precision and recall were approximately 75% to 80%.
The Hyperparameter LSTM model emerged as the most advanced classification method, demonstrating consistently high performance across all rhythm types. Notably, it achieved 89.34% precision and 94.78% recall for the bayati rhythm, while the jiharkah rhythm maintained impressive metrics with 91.88% precision and recall. The rast and nihawand rhythms also benefited from significant improvements, with precision and recall values ranging from 79% to 91%. The model's balanced performance across various rhythms highlights the effectiveness of the hyperparameter optimization strategy employed.

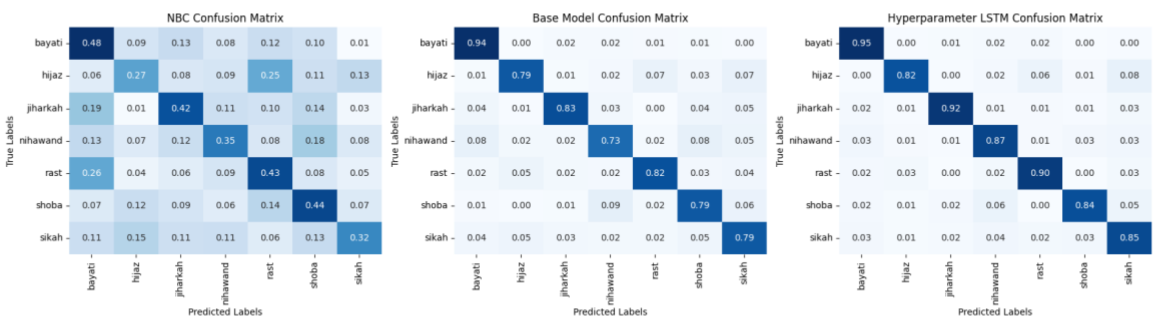

Fig. 2: Confusion Matrix for Each Model

Based on table 5, the progression from NBC to Hyperparameter LSTM demonstrates the powerful capabilities of deep learning techniques in capturing the intricate nuances of Quranic recitation rhythms, with the optimized LSTM model showing remarkable ability to distinguish between different rhythm types.

Table 5: Comparison of Precision and Recall scores

Labels	NBC		Base LSTM		Hyperparameter LSTM	
	Precision	Recall	Precision	Recall	Precision	Recall
Bayati	37%	48%	84.71%	93.91%	89.34%	94.78%
Hijaz	35%	27%	83.42%	79.43%	91.49%	82.30%

Jiharkah	34%	42%	83.65%	83.13%	91.88%	91.88%
Nihawand	33%	35%	74.56%	72.83%	79.89%	87.28%
Rast	47%	43%	89.44%	81.94%	91.08%	91.06%
Shoba	38%	44%	77.46%	79.33%	91.58%	83.65%
Sikah	48%	32%	75.22%	79.00%	80.26%	85.39%

4. Conclusion

This study introduces an innovative method for categorizing the rhythms of Quranic recitation through the use of an optimized LSTM neural network, attaining a remarkable classification accuracy of 88.07%. By methodically tackling the complexities associated with rhythm identification via sophisticated deep learning strategies, the research highlights the considerable promise in deciphering the complex melodic structures characteristic of mujawwad-style Quranic recitation. The suggested approach successfully addresses the shortcomings of conventional machine learning techniques, establishing a strong framework for rhythm classification that connects technological advancements with the analysis of cultural music.

While the current research offers promising results, several limitations and future research directions emerge. The dataset's confinement to a single Islamic boarding school suggests the need for expanded research involving more diverse reciters and recording environments. Future studies should explore more sophisticated deep learning architectures, develop practical applications for Quranic recitation learning, and refine rhythm classification techniques. Ultimately, this research contributes to the intersection of artificial intelligence and Islamic musical traditions, offering an innovative technological perspective for analyzing the complex rhythmic nuances of Quranic recitation and opening new avenues for cultural and technological exploration.

Acknowledgement

The authors express their sincere gratitude to the School of Information and Computer Management (STMIK) IKMI Cirebon for providing support and facilities throughout the research process. Special thanks are extended to the entire academic community of Al-Qur'an Al-Falah 2 Nagreg Islamic Boarding School, particularly the students and teachers who contributed to the *tilawah* recording data collection. Acknowledgment is also given to the *tilawah* experts who assisted in audio data validation and interpretation, and to the research colleagues who provided valuable insights and support during the research.

References

[1] S. Suarni and S. Syukrinur, “History of the Development of Nagham Al-Qur’an in Indonesia,” *J. Ilm. Al-Muashirah Media Kaji. Al-Quran Dan Al-Hadits Multi Perspekt.*, vol. 20, no. 2, pp. 288–298, Jul. 2023, doi: 10.22373/JIM.V20I2.18726.

[2] R. B. Qureshi and K. Nelson, “The Art of Reciting the Qu’ran,” *Ethnomusicology*, vol. 33, no. 3, p. 522, 1989, doi: 10.2307/851774.

[3] V. Y. Mafula, Abd. C. Fauzan, and T. R. Fernando, “Identifikasi Irama Tilawah al-Quran dengan Gaya Mujawwad Menggunakan Naive Bayes Classifier,” *Ilk. J. Comput. Sci. Appl. Inform.*, vol. 4, no. 2, pp. 242–251, Aug. 2022, doi: 10.28926/ilkomnika.v4i2.464.

[4] S. Shahriar and U. Tariq, “Classifying Maqams of Qur’anic Recitations Using Deep Learning,” *IEEE Access*, vol. 9, pp. 117271–117281, 2021, doi: 10.1109/ACCESS.2021.3098415.

[5] C. R. Wairata, E. R. Swedia, and M. Cahyanti, “PENGKLASIFIKASIAN GENRE MUSIK INDONESIA MENGGUNAKAN CONVOLUTIONAL NEURAL NETWORK,” *Sebatik*, vol. 25, no. 1, pp. 255–261, Jun. 2021, doi: 10.46984/sebatik.v25i1.1286.

[6] A. S. Ma`ruf, “IMPLEMENTASI DEEP LEARNING PADA KLASIFIKASI NATURE MUSIC DENGAN METODE EKSTRAKSI FITUR MEL-FREQUENCY CEPSTRAL COEFFICIENT (MFCC),” 2023, [Online]. Available: http://repositori.uin-alauddin.ac.id/id/eprint/27149

[7] B. Logan, “Mel Frequency Cepstral Coefficients for Music Modeling,” *Int. Symp. Music Inf. Retr.*, vol. 28, p. 11p., 2000, doi: 10.1.1.11.9216.

[8] D. Prabakaran and S. Sriuppili, “Speech Processing: MFCC Based Feature Extraction Techniques- An Investigation,” *J. Phys. Conf. Ser.*, vol. 1717, no. 1, p. 12009, Dec. 2021, doi: 10.1088/1742-6596/1717/1/012009.

[9] K. Zaman, M. Sah, C. Direkoglu, and M. Unoki, “A Survey of Audio Classification Using Deep Learning,” *IEEE Access*, vol. 11, pp. 106620–106649, 2023, doi: 10.1109/ACCESS.2023.3318015.

[10] F. Wolf-Monheim, “Spectral and Rhythm Features for Audio Classification with Deep Convolutional Neural Networks,” Oct. 2024, [Online]. Available: http://arxiv.org/abs/2410.06927

[11] M. Z. Adam, N. Shafie, H. Abas, and A. Azizan, “Analysis of Momentous Fragmentary Formants in Talaqi-like Neoteric Assessment of Quran Recitation using MFCC Miniature Features of Quranic Syllables,” *Int. J. Adv. Comput. Sci. Appl.*, vol. 12, no. 9, pp. 533–540, 2021, doi: 10.14569/IJACSA.2021.0120960.

[12] P. Anggeli, S. Suroso, and M. Z. Agung, “Klasifikasi Alat Musik Tradisional dengan Metode Machine Learning dengan Librosa dan Tensorflow pada Python,” *J-SAKTI J. Sains Komput. Dan Inform.*, vol. 5, no. 2, Art. no. 2, Sep. 2021, doi: 10.30645/j-sakti.v5i2.390.

[13] “Pondok Pesantren Al-Qur’an Al-Falah,” OfficialPonpesAlfalah. Accessed: Apr. 11, 2025. [Online]. Available: https://officialponpesalfalah.com

[14] B. McFee *et al.*, “librosa/librosa: 0.10.2.post1.” Zenodo, May 2024. doi: 10.5281/zenodo.11192913.

[15] G. Samara, E. Al-Daoud, N. Swerki, and D. Alzu’bi, “The Recognition of Holy Qur’an Reciters Using the MFCCs’ Technique and Deep Learning,” *Adv. Multimed.*, vol. 2023, pp. 1–14, Mar. 2023, doi: 10.1155/2023/2642558.

[16] T. O’Malley *et al.*, “KerasTuner.” 2019.

[17] I. Markoulidakis and G. Markoulidakis, “Probabilistic Confusion Matrix: A Novel Method for Machine Learning Algorithm Generalized Performance Analysis,” *Technol. 2024 Vol 12 Page 113*, vol. 12, no. 7, p. 113, Dec. 2024, doi: 10.3390/TECHNOLOGIES12070113.