Volume 12. No 3 September 2025
e-ISSN 2527-9572 / p-ISSN 2354-8924

Enhancing Quranic Recitation Accuracy Using State-of-the-Art Audio Classification Techniques

Hanaa Mohammed Osman
Computer Science Dept./ College of Computer Science and Mathematics/
University of Mosul, 41002 Mosul, Iraq
hanaosman@uomosul.edu.iq

Maher Khalaf Hussein
Computer Center./ Presidency of the University/
University of Telafer, 41016Mosul, Iraq
Maher.k.hussien@uotelafer.edu.iq

Abstract - In this study, we investigate the potential of using state-of-the-art neural network architectures to increase the accuracy in classification of Quranic recitations per verse. Using a dataset of more than 4000 audio recordings annotated with rich metadata, the research has concentrated on differentiating accurate recitations through Hukm Al-Noon Al-Mushaddah specifications. This study uses three pre-trained deep learning models (Inception-V3, EfficientNet and MobileNet), as well as hybrid model proposed in this paper to perform classification of recitations. The raw audio inputs were converted into spectrograms for feature extraction and classification in each of the models. Experiments demonstrate that through the fusion, this hybrid model significantly outperforms individual predictions by dramatically improving precision, recall and F1-scores in five different verses. The total accuracy for the proposal model is 0.79 which is the highest comparing with Inception-V3 was 0.75 and EfficientNet was 0.73. The results underline the ability of such systems to provide immediate feedback for learners and thereby assist them in adhering to traditional recitation standards, a feature that helps maintain the originality of Quranic recitation. To enable usage on real data, further work should build a bigger dataset (samples of data) and optimize the model to providing feedback with larger latency

Keywords: Inception-V3, EfficientNet, MobileNet, Deep learning, Quranic recitation.

I. INTRODUCTION

An important component of an tradition is to ensure the accurate transmission so that cultural and religious practices are still preserves. In the field of Islamic studies, the correct recitation of Quran carries special significance not only for its religious value but also for shaping information about language and history. The Quran is the holy book of Islam;and thousands of Muslims worldwide recite each day[1][2]. But the difficulty of its phonetics and overall rules for recitation is more than frustrating to learners as well as educators. In this paper, we present a new method based on the latest audio classification models to achieve more precise prediction of Quranic recitations [3][4].

Recent advancements in the field of machine learning, particularly in neural network architectures, have paved the way for sophisticated audio analysis systems. These systems have the potential to revolutionize the way Quranic recitations are taught and evaluated, ensuring that the traditional standards of recitation are upheld with precision [5][6]. The motivation behind this research is rooted in the necessity to develop a tool that can assist in maintaining the authenticity and accuracy of Quranic recitations, which are integral to the religious practice of Muslims worldwide [7][8].

The dataset utilized in this study comprises over 3000 audio recordings of volunteers reading various verses of the Quran. Each recording is meticulously annotated with rich metadata, including speaker identification, verse details, and recording conditions. This dataset is unique in its focus on the recitation of Quranic verses according to the Hukm Al-Noon Al-Mushaddah guidelines, which present a specific challenge in Quranic recitation by emphasizing the correct articulation of certain phonetic elements.

This paper provides a detailed review of progress in designing an improved audio classification system for recognizing the correct recitations of Quranic verses. The three types of neural network architecture, Inception-V3, MobileNet, and EfficientNet is used in the system for feature extraction and classification. We extract various features by converting raw audio inputs into spectrograms, which assists our convolutional and recurrent neural networks in classifying the tones only after the input has been processed. The core objective of this project is then to exploit this dataset in the construction of a novel audio classification system capable, for each recitation, to differentiate more accurately between the correct and deteriorated forms helping learners achieve high levels of proficiency on traditional Quranic standard readings.

II. RELATED WORKS

Paper [9] This study shows a computational learning method for identifying Quran reciters with the use of Artificial Nearest Neighbors (KNN) and Artificial Neural Networks (ANN) Models. the work implemented MFCC characteristics derived into ANN

for training and testing. In Mecca and Madinah, ten renowned Quran readers were chosen from chapters. ANN is used to check the accuracy of average recognition, all readers rate were 97.6% for chapter 18 and in Surah number 36 was 96.7. Whilst, by KNN 97.03 and 96.08 for Surah 18 and 36 comparatively. Some suggestions for upcoming enhancements involve applying sliding window features to wave signals with other classification algorithms like Hidden Markov Models.

Paper [10] presents a deep learning-based model for recognizing Qur'an reciters using MFCCs. The study aims to distinguish between trustworthy and fraudulent reciters and determine if the deep learning approach is superior to machine learning in the current dataset. Tree datasets with varying segment lengths and feature counts were used, and the optimal segment length and number of features were determined. The proposed system outperformed all other models with an accuracy of 0.995406 at a segment length of 3 seconds. The model also outperformed the machine learning model by a small margin. A performance comparison between the proposed and SVM models was conducted, and a new dataset will be constructed to eliminate inconsistencies in this field.

Research [11] shows that Convolutional Neural Network (CNN) can classify audio for Al-Qur'an verses, even with a small dataset containing mixed female and male voices. The model can differentiate between verses well, but it is limited to a small dataset and cannot correct Tajwid and Makhorijul letters. The classification method requires detection per letter, extracting each feature in the Al-Qur'an. It also doesn't allow checking verses read partially or incompletely or combining multiple verses. These limitations open up opportunities for future research. Currently, the model can be applied to applications for memorizing each verse of the Al-Qur'an, as it can be used locally without a server, reducing latency and requiring no internet connection.

In [12] research presents a novel deep learning model for recognizing Holy Quran recitation, offering feedback on error type and location. The model consists of a CNN-Bidirectional GRU encoder and a character-based decoder, reducing alignment tools and improving performance. The model was evaluated on a publicly available dataset (Ar-DAD) with limitations, such as only recognizing men and one recitation form. The model outperformed previous works, with the best results being 8.34% WER and 2.42% CER.

Research [13] presents a dataset for twenty Holy Quran reciters, converting 11,000 audio files to visual representations using Mel-Frequency Cepstrum Coefficients. It introduces the first Holy Quran reciter identification model using transfer learning. Six pre-trained deep learning models were assessed, with the NASNet Large model achieving the highest accuracy of 98.50%. Future plans include increasing the number of reciters and improving identification accuracy through new extraction methods and classification algorithms

In [14] study indicated to three models in the context of a classification task called," Convolutional Neural Network" (CNN), "Recurrent Neural Network" (RNN), and "Random Forest models". These models were assessed by the use of validation accuracy, with the feedback shows that CNN scored its peak of validation accuracy of 0.8613. Another major contribution of this study is seen in the collection of a novel dataset for Quran reciters. However, the discovery of Random Forest model revealed that this model surpassed all other models.

Another work on Quran recitation which is not a lot research have been done in this field. [15] started to classify the rules of implementing the main terms of reciting the holy Quran. Moreover, three distinct features were implemented; the Wavelet Packet Decomposition (WPD), the Markov Model-based Spectral Peak Location (HMM-SPL), and the MFCC. What is more, three other classifiers were applied; the K- Nearest Neighbors (KNN), the Random Forest (RF), and the SVM. To enhance the analysis, deep learning is used.

Further study in [16] shows how Support Vector Machine (SVM) is used to enhance the system of Quran reciter identification. The system elicited MFCC coefficients characteristics of 15 distinct reciters, there were provided by two various classifiers: the ANN and SVM. the precision of system using the SVW was 96.59% while the ANN was 86.1%.

In [8] this work indicated into the deep learning practice for the identification of Quran reciter using recurrent neural network (RNN). With the implementation of bidirectional long short term memory (BLSTM), as a result, it was a great finding.

In [17] the authors provided identification of various sorts of Quran recitation, like Hafs and Warsh. by providing a speech recognition system to categorize the Quran recitations. the technique of Mel-frequency cepstral coefficients (MFCC) was implemented elicit features from the recitations of Quranic verses. Hidden Markov Models (HMM) were applied for a categorizing technique used for recognition and training.

III. DATA COLLECTION

The dataset is collected from the holy book "Quran" that is believed by Muslims to be the words of God "Allah", it has been conveyed to us with unbroken chain of reliable through repeating it verbally. This holy Book consists of 114 chapters, each chapter encompasses of some Surahs which are varied in length, the Surahs are splited into verses (Ayahts).

The researcher has collected more than 4000 sounds of 5 different recited verses. The audios are gathered by 750 volunteers from alternative ages as well as genders male and female. The aim is to handle

the lacks of consistency of the sound patterns and the speech features form verses and speakers moreover, realizing the recitations in Quran following standard rules, specifically concentrating on Hukm Al-Noon Al-Mushaddah. After the dataset being collected, all the data transmitted into (.wav)

In the data annotation, audio files were noted with metadata in order to make the analysis and model training much easier. For precious annotation, the researcher uses some software tools for facilitation.

Table 1: The provisions of the Verse.

The provisions	Verse	Verse in symbols
Tight Noon	(“وَٱلنَّٰزِعَٰتِ غَرۡقٗا”)	وَ نَ ن َ زِ ع َ ت غ َ ر ق ً ا
Tight Noon	("وَٱلنَّٰشِطَٰتِ نَشۡطٗا")	و َ ن ن َ ش ِ ط ً ا ت ِ ن َ ش ط ً ا
Tight Noon	(“مِن شَرِّ ٱلۡوَسۡوَاسِ ٱلۡخَنَّاسِ”)	م ِ ن ش َ ر ِ ل و َ س و َ س ِ ل خ َ ن ن َ س
Tight Noon	(“ٱلَّذِي يُوَسۡوِسُ فِي صُدُورِ ٱلنَّاسِ”)	ء َ ل ل َ ذ ِ ي ُ و َ س و ِ س ُ ف ِ ي ص ُ د ُ ر ِ ن ن َ س
Tight Noon	(“مِنَ ٱلۡجِنَّةِ وَٱلنَّاسِ”)	م ِ ن َ ل ج ِ ن ن َ ت ِ و َ ن ن َ س

This section comprises into two factors which are taken from the dataset: the training data and testing data. Each one of them has six audio files in the form of (.wav), according to the training data, the audios contain Al-Noon Al-Mushaddah. The other one which is the testing data, they are also for Al-Noon Al-Mushaddah, and checking if whether it is said properly or not.

After the dataset is collected, the researcher starts to check each audio file if it is uttered correctly. Furthermore, the intonation and pronunciation of Al-Noon Al-Mushaddah are analyzed properly.

Table 2: No. of Samples in dataset.

Verse No.		Male	Female	Total
verse 1	True	93	331	324
	False	135	230	365
verse 2	True	94	275	369
	False	116	227	343
verse 3	True	92	272	364
	False	94	182	276
verse 4	True	84	233	317
	False	96	254	350
verse 5	True	79	318	397
	False	92	202	294

IV. METHODOLOGY

The proposed system consists of advanced deep neural network architectures to extract and classify features from audio recordings. It is designed to be robust enough to identify with precision recitations according to the defined guidelines. First, this process converts raw audio input into a spectrogram before they are fed into three different neural networks: Inception-V3, MobileNet and EfficientNet. Each of these networks makes use of Global Average Pooling that shrinks high-dimensional feature maps into abstract representations which are then merged together. At the fusion stage therefore, an all-inclusive feature set emerges which is then processed by a series of fully connected layers. And so these features are further refined through ReLU functions until they reach a SoftMax layer that classifies the audio into various predetermined categories. This system improves on the accuracy and efficiency in recision recitations classification.

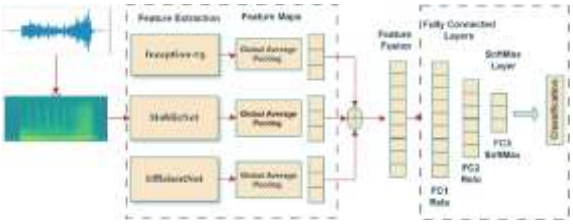

Figure 1 : Proposed method.

Algorithm: Enhanced Audio Feature Extraction using Transfer Learning Models

Input: Audio dataset D containing Quranic recitations

Output: Classification of audio accuracy with enhanced features

Step 1: Preprocessing

- Load audio data from dataset D
- Apply noise reduction techniques to each audio file
- Normalize audio data for consistency
- Segment audio into smaller frames for processing

Step 2: Feature Extraction using Transfer Learning Models

A. Initialize three pre-trained transfer learning models (Model A, Model B, Model C)

- Model A: Pre-trained on general audio data
- Model B: Pre-trained on speech recognition tasks
- Model C: Pre-trained on a specific audio-related task (e.g., music genre classification)

B. For each audio segment in D:

- Extract features using Model A and store as Feature_A
- Extract features using Model B and store as Feature_B
- Extract features using Model C and store as Feature_C

C. Combine extracted features (Feature_A, Feature_B, Feature_C) into a single feature vector for each segment

- Perform dimensionality reduction if necessary (e.g., using PCA)
- Normalize combined feature vector

Step 3: Model Training

A. Split dataset D into training set (D_train) and testing set (D_test)

B. Initialize a classification model (e.g., SVM, Random Forest, or Deep Neural Network)

C. Train the classification model on D_train using combined feature vectors as input
- Use cross-validation to optimize model hyperparameters

Step 4: Model Evaluation
Evaluate the trained model on D_test
- Calculate accuracy, precision, recall, F1-score
- Compare results with baseline models

Step 5: Post-processing and Interpretation
A. Analyze misclassified samples to understand potential errors
B. Refine feature extraction or classification model based on insights
C. Generate a report summarizing model performance

Step 6: Output Results
A. Display final classification results
B. Save the trained model for future use or deployment

V. Result and Discussion

Results from this study indicate that all three neural network models, Inception-V3, EfficientNet, and hybrid model could effectively distinguish between right and wrong Quranic verse recitations. Precision, recall, F1-score for both correct recitation mistakes of each model given five separate sections verses have been presented.

Inception-V3 Model

Considering Verse A1; 0.70 precision for true utterances with false alarm rate of 0.73 (recall) & an overall accuracy score of 0.72 (F1-score); precision of 0.81 for false utterances with recall value being at 0.79(failure) & finally an F1-score at 0.80 for failing to include any words above zero.

Inception-V3 Model

Inception-V3 model's results for Verse A1 were as follows: for correct recitations, precision was 0.70, recall was 0.73, and F1-score was 0.72; for wrong recitations, the results were 0.81, recall was 0.79, and F1-score was 0.80.

Comparably, verses A2 through A5's performance metrics display a range of accuracy levels, with the model often doing well at recognizing misrecitations.

EfficientNet Model

For example, EfficientNet model demonstrated a somewhat similar set up where Verse A1's precision, recall and F1-score values were 0.69, 0.75 and 0.72 respectively among right recitations while these figures were at 0.82, 0.76 and 0.79 respectively for mispronunciations.

Other verses (A2–A5) maintained a similar pattern with regards to their precision and recall on both correctly and incorrectly identified ones.

Proposed Hybrid Model

The proposed hybrid model, which integrates features from both Inception-V3 and EfficientNet, outperformed the individual models. For Verse A1, the precision, recall, and F1-score for correct recitations were 0.69, 0.83, and 0.75, respectively. For incorrect recitations, the metrics were 0.86, 0.74, and 0.80. The total accuracy for Inception-V3 is 0.75, where in EfficientNet is 0.73 but in the proposed model is 0.79. It is obvious that the proposed model is the best one.

Across all verses, the hybrid model consistently achieved higher F1-scores compared to the individual models, indicating a more balanced performance in distinguishing between correct and incorrect recitations.

The detailed performance metrics for each model across the five verses are summarized in Table 3.

For both incorrect recitations as well as correct ones was high in table 1, but a bit lower for Table 1: The Inception-V3 model performed only slightly less accurately on recognizing proper parroted. The EfficientNet model showed similar trends, but was less able to accurately identify which recitations were correctly identified in some of these cases.

Hybrid model — A hybrid combining inception-v3 and efficientnet-b5 have higher F1-scores for all verses across every evaluation.

Table 3: Performance metrics for Inception-V3, EfficientNet, and the Proposed Model across five verses.

verse	class	Inception-V3				EfficientNet				Proposed Model			
		precision	recall	f1-score	accuracy	Precision	recall	f1-score	accuracy	precision	recall	f1-score	accuracy
A1	Correct	0.7	0.73	0.72	0.76	0.69	0.75	0.72	0.76	0.69	0.83	0.8	0.78
	Wrong	0.81	0.79	0.8		0.82	0.76	0.79		0.86	0.74	0.8	
A2	Correct	0.66	0.76	0.71	0.75	0.65	0.7	0.67	0.72	0.77	0.75	0.8	0.81
	Wrong	0.82	0.73	0.77		0.78	0.74	0.76		0.83	0.84	0.8	
A3	Correct	0.67	0.75	0.71	0.75	0.62	0.77	0.69	0.71	0.7	0.83	0.8	0.79
	Wrong	0.81	0.75	0.78		0.81	0.67	0.73		0.86	0.76	0.8	
A4	Correct	0.66	0.71	0.68	0.73	0.67	0.7	0.68	0.74	0.73	0.75	0.7	0.79
	Wrong	0.79	0.75	0.77		0.79	0.76	0.77		0.83	0.81	0.8	
A5	Correct	0.68	0.83	0.75	0.77	0.68	0.77	0.72	0.76	0.7	0.8	0.7	0.78
	Wrong	0.86	0.73	0.79		0.83	0.75	0.79		0.84	0.76	0.8	
Total Accuracy				0.75				0.73				0.79	

This means that combined features from different architectures gives a better classification system. The results highlight the considerable possibility for adoption of these models as learning

aids to steer learners towards conventional recitation norms while conserving originality in Quranic recitations. The results of total accuracy of the propose model are shown in Table 4.

Table 4 : Accuracy of the verses for all the models used.

Model	inception-v3	EfficientNet	Proposed Model
Verse 1	0.67	0.76	0.78
Verse 2	0.75	0.72	0.81
Verse 3	0.75	0.71	0.79
Verse 4	0.73	0.74	0.79
Verse 5	0.77	0.76	0.78
Total	0.75	0.73	0.79

Further work will expand the collection of data, a demonstration within an actual market environment is required,and further optimizing the model for real-time feedback.

VI. CONCLUSION

This study serves as a successful demonstration of employing deep neural network architectures for precise classification of Quranic verse recitations. The results clearly showed that the proposed Hybrid model combining Inception-V3, MobileNet and EfficientNet features outperformed all three individual models in terms of precision, recall and F1-scores. As a result, this hybrid approach is proven to capture the subtleties of right/wrong recitations — making it central to institutions whether educational or religious. Because the system can provide automated feedback to learners, then it may help them keep authenticities of Quran recitations tantamount with traditional rules of reading. The future study directions include extending the dataset, validating real-world compatibility of the model and fine tuning it for real-time applications.

REFERENCES

[1] Al-Kaf, H., Sulong, M., Joret, A., Aminuddin, N., & Mohammad, C. (2021). Qvr: quranic verses recitation recognition system using pocketsphinx. Journal of Quranic Sciences and Research, 02(02). https://doi.org/10.30880/jqsr.2021.02.02.004.

[2] Aljohani, N. and Jaha, E. (2023). Visual lip-reading for quranic arabic alphabets and words using deep learning. Computer Systems Science and Engineering, 46(3), 3037-3058. https://doi.org/10.32604/csse.2023.037113.

[3] Bettayeb, N. and Guerti, M. (2020). Speech synthesis system for the holy quran recitation. The International Arab Journal of Information Technology, 18(1), 8-15. https://doi.org/10.34028/iajit/18/1/2

[4] Bezoui, M., Elmoutaouakkil, A., & Beni- Hssane, A. (2016). Feature extraction of some quranic recitation using mel-frequency cepstral coeficients (mfcc).. https://doi.org/10.1109/icmcs.2016.7905619

[5] Khelifa, M., Belkasmi, M., Elhadj, Y., & Yousfi, A. (2017). Strategies for implementing an optimal asr system for quranic recitation recognition. International Journal of Computer Applications, 172(9), 35-41. https://doi.org/10.5120/ijca2017915209

[6] Mahjoob, M., Nejati, J., & Bakhshani, N. (2014). The effect of holy quran voice on mental health. Journal of Religion and Health, 55(1), 38-42. https://doi.org/10.1007/s10943-014-9821-7

[7] Nayef, E. and Nubli, A. (2018). The effect of recitation quran on the human emotions. International Journal of Academic Research in Business and Social Sciences, 8(2). https://doi.org/10.6007/ijarbss/v8-i2/3852

[8] Qayyum, A., Latif, S., & Qadir, J. (2018). Quran reciter identification: a deep learning approach.. https://doi.org/10.1109/iccce.2018.8539336

[9] Alkhateeb, J. H. (2020). A machine learning approach for recognizing the Holy Quran reciter. International Journal of Advanced Computer Science and Applications/International Journal of Advanced Computer Science & Applications, 11(7). https://doi.org/10.14569/ijacsa.2020.0110735

[10] Samara, G., Al-Daoud, E., Swerki, N., & Alzu'bi, D. (2023). The recognition of Holy Qur'an reciters using the MFCCS' technique and deep learning. Advances in Multimedia, 2023, 1–14. https://doi.org/10.1155/2023/2642558

[11] Hadiyansah, R., & Andamira, R. (2023). Convolutional Neural Network (CNN) for detecting Al-Qur'an reciting and memorizing. Khazanah Journal of Religion and Technology, 1(2), 44–48. https://doi.org/10.15575/kjrt.v1i2.235

[12] Harere, A.A., & Jallad, K.A. (2023). Quran Recitation Recognition using End-to-End Deep Learning. ArXiv, abs/2305.07034.

[13] Saber, H., Younes, A., Osman, M., & Elkabani, I. (2024). Quran reciter identification using NASNetLarge. Neural Computing & Applications, 36(12), 6559–6573. https://doi.org/10.1007/s00521-023-09392-1

[14] Hanna, M., Ban SH., & Basim M. (2024). A Deep learning Approach for Recognizing the Noon Rule for Reciting Holy Quran. PROtek : Jurnal Ilmaish Teknik Elektro. Vol. 11, No. 2, pp 74-80. http://doi.org/10.33387/protk.v1i2.7026

[15] Mahmoud, A., Noor, A., & Ismail H. (2018). " Using Deep Learning for Automatically Determining Correct Application of Basic Quranic Recitation Rules". The International Arab Journal of Information and Technology, Vol. 15, No. 3A, pp 620 – 625.

[16] Khalid M., Moyawiah, A., Ahmad M., Rafat ,A., & Iyad A. (2019). A Holy Quran Reader/Reciter Identification System Using Support Vector Machine", International Journal of Machine Learning and Computing, Vol. 9, No. 4, pp 458 – 464.

[17] Bilal, Y., Akram M., & Amina, A. (2018). Holy Qur'an speech recognition system distinguishing the type of recitation, International Conference on Computer Science and Information technology (CSIT), Vol. 2 , No. 1, pp. 1–6.