Identification of Quran Reciters through Voice Analysis and Deep Learning

Mazin Mohamed Ashoor Al-Kathiri[1], Abdulqader Murad Abdulqader Basalama[2]

Abstract

In recent years, the field of Artificial Intelligence (AI) has witnessed tremendous progress, particularly in deep learning techniques that achieve remarkable results across various domains, including image processing, speech recognition, text processing, and computer vision. This research aims to leverage the capabilities of deep learning to develop an automatic identification system for Quran reciters based on their unique voices, utilizing deep convolutional neural networks (CNNs) and Log Mel Spectrograms as audio features extractor. Our study presents a series of models designed to classify Quran reciters based on their unique vocal characteristics. To enhance accessibility for users, we developed an Android application capable of running these models offline. Our model achieved an accuracy of 98% on pure sound signals for 23 reciters. However, we discovered that simply adding background noise to pure sounds, as suggested by previous studies, was inadequate for accurately representing real-world recordings. Due to the time-consuming nature of manual recordings, we recommend developing more advanced audio noise simulations that account for common signal distortions encountered in recordings from mobile devices. This could involve synthesizing typical background noises, recording artifacts, sound echo effects, and other real-world acoustic phenomena to create more realistic training data for the deep learning models.

Keywords: Artificial Intelligence, Deep Learning, Convolutional Neural Networks, Quran Recitation, Voice Identification, Mel Spectrograms, Audio Feature Extraction, Android Application, Acoustic Effects, Sound-based AI Applications.

[1] Department of Information Technology, College of Computers, Seiyun University
[2] Department of Computer Sciences, College of Computers, Seiyun University

I. INTRODUCTION

The identification of Quran reciters through voice is similar to the well-known problem faced by researchers in the field of machine learning known as " Speaker recognition". Speaker recognition is categorized into speaker identification and speaker verification[1] . Verification is the task of automatically determining if a person is a person. Identification is the process of automatically identifying an individual's identity from many speakers based on the unique characteristics of their voice. It is a vital technique that uses the patterns and distinctive features of a person's voice to determine their identity. Most research in this area focuses on English voices, and there is a need for more studies in the Arabic language, particularly in Quran recitation[2-6]. This need arises from the increasing demand for audio recordings of reciters and Quran applications due to the global spread of Islam. Achieving accurate identification of Quran reciters based on voice presents a significant challenge due to the increasing number of reciters, background noise, and recording artifacts [7]. These factors make it difficult to achieve high accuracy in recognizing a reciter's voice using traditional methods. Recently, several studies have focused on identifying reciters through voice recognition. However, many of these studies are limited by small datasets or rely on pure audio recordings[3, 4]. Additionally, they often overlook the challenges posed by low-quality audio, such as background noise and sound distortions. This makes it difficult to apply their findings in real-world situations and develop practical applications for users.

II. Literature Review

In this section, we provide a comprehensive summary of prior research, primarily building upon the findings presented in research paper [3]. This foundational paper offers an extensive survey of the challenges associated with identifying the reciters of the Holy Quran, detailing various proposed techniques and models within this domain. Our focus will be on the advancements made over the past decade, specifically from 2012 to 2024, highlighting key developments and trends that have emerged in this field of study. To review and analyze the relevant research papers on this problem, specifically focusing on studies that address the challenge of identifying Quran reciters through voice analysis. To ensure a thorough examination, we applied a set of inclusion and exclusion criteria to filter the retrieved papers, categorizing them into two distinct groups: relevant studies that utilize speaker identification techniques for this purpose, and those that do not meet our criteria. Each published paper was carefully assessed based on its title, abstract, and main content to confirm its relevance to our objectives. Through this rigorous process, we identified 17 papers that aligned with our research conditions and effectively contribute to the study's goals see Table 1 previous studies. This curated selection reflects the current state of research in the field and provides a solid foundation for further exploration.

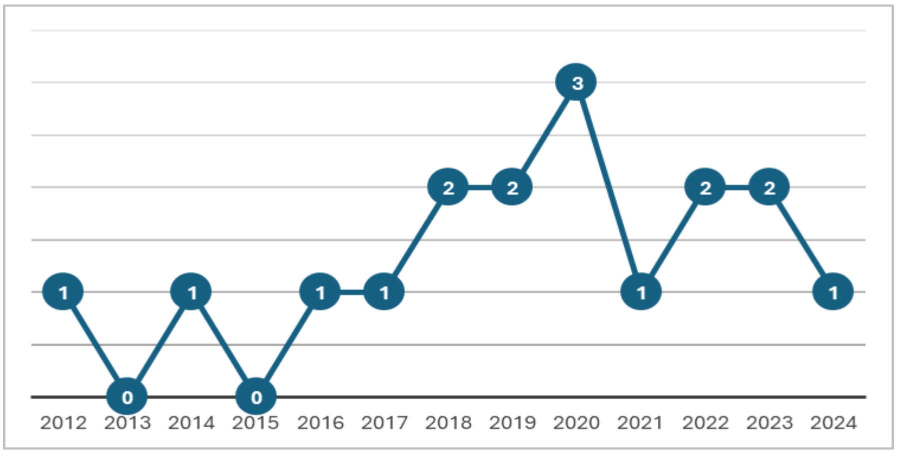


Figure 1 Papers' Publication Year

- **Statistical Results of Previous Studies**

Paper Reference	Publication Year	Number of Reciters	Features Extraction Technique	Classifiers
[8]	2012	20	MFCC	LBG-VQ (Linde-Buzo-Gray Vector Quantization)
[9]	2014	4	DWT (Discrete Wavelet Transform) and LPC (Linear Predictive Coding)	Random Forest (RF)
[10]	2016	5	MFCC	-
[11]	2017	5	MFCC	Gaussian Mixture Model (GMM)
[12]	2018	7	perceptual features	Support Vector Machine (SVM)
[13]	2018	-	MFCC	Bidirectional Long

				Short-Term Memory (BLSTM)
[14]	2019	15	MFCC	SVM, ANN
[15]	2019	12	MFCCs and Pitch Auto-correlograms	Naïve Bayes J48 RF
[16]	2020	30	perceptual and acoustic features	CNN, Decision Tree (DT), RF, LR, SVM-LINEAR, SVM-RBF, XG, GMM-UBM, BLSTM
[17]	2020	7	perceptual features, short time energy	XGBoost, SVM, SVM-RBF, DT LR, RF
[18]	2020	10	MFCC	ANN, K-Nearest Neighbors (KNN)
[19]	2021	10	wav2vec2.0, HuBERT	Multilayer Perceptron, RNN, CNN
[20]	2022	14	MFCC	LBG-VQ
[21]	2022	10	MFCC	KNN, ANN, SVM
[2]	2023	7	MFCC	CNN
[4]	2023	50	MFCC, TRILL-50, VGGish-50	CNN
[22]	**2024**	**20**	MFCC	NASNet**(CNN +** Controller Recurrent Neural Network CRNN **)**

Figure 1 illustrates the statistical distribution of the selected papers based on their publication year. The data spans from 2012 to 2024, showcasing the gradual increase in publications. The distribution indicates that the interest in the topic has steadily grown, reflecting an expanding body of knowledge and an increasing recognition of the importance of identifying Quran reciters through voice analysis.

Table 1 previous studies

In Figure 3 we observe a wide range in the number of reciters across the studies, with the highest being 50 reciters reported in study [23]. In contrast, some papers classified as few as 4 reciters like in [9].

Figure 2 illustrates the distribution of types of audio feature extractors used. The graph represents the frequency of different feature types and their respective counts. The prominent presence of MFCC (Mel-Frequency Cepstral Coefficients), with 11 instances, highlights its significance in capturing the spectral characteristics of audio signals. This prominence aligns with its widespread use in speech and music processing.

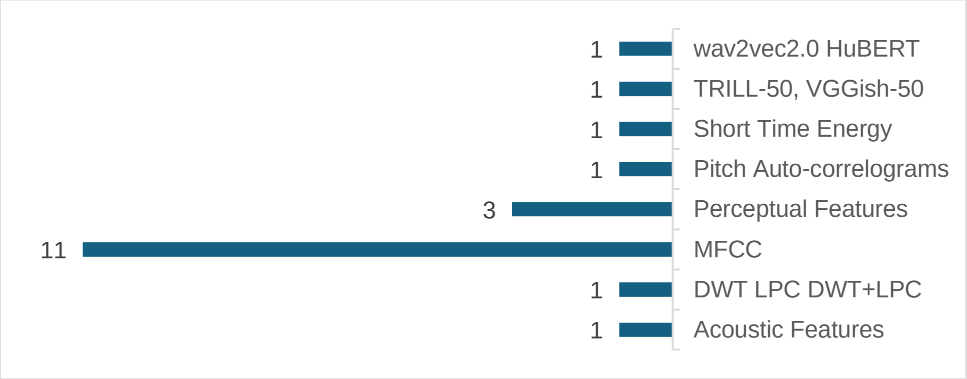
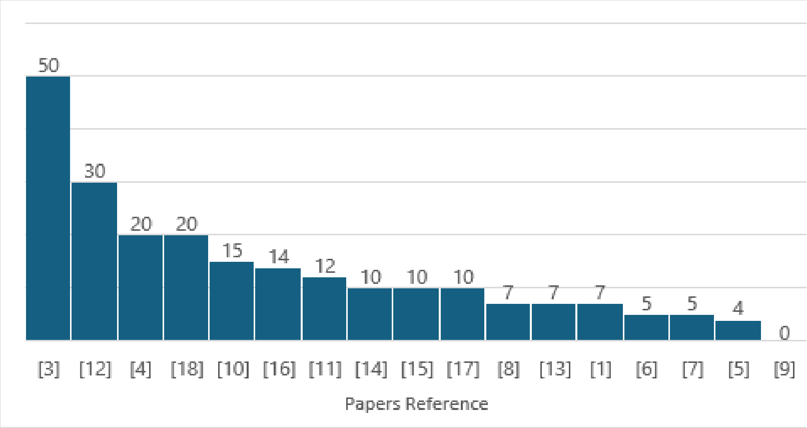


Figure 2 distribution of types of audio feature extractors used in previous studies



In Figure 4, we observe the frequency distribution of various classifiers used for Quran reciter identification. The Support Vector Machine (SVM) emerges as the most prevalent classifier, appearing 6 times, while the Convolutional Neural Network (CNN) follows closely with 5 occurrences. This trend highlights the increasing popularity of both classifiers in

Figure 3 Number of Reciters Classified in The Papers

audio classification tasks. However, in our study, we choose CNN for Quran reciter identification because CNNs have gained significant popularity and demonstrated superior precision compared to traditional Support Vector Machines (SVMs) in audio tasks, making them a compelling choice for this application [2, 24, 25]. The ability of CNNs to automatically learn features from audio data allows them to capture complex patterns that are crucial for distinguishing between different reciters of the Quran, thereby improving identification performance.

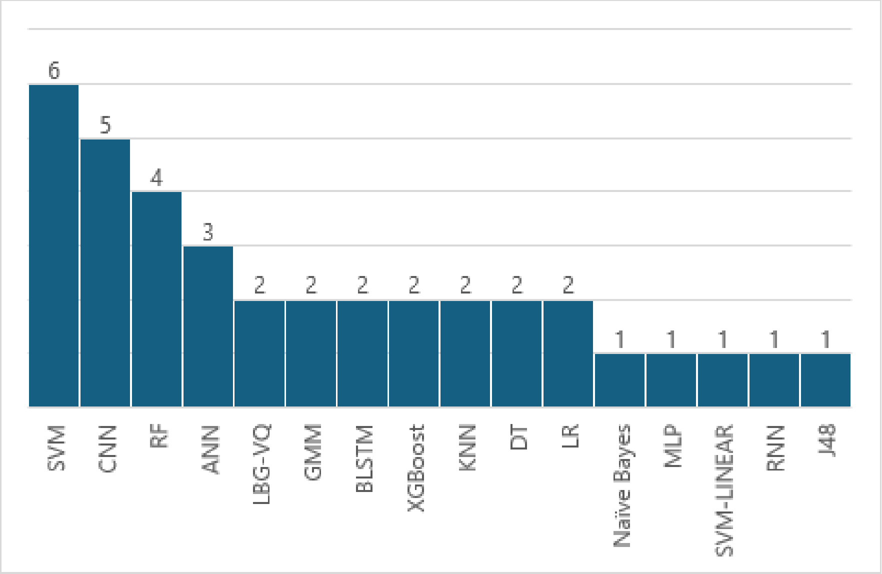


Figure 4 distribution of types of machine learning models used as classifiers in previous studies

- **Reciter Identification on Mobile Applications**

One of the most successful mobile applications for music recognition is Shazam, which is available on smartphones. The app works by analyzes the audio and compares it to a vast server database containing millions of songs and recordings.

To identify apps similar to Shazam that focus on recognizing Quran reciters, we conducted a search in both the Android Google Play Store and the iOS App Store. After reviewing the available applications, we found the "Ratel" app, developed by STC, the leading telecommunications provider in Saudi Arabia. This app serves as a comprehensive Quran application, offering features for reading the Quran and dhikr, and listening audio recitations, along with a feature for identifying Quran reciters.

However, we noted several limitations with the identification feature. It often struggles to recognize most recorded audio accurately and can only operate using the phone's built-in microphone. Additionally, an internet connection is required to utilize this feature. Moreover, this feature was recently removed in the latest updates.

- **Discussion**

Currently, there is no standard method to objectively evaluate the accuracy of models in the context of identifying Quran reciters, due to several reasons:

- Variations in the type and size of training and testing data, as data significantly impacts accuracy.
- Differences in the duration of model training and variations in computational resources.
- Discrepancies in feature extraction techniques.
- Differences in the number of reciters that models can identify.

The main issue with these studies is their reliance on small datasets or their dependence solely on pure audio recordings [3]. Furthermore, they often overlook the challenges associated with low-quality audio, including background noise and artifacts from mobile recordings. As a result, applying their findings to real-world scenarios and creating practical applications for users becomes challenging.

III. METHODOLOGIES

This section presents a machine learning-based system developed for the recognition of Qur'an reciters, with a particular emphasis on the use of Convolutional Neural Networks (CNNs) as the core classification model. The system processes recitation audio by first extracting features using Log-Mel spectrograms generated through the Fast Fourier Transform (FFT), which effectively captures the spectral properties of speech. The complete system is structured into six key phases: data acquisition, data preprocessing, feature extraction, model training, model evaluation, and model deployment.

- **Data Acquisition**

The most important step in any Deep learning system is collecting the data. Deep learning systems require large amounts of data to be successful. This data will be used to train and test the Deep learning models. In our system, a dataset will be created containing audio recordings of recitations. These recordings will be available in two types: original pure recordings and noisy recordings that contain some impurities, such as background sounds, or variations in volume. The original pure recordings were obtained from the mp3quran website. To automatically access the audio files on the website, we employed web scraping technology. As for the noisy data, it will be generated using two methods. The first method involves making manual recordings in a natural environment using mobile phones. The second method involves artificially mixing sounds—a synthetic approach that simulates manual recordings by adding noise to the pure audio. This process makes the data noisy, allowing the model to become more robust in handling recordings with background noise.

- **Data Preprocessing**

After collecting the audio dataset, it is converted into WAV format, which is an uncompressed audio format. And the sample rate is then reduced to 16,000 samples per second. Additionally, all audio channels are combined into a single channel to optimize resource usage. Each audio file is divided into equal segments of 8 seconds for training samples, and all training samples for each reciter are saved in a dedicated folder specific to that reciter.

- **Features Extraction [26]**

The acoustic characteristics of the recitation signals pose a challenge due to their variability and inconsistency Figure 5. This contrasts with music recognition, where clear, stable, and relatively short patterns emerge see Figure 6.

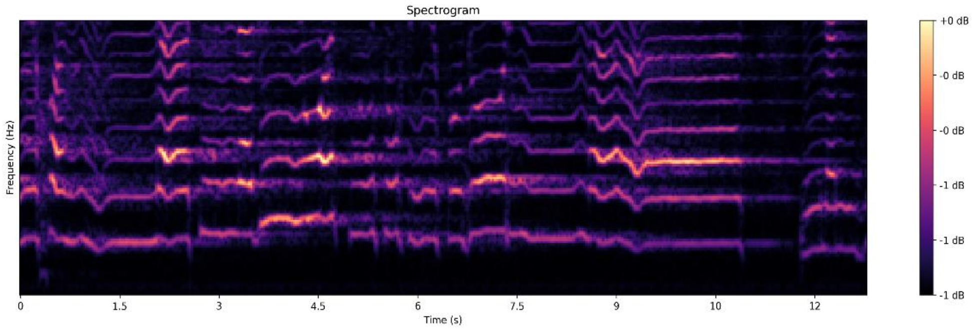


Figure 5 spectrogram of recitation audio signals

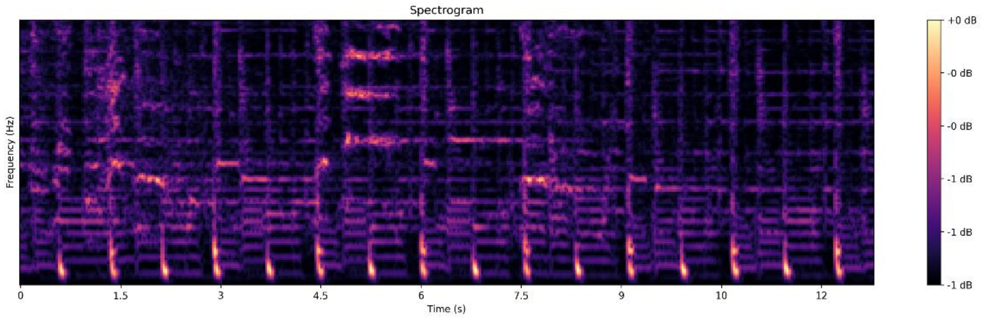


Figure 6 spectrogram of music audio signals

$(128{,}400) = \left(\text{mel bands}, \left(\frac{\text{Sample duration}}{\text{STFT hop seconds}}\right)\right)$Therefore, a longer analysis window is required to capture the patterns in recitation signals. In this project, a window length of 8 seconds was used for each sample. For feature extraction, we do not utilize Mel-Frequency Cepstral Coefficients (MFCCs); instead, we use Log Mel spectrograms, which serve as a precursor to MFCCs. Log Mel spectrograms provide a more detailed and interpretable representation. However, Mel-Frequency Cepstral Coefficients

(MFCC) features involve a loss of some fine-grained details due to their filter bank design [27].

The relationship between the Mel spectrogram and MFCCs is defined by the following equation:

MFCCs = DCT (log (Mel Spectrogram))

Here, the Discrete Cosine Transform (DCT) is applied to the logarithm of the Mel spectrogram to derive the MFCCs. Log Mel spectrograms utilize a logarithmic transformation of the frequency of a signal based on the Mel scale. This transformation is grounded in the principle that sounds perceived at equal intervals on the Mel scale are also experienced as equidistant by humans.

Here are The Parameters and Values Used for Features Extraction

Table 2 features extraction parameters

Properties	Values
Sample Rate in hertz	16000.0
Sample Duration	8.0
STFT window in seconds	0.08
STFT hop in seconds	0.02
Mel min in hertz	20
Mel max in hertz	2048
Mel bands	128

These operations result in an image of size (128,400) see Figure 7 log mel spectrogram

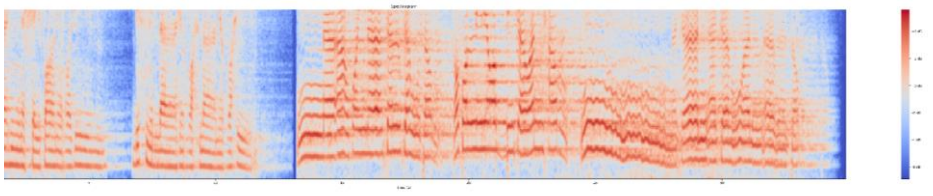


Figure 7 log mel spectrogram

And to enhance the intensity of the main signal and suppress the frequencies of side signals, normalization was performed on the spectral values between 0 and 1. This was followed by raising these values to the fourth power (X^4) to eliminate faint background signals. Then another normalization was then conducted, this time within a range of 1 to -1, as this range was found to be more suitable for convolutional networks

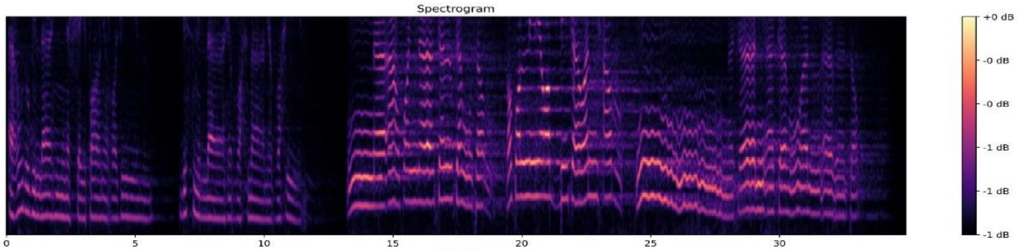


Figure 8 enhanced log mel spectrogram

- **Model Training**

Building and training a model for voice recognition of Quran reciters presents a significant challenge due to the large number of reciters. This requires a substantial amount of training data for each reciter, which in turn necessitates a very large storage capacity. Additionally, training the model with such a volume of data becomes difficult due to limited RAM and GPU memory during processing, which also increases the training time. Consequently, this necessitates the use of unconventional methods for training the model.

To train the model for a large number of reciters, a divide-and-conquer strategy was employed due to the challenges of training many reciters simultaneously. The training was conducted in two phases:

Phase 1

The reciters were divided into groups of ten, resulting in a total of 8 different models being trained, as 80 reciters were selected. For training these sub-models, each reciter contributed 6,000 samples, which were obtained from pure audio recordings mixed with random noise. This process led to a total of 60,000 samples for each group. The data was then divided into training and testing sets at a ratio of 80% to 20%.

Phase 2

After training each model for the individual groups, the convolutional networks were extracted, and the dense networks from the previous models were removed. This modification ensured that the output for each model consisted of 256 values derived from the convolutional network. These outputs were then merged into a unified model in parallel, facilitating the extraction of patterns and acoustic features see Figure 9.

This process resulted in an output of 256 × 8 = 2048 for each sample, significantly reducing the data space.

Subsequently, the acoustic features and patterns were extracted from the sounds using this unified model and stored in a two-dimensional vector for each reciter. This data was then used to train a dense model with the stored features.

Finally, the unified model was trained using 100,000 samples from 23 reciters, as shown in Table 3 This decision was made due to limited resources, as opposed to using samples from all 80 reciters. The samples were obtained from pure audio recordings that were mixed with random noise. The data was then divided into training and testing sets at a ratio of 80% to 20%.

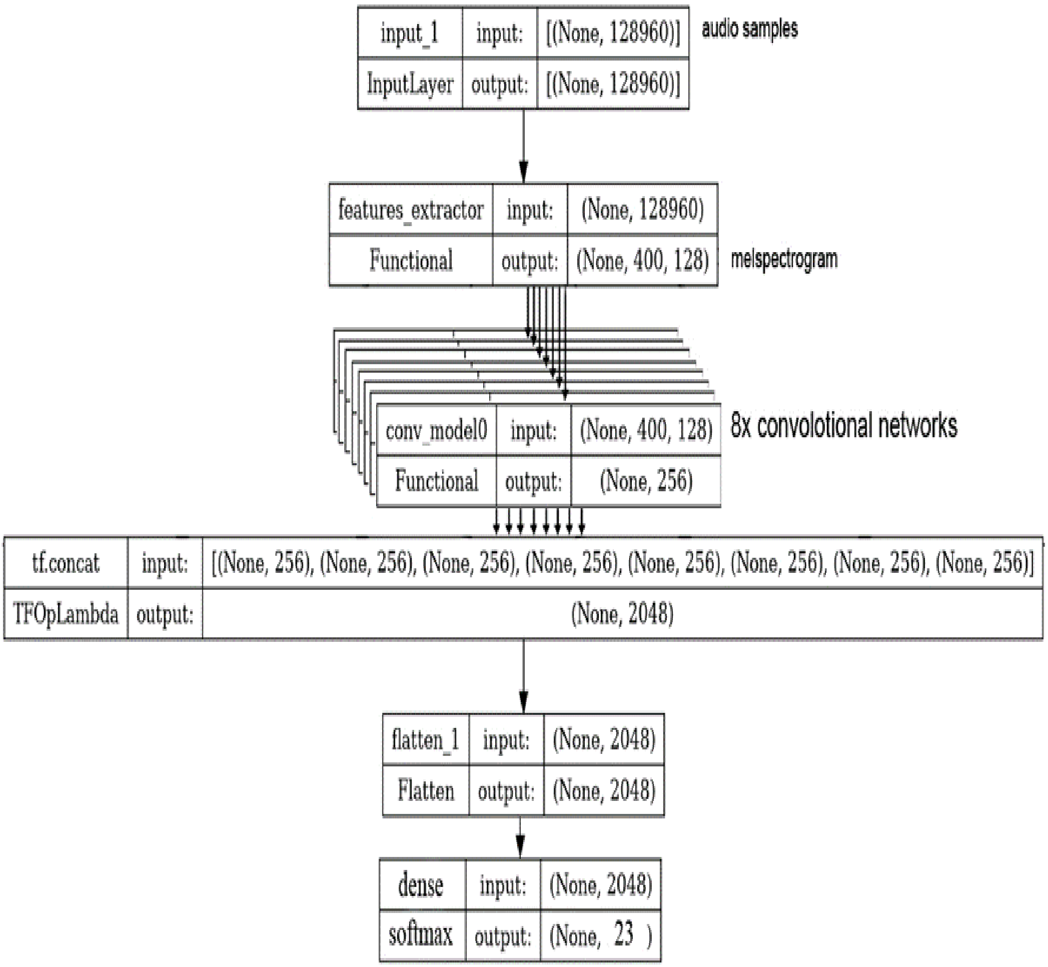


Figure 9 unified model architecture

	Key	Name	Arabic Name
0	a_jbr	Ali Jaber	علي جابر
1	a_klb	Adel Al-Khalbany	عادل الكلباني
2	a_swaiyd	Abdulrahman Alsuwayid	عبد الرحمن السويد
3	a_turki	Abdulaziz Alturki	عبد العزيز التركي
4	aabd-lkrym-lh-zmy-2	Abdulkareem Alhazmi	عبد الكريم الحزمي
5	aabd-lrhmn-lshh-t	Abdulrahman Alshahhat	عبد الرحمن الشحات
6	abkr	Idrees Abkr	إدريس ابكر
7	afs	Mishary Alafasi	مشاري العفاسي
8	ahmad_huth	Ahmad Alhuthaifi	أحمد الحذيفي
9	ahmd-aays-lmaasr-oy	Ahmad Issa Al Maasaraawi	أحمد عيسى المصراوي
10	ajm	Ahmad Al-Ajmy	أحمد العجمي
11	arkani	Abdulwali Al-Arkani	عبد الولي الاركاني
12	ayyub	Mohammed Ayyub	محمد أيوب
13	bader	Bader Alturki	بدر التركي
14	rashad	Mohammad Alshareef	محمد رشاد الشريف
15	s_bud	Salah Albudair	صلاح البدير
16	s_gmd	Saad Al-Ghamdi	سعد الغامدي
17	saad	Saad Almqren	سعد المقرن
18	saber	Ahmad Saber	أحمد صابر
19	salah_hashim_m	Salah Alhashim	صلاح الهاشم
20	salamah	Yasser Salamah	ياسر سلامة
21	sds	Abdulrahman Alsudaes	عبد الرحمن السديس
22	shaheen	Ahmad Shaheen	أحمد شاهين

Table 3 reciters key-names

- **Model Evaluation**

After training the unified model for 40 epochs, the performance results are as follows: the model achieved an accuracy

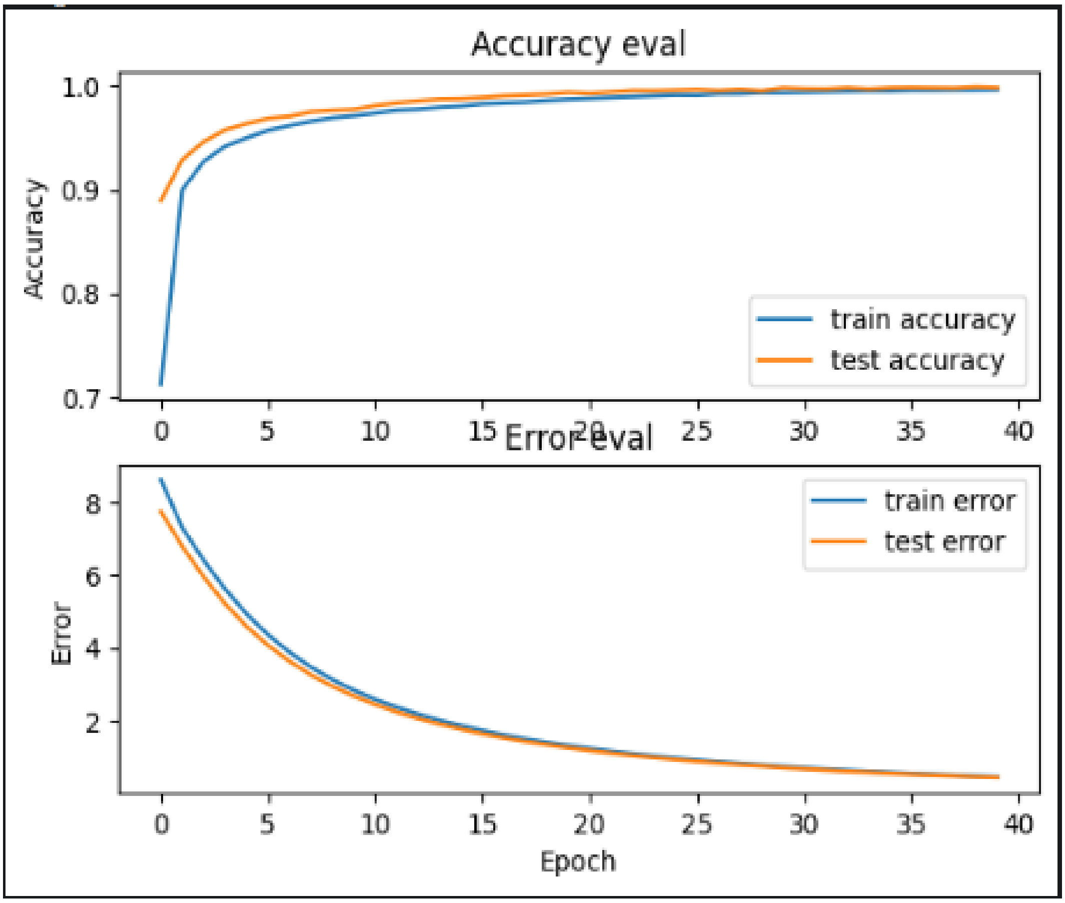


Figure 10 model evaluation graph

of 99.65% on the training dataset and 99.84% on the validation dataset.

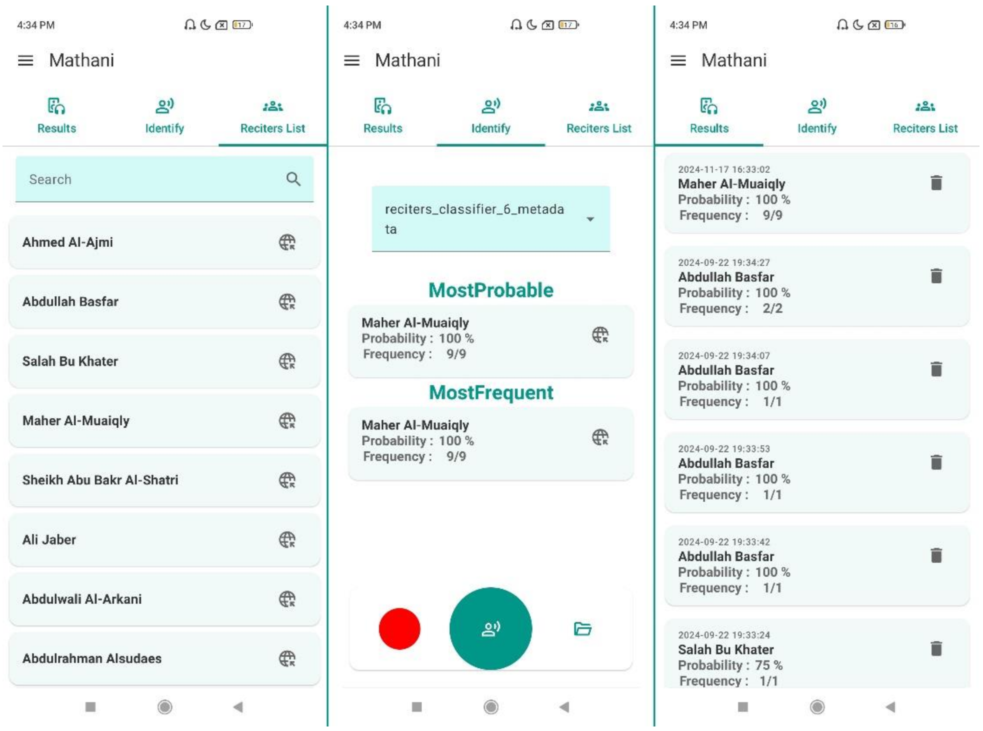
```
Epoch 40/40

1250/1250 [==============================] - 15s 12ms/step -
loss: 0.4494 - accuracy: 0.9965 - val_loss: 0.4324 - val_accuracy: 0.9984
```

- **Model Deployment**

The deployment of the model is a crucial phase in ensuring that the application functions effectively in real-world scenarios. The application has been developed using Jetpack Compose, a modern UI toolkit that utilizes the Kotlin programming language, and it employs

TensorFlow Lite (TFLite) for model functionality. The primary objective of this application is to run the model on Android devices, enabling users to identify reciters without needing an internet connection.



Search

Ahmed Al-Ajmi
Abdullah Basfar
Salah Bu Khater
Maher Al-Muaiqly
Sheikh Abu Bakr Al-Shatri
Ali Jaber
Abdulwali Al-Arkani
Abdulrahman Alsudaes

reciters_classifier_6_metadata

MostProbable
Maher Al-Muaiqly
Probability : 100 %
Frequency : 9/9

MostFrequent
Maher Al-Muaiqly
Probability : 100 %
Frequency : 9/9

2024-11-17 16:33:02
Maher Al-Muaiqly
Probability : 100 %
Frequency : 9/9

2024-09-22 19:34:27
Abdullah Basfar
Probability : 100 %
Frequency : 2/2

2024-09-22 19:34:07
Abdullah Basfar
Probability : 100 %
Frequency : 1/1

2024-09-22 19:33:53
Abdullah Basfar
Probability : 100 %
Frequency : 1/1

2024-09-22 19:33:42
Abdullah Basfar
Probability : 100 %
Frequency : 1/1

2024-09-22 19:33:24
Salah Bu Khater
Probability : 75 %
Frequency : 1/1

Figure 11 App UI

IV. RESULTS & DISCUSSION

The goal of a classification model is usually not just to achieve good classification on the given dataset, but to achieve good classification on new data that we haven't seen before. To test the model on new data, 44,000 samples that it hadn't been trained on were taken, and standard performance metrics were applied to them. In multi-class classification, we deal with multiple categories, so simply calculating the model's accuracy by dividing the number of correct predictions by the total number of predictions does not provide a complete picture of the model's performance, especially in cases where the classes are imbalanced or when the model is biased toward a particular category. To address these issues, we use additional metrics such as Precision, Recall, F1-Score, and the Confusion Matrix to gain a more comprehensive understanding of the model's performance. By considering these metrics collectively, we can gain insights into the model's strengths and weaknesses and make informed decisions about how to improve its performance.

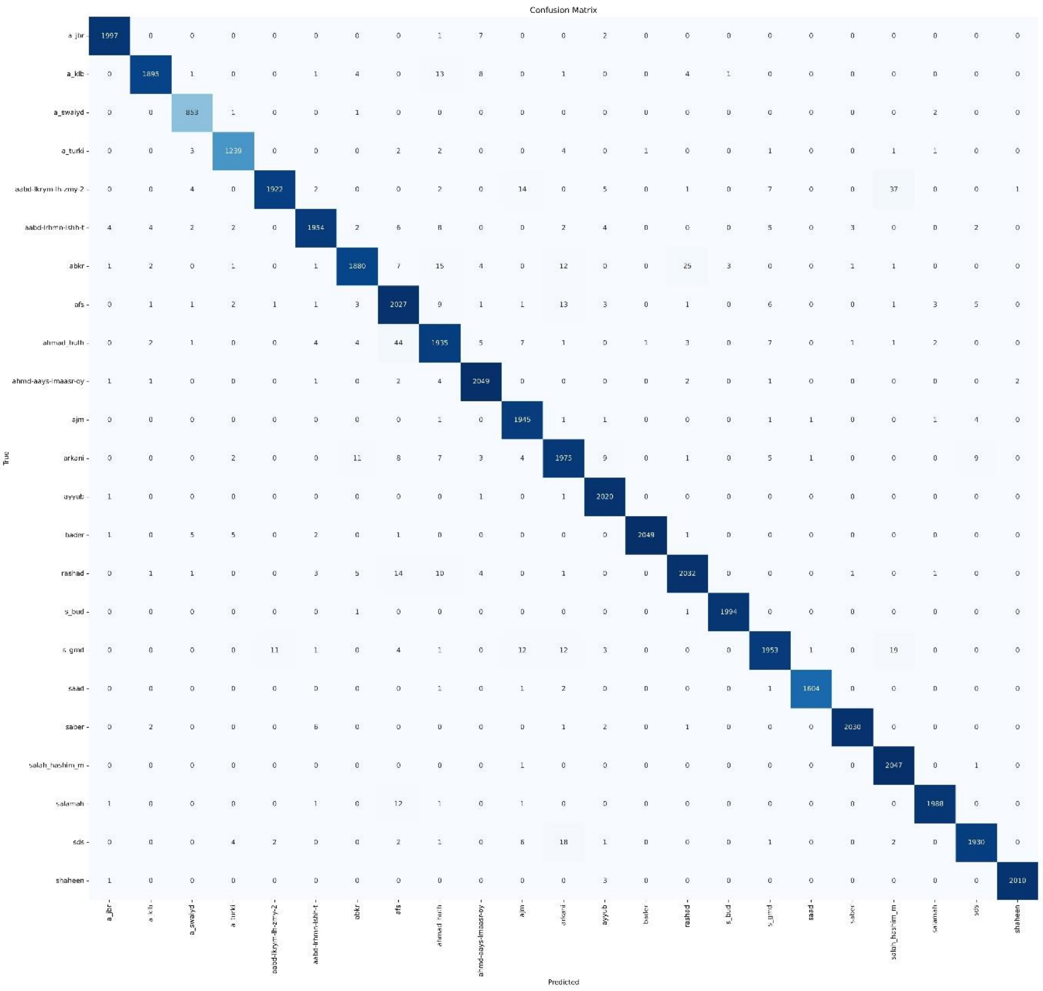


Figure 12 Confusion Matrix

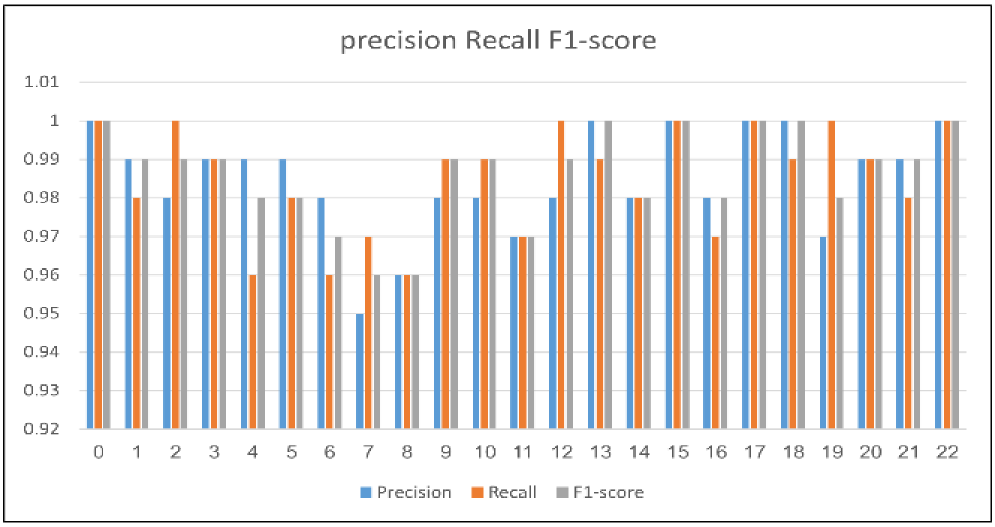


Figure 13 Precision, Recall, F1-Score

After applying the previously mentioned metrics to the test dataset, we obtained the following results:

Table 4 Results

	Precision	Recall	F1-score	Support
Macro Avg	0.98	0.99	0.98	44000
Weighted Avg	0.98	0.98	0.98	44000
Accuracy			0.98	44000

V. CONCLUSION

Although the model performs excellently with pure audio signals and with artificially noisy audio signals, it does not yield the same quality results when applied to recordings made in real noisy environments, despite being trained with artificially added noise. This indicates that the approach of merging clean audio signals with noise, as demonstrated in previous studies, may not be sufficient to accurately represent data in real-world scenarios. The inherent distortion and relative weakening of the signal are evident when comparing Figure 14, which displays the pure audio signal, to Figure 15, which illustrates the artificially corrupted signal created by blending the clean audio with background noise. Furthermore, Figure 16 presents the actual noisy audio signal recorded using a mobile device, highlighting the significant differences. These findings emphasize the need for more robust techniques that can better accommodate the complexities of real-world audio environments.

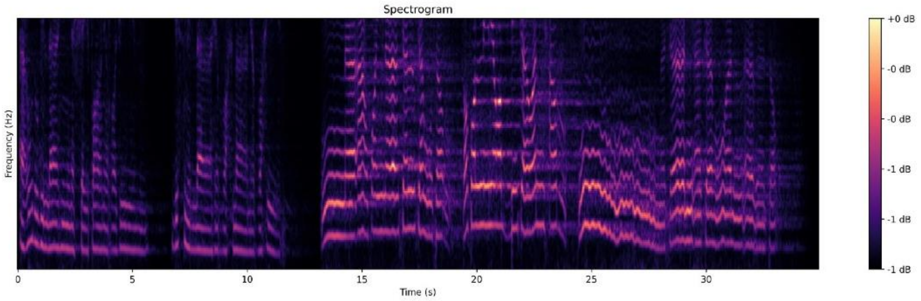


Figure 14 pure audio signal

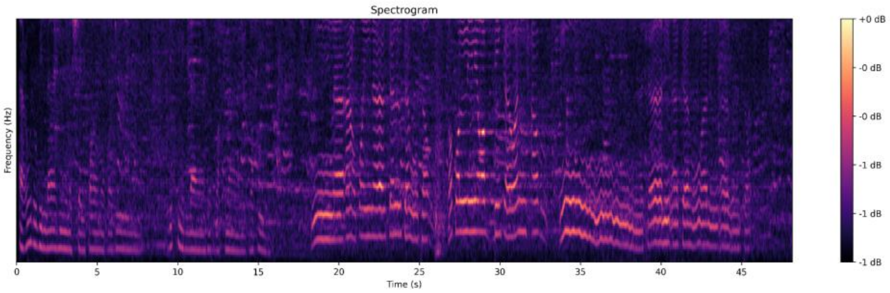


Figure 15 artificially noisy audio signal

In this figure, the main audio signal was not significantly affected despite the addition of background noise and remained similar to the pure signal.

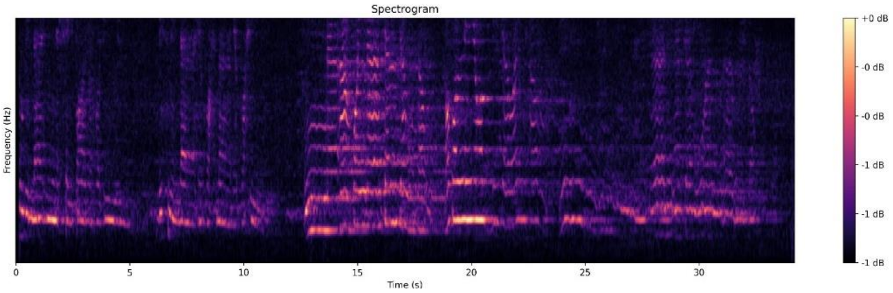


Figure 16 real noisy audio signal

As illustrated in the figure, the audio signals exhibit significant distortion and interference.

Future Works:

The proposed solutions aim to enhance audio signal processing in noisy environments. The first involves forming a manual recording team, which, while effective, may be costly and time-consuming, making it unsuitable for large-scale use. The second suggests developing an advanced audio simulation system that replicates real-world audio effects like distortions, background noise, and echoes, offering a more accurate and efficient alternative.

VI. REFERENCES

1. Brydinskyi, V., et al., *Comparison of modern deep learning models for speaker verification.* Applied Sciences, 2024. **14**(4): p. 1329.
2. Samara, G., E. Al-Daoud, N. Swerki, and D. Alzu'bi, *The Recognition of Holy Qur'an Reciters Using the MFCCs' Technique and Deep Learning.* Advances in Multimedia, 2023. **2023**: p. 2642558.
3. Alatiyyah, M., *QURAN RECITER IDENTIFICATION: TECHNIQUES AND CHALLENGES.* 2023.
4. Tall, M. *Deep learning for Quranic reciter recognition and audio content identification*. in *Deep Learning Indaba 2023*. 2023.
5. Mahmood, A., *Arabic speaker recognition system based on phoneme fusion.* Multimedia Tools and Applications, 2024. **83**(5): p. 15043-15060.
6. Balula, N.O.m., M. Rashwan, and S. Abdou, *Automatic speech recognition (ASR) systems for learning Arabic language and Al-quran recitation: a Review.* International Journal of Computer Science and Mobile Computing, 2021. **10**(7): p. 91-100.
7. Harere, A.A. and K.A. Jallad, *Quran recitation recognition using end-to-end deep learning.* arXiv preprint arXiv:2305.07034, 2023.
8. Hussaini, M.A. and R.W. Aldhaheri. *An Automatic Qari Recognition System*. in *2012 International Conference on Advanced Computer Science Applications and Technologies (ACSAT)*. 2012.
9. Shah, S.M. and S.N. Ahsan, *Arabic speaker identification system using combination of DWT and LPC features.* 2014 International Conference on Open Source Systems & Technologies, 2014: p. 176-181.
10. Bezoui, M., A. Elmoutaouakkil, and A. Beni-hssane. *Feature extraction of some Quranic recitation using Mel-Frequency Cepstral Coeficients (MFCC)*. in *2016 5th International Conference on Multimedia Computing and Systems (ICMCS)*. 2016.
11. Gunawan, T., N. Saleh, and M. Kartiwi, *Development of Quranic Reciter Identification System using MFCC and GMM Classifier.* International Journal of Electrical and Computer Engineering (IJECE), 2017. **8**.
12. Elnagar, A., R. Ismail, B. Alattas, and A.K.S.B.Q. Alfalasi, *Automatic Classification of Reciters of Quranic Audio Clips.* 2018 IEEE/ACS 15th International Conference on Computer Systems and Applications (AICCSA), 2018: p. 1-6.
13. Qayyum, A., S. Latif, and J. Qadir, *Quran Reciter Identification: A Deep Learning Approach.* 2018 7th International Conference on Computer and Communication Engineering (ICCCE), 2018: p. 492-497.
14. Nahar, K., et al., *A Holy Quran Reader/Reciter Identification System Using Support Vector Machine.* International Journal of Machine Learning and Computing, 2019. **9**: p. 458-464.
15. Khan, R., M. Hadwan, and A. Qamar, *Quranic Reciter Recognition: A Machine Learning Approach.* Advances in Science, Technology and Engineering Systems Journal, 2019. **4**.

16. Lataifeh, M., A. Elnagar, I. Shahin, and A.B. Nassif, *Arabic audio clips: Identification and discrimination of authentic Cantillations from imitations.* Neurocomputing, 2020. **418**: p. 162-177.
17. Elnagar, A. and M. Lataifeh, *Predicting Quranic Audio Clips Reciters Using Classical Machine Learning Algorithms: A Comparative Study*, in *Recent Advances in NLP: The Case of Arabic Language*, M. Abd Elaziz, M.A.A. Al-qaness, A.A. Ewees, and A. Dahou, Editors. 2020, Springer International Publishing: Cham. p. 187-209.
18. Alkhateeb, J., *A Machine Learning Approach for Recognizing the Holy Quran Reciter.* International Journal of Advanced Computer Science and Applications, 2020. **11**.
19. Moustafa, A. and S.A. Aly, *Towards an Efficient Voice Identification Using Wav2Vec2.0 and HuBERT Based on the Quran Reciters Dataset.* ArXiv, 2021. **abs/2111.06331**.
20. Al-Jarrah, M., et al., *Accurate Reader Identification for the Arabic Holy Quran Recitations Based on an Enhanced VQ Algorithm.* Revue d'Intelligence Artificielle, 2022. **36**: p. 815-823.
21. Meshal Mohammed Al Anazi, a.O.R.S., *A Machine Learning Model for the Identification of the Holy*
Quran Reciter Utilizing K-Nearest Neighbor and Artificial
Neural Networks. Information Sciences Letters, 2022.
22. Saber, H.-A., A. Younes, M. Osman, and I. Elkabani, *Quran reciter identification using NASNetLarge.* Neural Computing and Applications, 2024. **36**(12): p. 6559-6573.
23. Hadwan, M., H. Alsayadi, and S. Al-Hagree, *An End-to-End Transformer-Based Automatic Speech Recognition for Qur'an Reciters.* Computers, Materials and Continua, 2022. **74**: p. 3471-3487.
24. Liu, T., et al., *Identification of Fake Stereo Audio Using SVM and CNN.* Information, 2021. **12**(7): p. 263.
25. Zhang, B., C. Quan, and F. Ren. *Study on CNN in the recognition of emotion in audio and images*. in *2016 IEEE/ACIS 15th International Conference on Computer and Information Science (ICIS)*. 2016.
26. Kumar, D., *Feature normalisation for robust speech recognition.* arXiv preprint arXiv:1507.04019, 2015.
27. Saritha, B., et al., *CACRN-Net: A 3D log Mel spectrogram based channel attention convolutional recurrent neural network for few-shot speaker identification.* Computers and Electrical Engineering, 2024. **115**: p. 109100.

التعرف على قراء القرآن من خلال تحليل الصوت والتعلم العميق

د. مازن مُحَّد عاشور الكثيري[1]، عبدالقادر مراد عبدالقادر باسلامه[2]

الملخص

في السنوات الأخيرة، شهد مجال الذكاء الاصطناعي (AI) تقدمًا هائلًا، لا سيما في تقنيات التعلم العميق التي حققت نتائج مذهلة في مختلف المجالات، بما في ذلك معالجة الصور والتعرف على الكلام ومعالجة النصوص والرؤية الحاسوبية. يهدف هذا البحث إلى الاستفادة من إمكانيات التعلم العميق لتطوير نظام تلقائي للتعرف على قرّاء القرآن الكريم استنادًا إلى أصواتهم الفريدة، وذلك باستخدام الشبكات العصبية الالتفافية العميقة (CNNs) و Log Mel Spectrograms كمستخرج لميزات الصوت. تقدم دراستنا مجموعة من النماذج المصممة لتصنيف قرّاء القرآن بناءً على خصائصهم الصوتية الفريدة. ولتعزيز إمكانية الوصول للمستخدمين، قمنا بتطوير تطبيق أندرويد قادر على تشغيل هذه النماذج دون الحاجة إلى الاتصال بالإنترنت. حقق نموذجنا دقة بلغت 98% عند اختبار الأصوات النقية لـ 23 قارئًا. ومع ذلك، اكتشفنا أن إضافة الضوضاء الخلفية إلى الأصوات النقية، كما اقترحت دراسات سابقة، لم يكن كافيًا لتمثيل التسجيلات الحقيقية بدقة. نظرًا للوقت الطويل الذي تتطلبه عمليات التسجيل اليدوية، نوصي بتطوير محاكاة أكثر تقدمًا للضوضاء الصوتية، بحيث تأخذ في الاعتبار التشويشات الشائعة التي تحدث في التسجيلات عبر الأجهزة المحمولة. يمكن أن يشمل ذلك توليد ضوضاء الخلفية النموذجية، والعيوب الناتجة عن التسجيل، وتأثيرات الصدى الصوتي، والظواهر الصوتية الأخرى في العالم الحقيقي، مما يساهم في إنشاء بيانات تدريب أكثر واقعية للنماذج القائمة على التعلم العميق.

الكلمات المفتاحية: الذكاء الاصطناعي، التعلم العميق، الشبكات العصبية التلافيفية (أو: الشبكات العصبية الالتفافية)، تلاوة القرآن (أو: ترتيل القرآن)، التعرف على الصوت (أو: تمييز الصوت)، مخططات ميل الطيفية (أو: أطياف ميل)، استخراج خصائص الصوت، تطبيق أندرويد، التأثيرات الصوتية، تطبيقات الذكاء الاصطناعي القائمة على الصوت.

[1] قسم تقنية المعلومات، كلية الحاسبات، جامعة سيئون.

[2] باحث بقسم علوم الحاسوب، كلية الحاسبات، جامعة سيئون.