MSA Engineering Journal
MSA Engineering Journal (ekb.eg)
ISSN 2812-5339 (print); ISSN 2812-4928 (online)
E-mail: engjournal@msa.edu.eg

Convolutional Neural Networks for speech recognition case study: Recitation Rules of the holy Quran

Dahlia Omran*[†], Ahmed Kandil[†], Sherif Samy[†], Ahmed ElBialy[†], Sahar Fawzi[††]
[†]Systems and Biomedical Engineering, Faculty of Engineering, Cairo university, Giza, Egypt,
domran@msa.edu.egh, ahkandil@eng1.cu.edu.eg, Abialy_86@yahoo.com, shsamy@eng1.cu.edu.eg
[††]Information Technology & Computer Science, Nile university, Giza, Egypt
sfawzi@nu.edu.eg

Abstract

This work focused on applying the Convolutional Neural Networks (CNN) to recognize one recitation rule of the Holy Quran, the Qalqala recitation rule which is applied to letters (Ba/Dal/Jem/Qaf/Ta) of the Arabic Alphabet and it implies vibration of these letters when there is absence of vowels on them with sukun. The feature extraction technique used in the suggested system was the Mel Frequency Cepstral Coefficients (MFCC) its output was pre-processed then fed to the CNN as input to start the recognition process. Recognition process consists of two stages, the first stage was assigned with letter identification and it achieved 92% accuracy, the second stage was responsible of recognizing whether or not the identified letter is in Qalqalah status, it scored 99% for Baa, 93% for Daal, 95% for Jeem, 92% for Qaaf and 83% for Taa. The above mentioned Alphabet with sukun were used as the main dataset and they were annotated out of continuous audio signals for professional reader Ayman Swayed, each sample represent one of the Qalqalah letters with length of 300ms.

Keywords: Speech Recognition, Convolutional Neural Networks CNN, Tajweed Rules, Qalqalah, Mel Frequency Cepstral Coefficients MFCC.

1 INTRODUCTION

Human communication primarily relies on speech. However, for humans to communicate with machines, speech recognition technology is necessary. It is the process by which machines automatically recognize and interpret human spoken words, orders, and feelings, allowing them to respond accordingly [1-3]. Speech recognition has diverse significant applications such as in security, healthcare, handicapped devices, and daily activities like voice search, hands-free writing, voice commands for smart houses and cars, and has changed the way people interact with devices and home appliance [4].

The foundation of speech recognition is the audio signal of the spoken words or utterance, which consists of compounds called phonemes, with each phoneme wave holding three types of information: words, speaker identity, and emotional state of the speaker [5]. Speech recognition is a challenging task as these waves carry a great amount of diversity, such as speaker accent or style, utterance length, and environmental noises.

One promising and efficient approach to speech recognition is Machine Learning, particularly Deep Neural Networks (DNN) [6]. DNNs build a hierarchy of features to define high-level features in terms of lower-level features, with more hidden layers than regular neural networks [7]. DNNs predict the output with high accuracy based on learned old input data features, making it a powerful tool for speech recognition [8]. They come with different architectures, such as transform networks, Recurrent Neural Networks (RNN), and Convolutional Neural Networks (CNN) [9]. CNNs show excellent performance in recognizing similar patterns and predicting unknown data fields [10]. With pooling, weight sharing, local connections, and the use of several layers, CNNs can detect the same pattern in different locations, resulting in great reduction in the model's memory requirements without affecting the runtime [11]. By representing audio signals as spectrograms, CNNs can also be applied to audio processing, leading to remarkable performance in Automatic Speech Recognition [12]. CNNs improve the speaker invariance of the acoustic model and handle individual variations in the speech signal, leading to better performance and stability against noise[13].

In the context of the holy Quran, understanding the correct meaning of each word is essential, as it is considered the main guidance for all Muslims. The holy Quran was revealed in Arabic, and conveying the exact meaning of each word is essential, which requires following Tajweed rules [8]. Tajweed rules involve specific pronunciation, intonation, and articulation for letters in certain situations, such as when to stop or continue, when to merge two letters or not, when to strengthen or lighten a letter, and when to stretch letters and for how long [8]. Tajweed rules can be complicated and confusing, especially for those who are not familiar with Tajweed or who are non-Arabic speakers [14]. Therefore, it is crucial to maintain the integrity and authenticity of the Quran by ensuring that the correct Tajweed rules are followed [15].

2 Literature Review

Previous research on Tajweed has been carried out, but it is still considered to be within a small group of researchers.

Algrami et al. [8] conducted a study on Tajwed rules, specifically Edgham Mem, Ekhfa Mem, takhfeef Laam, and Tarqeek Laam. They collected a dataset containing approximately 80 audio recordings for each rule from universities, expert volunteers, and paid experts, with a total of 657 recordings for the four rules. The dataset included the correct pronunciation and different incorrect ways of pronouncing the rules. The researchers manually segmented all the audio files to contain only the part in which the rule is pronounced, with an average length of approximately 4 seconds per file. They used 70 filter banks as a baseline method for feature extraction and Support Vector Machine (SVM) for classification.

Hassan et al. [16] investigated the process of identifying correct pronunciations of Qalqalah recitation using Mel-frequency Cepstrum Coefficients (MFCC) for feature extraction and Multilayer Perceptron (MLP) for classification. They used one hundred samples, half of which were pronounced correctly and the other half were pronounced incorrectly, and acquired a threshold from Receiver Operating Characteristics (ROC) to compare the target and output. The results showed that the MLP was successful in distinguishing correct and incorrect pronunciations with high accuracy.

Arshad et al. [17] designed a filter to eliminate noise in Quranic recitation performed in a noisy setting. They used an adaptive filter based on Least Mean Square (LMS) and focused on seven letters out of the total of thirty. The filter effectively eliminated the noise. They also used a similar technique to create another filter to remove noise in a noisy environment where the pronunciation of Quranic letters was being carried out, and the filter worked effectively in eliminating the noise.

Damer et al. [18] focused on eight Tajwed recitation rules and tested multiple features extraction methods, including Mel-frequency Cepstral Coefficients (MFCC), Linear Predictive Code (LPC), and Wavelet Packet Decomposition (WPD), as well as various classification techniques, including k-Nearest Neighbors (KNN), Support Vector Machines (SVM), and Multilayer Perceptron (MLP) neural network. They found that MFCC was the most effective feature extraction method, while SVM was the best classification technique for identifying the correct and incorrect pronunciations of the Tajweed rules.

Hassan et al. [19] developed a system for recognizing Tajweed Al-Quran using sphinx tools, which are tools available for researchers to develop acoustic models. They used MFCC for feature extraction and Hidden Markov Models (HMM) for classification, achieving an accuracy of 85-90% when tested offline on a small chapter of Quran. They also created an automated delimiter that can extract verses from audio files using Sphinx and speech recognition techniques, which can be helpful for users searching for specific verses in audio files, but not for teaching the correct recitation of a specific verse.

Razak et al. [20] conducted a review of various techniques used for recognizing recitation of Arabic verses in the Quran. The authors compared the benefits and drawbacks of each approach and identified several methods for feature extraction from speech signals. However, the authors did not address the primary goal of recitation learning, and their suggested system may not be suitable for non-Arabic speakers.

Muhammad et al. [21] developed a system called E-Hafiz to help regular Quran readers learn how to recite it correctly. The system trains users on Tajweed rules based on expert readers and uses the Mel-Frequency Cepstral Coefficient (MFCC) technique to extract features from recorded voices of specific verses. The E-Hafiz dataset contains 10 experts with the first 5 Surahs of the Holy Quran in the database. However, the system does not provide specific feedback on which word was pronounced incorrectly or which rule was not followed correctly.

Ahsiah et al. [22] proposed a hybrid algorithm that combines the MFCC feature extraction technique with Vector Quantization (VQ) to improve the performance speed of a Tajweed Rule Checking Tool. The authors modified the conventional MFCC algorithm using VQ to reduce data while maintaining the same quality, thereby speeding up the computational time. The proposed hybrid MFCC-VQ algorithm was tested on recordings of the Qalqalah phoneme in Surah Al-Ikhlas and ten phonemes of all the Qalqalah letters from various parts of the Quran, recorded from 45 speakers. The results showed that the hybrid algorithm provided a significant improvement in real-time factor over the conventional MFCC.

Mahdy et al. [23] created the HAFSS system, a Computer Aided Pronunciation Learning (CAPL) system, to help non-native speakers learn Arabic pronunciation and later repurposed it for teaching the correct recitation of the holy Quran. The system includes a module for automatically generating pronunciation hypotheses and a phoneme duration classification algorithm to detect recitation errors related to phoneme durations. The system produces a confidence score along with its decision to reduce the impact of misleading system feedback on unpredictable speech inputs.

Performance evaluation using a dataset containing 6.6% wrong speech segments showed that the system correctly identified the error in 62.4% of pronunciation errors, requested a repeat for 22.4% of the errors, and falsely accepted 14.9% of total errors.

Objective

- Help people with less knowledge of Tajweed rules to recite the holy Quran correctly, especially when direct human aid is not available or affordable
- Aiding non- Arabic speakers learning the correct pronunciation of the holy Quran.
- Maintain the integrity and authenticity of the holy Quran.
- This work could be considered the seed of a complete fully automated Tajweed system covering all Tajweed rules that really can be a great help for all Muslims.

3 METHODOLOGY

This section clarifies the detailed plan that had been followed to reach the research objectives, which is the recognition of Qalqalah letters using deep learning technique as illustrated in Fig.1. Starting from Acquiring data and preprocessing it properly then preprocess the data to fit the training model requirements, going through the creation of the convolutional neural network model with a brief identification of all used parameters.

3. 1 Data Acquiring

The first step of work was establishing a representable dataset that can be used as the input of the CNN model.

Fig. 1. Work layout

In the beginning of acquiring the needed samples many obstacles had been raised, gone through a large number of trials and errors to reach the suitable limits of all research aspects such as: characteristics of the samples, the sufficient number of samples, the selection of the reference reciters and the adequate surah’s for each specific Qalqalah letter. Trials like testing the suitable utterance length that can carry the core of the syllable without extra unneeded information, testing the model with number of samples starting from 500 utterances and observing the results researchers kept increasing the number of samples gradually till reaching satisfying results and finally samples were acquired from all surah of the holy Quran.

3.1.1 Data collection

1- Audio version of the holy Quran was the main source of all used utterances .
2- Utterances were acquired from a professional certified reciter, sheihk Ayman swayed, all of collected samples are according to the narration of Hafs on the authority of Asim.

3- Utterances were collected from different surah of the Holy Quran. Table1 shows a sample of the number of utterances used from different surah for each Qalqalah sound.

Table 1.sample of number of utterances used from different surah of holy Quran for one reader

Sound	**Baa**	**Dal**	**Jeem**	**Qaf**	**Taa**
Surah	**Number of Utterances**				
2 Al-Bakara	42	43	40	42	11
3 AlyEmran	66	50	21	15	4
6 Al-Anaam	31	42	30	13	6
9 Atawba	29	13	10	15	1
50 Qaaf	34	20	4	5	2

Mus-haf Tajweed was used to spot out the words that contain one or more of the Qalqalah letters (Qaaf-Taa-Baa-Jeem-Daal). Mus-haf Tajweed is a written version of the holy Quran that contains all Tajweed rules, each of them is assigned to a certain color, it is a color coded Quran regarding recitation rules.

4- Sample's name consists of the name of the reciter followed by the number of the surah which the sample was taken from then the number of the verse and finally the Qalqalah letter name, for an example AlHussary-2-68-Baa.

3.1.2 Samples characteristics

1- Qalqalah letters samples were eliminated manually using audio editing program (oceanaudio 8.3).

2- Each Sample was set to be 300ms in length which was found after many trails to be the most adequate for collecting all of the letter characteristics and articulation. As less or more length of utterances the core information of the syllable was missed and the CNN model was not able to recognize it properly.

3- For the sample's quality to be improved and though enhance the model output, 100ms were added to the sample's beginning and ending so the main information of the sample is collected, centralized and protected during any process in the training stage.

4- Samples were originally aggregated in stereo audio form.

5- For each Qalqalah letter samples were gathered in two states the vowel and non-vowel states , this makes 10 classes in total

6- Each sample name consists of the reciter's name, the surah's number and the verse's number, followed by the name of the Qalqalah letter for an example: Ayman-2-4-Baa.

7- All samples have the same length 500ms, sampling rate 22050 Hz and single channels to preserve consistency that in essential parameter in the feature extraction step [24].

3. 2 Data Pre-processing:

This part of work involved arranging the dataset samples in a manner that facilitates handling them by the model, they must be preprocessed to be in an eligible format of the model

1. All Qalqalah letters' folders gathered in one file called Data.
2. Create dataset of these samples in the form of Comma-Separated Values (CSV) files which are type of files used to move data between programs that are not ordinarily able to exchange data files.
3. Each audio segment was converted from .mp3 form to .wav form.
4. All audio segments were converted from stereo to mono by averaging the reading of the two stereo channels and save it as single mono channel.
5. High pass Filter type alpha was used to reduce the noise and to flatten the speech signal spectrum [25].
6. Window hamming was used in the widowing stage to smooth the signal from each end side
7. Lifter of the sin type was used to emphasis central values.

3. 3 Training

This part involves acquiring the needed inputs for the speech recognition model and reshape it to suit the model requirements.

3.3.1 Feature extraction

For Feature extraction Mel-Frequency Cepstral Coefficients (MFCC) were used, with sampling rate of 44100 and with the first 12 coefficients to be gained.
The MFCC model used the first 12 coefficients of the signal after applying the Inverse Discrete Fourier transform (IDFT) operations. Along with the 12 coefficients, the energy of the signal sample will be taken as a feature that will help in identifying each signal. Along with these 13 features, the MFCC technique To increase the feature evidence of dynamic coefficients, delta and delta-delta can be devoted by adding the first order derivatives and second order derivative approximation to feature parameters which constitute another 26 features[2].
So overall MFCC technique will generate 39 features from each audio signal sample which are used as one of the inputs for the speech recognition model [26-27].

3.3.2 Data reshaping

1. Rearrange the resulting features to be listed in the form of the 39 features and a class label as there are 6 classes which are Qaaf-Taa-Baa-Jeem-Daal-No.
2. Encode class label using Label Encoder, as all information are needed to be represented in the form of ones and zeros in order to be understood by the neural network model.
3. Split the total available features into training set and testing set (80:20). The Train-test process of dividing the dataset into two subsets. The first is the training subset and it is

used to fit the model, the second is the test subset and it is used to estimate the performance of the model with totally new data. In order to reach the optimal train-test split percentage the training and computational cost and complexity need to be considered [28-29].

4. Reshape the input extracted features to fit the speech recognition model (histogram images) to be in the form of rows, columns and channels.

5. Set the random state value, the random state is a parameter that is used to initialize the internal random number generator, which will decide the splitting of data into train and test indices. It could be set to 0 or 1 or any other integer. If validation of the processing is needed over multiple runs, the same value must be used. This will guarantee that the same sequence of random numbers is generated each time the code is ran [30].

3.3.3 The Convolutional Neural Network (CNN)

The CNN audio recognition model comprises four key layers, each of which is composed of a convolutional layer containing two operations: a convolution operation (linear phase) and an activation operation (non-linear phase). The convolutional layer comes after it a pooling layer, and the final layer in the architecture is a flatten layer that uses a global average function . A dense layer, also known as a fully connected layer, is added at the end [7].
Reader can find the clarified Architecture of the used Convolutional Neural Network in Fig. 2.

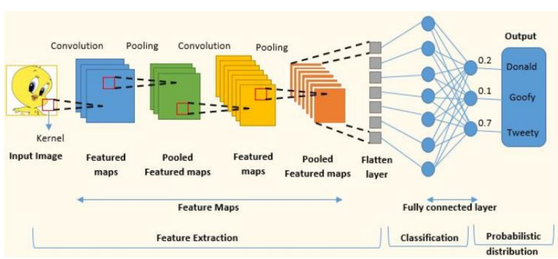

Fig. 2. Architecture of the used CNN model [31]

3.3.3.1 Convolutional layer

The CNN architecture shows the highest performance by manipulating two-dimensional data formats such as images, video that is way when working in audio signals those signals must be reshaped to take the 2D form [32], the first 2D convolutional layer uses 16 filters, the second layer uses 32 filter, the third layer uses 64 filters and the last layer (fourth) uses 128 filters, all filters used in all of the four convolutional layers are of size2. The used activation function in all of the four layers is the Rectified Linear Unit (ReLU) which adds non-linearity in the network and provides non-saturation of gradient for positive inputs [7].
The convolutional layer is a fundamental component of the CNN architecture that performs a stack of mathematical operations (linear) which implies filters (kernals) [32] as it scans the input signal which is an array numbers (tensor) to create a feature map that enables it to detect the presence or absence of a specific feature [12].

3.3.3.2 Pooling layer

The pooling process involves a 2D filter that slides over each feature map in the aim of decrease its dimension, down sample each feature map independently, shortens the training time and respectively reduces the number of learnable parameters while keeping the core information. In the proposed model all pooling layers use the Max-pooling method with filter size 2. The Max-pooling is the most popular form of pooling, it involves taking the maximum value of each patch and discarding all other values [11].

3.3.3.3 Dropout

After each pooling layer a Dropout step is executed, the dropout is the process of ignoring (dropping-out) some layer's outputs in the training phase. This action has the effect of making the training phase noisy and so, with each update this layer will have different number of nodes and sub-sequentially will have different connectivity to the prior layer, Also dropout helps avoiding the dead nodes phenomenon as almost all nodes are activated which is a very powerful advantage [12]. There is no absolute best dropout value, instead there is be value for each model. We can find that by testing values between 0.1 and 1 in increments of 0.1 for our model. The Value of The Dropout is set to be 0.2 in all four layers, the main effect of dropout is to prevent the Neural Network from over fitting as the output of any layer under dropout is randomly subsampled [33].

3.3.3.4 Flatten layer

A flatten layer is constructed Before the final layer of the model architecture. This layer significantly reduces the dimension of the feature map to a one-dimensional array by utilizing global average pooling, which involves computing the average of all the elements in each feature map [7].

3.3.3.5 Dense layer

The dense layer, which is the final layer of the model, is a fully connected layer where each input is connected to every output. It employs a soft-max activation function, which converts the non-normalized output of the network into a probability distribution over the predicted output classes [12].

The proposed model consists of two stages, the first stage is assigned to differentiate between the five letters of Qalqalah while the second stage is concerned with identifying if that letter is pronounced with or without Qalqalah.

3 RESULTS

In this section researcher is reviewing the classification report of the two stages' model, the first stage of the model works mainly to find out the characteristics of the Qalqalah letters and to be able to identify each one of them. Table 2 shows the classification report of this stage of the model metrics which are precision, recall and F1-score .

The precision represents the ratio of true positive tp to the sum of true positive tp and false positive fp, precision =tp / (tp + fp) . The precision is intuitively the ability of the classifier not to label a negative sample as positive.

The recall is the ratio tp / (tp + fn) where fn is the number of false negatives. The recall can be identified as the ability of the classifier to find all the positive samples.
The F-beta score can be interpreted as a weighted harmonic mean of the precision and recall, where an F-beta score reaches its best value at 1 and worst score at 0. The F-beta score weights recall more than precision by a factor of beta. Beta == 1.0 means recall and precision are equally important [34].

Table 2. Classification report of stage one of the model

Letter	Precision	Recall	F1-Score
Baa	0.81	0.95	0.87
Daal	0.95	0.79	0.86
Jeem	1.00	0.98	0.99
Qaaf	0.95	0.95	0.95
Taa	0.93	0.95	0.94

The model accuracy is 92 %
The second stage works as mini-models each stands for Qalqalah recognition of one letter of the five letters of Qalqalah. Tables 3-7) shows the results of Qalqalah status recognition for letters (Baa - Daal – Jeem – Qaaf – Taa) respectively.

Table 3. Letter Baa Qalqalah recognition

Status	Precision	Recall	F1-Score
No	1.00	0.94	0.97
Yes	0.98	1.00	0.99

The accuracy of Baa mini-model is 99%

Table 4. Letter Daal Qalqalah recognition

Status	Precision	Recall	F1-Score
No	0.77	0.77	0.77
Yes	0.96	0.96	0.96

The accuracy of Daal mini-model is 93%

Table 5. Letter Jeem Qalqalah recognition

Status	Precision	Recall	F1-Score
No	1.00	0.87	0.93
Yes	0.93	1.00	0.96

The accuracy of Jeem mini-model is 95%

Table 6. Letter Qaaf Qalqalah recognition

Status	Precision	Recall	F1-Score
No	0.91	0.87	0.89
Yes	0.93	1.00	0.96

The accuracy of Qaaf mini-model is 92%

Table 7. Letter Taa Qalqalah recognition

Status	Precision	Recall	F1-Score
No	0.78	0.99	0.84

Yes	0.89	0.77	0.83

The accuracy of Taa mini-model is 83%

5 CONCLUSION

In this paper a two stages Convolution Neural Network architecture model was suggested and executed to detect specific letters of the Arabic alphabet and then to verify if they were pronounced correctly according to one of the holy Quran Recitation rule which is the Qalqalah rule. The suggested model was able to recognize the five letters with 92% accuracy in the first stage , while the second stage which was responsible of recognizing whether or not the identified letter is in Qalqalah status, scored 99% for Baa, 93% for Daal, 95% for Jeem, 92% for Qaaf and 83% for Taa. These promising results showed the strength of the CNN technique in speech recognition field concerning the Arabic and Quranic recitation rules.

6 REFERENCES

1. V.Vineet, K.P. Aditya and P.Y. Satya, "Speech Recognition using Machine Learning .IEIE Transactions on Smart Processing and Computing" vol. 10, no. 3, June 2021.
2. V.Z. Këpuska and H.A. Elharati. "Robust Speech Recognition System Using Conventional and Hybrid Features of MFCC,LPCC, PLP, RASTA-PLP and Hidden Markov Model Classifier in Noisy Conditions" Journal of Computer and Communications, 1-9,2015.
3. D.Nagajyothi, P.Siddaiah, "Speech Recognition Using Convolutional Neural Networks .International Journal of Engineering & Technology",7 (4.6),133-137,2018.
4. J.M. Bhadragiri, N. Ramesh, "Speech Recognition using MFCC and DTW" ,the 1st International Conference on Advances in Electrical Engineering, Vellore Institute of Technology, India, January 2014.
5. A. Ayad et al, "Speech Recognition using Convolution Deep Neural Networks", Journal of Physics: Conference Series, 1973, 012166,2021.
6. H.Geoffrey et al,"Deep Neural Networks for Acousting Modeling in Speech Recognition", IEEE Signal Processing Magazine, pp 82-97, 2012.
7. R. Yamashita, M. Nishio, D.R. Kinh Gian and K. Togashi, "Convolutional neural networks: an overview and application in radiology. Insights into Imaging", 9:611–629, 2018.
8. Alagrami A, Eljazzar M (2020)SMARTAJWEED Automatic Recognition Of Arabic Quranic Recitation Rules. the International Conference on Computer Science, Engineering And Applications, 2020.
9. Papastratison I (2021) Speech Recognition: a review of the different deep learning approaches. https://theaisummer.com/speech-recognition/ , 2021-07.
10. Bhatia S, Devi A, Alsuwailem RI ,Mashat A (2022) Convolutional Neural Network Based Real Time Arabic Speech Recognition to Arabic Braille for Hearing and Visually Impaired. Frontiers in Public Health 10:898355. doi: 10.3389/fpubh.2022.898355

11. Heaton j (2018) Ian Goodfellow, Bengio Y, Courville A: Deep learning: The MIT Press, 2016, 800 pp, ISBN: 0262035618 .In Genetic Programming and Evolvable Machines19:305–307
12. Rady E,Hassen A, Hassan NM, Hesham M (2021) Convolutional Neural Network for Arabic Speech Recognition .Egyptian Journal of Language Engineering, Vol. 8, No.1
13. Abdel-Hamid O, Mohamed A, Jiang H, Deng L, Penn G, Yu D (2014) Convolutional Neural Networks for Speech Recognition. IEEE/ACM Transactions On Audio, Speech, And Language Processing, Vol. 22, No. 10
14. Shafaf Ibrahim, Farah Afiqah Abdul Rahim, Zaaba Ahmad (2019) Automatic Tajweed Rules Recognition using k-Nearest Neighbour (k-NN). International Journal of Recent Technology and Engineering (IJRTE)ISSN: 2277-3878, Volume-8, Issue-2S11
15. Ammar Mohammeda*, Mohd Shahrizal Sunara, Md. Sah Hj Salamb(2015) Quranic Verses Verification using Speech Recognition Techniques. Jurnal Teknologi (Sciences & Engineering) 73:2, 99–106
16. Hassan H, Nasrudin N H,Khalid M N, Zabidi A,Yassin A I(2012) Pattern Classification In Recognizing Qalqalah Kubra Pronuncation Using Multilayer Perceptrons. International Symposium on Computer Applications and Industrial Electronics 3-4, Kota Kinabalu Malaysia
17. Arshad N.W., S. N. A. Aziz, F. Naim, R. A. Karim, R. Hamid and N. F. Zakaria (2011) Speech processing for makhraj recognition: The design of adaptive filter for noise canceller. 7th International Conference on Information Technology in Asia, pp. 1-5, doi: 10.1109/CITA.2011.5999501.
18. Damer N, Al-Ayyoub M, Hmeidi I (2017) Automatically Determining Correct Application of Basic Quranic Recitation Rules. In: The International Arab Conference on Information Technology Yassmine Hammamet, Tunisia, 22-24 December 2017.
19. Tabbal H., W. El Falou and B. Monla (2006) Analysis and implementation of a "Quranic" verses delimitation system in audio files using speech recognition techniques .2nd International Conference on Information & Communication Technologies,pp. 2979-2984, doi: 10.1109/ICTTA.2006.1684889.
20. Razak, Z.B., Ibrahim, N.J., Idris, M.Y., Tamil, E.M., Yakub, M., Yusoff, Z.B., Rahman, N.N., & Lumpur., K. (2008). Quranic Verse Recitation Recognition Module for Support in j-QAF Learning: A Review.
21. Muhammad A,ul Qayyum Z, Mirza W M,Tanveer S, Martinez-Enriquez A M, Syed A (2012) E-Hafiz: Intelligent System to Help Muslims in Recitation and Memorization of Quran. Life Science Journal, 2012;9(1)
22. Ahsiah Ismail 1, Mohd Yamani Idna Idris 2, Noorzaily Mohamed Noor 3, Zaidi Razak4, Zulkiflimohd Yusoff(2014) MFCC-VQ Approachfor Qalqalah Tajweed Rule Checking .Malaysian Journal Of Computer Science. Vol. 27(4), 2014
23. Mahdy Sh ,Hamid S , Rashwan M, Samir A, Abd-Elhamid O, Shahin M, Nazih W (2006) Computer Aided Pronunciation Learning System Using Speech Recognition Techniques. International Conference on Speech and Language Processing INTERSPEECH 2006 - ICSLPSeptember 17-21, Pittsburgh, Pennsylvania.
24. Doshi K(2021) Audio Deep Learning Made Simple (Part 1): State-of-the-Art Techniques. https://ketanhdoshi.github.io/Audio-Intro/
25. Awaid M.,Fawzy S.,Kandil A.(2014) Audio Search Based on Keyword Spotting in Arabic Language

26. Mohammad K.,Fawzy S.,Kandil A,El-Bialy A.(20xx) Automated Detection And Consistency Analysis of Tajweed Recitation Rules In The Holy Quran
27. Kiran U (2021) MFCC Technique for Speech Recognition .part of the Data Science Blogathon
https://www.analyticsvidhya.com/blog/2021/06/mfcc-technique-for-speech-recognition
28. Brownlee J (2020) Train-Test Split for Evaluating Machine Learning Algorithms. Machine Learning Mastery in Python Machine Learning
https://machinelearningmastery.com/train-test-split-for-evaluating-machine-learning-algorithms/
29. Birba D E (2020) A Comparative study of data splitting algorithms for machine learning model selection. Dissertation, KTH Royal Institute of Technology computer science and engineering department Stockholm, Sweden
30. Aruchamy V (2021) How To Do Train Test Split Using Sklearn In Python – Definitive Guide. Stack Vidhya, Machine Learning-python
https://www.stackvidhya.com/train-test-split-using-sklearn-in-python/
31. Shah S. (2022)Convolutional Neural Network: An Overview
https://www.analyticsvidhya.com/blog/2022/01/convolutional-neural-network-an-overview/
This article was published as a part of the Data Science Blogathon
32. Taher K I ,Abdulazeez A M(2021) Deep Learning Convolutional Neural Network for Speech Recognition: A Review .International Journal of Science and Business Volume: 5, Issue: 3 Page: 1-14 2021
33. Brownlee J (2022) Dropout Regularization in Deep Learning Models with Keras. Machine Learning Mastery in Deep Learning
https://machinelearningmastery.com/dropout-regularization-deep-learning-models-keras/
34. https://scikitlearn.org/stable/modules/generated/sklearn.metrics.precision_recall_fscore_support.html#sklearn.metrics.precision_recall_fscore_support