Kinetik: Game Technology, Information System, Computer Network, Computing, Electronics, and Control
Journal homepage: http://kinetik.umm.ac.id
ISSN: 2503-2267
Vol. 9, No. 2, May 2024, Pp.181-192

Enhancing Qur'anic recitation experience with CNN and MFCC features for emotion identification

Lailis Syafa'ah*[1], Roby Prasetyono [1], Hariyady Hariyady[1]
Universitas Muhammadiyah Malang, Indonesia[1]

Article Info

Keywords:
Emotion identification, Qur'an murottal sound, MelFrequency Cepstral Coefficients (MFCC), Convolutional Neural Network (CNN)

Article history:
Received: March 22, 2024
Accepted: May 10, 2024
Published: May 31, 2024

Cite:
L. Syafa'ah, R. Prasetyono, and H. Hariyady, "Enhancing Qur'anic Recitation Experience with CNN and MFCC Features for Emotion Identification", KINETIK, vol. 9, no. 2, May 2024. Retrieved from https://kinetik.umm.ac.id/index.php/kinetik/article/view/2007

*Corresponding author.
Lailis Syafa'ah
E-mail address:
lailis@umm.ac.id

Abstract

In this study, MFCC feature extraction and CNN algorithms are used to examine the identification of emotions in the murottal sounds of the Qur'an. A CNN model with labelled emotions is trained and tested, as well as data collection of Qur'anic murottal voices from a variety of readers using MFCC feature extraction to capture acoustic properties. The outcomes show that MFCC and CNN work together to significantly improve emotion identification. The CNN model attains an accuracy rate of 56 percent with the Adam optimizer (batch size 8) and a minimum of 45 percent with the RMSprop optimizer (batch size 16). Notably, accuracy is improved by using fewer emotional parameters, and the Adam optimizer is stable across a range of batch sizes. With its insightful analysis of emotional expression and user-specific recommendations, this work advances the field of emotion identification technology in the context of multitonal music.

1. Introduction

In the current dynamic global landscape, the role of precise emotion recognition has become increasingly significant. Emotion is of utmost importance in different fields, such as disease detection, human-computer interface, and virtual reality [1]. Scientists have carried out multiple investigations on the identification of emotions based on human facial expressions. These research have emphasized the notable influence of changes in emotions on the characteristics of an individual's voice [2]. The process of identifying emotions from speech signals has become difficult due to the intricate and fluctuating nature of human vocalizations. To address this challenge, this study focuses on enhancing emotion identification in Qur'anic murottal sounds using a combination of CNN and MFCC features.

Emotions are important in the context of human interaction [3] [4], affecting how we perceive and respond to various stimuli, including spoken words and vocal expressions. Emotions are an essential part of human beings and play an important role in people's daily lives, including communication, interactions, learning, and so on [5][6][7]. The key to communication between humans is a reaction to emotion [3][4][8]. Emotions are feelings that describe the way we act in a particular case. [9]. Emotion detection enables interactions between people and smart technology and is critical in human-computer interaction (HCI) [10][11]. The way to control and regulate behavior is by studying the emotions of people [12][13]. To better meet human needs, an intelligent system capable of analysing emotional responses may best explain humans.[14]. Human emotion analysis is a critical component of the endowing of artifacts with humanized characteristics, which attracted increasing attention regarding the beneficial uses in interpersonal contact [15][16]. Emotions are extremely important in everyday life, whether it is engagement, interpretation, or decision-making. Emotions play a significant role in the behavior, cognition, and communication of humans [10][17]. The result is more objective when emotions are recognised on the basis of physiological signals [18][19].

In religious practices, such as reciting the Qur'an, it is crucial to comprehend the emotions expressed through sound [20][21]. It is an obligation for every Muslim to read the Holy Quran [22][23]. Reading the Holy Quran is a specific method of reading, where the readers are required to follow certain rules in order to read it. It is known as tajweed, the reading mechanism of the holy Quran [24][25]. The emotional Qur'an recitation, known as Murottal, holds profound significance as it connects with the hearts and minds of the listeners, invoking a spiritual and transformative experience.

Traditionally, the identification and interpretation of emotions in Qur'anic Murottal sounds have relied on human perception and subjective analysis [21][26]. However, the field of emotional technology is becoming more popular and the advancement of artificial intelligence techniques have paved the way for automated emotion recognition systems[27][28]. These systems employ machine learning algorithms and audio signal processing methods to extract relevant features and classify emotional states accurately.

In order to overcome these difficulties, this study suggests a method to improve the recognition of emotions in Qur'anic Murottal sounds by employing Convolutional Neural Networks (CNN) and MFCC characteristics. Convolutional neural networks (CNNs) have demonstrated their efficacy in capturing both spatial and temporal relationships in a range of tasks, such as speech and audio processing [29][30]. MFCC, on the other hand, is a widely used extracting features technique in audio filtering of signals, capturing the characteristics of speech and vocal expressions [31][32].

Our proposed methodology involves pre-processing the audio data to obtain MFCC features, which serve as input to the CNN model. The CNN architecture is trained using a dataset of labeled Murottal recordings encompassing various emotional states, including joy, sadness, fear, and tranquillity. By learning the patterns and representations within the MFCC features, the CNN model can accurately classify the emotional states conveyed in the Qur'an recitations.

The fundamental issue of this research arises from the constraints of prior studies, which predominantly depend on human perception and subjective analysis to identify emotions in Qur'anic Murottal sounds, instead of employing automated emotion recognition methods. Furthermore, there is a disparity between the suggested research and prior studies, mostly because of the restricted investigation of CNN and MFCC features for identifying emotions in Qur'anic Murottal sounds.

Although there have been improvements in emotion recognition approaches, there is a lack of study on using CNN and MFCC characteristics for identifying emotions in Qur'anic Murottal sounds [33] [34]. This research gap highlights the necessity for additional inquiry and advancement in this field. The successful completion of the essay hinges on clearly outlining the research aims, methodology, and conclusions of the study. The aim is to address the limitations and investigate the possibilities of automated emotion recognition systems in Qur'anic Murottal sounds.

This research aims to bridge this gap by proposing the use of CNN and MFCC features for emotion identification in Qur'anic Murottal sounds, with the goal of improving the accuracy and efficiency of emotion recognition in this specific context. The proposed research will contribute to the existing body of knowledge by exploring the effectiveness of CNN in combination with MFCC features for emotion identification in Qur'anic Murottal sounds. Additionally, the use of CNN and MFCC features can also provide valuable insights into the emotional aspects of Qur'anic recitations, enabling a deeper understanding of the emotional impact of these sounds on listeners.

2. Research Method

This study focuses on enhancing emotion identification in Qur'anic Murottal sounds using a combination of CNN and MFCC features. The use of CNN in emotion recognition has shown promising results, particularly in image-based emotion analysis. With the popularity of pictures and short videos, researchers have turned their attention to emotion recognition based on images and videos [35]. However, the application of CNN to analyze and identify emotions in audio signals, such as speech, has not been extensively explored. Recently, researchers have started to explore the use of CNN combined with MFCC features in speech and audio signal processing [36]. The MFCC is a widely used feature in speech signal processing due to its consistency and effectiveness, even for low-dimensional data [37]. To implement the proposed approach, the following steps are taken. First, the Qur'anic Murottal sounds are preprocessed to extract MFCC features. The extraction of MFCC features involves converting the audio signals into frequency-domain representations using the Fast Fourier Transform. Then, a Mel filterbank is applied to the power spectrum of the signal to obtain the Mel-scaled spectrogram. Next, the logarithm of the Mel-scaled spectrogram is taken, followed by applying the Discrete Cosine Transform to obtain the final MFCC features. The extracted MFCC features are then used as inputs to a CNN.

This model consists of multiple layers of convolutional and pooling operations, which effectively capture the local and global features in the MFCC representations. In addition, the model includes fully connected layers and softmax activation to perform emotion classification based on the extracted features. The use of CNN in conjunction with Mel Frequency Cepstral Coefficients features has proven to be effective in various speech and audio processing tasks, such as speech recognition, heart sound recognition, semantic analysis, emotion analysis, and keyword detection [36]. In the context of Qur'anic Murottal sounds, the application of CNN combined with Mel Frequency Cepstral Coefficients features can greatly enhance the identification of emotions. The combination of CNN and Mel Frequency Cepstral Coefficients features has shown promising results in automatic emotion recognition, as it leverages the ability of CNN to capture both local and global features in the audio data. Furthermore, the use of Mel Frequency Cepstral Coefficients features is particularly beneficial in emotion recognition tasks, as they have been shown to effectively capture the perceptual characteristics of sound, including the frequency content and spectral shape. By extracting MFCC features from Qur'anic Murottal sounds and feeding them into a CNN model, we can effectively capture the emotional characteristics of the recitations. In conclusion, the combination of CNN and Mel Frequency Cepstral Coefficients features presents a promising approach for enhancing emotion identification in Qur'anic Murottal sounds.

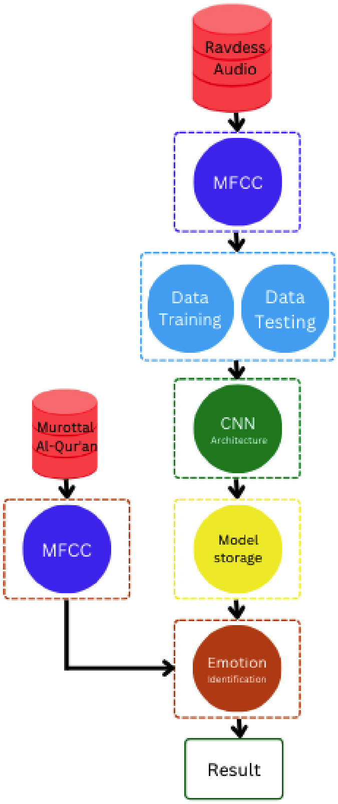


Figure 1. System Design Diagram

Figure 1 illustrates the research methodology used in this study, which entails designing and putting into use a system for identifying emotions using CNN and MFCC characteristics. The dataset used in this study was obtained from Kaggle and consists of.wav sound files from the RAVDESS dataset. Eight fundamental human emotions—calm, sad, angry, neutral, fear, surprise, and disgust—are represented by 1012 audio recordings in the RAVDESS dataset, which has a resolution of 16 bits and a frequency of 48 kHz. The development of the emotion identification model is based on this data.

1. Dataset Acquisition: The main dataset for this study is the RAVDESS dataset, which is accessible on Kaggle. As a supplemental dataset for testing the emotion recognition model specifically for Qur'anic recitations, recordings of Qur'anic murottal from the Yuk Ngaji Malang group are also gathered.
2. Dataset Pre-processing: The audio files for the RAVDESS dataset are first converted into a format appropriate for feature extraction and analysis. The audio files are made into mono-channel audio after being resampled to a 16 kHz frequency. The emotion labels that are part of each recording's metadata are retrieved and used for later model training and evaluation.
3. Feature Extraction: The pre-processed audio data is used to extract pertinent characteristics using the MFCC approach. This method gives a representative feature set for further analysis and simulates the frequency resolution of the human auditory system. The input representation for the CNN model is formed by the MFCC features, which capture the spectral characteristics of the audio signals.
4. CNN Model Development: Training, validation, and testing subsets of the RAVDESS dataset's feature-extracted data are created. The CNN model is trained and assessed using these subsets. To learn discriminative characteristics and properly categorise the emotional states, the CNN architecture incorporates numerous Convolution layers, accumulating, and completely linked layers. The training subset is used to train the model, and appropriate optimization procedures are then used to improve it.
5. Model Evaluation: The validation subset is used to test the trained CNN model's performance in identifying emotions. A number of evaluation metrics, including accuracy, precision, recall, and F1-score, are produced to assess how well the model categorizes emotions.
6. Emotion Identification on Qur'anic Murottal Data: An essential aspect of this work involves elucidating the process of identifying emotions through the analysis of Qur'anic Murottal data. This stage involves utilizing the trained Convolutional Neural Network (CNN) model on the Yuk Ngaji Malang Qur'anic Murottal dataset. The selection of suras Al-Ikhlas, Al-Falaq, and An-Nas is based on their relevancy and profound emotional impact. This phase will assess the CNN model's capacity to recognize emotional emotions found in the Qur'an. This is accomplished by

juxtaposing model predictions with human annotations. The approach involves examining the emotional content of the recitations, taking into account even subtle variations. To accurately detect emotions, it is essential to comprehend the context and emotional significance of Qur'anic Murottal recitations. This stage examines the CNN model's ability to understand and classify emotional reactions within the spiritual and cultural framework of the recitations. To assess the model's ability to recognize and comprehend emotions, we can compare its predictions with human annotations. This assessment demonstrates the model's ability to accurately represent the nuanced emotions present in Qur'anic recitations. Ultimately, the emotion identification portion of this study plays a crucial role in assessing the CNN model's proficiency in recognizing emotions in Qur'anic Murottal data. An in-depth elucidation of this approach will clarify the model's effectiveness and its applicability in religious audio analysis.

7. Result Analysis and Discussion: The outcomes of the technique used to identify emotions are carefully examined and interpreted. With particular emphasis on the possible contributions and consequences for Qur'anic recitation emotion analysis, the usefulness and limits of the proposed CNN and MFCC-based technique are examined.

A thorough foundation for recognizing emotions in both general audio recordings and, more especially, in Qur'anic murottal data is provided by the research methodology that has been offered. The creation of intelligent systems for assessing emotional nuance in audio signals is facilitated by the combination of CNN architecture with MFCC feature extraction, which enables accurate emotion recognition and categorization.

2.1 Labelling dataset

The dataset used in this investigation, shown in Figure 2, needs to be properly categorised before anything else. The correct emotion categories are applied to each Qur'anic Murottal sound recording in the dataset to create a label for it. Labels can be used to categorise a variety of emotions, including surprise, joy, and melancholy. The labels for each sound recording must accurately reflect the emotional expressions that may be heard in each one.

```
[ ] feeling_list=[]
    for item in mylist:
        if item[6:-16]=='02' and int(item[18:-4])%2==0:
            feeling_list.append('female_calm')
        elif item[6:-16]=='02' and int(item[18:-4])%2==1:
            feeling_list.append('male_calm')
        elif item[6:-16]=='03' and int(item[18:-4])%2==0:
            feeling_list.append('female_happy')
        elif item[6:-16]=='03' and int(item[18:-4])%2==1:
            feeling_list.append('male_happy')
        elif item[6:-16]=='04' and int(item[18:-4])%2==0:
            feeling_list.append('female_sad')
        elif item[6:-16]=='04' and int(item[18:-4])%2==1:
            feeling_list.append('male_sad')
        elif item[6:-16]=='05' and int(item[18:-4])%2==0:
            feeling_list.append('female_angry')
        elif item[6:-16]=='05' and int(item[18:-4])%2==1:
            feeling_list.append('male_angry')
        elif item[6:-16]=='06' and int(item[18:-4])%2==0:
            feeling_list.append('female_fearful')
        elif item[6:-16]=='06' and int(item[18:-4])%2==1:
            feeling_list.append('male_fearful')
        elif item[:1]=='a':
            feeling_list.append('male_angry')
        elif item[:1]=='f':
            feeling_list.append('male_fearful')
        elif item[:1]=='h':
            feeling_list.append('male_happy')
        #elif item[:1]=='n':
            #feeling_list.append('neutral')
        elif item[:2]=='sa':
            feeling_list.append('male_sad')
```

Figure 2. Emotion Labeling Program Listings

2.2 Feature Extraction

The MelFrequency Cepstral Coefficients (MFCC) method is used to extract features from each sound recording in the following step, as shown in Figure 3. The MFCC features have shown to be useful in characterising the acoustic properties of sound and can be applied to the analysis of vocal emotion. The selection of time frames, computation of the power spectrum, and cepstral transformation are all steps in the process of extracting these features.

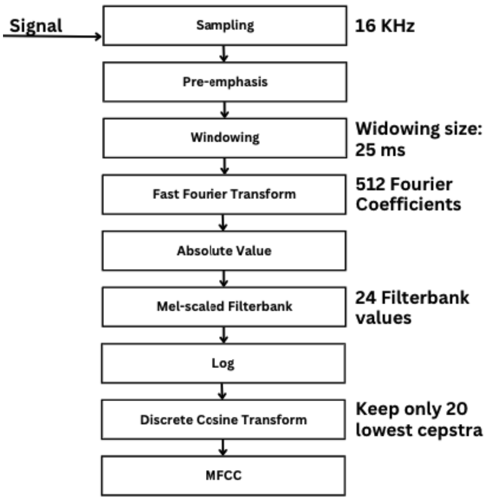


Figure 3. Diagram MFCC

2.3 Data Split

It is necessary to separate the labelled dataset and retrieved MFCC features into training sets and testing sets. This category is crucial for developing models and evaluating their effectiveness. A small portion of the data is typically utilised as a testing set, with the majority of the data serving as a training set.

2.4 Convolutional Neural Network

CNN approach that will be used at this time is applied to differentiate between emotions in the Murottal sounds of the Qur'an. Figure 4's illustration of a CNN neural network design demonstrates how well it may be used to infer spatial patterns from data, such as those observed in the frequency spectrum of sound. The CNN model will be trained using previously shared training information.

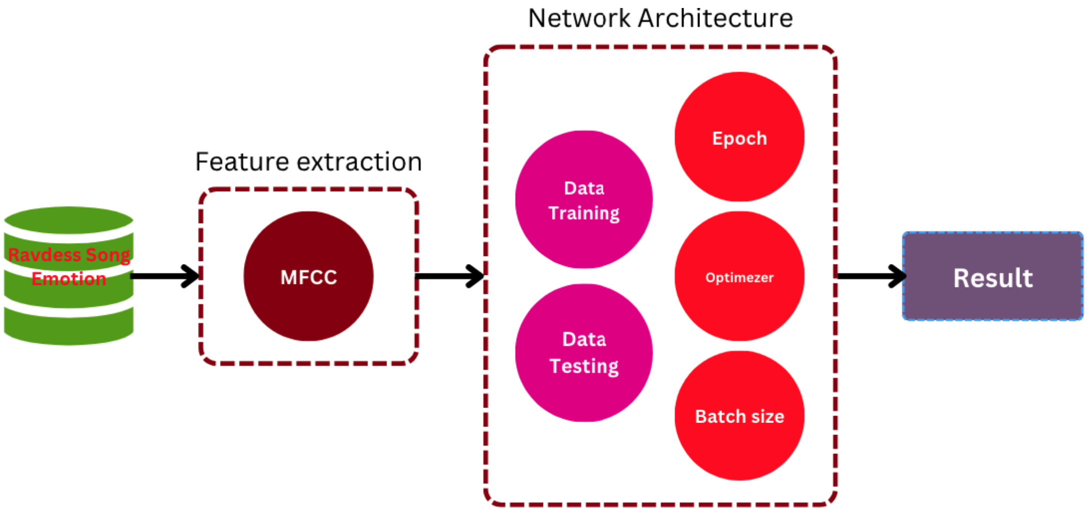


Figure 4. CNN Flowchart

2.5 Confusion Matrix

The effectiveness of the trained model at identifying emotions in Murottal sounds from the Qur'an will be evaluated using a confusion matrix, a common evaluation tool. The confusion matrix provides a comprehensive analysis of the model's capacity to classify emotions and differentiate between different emotional states. The confusion matrix will provide valuable insights into the model's ability to precisely classify emotions and identify potential development areas based on Table 1 - Confusion Matrix. This evaluation method will assist in identifying the model's strengths and weaknesses in identifying various emotional expressions, which will serve as a guide for future enhancements and optimizations of the emotion identification procedure.

Table 1. Matrix Confusion

		True value	True value
		TRUE	TRUE
Predicted Value	TRUE	TP (True Positive)	TP (True Positive)
	FALSE	FN (False Negative)	FN (False Negative)

Here's the description from the table:
TP (True Positive): The result obtained is correct and the actual value is correct.
TN (True Negative): The result is incorrect and the actual value is wrong.
FP (False Positive): The result obtained is correct, but the actual value is wrong.
FN (False Negative): The result is wrong, but the actual value is correct.

2.6 Model Storage

In the identification of Qur'anic emotions, it is essential to preserve the trained model for future use. Once the model has been trained using the training data and its performance has been evaluated using test data, it can be preserved in a manner that facilitates its use for identifying emotions in new Qur'anic Murottal sounds. The entire process is depicted in Figure 5. The Flow of Qur'anic Emotion Identification, including the crucial step of model storage. By preserving the trained model, researchers and practitioners can easily apply it to new audio recordings, enabling the efficient and accurate identification of emotions in Qur'anic recitations for a variety of applications and studies.

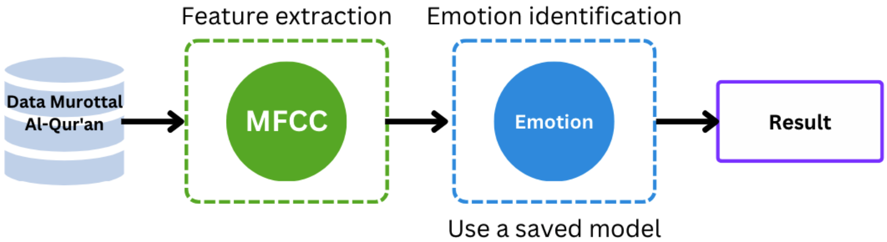


Figure 5. The Flow of Qur'anic Emotion Identification

2.7 Identification of Qur'anic Murottal Emotions

The new Qur'anic Murottal sound will be analysed at this point to identify emotions using the trained CNN model. The model will be given an unfamiliar Qur'anic Murottal sound as input, and based on the modelling completed during the training step, the model will categorise the emotions present in that sound.

The implementation of the aforementioned techniques make CNN an effective paradigm for recognising emotions in the Murottal sound of the Qur'an. These findings might be useful for identifying emotions, creating mobile apps, and researching the interplay between emotions and religion.

3. Results and Discussion

The technique utilized in this study offers a thorough framework for the recognition of emotions, with a specific emphasis on Qur'anic murottal data. In this discussion, we examine the fundamental elements of the methodology and their consequences for the identification of emotions in audio signals.

3.1 Dataset Acquisition

The RAVDESS dataset, along with recordings from the Yuk Ngaji Malang group, is used to create a broad and representative dataset for training and testing the emotion identification algorithm. This methodology enables the examination of emotions in different situations, encompassing both generic audio recordings and particular religious recitations.

3.2 Data Preprocessing:

Pre-processing procedures, such as transforming audio files into an appropriate format and adjusting their sample rate to a standardized frequency, are crucial for guaranteeing uniformity and compatibility. Extracting emotion labels from metadata allows for the correlation of emotions with audio samples, hence aiding the process of supervised learning in model training. These audio files contained aural data encoded in common formats such as WAV or MP3. Figure 6 - Ravdess Emotion Data illustrates the process of importing and incorporating the Qur'anic Murottal voice data into the research workflow. By importing the audio files, the study obtained access to the required data set for identifying emotions in Qur'anic recitations. This step established the groundwork for subsequent data preprocessing, feature extraction, and the development of the CNN and MFCC-based emotion identification model.

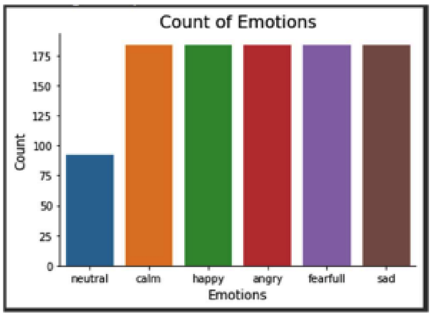


Figure 6. Ravdess Emotion Data

The Qur'anic Murottal voice dataset was meticulously annotated with corresponding emotions during the data labeling procedure. The researchers exhaustively analyzed each audio recording and identified the emotions conveyed by the sounds. The results of this meticulous labeling process are summarized in Table 2 - Labeling Emotions, which provides an exhaustive overview of the emotions present in the dataset as well as their respective counts. This meticulously labeled dataset serves as the basis for training and evaluating the CNN and MFCC-based emotion identification model. The accurate labeling of emotions guarantees the integrity of the dataset and enables the development of a robust model capable of identifying emotions in Qur'anic recitations.

Table 2. Labelling Emotions

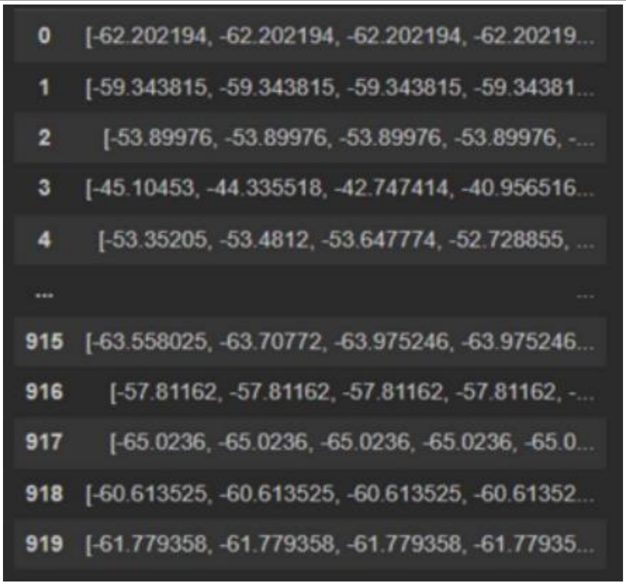
Emotion	Quantity
Male_calm	96
Male_happy	96
Male_angry	96
Male_fearful	96
Male_sad	96
Female_calm	88
Female_sad	83
Female_happy	88
Female_angry	88
Female_fearful	88
Male_neutral	48
Female_neutral	48

The labeled dataset includes 12 distinct emotional categories, such as happiness, anger, sadness, fear, and neutral expressions across male and female reciters. Each emotion category is adequately represented by a considerable number of audio samples, ranging from 48 to 96. This meticulous curation ensures the diversity and balance of the dataset, providing a firm foundation for training and evaluating the CNN and MFCC-based emotion identification model.

3.3 Extraction of Features

Extracting MFCC is an essential process for obtaining important attributes of audio sources. MFCC features offer a comprehensive depiction of the spectrum properties linked to various emotions by imitating the frequency resolution of the human auditory system. By using MFCC features as input to the CNN model, it guarantees that the model can efficiently learn distinctive patterns for identifying emotions.

The development of the CNN model focuses on extracting discriminative qualities from the MFCC data. By using numerous convolution layers and fully connected layers, the model can effectively capture the spatial and temporal dependencies present in the audio data. By training the model on smaller portions of the dataset and optimizing it using suitable techniques, we may verify that the model has the capability to reliably classify emotional states. As depicted in Figure 7 - Array MFCC, the extraction of MFCC features is carried out using well-defined and established methodologies. These numerical representations capture crucial spectral information, which is necessary for the subsequent stages of emotion identification. By employing MFCC feature extraction, the model acquires a rich and discriminative representation of the audio data, thereby enabling more accurate emotion recognition in Qur'anic Murottal recitations.



0	[-62.202194, -62.202194, -62.202194, -62.20219...
1	[-59.343815, -59.343815, -59.343815, -59.34381...
2	[-53.89976, -53.89976, -53.89976, -53.89976, -...
3	[-45.10453, -44.335518, -42.747414, -40.956516...
4	[-53.35205, -53.4812, -53.647774, -52.728855, ...
...	...
915	[-63.558025, -63.70772, -63.975246, -63.975246...
916	[-57.81162, -57.81162, -57.81162, -57.81162, -...
917	[-65.0236, -65.0236, -65.0236, -65.0236, -65.0...
918	[-60.613525, -60.613525, -60.613525, -60.61352...
919	[-61.779358, -61.779358, -61.779358, -61.77935...

Figure 7. Array MFCC

After feature extraction was complete, the Qur'anic Murottal sound dataset was split into two subsets: a labeled training set and a testing set. This division was performed at random in order to ensure an equal distribution of emotions in both subsets. The majority of the data was used to train the emotion identification model, while a smaller portion was set aside to evaluate the model's performance.

3.2.4 Model Evaluation:

Table 3 shows the performance metrics of a Convolutional Neural Network (CNN) model trained using Mel-frequency Cepstral Coefficients (MFCC) features to identify emotions in Qur'anic recitation. The model's performance was evaluated using two optimization algorithms: Adam and RMSprop. The findings suggest that the Adam optimizer consistently surpasses RMSprop in terms of accuracy, precision, and recall for most batch sizes. More precisely, when using a batch size of 8, the model attains its peak performance with the Adam optimizer, exhibiting an accuracy of 56%, precision of 58%, and recall of 59%. When the batch size is increased to 16 and 24, there is a modest decrease in accuracy; however, precision and recall stay generally consistent. RMSprop, however, demonstrates its optimal performance while using a batch size of 32, showcasing somewhat superior precision and recall compared to lesser batch sizes. These findings indicate that Adam is more effective for smaller batch sizes, while RMSprop may yield competitive results for larger batch sizes. This suggests that optimizing batch size and learning rates could improve the model's emotion identification capabilities in Qur'anic recitation contexts. Further exploration in this area is warranted. Significantly, our study provides distinct perspectives in this field, as prior studies [38][39] did not thoroughly investigate the influence of optimizer and batch size on the accuracy of emotion recognition. This highlights the importance of our discoveries and offers useful direction for future research aiming to enhance emotion identification algorithms for Qur'anic Murottal sounds. Significantly, our study provides distinct perspectives in this field, as prior studies [38][39] did not thoroughly investigate the influence of optimizer and batch size on the accuracy of emotion recognition. This highlights the importance of our discoveries and offers useful direction for future research aiming to enhance emotion identification algorithms for Qur'anic Murottal sounds.

Table 3. CNN Test Results

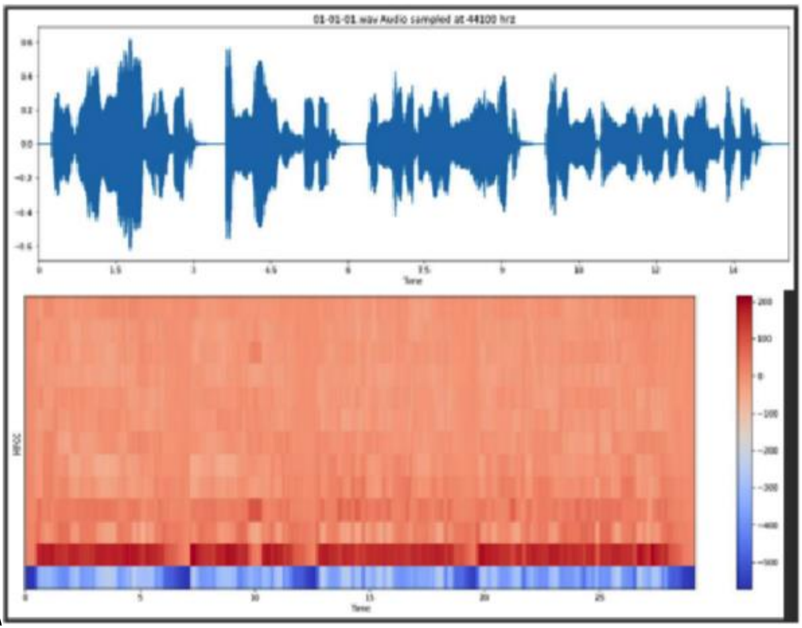
Batch size	**Result**					
	Accuracy %		**Precision %**		**Recall %**	
	Opt_adam	Opt_RMSprop	Opt_adam	Opt_RMSprop	Opt_adam	Opt_RMSprop
8	56%	45%	58%	47%	59%	52%
16	53%	50%	54%	49%	55%	56%
24	54%	52%	55%	52%	55%	55%
32	48%	51%	50%	53%	55%	53%

3.2.5 Emotion Recognition using Qur'anic Murottal Data

Utilizing the trained Convolutional Neural Network (CNN) model on Qur'anic murottal data enables the identification of emotional expressions in certain religious recitations. The efficacy of the model in discerning emotions in the context of Qur'anic recitations is assessed by analyzing emotional expressions in the recitations of suras Al-Ikhlas, Al-Falaq, and An-Nas.

The results of the emotion recognition approach are carefully examined and interpreted, with a specific emphasis on the implications for analyzing emotions in Qur'anic recitation. The debate analyzes the potential benefits and drawbacks of the proposed strategy that combines CNN and MFCC. It offers insights into the practicality and suitability of this approach.

Using the designated training dataset, the Convolutional Neural Network (CNN) underwent extensive training during the architecture training phase. During this process, crucial parameters including learning rate, number of epochs, and optimization methods were fine-tuned to optimize the performance of the model. During training, the CNN's performance was continuously evaluated using pertinent metrics, such as the loss function and accuracy. Figure 8 - Sample Data of Murottal Al-Qur'an Surah Al-Ikhlas depicts the essence of this training procedure visually. At each epoch, monitoring and analysis of the architectural performance were conducted, allowing for a deeper comprehension of how the model's performance evolved over the course of training. This rigorous training method ensures that the CNN architecture is capable of identifying voice emotions in Qur'anic Murottal recitations with a high degree of accuracy and dependability.



\

Figure 8. Sample Data of Murottal Al-Qur'an Surah Al-Ikhlas

Using the MFCC feature extraction and CNN algorithm, this research aims to produce significant results and substantial insights into emotion identification in the Murottal sound of the Qur'an by implementing the aforementioned proposed methods. It is anticipated that the comprehensive analysis and subsequent discussions will provide a deeper comprehension of emotion recognition in the Qur'an recitation voice, laying the groundwork for the development of relevant applications. Table 4 - Identify Murottal Emotions will demonstrate the culmination of this research by demonstrating the model's ability to accurately identify emotions within sacred recitations, contributing to the advancement of religious audio analysis and the development of intelligent systems in this domain.

Table 4. Identify Murottal Emotions

Voice Data	Identification results
01-01-01	Male_fearfull
01-02-03	Male_sad
01-03-05	Male_fearfull
01-03-07	Male_happy
02-03-02	Female_fearfull
02-02-04	Female_fearfull
02-01-06	Female_sad
02-01-06	Female_angry

4. Conclusion

Ultimately, this study focuses on the difficulty of human perception and subjective analysis when it comes to distinguishing emotions in Qur'anic Murottal sounds. It aims to overcome this obstacle by deploying automated emotion detection systems, notably employing CNN architecture and MFCC feature extraction. The results emphasize the importance of using sophisticated computational methods to improve the accuracy of emotion identification in religious audio settings.

By examining the CNN architecture and utilizing MFCC feature extraction, numerous significant findings have been uncovered. The selection of an optimizer inside the Convolutional Neural Network (CNN) architecture has a substantial impact on the accuracy of emotion identification. The Adam optimizer performs better than RMSprop in this regard. Moreover, directing attention towards distinct emotional factors might enhance the precision of the model in distinguishing between various emotions, underscoring the significance of focused investigation.

Nevertheless, the study also uncovers difficulties, namely in integrating gender identity with emotion detection, leading to decreased accuracy. This highlights the necessity for additional improvement and investigation in systems for identifying many tasks simultaneously.

In order to overcome these obstacles and improve the identification of emotions in Qur'anic Murottal sounds, this study suggests potential areas for further research. Incorporating datasets in several languages can enhance the model's ability to generalize, while including new datasets outside the Ravdess Song Emotion dataset can enhance the model's capability to identify emotions. In addition, investigating the simultaneous recognition of voices can enhance effectiveness and facilitate the use of applications in real-time.

References

[1] Z. Jia, Y. Lin, J. Wang, Z. Feng, X. Xie, and C. Chen, “HetEmotionNet: Two-Stream Heterogeneous Graph Recurrent Neural Network for Multi-modal Emotion Recognition,” *MM 2021 - Proceedings of the 29th ACM International Conference on Multimedia*. pp. 1047–1056, 2021. https://doi.org/10.1145/3474085.3475583
[2] R. Rafeh, R. Azimi Khojasteh, and N. Alobaidi, “Proposing A Hybrid Approach for Emotion Classification using Audio and Video Data,” pp. 31–40, 2019. https://doi.org/10.5121/csit.2019.91403
[3] F. Ye, “Emotion recognition of online education learners by convolutional neural networks,” *Comput. Intell. Neurosci.*, vol. 2022, p. 4316812, 2022. https://doi.org/10.1155/2022/4316812
[4] X. Liang, J. Liang, T. Yin, and X. Tang, “A lightweight method for face expression recognition based on improved MobileNetV3,” *IET Image Process.*, vol. 17, no. 8, pp. 2375–2384, 2023. https://doi.org/10.1049/ipr2.12798
[5] X. Xu, Y. Zhang, M. Tang, H. Gu, S. Yan, and J. Yang, “Emotion recognition based on double tree complex wavelet transform and machine learning in internet of things,” *IEEE Access*, vol. 7, pp. 154114–154120, 2019. https://doi.org/10.1109/access.2019.2948884
[6] Z. Han, H. Chang, X. Zhou, J. Wang, L. Wang, and Y. Shao, “E2ENNet: An end-to-end neural network for emotional brain-computer interface,” *Front. Comput. Neurosci.*, vol. 16, p. 942979, 2022. https://doi.org/10.3389/fncom.2022.942979
[7] C. Wei, L. lan Chen, Z. zhen Song, X. guang Lou, and D. dong Li, “EEG-based emotion recognition using simple recurrent units network and ensemble learning,” *Biomed Signal Process Control*, vol. 58, p. 101756, 2020. https://doi.org/10.1016/j.bspc.2019.101756
[8] H. Yang, J. Han, and K. Min, “A multi-column CNN model for emotion recognition from EEG signals,” *Sensors (Switzerland)*, vol. 19, no. 21, pp. 1–12, 2019. https://doi.org/10.3390/s19214736
[9] M. R. Islam *et al.*, “EEG Channel Correlation Based Model for Emotion Recognition,” *Comput Biol Med*, vol. 136, no. May, p. 104757, 2021. https://doi.org/10.1016/j.compbiomed.2021.104757
[10] J.-R. Zhuang, Y.-J. Guan, H. Nagayoshi, K. Muramatsu, K. Watanuki, and E. Tanaka, “Real-time emotion recognition system with multiple physiological signals,” *J. Adv. Mech. Des. Syst. Manuf.*, vol. 13, no. 4, pp. JAMDSM0075–JAMDSM0075, 2019. https://doi.org/10.1299/jamdsm.2019jamdsm0075
[11] P. Li *et al.*, “EEG Based Emotion Recognition by Combining Functional Connectivity Network and Local Activations,” *IEEE Trans Biomed Eng*, vol. 66, no. 10, pp. 2869–2881, 2019. https://doi.org/10.1109/TBME.2019.2897651
[12] N. Y. Weinstein, L. B. Whitmore, and K. L. Mills, “Individual differences in mentalizing tendencies,” *Collabra Psychol.*, vol. 8, no. 1, 2022. https://doi.org/10.1525/collabra.37602
[13] Y. Cimtay and E. Ekmekcioglu, “Investigating the use of pretrained convolutional neural network on cross-subject and cross-dataset eeg emotion recognition,” *Sensors (Switzerland)*, vol. 20, no. 7, pp. 1–20, 2020. https://doi.org/10.3390/s20072034
[14] Y. Liu *et al.*, “Multi-channel EEG-based emotion recognition via a multi-level features guided capsule network,” *Comput Biol Med*, vol. 123, no. March, p. 103927, 2020. https://doi.org/10.1016/j.compbiomed.2020.103927
[15] L. J. Z. Et.al, “Investigating the use of eye fixation data for emotion classification in VR,” *Turk. J. Comput. Math. Educ. (TURCOMAT)*, vol. 12, no. 3, pp. 1852–1857, 2021. https://doi.org/10.17762/turcomat.v12i3.1014
[16] T. Zhang, W. Zheng, Z. Cui, Y. Zong, and Y. Li, “Spatial-Temporal Recurrent Neural Network for Emotion Recognition,” *IEEE Trans Cybern*, vol. 49, no. 3, pp. 939–947, 2019. https://doi.org/10.1109/TCYB.2017.2788081
[17] S. K. Khare and V. Bajaj, “Time-Frequency Representation and Convolutional Neural Network-Based Emotion Recognition,” *IEEE Trans Neural Netw Learn Syst*, vol. 32, no. 7, pp. 2901–2909, 2021. https://doi.org/10.1109/TNNLS.2020.3008938
[18] L. Liu, Y. Ji, Y. Gao, T. Li, and W. Xu, “A data-driven adaptive emotion recognition model for college students using an improved multifeature deep neural network technology,” *Comput. Intell. Neurosci.*, vol. 2022, p. 1343358, 2022. https://doi.org/10.1155/2022/1343358
[19] L. Shu *et al.*, “Wearable emotion recognition using heart rate data from a smart bracelet,” *Sensors (Switzerland)*, vol. 20, no. 3, pp. 1–19, 2020. https://doi.org/10.3390/s20030718
[20] S. Syafril, N. E. Yaumas, E. Engkizar, A. Jaafar, and Z. Arifin, “Sustainable development: Learning the Quran using the tartil method,” *AL-TA LIM*, vol. 28, no. 1, pp. 1–8, 2021. https://doi.org/10.15548/jt.v1i1.673
[21] D. Daliman, “Ethical conduct-do and general well-being among university students, moderated by religious internalization: An Islamic perspective,” *Indigenous*, vol. 6, no. 2, pp. 14–24, 2021. https://doi.org/10.23917/indigenous.v6i2.14886
[22] N. Najiburrahman, Y. N. Azizah, J. Jazilurrahman, W. Azizah, and N. A. Jannah, “Implementation of the tahfidz Quran program in developing Islamic character,” *J. Obs. J. Pendidik. Anak Usia Dini*, vol. 6, no. 4, pp. 3546–3599, 2022. https://doi.org/10.31004/obsesi.v6i4.2077

[23] M. A. Al-Jarrah *et al.*, "Accurate Reader Identification for the Arabic Holy Quran Recitations Based on an Enhanced VQ Algorithm," *Revue d'Intelligence Artificielle*, vol. 36, no. 6, pp. 815–823, 2022. https://doi.org/10.18280/ria.360601

[24] Y. Hanafi *et al.*, "Student's and instructor's perception toward the effectiveness of E-BBQ enhances Al-qur'an reading ability," *Int. J. Instr.*, vol. 12, no. 3, pp. 51–68, 2019. https://doi.org/10.29333/iji.2019.1234a

[25] J. H. Alkhateeb, "A machine learning approach for recognizing the Holy Quran reciter," *International Journal of Advanced Computer Science and Applications*, vol. 11, no. 7, pp. 268–271, 2020. https://doi.org/10.14569/IJACSA.2020.0110735

[26] J. W. Watts, "Sensation and metaphor in ritual performance: The example of sacred texts," *Entangled Relig.*, vol. 10, 2019. https://doi.org/10.46586/er.10.2019.8365

[27] W. Zhang, "Intelligent recognition and analysis of negative emotions of undergraduates under COVID-19," *Front. Public Health*, vol. 10, p. 913255, 2022. https://doi.org/10.3389/fpubh.2022.913255

[28] B. Chakravarthi, S.-C. Ng, M. R. Ezilarasan, and M.-F. Leung, "EEG-based emotion recognition using hybrid CNN and LSTM classification," *Front. Comput. Neurosci.*, vol. 16, p. 1019776, 2022. https://doi.org/10.3389/fncom.2022.1019776

[29] Y. Cui and F. Wang, "Research on audio recognition based on the deep neural network in music teaching," *Comput. Intell. Neurosci.*, vol. 2022, p. 7055624, 2022. https://doi.org/10.1155/2022/7055624

[30] H. Geng, Y. Hu, and H. Huang, "Monaural singing voice and accompaniment separation based on gated nested U-Net architecture," *Symmetry (Basel)*, vol. 12, no. 6, p. 1051, 2020. https://doi.org/10.3390/sym12061051

[31] V.-T. Tran and W.-H. Tsai, "Speaker identification in multi-talker overlapping speech using neural networks," *IEEE Access*, vol. 8, pp. 134868–134879, 2020. https://doi.org/10.1109/access.2020.3009987

[32] M. Bandara, R. Jayasundara, I. Ariyarathne, D. Meedeniya, and C. Perera, "Forest sound classification dataset: FSC22," *Sensors (Basel)*, vol. 23, no. 4, 2023. https://doi.org/10.3390/s23042032

[33] H. K. Shin, S. H. Park, and K. W. Kim, "Inter-floor noise classification using convolutional neural network," *PLoS One*, vol. 15, no. 12 December 2020, 2020. https://doi.org/10.1371/journal.pone.0243758

[34] O. Ilina, V. Ziyadinov, N. Klenov, and M. Tereshonok, "A Survey on Symmetrical Neural Network Architectures and Applications," *Symmetry (Basel)*, vol. 14, no. 7, 2022. https://doi.org/10.3390/sym14071391

[35] Z. Huang and M. Liao, "Evidence-Based Research on Multimodal Fusion Emotion Recognition," pp. 594–601, 2023. https://doi.org/10.2991/978-94-6463-200-2_61

[36] S. C. Lai *et al.*, "Hardware Accelerator Design of DCT Algorithm with Unique-Group Cosine Coefficients for Mel-Scale Frequency Cepstral Coefficients," *IEEE Access*, vol. 10, pp. 79681–79688, 2022. https://doi.org/10.1109/ACCESS.2022.3194261

[37] Y. Sharma and Dr. B. Kumar Singh, "Depression analysis of voice samples using machine learning," *Journal of University of Shanghai for Science and Technology*, vol. 23, no. 11, pp. 429–438, 2021. https://doi.org/10.51201/jusst/21/10820

[38] A. Al Harere and K. Al Jallad, "Mispronunciation Detection of Basic Quranic Recitation Rules using Deep Learning," *arXiv*. pp. 1–16, 2023. https://doi.org/10.48550/arXiv.2305.06429

[39] A. S. Wibowo, I. D. M. Bayu, and A. Darmawan, "Iqra reading verification with mel frequency cepstrum coefficient and dynamic time warping," *Journal of Physics: Conference Series*, vol. 1722, no. 1. 2021. https://doi.org/10.1088/1742-6596/1722/1/012015