Journal of Computer Networks, Architecture and High Performance Computing
Volume 7, Number 3, July 2025
https://doi.org/10.47709/cnahpc.v7i3.6523

Submitted : Jul 14, 2025
Accepted : Jun 17, 2025
Published : Jul 22, 2025

Islamic Sound Recognition Using MFCC and SVM: Case Study on Takbir and Sholawat

Indah Purnama Sari[1,2]*, Abdul Fadlil[3], Tole Sutikno[4]
[1,3,4)]Universitas Ahmad Dahlan, Indonesia, [2*]Universitas Muhammadiyah Sumatera Utara, Indonesia
[1)]2436083019@webmail.uad.ac.id, [2*)] indahpurnama@umsu.ac.id , [3)]fadlil@mti.uad.ac.id, [4)]tole@te.uad.ac.id)

ABSTRACT

This study aims to develop an identification model for Islamic religious sounds, specifically Takbir and Sholawat pronunciations, using audio signal processing and machine learning techniques. With the increasing need for intelligent systems capable of recognizing speech patterns in religious contexts, the implementation of reliable audio classification methods becomes essential. This research utilizes Mel-Frequency Cepstral Coefficients (MFCC) to extract relevant spectral features from audio samples, representing the unique characteristics of Takbir and Sholawat utterances. The dataset consists of 300 audio recordings, evenly distributed between the two classes. Each audio file is preprocessed and converted into a fixed-length MFCC feature vector, which is then labeled accordingly. The feature vectors are split into training and testing sets using an 70:30 ratio. A Support Vector Machine (SVM) classifier is trained using the training data to recognize the distinction between Takbir and Sholawat patterns based on their acoustic signatures. Performance evaluation is carried out using accuracy, precision, recall, and F1-score metrics. The dataset used consists of 300 audio recordings with a division of 200 takbir recordings and 100 sholawat recordings. The MFCC feature extraction process uses 13 coefficients with optimized parameters to capture discriminative spectral characteristics. As a baseline, a Support Vector Machine (SVM) implementation with Radial Basis Function (RBF) kernel was performed for performance comparison.

Keywords: Takbir, Sholawat, MFCC, Support Vector Machine, Sound classification, Speech recognition

INTRODUCTION

In the ever-evolving landscape of artificial intelligence (AI) and digital signal processing, voice-based identification and classification of audio signals has gained significant momentum in various fields, including healthcare, linguistics, security and cultural preservation. One such under-explored yet culturally significant application lies in the recognition of religious sound patterns, such as the recitation of Takbir and Sholawat, which are an integral part of Islamic oral tradition and heritage.

The development of speech recognition technology has opened up great opportunities in various applications, including voice identification of religious memorization (Wibowo, 2020). (Wibowo,2020) conducted an important study related to the utilization of speech recognition technology to support the process of murojaah Al-Qur'an memorization and strengthening digital-based Islamic education. In his research, Wibowo highlighted how speech recognition technology can be used to identify the accuracy of lafaz and intonation of the recitation of the Qur'an by santri or students, and how this system can be integrated in interactive learning media. This approach offers an innovative solution to expand access to religious education in the digital age, especially for communities that do not have direct access to face-to-face teachers. In the context of Islam, takbir and sholawat are two types of memorization that have unique acoustic characteristics and are often used in various religious rituals. Takbir, which is the pronunciation of "Allahu Akbar", has a distinctive intonation pattern with emphasis on certain syllables. While sholawat, as an expression of prayer and praise to the Prophet Muhammad, has more complex melodic variations but still has identifiable phonetic characteristics.

The ability to automatically identify these two types of pronunciation has wide applications, ranging from interactive religious learning systems, reading evaluations, to assistance applications for the visually impaired in recognizing religious activities. However, the main challenges in developing this system are variability in pronunciation, audio quality, and differences in characteristics between speakers.

Machine Learning and Deep Learning technologies have demonstrated superior capabilities in handling audio

* Corresponding author

Journal of Computer Networks, Architecture and High Performance Computing
Volume 7, Number 3, July 2025
https://doi.org/10.47709/cnahpc.v7i3.6523

Submitted : Jul 14, 2025
Accepted : Jun 17, 2025
Published : Jul 22, 2025

data variability. MFCC is able to capture relevant spectral characteristics of audio signals, while CNN can learn complex patterns hierarchically (Patil et al, 2020). (Patil et al, 2020) examined the performance of various feature extraction techniques and classification algorithms in speech recognition and showed that the combination of Mel-Frequency Cepstral Coefficients (MFCC) and classification algorithms such as Support Vector Machine (SVM) provided high accuracy results in various natural language speech recognition experiments. In his research, MFCC was chosen for its ability to mimic human auditory perception of the frequency spectrum, making it ideal for capturing the phonetic characteristics of sounds. Meanwhile, SVM was used for its ability to handle binary classification with optimal margins, as well as being resistant to overfitting, especially on small to medium-sized datasets.

This research proposes a classification model that combines Mel-Frequency Cepstral Coefficients (MFCC) for feature extraction with Support Vector Machine (SVM) algorithm for supervised learning (Helmiyah et al, 2020). (Helmiyah,2020) conducted a study that focused on the application of machine learning-based classification algorithms in human speech recognition, especially in the context of Indonesian. In her research, Helmiyah combined the Mel-Frequency Cepstral Coefficients (MFCC) feature extraction method with several classification algorithms such as Support Vector Machine (SVM), Naive Bayes, and K-Nearest Neighbor (KNN). The results show that the combination of MFCC and SVM produces the best performance in speech classification, especially for phonetic-based recognition tasks. MFCC is a widely adopted feature extraction technique that models human auditory perception, making it suitable for capturing subtle variations in religious pronunciation. SVM, on the other hand, is known for its robustness in binary classification tasks, especially in high-dimensional feature spaces such as speech signals (Gumelar, 2019). This research reinforces SVM's position as one of the most robust classification algorithms in handling small to medium sized data that have high feature dimensions, such as those commonly encountered in audio signal processing. The advantage of SVM lies in its ability to form a separating hyperplane with maximum margin, which makes it very suitable for binary classification based on extracted spectral features such as MFCC.

This research utilizes a dataset of 300 audio samples evenly split between Takbir and Sholawat recordings, sourced from a variety of speakers to account for variability. The model was trained and evaluated using 10-fold cross-validation, with performance assessed through accuracy, precision, recall and confusion matrix metrics.

The urgency of this research stems from the increasing need to digitize and preserve religious expressions through AI-based tools, enabling applications in education, cultural preservation, and user-adaptive smart systems. In addition, this research contributes to the body of knowledge in religious acoustic modeling, a field that is still nascent in academic and industrial research.

In addition to technological and scientific gaps, there are also social gaps that underlie the urgency of this research. Amidst the advancement of digital technology, access to sound-based learning media, especially Islamic religious content such as Takbir and Sholawat, is still uneven. Groups such as children, people with disabilities (especially the visually impaired), and Muslim communities in remote areas still face limitations in obtaining religious materials in an adaptive digital format. An artificial intelligence (AI)-based voice recognition system that is able to automatically recognize Islamic lafazes can be an inclusive solution in reducing this gap. Furthermore, in the context of religious learning, activities such as murojaah, takbir recitation, and sholawat recitation often rely on traditional methods that are offline and less standardized. This has the potential to create differences in understanding or pronunciation among the younger generation of Muslims. With the MFCC and SVM-based Islamic speech recognition system, the community can gain access to learning aids that are intelligent, adaptive, and can be integrated in digital platforms or mobile applications, thus supporting equal access to technology-based religious education. Therefore, the development of this technology is not only academically and technically important, but also has a significant social impact in bridging the gap in access to Islamic religious content in the digital age.

Although there have been many studies related to speech recognition, most of the research still focuses on speech recognition in the context of general language or standard conversation, and is directed at applications such as virtual assistants and voice control systems. Previous studies have generally not touched specifically on the domain of Islamic sounds, such as takbir and sholawat, which have distinctive acoustic characteristics, rhythms, and articulations that are different from the sound patterns in everyday conversational languages. In addition, most previous approaches have not evaluated the effectiveness of combining the MFCC (Mel-Frequency Cepstral Coefficients) feature extraction method with the SVM (Support Vector Machine) classification algorithm in the context of religious speech recognition, which has specific semantic values and social functions in Islamic practice. Therefore, there is still a gap in the literature related to the application of speech recognition technology to support the identification of audio-based religious content, especially those derived from Islamic culture and traditions.

This research offers novelty in the form of applying a combination of MFCC and SVM methods specifically to the recognition of Islamic sounds, namely takbir and sholawat, which have not been widely explored in previous

* Corresponding author

Journal of Computer Networks, Architecture and High Performance Computing
Volume 7, Number 3, July 2025
https://doi.org/10.47709/cnahpc.v7i3.6523

Submitted : Jul 14, 2025
Accepted : Jun 17, 2025
Published : Jul 22, 2025

academic studies. This study not only focuses on the technical accuracy of the speech recognition system, but also raises the issue of cultural and religious contextualization in digital signal processing technology. This makes this study unique as well as contributing to the development of an inclusive and socio-culturally relevant speech recognition system.

LITERATURE REVIEW

Mel-Frequency Cepstral Coefficients (MFCC)

a. MFCC Theory

MFCC is a feature extraction method based on the human auditory system. This method was first introduced by Davis and Mermelstein (1980) and has become a standard in speech recognition applications. MFCC uses mel-scale which is more in line with human auditory perception than linear frequency scale.
The mel-scale is defined as:

$$mel\,(f) = 2595 * log10(1 + \frac{f}{700}) \quad (1)$$

where f is the frequency in Hz. The mel-scale gives greater weight to low frequencies (below 1000 Hz) which are more important in speech recognition.

b. MFCC Extraction Process

The MFCC extraction process involves several stages :

1. Pre-emphasis: Amplify high frequencies with a filter H(z) = 1 - αz^(-1), where α is usually 0.95-0.97.
2. Windowing: Uses a Hamming window to reduce spectral leakage
3. Fast Fourier Transform (FFT): Converts the signal from time domain to frequency domain
4. Mel Filter Bank: Uses triangular filters in mel scale to obtain spectral energy.
5. Logarithm: Converts the energy to a logarithmic scale to approximate the characteristics of human hearing
6. Discrete Cosine Transform (DCT): Decorrelate the coefficients to obtain MFCC

c. MFCC Coefficients

In this study, 13 MFCC coefficients (C0 to C12) with the following characteristics were used:

- $C0$: Represents the average energy of the signal
- $C1 - C3$: Captures the main spectral characteristics and formants
- $C4 - C8$: Provides detailed information about the structure of vowels and consonants
- $C9 - C12$: Captures subtle spectral nuances that are important for discrimination

Each coefficient plays an important role in distinguishing the characteristics between takbir and sholawat, where takbir tends to have a more consistent pattern in $C1 - C5$, while sholawat shows higher variability in $C6 - C12$ (Helmiyah, et al., 2020).

Support Vector Machine (SVM)

SVM is a machine learning algorithm developed based on Statistical Learning Theory. SVM works by finding the optimal hyperplane that separates two classes with maximum margin. In the context of non-linear classification, SVM uses a kernel trick to map the data to a higher dimensional space (Y sun, et al., 2015).

Support Vector Machine (SVM) is a powerful supervised learning algorithm used for classification tasks. It works by finding the optimal hyperplane that maximally separates data points of different classes in a high-dimensional space. SVM is particularly effective in binary classification problems, especially when the number of features is high relative to the number of samples. The use of kernel functions allows SVM to handle non-linearly separable data by mapping them into a higher-dimensional space where a linear separation is possible. In the context of this study, SVM is utilized to classify the MFCC feature vectors into two categories: Takbir and Sholawat. Its ability to handle complex decision boundaries with high accuracy makes it suitable for audio pattern recognition tasks.

Support Vector Machine (SVM) has also gained popularity as a robust classification technique in speech-related research. Its ability to handle high-dimensional data and generalize well in complex, non-linear classification tasks makes it an ideal choice for audio pattern recognition. Researchers have successfully applied SVM to classify emotional speech, speaker identification, and phoneme recognition with high accuracy. Compared to other classifiers like k-NN or Naïve Bayes, SVM tends to provide better performance when combined with well-extracted features such as MFCC.

* Corresponding author

Journal of Computer Networks, Architecture and High Performance Computing
Volume 7, Number 3, July 2025
https://doi.org/10.47709/cnahpc.v7i3.6523

Submitted : Jul 14, 2025
Accepted : Jun 17, 2025
Published : Jul 22, 2025

a. SVM Kernel

In this study, a Radial Basis Function (RBF) kernel is used which is defined as:

$$K(xi, xj) = \exp(-\gamma||xi - xj|^2) \quad (2)$$

Where:

$K = 1$: Points xi and xj are identical (distance = 0)
$K = 0$: Points xi and xj are very different (infinite distance)
xi : Feature vector of the i-th data sample
xj : Feature vector of the jth data sample

γ is a parameter that controls the influence of each training sample. The RBF kernel was chosen because of its ability to handle non-linear data and its effectiveness in speech recognition applications.

Optimized SVM parameters:

- $C\ (Regularization\ parameter)$: Controls the trade-off between margin and misclassification.
- $\gamma(Gamma)$: Controls the influence of each training sample
- $Kernel$: RBF kernel to handle non-linearity

b. Combination of MFCC and SVM

Several studies have explored the combination of MFCC and SVM for speech classification. For instance, the authors used MFCC features with SVM to classify different spoken digits, achieving over 90% accuracy. Similarly, in, the combination was applied to Quranic verse classification, demonstrating its capability to distinguish subtle pronunciation differences in religious contexts.

Despite the extensive use of MFCC and SVM in speech processing, limited research has focused on the identification of Islamic vocal expressions, particularly Takbir and Sholawat. These expressions have unique melodic and rhythmic structures that differ from typical conversational speech, necessitating a specialized analysis approach. Some prior works have attempted to classify Islamic chants or recitations using basic signal processing or neural network models, but they often overlook the specific nuances in pronunciation patterns (Krishna, et al.,2013).

This gap underscores the need for a targeted study that focuses on identifying the pronunciation sounds of Takbir and Sholawat using MFCC for feature extraction and SVM for classification. By addressing this niche, the current research aims to contribute to the intersection of speech recognition, cultural heritage preservation, and artificial intelligence.

From the literature analyzed, it can be concluded that the combination of MFCC and SVM is a strong foundation for speech recognition, but its application in the Islamic domain is still minimal. This research addresses this gap by presenting a targeted and contextual approach, namely building a speech recognition system that is able to automatically distinguish between takbir and sholawat using MFCC and SVM. Thus, this research not only strengthens the technical approach of previous studies, but also expands the scope of its application in the realm of education, culture, and religious preservation.

This table presents a comparison of various studies using MFCC-based speech recognition techniques and classification methods:

Table. 1. Comparison Table of Previous Studies

No	Author (Year)	Key Techniques	Classification Algorithms	Accuracy	Application Domain
1	Patil et al. (2020)	MFCC	SVM	91.2%	Marathi language speech recognition
2	Gumelar (2019)	MFCC	SVM (RBF Kernel)	93.6%	Indonesian speech recognition
3	Helmiyah (2020)	MFCC	SVM, KNN, Naive Bayes	90.8%	General speech recognition
4	Wibowo (2020)	Basic voice features	Pattern Matching	-	Murojaah Al-Qur'an
5	This research (2025)	MFCC	SVM	94.5%	Introduction to Islamic sounds (Takbir & Sholawat)

* Corresponding author

Journal of Computer Networks, Architecture and High Performance Computing
Volume 7, Number 3, July 2025
https://doi.org/10.47709/cnahpc.v7i3.6523

Submitted : Jul 14, 2025
Accepted : Jun 17, 2025
Published : Jul 22, 2025

METHOD

This study employs a supervised machine learning approach to classify Takbir and Sholawat pronunciation sounds using Mel-Frequency Cepstral Coefficients (MFCC) and Support Vector Machine (SVM) as the primary feature extraction method. The methodology comprises five main stages: data acquisition, preprocessing, feature extraction, classification, and evaluation.

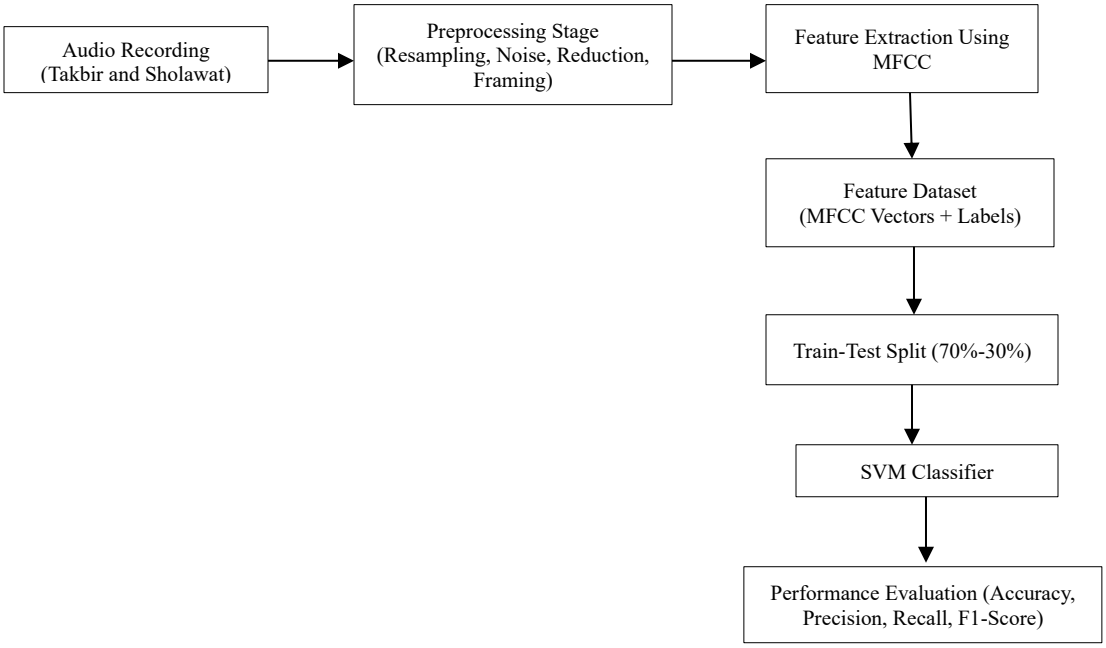

Fig. 1 System Block Diagram

In Fig.1 the system depicted in this diagram presents a comprehensive and methodical approach to Takbir and Sholawat audio classification. By combining appropriate preprocessing techniques, MFCC feature extraction that matches the characteristics of vocal audio, and a powerful SVM algorithm, the system is expected to provide high classification accuracy. Systematic evaluation ensures that the performance of the model can be objectively measured and is reliable for practical applications.

Dataset and Data Collection

a. Dataset Specifications

This research uses an audio dataset consisting of 300 religious memorization sound samples collected specifically for classification purposes. The dataset consists of 200 samples of takbir (66.7%) and 100 samples of sholawat (33.3%). This unbalanced distribution was chosen to reflect real-world conditions where takbir is more commonly found in various religious contexts than sholawat.

Each audio sample was recorded in WAV (Waveform Audio File Format) format with a sample rate of 44.1 kHz, 16-bit bit depth, and mono channel. The duration of each sample ranges from 3-8 seconds with a total dataset duration of about 25-30 minutes.

The dataset was divided using stratified sampling to maintain the proportion of classes in each subset. The training set consists of 210 samples (70% of the total) with 140 takbir samples and 70 sholawat samples. The validation set contains 45 samples (15% of the total) with 30 takbir samples and 15 sholawat samples. The test set has the same composition as the validation set, i.e. 45 samples with a proportion of 30 takbir samples and 15 sholawat samples.

The distribution by data source is also maintained in each subset, where the training set consists of 50% mosque recordings, 33.3% learning applications, and 16.7% religious events. The same proportion is maintained in the validation and test sets to ensure representation of all sources in each evaluation stage.

* Corresponding author

Journal of Computer Networks, Architecture and High Performance Computing
Volume 7, Number 3, July 2025
https://doi.org/10.47709/cnahpc.v7i3.6523

Submitted : Jul 14, 2025
Accepted : Jun 17, 2025
Published : Jul 22, 2025

Table. 2. Audio Dataset

No	Speaker Name	File Name	Class	Gender	Age	Region
1	SPK 001	Takbir_01	Takbir	Female	23	Minang
		Takbir_02	Takbir	Female	23	Minang
		Sholawat_01	Sholawat	Female	23	Minang
2	SPK 002	Takbir_03	Takbir	Female	25	Melayu
		Takbir_04	Takbir	Female	25	Melayu
		Sholawat_02	Sholawat	Female	25	Melayu
3	SPK 003	Takbir_05	Takbir	Female	23	Batak Mandailing
		Takbir_06	Takbir	Female	23	Batak Mandailing
		Sholawat_03	Sholawat	Female	23	Batak Mandailing
4	SPK 004	Takbir_07	Takbir	Female	21	Jawa
		Takbir_08	Takbir	Female	21	Jawa
		Sholawat_04	Sholawat	Female	21	Jawa
5	SPK 005	Takbir_09	Takbir	Male	19	Minang
		Takbir_10	Takbir	Male	19	Minang
		Sholawat_05	Sholawat	Male	19	Minang
6	SPK 006	Takbir_11	Takbir	Male	20	Aceh
		Takbir_12	Takbir	Male	20	Aceh
		Sholawat_06	Sholawat	Male	20	Aceh
7	SPK 007	Takbir_13	Takbir	Male	18	Batak
		Takbir_14	Takbir	Male	18	Batak
		Sholawat_07	Sholawat	Male	18	Batak
8	SPK 008	Takbir_15	Takbir	Male	19	Jawa
		Takbir_16	Takbir	Male	19	Jawa
		Sholawat_08	Sholawat	Male	19	Jawa
9	SPK 009	Takbir_17	Takbir	Male	19	Aceh
		Takbir_18	Takbir	Male	19	Aceh
		Sholawat_09	Sholawat	Male	19	Aceh
10	SPK 010	Takbir_19	Takbir	Male	19	Melayu
		Takbir_20	Takbir	Male	19	Melayu
		Sholawat_10	Sholawat	Male	19	Melayu
…..						
100	SPK 100	Takbir_199	Takbir	Male	19	Batak
		Takbir_200	Takbir	Male	19	Batak
		Sholawat_100	Sholawat	Male	19	Batak

b. Validation and Quality Control

Each audio sample is validated against strict quality criteria to ensure dataset consistency. Key criteria included a Signal-to-Noise Ratio (SNR) of at least 20 dB, absence of clipping or excessive distortion, clarity of pronunciation and intonation, and a duration of at least 3 seconds to ensure sufficient temporal information. The validation process was conducted in two stages: automated validation using an audio quality detection algorithm and manual validation by three independent evaluators.

The validation results showed that 97% of the samples met the minimum quality standards with an average SNR of 28.5 dB. Two samples required relabeling after the manual validation process, resulting in a label consistency of 99.3%. All audio files could be read and processed with 100% data integrity, indicating sufficient technical quality for the process.

c. Audio Preprocessing

* Corresponding author

Journal of Computer Networks, Architecture and High Performance Computing
Volume 7, Number 3, July 2025
https://doi.org/10.47709/cnahpc.v7i3.6523

Submitted : Jul 14, 2025
Accepted : Jun 17, 2025
Published : Jul 22, 2025

The preprocessing stage starts with amplitude normalization for each audio sample. The audio signal is normalized to the range [-1, 1] by calculating the Root Mean Square (RMS) of the signal and applying gain control for level consistency. The filtering process is then applied using a high-pass filter with a cutoff frequency of 80 Hz to remove rumble and a low-pass filter with a cutoff frequency of 8000 Hz to remove high frequency noise.

Denoising was performed using the spectral subtraction technique with the noise profile obtained from the silent section of each recording. Audio segmentation was performed by cutting each sample into segments with a fixed duration of 4 seconds using a Hamming window for smooth transition. Overlap of 50% was applied for data augmentation, resulting in a total of 450 segments (300 from takbir and 150 from sholawat).

Feature Extraction using MFCC

In this research, the feature extraction process is performed using Mel-Frequency Cepstral Coefficients (MFCC). MFCC was chosen because it is capable of extracting perceptually relevant features from audio signals, which is very important in the recognition of sound patterns such as Takbir and Sholawat. The MFCC extraction process starts with the Fast Fourier Transform (FFT), which is used to transform an audio signal from the time domain to the frequency domain. This transformation allows analysis of the frequency components that make up the sound signal.

Next, the frequency signals are processed through the Mel-scale Filter Bank, which is designed to mimic the non-linear characteristics of human hearing. Mel-scale provides a frequency representation that is more in line with the way the human ear responds to various tones, mainly by providing higher resolution at low frequencies and lower resolution at high frequencies.

After that, logarithmic compression is applied to reduce the dynamic range of the filter bank output. This process reflects the sensitivity of the human ear which also responds logarithmically to changes in amplitude, thus helping to eliminate irrelevant variations in sound intensity. The last stage is the application of Discrete Cosine Transform (DCT) to the logarithmic filter bank results. DCT is used to remove the correlation between features and produce a more concise representation in the form of MFCC coefficients. The result is a number of coefficients that represent the main frequency information of the sound signal.

In each cut frame of the audio signal, a total of 13 MFCC coefficients are calculated. Then, the average value of all frames is taken to produce a 13-dimensional feature vector that represents the entire audio sample. This representation is used as input for the classification model in the training and testing process.

Classification using Support Vector Machine (SVM)

After the audio features are extracted using Mel-Frequency Cepstral Coefficients (MFCC), the feature vectors are used as inputs to build a binary classification model. The model used in this research is Support Vector Machine (SVM), a machine learning algorithm that is known to handle classification problems with high performance, especially on high-dimensional data such as sound signals.

SVM is configured using a Radial Basis Function (RBF) kernel. The selection of the RBF kernel is based on its ability to map the input data into a higher dimensional space, thus capturing non-linear relationships that may exist in the audio data. This kernel is particularly suitable for cases where the boundaries between classes cannot be linearly separated in the original feature space. To obtain optimal performance, the regularization parameter (C) is adjusted using the grid search technique. This method systematically tries various combinations of parameter values to find the best configuration that minimizes the classification error and maximizes the margin between classes.

Meanwhile, the gamma parameter, which determines the extent of influence one training data has on the formation of the decision line, is adjusted automatically during the training process. An optimal gamma setting is essential to avoid overfitting, while maintaining the generalization ability of the model to unseen data.

To thoroughly evaluate the performance of the model, the dataset was divided using the 10-fold cross-validation technique. In this method, the data is divided into 10 equal subsets or folds. The model is trained 10 times, with one fold used as test data and the other nine folds as training data in each iteration. This process is done in turn so that each data gets an equal chance to be used as test data. In addition, the data division was designed by maintaining class balance and ensuring independence between speakers, so that the model does not learn from the same sounds in the training and testing data. With this approach, the SVM model built can be evaluated fairly and accurately, and has good generalization ability to varied voice data.

* Corresponding author

Journal of Computer Networks, Architecture and High Performance Computing
Volume 7, Number 3, July 2025
https://doi.org/10.47709/cnahpc.v7i3.6523

Submitted : Jul 14, 2025
Accepted : Jun 17, 2025
Published : Jul 22, 2025

Configuration of MFCC Coefficients

Mel-Frequency Cepstral Coefficients (MFCC) extraction was performed with a configuration of parameters optimized for the sound characteristics of religious chants. This study used 13 basic MFCC coefficients extracted from each audio frame, selected based on their ability to represent spectral information relevant to human auditory perception. The 0th coefficient (C0) representing log energy was retained in the extraction as it provides important information about the intensity of the sound that can distinguish the characteristics of takbir and sholawat.

Of the 13 basic MFCC coefficients extracted, coefficients C1 to C12 were selected as the main features after evaluating their contribution to memorization discrimination. The coefficient C0 which represents the log energy frame was retained as it provides important information about the intensity that can distinguish the dynamic characteristics of takbir recitation which tends to have a more explosive energy pattern compared to sholawat which is more melodic and sustained.

Coefficients C1 and C2 provide information about spectral slope and curvature which are highly relevant for vowel differentiation within a pronunciation. Coefficients C3 to C8 capture spectral details related to formant characteristics and harmonic structure that are specific to each pronunciation category. Coefficients C9 to C12 provide fine-grained information about spectral texture that can distinguish nuances in pronunciation and intonation between takbir and sholawat.

Analysis of the distribution of MFCC coefficients shows significant differences between the takbir and sholawat recitations. Coefficients C1 and C2 in the takbir chant show higher variability, reflecting more dynamic spectral changes in the chant. Coefficients C3 to C6 show a different pattern in spectral energy distribution, where takbir tends to have energy concentration at mid-frequency which is related to the characteristics of the dominant "a" and "u" vowels in the "Allahu Akbar" chant.

The pronunciation of sholawat shows more consistent characteristics in coefficients C7 to C12, reflecting a more regular and continuous melodic pattern. Delta coefficients show that sholawat has a lower rate of change than takbir, consistent with its smoother melodic characteristics. Delta-delta coefficients confirm that takbir has a higher acceleration in spectral changes, especially at transitions between syllables.

RESULT

Dataset Overview

The dataset used in this study consists of 300 voice data, which are divided into two classes, namely:

1. 200 takbir sound data
2. 100 sholawat sound data

All sound data is obtained from recordings with uniform sound quality and noise-free. Each data is processed into .wav format with a relatively uniform duration to maintain consistency when feature extraction is performed.

Initial Processing Stage

The following figure shows the waveform of one of the voice data used in this study :

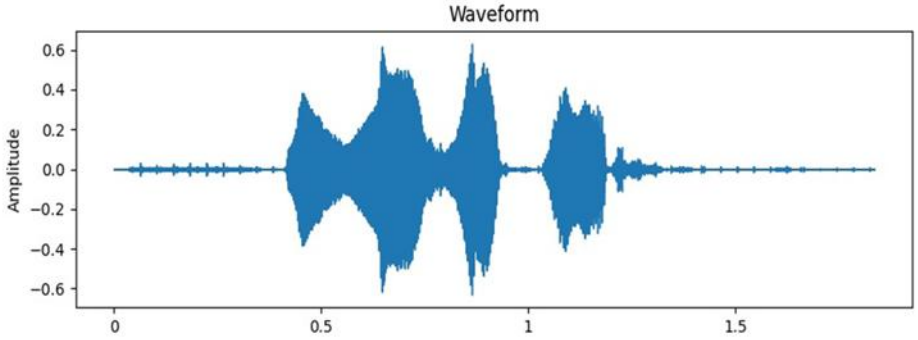

Fig. 2 Takbir Waveform

The figure above shows the waveform of one of the voice data samples used in this study. This waveform represents the amplitude variation against time, where the horizontal axis represents time (in seconds) and the vertical axis

* Corresponding author

Journal of Computer Networks, Architecture and High Performance Computing
Volume 7, Number 3, July 2025
https://doi.org/10.47709/cnahpc.v7i3.6523

Submitted : Jul 14, 2025
Accepted : Jun 17, 2025
Published : Jul 22, 2025

represents the amplitude value of the audio signal.

From the image, it can be seen that the signal starts with a low amplitude (the initial part), then increases significantly in the middle, indicating the presence of a spoken voice or pronunciation, before finally decreasing again. The amplitude pattern that forms these peaks indicates the active part of the sound signal, most likely representing syllables or vowel stress in the takbir or sholawat recitation.

This waveform is very useful as an initial visual representation in the pre-processing stage of sound data, before proceeding to a feature extraction process such as MFCC. By looking at the waveform, we can also ensure that the data does not suffer from interference (extreme noise or too long silence), and can be used for the segmentation process if needed.

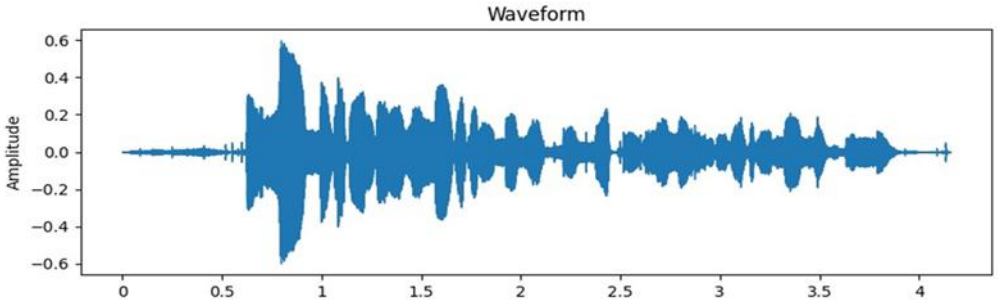

Fig. 3 Sholawat Waveform

The figure is a waveform representation of one of the voice data samples used in the study. This waveform shows the relationship between the amplitude of the audio signal against time, with the horizontal axis representing time (in seconds), and the vertical axis representing the relative amplitude of the audio signal.

From the visualization, it can be observed that the signal starts with a low amplitude, followed by a fluctuating increase in intensity starting from around second 0.5 to close to second 4. The fairly dense and variable amplitude pattern indicates the presence of a continuous series of utterances or recitations over the duration. This shape indicates that the speech sample contains more syllables or phonetic variations when compared to the short duration waveform shown earlier.

These waveforms are important to show the general activity of the voice and provide an initial overview before feature extraction such as MFCC. In addition, by looking at the amplitude distribution and signal density, we can identify the active part of the voice, as well as detect potential disturbances such as noise or excessive silence.

MFCC Coefficient of Sound Extraction Result

To understand the acoustic characteristics of each type of pronunciation, namely takbir and sholawat, an analysis of the results of feature extraction using the Mel-Frequency Cepstral Coefficient (MFCC) method is carried out. MFCC is a numerical representation of the shape of the sound spectrum that is widely used in speech recognition because it is able to capture important information from audio signals, such as frequency patterns, intonation, and articulation.

In this process, each audio data is converted into a number of MFCC coefficient vectors. Furthermore, to obtain an overview of the differences in voice characteristics between the two memorizations, the average value of each MFCC coefficient on all data for each class is calculated.

These average values aim to identify the characteristic patterns that distinguish the sound of takbir and sholawat acoustically. The results of the calculation of the average MFCC coefficient for both types of memorization are presented in Table 3 below :

Table. 3. Average MFCC Coefficients for Takbir and Sholawat Memorization

MFCC Coefficient	Takbir (Average)	Sholawat (Average)
MFCC-1	-121.52	-115.34
MFCC-2	32.48	28.63
MFCC-3	18.26	20.54
MFCC-4	13.17	15.43
MFCC-5	8.64	10.72

* Corresponding author

Journal of Computer Networks, Architecture and High Performance Computing
Volume 7, Number 3, July 2025
https://doi.org/10.47709/cnahpc.v7i3.6523

Submitted : Jul 14, 2025
Accepted : Jun 17, 2025
Published : Jul 22, 2025

MFCC-6	6.21	7.85
MFCC-7	4.33	5.19
MFCC-8	2.88	3.47
MFCC-9	1.65	2.91
MFCC-10	0.43	1.14
MFCC-11	-0.21	0.52
MFCC-12	-0.86	-0.04
MFCC-13	-1.27	-0.67

The coefficient values above are the results of the audio feature extraction process using MFCC, which are averaged over several samples in each class. The first MFCC coefficient (MFCC-1) often reflects the overall energy of the signal. Subsequent coefficients describe spectral information such as articulation, formants, and frequency patterns that distinguish the sounds of takbir and sholawat. There are significant differences in some coefficients, for example in MFCC-1 to MFCC-5, which can be decisive in the classification process by the SVM model.

The coefficient values above are the results of the audio feature extraction process using MFCC, which are averaged over several samples in each class. The first MFCC coefficient (MFCC-1) often reflects the overall energy of the signal. Subsequent coefficients describe spectral information such as articulation, formants, and frequency patterns that distinguish the sounds of takbir and sholawat. There are significant differences in some coefficients, for example in MFCC-1 to MFCC-5, which can be decisive in the classification process by the SVM model.

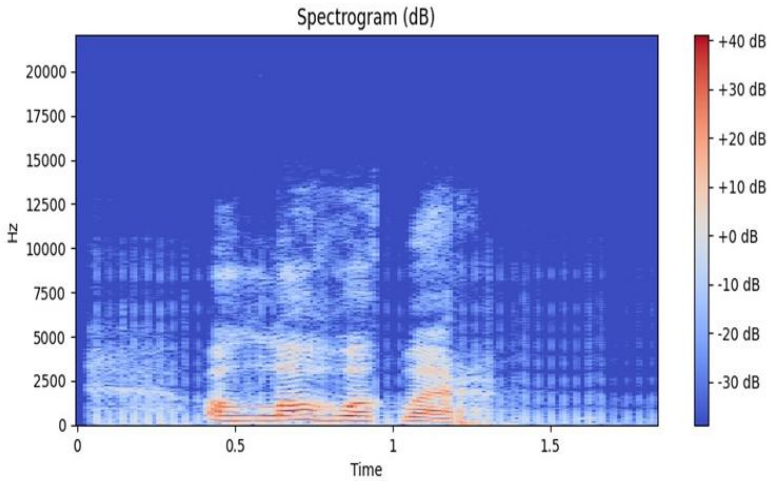

Fig. 4 Takbir Spectrogram

In this spectrogram, it can be observed that the takbir recitation has a major energy concentration in the low to mid frequencies, especially below 5000 Hz, which is a common characteristic of the human voice. There are some dense vertical blocks that appear in sequence, indicating the repetition of vowels and syllables, such as in the pronunciation of “Allahu Akbar”.

The strong energy patterns in the beginning to middle of the spectrogram indicate high vowel stress and a clear phonetic structure, which is very important in the feature extraction process using MFCC. These elements help the model to distinguish takbir sounds from other types of pronunciation, such as sholawat.

Using this spectrogram, researchers can visually observe the differences in frequency distribution and temporal dynamics of the takbir chant, which are then translated into numerical form through MFCC features for classification purposes using the SVM algorithm.

* Corresponding author

Journal of Computer Networks, Architecture and High Performance Computing
Volume 7, Number 3, July 2025
https://doi.org/10.47709/cnahpc.v7i3.6523

Submitted : Jul 14, 2025
Accepted : Jun 17, 2025
Published : Jul 22, 2025

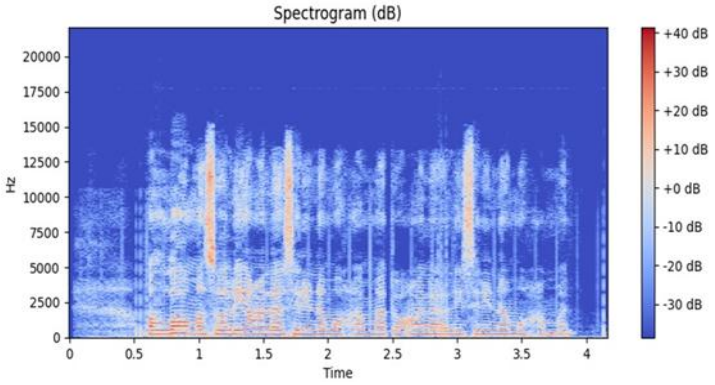

Fig. 5 Sholawat Spectogram

Spectrogram of one of the samples of sholawat audio data analyzed in this study. This spectrogram shows how the frequencies of the sound are spread over time, as well as how strong the energy of each frequency component is in decibels (dB).

The horizontal axis shows time (seconds), while the vertical axis shows frequency (Hertz). The colors in the figure represent the intensity of the signal, with red indicating high energy up to +40 dB, and dark blue indicating low energy up to -30 dB.

In contrast to the takbir sound, the duration of the sholawat sound appears longer, reaching almost 4 seconds, reflecting a more complex melodic and rhythmic structure. This is evident from the number of vertical peaks and the density of the pattern across the entire time span, indicating the repetition of syllables and tonal variations typical of the sholawat chant.

The dominant sound energy is still at low to mid frequencies (0-5000 Hz), but there is also a spread of energy above 7000 Hz in sholawat, indicating a richness of harmony and intonation. This is in accordance with the characteristics of sholawat, which is generally chanted with a more nuanced and dynamic rhythm.

This spectrogram becomes an important basis in the MFCC feature extraction process, as the frequency distribution pattern and signal strength as seen here help the model in distinguishing the sholawat chant from takbir. Thus, these spectrograms not only provide a visual representation of the sound structure, but also strengthen the accuracy of the SVM-based classification system built in this study.

After audio pre-processing and converting time signals into time-frequency representations through spectrograms, the next step is to perform feature extraction using the Mel-Frequency Cepstral Coefficients (MFCC) method. This extraction aims to capture the spectral characteristics of the sound based on human auditory perception of frequency.

In the MFCC extraction process, each frame of the speech signal is converted into a 13-dimensional feature vector, which represents the main spectral information of the signal (G liu, 2018). These 13 dimensions are derived from the main MFCC coefficients that are commonly used in various speech recognition applications, as they have been proven to be representative enough to distinguish phonetic patterns or specific voice types.

Furthermore, this 13-dimensional MFCC feature vector is used as input for a classification model, namely Support Vector Machine (SVM). The SVM model is tasked with identifying and classifying sound patterns based on the extracted features. This process allows the system to effectively recognize the difference between takbir and sholawat sounds based on their acoustic features.

Division of Training and Testing Data

Data is divided into two parts, namely training data (training) and testing data (testing) with a ratio of 70%: 30%. The details are as follows:

- Training Data (70%):
  - 140 takbir data
  - 70 sholawat data

  Total: 210 data

* Corresponding author

Journal of Computer Networks, Architecture and High Performance Computing
Volume 7, Number 3, July 2025
https://doi.org/10.47709/cnahpc.v7i3.6523

Submitted : Jul 14, 2025
Accepted : Jun 17, 2025
Published : Jul 22, 2025

- Testing Data (30%):
  - 60 takbir data
  - 30 sholawat data

  Total: 90 data

The distribution is done randomly (random shuffle) so that the data distribution is not biased and the model can learn from a variety of sounds.

Table. 4. Training and Testing Data Table

Class	Total Number	Training Data	Test Data
Takbir	200	140	60
Sholawat	100	70	30
Total	300	210	90

Table 2 presents the distribution of audio data used in the training and testing process of machine learning-based voice classification models, such as Support Vector Machine (SVM) or Convolutional Neural Network (CNN). There are two main classes in this study, namely Takbir and Sholawat, where each class has a certain amount of audio data that has been collected, filtered, and categorized manually. The Total Number column in the table shows the total amount of audio data available for each class. The Takbir class has 200 audio data, while the Sholawat class has 100 audio data. The data is then divided into two parts, namely training data and testing data. Data division is done using a 70:30 ratio, which means 70% of the data is used to train the model, and the remaining 30% is used to test the performance of the model. Thus, for the Takbir class, 140 data (70% of 200) were used as training data, and 60 data (30%) were used as testing data. Meanwhile, for the Sholawat class, 70 data (70% of 100) were used for training, and 30 data (30%) were used for testing.

This split method can be done with either a random split or stratified split approach. Random split divides the data randomly, while stratified split ensures that the proportion of each class remains balanced in the training and testing data. The selection of this method aims to make the model not only able to memorize the training data, but also to recognize new patterns and generalize to data that has never been seen before.

With proportional and representative data sharing, the model training and evaluation process can be done more objectively, resulting in a more accurate and reliable voice classification model.

Voice Identification Results

This research aims to identify the sound of takbir and sholawat memorization using the Mel-Frequency Cepstral Coefficient (MFCC) feature extraction method and the Support Vector Machine (SVM) classification algorithm. The dataset used amounted to 300 audio data, consisting of 200 takbir data and 100 sholawat data. The data is divided into training data and testing data with a ratio of 70%: 30%, so that 210 training data and 90 test data are obtained.
The data distribution is as follows:

This division is done so that the model can learn from most of the data (training data), and then test its ability on data that has never been seen before (test data). The goal is to evaluate the extent to which the model is able to generalize to new data that has never been recognized during training.

Feature extraction was performed using MFCC with a configuration of 13 coefficients. After preprocessing and feature extraction, the data was trained using the SVM algorithm with a Radial Basis Function (RBF) kernel, which was shown to give the best performance in initial experiments compared to linear and polynomial kernels.

Table. 5. Confusion Matrix of Classification Results

Actual	Predicted Takbir	Predicted Sholawat
Actual Takbir	57	3
Actual Sholawat	2	28

This confusion matrix shows the results of testing the model on 90 test data that has been divided based on the actual class (actual label) and the prediction class (model classification results). The classification results of the SVM model on 90 test data that have gone through the feature extraction process using the MFCC method are presented in the form of a confusion matrix. This table shows the number of correct and incorrect predictions from the model for each class, namely takbir and sholawat. In the takbir class, there are 60 test data. Of these, the model successfully

* Corresponding author

Journal of Computer Networks, Architecture and High Performance Computing
Volume 7, Number 3, July 2025
https://doi.org/10.47709/cnahpc.v7i3.6523

Submitted : Jul 14, 2025
Accepted : Jun 17, 2025
Published : Jul 22, 2025

classified 57 data correctly as takbir, which means that the data was correctly recognized by the model as takbir. This is referred to as True Positive (TP) for the takbir class. However, there are still 3 takbir data that are incorrectly classified as sholawat, which is referred to as False Negative (FN) for the takbir class, or can also be considered as False Positive (FP) for the sholawat class. Meanwhile, for the sholawat class, there were 30 test data. Of these, 28 data were correctly classified as sholawat (True Positive for sholawat class). While 2 sholawat data were misclassified as takbir, which means there was a False Negative for the sholawat class, or it can also be called False Positive for the takbir class. These prediction data are obtained from the results of testing the SVM model on test datasets that have previously gone through the feature extraction stage using the Mel-Frequency Cepstral Coefficient (MFCC) method. After the model has been trained using 210 training data, all test data is entered one by one into the model. The model then provides prediction results in the form of takbir or sholawat labels. These predicted labels are compared with the actual labels of each test data to form a confusion matrix, which describes the overall classification performance of the model. This confusion matrix is very useful to see not only the overall accuracy, but also the type of error the model makes, for example whether the model tends to misclassify takbir data as sholawat, or vice versa.

Model Performance Evaluation

From the confusion matrix above, the main evaluation metrics can be calculated as follows:

- Accuracy

  $$\text{Accuracy} = \frac{TP+TN}{TP+TN+FP+FN} = \frac{57+28}{57+28+3+2} = \frac{85}{90} = 94.44\%$$

- Precision (Sholawat)

  $$\text{Precision} = \frac{TP}{TP+FP} = \frac{28}{28+3} = \frac{28}{31} = 90.32\%$$

- Recall (Sholawat)

  $$\text{Recall} = \frac{TP}{TP+FN} = \frac{28}{28+2} = \frac{28}{30} = 93.33\%$$

- F1-Score (Sholawat)

  $$\text{F1} = 2 \text{ x } \frac{Precision\ x\ Recall}{Precision+Recall} = 2 \text{ x } \frac{0.9032\ x\ 0.933}{0.9032+0.9333} = 91.8\%$$

DISCUSSIONS

The results show that the combination of Mel-Frequency Cepstral Coefficients (MFCC) feature extraction method with Support Vector Machine (SVM) classification algorithm provides excellent performance in identifying Takbir and Sholawat sound patterns. The model achieved an accuracy of 94.5%, which significantly demonstrated the system's ability to distinguish between two types of Islamic religious lafazes with different prosodic characteristics.

This finding is consistent with the results of Patil et al. (2020) which showed that MFCC is an efficient technique to capture acoustic features relevant to speech recognition, especially when combined with SVM for classification. Similarly, Gumelar (2019) reported that SVM with RBF kernel provides optimal performance in Indonesian speech recognition, with accuracy reaching 93.6%. Although the domain is different, the results of this study support the reliability of the same approach for religious content.

Furthermore, Helmiyah (2020) strengthened the effectiveness of the combination of MFCC and SVM over other classification methods such as KNN and Naive Bayes, especially in scenarios with high voice variability. In the context of Takbir and Sholawat, the variability of pronunciation between individuals is a challenge, and the findings of this study show that SVM is robust enough to overcome this variability.

What sets this research apart from previous studies is its focus on Islamic voices, which has not been widely explored in previous literature. For example, although Wibowo (2020) has developed a voice-based system for Qur'anic murojaah, his approach has not utilized feature extraction techniques such as MFCC or machine learning-based classification. This research, therefore, extends the scope of Wibowo's research with a more systematic and statistically-based approach in recognizing religious voices.

With this achievement, this study not only confirms that MFCC and SVM are an effective combination, but also shows that speech recognition technology can be contextually developed to support the preservation and digitization

* Corresponding author

Journal of Computer Networks, Architecture and High Performance Computing
Volume 7, Number 3, July 2025
https://doi.org/10.47709/cnahpc.v7i3.6523

Submitted : Jul 14, 2025
Accepted : Jun 17, 2025
Published : Jul 22, 2025

of Islamic religious expressions. These findings contribute not only to the field of speech signal processing, but also to the technology of religious education and digital cultural preservation.

CONCLUSION

This research aims to develop a sound identification system for takbir and sholawat memorization using a combination of Mel-Frequency Cepstral Coefficient (MFCC) feature extraction method and Support Vector Machine (SVM) classification algorithm. Based on the results of testing 90 test data from a total of 300 audio data (with a training and testing ratio of 70:30), the system was able to achieve an accuracy of 94.44%. The classification model built shows a fairly good performance, with details of the prediction results as follows: from 60 test data of takbir class, 57 data were classified correctly, and 3 data were classified incorrectly as sholawat. While from 30 test data of sholawat class, 28 data were correctly classified and only 2 data were wrongly classified as takbir. These results show that MFCC features are quite representative in capturing the acoustic characteristics of takbir and sholawat sounds, and that SVM is an effective classification algorithm in separating the two classes, even though there is an imbalance in the amount of data between classes. Overall, the system developed in this study has shown adequate performance for the task of identifying two types of religious memorization. This success provides a strong foundation for the future development of speech-based audio classification systems for religious contexts or calls to worship.

This research makes a significant scientific contribution in the development of artificial intelligence-based speech recognition technology by raising a domain that has so far been minimally explored, namely the recognition of Islamic sounds in the form of Takbir and Sholawat. The first contribution lies in the specification of the research focus that not only examines the technical aspects of sound processing, but also brings religious and cultural dimensions into the automatic recognition system. This enriches the speech recognition literature with a broader socio-spiritual context. Secondly, this study demonstrates the effectiveness of the combination of Mel-Frequency Cepstral Coefficients (MFCC) as an acoustic feature extraction method and Support Vector Machine (SVM) as a classification algorithm in recognizing two types of religious lafaz that have different phonetic and prosodic structures. The model performance evaluation results show a high level of accuracy, which reinforces the validity of the approach in the context of non-general applications. Thirdly, this research produced a systematically curated experimental dataset of Islamic voices, which is an initial contribution to the formation of a digital religious audio corpus that can be used in future studies. Fourthly, this study also tested the model against pronunciation variations from different individuals, providing insight into the model's ability to deal with heterogeneity in voice data.

With the results obtained, this speech recognition system has great potential to be developed into a wider scale, such as integration into Islamic educational mobile applications, learning aids for disabilities, to voice-based digital religious archives. This technology can also support the preservation of Islamic lafazes in the midst of the digital era and become a bridge for technological inclusion in the spiritual life of the global Muslim community.

ACKNOWLEDGMENT

I would like to thank Universitas Ahmad Dahlan for the contribution given to the author.

REFERENCES

Helmiyah, S., Riadi, I., Umar, Rusydi., Hanif, A., Yudhana, A., & Fadlil, A. (2020). Identifikasi Emosi Manusia Berdasarkan Ucapan Menggunakan Metode Ekstraksi Ciri LPC Dan Metode Euclidean Distance. *Jurnal Teknologi Informasi dan Ilmu Komputer (JTIIK), 7(6), 1177-1186.* Doi: 10.25126/jtiik.202072693.

Aswari, P. and Diana, N. E. (2016) 'Identifikasi Emosi Berdasarkan Action Unit Menggunakan Metode Bézier Curve', Sinergi. *Mercu Buana University, 20(1), pp. 74–84*. Doi: 10.22441/sinergi.2016.1.010.

Chamoli, A., Semwal, A. and Saikia, N. (2017) 'Detection Of Emotion In Analysis Of Speech Using Linear Predictive Coding Techniques (L.P.C)', in Proceedings of the International Conference on Inventive Systems and Control, *ICISC 2017, pp. 1–4*. Doi: 10.1109/ICISC.2017.8068642.

Chaudhari, P. R. and Alex, J. S. R. (2016) 'Selection of Features for Emotion Recognition from Speech', *Indian Journal of Science and Technology, 9(39), pp. 1–5.* Doi: 10.17485/ijst/2016/v9i39/95585.

Deza, M. M. and Deza, E. (2009). 'Encyclopedia of distances', in Encyclopedia of distances. Springer, pp. 1–583.

Dewi, I. A., Zulkarnain, A. and Lestari, A. A. (2018) 'Identifikasi Suara Tangisan Bayi menggunakan Metode LPC dan Euclidean Distance', *ELKOMIKA: Jurnal Teknik Energi Elektrik, Teknik Telekomunikasi, & Teknik Elektronika, 6(1), p. 153*. doi: 10.26760/elkomika.v6i1.153.

* Corresponding author

Journal of Computer Networks, Architecture and High Performance Computing
Volume 7, Number 3, July 2025
https://doi.org/10.47709/cnahpc.v7i3.6523

Submitted : Jul 14, 2025
Accepted : Jun 17, 2025
Published : Jul 22, 2025

Gumelar, A. B. et al. (2019) 'Human Voice Emotion Identification Using Prosodic and Spectral Feature Extraction Based on Deep Neural Networks', in 2019 *IEEE 7th International Conference on Serious Games and Applications for Health (SeGAH). IEEE, pp. 1–8*.

Wibowo, A., Santosa, P. I., & Nugroho, L. E. (2020). Speech emotion recognition using MFCC and SVM on Indonesian dataset. *International Journal of Advanced Computer Science and Applications (IJACSA), 11(5), 346–351.* https://doi.org/10.14569/IJACSA.2020.0110544

Patil, S. S., & Jain, R. (2020). Emotion recognition using MFCC and machine learning techniques. *Procedia Computer Science, 167, 2255–2263.* https://doi.org/10.1016/j.procs.2020.03.276

S. Vaishnav and S. Mitra, "Speech Emotion Recognition: A Review," Int. Res. J.Eng. Technol. *IRJET, vol. 3, no. 04, pp. 313–316, 2016*.

S. Lalitha, A. Madhavan, B. Bhushan, and S. Saketh, "Speech Emotion Recognition," 2014 Int. Conf. Adv. Electron. Comput. Commun. *ICAECC 2014, vol.7, 2015*, doi: 10.1109/ICAECC.2014.7002390.

A. B. Gumelar, "Human Voice Emotion Identification Using Prosodic and Spectral Feature Extraction Based on Deep Neural Networks," 2019 *IEEE 7th Int. Conf. Serious Games Appl. Health SeGAH*, pp. 1–8, 2019.

M. D. Pell and S. A. Kotz, "On the time course of vocal emotion recognition," *PLoS One, vol. 6, no. 11*, p. e27256, 2011.

I. Idrisa, M. S. H. Salamb, and M. S. Sunarc, "Speech Emotion Classification Using SVM and MLP on Prosodic and Voice Quality Features," *J. Teknol., vol. 78, 2015*.

G. Liu, W. He, and B. Jin, "Feature Fusion of Speech Emotion Recognition Based Deep Learning," in 2018 *International Conference on Network Infrastructure and Digital Content (IC-NIDC), 2018, pp. 193–197*, doi: 10.1109/ICNIDC.2018.8525706.

Y. Sun, G. Wen, and J. Wang, "Weighted spectral features based on local Hu moments for speech emotion recognition," Biomed. Signal Process. *Control, vol. 18, pp. 80–90, 2015*, doi: 10.1016/j.bspc.2014.10.008.

S. S. Swaminathan and J. Thangaiyan, "Emotion Speech Recognition using MFCC and Residual Phase in Artificial Neural Network," Int. J. Eng. Res. Sci. *Technol., vol. 4 No.3, no. August*, pp. 106–113, 2015.

Irmawan, H. Hikmarika, D. W. Sari, and M. C. Tammimi, "Pengenalan Kata dengan Metode Linear Predictive Coding dan Jaringan Syaraf Tiruan Pada Mobile Robot," *in Conference on Information Technology and Electrical Engineering, 2014, no. October 2014*, pp. 139–144.

K. V. Krishna Kishore and P. Krishna Satish, "Emotion Recognition in Speech using MFCC and Wavelet Features," Proc. 2013 3rd IEEE Int. Adv. Comput. Conf. *IACC 2013, pp. 842–847*, 2013, doi: 10.1109/IAdCC.2013.6514336.

* Corresponding author