Jordanian Journal of Computers and Information Technology (JJCIT), Vol. 06, No. 04, December 2020.

AN EFFICIENT HOLY QURAN RECITATION RECOGNIZER BASED ON SVM LEARNING MODEL

Khalid M.O. Nahar, Ra'ed M. Al-Khatib, Moy'awiah A. Al-Shannaq and Malek M. Barhoush

(Received: 29-Jun.-2020, Revised: 17-Aug.-2020, Accepted: 6-Sep.-2020)

ABSTRACT

Holy Quran recitation recognition refers to the process of identifying the type of recitation, among those authorized styles of recitation ("Qira'ah" in Arabic). Several previous studies investigated the recitation rules ("Ahkam Al-Tajweed" in Arabic) that are applied by readers or reciters while reading the Holy Quran aloud, but no study has examined the problem of tracking the type of recitation used in the reading. Through this research, we can assist Holy Quran students to easily learn the perfect and accurate recitation by successfully applying Ahkam Al-Tajweed and help them distinguish between different recitations or "Qira'ah". In this paper, a recognition model is conducted to recognize the "Qira'ah" from the corresponding Holy Quran acoustic wave. This model was built upon three phases; the first phase is the Mel-Frequency Cepstrum Coefficients (MFCC) feature extraction of the acoustic signal and labeling it, the second phase is training Support Vector Machine (SVM) learning model the labeled features and finally, recognizing "Qira'ah" based on this trained model. To attain this, we have built our corpus, which has 10 categories, each of which is labeled as one type of Holy Quran recitation or "Qira'ah". Different machine learning algorithms were applied and compared. Experimental results proved the superiority of our proposed SVM-based recognition model for "Qira'ah" over other machine learning algorithms with a success rate of 96%.

KEYWORDS

Arabic language, Quran recitation, Artificial Neural Network, Support Vector Machine (SVM).

1. INTRODUCTION

The key principle of communication is mainly to exchange ideas among peers and friends. People generally communicate and understand each other through speech. However, this might prove difficult for some people due to the wide variety of languages spoken globally. Nowadays, many computer applications in the area of computational linguistics have been designed to take into consideration the problem of recognizing and translating spoken language [1]. The productivity of such software applications statistically enhances and enriches the many disciplines that exist within the natural language processing field. In the last decade, computer scientists have paid special attention to developing efficient algorithms to recognize spoken words in the Speech Recognition (SR) domain. Progress in this domain has been significant and it is now widely known as Automatic Speech Recognition (ASR) technology [2]-[3].

ASR has been developed to recognize voices, which can improve communication among humans. These applications of ASR can compensate for the difficulties which are caused by the existence of such a wide range of languages in the world [4]. Therefore, several techniques have been introduced in the area of ASR [5]-[6]. The Support Vector Machine (SVM) is a powerful technique that has been widely used for speech recognition [7]-[8]. An ASR-based system recognizes spoken words by detecting and analyzing the input voice in a waveform, as illustrated in Figure 1 which shows the framework of ASR.

In this research, we employed a speech recognition model for recognizing "Qira'ah" within the readings from the Holy Quran. We focused on adapting and deploying the SVM algorithm with a new data corpus and special attributes as well as new rules. The proposed model initially converts the reader's sound waves (i.e., the proposed data corpus) into MFCC features and then a features vector matrix is generated. Parts of the extracted features are utilized to train the adapted SVM-based algorithm. The trained SVM developed was tested again with the other part of the extracted features and yielded very promising

K. M. O. Nahar, R. M. Al-Khatib, M. A. Al-Shannaq and M. M. Barhoush are with Computer Sciences Department, IT Faculty, Yarmouk University, Irbid, Jordan. Emails: khalids@yu.edu.jo, raed.m.alkhatib@yu.edu.jo, moyawiah.s@yu.edu.jo and malek@yu.edu.jo

results. As far as we know, this is the first time that SVM has been used in recognizing features of Holy Quran recitations. Our proposed SVM-based model was evaluated using real-world data collected from famous reciters of the Holy Quran. The experiments showed very fruitful outcomes when analyzing and comparing the obtained results. For comparative evaluation, the results obtained from our proposed SVM-based model were compared with other well-known algorithms, using the same datasets of waves collected. Interestingly, our proposed SVM-based model outperforms other algorithms in terms of accuracy in almost all experiments.

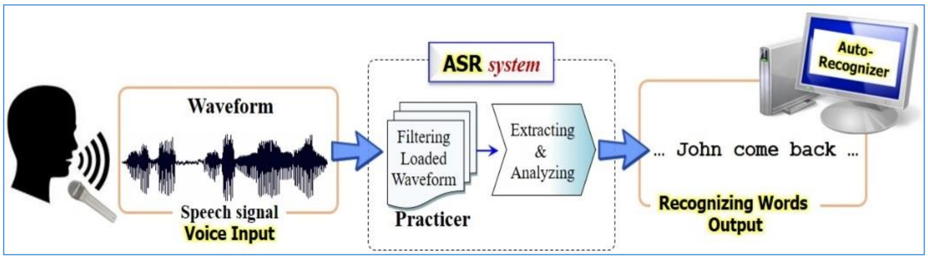


Figure 1. A general framework of Automatic Speech Recognition (ASR).

The rest of this paper is organized as follows. In Section 2, the literature review section, previous studies related to the current study are discussed and analyzed. In Section 3, we present the background of different speech recognition algorithms. The proposed approach and the methodological steps are fully presented in Section 4. Section 5 illustrates the experimental results. Finally, in Section 6, we show the conclusions of this study and propose several recommendations for further studies.

2. Literature Review

This section presents a basic description of a Speech Recognition System (SRS). It also provides more information about previous speech recognition systems used for Arabic.

2.1 Speech Recognition System (SRS)

Automatic Speech Recognition (ASR) enables the computer to identify the utterances of a person speaking into a microphone or telephone. Human Computer Interaction (HCI) is used, for example, as an authentication technique for user login *via* a voice recognition tool. Some ASR applications include voice interface as a command recognition application for computer users, dictation and written-text correction, interactive communication and voice response as an aid in learning foreign languages and for voice-controlled operation of machines. Additionally, ASR technology can improve the quality of life for disabled people, allowing them to communicate with others and interact in society [8]. In practice, the main principle of an ASR system is to recognize the appropriate word patterns for spoken utterances by applying a digitized analyzing process to enter analog sound waves [9]-[10].

2.2 Arabic Speech Recognition

One of the oldest Semitic languages in the world is the Arabic language. It is officially the sixth most spoken language and one of the official languages of the United Nations. There is an official Arabic linguistic form known as Modern Standard Arabic (MSA), which is mainly used in formal media, courtrooms, offices and by instructors for teaching in schools and universities [11]. Recently, many ASR systems have been developed in the domain of Arabic Speech Recognition to recognize the MSA version of Arabic. Unfortunately, recognizing traditional Arabic is still a challenge due to its lexical variety and the scarcity of data. Besides, the Arabic language is considered one of the most complex languages due to the morphological variations of its letters [12], [10].

ASR Arabic language research is still in its infancy age compared to the ASR already used in research related to other languages [13]-[14]. Consequently, we will review the top five studies that have been developed to improve Arabic speech recognition. In [15], the author dealt with continuous Arabic speech recognition, addressing the labeling of Arabic speech. Another model for Arabic speech recognition focused on prominent problems in recognizing conversational, dialectal and colloquial Arabic speech [13]. The authors reported significant improvement with respect to word error rate according to the 1997

NIST benchmark evaluations. An Arabic ASR system using ANN techniques was developed to improve the Arabic automatic recognition process [15]-[16]. Another Arabic ASR based on Hidden Markov Model (HMM), SVM or a hybrid of both was also developed [14], [17]. The last area relates to the work of some Arabic ASR researchers who took into consideration pronunciation variations to improve the performance of Arabic ASR systems [18].

2.3 Holy Quran Recitation Recognizers

Computer-aided Pronunciation Learning (CAPL) was considered early in the twentieth century when great efforts were made and improvements achieved by researchers. Recognizing the Holy Quran Tajweed rules and tracking reading errors presented a great challenge [19]. The authors in [20], developed an intelligent Tajweed-rules tracker system. The system listens to the Holy Quran, recited by a learner and then suggests a correction to his/her recitation. The name of the system they developed is “Hafize”. The “Hafize” system is trained to recognize the recitations of 10 reciters, including women, men and children. The accuracy of “Hafize” recitation was measured against each reader’s recitation with the average result in the region of 89%. The “Hafize” system has a limitation in that it relates to a pronunciation based on phonetic rules which are not given. “Hafize” does not only consider the verse as a whole, but it can also recognize mistakes at word level. The authors in [21], developed another recitation model to help Malaysian primary school students pronounce the Quran verses correctly. Unfortunately, we have not found sufficient information about the implementation of this model to be able to test it. A new recitation system named “Makhraj” was developed by [22] to make the recitation of the Holy Quran less dependent on expert reciters. The accuracy of this model was calculated based on the False Rejection Rate (FRR). The authors used MFCC for feature extraction. Two modes for recognition process are used in the Makhraj-based system: the one-to-one mode and the one-to-many mode. However, the one-to-one mode is 98% accurate, which is not considered very accurate in this mode due to the utilization of a simple matching technique.

The authors in [23] introduced an SVM-based learning model, which recognizes Quranic words from online resources. In [24], the researchers implemented a virtual learning system called “Electronic Miqra’ah”. The system can independently receive voice commands and allows blind students to interact with the voice of Holy Quran scholars. However, the system has a low recognition rate due to the use of the Google API. Nonetheless, as a model, it works well. In [25], the authors developed a high-performance phoneme-based and speaker-independent system for Holly Quran recitation based on HMM with 3-emitting states. The accuracy of this system reaches 92%. In [26], the authors introduced a powerful training system for Quran recitation using a continuous Arabic speech recognition system that depends on HMM to recognize Quran recitation and detect recitation errors. Their Arabic speech recognition system provides phonetic time alignments and its classifier is used to distinguish between confusing phones. This Arabic speech recognition system was built on the results produced by WEKA tool. In this system, The SVM classifier was used with an accuracy rate of 91.2% at word level.

In [27], the authors used the CMU-Sphinx4 tool to produce a new recitation recognition system for the Holy Quran based on an HMM algorithm. In this HMM-based recognition system, a simplified set of phonemes was used to build the language model, as well as to train the recognizer. The authors in [28] developed a system for tracking Holy Quran basic Tajweed rules using deep learning techniques. They used MFCC, WPD and HMM-SPL feature extraction algorithms and considered them as the best features. Moreover, they reported that their system reached a 97.7% rate of accuracy. A limited number of Quranic chapters were tested and the accuracy was high, reaching 98%. Table 1 summarizes the studies that have been presented in this section.

Table 1. Summary of Holy Quran recitation research.

AUTHORS & DATE	SYSTEM DESCRIPTION & REF.S	Strengths	Weakness	Feature Extraction Algorithm
Muhammad et al. (2012)	Intelligent Tajweed-rules tracker system (Hafize) [20]	It can discover mistakes in the recitation of verses from the Holy Quran.	It did not work at the phonetic level.	MFCC

Mssraty & Faryadi (2012)	A recitation model to help Malaysian primary school students [21]	An initial analysis denotes its potential usefulness.	It has not yet been implemented.	N/A
Arshad et al. (2013)	A recitation system: "Makhraj" [22]	It produces a good level of accuracy on one-to-one mode.	Accuracy level is low in the one-to-many mode due to the simple match technique used.	MFCC
Sabbah & Selamat (2014)	A learning model based on SVM to recognize Quranic words [23]	Good recognition accuracy.	The sparsity of the feature matrix; and the number of features increases the time for building the classification model.	Statistical Features
Mohamed et al. (2014)	A virtual learning system: (Electronic Miqra'ah) [24]	Robust application.	Moderate recognition rate due to the use of Google API.	N/A
Elhadj et al. (2014)	Phoneme-based speaker-independent system for Holy Quran recitation [25]	Good accuracy level reaching 92%.	N/A	MFCC
Tabbaa & Soudan (2015)	Quran recitation based on HMM and SVM classifier [26]	Accuracy reached 91.2% at the word level.	It suffers from high confusing sounds.	MFCC
El Amrani et al. (2016)	Limited Holy Quran recitation based on HMM model [27]	Accuracy reaches 98%, but limited corpus.	Not all Arabic phonemes are included in the recognition.	MFCC
Al-Ayyoub, Damer & Hmeidi (2018)	Verifying the proper use of Tajweed-rules of Holy Quran [28]	Accuracy of the system reached 97.7%.	It only tackled the basic Tajweed rules without using correct recitation verification.	MFCC, WPD & HMM-SPL

Despite the examples given in the literature review, only limited attempts have been made in Quranic recitation recognition to date. Thus, the automatic identification and recognition of aspects of a Quran recitation is still a fresh field for study. In this paper, a new Quranic recitation recognition model is proposed based on the SVM learning algorithm, which is directed at recognizing and detecting aspects of Holy Quran recitation using what we consider to be an enhanced approach.

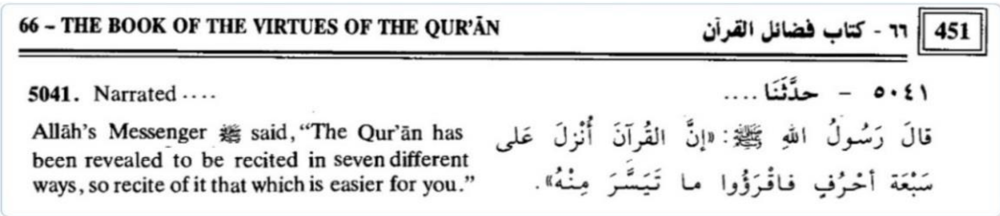
3. BACKGROUND

In this section, the types of Holy Quran reciters are discussed. This section also presents an overview of the most outstanding learning algorithms.

3.1 Types of Holy Quran Reciters Based on Narration "Qira'ah"

In various Arab communities, there are many different regional accents,"lahjah" "لهجة" in Arabic. These numerous accents are seen as a challenge for the ASR-based applications, which need to recognize the correct accent among multiple local accents. It is noticeable that some of these applications are unfamiliar with local accents [30]-[31]. Arabic is one of the most prevalent languages in the world, because of its number of speakers and because the Holy Quran was given through divine revelation in the Arabic language [32]. As a result, not only do the Arab communities in 23 Arabic countries speak Arabic, but also all non-Arabic speakers who share this faith aim at learning Arabic to recite the Holy Quran correctly [33]-[34]. Their main goal is to understand and read the Holy Quran properly, based on the officially established rules of readings [35]. Unlike any other subject of voice or speaker identification systems, the way the reader recites Holy Quran is different from other types of speech. The difference resides in the existence of certain acoustic rules, which are known as “Ahkam Al-Tajweed” (“أحكام التجويد”), previously mentioned, that have to be applied and maintained when reading the Holy Quran. Furthermore, emotional features are added by the reader. Moreover, the transition from one acoustic level to another when reading also distinguishes this type of reading or recitation from normal speech. The transition from one acoustic level to another in the Arabic language is known as “Maqam” “مقام” [19].

The Holy Quran has seven main designated styles of reading or “Qira’at” “قراءات”, which are accepted as the most popular ways of reading or recitation extracts from the Holy book. This is what is stated by Prophet Mohammed in the 4th century AH, according to the narrative “hadith” “حديث” No. 5041, extracted from [36], as shown in Figure 2 [37].


66 – THE BOOK OF THE VIRTUES OF THE QUR’ĀN ٦٦ - كتاب فضائل القرآن 451

5041. Narrated …. ٥٠٤١ - حدَّثنا ....

Allāh’s Messenger ﷺ said, “The Qur’ān has been revealed to be recited in seven different ways, so recite of it that which is easier for you.”

قالَ رَسُولُ اللهِ ﷺ: «إنَّ القُرآنَ أُنزِلَ على سَبْعَةِ أحْرُفٍ فاقْرَؤُوا ما تَيَسَّرَ مِنْهُ».

Figure 2. Extracted from the Hadith 5041 (the rendering of the quotation by Sahih Al-Bukhari).

Subsequently, three new reading styles were added to the seven “Qira’at”. The first seven of the accepted readings are known in the Arabic language as “Mutilator Qira’at”, meaning "successive readings", in Arabic “قراءات متواترة” [38]. The last three readings are called “Mashhur Qira’at” “قراءات مشهورة”. Here, the term “Qira’at” “قراءات” relates to the different recitations that represent changes in Holy Quran reading styles that occur mainly in the areas of pronunciation and tone uttering Quranic extracts. In the science of reading of the “Qira’at”, the ten styles of reading or recitation are attributed to the original readers known as “Imam” “إمام” and designated by the name of the Imam. The most popular “Qira’at” variants are the “Asim” “قراءة عاصم” and the “Nafi” reading styles “قراءة نافع”. All recitations are listed in Table 2. The differences in the recitation styles can be limited to three main distinctions:

a) The prolongation “Mudud” “مدود” and the shortening of words called “Kasser” “قصر”.
b) Adding the punctuation (in Arabic “Harakatt” “حركات”) of the written text of the Holy Quran.
c) The pronunciation and reading of the Quranic extracts and parts, depending on the varying styles of recitation [21][8].

Many researchers have addressed the various challenges people meet in approaching the Arabic language, in addition to the intricacies of the Holy Quran and the way they should recite its verses. One of the main drivers of this work is to facilitate the recitation by non-Arabic speakers [21][39]. It is recognized that beginners face many difficulties in reciting, distinctly for the following three reasons:

a) They lack oral practice and monitoring for correction of errors.
b) Many learners are unfamiliar with Arabic, but the revealed text is in the Arabic language.
c) Most software related to Tajweed lacks any follow-up for readers to improve in their future recitation.

As previously stated, some of the difficulties readers find in the reciting aloud of the Holy Quran are the result of the need to base their reading on strict rules of recitation “Ahkam Al-Tajweed” [35][10]. The science of “Tajweed” teaches the reciter the basic rules that help him/her to pronounce the words of the Holy Quran as Prophet Muhammad, peace be upon him, recited them. A teacher of Tajweed must be authorized and certified to do so.

Since the teaching of the correct pronunciation of the words and the appropriate rendering of Quranic verses is central to Islam, nowadays we have the opportunity to check the accuracy of the recitation automatically using applications related to ASR systems as noted earlier. Recitation from the Holy Quran is not like any other type of reading. The rules of recitation (“Ahkam Al-Tajweed”) must be followed to be a faithful rendering of the verses. Because these rules of pronunciation are followed by multiple speakers, distinguishing between different speakers is not easy and becomes an important issue for researchers. However, no study has covered the area comprehensively until now.

Different approaches in the literature use MFCC features along with some learning algorithms, such as Hidden Markov Model (HMM), Support Vector Machine (SVM) and Artificial Neural Network (ANN), in an attempt to recognize a voice among a set of different voices [14], [17], [27]-[28] . A recent investigation was conducted to build a recognizer to identify the different types of Holy Quran reading “Qira’ah” [37]. However, the authors built a limited corpus of only two types of reading styles: “Eldori” “قراءة الدوري” and “Hafs” “قراءة حفص”. In addition to that, the research depends on one reciter, who recites using the two styles of recitation in the study. Another limitation of the study is that only three chapters (“Surahs” “السور”) were included in their corpus. We believe that these limitations may not produce objective results.

3.2 Artificial Neural Network (ANN)

The Artificial Neural Network (ANN) is inspired by the human biological nerve system [40]. ANN is a collection of many artificial neurons that are connected together for the purpose of learning. The main purpose of ANN is to map the inputs into meaningful outputs. To illustrate this further, ANN could be found in many topologies, such as Feed Forward and Back Propagation. Figure 3 shows the architecture of a feed-forward ANN.

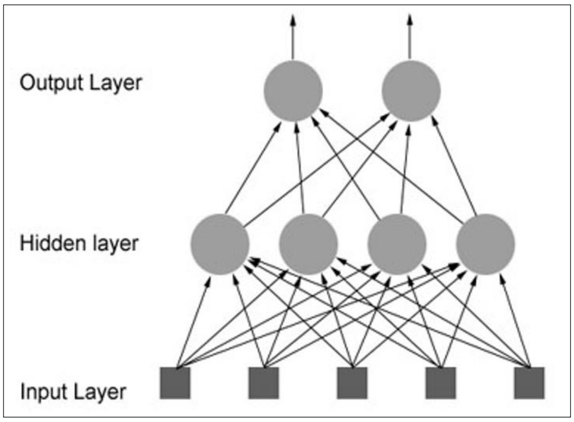


Figure 3. Feed-forward artificial neural network.

Each input from the input layer feeds each node in the hidden layer and then the hidden layer feeds the next layer until it reaches the output layer. In most cases, ANN consists of multiple hidden layers that must be passed through before ultimately reaching the output layer. The ANN in Figure 3 is called a feed-forward ANN, due to the fact that signals are passed through the layers of the neural network in a single forward direction. However, the ANN can be feedback networks, where the architecture allows signals to travel in both directions [41].

3.3 Support Vector Machine (SVM)

The Decision Tree, Radial Bases, Forest Decision Trees, Nearest Neighbor, Fuzzy Classifier, Deep Learning Classifier and Support Vector Machine are some of the well-known machine learning algorithms used in the literature for classification and categorization [40], [42], [46]. The most significant of these algorithms is SVM, which is considered the best of learning algorithms [42].

SVM is a common classifier which separates instances through the use of a hyperplane. In supervised learning, SVM produces an optimal straight line that separates categories, as shown in Figure 4. In essence, the SVM algorithm finds a decision boundary with maximum margins between categories, because the optimal separating hyperplane maximizes the margin of the training data [43]. The SVM is a built-in function in a variety of software. The Sequential Minimal Optimization (SMO) algorithm is an SVM implementation of WEKA open source software, which was implemented by [44]-[45]. The SMO is one of the most efficient solutions for the SVM algorithm. It is based on solving a series of small quadratic problems, which in each iteration uses only two variables in the working set in order to save time [46].

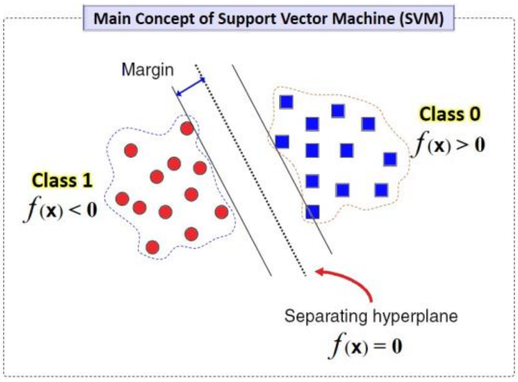


Figure 4. SVM classification hyperplane.

The SVM as a classifier is considered to be one of the most powerful statistical learning techniques [46][42]. SVM successfully addresses and solves different kinds of problems, for three key reasons. The first reason is that the tangent of SVM which is seen in its mathematical and theoretical foundations supports classification and problem-solving. The second reason is that SVM has been confirmed as being suitable to manage complex data, including data of high dimensions, such as text and image data [40][14]. The third one is that SVM has the potential for success in the pattern recognition domain and its effectiveness has already been confirmed in the image processing field [47]. In brief, the SVM consists of class classifiers, which are mathematically constructed from the summation of a kernel function as stated in the following equation:

$$f(x) = \sum_{i=1}^{N} \alpha_i t_i K(x, x_i) + d \,, \tag{1}$$

where $K(x, x_i)$ constructs the summation of the kernel function modeled by [48]. t_i denotes the ideal outputs when $\sum_{i=1}^{N} \alpha_n t_i = 0$ and $\alpha_i > 0$. From the training set of optimization process in [49], x_i denotes the support vectors. The ideal and optimal results may be 1 if the corresponding support vector is in class 0 or may be -1 if the corresponding support vector is in class 1. In terms of classification, the class decision happens when the value of $f(x)$ is above a specific threshold or below it. The kernel function $K(.,.)$ is constrained and limited to specific properties, which are known as the Mercer condition and can be expressed as follows:

$$K(x, y) = b(x)^t . b(y) \,, \tag{2}$$

where $b(x)^t$ implies a mapping according to the input space. Here, x indicates what could possibly be an infinite dimensional space. Finally, the Mercer condition here is responsible for guaranteeing that the validation of the margin concept and the optimization of the SVM limited to definite and particular boundaries [50].

Specifically, the optimization condition depends on a maximum margin concept, as depicted in Figure 4. The SVM chooses an appropriate high-dimensional space to put in place the best hyperplane that has the maximum margin. As a result, the training of the input data points set will be located on the boundaries of the support vectors that are based on Equation (1). These boundaries are represented by two solid lines, as shown in Figure 4. Modeling of these two boundaries is the main aim of the SVM training process.

3.4 Hidden Markov Model (HMM)

The HMM model makes a chain called a Markov chain usually used in stochastic processing [51]. The HMM has the capability to heuristically address the variability using such stochastic modeling. Furthermore, HMM model was efficiently used to improve the behavioral performance of metaheuristics [52]. Since the HMM is a time series learning algorithm, it is not adapted to the comparative outputs to our problem and consequently not used as a model in this research.

In this paper, after due consideration, it was decided to use the SVM learning algorithm, because SVM generates the hyperplane that classifies the training instances with high speed and more accuracy than other traditional clustering methods. Those other traditional clustering methods mainly depend on probability distributions when training data is classified as has been demonstrated by [53].

4. Proposed SVM-based Approach

The proposed approach is mainly based on the feature extraction process and recitation modeling. Figure 5 illustrates the proposed reading/recitation recognition system, where some parts of this figure are adapted from [54]. Initially, the proposed system is trained and tested using SVM, ANN, among others. After building the corpus of recitation of the Holy Quran, the proposed recognizer is built by applying three phases. The first one is to use the Mel-Frequency Cepstrum Coefficients (MFCCs) algorithm to extract a range of informative media features from the trained corpus. The second one is to formulate a matrix of training features and build a learning model across the SVM learning algorithm. Lastly, testing and comparing the results obtained from the previous phase to show the superiority and relevance of the SVM algorithm to the problem of Quran recitation recognition. In the testing phase, some external data (data totally outside the corpus) will be used to test and evaluate our proposed SVM-based approach. The following subsections show more details about our methodology.

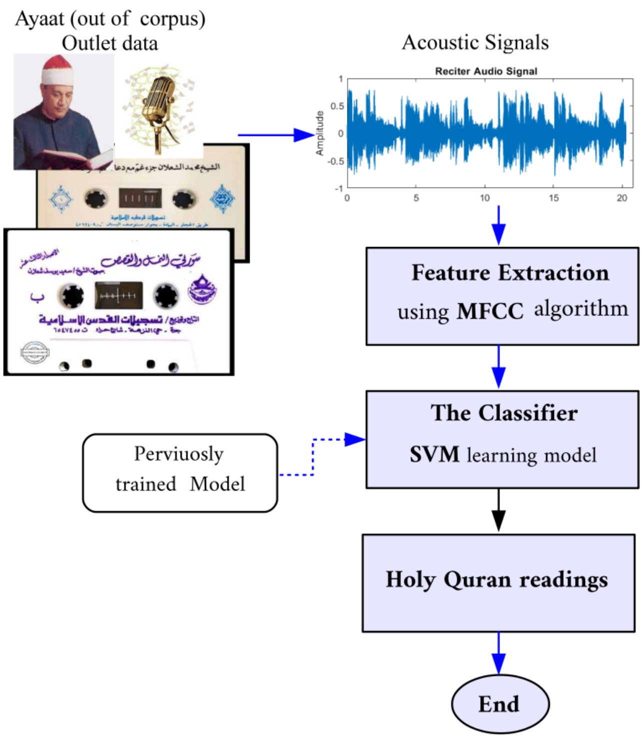


Figure 5. Overview of proposed SVM-based approach.

4.1 Building Corpus

Since the proposed approach aims to identify Holy Quran reading types, which are well-known as "Qira'ah" "قراءة", acoustic samples from Holy Quran need to be collected in order to build a relevant corpus. The corpus contains a number of acoustic waves that are labeled based on the "Qira'ah". A pictorial view of an acoustic wave and its features are illustrated in Figure 6 and in Figure 7, respectively. Acoustic waves for each reading were collected and the feature vectors of these acoustic waves were extracted using the MFCC extraction algorithm. Figure 6 represents a sample of the acoustic wave used in building up the corpus along with its full energy spectrum. Each acoustic wave has its own phase (starting wave angle), amplitude and frequency, but all of them are limited within a specific range and duration.

Figure 7 represents a sample of the MFCC features and below is the spectrum of the acoustic wave after removing weak energy. Almost all spectra of all waves are normally distributed. In the spectrum, each feature has a value which represents part of the wave attributes.

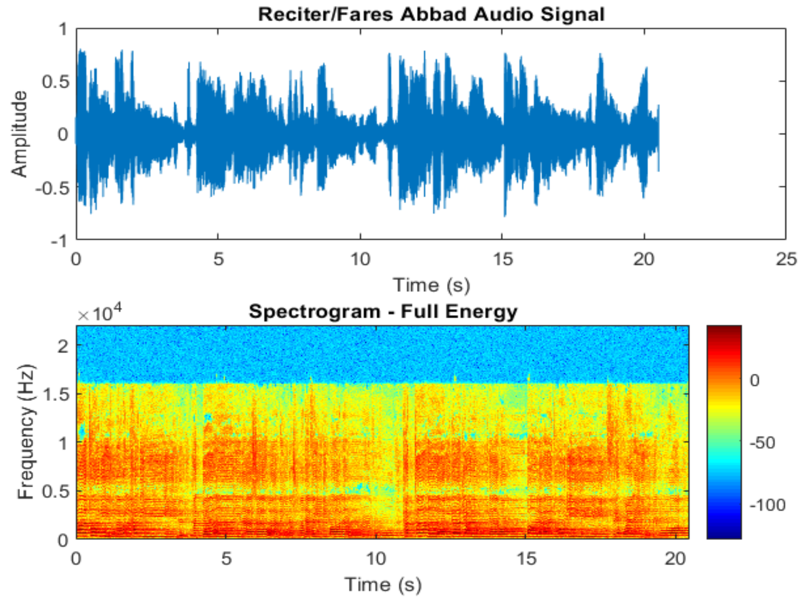


Figure 6. Sample of acoustic wave with its full spectrum [54].

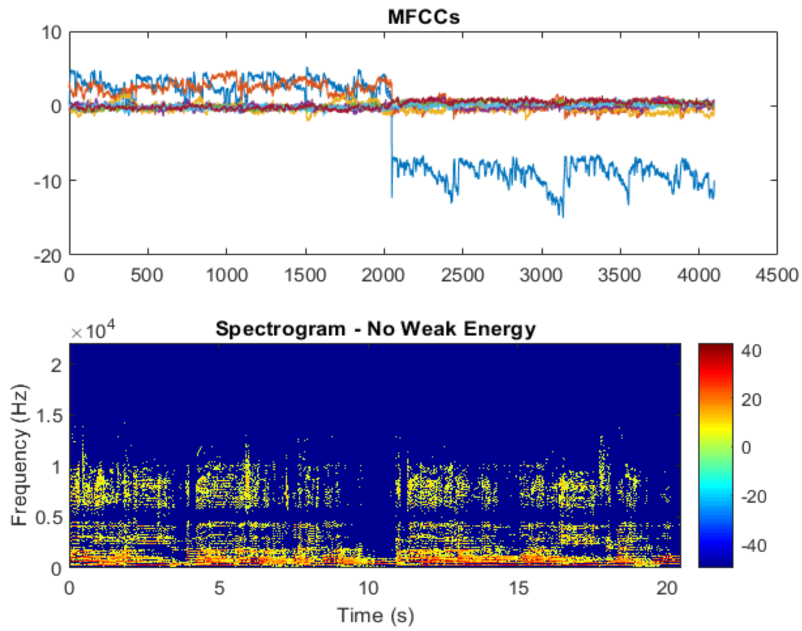


Figure 7. Sample of MFCC file without weak energy spectrum [54].

4.2 Orders and Categories for Labeling Reciters

This sub-section introduces details of the data acquisition and corpus construction related to our proposed method. In order to build a recitation recognition system for Holy Quran “Qira’ah”, the main features of the “Qira’ah” waves should be isolated from those of other waves. In brief, the Holy Quran is a collection of verbal revelations given to Prophet Muhammad over a period of twenty-three years. The Holy Quran has 114 Chapters (“Surah” in Arabic “سورة”) of varying lengths and each Chapter “سورة” consists of a number of individual verses “Ayaat” “آيات”. There are 6,348 different verses in the Holy Quran. To date, there is no corpus for Holy Quran verses “Ayaat” “آيات” based on the style of reading or recitation or “Qira’ah” “قراءة”. Therefore, in this research, we have collected all possible “Qira’ah” waves and placed them in a folder, regardless of the identity of the reader. Ten different folders have been created to represent ten different types of reading (“Qira’ah”) waves. Table 1 shows four columns. In the first, the name of the Imam who established the specific “Qira’ah” is mentioned in

Arabic. In the second column, the English translation of the imam name is given and it is abbreviated in the third column. In the last column, the number of waves that were collected with respect to each "Qira'ah" is given. It should be mentioned here that the collected wave files are available and can be downloaded from the official website of our proposed method.

Ten different types of reading ("Qira'ah") were used in the construction of the corpus, as detailed in Table 2. Each Holy Quran reading type has a different number of wave files. As a result, we have a total of 258 wave files. Holy Quran readers often used the same styles of reading or recitation and rarely use the other available styles. Therefore, some types of reading are represented by a large number of sound waves, while others do not.

Table 2. Reading types, code and number of audio files.

The Imam of the "Qira'ah"	Translated title in English	Key-Class	Number of Wave Files
ابن عامر	Ebin-Amer	EA	6
ابن كثير	Ebin-Khatheer	EK	6
ابو جعفر المدني	Abee-Jafar-Almadani	AJ	18
ابو عمرو البصري	Abee-Amro-Albasree	AA	51
الكسائي	Al-Kesaee	K	51
حمزة الكوفي	Hamzah-Alkofee	HK	18
خلف العاشر	Khalaf-AlAsher	KH	3
عاصم الكوفي	Aseem-Alkofee	AK	48
نافع المدني	Nafee-AlMadani	NM	51
يعقوب الحضرمي	Yakoob-Alkhathramee	YKH	6
Ten "Qira'ah"	Total		258 files

4.3 Extracting Features and Building the Feature Vectors

Many previous studies have been carried out in converting sound waves, including the extraction of statistical information from acoustic signals. These studies have resulted in many valuable methods for interpreting the wave to provide information that could be processed easily. A comparative study between those methods was performed by [55]. The methods that were examined in Shrawankar's study include: Linear Predictive Coefficients (LPCs), Linear Predictive Cepestral Coefficients (LPCCs), Perceptual Linear Predictive Coefficients (PLPs), Mel-frequency Cepestral Coefficients (MFCCs), Mel

Table 3. Ranking of feature extraction methods.

Feature Extraction Function	Best Classification Algorithm	Accuracy
MFCC	SVM	89.16%
LPC	Random Tree	79.15%
LPCC	Random Tree	76.99%
LAR	Decision Table	83.83 %
SSC	Function, Logistic	66.05%
LSF	Random Tree	82.54%
PLP	Decision Tree	71.23%
FFT	ANN	78.34%
MEL	SVM	87.61%
RASTA	ANN	85.23%
DELTA	Random Forest and SVM	82.50%

Jordanian Journal of Computers and Information Technology (JJCIT), Vol. 06, No. 04, December 2020.

Scale Cepestral Analysis (MEL), Relative Spectra Filtering of Log Domain Coefficients (RASTA), First Order Derivative (DELTA), Perceptual Linear Prediction (PLP), Fast Fourier Transform (FFT), Line Spectral Frequencies (LSFs), Spectral Subband, Centroid (SSC) and Log Area Ratio Coefficients (LARs) [55]. Finally, a comprehensive analysis and wide investigation proved that the MFCC method produces a small number of coefficients that represent frequency information [55]. It replicates the perception of loudness in the human auditory system and is considered as a simplified auditory model that eases computation and is relatively faster than others.

Before proceeding to our experimental part, we undertook a brief investigation regarding the accuracy that could be gained from each type of feature extraction mechanism. Table 3 shows the best results of this combination based on a brief initial test. From Table 3, it is clear that MFCC method produced the maximum accuracy when used with SVM learning algorithm. A pictorial view of Table 3 is shown in Figure 8.

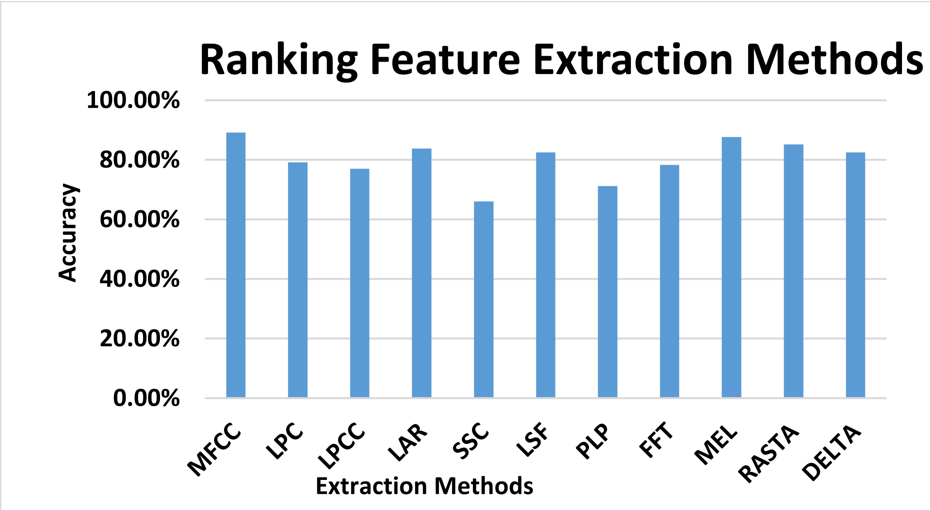


Figure 8. Feature extraction methods and levels of accuracy.

Algorithm 1 MFCC Method

Inputs: s: the signal, $\boldsymbol{fs}$: the sampling rate of the signal.
Outputs: $\boldsymbol{r}$ = transformed signal.
function $\boldsymbol{r} = \text{MFCC}(\boldsymbol{s}, \boldsymbol{fs})$ ▷ Frames blocking phase.
  $m = 100$ ▷ m: the distance between the begging of two frames.
  $n = 256$ ▷ n: frame length.
  $l = \text{length}(s)$
  $nbFrame = \text{floor}((l - n)/m) + 1;$ ▷ $nbFrame$: # of frames.
  for each integer $i = 1 : n$ **do**
    for each integer $j = 1 : nbFrame$ **do**
      $M(i, j) = s(((j - 1) * m) + i);$
    end for
  end for
  $h = \text{hamming}(n);$
  $M2 = \text{diag}(h) * M;$ ▷ *Windowing phase*: windowing all frames via multiply each individual frame by windowing function.
  for each integer $i = 1 : nbFrame$ **do**
    $\text{frame}(:, i) = \text{FFT}(M2(:, i));$
  end for ▷ *FFT Phase*: to convert each frame from time domain into frequency domain to removes the redundancy of mathematical calculations.
  $m = \text{melbankm}(20, n, fs);$ ▷ mel-spaced filterbank.
  $n2 = 1 + \text{floor}(n/2);$ ▷ Length of mel-spaced FFT.
  $z = m * \text{abs}(\text{frame}(1 : n2, :))^2;$
  $r = \text{DCT}(\log(z));$ ▷ Take log and then the DCT conversion.
end function

After using the MFCC, the speech signal was divided into segments of 15 ms frames with the use of a hamming window for further analysis [56]. We experimentally determined that 15 ms frames generated better recognition than others, because Holy Quran readers generally have a slow rate of recitation. A larger frame size does not enhance the ability of the recognition system to learn the characteristics of the signal. Feature vectors were extracted by MFCC algorithm as illustrated in Algorithm 1.

MFCC is a short-period power spectrum that is used to represent sound waves [57]. Mel frequencies are based on the critical bandwidth of the human ear recognized as a variation with frequency filters, which includes two types of frequencies [58]; the first one at frequencies below 1 kHz and the second one logarithmic filters at frequencies higher than 1 kHz to capture phonetically important characteristics [59]. In recitations from the Holy Quran, the correct pronunciation depends on the context, controlled by the voice of the reader and the reader's ability to move from one acoustic level to another "Maqam" "مقام". The stages involved in MFCC extraction are illustrated in Figure 9.

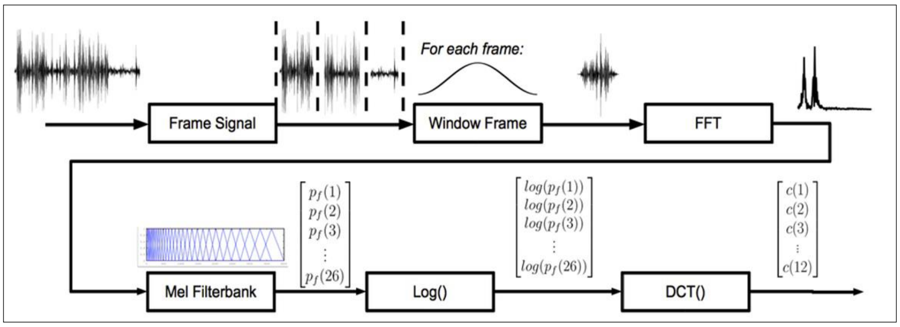


Figure 9. MFCC feature extraction stages [54].

The MFCC is calculated using Equation (3) and its implementation is in Algorithm 1.

$$C_i = \sum_{k=1}^{N} X_k cos\left(\frac{[\pi_i(k-0.5)]}{N}\right), \quad for\ i = 1,2,\dots,p \tag{3}$$

where C_i denotes the Cepstral coefficients, p is the order, k is the number of discrete Fourier transformations magnitude coefficients, X_k is the k^{th} order log-energy output of the filter bank and N is the number of filters (usually 20). Thus, 19 coefficients and an energy feature were extracted, generating a vector of 20 coefficients perframe. In this research, the first 20 orders of the MFCC were extracted. It was proven by Chaudhari (2015) that the increasing number of filters would raise the recognition rate, though the recognition and computational time for both training and testing would be negatively affected by the increasing number of filters [38].

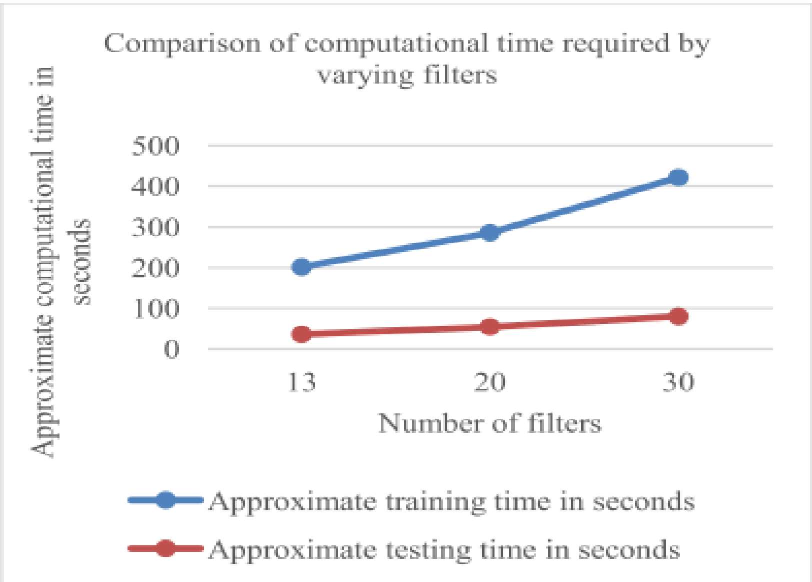


Figure 10. Comparison of computational time required by varying filters (taken from [38]).

Jordanian Journal of Computers and Information Technology (JJCIT), Vol. 06, No. 04, December 2020.

It can be concluded from Figure 10 and Chaudhari (2015) that the recognition rate will be decreased either for testing or training as the number of filters increased and *vice versa* [38]. In our paper, 20 filters will be used, the number used in most speech recognition tools, in accordance with the belief that this number of filters provides a better rate of recognition and an acceptable computation time for training and testing.

Since the verses are of different lengths, 20 feature representations of each verse ("Ayah" "آية") were extracted for the 258 acoustic waves. The MFCC files correspond to the acoustic waves, extracting the feature vectors of the total number of verses in a vector matrix for all readers with a size equal to (20×22952) features. After that, these 20-feature matrices were combined into one file for the ten different styles of recitation. Each verse vector is transposed and labeled according to the reading style in one CSV-file that includes a labeled matrix for reading. Furthermore, we have generated six feature vector matrices using six different methods for the purpose of comparison between the levels of accuracy of the results of different feature extraction methods.

4.4 Formulating the Training Matrix and Its Feature Representation Using SVM

The CSV labeled matrix extracted and generated in the previous phase is divided into two parts. Part I usually takes 70% of this matrix which is used for training on the proposed model. The rest of the matrix (30%) is used to test and evaluate the proposed model. Figure 11 shows the general framework of the training phase, while Figure 12 shows the general testing framework.

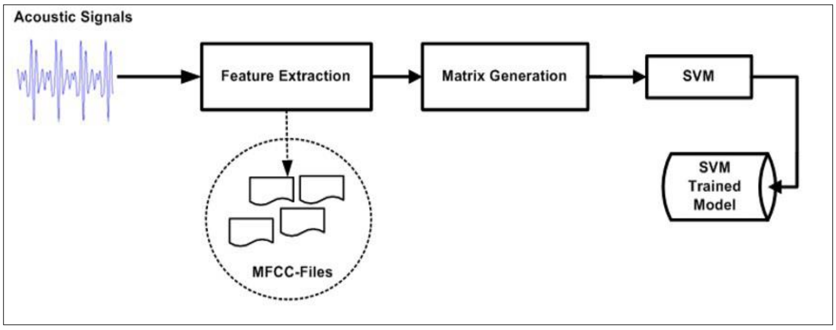


Figure 11. Proposed SVM-based training model.

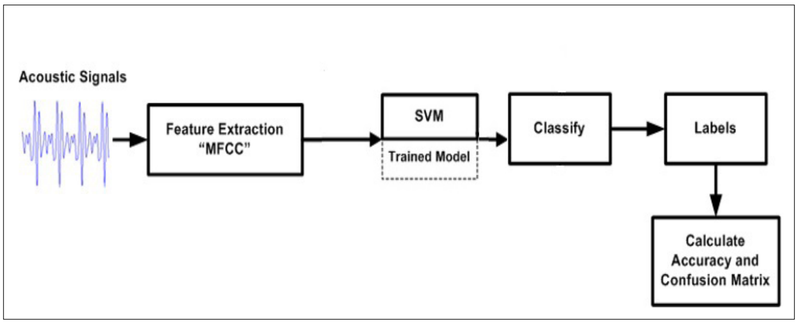


Figure 12. Proposed testing SVM-based model.

The selection of the features was homogeneous and normally distributed. Figure 13 shows the homogeneity in selecting the features drawn by the WEKA tool. Features one and two for all the readings were biased a little to the right, while the other features were normally distributed in all the readings of the audio waves. This is due to the limited examples of some types of reading, leading to a limited number of features. Figure 14 is the same as Figure 13, but expressed in a graphical representation using colored dots.

A built-in randomized filter is used to rearrange the features in randomly chosen rows. This randomized filter will enhance the learning process for any learning algorithm and will affect the accuracy of the results.

Figure 13. Feature distribution over audio wave files for readings.

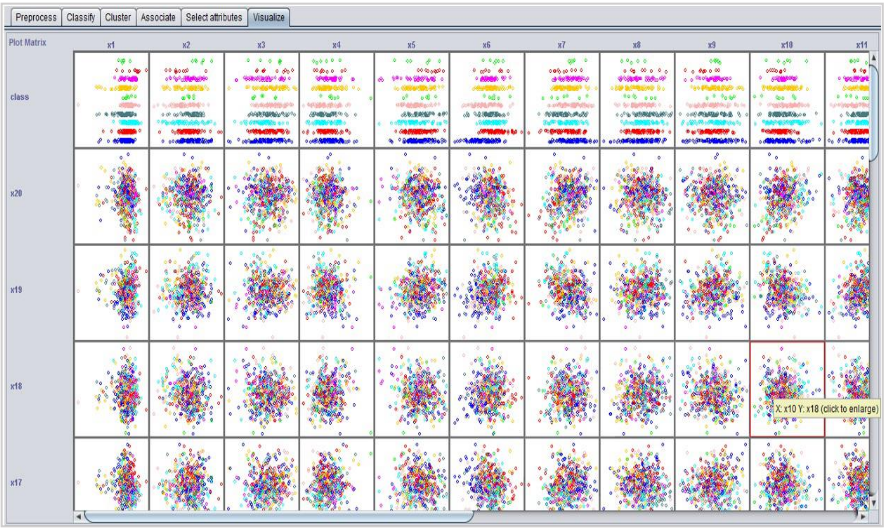


Figure 14. Representation of feature homogeneity.

4.5 Testing Phase

After obtaining the training model through SVM learning, the rest of the feature matrix (30%) is used for testing, as shown in Figure 12. The testing results obtained are represented by the confusion matrix of SVM, which shows the hit and the misclassification of the waves. Note that the SVM confusion matrix will be fully discussed in Section five. In addition, the proposed methodology stages will be applied to different learning algorithms other than the SVM. The results obtained from these learning algorithms will be compared for analysis and discussion.

5. Experimental Results and Discussion

The Holy Quran recitation recognition system is based primarily on the characteristics extracted from acoustic waves in conjunction with learning and classification algorithms. The SVM was trained with 10 Holy Quran readings, as listed in Table 1. For comparison purposes, the same experiment was repeated with the Least Square Support Vector Machine (LSSVM) and ANN models. For training purposes, we chose 70% of the features to train SVM, LSSVM, ANN and other learning algorithms. The remaining 30% of the features were cropped for testing purposes. When the attribute matrix is loaded into WEKA, attribute classifications and properties (features) from X_1 to X_{20} are displayed, so

Jordanian Journal of Computers and Information Technology (JJCIT), Vol. 06, No. 04, December 2020.

that the user can verify them. The attributes X_1, X_2, …, X_{20} are the feature vector values or predictors and the “Readings” are the target classes. We started training SVM on 70% of the features and testing on 30%.

The output of the SVM text file contains the testing confusion matrix, accuracy matrix for all classes and summary of the overall results. The confusion matrix is a data structure used to show the classification results. The rows of this matrix represent the desired classification, while the columns represent the predicted classification. Table 4 represents the confusion matrix of the testing phase with SVM. It is obvious that some samples go below the hyper plane of the SVM (main diagonal), while others are above it. In either case, it must be regarded as a misclassification of the sample. When the samples are on the main diagonal, they are correctly classified. From Table 4, most of the samples are correctly classified.

Table 4. Testing confusion matrix of SVM.

a	b	c	d	e	f	G	h	i	j	<-- classified as		
22239	338	102	231	162	5	181	132	14	28	**a**	=	HK
81	**23169**	51	175	82	9	117	33	13	4	**b**	=	AJ
116	160	**16752**	113	42	0	74	41	6	7	**c**	=	AK
206	177	68	**20713**	98	23	126	30	17	6	**d**	=	AA
221	115	35	131	**22383**	1	121	29	12	18	**e**	=	K
36	66	10	70	18	**1679**	40	3	30	11	**f**	=	YKH
180	121	52	173	82	1	**20411**	18	2	12	**g**	=	NM
80	96	36	57	34	0	33	**12310**	24	10	**h**	=	KH
50	140	40	99	58	28	22	40	**2292**	0	**i**	=	IA
46	18	12	50	42	12	74	31	6	**2301**	**j**	=	IK

To calculate the accuracy from the confusion matrix, Equation (4) is used:

$$\text{Accuracy} = \frac{\sum \text{Diagonal Sample of confusion matrix}}{\text{Total Sample}} \tag{4}$$

The total accuracy of the recitation’s recognition is 96.59% when using SVM. About 4% of the total samples were poorly classified. Misclassified samples may arise as a result of the way in which the Holy Quran reader recites different verses. The similarity between the reader’s sound and its emotional tone is critical. Most of Holy Quran readers follow the same rules for the recitation, Ahkam Al-Tajweed “أحكام التجويد”, when they are reciting the various verses of the “surah” “السور”. Applying “Ahkam Al-Tajweed” appropriately will increase the chance of the similarity of the acoustic waves produced by different readers. Finally, every reader has his/her own shifts in tone “Makam” “مقام”, but these might sound very similar when the rules are applied. The detailed levels of accuracy of SVM at the class are clearly established in Table 5. The weighted average for each measurement is listed in the last row. As

Table 5. Detailed SVM accuracy by class.

TP Rate	FP Rate	Precision	Recall	F-Measure	MCC	ROC Area	PRC Area	Class
0.949	0.008	0.956	0.949	0.953	0.944	0.971	0.916	HK
0.976	0.01	0.95	0.976	0.963	0.956	0.983	0.931	AJ
0.968	0.003	0.976	0.968	0.972	0.968	0.982	0.949	AK
0.965	0.009	0.95	0.965	0.957	0.95	0.978	0.921	AA
0.97	0.005	0.973	0.97	0.972	0.967	0.983	0.949	K
0.855	0.001	0.955	0.855	0.902	0.903	0.927	0.819	YKH
0.97	0.006	0.963	0.97	0.966	0.961	0.982	0.938	NM
0.971	0.003	0.972	0.971	0.971	0.969	0.984	0.946	KH
0.828	0.001	0.949	0.828	0.884	0.884	0.913	0.788	IA
0.888	0.001	0.96	0.888	0.922	0.922	0.944	0.854	IK
0.961	**0.006**	**0.961**	**0.961**	**0.961**	**0.955**	**0.978**	**0.929**	**←Weighted Average**

can be seen from Table 5, the weighted average of precision reached 96%, while the weighted average of the false positive rate (FP) reached 0.006 which is a very good indicator regarding the recognition rate.

The TP, FP, Precision, Recall, F-Measure, MCC, ROC and PRC measurements are mentioned in Table 5, while the details of these measurements are reported in Table 6, which is taken from the URL[1]. Note that the weighted arithmetic mean is similar to an ordinary arithmetic mean (the most common type of average), except that instead of each of the data points contributing equally to the final average, some data points contribute more than others, see the (Taken from [2]).

Table 6. Summary of measurements and accuracy metrics.

Measurements	Meaning
TP	Rate of true positives (instances correctly classified as a given class).
FP	Rate of false positives (instances falsely classified as a given class).
Precision	Proportion of instances that are truly of a class divided by the total instances classified as that class
Recall	Proportion of instances classified as a given class divided by the actual total in that class (equivalent to TP rate).
F-Measure	A combined measure for precision and recall calculated as: $F - Measure = \frac{2*Precision*Recall}{Precision+Recall}$. It is the weighted harmonic mean (sometimes called the subcontrary mean), which is one of several kinds of averages appropriate for situations when the average of rates is desired.
MCC	The Matthews Correlation Coefficient whichis used in machine learning as a measure of the quality of binary (two-class) classifications.
ROC Area	Receiver Operating Characteristic: A plot of a true positive fraction (= sensitivity) *vs.* a false positive fraction (= 1 – specificity) for all potential cut-offs for a test.
PRC Area	Precision-recall curve: A plot of precision (= PPV) *vs.* recall (= sensitivity) for all potential cut-offs for a test. PRC might be a better choice for unbalanced datasets.

The pictorial view of the previous table is represented in Figure 15, which illustrates the graphical representation of SVM results. It is clear that all classes of readings are recognized with a high probability of accuracy. K and AK classes reach the highest recognition rate, which is mapped to "Aseem" "عاصم" and "Al-Kesaee" "الكسائي" readings. Table 7 summarizes the SVM results and measurements of error. Clearly, the absolute error rate is very low which is considered satisfactory in this kind of recognition problem. The overall accuracy as seen from Table 7 reaches 96%.

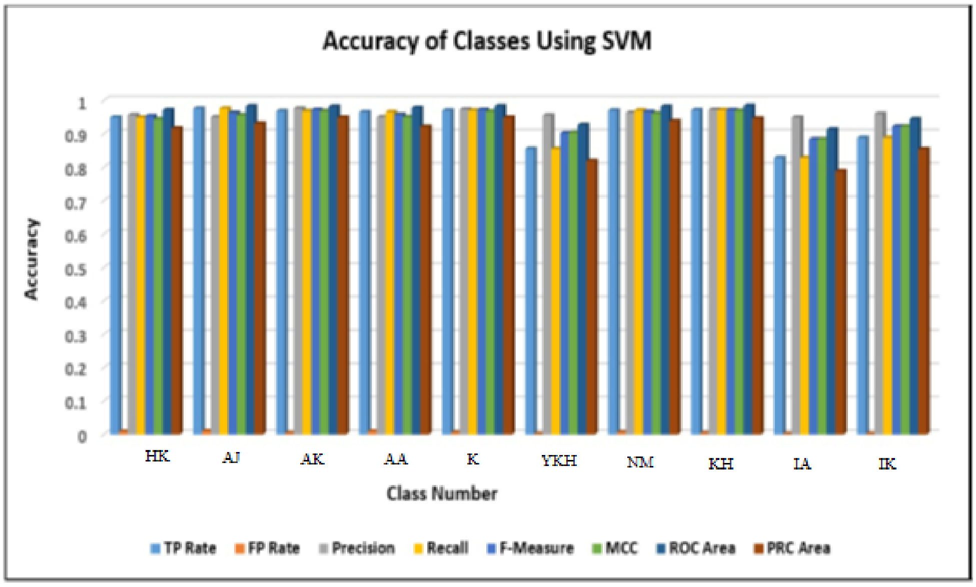


Figure 15. Graphical representation of SVM results.

[1] https://acutecaretesting.org/en/articles/precision-recall-curves-what-are-they-and-how-are-they-used
2 https://en.wikipedia.org/wiki/Weighted_arithmetic_mean

Jordanian Journal of Computers and Information Technology (JJCIT), Vol. 06, No. 04, December 2020.

Table 7. Summary of SVM results.

Correctly Classified Instances	144249 or 96.1256 %
Incorrectly Classified Instances	5814 or 3.8744 %
Kappa Statistic	0.9552
Mean Absolute Error	0.0077
Root Mean Squared Error	0.088
Relative Absolute Error	4.4758 %
Root Relative Squared Error	29.9198 %
Total Number of Instances	150063

The same experiment was repeated using multi-perceptron ANN with 20 inputs representing the features. Table 8 shows the confusion matrix of the ANN. Compared to Table 7, it is obvious that most of the instances are incorrectly classified.

Table 8. Testing confusion matrix of ANN.

a	b	c	d	e	f	G	h	i	j	<-- classified as		
15121	1356	751	1441	2028	2	1594	1120	4	15	**a**	=	HK
856	**17269**	658	1902	1007	3	1067	904	25	43	**b**	=	AJ
410	743	**13287**	801	526	2	477	993	33	39	**c**	=	AK
1312	790	1513	**14955**	1105	3	1364	371	41	10	**d**	=	AA
1417	552	541	876	**17671**	28	1425	532	5	19	**e**	=	K
136	341	156	656	185	**120**	197	155	17	0	**f**	=	YKH
942	734	1041	1694	1256	26	**14838**	518	3	0	**g**	=	NM
777	359	624	950	255	3	365	**9322**	9	16	**h**	=	KH
64	622	243	601	285	2	157	302	**493**	0	**i**	=	IA
203	52	275	220	511	0	357	439	0	**535**	**j**	=	IK

Moreover, by looking at Table 9, we can see that the maximum classification accuracy is 78%. Some classes, such as IA and IK, reached only 22.7% and 38.9%, respectively, which are very low compared to their accuracy when SVM is used. The average accuracy for all classes reaches only 69%, which indicates that the ANN is inadequate for this kind of recognition. This signifies that the absolute error was 8% when using ANN, while it was 0.0077 when using SVM. This result indicates that SVM is more appropriate than the ANN algorithm.

Table 9. Detailed ANN accuracy by class.

TP Rate	FP Rate	Precision	Recall	F-Measure	MCC	ROC Area	PRC Area	Class
0.645	0.048	0.712	0.645	0.677	0.622	0.896	0.712	HK
0.728	0.044	0.757	0.728	0.742	0.695	0.932	0.78	AJ
0.768	0.044	0.696	0.768	0.73	0.694	0.94	0.754	AK
0.697	0.071	0.621	0.697	0.656	0.597	0.909	0.649	AA
0.766	0.056	0.712	0.766	0.738	0.689	0.916	0.772	K
0.061	0	0.635	0.061	0.112	0.194	0.56	0.118	YKH
0.705	0.054	0.679	0.705	0.692	0.641	0.916	0.701	NM
0.735	0.039	0.636	0.735	0.682	0.652	0.898	0.719	KH
0.178	0.001	0.783	0.178	0.29	0.369	0.782	0.227	IA
0.206	0.001	0.79	0.206	0.327	0.399	0.861	0.389	IK
0.69	**0.049**	**0.695**	**0.69**	**0.682**	**0.639**	**0.908**	**0.704**	**←Weighted Average**

A summary of ANN results and error measurements is shown in Table 10. Clearly, the absolute error rate is very high compared to SVM. The overall accuracy according to Table 10 reaches 69%.

For comparison between ANN and SVM, Table 11 shows a summary of the recognition rate between ANN and SVM when applied on testing data from the same corpus and data from outside the corpus (outlet data).

Our proposed SVM-based system obtained better results compared with ANN, having higher accuracy and lower Mean Square Error (MSE) compared with ANN. A stacked bar graph is shown in Figure 16 to illustrate the results for both SVM and ANN with different measurements.

Table 10. Summary of ANN results.

Correctly Classified Instances	103611 or 69.045 %
Incorrectly Classified Instances	46452 or 30.955 %
Kappa Statistic	0.6409
Mean Absolute Error	0.0831
Root Mean Squared Error	0.2155
Relative Absolute Error	47.9745 %
Root Relative Squared Error	73.2585 %
Total Number of Instances	150063

Table 11. SVM and ANN result summary.

Accuracy	**Learning Algorithms**	
	SVM	**ANN**
Accuracy-Normal Corpus	**96 %**	69%
Accuracy-Outlet Data	**95%**	60%
Average Accuracy	**95.5%**	64.5%
Average MSE	0.0077	0.08

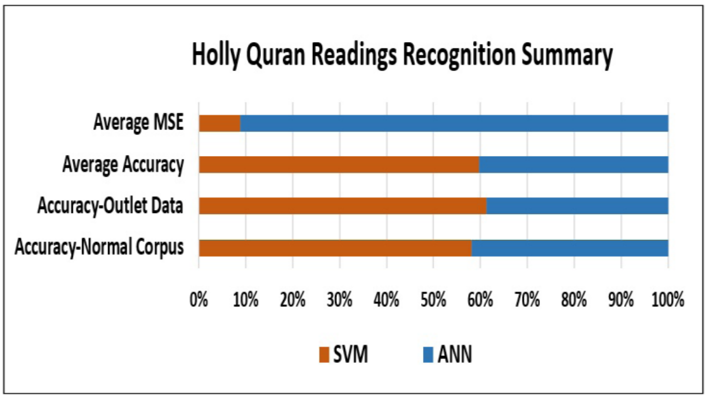


Figure 16. SVM and ANN result summary.

Table 12. Summary of results obtained from different learning algorithms.

Number of Instances = 500211 (70% for Training and 30% for Testing)				
Classifier	**Training time in seconds**	**Testing time in seconds**	**Mean Square Error (MSE)**	**Accuracy**
NB-Tree (Decision Tree)	72657.88	8261.66	0.2641	47.4118%
RBF (Radial Bases)	3211.61	2270.68	0.2796	31%
Random Forest (Forest Decision Trees)	1500.33	300.65	0.2293	68.3529%
NNge (Nearest Neighbor)	1200.25	300.21	0.3147	50.4706%
Multi-Objective Evolutionary Fuzzy Classifier	5000.66	1700.12	0.3163	20.8235%
Deep Learning Classifier	1900.33	460.89	0.2827	40.7647%
SVM (Support Vector Machine)	**42357.75**	**5211.73**	**0.08800**	**96.12 %**
LSSVM (Least Squares Support Vector Machines)	**32133.60**	**3332.00**	**0.09100**	**95.16 %**
Multi-layer Perceptron (ANN)	2677.02	10.20	0.2435	62.0588%

For the sake of consistency, further experiments were conducted using other learning algorithms, such as: NB-Tree (Decision Tree), RBF (Radial Bases), Random Forest (RF), NNge (Nearest Neighbor), Multi-Objective Evolutionary Fuzzy Classifier (MOEFC), Deep Learning Classifier (DLC) and Least Square Support Vector Machine (LSSVM). We focused on the training time, test time, MSE and accuracy measurements. The results obtained from those algorithms are shown in Table 12.

It is clear that SVM and LSSVM have the lowest MSE and the highest accuracy, but they need more time for training and testing compared to other algorithms. Figure 17 shows a chart of Table 12.

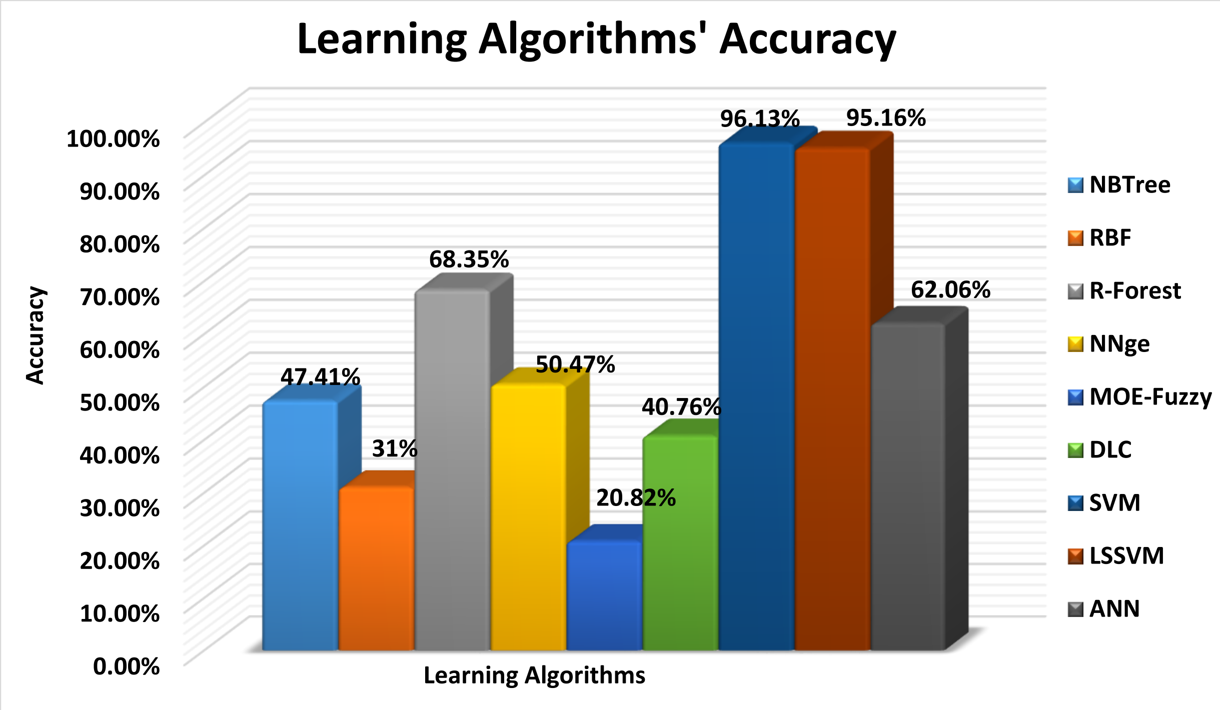


Figure 17. Accuracy of different learning algorithms.

6. Conclusion and Future Work

In this research, we have addressed the problem of determining the efficiency of recitation of the Holy Quran based on established rules of recitation. We have investigated the recognition of the type of Holy Quran recitation based on the SVM learning algorithm. Moreover, we compared the results with other learning algorithms. Acoustic waves for the ten Holy Quran reading styles ("Qira'ah" "قراءة") were collected in a corpus of ten recitation variants "Qira'at" "قراءات" and a variety of readers were included in the study. Subsequently, MFCC properties were extracted from wave signals and labeled according to the appropriate reading ("Qira'ah") style. The labeled matrix of the "Qira'ah" was fed to SVM using the WEKA tool. 70% of the classified matrix was given to SVM for training. The remaining 30% of the labeled matrix was used for testing purposes. Additional outlet data was used to test the proposed model in order to demonstrate the validity and reliability of our proposed system.

A comparison between our adopted learning algorithm (SVM) and other learning algorithms (Table 11) was made to validate the adequacy of SVM. Briefly, the results obtained using SVM outperformed the results obtained when using other learning algorithms. The results reveal that the identification accuracy using SVM is higher in comparison to those produced by the ANN and other learning algorithms. The accuracy using SVM reaches approximately 96%, while ANN accuracy is 62%. Some other learning algorithms, such as FDT, reached 68%, which is greater than ANN's level of accuracy. When comparing the results obtained from Table 11, we can see that SVM demonstrated its superiority over all other learning algorithms used. However, SVM requires more time in the training phase, while proving to be faster in testing input data.

The promising outcomes of this research have encouraged us to undertake further investigation in recognizing Holy Quran recitation. Several future directions could be taken in this area. For example, new deep learning algorithms could be applied to compare the results obtained from the application of such algorithms with the results of this study. Another example is hybridizing some powerful learning algorithms, like Learning Vector Quantization (LVQ) with the SVM model to improve the results. Furthermore, expanding the corpus to include other reciters could be carried out to validate our system. Finally, we may need to extend the evaluation criteria for more accurate measurement of SVM performance.

REFERENCES

[1] Y. Kato, "Speech Recognition System," Journal of the Acoustical Society of America, vol. 107, no. 5, p. 2326, 2000.

[2] L. R. Rabiner and B.-H. Juang, Fundamentals of Speech Recognition, PTR Prentice Hall Englewood Cliffs, 1993.

[3] T. K. Das and K. M. O. Nahar, "A Voice Identification System Using Hidden Markov Model," Indian Journal of Science and Technology, vol. 9, no. 4, DOI: 10.17485/ijst/2016/v9i4/83894, 2016.

[4] J. M. Baker et al., "Developments and Directions in Speech Recognition and Understanding, Part 1 [DSP Education]," IEEE Signal Processing Magazine, vol. 26, no. 3, pp. 75-80, 2009.

[5] J. Chong, E. Gonina, D. Kolossa, S. Zeiler and K. Keutzer, "An Automatic Speech Recognition Application Framework for Highly Parallel Implementations on the GPU," Technical Report No. UCB/EECS-2012-47, [Online], available: http://www.eecs.berkeley.edu/Pubs/TechRpts/2012/EECS-2012-47.html, 2012.

[6] K. Nahar, H. Al-Muhtaseb, W. Al-Khatib, M. Elshafei and M. Alghamdi, "Arabic Phonemes Transcription Using Data Driven Approach," Int. Arab J. of Information Technology, vol. 12, no. 3, pp. 237–245, 2015.

[7] Y.-H. Shao and N.-Y. Deng, "A Coordinate Descent Margin Based-twin Support Vector Machine for Classification," Neural Networks, vol. 25, pp. 114–121, 2012.

[8] K. Nahar, M. Abu Shquier, W. G. Al-Khatib, H. Al-Muhtaseb and M. Elshafei, "Arabic Phonemes Recognition Using Hybrid LVQ/HMM Model for Continuous Speech Recognition," International Journal of Speech Technology, vol. 19, no. 3, pp. 495–508, 2016.

[9] J. Allen, D. Byron, M. Dzikovska, G. Ferguson, L. Galescu and A. Stent, "An Architecture for a Generic Dialogue Shell," Natural Language Engineering, vol. 6, no. 3–4, pp. 213–228, 2000.

[10] K. M. O. Nahar, W. G. Al-Khatib, M. Elshafei, H. Al-Muhtaseb and M. M. Alghamdi, "Data-driven Arabic Phoneme Recognition Using Varying Number of HMM States," Proc. of the 1st IEE International Conference on Communications, Signal Processing and Their Applications (ICCSPA 2013), pp. 1–6, Sharjah, United Arab Emirates, 2013.

[11] M. Abdul-Mageed, M. Diab and S. Kübler, "SAMAR: Subjectivity and Sentiment Analysis for Arabic Social Media," Computer Speech and Language, vol. 28, no. 1, pp. 20–37, 2014.

[12] F. Diehl, M. J. F. Gales, M. Tomalin and P. C. Woodland, "Morphological Decomposition in Arabic ASR Systems," Computer Speech and Language, vol. 26, no. 4, pp. 229–243, 2012.

[13] K. Kirchhoff et al., "Novel Approaches to Arabic Speech Recognition: Report from the 2002 Johns-Hopkins Summer Workshop," Proceedings of IEEE International Conference on Acoustics, Speech and Signal Processing (ICASSP'03), vol. 1, pp. I344-I347, Hong Kong, China, 2003.

[14] E. Zarrouk, Y. Ben Ayed and F. Gargouri, "Hybrid Continuous Speech Recognition Systems by HMM, MLP and SVM: A Comparative Study," International Journal of Speech Technology, vol. 17, no. 3, pp. 223–233, 2014.

[15] A. Al-Otaibi, "Speech Processing," Britsh Library in Association with UMI, 1988.

[16] M. M. El Choubassi, H. E. El Khoury, C. E. J. Alagha, J. A. Skaf and M. A. Al-Alaoui, "Arabic Speech Recognition Using Recurrent Neural Networks," Proceedings of the 3rd IEEE International Symposium on Signal Processing and Information Technology (IEEE Cat. No.03EX795) (ISSPIT 2003), pp. 543–547, Darmstadt, Germany, 2003.

[17] A. Ali et al., "Automatic Dialect Detection in Arabic Broadcast Speech," arXiv Paper, arXiv1509.06928, [Online], available: https://arxiv.org/pdf/1509.06928.pdf, 2015.

[18] F. Biadsy, N. Habash and J. Hirschberg, "Improving the Arabic Pronunciation Dictionary for Phone and

Word Recognition with Linguistically-based Pronunciation Rules," Proceedings of Human Language Technologies: The 2009 Annual Conference of the North American Chapter of the Association for Computational Linguistics, pp. 397–405, United States, 2009.

[19] N. J. Ibrahim, M. Y. Zulkifli and H. Mohd, "Improved Design for Automated Tajweed Checking Rules Engine of Quranic Verses Recitation : A Review," International Journal on Quranic Research, no. January, pp. 39–50, 2011.

[20] A. Muhammad, Z. Ul Qayyum, M. Waqar Mirza, S. Tanveer, A. M. Martinez-Enriquez and A. Z. Syed, "E-hafiz: Intelligent System to Help Muslims in Recitation and Memorization of Quran," Life Science Journal, vol. 9, no. 1, pp. 534–541, 2012.

[21] T. Mssraty and Q. Faryadi, "Teaching the Qur'anic Recitation with Harakatt: A Multimedia-based Interactive Learning Method," International Journal of Scientific and Enginerring Research, vol. 3, no. 8, pp. 1–4, 2012.

[22] N. W. Arshad et al., "Makhraj Recognition for Al-Quran Recitation Using MFCC," International Journal of Intelligent Information Processing, vol. 4, no. 2, pp. 45–53, Jun. 2013.

[23] T. Sabbah and A. Selamat, "Support Vector Machine-based Approach for Quranic Words Detection in Online Textual Content," Proc. of the 8th Malaysian Software Engineering Conference (MySEC 2014), pp. 325–330, Langkawi, Malaysia, 2014.

[24] S. A. E. Mohamed, A. S. Hassanin and M. T. Ben Othman, "Virtual Learning System (Miqra'ah) for Quran Recitations for Sighted and Blind Students," Journal of Software Engineering and Applications, vol. 07, no. 04, pp. 195–205, 2014.

[25] Y. O. M. Elhadj, M. Alghamdi and M. Alkanhal, "Phoneme-based Recognizer to Assist Reading the Holy Quran," Recent Advances in Intelligent Informatics, Part of the Advances in Intelligent Systems and Computing Book Series (AISC), vol. 235, pp. 141–152, 2014.

[26] H. M. A. Tabbaa and B. Soudan, "Computer-aided Training for Quranic Recitation," Procedia-Social and Behavioral Sciences, vol. 192, pp. 778–787, 2015.

[27] M. Y. El Amrani, M. M. H. Rahman, M. R. Wahiddin and A. Shah, "Building CMU Sphinx Language Model for the Holy Quran Using Simplified Arabic Phonemes," Egyptian Informatics Journal, vol. 17, no. 3, pp. 305–314, 2016.

[28] M. Al-Ayyoub, N. A. Damer and I. Hmeidi, "Using Deep Learning for Automatically Determining Correct Application of Basic Quranic Recitation Rules," International Arab Journal of Information Technology, vol. 15, no. 3, pp. 620–625, 2018.

[29] M. Y. El Amrani, M. R. Wahiddin, M. M. H. Rahman and A. Shah, "Towards Using CMU Sphinx Tools for the Holy Quran Recitation Verification," International Journal on Islamic Applications in Computer Science and Technol., vol. 4, no. 2, pp. 10–15, 2016.

[30] A. Hughes, P. Trudgill and D. Watt, English Accents and Dialects: An Introduction to Social and Regional Varieties of English in the British Isles, Routledge, 2013.

[31] O. F. Zaidan and C. Callison-Burch, "Arabic Dialect Identification," Computational Linguistics, vol. 40, no. 1, pp. 171–202, 2014.

[32] H. K. Tayyeh, M. S. Mahdi and A. S. A. Al-Jumaili, "Novel Steganography Scheme Using Arabic Text Features in Holy Quran," International Journal of Electrical and Computer Engineering, vol. 9, no. 3, pp. 1910-1918, 2019.

[33] L. C. Moore, "Learning by Heart in Qur'anic and Public Schools in Northern Cameroon," Social Analysis, vol. 50, no. 3, pp. 109–126, 2006.

[34] A. Rasmussen, Women, the Recited Qur'an and Islamic Music in Indonesia, Univ. of California Press, 2010.

[35] Z. A. Adhoni, H. Al Hamad, A. A. Siddiqi, M. Parvez and Z. A. Adhoni, "Cloud-based Online Portal and Mobile Friendly Application for the Holy Qur'an," Life Science Journal, vol. 10, no. 12, 2013.

[36] M. M. Khan, The Translation of the Meanings of Sahih Al-Bukhâri, vol. 5, Kazi Publications, 1997.

[37] T. Wahbi and A. Gadeed, "The Recognition of Holy Quran Reading Types 'Rewaih," International Journal of Advanced Research in Computer Science, vol. 5, no. 3, pp. 37-40, 2014.

[38] A. A. Chaudhari, "Effect of Varying MFCC Filters for Speaker Recognition," International Journal of Computer applications, vol. 128, no. 14, pp. 7–9, 2015.

[39] M. N. Al-Kabi, G. Kanaan, R. Al-Shalabi, M. O. K. Nahar and M. B. Bani-Ismail, "Statistical Classifier of the Holy Quran Verses (Fatiha and Yaseen Chapters)," Journal of Applied Sciences, vol. 5, no. 3, pp. 580–583, 2005.

[40] H. Okuyucu, A. Kurt and E. Arcaklioglu, "Artificial Neural Network Application to the Friction Stir Welding of Aluminum Plates," Materials and Design Journal, vol. 28, no. 1, pp. 78–84, 2007.

[41] M. Al-Abri and N. Hilal, "Artificial Neural Network Simulation of Combined Humic Substance Coagulation and Membrane Filtration," Chemical Engineering Journal, vol. 141, no. 1, pp. 27–34, 2008.

[42] J. Huang, J. Lu and C. X. Ling, "Comparing Naive Bayes, Decision Trees and SVM with AUC and Accuracy," Proc. of the 3rd IEEE International Conference on Data Mining (ICDM 2003), pp. 553–556, Melbourne, FL, USA, 2003.

[43] A. Statnikov, C. F. Aliferis, D. P. Hardin and I. Guyon, A Gentle Introduction to Support Vector Machines in Biomedicine: Volume 2: Case Studies and Benchmarks, World Scientific, 2013.

[44] J. Platt et al., "Probabilistic Outputs for Support Vector Machines and Comparisons to Regularized Likelihood Methods," Adv. Large Margin Classif., vol. 10, no. 3, pp. 61–74, 1999.

[45] Z.-Q. Zeng, H.-B. Yu, H.-R. Xu, Y.-Q. Xie and J. Gao, "Fast Training Support Vector Machines Using Parallel Sequential Minimal Optimization," Proc. of the 3rd International Conference on Intelligent System and Knowledge Engineering, vol. 1, pp. 997–1001, Xiamen, China, 2008.

[46] L. Wang, Support Vector Machines: Theory and Applications, vol. 177, Springer Science & Business Media, 2005.

[47] S. Tong and E. Chang, "Support Vector Machine Active Learning for Image Retrieval," Proceedings of the 9th ACM International Conference on Multimedia, pp. 107–118, [Online], available: https://doi.org/10.1145/500141.500159, 2001.

[48] M. A. Hearst et al., "Support Vector Machines," IEEE Intelligent Systems, vol. 13, no. 4, pp. 18–28, 1998.

[49] R. Collobert and S. Bengio, "SVMTorch: Support Vector Machines for Large-scale Regression Problems," Journal of Machine Learning Research, vol. 1, no. Feb, pp. 143–160, 2001.

[50] C. J. C. Burges, "A Tutorial on Support Vector Machines for Pattern Recognition," Data Mining and Knowledge Discovery, vol. 2, no. 2, pp. 121–167, 1998.

[51] M. Lalaoui, A. El Afia and R. Chiheb, "A self-tuned Simulated Annealing Algorithm Using Hidden Markov Model," International Journal of Electrical and Computer Engineering, vol. 8, no. 1, p. 291, 2018.

[52] A. El Afia, M. Sarhani and O. Aoun, "Hidden Markov Model Control of Inertia Weight Adaptation for Particle Swarm Optimization," IFAC-PapersOnLine, vol. 50, no. 1, pp. 9997–10002, 2017.

[53] W. M. Campbell, J. P. Campbell, D. A. Reynolds, E. Singer and P. A. Torres-Carrasquillo, "Support Vector Machines for Speaker and Language Recognition," Computer Speech and Language, vol. 20, no. 2, pp. 210–229, 2006.

[54] K. M. O. Nahar, M. Al-Shannaq, A. Manasrah, R. Alshorman and I. Alazzam, "A Holy Quran Reader/Reciter Identification System Using Support Vector Machine," International Journal of Machine Learning and Computing, vol. 9, no. 4, pp. 458-464, 2019.

[55] U. Shrawankar, "Techniques for Feature Extraction in Speech Recognition System: A Comparative Study," arXiv Prepr, arXiv1305.1145, [Online], Available: https://arxiv.org/ftp/arxiv/papers/1305/1305. 1145.pdf, 2013.

[56] T. Kinnunen, R. Saeidi, J. Sandberg and M. Hansson-Sandsten, "What Else Is New Than the Hamming Window? Robust MFCCs for Speaker Recognition *via* Multitapering," Proc. of the 11th Annual Conference of the International Speech Communication Association, Japan, [Online], Available: https://citeseerx.ist.psu.edu/viewdoc/download?doi=10.1.1.178.5674&rep=rep1&type=pdf, 2010.

[57] S. Chowdhury, N. Mamun, A. A. S. Khan and F. Ahmed, "Text Dependent and Independent Speaker Recognition Using Neural Responses from the Model of the Auditory System," Proc. of the IEEE International Conference on Electrical, Computer and Communication Engineering (ECCE), pp. 871–874, Cox's Bazar, Bangladesh, 2017.

[58] R. Bharti and P. Bansal, "Real Time Speaker Recognition System Using MFCC and Vector Quantization Technique," International Journal of Computer Applications, vol. 117, no. 1, pp. 25-31, 2015.

[59] S. A. Majeed, H. Husain, S. A. Samad and T. F. Idbeaa, "Mel Frequency Cepstral Coefficients (MFCC)

Feature Extraction Enhancement in the Application of Speech Recognition: A Comparison Study," Journal of Theoretical and Applied Information Technology, vol. 79, no. 1, pp. 38–56, 2015.

ملخص البحث:

يشير تمييز قراءات القرآن الكريم الى تحديد نوع القراءة ضمن الأنماط المعتمدة (بالعربية: القراءات). لقد بحثت دراسات سابقة عديدة في قواعد قراءة القرآن الكريم المعروفة بأحكام التجويد، التي تطبق من قبل القُرّاء عند القراءة بصوتٍ عالٍ. إلا أنه ليست هناك دراسات سابقة تناولت مسألة تتبُّع نوع القراءة. ومن خلال هذا البحث، يمكننا مساعدة الطلبة في تعلُّم قراءة القرآن الكريم بسهولة ودقة عبر التطبيق الدقيق لأحكام التجويد والتمييز بين أنواع القراءة المختلفة.

في هذه الدراسة، تم تطبيق نموذج تمييز مقترح من أجل تمييز نوع قراءة القرآن الكريم من الموجة الصوتية الخاصة بتلك القراءة. وقد بُني النموذج المقترح على ثلاث مراحل؛ الأولى هي مرحلة استخلاص السِّمات من الإشارة الصوتية ووسْمها. أما المرحلة الثانية فهي تدريب النموذج (SVM) على السِّمات المستخلصة والموسومة، في حين تتمثل المرحلة الثالثة والأخيرة في تمييز نوع القراءة بناءً على ذلك النموذج. ولتحقيق ذلك، قمنا ببناء حُزمةٍ تتكون من عشر فئاتٍ كل واحدة منها توسم كنوع من أنواع القراءة. كذلك تم تطبيق مجموعة من خوارزميات تعلم الآلة ومقارنتها. وقد أثبتت النتائج التجريبية تفوُّق نموذج التمييز المقترح القائم على آلة متّجهات الدّعم (SVM) على غيره من الخوارزميات الأخرى التي تم تطبيقها ومقارنتها؛ إذ بلغت نسبة نجاح النموذج المقترح (96%).