Intelligent Categorization of Arabic Commands Utilizing Machine Learning Techniques with Short Effective Features Vector

Azza E.B. Ibrahim
Computers and Systems Dept.
Electronic Research Institute, Egypt

Rana S.M. Saad
Computers and Systems Dept.
Electronic Research Institute, Egypt

ABSTRACT
Different technologies are now being employed to improve the quality of life, particularly for the disabled and elderly. Speech is the quickest and most convenient method of communicating with people and technology. The majority of the works have focused on English speech; however, there is some interest in Arabic. In this study, an Arabic dataset is created, which will eventually be used to control a mobile assistant robot. Arabic is a challenging language to learn because of its many dialects, each of which has its own impact on the spoken word. The Egyptian Arabic Speech Commands (EASC) dataset was compiled from people of different backgrounds, ages, and genders who spoke in colloquial dialects. The Arabic recognition test is made more difficult by this fluctuation. Using various machine learning techniques, Arabic speech commands were classified. Mel Frequency Cepstral Coefficients were used to create an effective feature vector (MFCC). Spectral centroids and signal power are combined with MFCC to generate an enlarged features vector, which improves recognition accuracy. Because these commands will drive a robot in real time, they must be classified quickly. As a result, the training features vector's dimension is lowered by performing some statistical calculations on it. Support Vector Machines (SVM), Random Forest decision tree (RFT), Neural Network (Multi-Layer Perceptron, MLP), and k-nearest neighbors (KNN) approaches were employed as intelligent classifiers. A thorough examination of the classifiers' various parameters was carried out. With a classification accuracy of 94.84 percent, the SVM approach outperformed other techniques. We concluded in this research that the enlarged features vectors with a lower dimension are more effective for this problem and can be employed in real applications.

General Terms
Pattern Recognition, Machine Learning Algorithms

Keywords
Machine-Learning techniques, (SVM), KNN, MLP, Random Forest Trees, MFCC

1. INTRODUCTION
The COVID-19 epidemic began in Wuhan, China, and has since spread throughout the world. It spreads quickly among humans. So far, the World Health Organization (WHO) has documented more than 147 million confirmed COVID-19 infections and more than 3 million deaths in 223 countries [1]. Several vaccinations have been released to combat COVID-19 [2]. Despite the presence of vaccines, the virus has no end. Masks and maintaining social/physical distances remain the most crucial key principles for everyone fighting COVID-19 [3]. Elders and children, in particular, have a high priority to stay safe during the pandemic because they are the most vulnerable to the virus and are unlikely to recover.Robotics is one of the most essential technologies that could help the healthcare system, and its importance has grown significantly because of the COVID-19 Pandemic. Robots have recently been employed in tele assistance for many hospitals across the world for various services such as supplying basics, prevention, screening, diagnosing, treatment, home care, and sterilizing many locations in a separate and safe manner. As a result, robots may be able to assist in the reduction of social distance, as recommended by WHO for COVID-19 awareness. [4] [5] and [6]. Designing a socially interactive robot makes it easier for robots and humans to communicate. They may learn things spontaneously and perform well [7]. Several artificial intelligence-based Human Machine Interacting (HMI) approaches have been discussed in the literature, including EEG signals and eye blinks [8], Sip and puff, tongue, touch screen, head movement, gaze tracking, and voice instructions are some of the methods used [9,10].Speech is the simplest form of interactive communication. As a result, rather than traditional robots, building a speech-controlled robot that responds to a set of orders is more beneficial to the society. Different languages have been used to address different voice-controlled robots in the literature [12]. Arabic is one of the most commonly spoken languages in the world, with over 400 million people speaking it in various dialects. In comparison to other languages, it receives minimal attention in literature. This is owing to the nature of Arabic and the scarcity of datasets. As a result, gathering an adequate dataset containing the Egyptian dialect is required in order to develop a durable Arabic voice-controlled robot. The data is gathered by moving the robot in four directions: forward, backward, right, left, and to stop. For the commands in our dataset, we employ a variety of Egyptian colloquial words: advance, stop, and left. Table 1 lists the most commonly used Arabic commands and their English translations. The goals of this work are as follows:

1. Create an original Arabic speech dataset and ensure that the data is properly positioned and prepared.
2. Select the most effective features vector for capturing the acoustic properties of Arabic voice samples.
3. Present an analysis of several voice command classification approaches based on machine learning techniques.
4. Finally, conduct a comparison and analysis of all machine-learning techniques that have been used. The next is how the paper is structured: The relevant

work is introduced in Section 2. Section 3 is a summary of the proposed classification system. In Section 4, the feature extraction method is illustrated. Section 5 explains statistical and feature reduction. In section 6, we discussed machine learning. The experimental data and comments are presented in Section 7. Finally, Section 8 has the conclusion.

2. RELATED WORK

Hassan, Hani S. et al. [10] designed a recognizer to assist a wheelchair robot. The recognizer mission is to classify four Arabic speech commands " يسار"و "يمين"و "قف "اذهب""". They have constructed their own dataset, preprocessed, and trained over Artificial NN. They acquired a 94.4% classification accuracy. To that end, they could not evaluate their classification results over other algorithms. In addition, they used limited commands for the wheelchair without any consideration for the different dialects. Qidwai, Uvais et al. [11] design a voice-controlled module to be built in a robot. They used the EasyVR module that receive the data from the UART port found in ATMega module. Their work does not use artificial intelligence for recognition task. UvaisQidwai in [13] used fuzzy logic to control a wheel chair using Arabic commands. They measured the input signal energy level resulting from filter bank response. FIS Rule-based method is further used to differentiate between different commands. They did not consider different speakers to get each commend suitable band ranges. Thus, the dedicated band could not be generalized for any speaker especially in noisy environment. In a plus attaching, a notebook to the chair could affect the total cost.Several researches proposed the speech recognition problem for either Arabic or other languages. In addition, the classification task may be done on certain commands or over different dialects. Alasadi, Abdulmalik A., et al [14] compare between several features to develop Arabic ASR system. They compared between Mel-Frequency Cepstral Coefficients (MFCC), Power- Normalized Cepstral Coefficients (PNCC), and Modified Group Delay Function (ModGDF) with using SVM classifier. Their system shows that PNCC is the best feature over MFCC with 9% drop in classification accuracy. Deshwal et al. [15] have to build language identification system. They collected a database from 50 utterances with four languages: Tamil, Malayalam, Hindi and English. In order build their system, they used different hybrid features combinations to reach the highest accuracy. The different features combinations are such: MFCC with (MFCC and delta features) DMFCC, perceptual linear prediction features (PLP) with Delta D PLP, RASTA-PLP, and Shifted delta cepstrum (SDC). The classification over Feed-forward Artificial NN shows that the best feature combination is for MFCC-RASTA-PLP.Marlina, Lina, et al.16]] developed HijaiyahMakhraj system to differentiate between Arabic Quran recitation pronunciations. They extracted MFCC features from the Quran Audio then classified by SVM classifier with RBF kernel. Al-Omari, Ayoub Abdelrahman [17] used to build ASR system. They used MFCC features with three different classifiers: Data Time Wrapping (DTW), Hidden Markov Model (HMM) and Dynamic Bayesian Network (DBN). The results show the highest word error rate is achieved with DTW.LubnaEljawad et al. used discrete wavelet coefficient to recognize the Arabic speech commands. The input voice commands are preprocessed for removing noise, sample resizing, removing DC value. They used both neural network and Suggeno fuzzy logic to classify the wavelet coefficient. The results are applied on male and female speech that shows superiority of NN over fuzzy logic by 17% in classification accuracies [18].Elvira SukmaWahyuni used to classify the hijaiyah of three Arabic speech letters with almost near Makharij "sa, sya, and tsa". They extracted MFCC that classified by ANN to recognize each voice. They achived 92.4% average classification accuracy [19].

3. THE PROPOSED SYSTEM

Figure 1 depicts the proposed methodology, which is made up of various modules. The Egyptian Arabic dataset is collected in the first step from a variety of speakers. Using a mobile cell, the sound waves of eight orders are continually recorded. The next stage is to preprocess and clean our data, which may be done in a variety of ways. In the third module, the appropriate features from the sound waves are picked and tested. The dimension of the training data should be lowered for quick categorization. Various machine-learning approaches are used to build and test the data model.

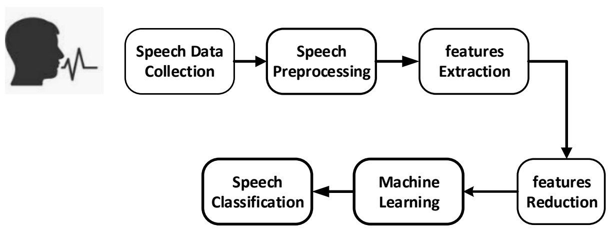

Fig. 1: The Schematic Diagram of the Proposed Methodology

The following sections will go through each block in the previous block diagram in detail.

3.1 Egyptian Arabic Speech Commands (EASC) Data collection process

The data we used in this study was originally gathered from ordinary people. Arabic is a challenging language to learn because of various dialects, which have an impact on how it is pronounced. Various speakers' speech signals differ depending on their age, gender, dialects, speaking style (slow or quick), and the surroundings (such as streets, homes, and labs), all of which have different background noises. These variations contribute to the difficulty of the Arabic recognition task. As a result, in order to develop a more generic model, we gathered the Egyptian Arabic Speech Commands (EASC) dataset, which covers all of these variables.As mentioned earlier, the purpose of these dataset is to be used in controlling a mobile assistant robot. Thus, we concerned with collect a motion-based commands such as forward, backward, stop, right, left in Arabic. Due to the Egyptian dialects, some meaningful command was recording twice as forward will be (اوادم ,اماك)) and stop will be (قف),(أوقف). Table. 1 shows each recorded command meaning.

Table 1. Egyptian Arabic Commands

Arabic Command	Translation in English	Arabic Command	Translation in English
"قف"	Stop	"امام"	Forward
"أوقف"	Stop	"اودام"	Forward
"يسار"	Left	"يمين"	Right
"شمال"	Left	"ارجع"	Backward

The dataset is recorded from 90 speakers. Each one records eight words in Arabic (أمام-أدام-يمين-يسار-شمال-قف-أوقف-ارجع) in one record. The commands are recorded by a mobile cell phone and stored with wave format files with 44100 Hz sampling rate. As eight words have been recorded from each speaker for the 90 speakers, the whole set becomes 720 Commands. Figure 2 shows samples of speech signals of the eight voice commands.

3.2 Data Analysis

Data is only meaningful with the proper preprocessing and preparation. Before extracting the signal features, the speech signal passed through many steps regarding the length and the amplitude. The start and the end of each recorded command were manually cut to isolate it from the other commands such that the volunteers say all commands in one shot. Then the speech signal was read by MATLAB software and saved as a raw data. Padding was done to keep the same length for all the recorded words. The magnitude of some recorded signals was higher than other signals depending on the speaker’s volume and intensity, so the signals were normalized to have the same scale. Finally, the resulted speech signals are entered to the next module, which is the module for extracting features.

4. FEATURES EXTRACTION

The waveform of a speech signal is quite complicated; it contains a huge dynamic range of numerous frequency components in the short-time spectrum, and it has a lot of variability. Some sort of feature extraction is used to reduce the variability and data size of the voice signal. Many studies have concentrated on identifying useful elements for modelling speech patterns. There are several methods for obtaining these characteristics. Mel-Frequency Cepstral Coefficients are the most commonly used (MFCC). MFCC, Centroid, and signal power are used to extract features in this research.

4.1 Mel-Frequency Cepstral Coefficients

Melodic frequency Coefficients of Cepstral Position MFCCs are one of the most widely used frequency domain feature extraction techniques in speech recognition, based on the Mel scale, which is based on the human ear scale. MFCCs are much more accurate than time domain features when considered as frequency domain features [20]. The following steps are commonly used to derive MFCCs from a speech signal: Divide the signal into frames (typically 20-25 ms with 10 ms overlap) using a windowing function (such as a Hamming window, as shown in equation (1)).

$$W(n) = 0.54 - 0.46 \cos\left(\frac{2\pi n}{N-1}\right),\ 0 \le n \le N-1 \qquad (1)$$

The Fourier transform is then used to compute the spectrum of each frame. The power spectra is then subjected to a mel-filter bank. The Mel is made up of triangular overlapping windows that span the entire frequency range. These filter banks simulate the non-linear sound perception of the human ear. Lower frequencies are more discriminative, while higher frequencies are less discriminative. Each filter output is passed to the log function, and the features processed in this manner are referred to as FBANKs. To further select a subset (usually 13) of the transformed features, a discrete cosine transform of the FBANK features is used. Figure 3 depicts a sample block diagram of the MFCC feature generation steps.

4.2 Spectral Centroids

The spectral centroid, like the center of mass in physics, is thought to be the 'gravity' of the spectrum. Its value is determined by computing the "center of gravity" using the frequency and magnitude information from the Fourier transform. The average frequency weighted by amplitudes, divided by the total of the amplitudes, is the individual centroid of a spectral frame, as determined by equation (2).

$$spectral\ centroid = \frac{\sum_{k=1}^{N} k\, F[k]}{\sum_{k=1}^{N} F[k]}$$

(2)Where, F [k] is the amplitude corresponding to bin k in discrete Fourier transform. In practice, centroid finds this frequency for a given frame, and then finds the nearest spectral bin for that frequency [21], [22].Feature vector reduction was performed by applying some mathematical statistical operation on MFCC coefficients.The Maximum, Minimum, Mean, Standard deviation, Skewness, Kurtosis, Median values are computed. For example, the MFCC coefficients of command أمام, has 25365 vector length) was found as 55*14 and after getting the seven statistics, the sample dimension becomes 7*14. This matrix was converted to a row matrix to get the feature vector, the database becomes with dimensions of 720 * 98 = 70,560. The same seven statistics were also computed for the centroid feature to represent its pattern with lower dimension, and the seven statistics of centroid were concatenated to the preceding MFCC to produce a new feature vector with dimension 98+7 = 105. In addition, the value of the power of each signal is added to get the final feature vector as, MFCC, Centroid, and the signal Power, “MFCC-Cent. -P” with 105 + 1 = 106 length.

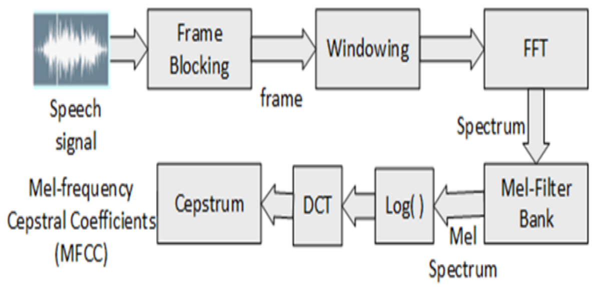

Fig 3. Block Diagram of MFCC Process

5. MACHIN LEARNING TECHNIQUES

Four different machine learning classification approaches are used to classify the eight spoken requests. K closest neighbor (KNN), Support Vector Machine (SVM) with different kernel functions (SVM), Random Forest Trees, and Multi-Layer Perceptron are some of them (MLP).

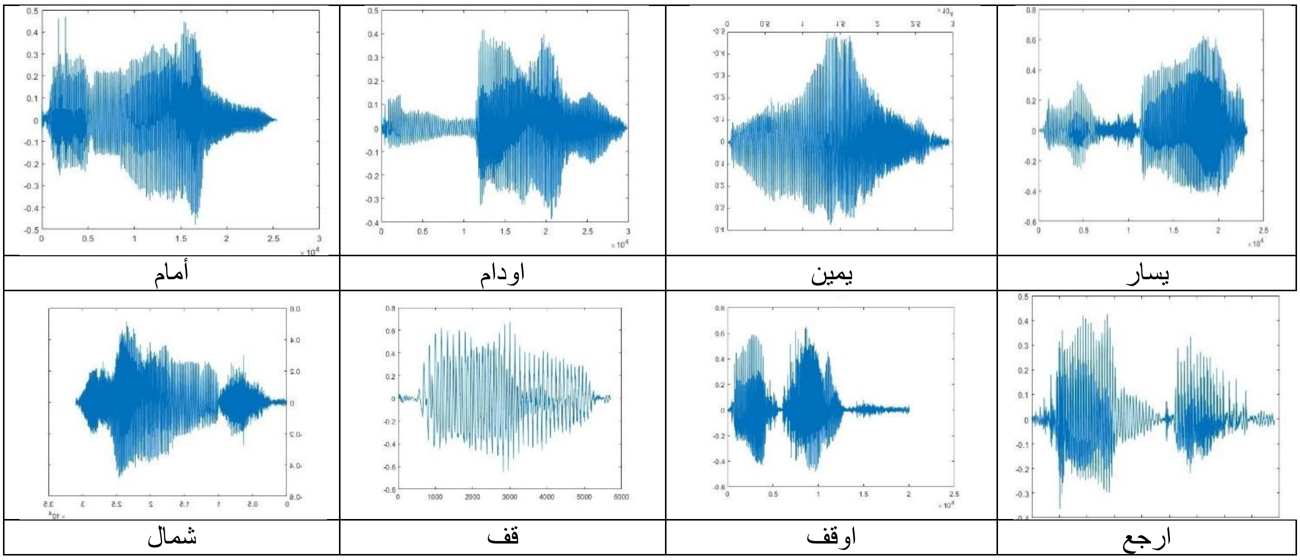

Fig 2: Sample of recorded waveforms of the Eight Voice Commands

5.1 K-Nearest-Neighbor Classifiers

KNN is a non-parametric simple classifier that can be selected for a classification study when there is no prior knowledge about the data distribution [23]. In the method the similarity between the test pattern and every pattern in the training set are measured and classified, k is the number of neighbors to be considered. Thus, k is the parameter required to be determined. KNN uses different types of distance measure between the test point and the training points assuming the two points are p (x1, y1) and q (x2, y2). The three types of distance measure are computed as in equations (4-6).

$$Euclidean\ distance = \sqrt{(x_1 - x_2)^2 - (y_1 - y_2)^2} \quad (4)$$

$$Manhattandistance = (|x_1 - x_2| + |y_1 - y_2|) \quad (5)$$

$$Chebyshevdistance = \max(|x_1 - x_2|, |y_1 - y_2|) \quad (6)$$

The performance of the KNN algorithm may be degraded by the existence of noise.

5.2 Support Vector Machine (SVM)

SVM is used for many applications especially for classification problems [24]. In SVM, the training data is represented as points in space separated into groups by gaps. The concept of SVM is to get the hyperplane (decision boundary) that separates these groups with maximum separation margin. New observation is then mapped into that space and its class is estimated based on region of the test data it falls. This happened for linear separable data but for nonlinear data, the points are inseparable so kernels trick are suggested [25]. By using Kernels, the data is transformed into another representation space makes it linearly separable. There are many standard Kernels such as polynomial, radial basis function (RBF) [26], PUK kernel functions [27]. Selecting the best kernel for a given training data is a challenge. Let x be the data set and let yi be the class label of xi. The polynomial kernel with exponent d, the RBF kernel, and the Pearson VII function-based, and universal kernel PUK are described by equations (7), (8), and (9).

$$k(x, y)_{poly.} = (x^T y + 1)^d \quad (7)$$

$$k(x, y)_{RBF} = \exp\left(-\frac{\|x - y\|^2}{2\gamma^2}\right) \quad (8)$$

$$k(x, y)_{PUK} = \frac{1}{1 + \left[\left(\frac{2\sqrt{\|x - y\|^2}\sqrt{2^{(1/\omega)} - 1}}{\sigma}\right)^2\right]^{\omega}} \quad (9)$$

There are two approaches to multi-class SVM: one-vs-all, in which each class discriminates against the remaining classes, and (one-vs-one), in which each class discriminates between every pair of classes, for all possible pairings [28]. This work is based on a single vs. single multiclass classifier.

5.3 Random Forest Decision Trees

One of the most powerful supervised machine learning algorithms is Random Forest. It is made up of groups of decision trees that work together to make predictions. It builds several decision trees based on observations and certain attributes at random, and the outcomes are then averaged or based on the majority vote. Each tree chooses a subset of features at random (feature bagging) from all features, and the corresponding feature space is distinct (but fixed) for each tree. The number of trees and characteristics utilized for training at each node split are the method's key setup parameters [29]. The Random Forest technique can accommodate a wide variety of data items and provide extremely high accuracy; there is no need to scale data in the Random Forest algorithm, even when a major amount of the data is missing, it provides high accuracy. However, complexity and its execution are more difficult and time-consuming, necessitating the use of computer resources.

5.4 Multi-Layer Perceptron Neural Network

The artificial neural network is the most well-known of the traditional classification methods. The ability to tackle complicated issues is the major benefit of neural network approaches. A feed forward supervised machine learning algorithm is a multilayer perceptron (MLP). MLP is made up of an input layer that receives the signal, an output layer that chooses what to do with it, and one or more hidden layers in between [30].

6. EXPERIMENTAL STUDY AND ANALYSIS

This study used WEKA software which is an open source available tool for several Machine Learning algorithms[1]. The eight speech commands are identified using various classifiers.

6.1 KNN Classifier Results

Table 2 shows the highest classification accuracy for Euclidean, Manhattan, and Chebychev measured distances for various values of k (ranging from 1 to 11). It demonstrates that utilizing an enhanced feature vector (MFCC-Cent. -P) increases accuracy over just using MFCC features. If just MFCC is employed, the greatest accuracy value is 77.7% for Manhattan distance, but it rises to 82.29 percent for k = 8 with Euclidean distance for MFCC-Cent. -P feature vector.

Table 2: KNN best parameters, and accuracies for MFCC and MFCC-C-P

Features Extraction	**Euclidean distance**		**Manhattan distance**		**Chebychev distance**	
	k	**Accuracy**	**k**	**Accuracy**	**k**	**Accuracy**
MFCC	9	76%	10	77.7%	9	56.9%
MFCC-Cent. -P	**8**	**82.29 %**	8	81.94%	9	62.5%

6.2 Random Forest Classifier Results

In two steps, the random forest algorithm is trained. To begin, a fixed number of trees is chosen (100 trees as a starting value), and the number of subset characteristics is varied with a range of 2:16 to capture the best results. For both MFCC and MFCC-Cent-P, it was discovered that subset characteristics 4 and 8 provide superior accuracy as shownin Table 3. Second, as indicated in Table 4, the number of subset features is constant while the number of trees is varied until the highest Accuracy is obtained.

Table 3: Random Forest best parameters selection for MFCC and (MFCC-Cent. –P)

Random features	**MFCC (100 trees)**	**MFCC-C-P (100 trees)**
2	86.45 %	89.2 %
4	86.45 %	90.6 %
8	87.84 %	91.3 %
16	84.7 %	88.5 %

Table 4: Random Forest best percentage classificationaccuracies for MFCC and (MFCC-Cent. -P)

Features vector	**Random subset feature**	**100 trees**	**200 trees**	**300 trees**	**400 trees**	**500 trees**	**600 trees**
MFCC	4	86.45	87.2	86.8	87.15	87.15	87.15
	8	**87.84**	86.1	85.7	85.76	85.76	85.76
MFCC-Cent. -P	4	90.6	91.3	92.3	91.3	91.3	91.3
	8	**91.3**	90.9	91.3	85.76	91.6	91.3

Random forest yields 87.84 percent classification percentage for MFCC under 8 subset features and 100 trees, and 91.3 percent for (MFCC-Cent. –P) features, according to the trials.

6.3 MLP Classifier Results

At the beginning, one hidden layer of the MLP structure is created, and various hidden node values are chosen as given in table 5. The structure of two hidden layers is then examined, as well as different values of hidden nodes in the first and second levels, with the results presented in Tables 5 and 6. (Note that the symbol "a" is equal to (number of features + number of classes) / 2).

Table 5: Classification accuracy percentage for one hidden layer MLP for MFCC, and (MFCC-Cent. -P), learn-rate=0.3, momentum=0.2, epoch =500 iteration.

No. of hidden nodes	7	20	25	50	a
MFCC	84.7	88.88	88.19	87.85	89.23
MFCC-Cent. -P	90.9	93.09	93.44	93.44	**94.44**

Table 6:Classification accuracy percentage for two hidden layers MLP for MFCC

Number of hidden nodes in second hidden layer	Number of hidden nodes in first hidden layer				
	7	**20**	**25**	**50**	**a**
5	77.77	85.09	82.6 5	84.72	83.33
10	83.33	85.06	85.41	84.02	84.72
20	84.72	88.88	86.11	82.63	87.5
50	84.72	86.8	88.88	86.46	86.45

[1]https://www.cs.waikato.ac.nz/ml/weka/

Table 7. Classification accuracy percentage for two hidden layers MLP for (MFCC-Cent. –P)

Number of hidden nodes in second hidden layer	Number of hidden nodes in first hidden layer				
	7	**20**	**25**	**50**	**a**
5	86.8	92	92	86.45	88.88
10	90.97	86.8	91.31	89.58	87.5
20	90.2	91.66	92.36	92.01	89.93
50	88.5	90.97	91.3	90.97	93.75

The optimum MLP structure is chosen, as shown in Table 8.

Table 8: MLP for MFCC and (MFCC-Cent. –P), learn-rate=0.3, momentum=0.2, epoch =500 iteration.

	One hidden layer		two hidden layers		
	No. of hidden nodes	Accuracy	No. of nodes in *first* hidden layer	No. of nodes in *second* hidden layer	Accuracy
MFCC	a	89.23%	20	20	88.88%
MFCC-Cent. -P	a	94.44%	a	50	93.75%

6.4 SVM Classification Results

Many tests are carried out in order to find the optimum kernel function and its parameters. The value of exponent parameter d was altered from 0.2 to 5 for polynomial kernel, and C was changed from 1 to 10, resulting in classification accuracy of 89.58 percent for MFCC and 94.44 percent for (MFCC-Cent. –P), as shown in Table 9. RBF kernel with C=10, and γ=0.45 gives 95.48% for MFCC-C-P features vector. PUK kernel with parameters C=1, ω=1, and σ = 4 gives the good classification accuracy for MFCC features only.

Table 9. The best Classification Accuracy percentage for SVM with Different Kernel functions.

Feature vector	**Polynomial**			**PUK**				**RBF**		
	d	**C**	**Acc.**	**σ**	**ω**	**C**	**Acc.**	**γ**	**C**	**Acc.**
MFCC	0.7	1	89.58	4	1	1	**89.9**	0.1	10	89.58
MFCC-Cent. -P	0.7	1	94.44	4	0.5	1	92.7	0.45	10	**95.48**

It is clear from Table 9 that SVM with radial basis function for (MFCC-Cent. -P) is more accurate such that it gives 95.48%.

6.5 Comparison Study

Table 10 shows a comparison of all classifiers at the conclusion. The support vector machine approach clearly wins the competition and produces the highest accurate classification accuracy of 95.48 percent.

Table 10: Classification Precisions for of All techniques

Feature vector	**KNN**	**Random Forest trees**	**ANN**		**SVM**		
			1-hidden layer	**2-hidden layer**	**(RBF)**	**(poly.)**	**(PUK)**
MFCC	77.71%	87.84%	89.23%	88.88%	89.58%	89.58%	89.9%
MFCC-Cent. -P	82.29%	91.3%	94.44%	93.75%	**95.48%**	94.44%	92.7%

According to the experimental findings in Table 10, the greatest accuracy using just MFCC features is 89.9%, while the best accuracy using augmented features of MFCC plus Centroid and signal Power (MFCC-Cent. -P) is 95.48 percent. As a result, combining the Centroid and Power characteristics with MFCC improves command recognition. As a conclusion, the SVM classifier is suggested. It should be noted that this promised result is attained despite the fact that the used dataset contains varying noise levels and noise kinds.

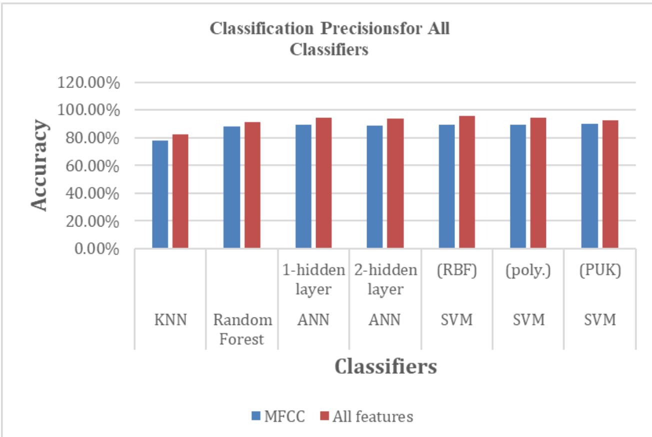

Fig.3: A comparative chart of all classifiers

7. CONCLUSION

A multi-class classification problem is tackled in this paper. Four classification approaches (KNN, RFT, MLP, and SVM) are compared. Each algorithm is fine-tuned, trained, and tested independently. The Egyptian Dialects Arabic dataset (EASC) was created and preprocessed with the goal of training machine learning models with it. For a variety of reasons, categorizing Arabic instructions is challenging. Both signals are no longer the same if the same word is spoken at various speeds (speech rate) and in different settings. Furthermore, the audio stream is impacted by gender, age, and health state when the speaker is changed.

To cope with all of these variables and influences, and to capture the Arabic acoustical properties of the sounds, a valuable features vector should be chosen, and the data should be modelled using an intelligent classifier. A well representative integrated feature vector is examined using an effective combination of Spectral Centroid Frequency and signal power with Mel-Frequency Cepstral Coefficients (MFCC-Cent. -P) feature vectors. By employing statistical approaches, the feature vector dimension is lowered, allowing the classification process to be finished fast and employed in real-time applications. The suggested short integrated features vector enhances classification accuracies while reducing computational complexity, according to the results. The SVM classifier was likewise shown to be superior in this study, with a promise accuracy of 94.84 percent. In the future, the results of this study will be applied to our project "Arabic speech-controlled assistance robot".

8. REFERENCES

[1] Mohan BS, Nambiar V (2020) COVID-19: An Insight into SARS-CoV-2 Pandemic Originated at Wuhan City in Hubei Province of China. J Infect Dis Epidemiol 6:146. doi.org/10.23937/2474-3658/1510146

[2] RS Mahla, LB Dustin, "Lessons from a large-scale COVID-19 vaccine trial ", The Journal of Clinical Investigation, https://doi.org/10.1172/JCI163202.

[3] Fauci, Anthony S. "The story behind COVID-19 vaccines." (2021): 109-109

[4] Pfefferbaum, Betty, and Carol S. North. "Mental health and the Covid-19 pandemic." New England Journal of Medicine 383.6 (2020): 510-512.

[5] Di Lallo, Antonio, et al. "Medical Robots for Infectious Diseases: Lessons and Challenges from the COVID-19 Pandemic." IEEE Robotics & Automation Magazine 28.1 (2021): 18-27.

[6] Vindegaard, Nina, and Michael Eriksen Benros. "COVID-19 pandemic and mental health consequences: Systematic review of the current evidence." Brain, behavior, and immunity 89 (2020): 531-542.

[7] Islam, Mohammad Shamim, et al. "Design of A Social Robot Interact with Artificial Intelligence by Versatile Control Systems." IEEE Sensors Journal (2021).

[8] Al-Wesabi, Fahd N., et al. "Smart Mind-based Approach to Control Wheelchair Wirelessly." Sensors and Materials 32.4 (2020): 1533-1555.

[9] Korayem, M. H., et al. "Design and Implementation of the Voice Command Recognition and the Sound Source Localization System for Human–Robot Interaction." Robotica: 1-12.

[10] Hassan, Hani S., and S. Jammila Harbi. "Arabic Command Based Human Computer Interaction." Journal of Physics: Conference Series. Vol. 1530. No. 1. IOP Publishing, 2020.

[11] Qidwai, Uvais, and Mohamed Shakir. "Ubiquitous Arabic voice control device to assist people with disabilities." 2012 4th International Conference on Intelligent and Advanced Systems (ICIAS2012). Vol. 1. IEEE, 2012.

[12] Lukas Grasse, Sylvain J. Boutros, and Matthew S. Tata, "Speech Interaction to Control a Hands-Free Delivery Robot for High-Risk Health Care Scenarios", Published online 2021 Apr 8., PMCID: PMC8060474, doi: 10.3389/frobt.2021.612750.

[13] UvaisQidwai; Fatma Ibrahim, "Arabic speech-controlled wheelchair: A fuzzy scenario", 10th International Conference on Information Science, Signal Processing and their Applications (ISSPA 2010) DOI:10.1109/ISSPA.2010.5605493.

[14] Alasadi, Abdulmalik A., et al. "Efficient Feature Extraction Algorithms to Develop an Arabic Speech Recognition System." Eng. Technol. Appl. Sci. Res 10.2 (2020): 5547-5553.

[15] Deshwal, Deepti, Pardeep Sangwan, and Divya Kumar. "A language identification system using hybrid features and back-propagation neural network." Applied Acoustics 164 (2020): 107289.

[16] Marlina, Lina, et al. "Makhraj recognition of Hijaiyah letter for children based on Mel-Frequency Cepstrum Coefficients (MFCC) and Support Vector Machines (SVM) method." 2018 International Conference on Information and Communications Technology (ICOIACT). IEEE, 2018.

[17] Al-Omari, Ayoub Abdelrahman. A Comparative Study of Classification Techniques for English to Arabic Speech Recognition. Diss. Middle East University, 2016.

[18] LubnaEljawad, Rami Aljamaeen, Mutasem K. Alsmadi, Ibrahim Al-Marashdeh, HayamAbouelmagd, Sanaa Alsmadi, Firas Haddad, Raed A. Alkhasawneh, Mohmmed Alzughoul and 5Malik B. Alazzam, " Arabic Voice Recognition Using Fuzzy Logic and Neural Network" , International Journal of Applied Engineering Research ISSN 0973-4562 Volume 14, Number 3 (2019) pp. 651-662 © Research India Publications. http://www.ripublication.com

[19] Elvira SukmaWahyuni, "Arabic speech recognition using MFCC feature extraction and ANN classification" , 2017 2nd International conferences on Information Technology, Information Systems and Electrical Engineering (ICITISEE). DOI: 10.1109/ICITISEE.2017.8285499.

[20] Elvira SukmaWahyuni "Arabic Speech Recognition Using MFCC Feature Extraction and ANN Classification. " , 2017 2nd International Conferences on Information Technology, Information Systems and Electrical Engineering (ICITISEE), 978-1-5386-0658-2/17/$31.00 ©2017 IEEE.

[21] https://www.cs.cmu.edu/~music/icm/slides/05-algorithmic-composition.pdf.

[22] Jia Min Karen Kua, TharmarajahThiruvaran ,MohaddesehNosratighodsEliathambyAmbikairajah, Julien Epps, "Investigation of Spectral Centroid Magnitude and Frequency for Speaker Recognition ", Odyssey 2010 The Speaker and Language Recognition Workshop 28 June – 1 July 2010, Brno, Czech Republic. file:///C:/Users/azza/Downloads/InvestigationofSpectral CentroidMagnitudeandFrequencyforSpeakerRecognition.pdf

[23] Bayraktar, R., Akgul, B. A. & Bayram, K. S. (2020). Colour recognition using colour histogram feature extraction and K-nearest neighbour classifier. New Trends and Issues Proceedings on Advances in Pure and Applied Sciences. (12), 08–14. DOI: 10.18844/gjpaas.v0i12.4981.

[24] Ur¨unDo˘gan, Tobias Glasmachers, and Christian Igel. "A Unified View on Multi-class Support Vector Classification ", Journal of Machine Learning Research 17 (2016) 1-32.

[25] Muhammad Achirul Nanda, KudangBoro Seminar, Dodi Nandika, and AkhiruddinMaddu "A Comparison Study of Kernel Functions in the Support Vector Machine and Its Application for Termite Detection', Information 2018, 9, 5; doi:10.3390/info9010005.

[26] Khalid AA Abakar&chongwen Yu, " Performance of SVM based on PUK in comparison to SVM based on RBF kernel in prediction of yarn tenacity", Indian Journal of Fibre& Textile Research, Vol.39, March 2014, pp. 55-59.

[27] Santosh Chapaneri, Renia Lopes, and DeepakeJayaswel "Evaluation of Music Features for PUK Kernel based Genre Classification.", International Conference on Advanced Computing Technologies and Applications (ICACTA-2015), © 2015 Published by Elsevier B.V. This is an open access article under the CC BY-NC-ND license.

[28] A. Mathur; G. M. Foody, "Multiclass and Binary SVM Classification: Implications for Training and Classification Users ", IEEE Geoscience and Remote Sensing Letters (Volume: 5 , Issue: 2 , April 2008), DOI: 10.1109/LGRS.2008.915597.

[29] HanyangZhong; Xin Song; Lei Yang "Vessel Classification from Space-based AIS Data Using Random Forest", 2019 5th International Conference on Big Data and Information Analytics (BigDIA), DOI: 10.1109/BigDIA.2019.8802792.

[30] Mohammed BEN NASR,& Mourad TALBI, AdnaneCHERIF, "Arabic Speech Recognition by MFCC and Bionic Wavelet Transform using a Multi-Layer Perceptron for Voice Control", 2012 6th International Conference on Sciences of Electronics, Technologies of Information and Telecommunications (SETIT).