DOI: https://doi.org/10.33103/uot.ijccce.21.3.10

Impostor Recognition Based Voice Authentication by Applying Three Machine Learning Algorithms

Ashraf Tahseen Ali[1], Hasanen S. Abdullah[2] , Mohammad N. Fadhil[3]
[1,2,3] *Computer science Dept., University of Technology, Baghdad, Iraq.*
[1]*cs.19.12@grad.uotechnology.edu.iq*, [2]*110014@uotechnology.edu.iq*
[3]*mohammad.n.fadhil@uotechnology.edu.iq*

Abstract – As compared to other conventional biometrics systems, voice is a unique and important metric, where it is used in many vital fields as the security and communication domains that do not need to be expensive to achieve. The purpose of this article is to see how machine learning (ML) algorithms perform for speaker Authentication to recognize impostors. To boost the audios usable in real environments, it was suggested the preprocessing of audio, like noise decreasing and voiced improving. Mel Frequency Cepstral Coefficients (MFCC) and the four features (Amplitude, Zero Crossing, Mean, and Standard Division) are extracted for all audio metrics, straight beside their differentials and accelerations. Then, Vector Quantization (VQ) is applied to these files. The algorithms were prepared and examined on two datasets, by applying k-fold cross-validation. The preparation for testing and comparing the three (ML) approaches are as follows: Support Vector Machine (SVM), One Rule (One-R), Linear Regression (LR). The result of the (SVM) algorithm average accuracy of 96.33 percent was superior.
Index Terms*–* *ML, Speaker Authentication, SVM, LR, ONE.R, Impostor*.

I. INTRODUCTION

Speaker Authentication has been around since the eighth decade of the last century. The audio features are used to identify the speaker's voice that has been found to vary among human [1]. Speak identification methods are economical and effortless to be used. In present time, speech verification is essential in a mixture of approaches. Voice-Activated (VA) business, house computerization, and (VA) machines are only some of the various applications for speech identification. The operation of identifying persons based on their sound signal is identified as speaker identification. In order that of changes in that frames of the sounds field, the extent of the larynx, and additional parts of the speech generation organs, every person's voice may be distinguished. Considering voice recognition need to be handled in different conditions, the features extracted must also be resistant to surrounding noise and sensor failure. The identification technology enables the speaker's voice to be applied to authenticate their identification and monitor entrance to places, companies and institutions to which they belong [2]. The speaker recognition sections can be represented as shown in *Fig.*1, speaker recognition system stages can be described as shown in *Fig.* 2 [3]. The preprocessing is the initial stage of other levels in speaker recognition to separate the voiced or other signal and generate feature vectors. The important and common steps used are such as surrounding noise removal, pre-emphasis, voice activity discovery, windowing and framing. It can be clarified that the method of processing speaker recognition differs according to the type of data set and its format, as if (wave or mp3) and both are the most common. In terms of the amount of noise existed in each audio clip, the preprocessing is needed to remove noise and useless data. As for feature extraction, many various methods were used by researchers which achieved their purpose such as MFCC by

Received 31/5/2021; Accepted 14l8l2021

@ 2017 University of Technology, Iraq **ISSN (Print) 1811-9212** **ISSN (Online) 2617-3352**

DOI: https://doi.org/10.33103/uot.ijccce.21.3.10

preform appropriate parameters. Using a classifier from machine learning that leading to achieve powerful results. Briefly, these are the main important keys in the method of identifying the speaker.

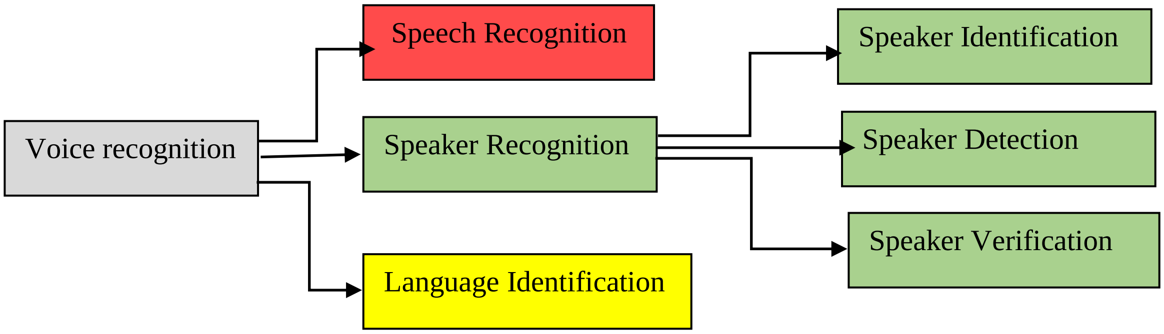

FIG. 1. SPEAKERS RECOGNITION SECTIONS

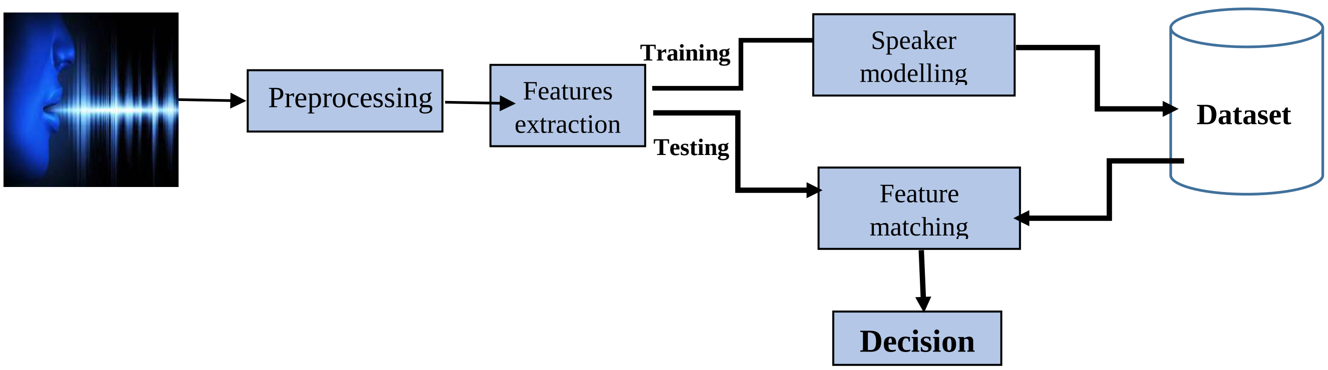

FIG. 2. DESIGN OF THE SPEAKER RECOGNITION METHOD

In addition, *Fig*.3. below will show the feature extraction methods, the most common and efficient in speaker recognition technology in the last ten years.

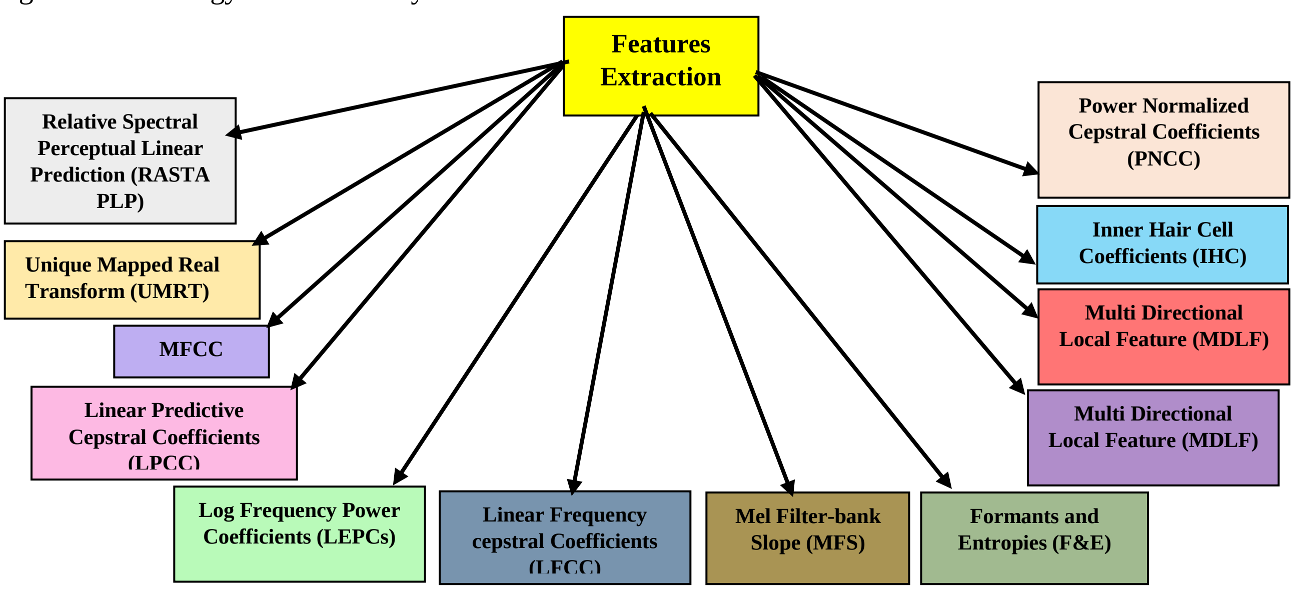

FIG 3. FEATURES EXTRACTION METHODS

Received 31/5/2021; Accepted 14l8l2021

DOI: https://doi.org/10.33103/uot.ijccce.21.3.10

II. LITERATURE REVEIW

Speaker Recognition (SR) is a computerized system of recognizing persons on the basis of their voice signal, which is a biometric approach like different biometrics such as finger veins, Palm, Retina, Iris, and Face recognition. The main factor which makes (SR) from other biometrics is that (SR) can be determined as the only technique that prepares voiced data, in contradiction with else methods, that generally use image information [3]. In this section, some of the past literature linked to this research will be discussed:

Paulose, et al. [4], suggested that voices-sources features and spectro-temporal features be applied to introduce identification methods. Method is done with two various classifiers for the I-Vector (I.V) approach, and the accuracy percentages are compared. The used for methods for features extraction were Hair Cell Coefficients (IHC) and MFCC. The performance of a couple of different speech identification methods was determined to be compared. The study shows that Gaussian Mixture Modeling (GMM) outperforms i-vectors for short speech, with an accuracy rate of 94.33 percent Thiruvengatanadhan. [5], developed a speak identification system applying SVM. Voice Activity Detection is used to distinguish unique words from connected conversations. The features of every separate term were selected, and the forms were fortunately trained. SVM was applied to model each human utterance. The MFCC is a list of features that are used to define audio material. SVM was applied to identify speaking by learning from training data. According to experimental findings, the audio SVM method has a high performance in 95 percent speaking identification. Chauhan, et al. [6] suggested speaker recognition system using SVM algorithm. Linear Predictive Coding (LPC) and Mel Frequency Cepstral Coefficients (MFCC) applied for features extraction this method achieved accuracy 80.6 percent. M. Subba, et al. [7], suggested divided the study into the following sections: preprocessing of audio, feature extraction in which (MFCC) are extract for each voice, and applying the Random Forest (RF) algorithm, accuracy is achieved 84 percent. Rao, et al. [8], presented various techniques of audio preprocessing such as trimming, split and merge, noise reduction, and voice improvements to improve the audios taken from real-world places. (MFCC) are extracted for each audio, along with their differentials and accelerations to evaluate the machine learning k-Nearest Neighbor (KNN) Algorithm, the accuracy was reached 68.1 percent. Huh, et al. [9], suggested an augmentation adversarial training plan to train active speaker embedding's with self-supervision. The technique employs an Augmentation Classifier (AC) and Gradient Reversal (GRL) Layer to block the speaker embedding extractor from learning the channel data; the accuracy reached 91.35 percent. Nawas, et al. [10], presented speaker recognition model based on Reconstructed Phase Space (RPS) for features extraction, and choose Timit dataset. To be tested by random forest classifier, the result was 71 percent. Karthikeyan, et al. [10], proposed a voice recognition system where used Matthews correlation coefficient (MCC) for features extraction and the important point of their approach is to employ the hybrid AdaBoost (AB) classifiers and random forests wherein the first stage it employs (RF), and then strengthens it using (AB), the results were obtained 92 percent.

Received 31/5/2021; Accepted 14l8l2021

DOI: https://doi.org/10.33103/uot.ijccce.21.3.10

III. THE STRUCTURE OF THE PROPOSED METHOD

Based on (ML) algorithms, the proposed biometrics-based system would recognize speech. The Database description, Preprocessing, Feature extraction, k-fold cross validation stages and classification stages are all included in this section. *Fig*.4 depicts the proposed method design.

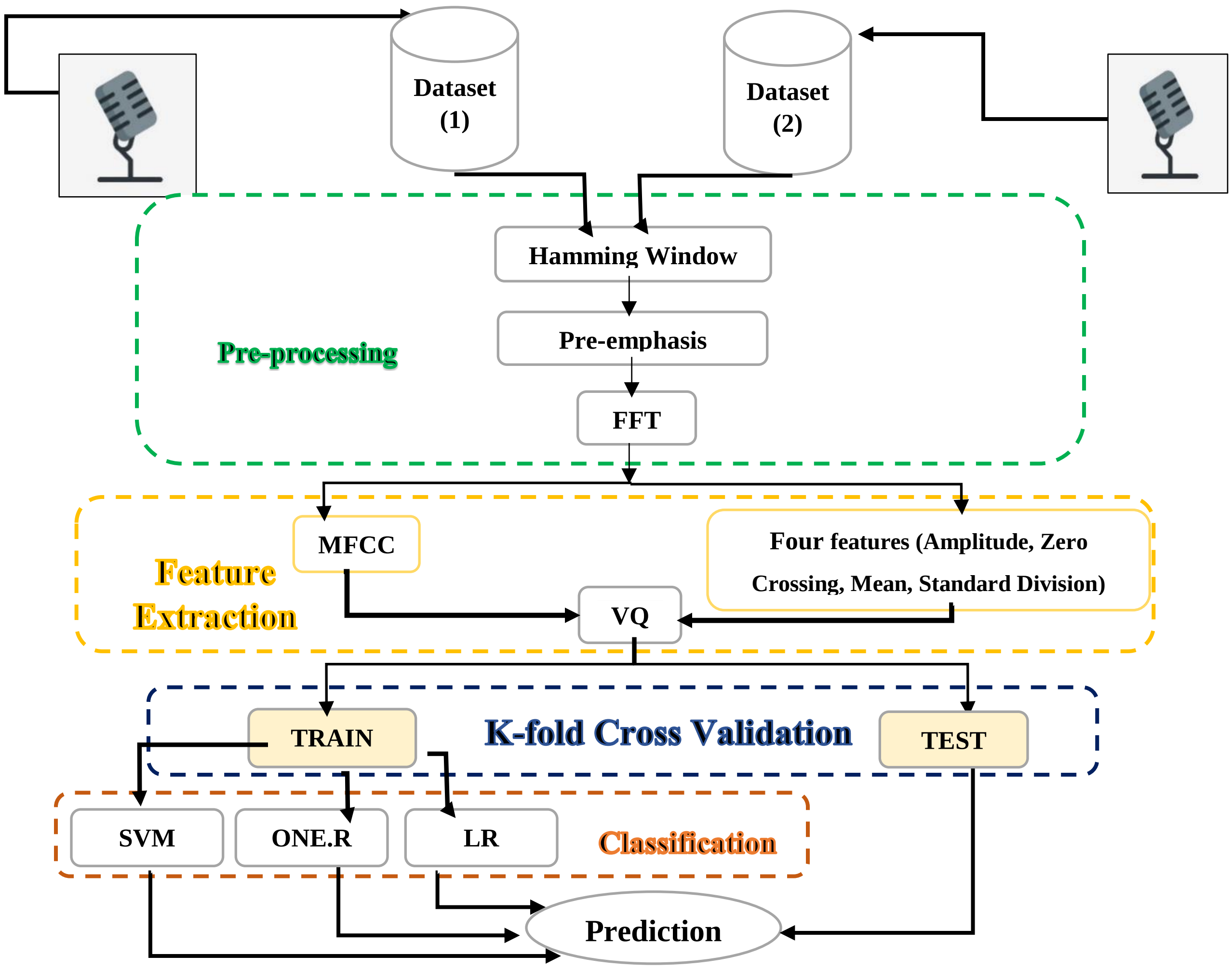

FIG. 4. THE PROPOSED SPEAKER RECOGNITION MODEL STRUCTERS

A. Database Description

Two voice datasets were selected. The details that include extension, capacity, noise intensity, dataset record place, number of persons, and number of samples are mentioned in Table (1) below.

Received 31/5/2021; Accepted 14l8l2021

DOI: https://doi.org/10.33103/uot.ijccce.21.3.10

TABLE 1. DATASETS OF SYSTEMS DETAILS

Dataset name	Dataset Recorde d Place	Noise intensity	Dataset Recorded Place	File format	File capacity	No. of persons	No. of instances	Obtained from
The speeches of the five leaders	Outside	medium	Outside	.Wav	16khz	5	7500	Kaggle web side
Speaker Authentica tion	Inside	low	Inside	.Wav	16khz	50	2226	Kaggle web side

B. Pre-processing

The principal benefit of preprocessing level is that it fits the data, producing identification more comfortable. It makes the data arranged and comfortable to deal with and deduce useful information in the following stages.

C. Extraction of Features using MFCC

The procedure of computing a set of the feature vectors which provides a compact representation of a specific speaking signal is known as feature extraction. (MFCC) is an idea for extracting features maintained by an acoustic signal. Based on person hearing that cannot sense frequencies of higher than 1 KHz, computations carried out by MFCC are conditional on the process of changing signals from analog to digital. MFCC offers Computations ranging from the length of the wave height, noise and other things so that the words that are well spoken by the user are received [12]. .All these steps will be clearly in *Fig*.5 [13]:

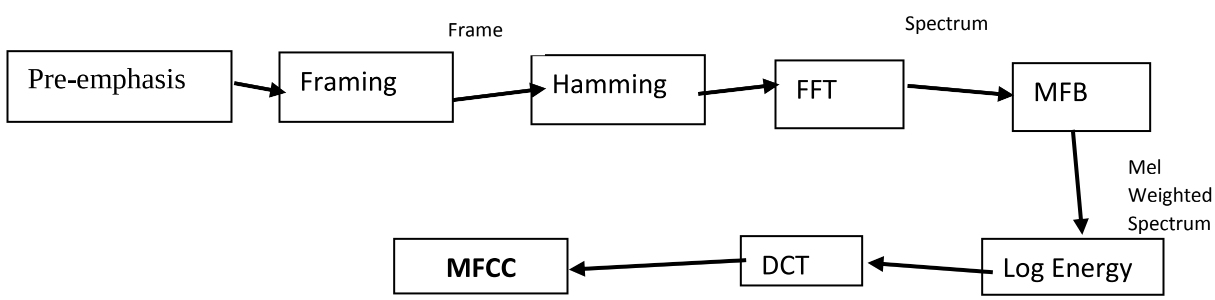

FIG. 5. FETURES EXTRACTION METHODS

1. **Pre-emphasis**. The most important stage of preprocessing which is an implicit in MFCC is the first step that increases the amount of energy in high frequencies by applying filters. The principal goal of the pre-emphasis filters is to adjust the speech signal frequency. Involvement of the pre-emphasis filter in the time domain is provided by:

$$y(n) = x(n) - 0.96\, x(n-1) \qquad (1)$$

where the value of x ranges from 0.9 to 1.

Received 31/5/2021; Accepted 14l8l2021

DOI: https://doi.org/10.33103/uot.ijccce.21.3.10

2. **Framing:** In these steps, voice instances are formed into short frames. Normally, the length of every frame is in the range of 20 to 40 m.sec. (25 msec is taken for the suggested framework). i.e., the voice signal is divided into N frames of M instances.

3. **Windowing:** High sound signals are surveyed in these steps by taking enough samples' parts. The principal goal of this step is to decrease the edge effect, to make it smoother and to improve the harmonics during taking the DFT on the signal. Normally, the Hamming window is used.

If the window w(n) is defined for $0 \leq n \leq M - 1$:

$$y(n) = x(n) * w(n) \tag{2}$$

The Hamming window, w(n) is achieved by:

$$w(n) = 0.54 - 0.46\cos(2\pi nM - 1)\big(0 \leq n \leq (M - 1)\big) \tag{3}$$

$$X(k) = \sum_{i=0}^{n-1} x(i)e\,\frac{-j2\pi ik}{N} \quad 0 \leq k \leq N - 1 \tag{4}$$

4. **Mel spectrum:** The FFT converted signal is squared to get the DFT power spectrum and it is transferred through the band-pass filters named Mel-filter bank to get mel spectrum.

$$\text{fmel} = 2595\,\text{log10}(1 + \frac{\text{f}}{700}) \tag{5}$$

$$\text{y(n)} = \sum_{\text{i}=0}^{\text{n}-1} [|\text{X(i)}|2\text{Wj (i)}]0 \leq \text{j} \leq \text{J} - 1 \tag{6}$$

Where j is the total number of mel weighing filters

5. Log: During this step, get the log of power at each of mel frequencies.
6. DCT: Log mel spectrum is transformed to the time domain in this step. Following transforming it into the time domain, MFCCs are obtained. This collection of coefficients are estimated acoustic vectors.

D. Four Features (FF)

Amplitude, Zero Crossing, Mean, and Standard Division are mathematical and statistical features that can be extracted from the data and help in inferring information that expresses the description of that data. The first feature Zero-Crossing (ZC) is the rate at which the signal's sign shifts during the frame. It can also be explained as the number of times the signal's value moves from positive to negative and back, divided by the frame's duration. It is given by the number of time the signal amplitude crosses the zero value. The Second feature is Sound's Amplitude (A) which is defined as the loudness or the amount of maximum displacement of vibrating particles of the medium from their mean position when the sound is performed. It is the distance between the top or trough and the mean position of the wave. The third feature, the Mean

Received 31/5/2021; Accepted 14l8l2021

DOI: https://doi.org/10.33103/uot.ijccce.21.3.10

(M), is the average used to, where add up all the numbers and then divide by the number of digits. The fourth feature Standard Deviation (SD) is a statistic that measures a dataset's distribution of an average. Through measuring each data point's deviation of an average, the standard deviation is measured as the square root of variation.

E. Apply Vector Quantization (VQ)

VQ is the ability of a speaker recognition method to evaluate probability allocations of the determined feature vectors. In addition, it is not conceivable to keep all single generated vector by training-mode; meanwhile such allocations are well-defined over a high-dimensional space. It seems easy to initiate that every single feature vector is quantized to one of smallest part of template vectors. VQ is the method of mapping vectors from large spaced vector to various regions in same space [14].

F. The Suggested Method Classifiers

In evaluating the talker's gross performance identification design, ML algorithms, preceded by the extraction of feature, are important. The goal is to identify audios and find out who is speaking in them, so this is a classification issue. As a consequence, the following machine learning algorithms for supervised classification will be applied.

1. SVM

SVM (Support Vector Machine) is a supervised machine learning (SML) algorithm-based classification model. This method affords excellent performance outcomes in classifying and regression, it suits the interest of researchers in the (ML) domain. SVM classified data into two classes, which implement the classification by separating the data, with a hyperplane, into two classes [15]. SVM is a very strong and resilient algorithm that can achieve linear or nonlinear outlier detection. SVM can especially fit all small- or medium-sized datasets. SVM algorithm is a representation of the samples as points in space, mapped so that the samples of the individual categories are divided by a clear gap that is as large as possible [16]. The idea that makes SVM more efficient in (ASR) applications than different methods depending on linear discriminants is its learning rule. The purpose of any classifier must minimize the number of misclassifications in any possible set of samples. This is called Risk Minimization (RM). The reason for using SVM is it has a unique solution and its convergence is confirmed (the solution is obtained by minimizing a convex function). The solution is that the maximum margin makes these machines robust and, in our opinion, it is very well suited for applications such as ASR in noisy environments, and when considering the minimization method, particularly the required kernel matrix, it can be handled with input vectors of high dimensionality, as long as it is capable of calculating their corresponding kernels. In practice, one can deal with vectors of thousands of dimensions [17].

2. ONE.R

ONE.R is a simplistic method. The ONE.R creates one rule for each attribute in the training data and the next chooses the rule with the least error rate as its one rule. The method is based on ranking all the attributes based on the error rate. To produce a rule for an attribute, the most frequent class for each attribute value must be determined. The most frequent class is simply the class that appears most often for that attribute value. A rule is simply a set of attribute values bound to their majority class. One-R selects the rule with the

Received 31/5/2021; Accepted 14l8l2021

DOI: https://doi.org/10.33103/uot.ijccce.21.3.10

lowest error rate. In the event that two or more rules have an equal error rate, the rule is chosen at random [18]. The reason for choosing it is the speed and simplicity of implementation.

3. Linear Regression (LR)

Linear Regression (LR) is a case model with a single independent variable. Linear regression defines the dependence of the variable, and it distinguishes the influence of independent variables from the interaction of dependent variables. (LR) requires the two variables on the x-axis and y-axis to be linearly correlated [19]. In other words, Linear Regression is a sample model with a single independent variable. Linear regression describes the dependency of the variable. $y = \beta 0 + \beta 1 x +$. Simple regression identifies the influence of independent variables from the interaction of dependent variables [20]. The reason for choosing the Linear Regression is that more adaptable and has wide applicability, a simpler model makes it easier to describe how the model works and how to interpret model findings, and Learning Regression Analysis will provide you with a better general understanding of statistical inference.

IV. THE SUGGESTED SYSTEM

FIG.6 explains the details and stages of the proposed system from the inputs through training and testing, after the initial processing, then extracting the features and using the three classifiers, ending with the evaluation of the results of the system.

Input -: Voice dataset(1) or Voice dataset(2)

Output: - Best classifier performance

Begin

1: Load dataset (1) or (2) **// Input //**

2: Removing noise using Hamming window

3: Make the spectral form of the speech signal frequency more smooth using Pre-emphasis

4: Signals converting into a frequency spectrum using (FFT)

5: Human hearing activity to detect frequencies using Mel-Frequency Cepstral Coefficients (MFCC)

Pre-processing Phase

6: Combine the four features with MFCC

7: Digital representation of signals using Vector Quantization (VQ)

Feature Extraction Phase

8: Shuffle the dataset randomly and divide it into (k=8) groups using Cross-Validation, a (k –1) sub-instances were applied for training

9: The remaining data single sub-instance will be applied as the validation data for testing

10: Classify instances based (SVM, LR, ONE.R) Classifiers

Data Split

11: Classifiers Evaluation

12: Best Classifier performance (Accuracy Measurements) **//Output//**

End

FIG. 6. ALGORITHM FOR THE SUGGESTED SPEAKER AUTHENTICATION METHOD

Received 31/5/2021; Accepted 14l8l2021

DOI: https://doi.org/10.33103/uot.ijccce.21.3.10

V. SYSTEM EVALUATION

Certain parameters are utilized to determine the behavior of a model when evaluating its performance. The size of the training data, the quality of the audio recordings, and, most crucially, the type of machine-learning algorithm utilized all have an impact on the results. The models' efficacy is evaluated using the following criteria [21]:

Accuracy (Acc): $$\text{Acc} = \frac{\text{PT+NT}}{\text{PT+NT+PF+NF}} \quad (7)$$

Where

- **PT** = correct positives: total of instances foretold positive which really positive
- **PF** = incorrect positives: total of instances foretold positive which really negative
- **NT** = correct negatives: total of instances foretold negative which really negative
- **NF** = incorrect negatives: total of instances foretold negative which really positive

- *Precision (Pr):* $\boldsymbol{Pr = \frac{PT}{PT+PF}}$ (**8**)
- *Recall (R.cll):* $\boldsymbol{R.cll = \frac{PT}{PT+NF}}$ (**9**)
- *F- measure (F1):* $\boldsymbol{F_1 = 2 * \frac{precision*recall}{precision+recall}}$ (**10**)
- *Error rate (Er.Ra):* $\boldsymbol{Er.Ra = \frac{PF+NF}{PT+NT+PF+NF}}$ (**11**)
- *Specificity (TNR) :* $\boldsymbol{TNR = \frac{NT}{NT+PF}\ 100\%}$ (**12**)

VI. THE EMPIRICAL RESULTS

This research examined the two datasets with (SVM, LR, and ONE.R). The results of evaluating the voice dataset (1) and dataset (2) as input are showed in Table (1) and Table (2). One can notice that the (SVM) provides the best accuracy Measurements, while (ONE.R) classifier gives the minimum accuracy Measurements.

TABLE 2. RESULTS OF ML CLASSIFIERS

Measurements	**ONE.R**		**LR**		**SVM**	
	Dataset(1)	Dataset(2)	Dataset(1)	Dataset(2)	Dataset(1)	Dataset(2)
Overall samples	7500	2226	7500	2226	7500	2226
Overall positive	3879	523	6961	1940	7321	2117
Overall negative	362	1703	539	286	179	109
Acc	0.517	0.235	0.9283	0.8715	0.977	0.951
Pr	0.51	0.19	0.9282	0.8739	0. 9763	0.95

Received 31/5/2021; Accepted 14l8l2021

DOI: https://doi.org/10.33103/uot.ijccce.21.3.10

R.cll	0.51	0.234	0.928	0.87	0.976	0.95
F1	0.51	0.195	0.9281	0.871	0.976	0.948
Er.Ra	0.482	0.765	0.0718	0.1284	0.0238	0.049
TNR	0.879	0.973	0.982	0.9963	0.994	0.998
Achievement Time	1.45 msec	1.38 sec	1.22 msec	1.15 sec	2.36 msec	1.58 msec

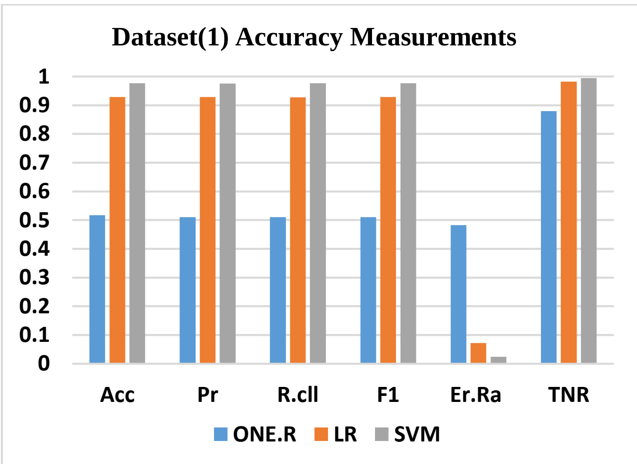

FIG. 7. DATASET (1) ACCURACY MEASUREMENTS

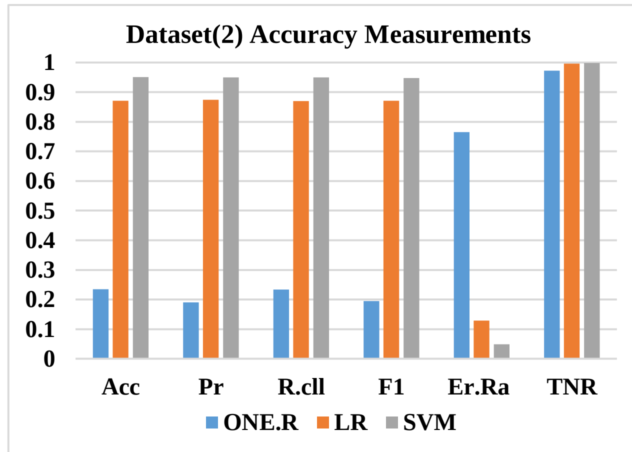

FIG. 8. DATASET(2) ACCURACY MEASUREMENTS

It is clear to us from *Figs*. (7 and 8) that the SVM algorithm has achieved the highest accuracy among the remaining two algorithms, reaching 97.7 percent for dataset (1), and presented 95.1percent for dataset (2). As for measuring the error, the SVM algorithm is also outperformed by achieving the least error rate, it achieved 2.3 percent for dataset (1), and presented 4.9 percent for dataset (2).Not to lose sight of the time factor because of its great importance in this field as it will appear below.

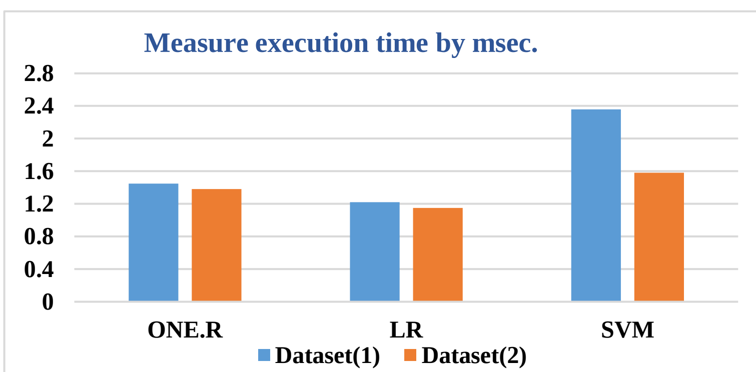

FIG. 9. MEASURE EXECUTION TIME

To measure the time, the task must be completed in seconds. It seems to us that the advantage in terms of implementation time was in favour of an (LR) algorithm, as seen above from Table (2) and *FIG*. 9.

Received 31/5/2021; Accepted 14l8l2021

DOI: https://doi.org/10.33103/uot.ijccce.21.3.10

VII. THE COMPARISON OF RESULTS

In Table (3), the competition is between the best two achievements of the suggested method algorithms with the relevant studies. In order for the comparison to be fair and objective, the difference in the data set must be defined for calculating the accuracy rate. The average of accuracy Rate of the suggested method for the two classifiers (SVM, LR) is determined by Eq. (13):

$$Average\ of\ accuracy = \frac{\boldsymbol{Dataset(1)\ accuracy + Dataset(2)\ accuracy}}{\mathbf{2}} \qquad (\mathbf{13})$$

TABLE 3. RESULTS COMPARISON WITH RELEVANT STUDIES

Title	Year	Features extraction	Technique used	Accuracy
Ref. [4]	2017	IHC & MFCC	GMM & I.V	94.33%
Ref.[5]	2018	MFCC	SVM	95%
Ref.[6]	2019	LPC & MFCC	SVM	80.60%
Ref.[7]	2020	MFCC	RF	84.40%
Ref.[8]	2020	MFCC	KNN	68.10%
Ref.[9]	2020	GRL	(AC)	91.35%
Ref.[10]	2021	RPS	RF	71%
Ref.[11]	2021	MCC	AB+RF	92%
SVM PROPOSED	**2021**	**MFCC & FF**	**SVM**	**96.36%**
LR PROPOSED	2021	MFCC & FF	LR	89.99%

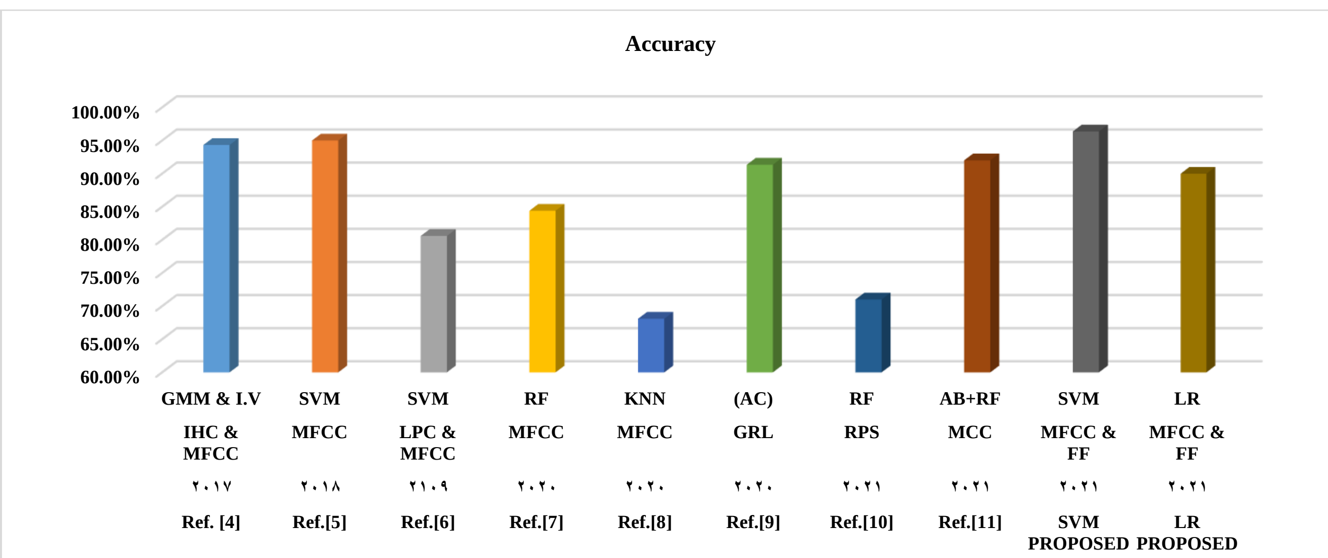

FIG. 10. RESULTS COMPARISON WITH RELATED WORK.

Received 31/5/2021; Accepted 14/8/2021

DOI: https://doi.org/10.33103/uot.ijccce.21.3.10

The results indicate that the method proved to be advantageous in applying (the Proposed SVM) in expressions of the accuracy rate of speak identification over all literature review systems which analyzed in Table (3) and observed in *Fig*. 10.

VIII. CONCLUSIONS

The three key elements of this research are, voice preprocessing, feature extraction and ML algorithms. Since the audio samples were not recorded in enclosed spaces, preprocessing of audio was an important part of the research. The two most important aspects of pre-processing were to reduce ambient noise and to emphasize human voice. The reason that the analysis would be not sufficient is when only the MFCC gets used, so the four features (Amplitude, Zero Crossing, Mean, and Standard Division) were extracted and added. Ere merging the product matrices of the other instruments, (VQ) was applied to transform the binary matrix produced by MFCC and (FF) to a one-row matrix. There were two datasets and three ML algorithms used in this research when testing two datasets of different quality and format, to increase reliability and dependability. The K-fold cross-validation was chosen for training and evaluation, and the value of K = 8, the reason was when increasing K more than 8, there would be no increase in the accuracy of the results and it would cause a delay in obtaining the results, leading us to loss time and cost. The measures of the results accuracy were discussed from several aspects, in addition to measuring the time factor by finding the result of implementation time for each classifier. The process of (ML) Techniques of model method improved accuracy, with (SVM) achieving a higher average of accuracy which was 96.36 %. In terms of performance in relation to the time taken at the speed of execution, the algorithm (LR) was the best with an average execution time of 1.27 seconds.

REFERENCES

[1] A. Tahseen Ali, H. Abdullah and M. Fadhil, "Intelligent system for imposter detection: Asurvey", Materials Today: Proceedings, 2021. Available: 10.1016/j.matpr.2021.04.074.

[2] A. Tahseen Ali, H. Abdullah and M. Fadhil, "Voice recognition system using machine learning techniques", Materials Today: Proceedings, 2021. Available: 10.1016/j.matpr.2021.04.075.

[3]A. M. Ahmed and A. K. Hassan, "Speaker Recognition Systems in the Last Decade – A Survey", Engineering and Technology Journal, vol. 39, no. 1, pp. 30-40, 2021. Available: 10.30684/etj.v39i1b.1589.

[4]S. Paulose, D. Mathew and A. Thomas, "Performance Evaluation of Different Modeling Methods and Classifiers with MFCC and IHC Features for Speaker Recognition", Procedia Computer Science, vol. 115, pp. 55-62, 2017. Available: 10.1016/j.procs.2017.09.076.

[5]R. Thiruvengatanadhan, "Speech Recognition using SVM", International Research Journal of Engineering and Technology (IRJET), vol. 5, no. 9, 2018.

[6]N. Chauhan, T. Isshiki and D. Li, "Speaker Recognition Using LPC, MFCC, ZCR Features with ANN and SVM Classifier for Large Input Database", International Conference on Computer and Communication Systems, no. 4, pp. 130-133, 2019.

[7]M. Subba Rao, G. Bhagya Lakshmi, P. Gowri and K. Bharath Chowdary, "RANDOM FOREST BASED AUTOMATIC SPEAKER RECOGNITION SYSTEM", The International journal of analytical and experimental modal analysis, vol., no., pp. 526-535, 2020.

Received 31/5/2021; Accepted 14l8l2021

DOI: https://doi.org/10.33103/uot.ijccce.21.3.10

[8]M. Subba Rao, Y. Haritha and A. Harsha Vardhan Reddy, "AUTOMATIC SPEAKER RECOGNITION SYSTEM USING K-NEASREST NEIGHBOR ALGORITHM", Journal of Engineering Sciences, vol. 11, no. 4, pp. 576-583, 2020.

[9]J. Huh, H. Soo Heo, J. Kang, S. Watanabe and J. Son Chung, "Augmentation adversarial training for self-supervised speaker recognition", 34th Conference on Neural Information Processing Systems, 2020.

[10]K. Khadar Nawas, M. Kumar Barik and A. Nayeemulla Khan, "Speaker Recognition using Random Forest", ITM Web of Conferences, vol. 37, p. 01022, 2021. Available: 10.1051/itmconf/20213701022.

[11]V. Karthikeyan and S. Suja Priyadharsini, "A strong hybrid AdaBoost classification algorithm for speaker recognition", Sādhanā, vol. 46, no. 3, 2021. Available: 10.1007/s12046-021-01649-6..

[12]A. Haq, M. Nasrun, C. Setianingsih and M. Murti, "Speech Recognition Implementation Using MFCC and DTW Algorithm for Home Automation", Proceeding of the Electrical Engineering Computer Science and Informatics, vol. 7, no. 0, 2020. Available: 10.11591/eecsi.v7.2041.

[13]E. Fathima Bareeda, B. Shajee Mohan and K. Ahammed Muneer, "Lie Detection using Speech Processing Techniques", Journal of Physics: Conference Series, vol. 1921, no. 1, p. 012028, 2021. Available: 10.1088/1742-6596/1921/1/012028.

[14]F. n ur Rehman, C. Kumar, S. Kumar, A. Mehmood and U. Zafar, "VQ BASED COMPARATIVE ANALYSIS OF MFCC AND BFCC SPEAKER RECOGNITION SYSTEM", nternational Conference on Information and Communication Technologies (ICICT), pp. 28-32, 2017. [Accessed 23 February 2022].

[15]K. Jalil Taher and H. D.Majeed, "Recognition of Handwritten English Numerals Based on Combining Structural and Statistical Features", Iraqi Journal of Computer, Communication, Control and System Engineering, vol. 21, no. 1, pp. 73-83, 2021. Available: 10.33103/uot.ijccce.21.1.7.

[16]Z. Ali Mohammed, M. Najm Abdullah and I. Hussain Al-Hussaini, "Predicting Incident Duration Based on Machine Learning Methods", Iraqi Journal of Computer, Communication, Control and System Engineering, vol. 21, no. 1, pp. 1-15, 2021. Available: 10.33103/uot.ijccce.21.1.1.

[17]R. Solera-Ure˜na, J. Padrell-Sendra, D. Mart´ın-Iglesias, A. Gallardo-Antol´ın, C. Pel´aez-Moreno and F. D´ıaz-de-Mar´ıa, "SVMs for Automatic Speech Recognition: A Survey", Progress in nonlinear speech processing, pp. 190-216, 2007.

[18]Mohammed, Yasir, A. Ali Mohammed and W, "Comparison of Classification Algorithm in the Educational Data Mining (Bayes Network, One Rule, Random Forest.", University Of Science And Technology, , Master's thesis, 2018.

[19]D. Maulud and A. Abdulazeez, "A Review on Linear Regression Comprehensive in Machine Learning", Journal of Applied Science and Technology Trends, vol. 1, no. 4, pp. 140-147, 2020. Available: 10.38094/jastt1457.

[20]P. Bartlett, P. Long, G. Lugosi and A. Tsigler, "Benign overfitting in linear regression", Proceedings of the National Academy of Sciences, vol. 117, no. 48, pp. 30063-30070, 2020. Available: 10.1073/pnas.1907378117.

[21] Tahseen Ali, A., Abdullah, H., & Fadhil, M. (2021). Finger veins recognition using machine learning techniques. Materials Today: Proceedings. doi: 10.1016/j.matpr.2021.04.076.

Received 31/5/2021; Accepted 14l8l2021