Turkish Journal of Forecasting vol. 07 no. 2 (2023) pp. 17-22

Content list available at JournalPark

Turkish Journal of Forecasting

Journal Homepage: tjforecasting.com

Hybrid AI-based Voice Authentication

B. Bora[1], A.E. Emanet[1], E. Elmaci[2], D. Kandaz[1,*], M.K. Ucar[1]

[1]*Sakarya University, Faculty of Engineering, Department of Electrical Electronics Engineering, Sakarya Campus, Sakarya, Turkey*

A R T I C L E I N F O

Article history:

Received	04	March	2023
Revision	07	September	2023
Accepted	17	December	2023
Available online	18	December	2023

Keywords:

Biometric Authentication
Correlation Score
Voice Biometric
Hybrid AI

R E S E A R C H A R T I C L E

A B S T R A C T

Biometric authentication systems reveal individuals' physical or behavioral uniqueness and identify them by comparing them with existing records. Today, many biometric recognition systems, such as fingerprint reading, palm reading, and face reading, are being studied and used. The human voice is also among the techniques used for this purpose. Due to this feature, the human voice performs secure transactions and authentication in various fields. Based on these voice features, we used a dataset of 66,569 voice recordings. The voice recordings were revised to include six sentences of at least six words each from 24 different people to get the maximum benefit from the dataset. The voices in the reduced dataset were labeled as sentences belonging to the same person and sentences belonging to different people and converted into matrix form. A biometric recognition study resulted in a correlation score of 0.88. As a result of these processes, the feasibility of a voice biometric recognition system with artificial intelligence has been demonstrated.

1. Introduction

Identity verification is a security function used within an information system to control a logical server or device based on individuals' biological or behavioral characteristics [1]–[3]. These systems can be categorized into three groups: knowledge based, physiological, and behavioural biometric based. Knowledge-based biometric identification requires an individual to enter identification information to verify their identity. Physiological biometric-based identity verification creates an artificial intelligence-based identification system utilizing an individual's physiological features. These features encompass biometric data like fingerprints, hand geometry, palm veins, height, retina, iris color, and size. Behavioral biometrics, on the other hand, creates a behavioural model by analysing a person's unique habits and movements. Behavioural traits include biometric data like voice and signature. In this study, the focus is on voice identity verification, which is one of the behavioural biometric recognition methods [1], [2]. Voice identity verification encompasses the process of identifying sounds produced by the vocal cords. Each sound has distinctive identifying features, which are differentiated by the speaker's anatomy and behavioural speech patterns. This method, containing numerous distinctive characteristics, can authenticate a speaker's biometric identity by matching their voice. This technique, requiring no memory, being non-removable, and having an easy application, is preferred. Among biometric technologies for identity verification, voice biometrics are more convenient and reliable for users. Additionally, users can employ this method without the need to carry or remember something, and

* Corresponding author.
E-mail addresses: deryakandaz@sakarya.edu.tr (Derya Kandaz)

https://doi.org/10.34110/forecasting.1260073

they don't worry about risks like identity card theft or password hacking. However, alongside these benefits, the various systems used in voice identity verification introduce different algorithmic structures [4]. In 1987, Worlds of Wonder Company developed speech recognition application through a baby crib for the first time. Texas Instruments developed an algorithm using digital signal processing to identify eight different sentences [5]. Moreover, work has been carried out for many years on voice recognition systems with the aim of assisting disabled individuals [6]. In this context, a high-security identity verification system has been developed using a dataset consisting of 66,569 human voice recordings.

A review of the existing literature studies indicates that the voice signals used as datasets have short-duration values [7]. To overcome this limitation, longer speech signals were selected from the dataset, resulting in a reduced dataset. The speech recordings in the reduced dataset consist of six sentences from 24 different individuals, each containing a minimum of six words. Additionally, previous studies have shown that the processing stages of speech signals are multiple and complex [7]. This study, from this perspective, is simpler and more comprehensible. As opposed to using only 6 to 7 voice parameters, as compared to studies in the literature, this work employs 25 statistical voice features [8], [9]. Finally, the speech samples in the dataset were labelled as sentences from the same person or from different individuals, and they were transformed into matrix form. The proposed machine learning algorithms were applied to the dataset, thereby verifying biometric identity through voice.

2. Material and Method

The application of the study is summarized in Figure 1. According to the flow diagram, the collected data was pre-processed, and labelled matrices were created. Statistical features were extracted from the records and classified selectively. Feature extraction was performed. After all these steps, feature selection was performed on the available data. Various machine learning algorithms such as Support Vector Machine (SVM), decision tree (DT), and ensemble were used in these stages of the study.

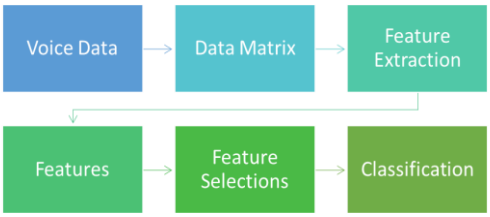

Figure 1. Application flowchart

2.1. Data Acquisition

The data was obtained and used from Mozilla, an open source sharing platform [10] There are 144 voice records from 24 people in the dataset, six for each person. The dataset is 76 hours long and contains 65,639 voice recordings of 1,299 different people. Primarily, these records were combined, and the data matrix was created.

2.2. Data Pre-Processing

The voice recordings were labelled with each other by cross-validation. Two different voice recordings of the same person are labelled 1, while voice recordings of different people are labelled 0. The data labelling processes are indicated in the Table 1 and Table 2. Due to the combination, the sizes of the matrices labelled 1 and 0 are different. For this reason, data balancing was applied, and the labels were balanced. A total of 10,296 records were obtained from both labels. 360 of these records belong to the same person, and 9,936 belong to different persons. Data balancing was performed using systematic sampling theory to avoid imbalance. Systematic sampling theory has been used to ensure that the balancing process is as accurate as possible.

2.3. Feature Extraction

Feature extraction aims to reduce the number of features in a dataset by creating new features from the existing ones. After data preprocessing and balancing, 25 statistical features of the records were extracted. These features are descriptive statistical parameters that are frequently used in statistics. The mathematical parameters related to the features are presented in Table 3.

Table 1. Label 1 sample matrix

Number of Comparisons	First Person		Second Person		Label
	Person Number	Person Feature Number	Person Number	Person Feature Number	
1	1	1	1	2	1
2	1	1	1	3	1
3	1	1	1	4	1
4	1	1	1	5	1
5	1	1	1	6	1
6	1	2	1	3	1
7	1	2	1	4	1
8	1	2	1	5	1
9	1	2	1	6	1
10	1	3	1	4	1
11	1	3	1	5	1
12	1	3	1	6	1
13	1	4	1	5	1
14	…	…	…	…	1
360	24	5	24	6	1

Table 2. Label 0 sample matrix

Number of Comparisons	First Person		Second Person		Label
	Person Number	Person Feature Number	Person Number	Person Feature Number	
1	1	1	2	1	0
2	1	1	2	2	0
3	1	1	2	3	0
4	1	1	2	4	0
5	1	1	2	5	0
6	1	1	2	6	0
7	1	1	3	1	0
8	1	1	3	2	0
9	1	1	3	3	0
10	1	1	3	4	0
11	1	1	3	5	0
12	1	1	3	6	0
13	1	1	4	1	0
14	…	…	…	…	0
9936	23	6	24	6	0

2.4. Feature Selection

Feature selection is the process of isolating the most consistent, non-redundant, and relevant features to use in model construction. The main goal of feature selection is to improve the performance of a predictive model and reduce the computational cost of modelling. Spearman correlation coefficients were used for selected features. The relationship between familiar signal comparison tags and unfamiliar signal comparison tags was determined by Spearman correlation coefficients (r_s). The data set was divided into 11 groups using certain percentage values. These values were ranked according to performance. The features with the best correlation value between them were used.

3. Results

The study aims to realize a biometric recognition system with voice signals. For this purpose, machine learning models with high performance values were evaluated. The machine learning outputs were compared with the real database outputs. The model outputs 1 if two voice recordings belong to the same person and 0 if they belong to different people. After extracting 25 features of the voice signal, 11 data sets with different percentages were

produced. These datasets were evaluated with the classification method in machine learning models and the highest success percentage was achieved by the SVM algorithm with 95.14% in the Table 4. This success percentage can be considered as a very good result when compared to other success percentages found in the literature.

Table 3. Mathematical equations for statistical properties

Number	Features	Equation
1	Kurtosis	$x_{kur} = \frac{\sum_{i=1}^{n}(x(i)-\bar{x})^4}{(n-1)S^4}$
2	Skewness	$x_{ske} = \frac{\sum_{i=1}^{n}(x_i-\bar{x})^3}{(n-1)S^3}$
3	* * Interquartile Range	$IQR = iqr(x)$
4	Coefficient of Variance	$DK = (S/\bar{x})100$
5	Geometric Mean	$G = \sqrt[n]{x_1+\cdots+x_n}$
6	Harmonic Mean	$H = n/(\frac{1}{x_1}+\cdots+\frac{1}{x_n})$
7	Activity-Hjort Parameter	$A = S^2$
8	Mobility-Hjort Parameter	$M = S_1^2/S^2$
9	Complexity-Hjort Parameter	$C = \sqrt{(S_2^2/S_1^2)^2-(S_1^2/S^2)^2}$
10	* Maximum	$x_{max} = max(x_i)$
11	Median	$\tilde{x} = \begin{cases} x_{\frac{n+1}{2}} & : x \text{ odd} \\ \frac{1}{2}(x_{\frac{n}{2}}+x_{\frac{n}{2}+1}) & : x \text{ even} \end{cases}$
12	* Mean Absolute Deviation	$MAD = mad(x)$
13	* Minimum	$x_{min} = min(x_i)$
14	* Central Moments	$CM = moment(x,10)$
15	Average	$\bar{x} = \frac{1}{n}\sum_{i=1}^{n} = \frac{1}{n}(x_1+\cdots+x_n)$
16	Average Curve Length	$CL = \frac{1}{n}\sum_{i=2}^{n}\lvert x_i - x_{i-1}\rvert$
17	Average Energy	$E = \frac{1}{n}\sum_{i=1}^{n}x_i^2$
18	Root Mean Square Value	$X_{rms} = \sqrt{\frac{1}{n}\sum_{i=1}^{n}\lvert x_i\rvert^2}$
19	Standard Error	$S_{\bar{x}} = S/\sqrt{n}$
20	Standard Deviation	$S = \sqrt{\frac{1}{n}\sum_{i=1}^{n}(x_i-\bar{x})}$
21	Shape Factor	$SF = X_{rms}/(\frac{1}{n}\sum_{i=1}^{n}\sqrt{\lvert x_i\rvert})$
22	*Singular value decomposition	$SVD = svd(x)$
23	* 25% trimmed average	$T25 = trimmean(x,25)$
24	* 50% trimmed average	$T50 = trimmean(x,50)$
25	Average Teager energy	$TE = \frac{1}{n}\sum_{i=3}^{n}(x_{i-1}^2 - x_i x_{i-2})$

Table 4. Performance evaluation results

Information				Performance Evaluation Criteria					
L	**NF**	**FP**	**Model**	**ACC**	**SENS**	**SPEC**	**F1**	**KAPPA**	**AUC**
1	1	5	DT	65,10	0,39	0,90	0,54	0,30	0,65
			SVM	63,11	0,27	0,83	0,42	0,23	0,58
			EBT	62,25	0,24	0,81	0,41	0,20	0,53
2	3	10	DT	70,65	0,55	0,85	0,67	0,41	0,70
			SVM	65,65	0,47	0,80	0,61	0,36	0,60
			EBT	63,54	0,44	0,76	0,59	0,32	0,58
3	4	15	DT	76,38	0,64	0,87	0,74	0,52	0,76
			SVM	72,93	0,54	0,77	0,65	0,43	0,64
			EBT	74,41	0,60	0,81	0,72	0,47	0,72
4	5	20	DT	76,73	0,63	0,89	0,74	0,53	0,76
			SVM	71,73	0,56	0,79	0,63	0,47	0,71
			EBT	69,73	0,51	0,82	0,67	0,50	0,72
5	6	25	DT	80,90	0,72	0,89	0,79	0,61	0,80
			SVM	71,80	0,65	0,82	0,73	0,56	0,74
			EBT	70,77	0,63	0,80	0,69	0,55	0,72
6	8	30	DT	80,38	0,71	0,88	0,79	0,60	0,80
			SVM	75,69	0,65	0,74	0,73	0,53	0,77
			EBT	74,44	0,61	0,70	0,71	0,52	0,74
7	9	35	DT	80,20	0,71	0,88	0,79	0,60	0,80
			SVM	74,48	0,65	0,81	0,74	0,50	0,73
			EBT	76,43	0,68	0,83	0,76	0,54	0,76
8	10	40	DT	81,77	0,74	0,89	0,81	0,63	0,81
			SVM	77,49	0,71	0,84	0,75	0,60	0,76
			EBT	72,51	0,65	0,80	0,71	0,56	0,73
9	11	45	DT	83,15	0,75	0,90	0,82	0,66	0,83
			SVM	79,14	0,71	0,83	0,77	0,61	0,81
			EBT	76,41	0,72	0,80	0,73	0,59	0,77
10	13	50	DT	85,59	0,79	0,91	0,85	0,71	0,85
			SVM	82,93	0,75	0,88	0,82	0,67	0,81
			EBT	81,77	0,74	0,86	0,81	0,65	0,81
11	25	100	DT	95,13	0,91	0,98	0,95	0,90	0,95
			SVM	90,16	0,89	0,93	0,91	0,85	0,88
			EBT	91,11	0,88	0,91	0,89	0,87	0,91

L: Level, NF: Number of Features, FP: Feature Percentage

DT: Decision Tree, SVM: Support Vector Machine, EBT: Ensemble Bagged Tree

4. Discussion

Although artificial intelligence-based systems based on biometric voice recognition in the literature have high accuracy, hybrid artificial intelligence-based voice authentication is new. It is applicable where control and boundary are not a problem. Some of these methods rely heavily on mathematical expressions and are time-consuming to run on the system [11], [12]. In addition, most of them have a similar mathematical basis and benefit from the morphological features of the voice to extract features [13], [14].

The datasets used in the studies are limited, and a comprehensive evaluation is impossible [12]. This study, unlike the others, made it possible to distinguish the voice from the extensive data set with high accuracy. As working constraints, situations such as time segmentation or not being able to examine in the time domain, not benefiting from signal properties, or trying to eliminate the problems predicted by another identification system may be encountered. However, from different perspectives, the study attempted to achieve the desired goals by using the classification model and by making statistical feature extraction. For this reason, it has taken its place in the literature by performing high-accuracy identification with voice authentication.

5. Conclusion

In this paper, studies on voice biometric recognition systems have been conducted. The results show that this method can be used in biometric recognition systems. With an accuracy rate of 95.14%, it has been shown that it can be used in high authentication methods that should not be exceeded. This study provides a high accuracy solution for machine learning based voice recognition. By utilizing this work, more advanced systems can be developed in the future. New models can be created by using larger data sets and longer speech signals. It is a system that can be improved by experimenting under different conditions.

References

[1] M. Nizam Kamarudin, H. Nizam Mohd Shah, M. Zamzuri Ab Rashid, M. Fairus Abdollah, C. Kok Lin, and Z. Kamis, “Biometric Voice Recognition in Security System,” 2014, Accessed: Aug. 28, 2023.

[2] K. Fatima, S. Nawaz, and S. Mehrban, “Biometric Authentication in Health Care Sector: A Survey,” 3rd International Conference on Innovative Computing, ICIC 2019, Nov. 2019, doi: 10.1109/ICIC48496.2019.8966699.C. Berghoff, M. Neu, and A. von Twickel, “The Interplay of AI and Biometrics: Challenges and Opportunities,” Computer (Long Beach Calif), vol. 54, no. 09, pp. 80–85, Sep. 2021, doi: 10.1109/MC.2021.3084656.

[3] C. Berghoff, M. Neu, and A. von Twickel, “Vulnerabilities of Connectionist AI Applications: Evaluation and Defense,” Front Big Data, vol. 3, p. 544373, Jul. 2020, doi: 10.3389/FDATA.2020.00023/BIBTEXA. Boles and P. Rad, “Voice Biometrics: Deep Learning-based Voiceprint Authentication System,” 2017, doi: 10.1109/SYSOSE.2017.7994971.

[4] S. Albalawi, L. Alshahrani, N. Albalawi, R. Kilabi, and A. Alhakamy, “A Comprehensive Overview on Biometric Authentication Systems using Artificial Intelligence Techniques,” International Journal of Advanced Computer Science and Applications, vol. 13, no. 4, pp. 782– 791, 2022, doi: 10.14569/IJACSA.2022.0130491.

[5] J. Noyes and C. Frankish, “Speech recognition technology for individuals with disabilities,” Augmentative and Alternative Communication, vol. 8, no. 4, pp. 297–303, 1992.

[6] F. Alcantud, I. Dolz, C. Gaya, and M. Martín, “The voice recognition system as a way of accessing the computer for people with physical standards as usual,” Technol Disabil, vol. 18, no. 3, pp. 89–97, 2006, doi: 10.3233/TAD-2006- 18301.

[7] A. Boles and P. Rad, “Voice biometrics: Deep learning-based voiceprint authentication system,” 2017 12th System of Systems Engineering Conference, SoSE 2017, Jul. 2017, doi: 10.1109/SYSOSE.2017.7994971.

[8] H. H. Zhu, Q. H. He, H. Tang, and W. H. Cao, “Voiceprint-biometric template design and authentication based on cloud computing security,” 2011 International Conference on Cloud and Service Computing, pp. 302–308, 2011, doi: 10.1109/CSC.2011.6138538.

[9] S. B. Sadkhan, B. K. Al-Shukur, and A. K. Mattar, “Biometric voice authentication autoevaluation system,” 2017 Annual Conference on New Trends in Information and Communications Technology Applications, NTICT 2017, pp. 174– 179, Jul. 2017, doi: 10.1109/NTICT.2017.7976100.

[10] “Common Voice.” https://commonvoice.mozilla.org/en/datasets (accessed Dec. 20, 2022).

[11] H. Shahid, S. Aziz, A. Aymin, M. U. Khan, and A. N. Remete, “A Survey on AI-based ECG, PPG, and PCG Signals Based Biometric Authentication System; A Survey on AI-based ECG, PPG, and PCG Signals Based Biometric Authentication System,” 2021, doi: 10.1109/ICECube53880.2021.9628307.

[12] C. Berghoff, M. Neu, and A. von Twickel, “Vulnerabilities of Connectionist AI Applications: Evaluation and Defense,” Front Big Data, vol. 3, p. 23, Jul. 2020, doi: 10.3389/FDATA.2020.00023/BIBTEX.

[13] S. B. Sadkhan, B. K. Al-Shukur, and A. K. Mattar, “Biometric voice authentication autoevaluation system,” 2017 Annual Conference on New Trends in Information and Communications Technology Applications, NTICT 2017, pp. 174– 179, Jul. 2017, doi: 10.1109/NTICT.2017.7976100.

[14] J. Gałka, M. Masior and M. Salasa, "Voice authentication embedded solution for secured access control," in IEEE Transactions on Consumer Electronics, vol. 60, no. 4, pp. 653-661, Nov. 2014, doi: 10.1109/TCE.2014.7027339.