Study of Machine Learning Methods for Voice Biometric Identification in Data Protection Systems

Serhii Semenov, Viacheslav Davydov and Denys Grynov

December 16, 2024

The 14th IEEE International Conference on Dependable Systems, Services and Technologies, DESSERT'2024
11-13 October, 2023, Athens, Greece

Study of Machine Learning Methods for Voice Biometric Identification in Data Protection Systems

Serhii Semenov
Institute of Security and Computer Science,
University of the National Education Commission
Krakow, Poland
serhii.semenov@uken.krakow.pl

Viacheslav Davydov
Science Entrepreneurship Technology University
Kyiv, Ukraine
vyacheslav.v.davydov@gmail.com

Denys Grynov
National Technical University«Kharkiv Polytechnic Institute»
Kharkiv, Ukraine
dgrynov@gmail.com

***Abstract* — This research investigates the effectiveness of machine learning methods in biometric voice identification, focusing on their application in data protection systems. Biometric voice identification has emerged as a promising approach due to its ability to analyze unique voice characteristics, enhancing authentication reliability and resistance against spoofing attacks. However, factors such as emotional state and physical health can impact voice consistency, posing challenges to accurate identification. The study involves a comparative analysis of various classifiers, including Gaussian Naive Bayes, support vector machine, and k-nearest neighbors, utilizing a prepared voice biometric dataset. The findings reveal that the Gaussian Naive Bayes classifier outperforms the others in terms of accuracy, achieving a mean accuracy of approximately 88.4%. Despite this, the research acknowledges the necessity for further refinement in preprocessing techniques and dataset enhancements to address existing identification accuracy limitations. Overall, this work emphasizes the potential of machine learning to bolster voice biometric systems while identifying areas for ongoing development..**

Keywords — Biometric voice identification, machine learning, Gaussian naive Bayes classifier, data protection, identification accuracy, authentication systems.

I. Introduction

The biometric identification and authentication technologies have recently become increasingly popular in data protection systems. One of the biometrics most perspective areas is biometric voice identification, which is based on the voice data unique characteristics.

The voice identification systems main advantages in the information security area include the following:

– several voice characteristics control, which provides more reliable recognition;

– the recognition of playback through miniature loudspeakers, which complicates the identification from the recorded speech;

– the possibility of the authentication reliability enhancement through the voice identification and speech recognition methods combination.

However, it should be noted that voice recognition has its own limitations. A person's voice changes over time and under the influence of various factors, such as physical and emotional state. For example, being drunk or chewing gum would make it difficult to recognize the voice. Disease conditions such as laryngitis or influenza can also distort the voice and create barriers for the identification.

These factors negatively affect the performed operations quality. Regarding this, the artificial intelligence technologies and the advanced machine learning methods usage becomes a relevant issue for a person high-quality identification and authentication.

The purpose of the research is the comparative analysis and the assessment of the machine learning methods for biometric voice identification in order to identify the most accurate and effective approach.

Research objectives:

1. Review the literature on biometric voice identification methods and existing machine learning models used in this area.
2. Collect and prepare the voice biometric dataset for the models training and testing.
3. Implement a few machine learning methods for the voice data classification, including Gaussian Naive Bayes classifier, support vector machine (SVM), perceptron, and k-nearest neighbors.
4. Train each model on the prepared data and assess their accuracy.
5. Analyze the results and to make conclusions about the most effective machine learning method for biometric voice identification.

Research methods:

– data collection and preparation: the voice biometric dataset preparation, including splitting into training and testing datasets;

– machine learning models implementation: a few machine learning models for the voice data classification implementation and training;

– model performance assessment: the usage of various metrics to assess each model performance;

– comparative analysis: the results comparison and the advantages and disadvantages of each machine learning method analysis.

II. State-of-the-Art

Deep Neural Networks (DNN) have become the basis for many modern biometric identification systems, including voice technologies. In [1], for example, an identification system using convolutional neural networks (CNNs) and recurrent neural networks (RNNs) is proposed. CNNs process spectrograms derived from voice data, while RNNs take into account temporal dependencies in the voice signals. These approaches provide high recognition accuracy compared to traditional methods such as GMM (Gaussian Mixture Models).

In the context of the paper [2] for small datasets, this approach can be useful for modelling systems where there are a limited number of transitions between states and the dwell time in each state cannot be accurately predicted. The use of weak learning and approximation based on Erlang distributions helps to better understand temporal dynamics even when data is scarce, making the model applicable to complex systems such as voice biometric identification where collecting sufficient data can be difficult.

The article [3] identification various methods analysis, including acoustic, linguistic and physiological approaches. This research includes a review of existing technologies, as well as a comparative analysis of their advantages and limitations. However, the article does not contain the practical aspects of the dataset usage for the machine learning methods analysis.

Studies in recent years demonstrate the growing popularity of Transformer models for speech recognition and biometric identification tasks. The paper [4] presents a model based on the Transformer architecture adapted for analysing voice characteristics, which allowed to improve the quality of identification in conditions of noise and low quality of recording.

The analytical research of the machine learning methods usage in the biometric voice identification systems development is presented in article [5]. The article discusses each method advantages and disadvantages, their application areas, and the performance comparative analysis. The disadvantage of this research is the lack of real examples and datasets in the analysis process.

The article [6] presents the performance research of the various machine learning models in the biometric voice identification context. In this work, the analysis of the speed and stability of the models based on experimental data analysis is presented. However, the identification accuracy is not assessed in this article.

The article [7] contains the research of the voice features variety used for biometric identification. The features collection and analysis methodology, their impact on system performance, and possible optimization methods are described. At the same time, the work does not contain the usage of machine learning methods for voice identification issues.

The article [8] analyzes the security and privacy issues connected to the biometric voice identification usage. The article reveals the potential system vulnerabilities, the attack and the user data protection methods. However, similarly to the previous case, the disadvantage is the letting the usage of machine learning methods for voice identification ride.

The article [9] explores the biometric voice identification usage in various areas, such as finance, medicine, security and others. Specific use cases are discussed and the effectiveness of the methods in each area is assessed. However, cybersecurity issues are not disclosed in this work.

The article [10] analyzes the current trends and directions of the biometric voice identification technologies development. Recent achievements in research area, development perspectives, and potential challenges in the system efficiency and reliability improvement are discussed. The usage of artificial intelligence in general and machine learning methods in particular is defined as one of the main research areas.

The paper [11] addresses the processing of information regarding the state of a computer system using probabilistic automata. It proposes a model for an intelligent system aimed at detecting and classifying malware by comparing characteristic feature sets of various virus classes with the system’s multiple states.

The paper [12] discusses the possibility of integrating voice biometrics with blockchain technologies to improve security. This allows for decentralised storage of biometric data and prevents it from being tampered with or leaked.

In the context [13] of the noise robustness problem, the approach proposed in this paper can be adapted to model voice biometric identification processes under data uncertainty caused by noise or compression loss. Fuzzie models, such as the proposed GERT network, can account for both temporal and probabilistic characteristics of the system, making them suitable for scenarios where the input data is incomplete or distorted. Improving modelling accuracy by accounting for uncertainty and network capabilities can increase the resilience of a biometric system to external influences and errors.

In the context of protecting voice biometric systems from spoofing attacks [14], the proposed approach can be adapted to detect malicious activities such as spoofed voice signals based on user behaviour analysis and data clustering. Using graph clustering techniques, the system could identify anomalous activities related to voice spoofing, which would improve the defence against attacks on the biometric system. Developing such algorithms for trend analysis and action clustering could increase the system's resistance to spoofing attacks and improve the overall security of biometric identification.

III. COLLECTION AND PREPARATION OF A SET OF VOICE BIOMETRIC DATA FOR TRAINING AND TESTING MODELS

The research conducted has shown that there are several methods for the voice biometric data for training and testing models collection:

– interview and dialogue audio recording: using the interviews or dialogues between two or more people audio recordings. This method allows to collect the voice data variety, including different intonations, emotional states, and speech patterns;

– text to read or tell: providing participants with texts to read or tell into the microphone. It allows to collect the data with different types of speech (e.g., reading, telling etc.) and different styles of expression;

– command and phrase for recognition: using a set of standard commands or phrases that participants must say into the microphone. This method allows to collect the data can be used for the voice recognition in specific scenarios, such as devices voice control;

– phone call and conversation audio recording: using audio recordings of the phone calls or conversations to collect voice data. This method allows to collect the data can be used for the voice identification in real-world use cases, such as telephone authentication systems;

– real-time audio recording: real-time voice data collection in a controlled environment, such as a laboratory setting. This method may include recording participants' speech in response to questions or tasks asked;

– online platforms: using the online platforms where users can provide audio recordings of their voices for research purposes. It allows to access a huge volume of data from different sources and different geographical areas.

The human voice uniqueness is conditioned by a variety of physiological features, such as the structure of the vocal cords, trachea, nasal cavities, the manner of the sounds pronunciation and the location of the teeth. Each person has a unique combination of these features, similarly to the fingerprints uniqueness. However, even the usage of unimodal biometric identification systems, including voice, does not guarantee the 100% perfect accuracy.

The main error sources of the speaker identification include the followings:

– the recording environment, including the level and type of noise, and the level of reverberation;

– the presentation factors, such as speech duration, psychophysiological state of the speaker (e.g., illness, emotional state etc.), speech language and vocal effort changes;

– the transmission channel characteristic, including interference (e.g., pulse, tonal) and distortion (e.g., microphone and transmission channel amplitude-frequency characteristics, channel encoding type etc.).

To minimize these error sources impact on the voice biometric dataset for training and testing model, robust automatic methods and algorithms that carry out the following stages of speech signal processing are developed:

– speech signal pre-processing, including the speech areas identification and the speech material quality assessment;

– soundtrack speakers’ automatic segmentation;

– voice and speech biometric features automatic extraction;

– speaker identification;

– multi-algorithmic and multi-modal mixing.

The voice biometric data collection and preparation next task is to calculate the data required for audio signal analysis. The following dataset is offered:

– meanfreq, calculated as the average frequency of the signal;

– sd (frequency standard deviation), calculated as the signal frequency standard deviation;

– median, calculated as the value separating the frequency spectrum upper and lower halves;

– Q25 (first quantile), calculated as the value that separates the data first quarter from the remaining 75%;

– Q75 (third quantile), calculated as the value that separates the data first three-quarters from the remaining quarter;

– IQR (interquartile range), calculated as the difference between the third and first quartiles;

– skew (skewness), calculated as a frequency distribution skewness measure;

– kurt (kurtosis), calculated as a frequency distribution peak sharpness measure;

– sp.ent (spectral entropy), calculated as a frequency spectrum chaos or diversity measure;

– sfm (spectral plane), calculated as a spectrum flatness measure;

– mode, calculated as the most frequently occurring frequency;

– centroid (frequency centroid), calculated as a frequency spectrum mass center;

– peakf (peak frequency), calculated as the frequency with the highest energy in the signal;

– meanfun (mean fundamental frequency), calculated as the signal fundamental frequency average;

– minfun (minimum fundamental frequency), calculated as the signal fundamental frequency minimum value;

– maxfun (maximum fundamental frequency), calculated as the signal fundamental frequency maximum value;

– meandom, calculated as the signal dominant frequency average;

– mindom (minimum dominant frequency), calculated as signal dominant frequency minimum value;

– maxdom (maximum dominant frequency), calculated as the signal dominant frequency maximum value;

– dfrange (dominant frequency range), calculated as the difference between the dominant frequency maximum and minimum values;

– modindx (modulation index), calculated as the accumulated absolute difference between adjacent fundamental frequency measurements divided by the frequency range.

The following audio data characteristics are a key for a person’s voice identification.

The specified characteristics of the voice biometric dataset frequency distribution charts are presented on the Fig. 1-5.

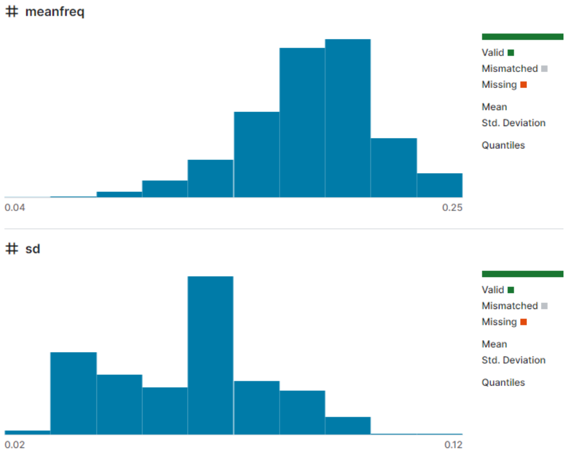

Fig. 1. The characteristics of the voice biometric dataset frequency distribution charts for meanfreq and sd datasets

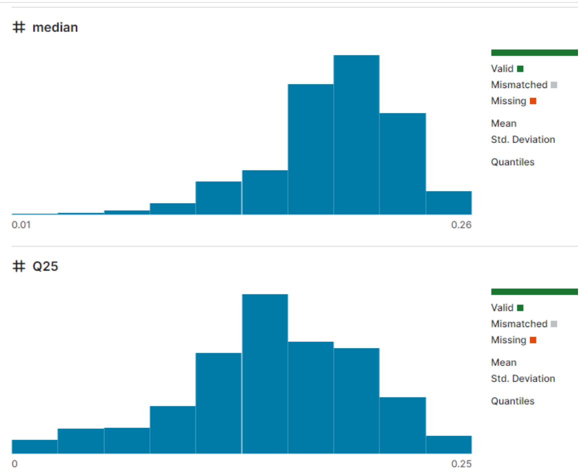

Fig. 2. The characteristics of the voice biometric dataset frequency distribution charts for median and Q25 datasets

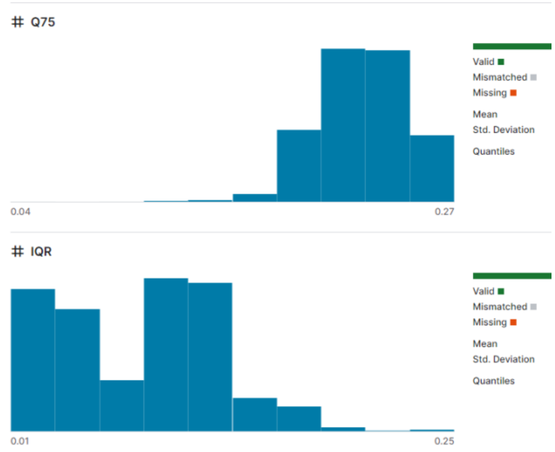

Fig. 3. The characteristics of the voice biometric dataset frequency distribution charts for Q75 and IQR datasets

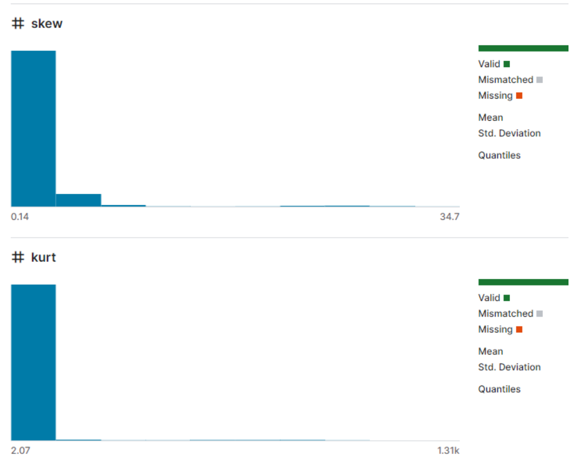

Fig. 4. The characteristics of the voice biometric dataset frequency distribution charts for skew and kurt datasets

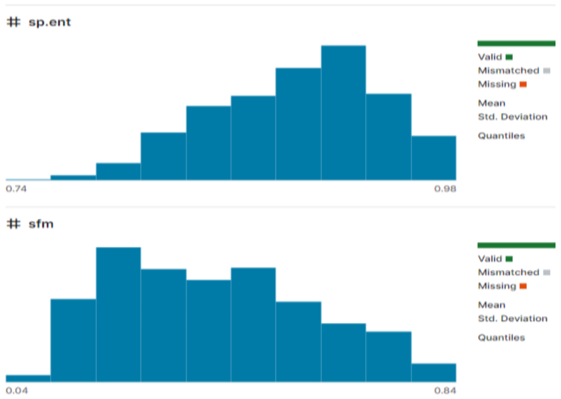

Fig. 5. The characteristics of the voice biometric dataset frequency distribution charts for sp.ent and smf datasets

IV. Implementation of Machine Learning Methods for Classification of Voice Data

For the machine learning methods implementation and the data classification studies conduction, the authors have developed software, which is a script in Python, using the following algorithm.

The program algorithm assesses the performance of various machine learning classifiers based on a data set. Each of the classifiers presented in the program can be selected for the model training and testing. Let's describe each algorithm step in details:

1. Libraries import. The program begins from the necessary libraries and classifiers import. In this research, the modules used for various machine learning methods are: ‘Perceptron’, ‘svm.SVC’, ‘KNeighborsClassifier’ (‘sklearn’ library).
2. Data reading. Next, the program reads the data from the ‘voice.csv’ file. The file contains a voice recording containing an evidence and class labels set.
3. Data partitioning. The program divides the data into training and testing datasets. Typically, cross-validation or data random splitting is used for this purpose.
4. Model training. The selected classifier (one of the three presented: Perceptron, SVC, KNeighborsClassifier) is trained on the training dataset. This happens using the ‘fit()’ method, which fits the model to the data.
5. Model testing. After the training, the model is tested on the testing dataset. This happens using the ‘predict()’ method, which predicts the class labels for the test data based on the trained model.
6. Performance assessment. The program calculates the model performance by comparing the predicted class labels with the true labels in the testing dataset. It counts the number of correctly and incorrectly predicted labels, and then calculates the model's accuracy (the percentage of correct predictions).
7. Results output. At the end of the program, the model's performance results, such as the number of correct and incorrect predictions, and the accuracy of the model are output.

This algorithm allows to quickly compare the different machine learning classifiers based on the provided data results accuracy, and select the most suitable one for a specific task. The number of experiments may vary depending on the task complexity.

The results of the research using the specified software model are presented on the Fig. 6. The condition of this model is that the first 40% of the data is used for training, and the remaining 60% of the data is used for testing.

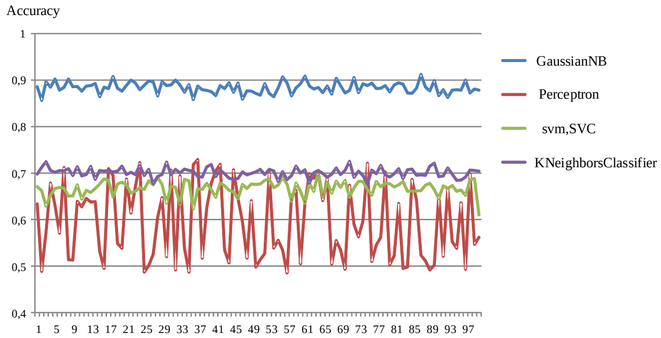

Fig. 6. The changes in the accuracy rate of the voice data classification using machine learning methods charts (Gaussian Naive Bayes classifier, support vector machine (SVM), perceptron and k-nearest neighbors)

As can be seen on the charts, for the presented example the most rational classification option is to use a Gaussian naive Bayes classifier. The classifier average accuracy is 0.883764799.

Another research example is a randomly selected for training grouped dataset.

The research results for this software model use case are presented on the Fig. 7.

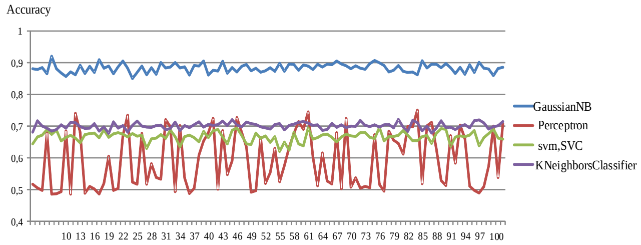

Fig. 7. The changes in the accuracy rate of the voice data classification using machine learning methods charts (Gaussian Naive Bayes classifier, support vector machine (SVM), perceptron and k-nearest neighbors) for the research second use case

As can be seen on the charts, for the presented example case, as in the previous case, the most rational classification option is to use a Gaussian naive Bayes classifier. The classifier average accuracy is 0.882991.

V. CONCLUSION AND FUTURE WORK

Thus, the comparative analysis and the assessment of the machine learning methods for biometric voice identification in data protection systems have been carried out.

To minimize the error sources impact on the voice biometric data on the preparation stage for training and testing models, robust automatic methods and algorithms have been developed which perform the speech signal pre-processing, the voice and speech biometric features automatic segmentation and selection, the multi-algorithmic and multimodal mixing and speech areas selection, and the speech material quality assessment.

Several machine learning methods for the voice data classification, including Gaussian Naive Bayes classifier, support vector machine (SVM), perceptron, and k-nearest neighbors, have been implemented.

The models developed have been trained on the prepared data and their accuracy have been assessed.

The results showed the advantages of the Gaussian Naive Bayes classifier. At the same time, for the presented dataset, even such a promising method revealed insufficient identification accuracy.

Further research is going to be aimed on the datasets pre-processing methods improvement and the Gaussian Naive Bayes classifier refinement.

REFERENCES

1. Elbayoumi, Yousef. (2024). Applying Machine Learning and Deep Learning in The Voice Biometrics Technology. 10.13140/RG.2.2.25693.42726.

2. Meleshko, Y., Raskin, L., Semenov, S., & Sira, O. (2019). Methodology of probabilistic analysis of state dynamics of multidimensional semiMarkov dynamic systems. Eastern-European Journal of Enterprise Technologies, 6(4 (102), 6–13. https://doi.org/10.15587/1729-4061.2019.184637

3. Rousan, Mohammad & Benedetto, Intrigila. (2020). A Comparative Analysis of Biometrics Types: Literature Review. Journal of Computer Science. 16. 1778-1788. 10.3844/jcssp.2020.1778.1788.

4. Ravishankar Mehta, Sindhuja Shukla, Jitesh Pradhan, Koushlendra Kumar Singh, Abhinav Kumar, A vision transformer-based automated human identification using ear biometrics, Journal of Information Security and Applications, Volume 78, 2023, 103599, ISSN 2214-2126, https://doi.org/10.1016/j.jisa.2023.103599.

5. Hooda, Susheela & Shrivastav, Supriya & Sharma, Preeti. (2023). A Study on Biometrics and Machine Learning. 1-5. 10.1109/RMKMATE59243.2023.10368885.

6. Li, Huining & Xu, Chenhan & Rathore, Aditya & Li, Zhengxiong & Zhang, Hanbin & Song, Chen & Su, Lu & Lin, Feng & Ren, Kui & Xu, Wenyao. (2020). VocalPrint: exploring a resilient and secure voice authentication via mmWave biometric interrogation. 312-325. 10.1145/3384419.3430779.

7. Lucia, C., Zhiwei, G. & Michele, N. Biometrics for Industry 4.0: a survey of recent applications. J Ambient Intell Human Comput 14, 11239–11261 (2023). https://doi.org/10.1007/s12652-023-04632-7

8. Alharbi B, Alshanbari HS. Face-voice based multimodal biometric authentication system via FaceNet and GMM. PeerJ Comput Sci. 2023 Jul 11;9:e1468. doi: 10.7717/peerj-cs.1468. PMID: 37547388; PMCID: PMC10403184.

9. Wang, J.S. Exploring biometric identification in FinTech applications based on the modified TAM. Financ Innov 7, 42 (2021). https://doi.org/10.1186/s40854-021-00260-2.

10. Feng, Zhao. (2018). Biometric Identification Technology and Development Trend of Physiological Characteristics. Journal of Physics: Conference Series. 1060. 012047. 10.1088/1742-6596/1060/1/012047.

11. S. G. Semyonov, S. Y. Gavrilenko and V. V. Chelak, "Information processing on the computer system state using probabilistic automata," *2017 2nd International Ural Conference on Measurements (UralCon)*, 2017, pp. 11-14, doi: 10.1109/URALCON.2017.8120680.

12. Youn Kyu Lee, Jongwook Jeong, Securing biometric authentication system using blockchain, ICT Express, Volume 7, Issue 3, 2021, Pages 322-326, ISSN 2405-9595, https://doi.org/10.1016/j.icte.2021.08.003.

13. Semenov, S., Zhang, L., Cao, W., Bulba, S., Babenko, V., & Davydov, V. (2021). Development of a fuzzy GERT-model for investigating common software vulnerabilities. Eastern-European Journal of Enterprise Technologies, 6(2 (114), 6–18. https://doi.org/10.15587/1729-4061.2021.243715

14. Yelyzaveta Meleshko, Mykola Yakymenko, Serhii Semenov. "A Method of Detecting Bot Networks Based on Graph Clustering in the Recommendation System of Social Network." International Conference on Computational Linguistics and Intelligent Systems (2021) 1249-1261