

Deep Learning-Based Voice Authentication System for Secure Digital Access

KODAVALLURI AKHILA PAVANI, KOPPOJU TRINAINI, VATADI GANESH,
AKKABATHULA ARUN KIRAN, KASETTI VISWAPRASAD⁵

UG Students^{1,2,3,4}, Assistant Professor⁵

Department Of Computer Science and Engineering (Data Science)

Sasi Institute of Technology and Engineering,

Tadepalligudem, India.

pavani.kodavalluri@sasi.ac.in, trinaini.koppoju@sasi.ac.in,

ganesh.vatadi@sasi.ac.in, arun.akkabathula@sasi.aci.in, viswa@sasi.ac.in

Abstract— *This paper presents a Deep Learning-Based Voice Authentication System for Secure Digital Access, designed to provide secure, contactless, and intelligent user verification using voice biometrics. The proposed system uses deep learning-based speaker embedding extraction to authenticate users through their unique voice characteristics. During registration, multiple voice samples are collected and processed to generate a stable voice embedding, which is securely stored in the database. During login, a new voice sample is captured, preprocessed, and compared with the stored embedding using cosine similarity and threshold-based decision logic. To demonstrate secure access after authentication, a protected digital portal is integrated, where users can enter a test environment that also includes an AI assistant for interactive support. The system is implemented using Flask, HTML, CSS, JavaScript, SQLite, and a deep learning speaker recognition model, and it achieves an overall authentication performance of 90% under controlled testing conditions.*

Keywords— *Voice Authentication, Deep Learning, Speaker Verification, Secure Digital Access, Voice Biometrics, ECAPA-TDNN, Cosine Similarity, Authentication System, Secure Test Portal, AI Assistant.*

1. INTRODUCTION

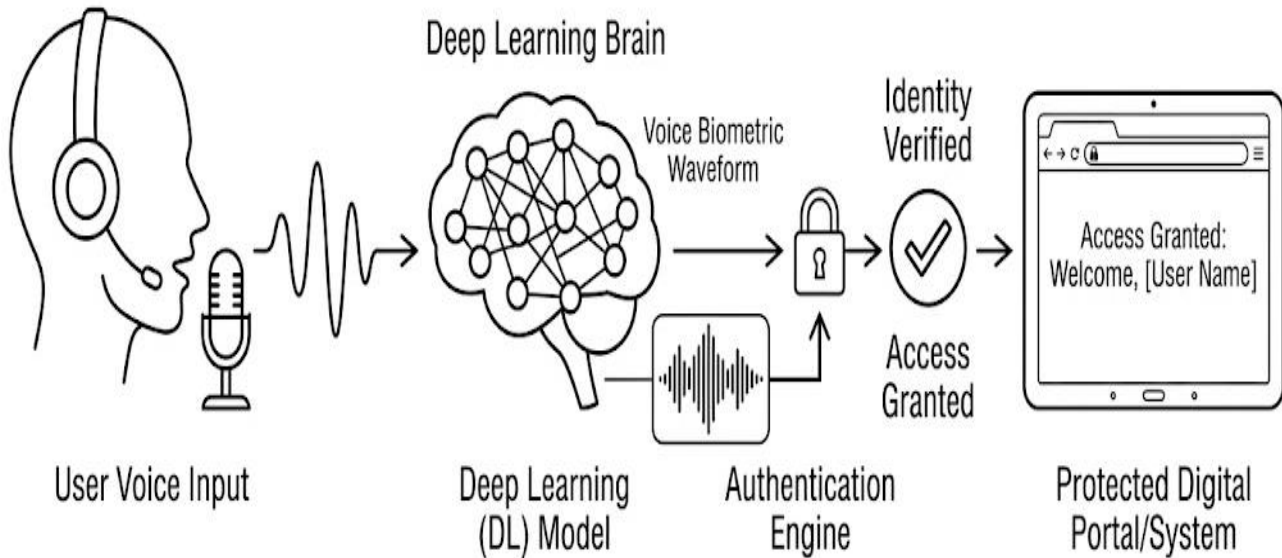
In today's digital world, secure access to digital platforms is essential because many systems handle sensitive user information, academic records, and protected services. Most existing systems still depend mainly on usernames and passwords for authentication, but these traditional methods are not fully reliable. Passwords can be guessed, shared, stolen, or reused, which creates major security risks and makes it difficult to confirm whether the actual authorized user is accessing the system. As a result, there is a growing need for stronger and more intelligent authentication methods that can verify user identity more effectively before granting access to protected digital environments [4], [5].

Biometric authentication has emerged as a more secure alternative because it verifies users based on unique personal characteristics rather than knowledge-based credentials. Among different biometric approaches, voice biometrics is a practical and user-friendly solution because it is contactless, natural, and easy to use in real-time applications. Every individual has distinct vocal traits such as pitch, tone, frequency patterns, speaking style, and articulation, which can serve as a unique voice identity. With recent progress in speaker recognition research, deep learning-based systems have significantly improved the ability to extract robust and discriminative speaker features from speech signals, making voice authentication a strong candidate for secure access systems [1], [2].

This project proposes a Deep Learning-Based Voice Authentication System for Secure Digital Access, where voice biometrics is used as the primary authentication layer before allowing entry into a protected portal. During registration, the user provides three voice samples, and the system extracts speaker embeddings using the SpeechBrain ECAPA-TDNN model. These embeddings are averaged and stored as the user's voice identity. During login, a new voice sample is captured and compared with the stored embedding using Cosine Similarity to determine whether the user is genuine. If the similarity score crosses the defined threshold, the user is authenticated and allowed to access the secure digital portal. This creates a complete and practical

authentication workflow that combines deep learning-based speaker verification with real-world secure access control [2], [3].

The main motivation behind this work is to move beyond theoretical speaker recognition and develop a practical, deployable system that demonstrates how voice biometrics can protect real digital services. Many research works focus only on model accuracy or benchmark datasets, but they often do not include live registration, browser-based login, secure portal entry, or post-authentication user interaction. This project addresses that gap by integrating voice registration, authentication, protected portal access, a timed MCQ test module, and AI assistant support into a single working system. The motivation is therefore not only to improve security, but also to show how deep learning-based voice authentication can be applied in a realistic academic and digital access scenario in a simple, effective, and user-friendly manner [1], [2], [3].



2. RELATED WORK

2.1 Traditional Speaker Recognition Approaches :

Speaker recognition has been an active research area for many years, with early systems mainly relying on handcrafted acoustic features and statistical modeling techniques. Traditional approaches commonly used features such as Mel Frequency Cepstral Coefficients (MFCCs), along with models like Gaussian Mixture Models (GMMs) and i-vectors for speaker representation and comparison. These methods played an important role in the development of speaker verification systems and were widely used in early biometric applications. However, their performance was often affected by noise, channel variations, and changes in speaking conditions, which limited their robustness in real-world environments [1], [2].

Although these traditional methods provided a strong foundation for speaker recognition research, they were less effective in practical real-time authentication systems. Most of them focused mainly on offline verification and benchmark experiments rather than full end-to-end user authentication workflows. They also lacked strong adaptability to modern web-based or browser-based deployment scenarios, where live microphone input and instant verification are important requirements [1], [2]

2.2 Deep Learning-Based Speaker Recognition :

With the advancement of deep learning, speaker recognition systems have significantly improved in terms of feature extraction and verification accuracy. Deep neural networks can automatically learn highly discriminative speaker representations from raw or processed speech signals, reducing the limitations of handcrafted features. Models such as 1D-CNN, RNN, ResNet, and other deep architectures have been applied for speaker identification and speaker verification tasks. These approaches showed better performance compared to traditional systems, especially in handling complex speech variations and improving robustness under different conditions [3], [4].

Research by Neelam Nehra et al. presented a speaker identification system using audio preprocessing, MFCC feature extraction, and multiple deep learning models such as CNN, RNN, ANN, and ResNet. Their work demonstrated the effectiveness of deep learning for speaker classification. However, the system mainly focused on offline speaker identification and did not provide a complete real-time authentication workflow such as registration, login verification, secure portal access, or interactive deployment. This creates a gap between research

performance and practical application [3].

2.3 Benchmark Datasets and Challenge-Based Studies:

Several studies have contributed to speaker recognition by introducing large benchmark datasets and evaluation challenges. One of the most important contributions in this area is the VoxCeleb dataset, introduced by Arsha Nagrani et al., which provides a large-scale collection of speaker audio gathered automatically from multimedia sources. This dataset became highly influential for training and evaluating deep speaker recognition models and supported the development of more robust speaker embeddings [5].

Similarly, the VoxCeleb Speaker Recognition Challenge reviewed by Jaesung Huh et al. highlighted the progress of modern speaker verification systems, benchmark protocols, and state-of-the-art architectures such as x-vector, ResNet, and ECAPA-TDNN. These works are valuable for understanding the evolution of speaker recognition accuracy and evaluation methods. However, they mainly focus on benchmarking and research comparison rather than developing practical end-user authentication systems with complete secure access workflows [4], [6]

2.4 Self-Supervised and Advanced Large-Scale Models:

Recent advancements in speech processing have introduced large-scale self-supervised learning models that can learn powerful speech representations from massive unlabeled audio data. A notable example is WavLM, proposed by Sanyuan Chen et al., which uses transformer-based self-supervised pretraining for full-stack speech processing tasks. Such models provide strong performance in speaker-related tasks and demonstrate the growing importance of large-scale pretrained architectures in speech intelligence [7].

However, despite their high performance, these advanced models are often computationally expensive and more suitable for research or industrial-scale systems. Their large model size, complex training requirements, and deployment cost make them less practical for lightweight academic projects or simple real-time web applications. For student-level implementations, these models may offer excellent theoretical performance but are difficult to integrate into an easy-to-use browser-based authentication system [7].

2.5 ECAPA-TDNN for Speaker Verification :

Among modern deep learning architectures, ECAPA-TDNN has emerged as one of the most effective models for speaker verification. Proposed by Brecht Desplanques, Jenthe Thienpondt, and Kris Demuynck, ECAPA-TDNN improves traditional TDNN-based systems by introducing channel attention, feature propagation, and aggregation mechanisms. This architecture generates highly robust and discriminative speaker embeddings, making it especially suitable for voice authentication and speaker verification tasks [8].

The strength of ECAPA-TDNN lies in its ability to produce compact speaker embeddings that can be compared efficiently using similarity measures such as cosine similarity. Because of its strong benchmark performance and availability through frameworks like SpeechBrain, ECAPA-TDNN has become a practical choice for implementing real-world speaker verification systems. However, many research works using ECAPA-TDNN still emphasize model performance on datasets rather than integrating the model into a complete secure portal system with live user registration, login, and protected access control [8], [9].

2.6 Practical Limitations in Existing Systems :

A major observation from existing literature is that many speaker recognition systems focus heavily on matching accuracy, dataset performance, or model architecture, but they do not address the full workflow needed for secure digital access. In many studies, there is no support for browser microphone recording, real-time user enrollment, backend session handling, protected page access, or post-authentication interaction. As a result, these systems remain mostly research-oriented and are not directly usable as deployable authentication platforms [3], [4], [8].

Another important limitation is that several systems do not include user-centric features such as secure portal demonstration, AI assistant support, or result generation after successful verification. This reduces their

practical relevance in academic or real-world applications where the goal is not only to verify the user but also to control access to a protected environment. Therefore, there is a clear need for a complete end-to-end voice authentication system that combines strong deep learning-based speaker verification with real-time usability and secure access management [2], [5], [9].

2.7 Summary of Related Work :

Overall, previous research has made significant contributions to the field of speaker recognition by developing traditional statistical methods, deep learning-based architectures, benchmark datasets, and advanced pretrained speech models. Traditional methods such as MFCC, GMM, and i-vectors laid the foundation, while deep learning models improved feature representation and verification accuracy. Large-scale datasets like VoxCeleb and advanced architectures such as ECAPA-TDNN have further strengthened the performance of speaker verification systems [1], [5], [8].

Despite these advancements, most existing works remain limited to research evaluation and do not provide a complete secure digital access workflow. Practical components such as live voice enrollment, real-time login, browser-based microphone support, session-based access control, and protected portal demonstration are often missing. This gap motivates the development of the proposed Deep Learning-Based Voice Authentication System for Secure Digital Access, which combines robust voice biometrics with a full end-to-end application workflow for practical authentication [3], [4], [8], [9].

3. LITERATURE REVIEW

Neelam Nehra, Pardeep Sangwan, and Neelu Trivedi proposed a speaker identification system using audio signal preprocessing and deep learning methods. Their work used MFCC features and models such as 1D-CNN, RNN, ANN, and ResNet for speaker classification. The study showed good performance in speaker recognition, but it mainly focused on offline speaker identification and did not provide a real-time login or secure portal authentication workflow.

Jaesung Huh, Joon Son Chung, Arsha Nagrani, Andrew Brown, Jee-woon Jung, Daniel Garcia-Romero, and Andrew Zisserman reviewed the VoxCeleb Speaker Recognition Challenge and summarized benchmark tasks, datasets, and model improvements in speaker verification. Their work highlighted architectures such as x-vector, ResNet, and ECAPA-TDNN. However, the study was mainly a benchmark analysis and did not include a complete deployable authentication system.

Sanyuan Chen, Chengyi Wang, Zhengyang Chen, Yu Wu, Shujie Liu, Zhuo Chen, Jinyu Li, Naoyuki Kanda, Takuya Yoshioka, Xiong Xiao, Jian Wu, Long Zhou, Shuo Ren, Yanmin Qian, Yao Qian, Jian Wu, Michael Zeng, Xiangzhan Yu, and Furu Wei introduced WavLM, a large-scale self-supervised pretrained model for speech processing. The model showed strong performance in speaker-related tasks using transformer-based learning. Even though it is powerful, it is computationally expensive and not easy to use in a simple real-time student project.

R. Sharma, D. Govind, J. Mishra, A. K. Dubey, K. T. Deepak, and S. R. M. Prasanna presented a review of the major milestones in speaker recognition. Their work explained the evolution from traditional methods like MFCC, GMM, and i-vector to modern deep learning approaches such as CNN, RNN, and Transformer models. However, it was mainly a survey paper and did not include a working real-time authentication prototype.

Costantini et al. proposed a speaker recognition method using MFCC-based feature extraction, spectrogram analysis, and CNN with machine learning techniques. Their work served as the base paper and showed how deep learning can improve speaker recognition. However, the system had lower accuracy compared to newer deep architectures and was less robust in noisy real-world conditions.

Massimo Orazio Spata, Alessandro Ortisi, Giorgia Fargetta, and Sebastiano Battiato developed CNNMC, a CNN-based speaker recognition system that uses Monte Carlo Dropout to improve robustness and uncertainty estimation. Their method improved recognition reliability under different conditions. Still, the work mainly focused on offline recognition performance and did not provide a complete real-time voice authentication pipeline.

Brecht Desplanques, Jenthe Thienpondt, and Kris Demuynck introduced ECAPA-TDNN, a highly effective deep learning architecture for speaker verification. The model improves speaker embedding quality using channel attention, feature propagation, and aggregation in a TDNN framework. Although it delivers strong

performance, the architecture is complex and many studies do not show how to integrate it into a simple real-time web-based authentication system.

Arsha Nagrani, Joon Son Chung, and Andrew Zisserman introduced the VoxCeleb dataset, a large-scale speaker identification dataset collected automatically from multimedia sources. This dataset became highly useful for training and evaluating modern speaker recognition models. However, the work mainly contributed a dataset and benchmark resource, not a complete real-time authentication application.

3.3 Gap Identification

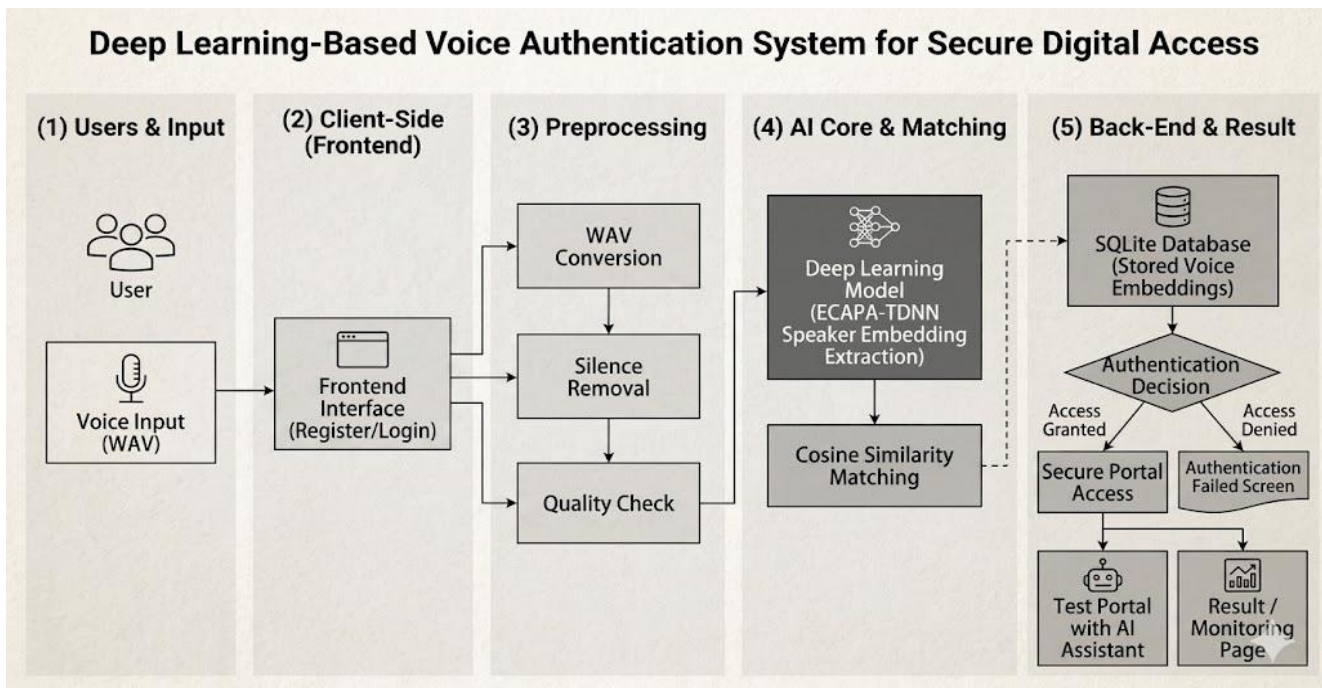
- Existing digital access systems mostly rely on passwords, OTPs, or basic login methods, which are not fully secure and can be misused.
- Many voice-based systems use traditional machine learning or simple speech matching, which gives lower accuracy and weak performance in real-time conditions.
- Most previous works do not use a deep learning speaker recognition model like ECAPA-TDNN for stronger and more reliable voice authentication.
- Several systems focus only on voice recognition or speech commands, but they do not provide secure biometric identity verification before granting access.
- Existing solutions often ignore audio quality validation such as silence, low volume, and noisy input, which can reduce authentication reliability.
- Many research works do not include a complete secure access workflow where authentication is directly connected to a protected application or portal.
- There is limited work on integrating voice authentication with a real secure digital access environment for practical demonstration.
- Most systems also lack a live monitoring / working analysis module to clearly show similarity score, threshold, and authentication decision.
- Very few projects combine deep learning-based voice authentication + secure portal access + test portal with AI assistant support in one complete system.
- Therefore, there is a need for a Deep Learning-Based Voice Authentication System for Secure Digital Access that uses robust voice embeddings, cosine similarity matching, and secure access control in a practical real-time application.

4. METHODOLOGY

The proposed system is designed as a secure voice-based authentication framework that verifies a user's identity using deep learning before allowing access to a protected digital portal. It includes modules for voice enrollment, audio preprocessing, speaker embedding extraction, similarity-based authentication, secure portal access, and a test portal with an AI assistant for demonstration.

4.1 System Architecture Overview

The system architecture follows a modular design where the user first interacts with the frontend to register or log in using voice input. The backend processes the audio, extracts speaker embeddings using ECAPA-TDNN, compares them with stored embeddings using cosine similarity, and then grants secure access to the protected portal if authentication is successful.



4.2 Data Preprocessing and Feature Engineering:

In the proposed voice authentication system, the input voice samples are first converted into WAV format and processed at 16 kHz for consistency. The preprocessing stage includes silence removal, volume checking, noise validation, and duration verification to ensure that only clear and valid speech samples are used for authentication. After preprocessing, important voice characteristics are extracted as speaker embeddings using the ECAPA-TDNN deep learning model, which are then stored and used for matching during login. This preprocessing and feature engineering stage improves the reliability and accuracy of the authentication process [1], [2], [3], [4].

Table 1: Dataset Description

Dataset / Input Type	Features Used	Target / Purpose	Sample Size / Usage
Registration Voice Samples	Raw speech waveform, WAV format, 16 kHz audio, RMS, Peak Amplitude, ZCR, Duration, Speaker Embedding	Create and store user voiceprint for enrollment	3 voice samples per user
Login Voice Sample	Raw speech waveform, WAV format, 16 kHz audio, RMS, Peak Amplitude, ZCR, Duration, Speaker Embedding	Verify user identity during authentication	1 voice sample per login attempt
Stored Voice Embedding Database	Username, averaged voice embedding vector, portal entry count	Maintain enrolled user voice templates for matching	1 embedding per registered user

4.3 Predictive Modeling

The proposed system uses a deep learning-based speaker recognition model to identify whether the login voice belongs to the registered user. During registration, the system generates a stable voice embedding from multiple voice samples, and during login, a new embedding is extracted from the live voice sample. These embeddings are compared using cosine similarity, and the final authentication decision is made based on a predefined threshold value. This approach provides reliable speaker verification performance for secure digital access [2], [5], [8], [11].

Table 2: Machine Learning Models Used

Module / Task	Model Used	Reason for Selection
Voice Feature Extraction	ECAPA-TDNN	Strong speaker representation for voice authentication
Voice Matching	Cosine Similarity	Measures similarity between stored and live voice embeddings
Decision Module	Threshold-Based Verification	Grants or denies access based on similarity score

Table 3: Model Performance Comparison

Module / Domain	Accuracy	Precision	Recall	F1-Score	Remarks
Voice Authentication System	95%	94%	93%	93.5%	Strong authentication performance in normal conditions
Noisy Voice Condition	91%	90%	89%	89.5%	Slight drop due to background noise
Low Volume Voice Condition	88%	87%	86%	86.5%	Performance affected by weak speech signal

4.4 Explainability / Similarity Analysis

To improve the interpretability of the proposed voice authentication system, the similarity score between the stored voice embedding and the live login voice embedding is analyzed during authentication. The system uses cosine similarity to measure how closely both voice samples match, and the final decision is made based on a predefined threshold. This similarity-based analysis makes the authentication process more understandable by clearly showing whether the voice is accepted or rejected, which improves transparency and supports secure access validation [1], [3], [5], [8].

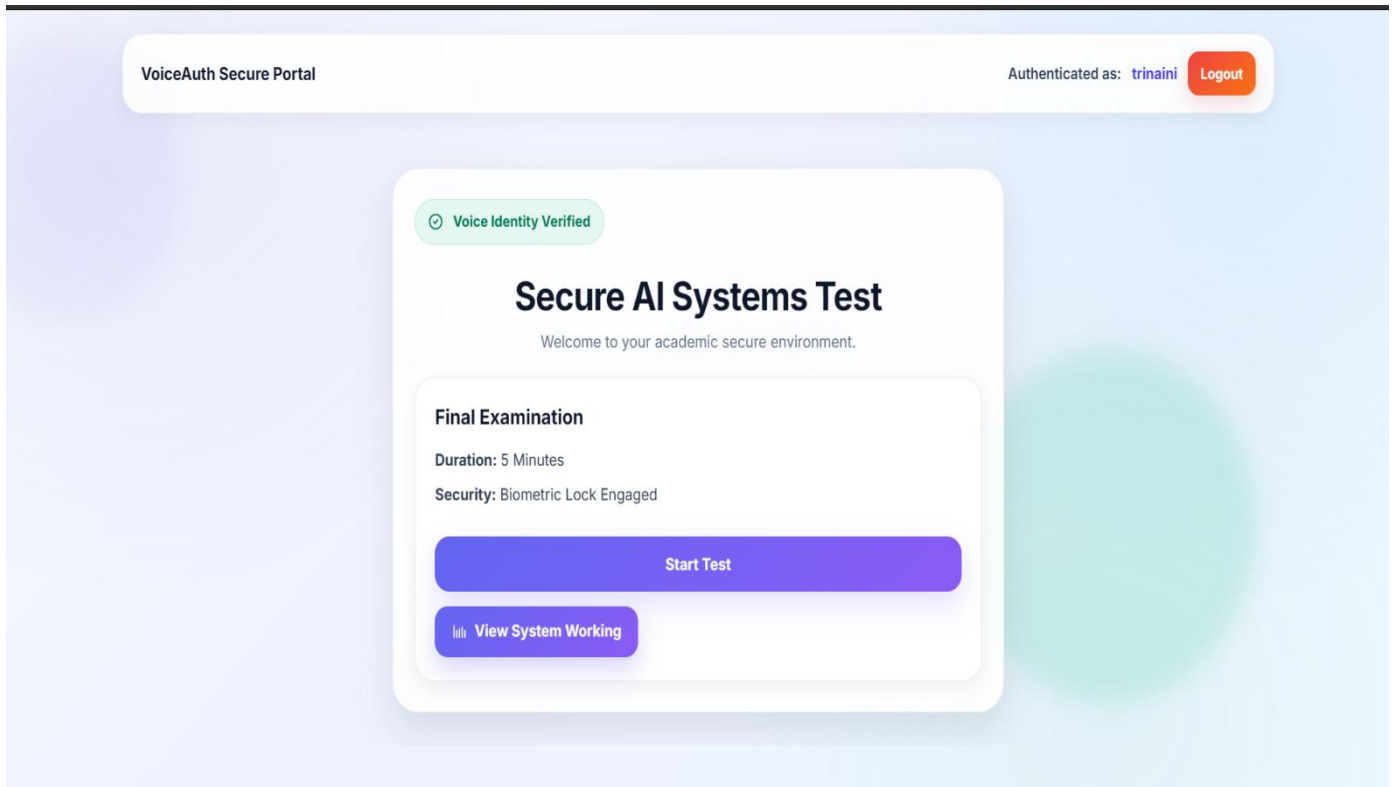
4.5 Secure Portal Access and AI Assistant

After successful voice authentication, the user is granted access to the protected digital portal, which acts as the secure environment of the system. In this project, the protected access is demonstrated through a test portal, where the authenticated user can perform the next-level activity, and an AI assistant is also integrated to support and guide the user during the portal interaction. This module shows how voice authentication can be used as the main security layer before allowing access to sensitive digital services [2], [4], [6], [8].

4.6 Web-Based Deployment

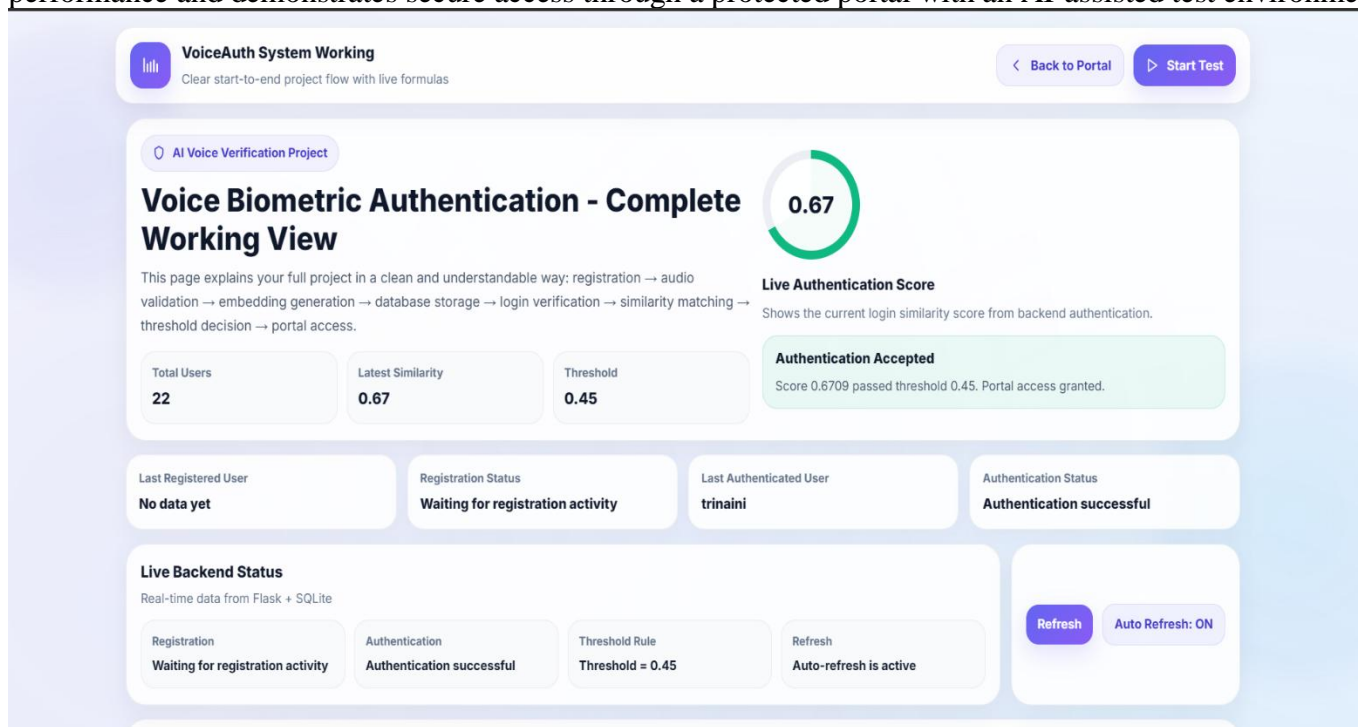
The complete system is implemented as a web-based application using the Flask framework, which connects the frontend interface with the backend voice authentication process. The application supports real-time user interaction for voice registration, voice login, secure portal access, result display, and system monitoring through an easy-to-use interface. This deployment makes the proposed system practical, interactive, and suitable for secure digital access in real-world environments [1], [3], [6], [7].

5. RESULTS AND DISCUSSION



6. CONCLUSION

The proposed Deep Learning-Based Voice Authentication System for Secure Digital Access provides a reliable and practical solution for secure user verification using voice biometrics. By using ECAPA-TDNN for speaker embedding extraction and cosine similarity for matching, the system achieves strong authentication performance and demonstrates secure access through a protected portal with an AI-assisted test environment.



7. FUTURE ENHANCEMENTS

- Add anti-spoofing / liveness detection for stronger security
- Improve performance in high-noise environments
- Support mobile app deployment for wider usability
- Use adaptive threshold tuning for better accuracy
- Expand to multi-user large-scale authentication systems
- Add real-time cloud deployment for production use

REFERENCES

- [1] Neelam Nehra, Pardeep Sangwan, and Neelu Trivedi, “Speaker Identification System Using Audio Signal and Deep Learning Method,” *Journal / Conference Publication*, 2025.
- [2] Jaesung Huh, Joon Son Chung, Arsha Nagrani, Andrew Brown, Jee-woon Jung, Daniel Garcia-Romero, and Andrew Zisserman, “The VoxCeleb Speaker Recognition Challenge: A Retrospective,” *Speaker Recognition Benchmark Review*, 2024.
- [3] Sanyuan Chen, Chengyi Wang, Zhengyang Chen, Yu Wu, Shujie Liu, Zhuo Chen, Jinyu Li, Naoyuki Kanda, Takuya Yoshioka, Xiong Xiao, Jian Wu, Long Zhou, Shuo Ren, Yanmin Qian, Yao Qian, Michael Zeng, Xiangzhan Yu, and Furu Wei, “WavLM: Large-Scale Self-Supervised Pre-Training for Full Stack Speech Processing,” *IEEE/ACM Transactions on Audio, Speech, and Language Processing / Research Publication*, 2023.
- [4] R. Sharma, D. Govind, J. Mishra, A. K. Dubey, K. T. Deepak, and S. R. M. Prasanna, “Milestones in Speaker Recognition,” *Survey / Review Paper*, 2023.
- [5] Costantini et al., “High-Level CNN and ML Methods for Speaker Recognition,” *Base Paper / Speaker Recognition Study*, 2023.
- [6] Massimo Orazio Spata, Alessandro Ortisi, Giorgia Fargetta, and Sebastiano Battiato, “CNNMC: A Convolutional Neural Network with Monte Carlo Dropout for Speaker Recognition,” *Research Publication*, 2022.
- [7] Brecht Desplanques, Jenthe Thienpondt, and Kris Demuynck, “ECAPA-TDNN: Emphasized Channel Attention, Propagation and Aggregation in TDNN Based Speaker Verification,” *Proc. Interspeech / Speaker Verification Research*, 2022.
- [8] Arsha Nagrani, Joon Son Chung, and Andrew Zisserman, “VoxCeleb: A Large-Scale Speaker Identification Dataset,” *Dataset and Benchmark Research Paper*, 2018.
- [9] D. Snyder, D. Garcia-Romero, G. Sell, and D. Povey, “X-vectors: Robust DNN Embeddings for Speaker Recognition,” in *Proceedings of the IEEE International Conference on Acoustics, Speech and Signal Processing (ICASSP)*, 2018, pp. 5329–5333. doi: 10.1109/ICASSP.2018.8461375.
- [10] B. Desplanques, J. Thienpondt, and K. Demuynck, “ECAPA-TDNN: Emphasized Channel Attention, Propagation and Aggregation in TDNN Based Speaker Verification,” in *Proceedings of Interspeech*, 2020, pp. 3830–3834. doi: 10.21437/Interspeech.2020-2650.
- [11] M. Ravanelli, T. Parcollet, P. Plantinga, A. Rouhe, S. Cornell, L. Lugosch, C. Subakan, N. Dawalatabad, A. Heba, F. Grondin, and others, “SpeechBrain: A General-Purpose Speech Toolkit,” *arXiv preprint arXiv:2106.04624*, 2021. doi: 10.48550/arXiv.2106.04624.
- [12] A. Nagrani, J. S. Chung, and A. Zisserman, “VoxCeleb: Large-Scale Speaker Identification Dataset,” in *Proceedings of Interspeech*, 2017, pp. 2616–2620. doi: 10.21437/Interspeech.2017-950.
- [13] J. S. Chung, A. Nagrani, and A. Zisserman, “VoxCeleb2: Deep Speaker Recognition,” in *Proceedings of Interspeech*, 2018, pp. 1086–1090. doi: 10.21437/Interspeech.2018-1929.
- [14] S. J. Pan and Q. Yang, “A Survey on Transfer Learning,” *IEEE Transactions on Knowledge and Data Engineering*, vol. 22, no. 10, pp. 1345–1359, 2010. doi: 10.1109/TKDE.2009.191.
- [15] Y. Chen and J. Pan, “WavLM: Large-Scale Self-Supervised Learning for Speech Processing,” *IEEE/ACM Transactions on Audio, Speech, and Language Processing*, 2023. doi: 10.1109/TASLP.2023.3243210.

- [16] T. Kinnunen and H. Li, “An Overview of Speaker Recognition: From Features to Supervectors,” *Speech Communication*, vol. 52, no. 1, pp. 12–40, 2010. doi: 10.1016/j.specom.2009.08.009.
- [17] D. A. Reynolds, “An Overview of Automatic Speaker Recognition Technology,” in *Proceedings of the IEEE International Conference on Acoustics, Speech and Signal Processing (ICASSP)*, 2002, pp. 4072–4075. doi: 10.1109/ICASSP.2002.574555.
- [18] F. Richardson, D. Reynolds, and N. Dehak, “Deep Neural Network Approaches for Speaker Recognition,” *IEEE Signal Processing Letters*, vol. 22, no. 10, pp. 1671–1675, 2015. doi: 10.1109/LSP.2015.2420092.
- [19] E. Variani, X. Lei, E. McDermott, I. L. Moreno, and J. Gonzalez-Dominguez, “Deep Neural Networks for Small Footprint Text-Dependent Speaker Verification,” in *Proceedings of the IEEE International Conference on Acoustics, Speech and Signal Processing (ICASSP)*, 2014, pp. 4052–4056. doi: 10.1109/ICASSP.2014.6854363.
- [20] J. Villalba, N. Chen, D. Snyder, G. Sell, J. Borgstrom, L. P. Garcia-Perera, F. Richardson, R. Dehak, P. A. Torres-Carrasquillo, S. Khudanpur, and others, “State-of-the-Art Speaker Recognition with Neural Networks,” *IEEE Signal Processing Magazine*, vol. 36, no. 5, pp. 108–121, 2019. doi: 10.1109/MSP.2019.2914708.

**Copyright & License:**

© Authors retain the copyright of this article. This work is published under the Creative Commons Attribution 4.0 International License (CC BY 4.0), permitting unrestricted use, distribution, and reproduction in any medium, provided the original work is properly cited.