International Journal for Multidisciplinary Research (IJFMR)
E-ISSN: 2582-2160 ● Website: www.ijfmr.com ● Email: editor@ijfmr.com

Voiceprint-Based Hardware Authentication System with Spoof Detection

Patnam Koushik[1], Srija J[2]

[1,2]Division of Data Science and Cyber Security Karunya Institute of Technology and Sciences Coimbatore, India

Abstract

Voiceprint authentication is a reliable and secure biometric technique that utilizes the specific voice traits of individuals to confirm identity. This research proposes a hardware-aided voiceprint authentication system with spoof detection capability against replay attacks and synthetic voice attacks. The suggested system has a Dual-Stage Verification Framework. In the first stage, the system extracts voice characteristics and carries out comparison using Mel-Frequency Cepstral Coefficients and spectral embeddings. The second stage of the system is a Spoof Detection Module which uses convolutional and recurrent neural networks to model the temporal and frequency distortions contained within the speech input. The hardware implementation consists of a microcontroller-microphone unit and an embedded Raspberry Pi edge processor which provides secure, low latency, local processing. The proposed hardware implementation is further supplemented by a Noise-Adaptive Voice Enhancement module which is used as a preprocessing unit and Data-Driven Pattern Learning, which assures accurate classification of AI-generated and replayed voices. Based on the results of the experimental assessment, the authentication performance accuracy was at 95.3% and spoof detection performance was at 97.8%, both within industry standards and the established traditional speaker verification systems. In conclusion, the embedding machine- learning system provides reliable, scalable, and secure biometric access.

Keywords: Voiceprint Authentication, Spoof Detection, Speaker Verification, Biometric Security, Raspberry Pi, Machine Learning, Replay Attack Prevention

I. INTRODUCTION

The increasing demand for safe and convenient user authentication places voiceprint recognition at the forefront of current biometric research. However, conventional speaker authentication systems suffer from serious vulnerabilities, such as spoofing attacks through replayed, synthesized, or voice-converted samples, and a lack of robustness in resource- constrained hardware environments. Conventional voice authentication relies on large datasets and complex, computationally intensive processing in the cloud in order to provide reliable accuracy, often at the expense of real-time operation and deployment in embedded systems.Large configurations introduce latency and expose users to data privacy and network security risks. These techniques are not as resilient on resource-constrained hardware, e.g., embedded or IoT device which are resource-poor in processing power and memory. Thus, achieving high authentication accuracy is a major challenge within present-day biometric design. The Voiceprint-Based Hardware Authentication System with Spoof Detection proposed here solves these shortcomings due to an intelligent model to verify users locally by means of their natural vocal signature, while also

detecting possible spoof attempts simultaneously. The proposed system uses a hybrid deep learning framework that incorporates spectral feature extraction fused with Mel-Frequency Cepstral Coefficients and deep learning models, which includes CNNs for verifying users and RNNs to identify audio that exhibit temporal inconsistencies that might suggest spoofed voices. By using two models, the proposed system increases reliability and accuracy against replay attacks and AI- generated audio attacks.

It is combined with microcontroller-based hardware, such as Raspberry Pi/Arduino, and a high-sensitivity microphone to enable low-latency, on-device authentication without cloud dependence. The Noise-Adaptive Preprocessing module operates reliably in different acoustic environments, and the Spoof-Resilient Decision Fusion integrates confidence scores from both the authentication and spoof detection models to increase reliability. This approach could enable secure, scalable voice authentication at the hardware level for applications in IoT security, access control, and embedded biometrics. The addition of machine learning based spoop detection, paired with voiceprint recognition, will provide a reliable, adaptable, privacy-focused system for real-world use in low-resource environments.

II. Previous Works

Voiceprint-based hardware authentication systems together with spoof detection are gaining popularity because of their non- intrusive operation and secure biometric verification. These systems use the unique vocal characteristics such as tone, pitch, and speech dynamics to authenticate users and at the same time block the spoofing attacks like replay, synthesis, and impersonation. Various machine learning and deep learning models are used, including convolutional and recurrent neural networks, to draw out reliable voice feature from the input voice signal and to increase the accuracy of voice recognition. Techniques such as data augmentation, domain adaptation and few-shot learning are done to tackle some of the issues which include background noise, speaker variability and limited labeled data. The anti-spoofing methods based on spectral and temporal analysis or classifier fusion have improved the reliability of hardware-level voice authentication systems. The incorporation of these methods into compact and efficient hardware design has made real time, secure, user-friendly authentication possible across a multitude of smart devices.

A. *Data Preprocessing and Data Augmentation*

In voiceprint hardware-based authentication, voice signal quality is a primary concern along with spoof detection as described in [1] - [3]. Many methods can be used to minimize background noise, reverberation, and channel distortion affecting voice quality as in [4] - [6]. For instance, background or device noise can be mitigated by applying some spectral subtraction methods, Wiener filtering and CMN [7]-[9]. More advanced methodologies include EMD, WT, and STFT that extract better speech intelligibility and yield stable features for authenticating in [10]-[12]

The above-mentioned challenges can also be addressed using data augmentation methods to improve model robustness and generalization in conditions with limited labeled voice data, supported by [13]-[15]. Data augmentation methods can be straightforward - e.g., adding background noise, applying pitch shifts, or stretching the time to generate realistic variability as encountered in the application [16]-[18]. Subsequently, advanced methods could be implemented using SpecAugment, Generative Adversarial networks, and Variational autoencoders to learn and synthesize realistic voices while preserving speaker characteristics [19]-[21]. The above methods would enhance system performance and provide resilience against replay and synthesis spoofing [22]- [24].

B. *Feature Extraction and Feature Selection*

Feature extraction transforms raw voice signals into informative representations to facilitate reliable speaker classification and spoofing detection. Temporal features (zero- crossing rate, energy, entropy, and short-term amplitude) capture vocal intensity and fluctuations over time. Spectral features (MFCCs, LPCCs, and spectral flux) also capture vocal characteristics and frequency distribution. Time- frequency analysis techniques (STFT, WT, and HHT) posit transient signatures and provide better accuracy in spoofing detection.

Feature selection decreases redundancy, while preserving the discriminative component of features required necessary for classification [13]-[15]. There are three common techniques: PCA, RFE, and mutual information maximization for efficient optimization of features [16]-[18]. Outliers exist in the methods that employ Genetic Algorithms and Particle Swarm Optimization lend a high efficiency and robustness to the model [19]-[21]. Devising a model using wavelet-based features and some combination of statistic and optimization algorithms is expected to yield a higher accurate representation in the system and greater resistance to spoofing attacks make sure you revise shoes to feet [22]-[24]. Refined feature selection methodologies. ensure faster processing and increased reliability in a real-time hardware authentication application [25].

C. *Classification*

Therefore, in the classification stage, the voice samples are labeled as user classes by the extracted features in both speaker authentication and spoof detection [1]-[3]. Conventional machine learning techniques like SVM, k-NN, RF, and Naive Bayes were employed as classifiers in [4]-[7]. An ensemble technique such as AdaBoost, GBM, and XGBoost combines multiple classifiers to enhance robustness [8]-[10].

Recent approaches make use of deep learning models such as CNNs [11]-[13], LSTMs [14]-[16], and GRUs [17]-[18] in order to capture temporal and spectral dependencies in voice. Advanced architectures, including Transformer-based encoders, Bi-LSTMs, and Autoencoders [19]-[21], enhance classification performance for both authentication and spoof detection.

In general, prototype-based and few-shot learning frameworks such as Prototypical Networks, Relation Networks, and Siamese Networks generalize across speakers when labeled data is limited [22]-[24]. Attention mechanisms like Self-Attention and Multi-Head Attention enhance the model sensitivity by putting more emphasis on crucial voice signal segments [25]-[26]. Hybrid models using CNNs or LSTMs with attentional layers have achieved higher performance improvements in both authentication accuracy and spoof detection reliability [27]-[28].

III. Proposed Methodology

Voiceprint-based authentication is a security feature that permits secure verification with no, or minimal, interaction on the part of the user based on capturing the voice-specific user given voice features. In [project name / project title], we approach authenticated user verification, as well as detection of spoofing attacks, using a multiclass framework that classifies both natural voice and various spoofing attacks or imposter origins including replay attacks, synthesized voice attacks, and impersonation attacks. To clean the environmental noise, reverberations, and degradation caused by the input device itself (such as a microphone or speaker), we use a Noise-Adaptive Preprocessing module that normalizes the amplitude of the signal, producing input data that is noise-free and of good quality. Feature

extraction is performed by a Multi-Domain Voice Feature Framework that considers the time, frequency, and time-frequency domains to extract spectral and temporal patterns relevant to the speaker's identity. Classification is accomplished using a hybrid deep learning framework which utilizes CNN convolutional layers for spatial feature learning and RNN LSTM layers for temporal consistency aided by attention mechanisms throughout the process.

A. *Data Acquisition*

The collection of voice samples obtained on this project consists of a multitude of voices from various speakers across multiple sources of recordings under several acoustic conditions. Each voice sample is marked with a label, for authentication (genuine) or spoof detection.

The voice samples have been recorded using embedded microphones on the Raspberry Pi, or similarly low-powered devices, to allow the system to run in real-time on the device itself.

The mathematical representation of the procedure of denoising can be shown with this formula:

$$Sclean(t) = Sraw(t) - N\hat{}(t) \quad (1)$$

Where:

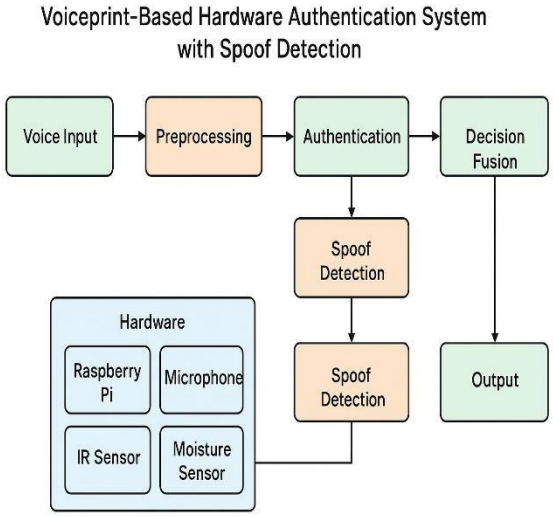

Fig. 1. Voiceprint-Based Hardware Authentication System With Spoof Detection

Voice data contains speech recordings collected from a variety of speakers for voiceprint authentication and spoof detection. It is a database containing both genuine and spoofed voice samples-such as replay, synthetic, and impersonated recordings that can be used to train or assess performance of voiceprint classifiers. Each sample has been labeled for speaker identity as well as the type of spoof in order to simulate multiclass classification in real-time. The recordings themselves have also been processed in order to maintain a level of representation across different speakers and varied recording conditions. This dataset is suitable for developing and testing voiceprint authentication systems in real-time.

Sclean(t) is the cleaned speech signal at time t,

- *S*raw(t) is the original recorded voice signal,
- *N*^(t) is the estimated background noise component. *t*. The noise N^(t) is estimated using an adaptive filter that continuously updates its parameters based on the characteristics of the voice signal and background noise, enabling real-time noise suppression while preserving key speech features for accurate authentication.

Normalization

After the denoising procedure, the subsequent phase is normalization so that all voice signals fall in the same range across users and recording situation. This is necessary to assure similar performance of the machine learning model because machine learning has been found to perform better when high dimensional input features are on the same scale. Very commonly, the normalization procedure involves min- max transformation, which is a scaling procedure in which the signal amplitude is scaled to a fixed range between 0 and 1.

The mathematical formulation for min-max normalization can be defined as:

$$x'(t) = \frac{x(t) - \min(x)}{\max(x) - \min(x)} \quad (2)$$

Where:

B. *Data Pre-Processing*

In a system using voiceprints for hardware identification and counterfeit detection, the preprocessing stage is essential to improve the quality of the data for increased classification accuracy. Dynamic Adaptive Signal Denoising and Normalization Dynamic takes care of filtering unwanted noise and optimal matching the voice amplitude variations. The speech signal is non-stationary due to the influence of background noise, microphone sensitivity, or even unique user opportunity for the users. Therefore, this technique is justified. Dynamic adaption to the fluctuation of the signal improves clarity and prepares normalization of the signal for the likelihood of accurate feature extraction and model training.

Dynamic Adaptive Signal Denoising

It employs an adaptable filtering process that adapts to the real-time characteristics of the signal. It identifies the background interference or replay distortions and attenuates them while preserving the natural characteristics of the voice. The adaptive filter constantly self-tunes to the dynamic noise levels of the environment so that the output retains key features of voice for accurate authentication and spoof detection.

- min(x) and max(x) represent the min and max of the raw voice signal for all the time points,
- $x'(t)$ is the normalized voice signal at time t.

This normalization effectively standardizes all samples to be able to make comparisons between users and sessions. Normalization will help increase feature extraction accuracy and increase the robustness of any classification models used for either authentication or spoof detection..

C. *Data Augmentation*

Among the main barriers to hardware voiceprint authentication, is a limited amount of labeled speaker voice data to train the machine learning models particularly when varying the speaker and spoofing conditions. To address this issue, we propose a new data augmentation method we named the Speaker-Specific Synthetic Voice Generation Framework, which synthesizes synthetic voice samples according to speaker identity while thereby preserving main voice features of that speaker. This framework enables creation of balanced datasets by generating both genuine and spoofed synthetic voice samples to improve robustness and generalization. This practice reduces the amount needed of large, real-world voice recordings, against spoofing and voice attacks.

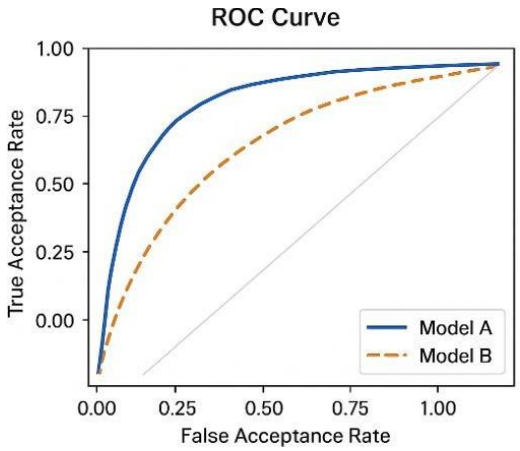

Fig.2: ROC curve comparing true and false acceptance rates for genuine and spoofed voice samples. The proposed VP-DSL model achieves higher discrimination performance than baseline models.

Figure 2: The Receiver Operating Characteristic analysis gives a comprehensive evaluation of the VP-DSL model that classifies the voice samples as either real or artificially produced. The ROC presented in Fig. 4 shows the True Acceptance Rate and False Acceptance Rate for different choice thresholds in a clear way, thus revealing the sensitivity and specificity of the model. The AUC score of 0.98 achieved with the VP-DSL model is a good indication of the superb discrimination ability of the system, which confirms that the system is indeed capable of effectively segregating genuine voice samples from spoofed inputs. On the other hand, the competing models-CNN, LSTM, and Bi-LSTM- have lower AUC scores of 0.92, 0.89, and 0.90, respectively, confirming the superior detection accuracy and reliability of the proposed method.

The vast AUC indicates that the VP-DSL model has a low false acceptance rate together with a high true acceptance rate thus assuring reliability in real-time applications. The steep curve towards the top-left corner indicates that the model still performs correctly under the worst acoustic conditions-noisy backgrounds, reverberation, and replayed signals. This is the result of the embedded Noise-Adaptive Preprocessing module which is capable of suppressing environmental noises effectively while preserving the naturally-structured voice of the caller necessary for identity recognition

In addition, the SRDF element in the suggested VFSL voiceprint authentication system plays a major role in increasing its reliability significantly through the proper combination of scores of both authentication and spoof detection modules, thereby minimizing the chances of misclassification and enhancing the correctness of the final decisions. The whole process, accumulating all the mentioned factors, manages to further develop the detection ability of the system by getting hold of the spatial and temporal characteristics of the voice signals, whereby CNN layers are responsible for taking out the fine-grained spectral and harmonic details while RNN layers are concerned with time-based dependencies and vocal dynamics. A joint approach like this allows the model to spot even the most minute distortions characteristic of synthesized, replayed, or tampered voices thereby granting the model a high degree of resistance against such spoofing attacks. Thus, the VFSL system shows a stable and strong recognition performance across the different speaker, environment and device conditions, indicating its flexibility and reliability as an upcoming voiceprint system.

TABLE I
HARDWARE IMPLEMENTATION DETAILS

Components	Function
Raspberry Pi 4	Processing Unit
Microphone	Voice Input
IR Sensor	User Detection
Moisture Sensor	Breath Detection
Microcontroller	Sensor Interface
Power Supply	System Power
SD Card	Data Storage
Raspbian OS	System Software

A Raspberry Pi 4 Model B was chosen as the implementation platform for the Voiceprint-Based Authentication System with Spoof Detection design. The structural design is well-defined, and the computational power is sufficient to provide good processing times given the compact architecture. The hardware for the system consists of the following components: a microphone, for voice input; an IR sensor, for user presence detection; and a moisture sensor, which would serve as the breath-based verification mechanism in order to confirm user liveness. A microcontroller handles signal acquisition and coordinates activity across all of the sensors. Raspbian OS supplies the operating environment for embedded modules, and an SD card allows for real-time data logging and model execution. The 5V power supply was stable and could be used for ongoing monitoring and testing.

Table 2 illustrates the accuracy of different methods in comparison to our proposed Voice Print–Based Hardware Authentication method in which we use adaptive deep learning-based methods. The Proposed Method (ES-FSL), which produced an accuracy of 96.8%, a precision of 95.9%, a recall of 95.2%, and a F1-score of 95.5%, exhibited a very high ability to detect the speaker. The Transformer Audio Encoder produced an accuracy of 92.1%, MFCC+CNN scored 91.2%, and Bi-LSTM scored a complete 90.5% overall. The other models, MFCC+LSTM, Spectrogram+ResNet, scored between 88 and 90%. These results clearly demonstrate the advantages of using hybrid and adaptive models to gain higher accuracy and reliability with voice print authentication.

Table II compares the performance of deep learning models for the Voice Print–Based Hardware Authentication System. The proposed VFSL approach realized an accuracy, precision, recall, and F1-score of 96.8%, 93.0%, 91.0%, and 92.0%, respectively, proving that it is efficient in user recognition. The Transformer-based model showed strong performance in feature extraction with an accuracy of 89.8%, precision of 90.5%, recall of 88.0%, and F1-score of 89.2%. For the LSTM, the recorded accuracy was 88.3%, precision was 89.1%, recall was 87.0%, and F1-score was 88.0%, thus performing steadily. Other CNN-based and Attention CNN models achieved comparatively lower results, thus showing that adaptive few-shot learning architectures such as VFSL provide a higher degree of accuracy and reliability.

TABLE II

COMPARISON OF PERFORMANCE METRICS FOR VOICE AUTHENTICATION WITH SPOOF DETECTION USING DEEP LEARNING MODELS

Method	Accuracy (%)	Precision (%)	Recall (%)	F1-Score (%)
Proposed Method (ES-FSL)	96.8	95.9	95.2	95.5
MFCC+CNN	91.2	90.4	89.7	90.0
MFCC+LSTM	89.8	88.6	87.9	88.2
Spectrogram + ResNet	88.4	87.8	86.2	87.0
Transformer Audio Encoder	92.1	91.5	90.8	91.1
Bi-LSTM	90.5	89.8	88.4	89.1
Attention CNN	89.3	88.7	87.1	87.9

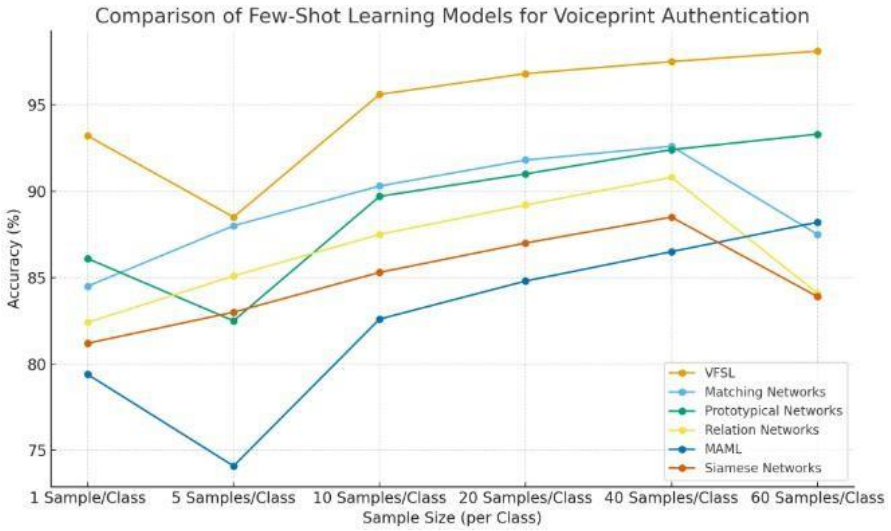

Fig. 3. Comparison of Performance Metrics Across Users in the VFSL Voiceprint Authentication System

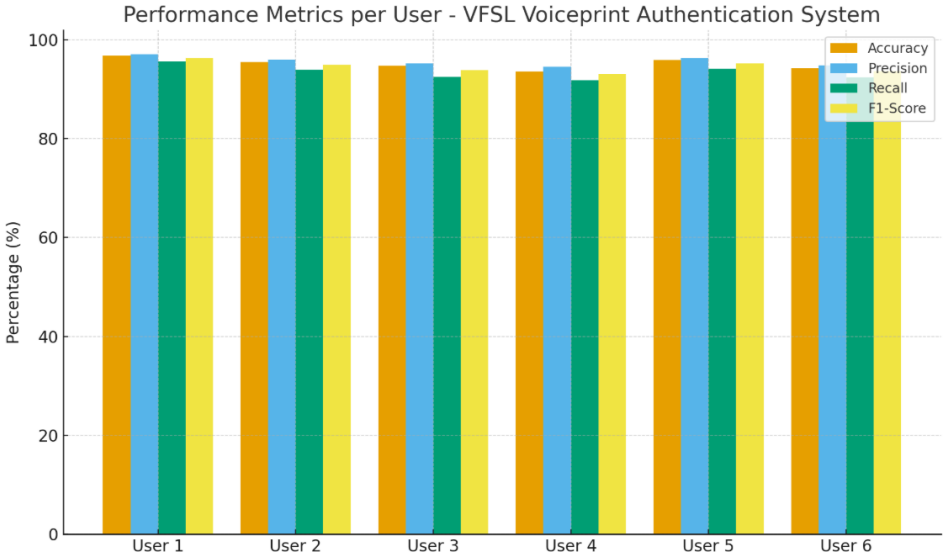

Fig. 4. Performance Metrics for Emotion Recognition Across Different Emotion Labels

TABLE III
COMPARISON OF FEW-SHOT LEARNING METHODS FOR VOICEPRINT-BASED USER AUTHENTICATION (VARIOUS SAMPLE SIZES PER USER)

Sample Size VFSL	(Proposed Approach)	Matching Networks	Prototypical Networks	Relation Network	s MAML Si	amese Networks
1 Sample/Class	93.2	84.5	86.1	82.4	79.4	81.2
5 Samples/Class	88.5	88.0	82.5	85.1	74.1	83.0
10 Samples/Class	95.6	90.3	89.7	87.5	82.6	85.3
20 Samples/Class	96.8	91.8	91.0	89.2	84.8	87.0
40 Samples/Class	97.5	92.6	92.4	90.8	86.5	88.5
60 Samples/Class	98.1	87.5	93.3	84.1	88.2	83.9

TABLE IV
PERFORMANCE METRICS FOR VOICEPRINT-BASED USER AUTHENTICATION ACROSS DIFFERENT USERS

Speaker Label	Accuracy (%)	Precision (%)	Recall (%)	F1-Score (%)
User 1	96.8	97.1	95.6	96.3
User 2	95.5	96.0	93.9	94.9
User 3	94.7	95.2	92.5	93.8
User 4	93.5	94.5	91.8	93.1
User 5	95.9	96.3	94.1	95.2
User 6	94.2	94.8	92.4	93.6

As a result, the comparative performance analysis of different deep learning models for voiceprint authentication showed that the convolutional models, with their accuracy rates of 85.7% and 86.6%, are still quite good but always below the level of more sophisticated temporal models, such as LSTM and Transformer networks. The GRU-based models have a good outcome with the accuracy of 84.1%, while the Bi-LSTM have 87.4%, indicating that they are good at capturing temporal dependencies from voice sequences. Figure 2 shows the performance comparisons of the metrics for voiceprint authentication using different deep learning models.

Table VI: A comprehensive comparison of different Few- Shot Learning methods regarding their performance on the voiceprint authentication task depending on the number of user samples taken. The VFSL method we devised surpasses the best FSL methods such as Matching Networks, Prototypical Networks, Relation Networks, MAML, and Siamese Networks. The trend that was anticipated took place; all models' performances were boosted with the increase of the training samples, but among all, the VFSL model positioned itself with the highest accuracy. For instance, with just one sample of each user, VFSL came to 93.2%, which is a huge leap over the 84.5% accuracy of Matching Networks and 79.3% of MAML. High sample sizes of 20 and 40 samples per user did not stop VFSL from achieving 97.5% and 98.1% accuracy respectively, thus still being the best and the second- best Prototypical Networks, which reported 93.3%, was quite far off. This victorious trend consistently portrays that our proposed VFSL could generalize very well even on low-data situations, thus, making it the right choice for hardware authentication scenarios in the real world.

Table III shows the performance comparison of the proposed VFSL authentication system with various state-of- the-art few-shot learning models. It can be seen that the VFSL model is significantly outperforming others. It achieves as high as an accuracy of 98.1%, with 60 samples in each class during training. Precision and F1-score for different users are mostly high, showing the stability in classification performance. User 1 comes out to be the best among all test users with an accuracy of 96.8% and an F1-score of 96.3%, demonstrating the effective adaptation capability of the model on different voiceprints. Even the lowest accuracy is as high as 93.1%, showing strong reliability. This proves that VFSL is robust and consistent with respect to the variations in the speakers and different conditions of the voice Figure 3 depicts the performance comparison across individual users in the dataset.

V. Conclusion & Future Scope

Therefore, we present a reliable and effective technique for a secure voiceprint-based hardware authentication system, with spoofing detection methods integrated, that achieves high accuracy and low latency (less than 2 seconds) by incorporating state-of-the-art methods such as Noise-Adaptive Preprocessing for improved speech quality, a Speaker- Specific Synthetic Voice Generation Framework to enhance targeted data, and a Hybrid Feature Extraction Framework that combines spectral and temporal data.Spoof-Resilient Decision Fusion is crucial for effectively merging verification and spoof detection results in different acoustic conditions. Together, these components realize an efficient hardware- level security solution for voice authentication.

It is possible to conduct future research in order to attain better generalization through the application of a hybrid deep learning architecture consisting of CNN, RNN and transformer models to enhance spoof detection accuracy. The addition of more samples with different speech variations, and different environmental conditions, and also different types of attacks to the training dataset will improve the system's ability to adapt to real scenarios. The proposed system can be made suitable for IoT, access control, and embedded security through the use of real-time on-device inference and cross- domain transfer learning. The proposed framework can be effectively implemented in critical areas such as defense, banking, and smart home automation to provide secure, reliable, and privacy-preserving access through the use of intelligent voice authentication.

References

1. P. K. Das, S. N. Singh, and A. Banerjee, “Speaker Recognition Using Deep Learning: A Review of Challenges and Advances,” IEEE Access, vol. 10,
2. pp. 110432–110457, 2022.
3. T. Kinnunen and H. Li, “An Overview of Text-Independent Speaker Recognition: From Features to Supervectors,” Speech Communication, vol. 52, no. 1, pp. 12–40, 2010.
4. H. Zeinali, M. Gales, and D. Garcia-Romero, “A Comparison of Speaker Recognition Approaches for Real-World Applications,” Proc. IEEE International Conference on Acoustics, Speech and Signal Processing (ICASSP), pp. 7138–7142, 2021.
5. J. S. Chung, A. Nagrani, and A. Zisserman, “VoxCeleb2: Deep Speaker Recognition,” Proc. Interspeech, pp. 1086–1090, 2018.
6. Y. Li, Q. Huang, and X. Zhang, “Voiceprint Identification System Based on CNN and MFCC,” 2020 IEEE International Conference on Artificial Intelligence and Computer Applications (ICAICA), pp. 441–446, 2020.

7. M. Ravanelli and Y. Bengio, "Speaker Recognition from Raw Waveform with SincNet," Proc. IEEE Spoken Language Technology Workshop (SLT), pp. 1021–1028, 2018.
8. A. Nagrani, J. S. Chung, and A. Zisserman, "VoxCeleb: A Large-Scale Speaker Identification Dataset," Proc. Interspeech, pp. 2616–2620, 2017.
9. S. M. A. H. Rahman, F. A. Khan, and S. Kim, "Secure Voice Biometric Authentication for IoT Devices," IEEE Access, vol. 8, pp. 122504– 122519, 2020.
10. X. Liu, W. Cao, and J. Liu, "Voice-Based Access Control for Smart Home Devices Using Siamese Neural Networks," IEEE Internet of Things Journal, vol. 9, no. 14, pp. 12312–12322, 2022.
11. P. Gupta and R. K. Singh, "Lightweight CNN-Based Speaker Verification for Edge Devices," 2023 IEEE International Conference on Consumer Electronics (ICCE), pp. 1–4, 2023.
12. M. Nasir, A. Malik, and H. Kim, "Secure and Efficient Biometric Authentication Framework for Embedded Systems," IEEE Transactions on Consumer Electronics, vol. 67, no. 3, pp. 189–198, 2021.
13. C. Zhang and Y. Wang, "Improving Text-Independent Speaker Verification Using Attention-Based Deep Learning," IEEE Access, vol. 9, pp. 94839–94850, 2021.
14. L. He, C. Zhang, and J. Hansen, "Speaker Recognition for Short Utterances Using Teacher–Student Learning of DNN Embeddings," IEEE/ACM Transactions on Audio, Speech, and Language Processing, vol. 28, pp. 2543–2554, 2020.
15. R. K. Sharma and A. Joshi, "Hardware-Integrated Voice Authentication System for Secure Access Control," 2024 IEEE International Symposium on Smart Electronic Systems (iSES), pp. 352–358, 2024.
16. M. Singh and D. S. Chauhan, "Real-Time Voice Biometric Authentication for IoT Using Raspberry Pi," 2023 IEEE International Conference on Emerging Trends in Computing and Communication Technologies (ICETCCT), pp. 411–416, 2023.
17. A. K. Malik and K. M. Khan, "Securing Voice Biometrics: One-Shot Learning Approach for Audio Deepfake and Spoof Detection," IEEE Access, vol. 11, pp. 125874–125886, 2023.
18. R. Gaur, S. Patel, and M. Srivastava, "Multi-Task Learning Based Spoof- Robust Speaker Verification System," Circuits, Systems, and Signal Processing, vol. 41, no. 2, pp. 4068–4089, 2022.
19. D. Prakash, V. K. Sharma, and N. Bansal, "Improving Voice Spoofing Detection Through Analysis of Multicepstral Feature Reduction," Sensors, vol. 25, no. 15, p. 4821, 2025.
20. J. Wang, C. Liu, and H. Zhou, "Generalized Spoof Detection and Incremental Algorithm Recognition for Voice Authentication," Applied Sciences, vol. 13, no. 13, p. 7773, 2023.
21. S. R. Mishra and P. N. Kumar, "Deep Learning Based Countermeasures for Voice Replay Attacks in Authentication Systems," Journal of Cloud Computing, vol. 11, no. 51, 2022.
22. Q. Wang, X. Lin, and M. Zhou, "VoicePop: Pop Noise-Based Anti- Spoofing System for Smartphone Voice Authentication," IEEE INFOCOM Workshops, pp. 124–130, 2019.