International Journal of
INTELLIGENT SYSTEMS AND APPLICATIONS IN ENGINEERING
ISSN:2147-6799 www.ijisae.org Original Research Paper

Enhancing UPI Security Using Deep Learning Based Voice Authentication Systems

Purav Nirav Doshi[1], Ganesh Khekare*[2], Uddhav Khetan[3]

Submitted: 28/01/2024 **Revised**: 06/03/2024 **Accepted**: 14/03/2024

Abstract: The identity of a person can be determined based on their voice, through the process of speaker identification. It can be used to improve the security of the United Payments Interface (UPI) framework. The process involves capturing and analyzing the acoustic features of the user's speech and comparing them to certain voice profiles that are stored in a database to find a match. Once a match is found the transaction can proceed smoothly. A model is built using the Fast Fourier Transform (FFT) and a 1-D CNN and it shows 98.46% accuracy on the data provided to it from the beginning and 98% on the validation data. This model is then compared with other existing models using different methods to obtain important attributes like Mel Spectrogram and MFCC. A process to integrate the model into the UPI ecosystem is successfully developed. This involves designing a protocol and developing an Application Programming Interface (API) for integration with UPI and Security Layering for additional threats. This paper addresses the current security concerns as well as paving the way for further research to improve security in the UPI ecosystem.

Keywords: *Speaker Identification, UPI, Deep Learning, Voice Authentication, Convolutional Neural Networks, Fast Fourier Transform*

1. Introduction

Speaker Identification uses features like pitch, tone, and other vocal traits to create a unique voice profile [1]. The voice profile can later be used to identify the speaker. As speech recognition is being used in upcoming technologies, speaker recognition will prove to be a natural authentication mechanism [2]. Speaker Identification utilizes voice biometrics and thus it gives better authenticity [3]. UPI was created by the National Payments Corporation of India (NPCI), to exchange funds between different bank accounts [4] and it is managed by the Reserve Bank of India (RBI). According to the NPCI, as of January 2024, 12.20 billion UPI transactions worth $222.17 billion have been processed. This increased the transaction value by 41.72 % compared to January 2023. The UPI ecosystem has the following advantages:

- UPI allows real-time transfer of funds which makes it faster and more efficient than the traditional methods like NEFT, RTGS, or IMPS.
- UPI services are available 24/7 so that the users can make transactions at any time, irrespective of the working hours of the bank.
- UPI uses a UPI ID instead of the IFSC Codes and bank account numbers. This helps to simplify the transaction process and allows the users to have more privacy.
- Users can link and manage multiple bank accounts from a single UPI app which makes it easier to track and manage finances.
- Most transactions that are done through UPI are free of cost or have minimum charges compared to other methods of payment. This makes it an affordable option for users.

There are certain problems in the security of the UPI framework that can be prevented with the help of innovative solutions. One of them is the risk of unauthorized access and fraudulent activities, where malicious actors exploit the weaknesses in the UPI system. For example, a person may install a malicious app by mistake and the hacker can launch an attack and transfer the funds out of the user's bank [5]. This can lead to security breaches and financial losses. Personal devices such as smartphones can also introduce security risks related to device theft, malware, and phishing attacks. Another challenge is the increasing demand for robust authentication mechanisms because traditional methods like PINs and passwords are subject to theft, sharing, or forgetfulness. Deep learning-based voice authentication systems can serve as a potential solution to these challenges [6]. Since they use unique vocal characteristics for secure user verification, security risks that are associated with identity theft and unauthorized access will decrease immensely. Voice authentication is user-friendly and can be easily integrated into the UPI framework. This will remove the problem of needing complex passwords and PINs because voice authentication can help transactions to be done securely with the help of a simple voice command [7]. Deep learning-based voice authentication systems can include different kinds of access, pitch, and tone and thus address the concerns of inclusivity

[1] *SCOPE, Vellore Institute of Technology, Vellore, 632014, India*
ORCID ID: 0009-0009-9640-044X
[2] *SCOPE, Vellore Institute of Technology, Vellore, 632014, India*
ORCID ID: 0000-0002-1687-4699
[3] *SCOPE, Vellore Institute of Technology, Vellore, 632014, India*
ORCID ID: 0009-0005-4370-2517
** Corresponding Author Email: khekare.123@gmail.com*

in UPI security. Thus, voice authentication can be seen as a user-friendly, efficient, and advanced method to improve security in the UPI ecosystem.

2. Literature Review

According to Balakrishnan M. [8], UPI is an important element in India's digitized ecosystem because it has significantly reduced barriers to digital payment. The user-friendly interface and real-time money-transfer ability have made it popular across many banking platforms. The paper focuses on UPI security and examines its multi-layered framework, which includes the interface layers, communication protocols, and data encryption standards. UPI transactions can also be strengthened by security practices like virtual payment addresses and two-factor authentication. The RBI and the NPCI oversee regular updates and audits in the UPI ecosystem. The paper also addresses common threats like phishing attacks, SIM cloning, and device theft and shows their impact on the confidence of consumers. At the same time, the paper emphasizes the effectiveness of biometric authentication and transaction encryption and helps to identify weaknesses in the current UPI security. Overall, this paper gives an overall understanding of the importance of UPI and gives a detailed explanation of its security infrastructure, current threats, and the efforts to strengthen this system.

According to Popa, D., and Simion, E. [9], biometrics utilizes unique physical and behavioral characteristics of an individual for identity verification and plays an important role in improving security. Integrating biometrics into multi-factor authentication systems provides a reliable means of verification and thereby reduces the risk of unauthorized access. The paper also focuses on integrating biometrics and cryptography to improve information security. By using biometric features, like fingerprints and facial patterns, along with cryptographic key generation and authentication, the security of digital systems can be greatly improved. The paper discusses applications like key release, key binding, and key generation and addresses such as secure storage and remembering different cryptographic keys. It also shows the potential of biometric cryptosystems in different domains such as authentication, access control, and secure communication. Additionally, the paper shows the challenges related to template protection, liveness detection, and privacy regulations and thus gives valuable insights for future research and development.

Khan, S. A., and Naaz, S. [10], have explored the advantages of biometric approaches in information security systems. The human finger vein uses unique patterns under the skin's surface, making it resilient against forgery and offering strong security without any direct client-system contact. It can even distinguish between identical twins and thus serves as a potential biometric device for identity verification. It has applications in medical sciences, law enforcement, and security. The iris identification method relies on the stable and distinctive texture of the iris. The iris is formed during fetal development and does not change. Each iris is unique, even among identical twins, and thus can be used to enhance security. Iris recognition systems are extremely accurate and recent advancements have also made them affordable and convenient. The animal body odor identification method uses the chemical composition of body odors as a suitable identification method. The unique primary Odors that are not affected by diet or environment can be used for differentiation.

Vassilev, V. et al. [11], have introduced an approach to secure digital banking and online payments through a two-factor, voice-controlled authentication. The research paper combines cloud technology and voice assistants to create a prototype to ensure secure authentication. A voice biometric system is implemented within a web application for two-factor authentication at the time of user logins. The prototype addresses security concerns in digital banking by combining cloud technology and devices with a voice assistant. The paper shows the need for added security in financial operations and shows voice biometrics as an optimal solution that can contribute to the development of secure authentication in digital banking.

In a comprehensive study conducted by Yerramreddy et al. [12], multiple models, like the Gaussian Mixture Model, CNN, LSTM, K-Nearest Neighbour, and Random Forest Classifier, were reviewed for identifying speakers using MFCC feature selection. GMM outperformed the LSTM model with a 98.68% accuracy. Prachi et al. [13] delved into advanced speaker recognition methods applying voice biometrics for security and authentication. Their method, utilizing MFCC on TIMIT and LibriSpeech datasets, highlighted the predominance of closed set over the open set approach, with CNN showing more accuracy compared to LSTM. These research studies offer invaluable insights into choosing efficient models for real-world applications in speaker identification and provide remarkable contributions to the field.

According to Barhoush, M. et al. [14], recognition models based on MFCC attribute selection are remarkable under the constraint of clear speech, but a noticeable decline occurs in the system's robustness when confronted with noisy environment and short voice snippets. To combat its shortcomings, the paper introduces scattered MFCC (SHMFCC), coupled with an innovative data manipulation tactic, resulting in outstanding performance across diverse datasets and varying environmental and noise conditions, regardless of the plethora of speakers involved.

Jaffino G. et al. [15] emphasize the effectiveness of MFCC and Dynamic MFCC (DMFCC) in representing the repetitive nature of speech signals. The study makes use of the Gaussian Mixture Model (GMM) and Bayesian

Classifier, showing that the GNN system outperforms the Bayesian classifier model. Singh, M. K. [16] introduces a CNN model to identify speakers which addresses issues like a long training time. By utilizing MFCC, PNCC, and GNCC for feature extraction, the proposed CNN architecture surpasses current models on the VoxCeleb1 database, reaching an accuracy of 96.90%. The model, including a Recurrent Neural Network, significantly betters the acceptance rates, displaying superior accuracy even in challenging environments and low signal-to-noise ratios when tested on TIMIT data. Table 1 illustrates the challenges in UPI Security and the corresponding solutions offered by Voice Authentication.

Table 1. Challenges in UPI Security and their Solutions by Voice Authentication

Challenges in UPI Security	Solutions offered by Voice Authentication
Phishing Attacks	Unique voiceprints: Harder to steal or replicate than traditional passwords
Sim Swap Fraud	Continuous Authentication: Continuously verifies the user's identity throughout the session to prevent unauthorized access even if the sim card gets swapped
Shoulder Surfing	Hands-free Authentication: It allows the user to authorize the transaction through his voice, instead of a PIN or password that can be observed by others
Password Theft and Reuse	Dynamic Tokens: Each session is authenticated with a unique voice token, thus eliminating the risks associated with static passwords
Man-in-the-middle Attacks	Encrypted Voice Templates: Deep Learning algorithms can create encrypted voice prints that secure the voice data even if it is intercepted during transmission
Replay Attacks	Liveness Detection: Advanced algorithms can differentiate between a live voice and a recorded/playback voice to prevent replay attacks
Lost/Stolen Phone Scenarios	Device Independent: As the voice authentication system depends on the user's voice, it remains secure even if the phone is lost or stolen.

Shihab, M. et al. [17] have demonstrated a speaker recognition model using a hybrid GRU [18] and CNN technique [19] for attribute selection to optimize the loss and select the optimal feature vector [20]. A feature extraction method [21], based on statistics, is later applied to select and combine the best features. This model is tested on the VoxCeleb dataset that consists of 6000 speakers with different voices.

3. Methodology

The research aims at building a robust voice authentication model using deep learning techniques and integrating it into the existing Unified Payments Interface (UPI) Ecosystem to increase security measures. Before developing the methodology several papers, that discussed different deep learning algorithms for speaker identification as well as other biometric methods used in payment systems, were examined. The insights gathered from these papers show the accuracy of the model and the need to add security measures in the real-time processing of financial transactions.

3.1. Model Development

The proposed solution involves the construction of a deep-learning model with the following steps:

1. Data Collection

• Gather unique voice samples from many individuals across different genders, age groups, and linguistic backgrounds to train the model based on variety.

• Background noises from different environments should also be added to improve its performance.

2. Data Preprocessing

• Adjust the sample rate of the audio file to be fixed at a desired rate to maintain uniformity in the dataset.

• The Fast Fourier Transformation is used to transform the audio waveforms into the frequency domain.

• The data is then shuffled and split into different sets for enhanced performance.

3. Model Architecture

• CNN is built to classify the different voice samples. The CNN layers will help to find hidden patterns and features in the audio samples and will thereby provide a more accurate classification.

• The model will also have mechanisms to address overfitting such as dropout layers and data augmentation techniques.

3.2. Integrating into the UPI system

The model is now integrated into the UPI ecosystem by following these steps:

1. Protocol Design

• A secure voice authentication protocol is defined to integrate with the current UPI architecture.

• The system will ask the users for a voice command before the initiation of any UPI transaction which will work along with the other existing security measures like the UPI PIN.

2. API Development

• An Application Program Interface (API) is developed that bridges the voice authentication model with the current UPI framework.

• The API should use secure methods for data transmission, and it should comply with the security standards laid down by the NCPI.

3. Security Layering

• The voice authentication system will be added on top of the current security measures as an additional factor in the authentication process.

• Secure session handling and timeout mechanisms are employed to protect the authentication process from session hijacking and replay attacks.

The above methodology aims to enhance the current financial security with a robust voice authentication system using a planned sequence of development and integration. It is vital to address the security vulnerabilities associated with the UPI ecosystem in the current day and age. To enhance UPI security using a deep-learning-based voice authentication system, it is crucial to conduct in-depth research on the latest advancements in voice detection and authentication to ensure that the system remains secure against potential security threats. Rigorous testing, validation, and updates to the voice authentication model are also important for maintaining UPI security. By addressing these aspects, building a deep learning-based voice authentication model in the UPI framework can help safeguard against existing threats and vulnerabilities. Figure 1 summarizes the key steps of the proposition.

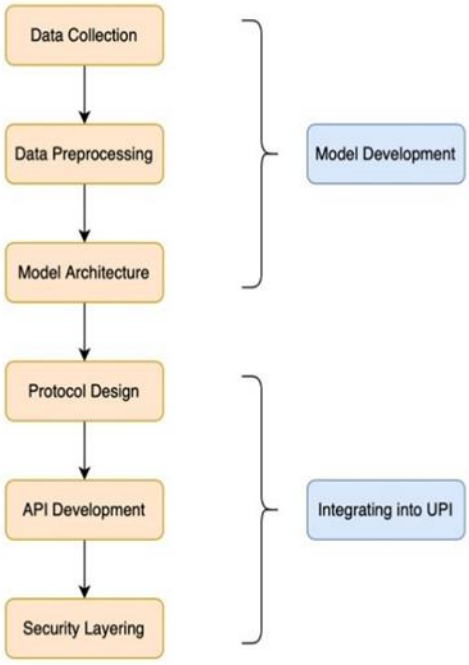

Fig. 1. Flowchart of the Methodology

3.3. System Composition

CNN is built to classify the different audio samples. The main components of the model are given in Figure 2 and are as follows:

1. Residual Block Function

The residual blocks are built to enhance the training of CNN for audio classification. The residual block consists of 1-D convolutional steps with a “skip connectivity”, which mitigates the exploding gradient issue and helps in training deeper layers. The local data block function accepts a tensor, applies convolutions with a specified number of filters and a ReLU activation function, and applies a series of shortcut connections to ease the flow of gradients during backpropagation. The residual connections help to simplify optimization, thereby helping the model to converge faster during training.

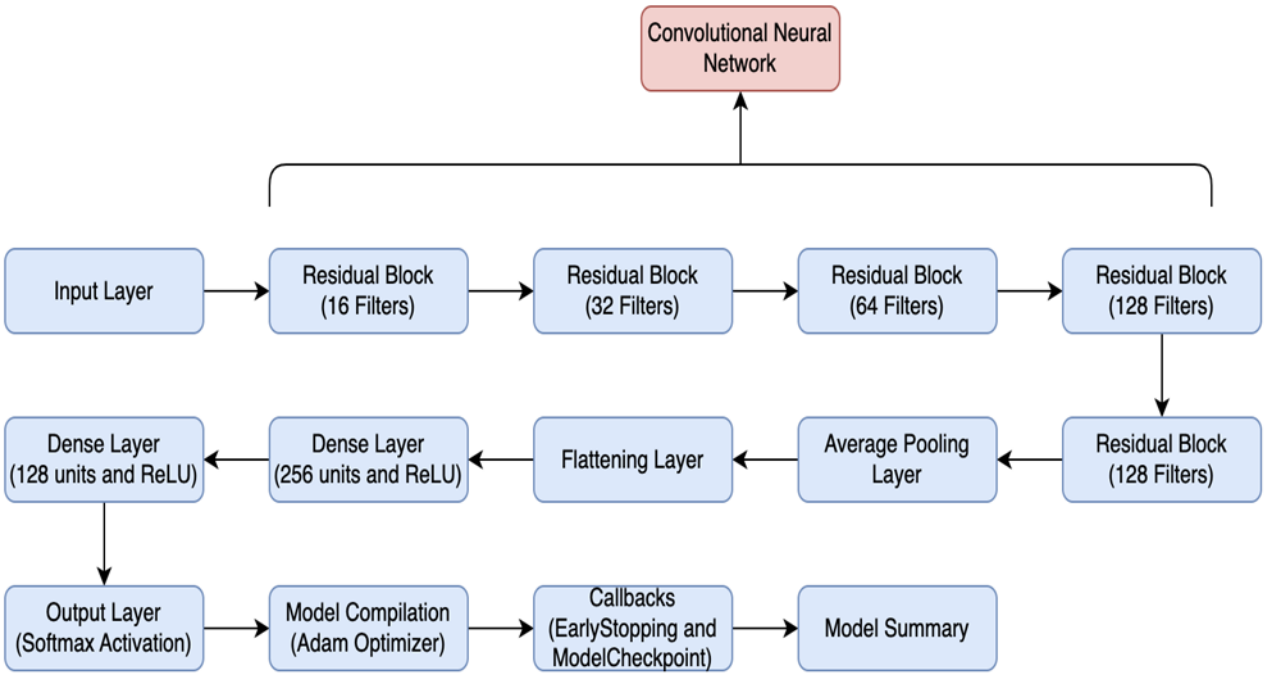

Fig. 2. Flowchart of the Model Architecture

2. Building the Model

The overall architecture is constructed by stacking multiple residual blocks. The model starts with an input layer, followed by four residual blocks whose filter sizes keep on increasing and are of sizes (16,32,64,128) and this is followed by a final block with 128 filters repeated two times. Average Pooling and Flattening layers are applied to reduce the spatial dimensions. Dense levels having 256 and 128 units, respectively, with ReLU function, are added to further process features that have been extracted from convolutional layers. An output layer with a SoftMax function is added at the end. This helps the model to find hidden attributes within audio data with ease and to provide a more accurate classification.

3. Model Compilation

The model is compiled using Adaptive Moment Estimation (ADAM). Sparse category-wise cross-entropy is chosen as the loss function and the authenticity metric determines the accuracy of the model.

4. Callbacks

Two callbacks are included to train the model efficiently. Early Stopping callback monitors the authenticity and stops the training if there is no improvement in the accuracy after a specified number of epochs, thus preventing the model from overfitting and thereby optimizing the training time of the model. Model Checkpoint callback saves the model that has the highest accuracy and thus ensures that the final model is the one that displays the highest accuracy on the data.

5. Model Summary

It provides an overview of the architecture of the model and gives details about layers, output, and number of parameters. This helps to understand the model's depth and the attributes involved. The system is now ready for training on the audio dataset.

3.4. Model Training

The model is trained on the available data for a certain number of epochs. After this, the testing data shows the system's performance on unknown voice samples. The two callbacks are used to avoid overfitting and to save the best-performing model. The training history, including metrics such as training accuracy, validation accuracy, and loss are stored separately. This information can be used for performance analysis and visualization and to tune the model even more.

3.5. Evaluating the Model and Predicting Values

The usefulness of the model depends on the following:

1. Loss Function: The sparse category cross-entropy loss is chosen, and it outperforms both MSE and RMSE because it punishes errors more heavily than MSE and RMSE. The cross-entropy loss is calculated by Equation 1 given below:

$$\text{Cross} - \text{Entropy Loss} = -\sum_{i=1}^{n} t(i).log(p(i)) \quad (1)$$

where:

pi is the softmax probability of the i^{th} class.

2. Metric: Accuracy is the ratio of accurate observations to the total number of observations. When the accuracy is multiplied by 100, it gives the accuracy expressed as a percentage (as shown in Equation 2) which shows the model's performance in speaker identification.

$$Accuracy = \frac{Number\ of\ Correct\ Observations}{Total\ Number\ of\ Observations} \times 100 \quad (2)$$

3. Optimizer: The Adam optimizer has been used, which uses a combination of AdaGrad and RMSProp to improve the model's accuracy and is best suited to the problem definition taken.

3.6. Integrating with UPI

3.6.1. Design for Enhancing UPI security through voice authentication.

The designed UPI protocol aims to easily integrate voice authentication systems as an additional security layer that will complement the existing UPI mechanism effectively. The enhanced protocol uses deep learning techniques such as Fast Fourier Transforms (FFT) and 1-D Convolutional Neural Networks (CNN) to ensure efficient voice authentication and build trust among the users.

a. User Enrolment and Authentication

The process begins with the enrolment of a user's voice biometric data. After registration with an existing UPI service provider, the users are requested to provide voice samples for their authentication. These voice samples are processed through a Fast Fourier Transform to extract the audio attributes and input them to the 1-D CNN.

b. Initiating transactions with Voice Authentication

When a user wants to initiate a UPI transaction, apart from the existing UPI pin, they are also provided with the option to initiate with their voice. If they select the option for voice authentication, the UPI application will prompt the user to speak a predetermined sentence that gets captured and processed by the voice authentication model. Then 1-D CNN analyses the voice input and compares it against the enrolled voice biometric of the user. If it is a match, the identity of the user is verified, and the transaction proceeds to the next stage.

c. Integration with UPI transaction validation

Once the voice authentication process has been completed successfully, the voice authentication module communicates with the UPI server providing the authentication status and the user identification. Consequently, the UPI transaction validation involving the transaction amount, recipient details, and other data is processed.

d. Fallback Mechanisms and User Experience

If there is an unsuccessful voice authentication, the UPI protocol reverts to the PIN authentication method. This ensures a smooth transaction for the user. The user is also provided with clear prompts to guide them through the authentication process which enhances the user experience and accessibility.

e. Security Measures and Protection of Data

Extensive security measures have been taken to safeguard the user's voice biometric data and to ensure secure payment transactions. The voice data captured for authentication purposes is encrypted using the industry-standard cryptographic protocols during transmission and storage. Additional security measures such as the implementation of session tokens and dynamic timeout mechanisms are also added to protect the user's data from session hijacking and replay attacks.

f. API Architecture and Data Transmission Security

The development of a robust API will serve as the interface between the voice authentication model and the existing UPI architecture. The API architecture will strictly adhere to security standards and will implement SSL/TLS encryption for secure data transmission, thereby following the information given by the NCPI to ensure compliance with the rules and integrity of the user's data. Figure 3 given below illustrates the key steps under Protocol Design.

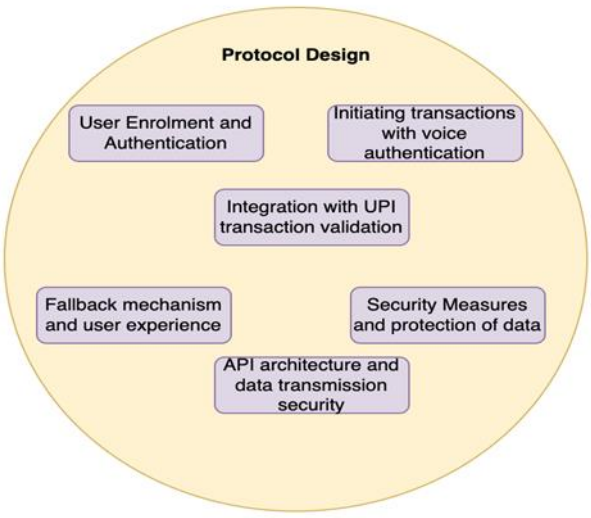

Fig. 3. Steps under Protocol Design

3.6.2. Developing an API to integrate the model and UPI framework.

The implementation of the API for integrating the voice authentication module with the existing UPI framework has many important steps. It involves the design of a robust and

secure communication interface that bridges the voice authentication module with the UPI servers adhering to strict security measures and regulations.

a. Architecture and Endpoint Design

The API is implemented following a well-defined architecture that consists of a set of endpoints and the corresponding request-response patterns. These endpoints have been carefully designed to handle different aspects of the authentication and transaction validation process and they ensure standardized and efficient data exchange in line with the RESTful principles. Each endpoint has been structured carefully, specifying the input parameters, the expected response, and the business logic that oversees the interaction between the voice authentication module and the UPI servers.

b. Secure Data Transmission

The implementation of the API prioritizes secure data transmission by using SSL/TLS encryption for HTTP requests. This encryption protocol ensures that sensitive data, such as voice biometric samples, are confidential during transactions. Additionally, cryptographic protocols and secure hashing algorithms such as SHA-256 are implemented to validate the integrity of the data that has been transmitted, which helps to safeguard the data against unauthorized access and data tampering during its transmission and reception.

c. Conforming to Regulatory Standards

The implementation aligns with the security standards and guidelines that have been laid down by the NPCI and ensures whole data transmission and authentication processes are in full compliance with the regulatory requirements. The API implementation adheres to the security measures to meet the criteria for safe and secure data authentication within the UPI ecosystem.

d. Scalability and Maintenance

Scalability and maintenance are an important part of the UPI protocol as they ensure that the system can efficiently handle a growing user database and increasing volumes of transactions. The API architecture is designed to accommodate scalability, with efficient load balancing and failover strategies to avoid saturation of performance. The maintenance framework also includes version control, continuous monitoring, and regular updates to strengthen security and adjust to the current security trends.

e. User Authentication and Authorization

The implementation of the API facilitates secure user authentication and authorization processes and thus enables seamless integration with the UPI framework to initiate and validate transactions. It provides secure mechanisms for verifying the user's identity through voice biometric authentication, validating transaction requests, and communicating the authentication details and transaction details to the UPI servers. This ensures that the voice authentication process is robust, and efficient and aligns with the best practices in the industry for secure authorization and exchange of data.

f. Error handling and logging

Error handling mechanisms are implemented within the API to ensure that the integration process can function smoothly. The implementation consists of tracking the various stages of the authentication and transaction validation process and enabling efficient debugging and diagnostics in case of any unexpected errors or exceptions. This ensures that any potential threat is rapidly identified and addressed, thereby increasing the stability of the API integration with the UPI ecosystem. Figure 4 given below illustrates the key steps under API Development.

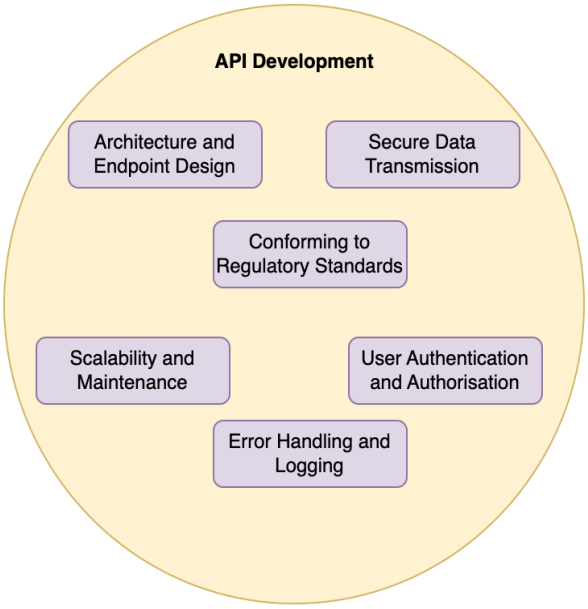

Fig. 4. Steps under API Development

3.6.3. Security Layering against Potential Threats

When designing the security framework for the UPI protocol and its associated API, many components contribute to the protection against threats, vulnerabilities, and attacks. The protection also includes secure session handling, timeout mechanisms, replay attack prevention, and risk assessments. A detailed explanation of these aspects has been provided, emphasizing the important role each component plays in enhancing UPI security.

a. Secure Session Handling

Secure Session Handling is the foundation for safeguarding user interaction within the UPI ecosystem. It includes the implementation of secure session management techniques, providing for the creation, maintenance, and termination of

user sessions, and ensuring their confidentiality and integrity. Using Transport Layer Security (TLS) protocols such as HTTPS, along with secure cookies and tokens, facilitates secure session initiation for user interaction and effectively prevents session hijacking and eavesdropping.

b. Timeout Mechanisms for additional layering

Timeout Mechanisms serve as an additional layer of defense against unauthorized access and misuse of user sessions. When we utilize the timeout configurations, the sessions become invalid and are terminated after a period of inactivity. This reduces the chances for exploitation of the user's sessions. The implementation of the session timeout mechanism complies with industry standards, such as OWASP's session management best practices, and thus strengthens the UPI framework's resistance against unauthorized access and session-based attacks.

c. Protection against Replay Attacks

By using temporal and cryptographic measures, the UPI ecosystem is safe from replay attacks. Implementing nonce-based protocols within the API layer, along with timestamp validation and verification, prevents the risk of replay attacks by ensuring that each transaction request is identified uniquely and is temporarily constrained. The integration of cryptographic techniques, such as digital signatures and message authentication codes, also strengthens the ability of the UPI framework to detect and discard replayed transaction requests and thus safeguards against unauthorized re-transmission of intercepted data.

d. Assessing Risks and Constant Authenticity Audits

Ongoing risk assessments and security audits are an important part of the UPI security framework because they facilitate the continuous evaluation and enhancement of the security posture. Using techniques such as threat modeling, vulnerability assessments, and penetration testing, we can ensure that the UPI ecosystem identifies and prevents emerging risks and vulnerabilities effectively. Additionally, regular security audits that are aligned with industry standards like ISO/IEC 27001 or NIST SP 800-53 can be used to validate the robustness of the security controls and thus promote continuous security improvement within the UPI ecosystem. Figure 5 given below illustrates the key steps under Security Layering.

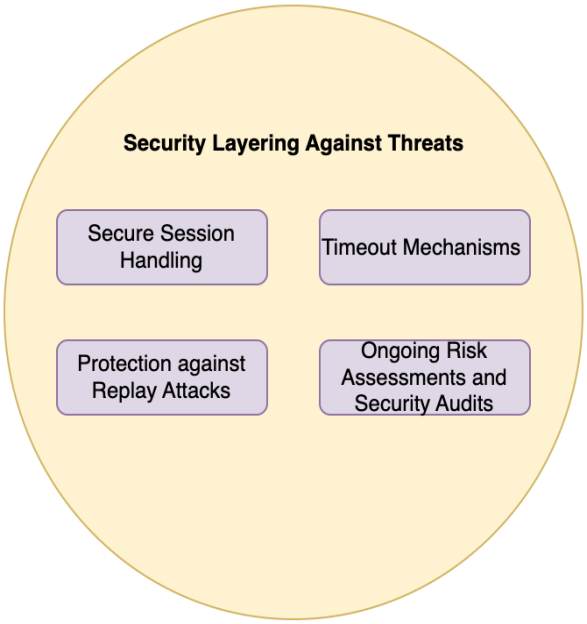

Fig. 5. Steps under Security Layering against additional threats

4. Result and Discussion

A. Accuracy of the model

A 98.46 % accuracy has been achieved on the training set and a 98 % accuracy has been achieved on the validation data. The graphs representing these accuracies corresponding to the number of epochs are given below in Figure 6 and Figure 7.

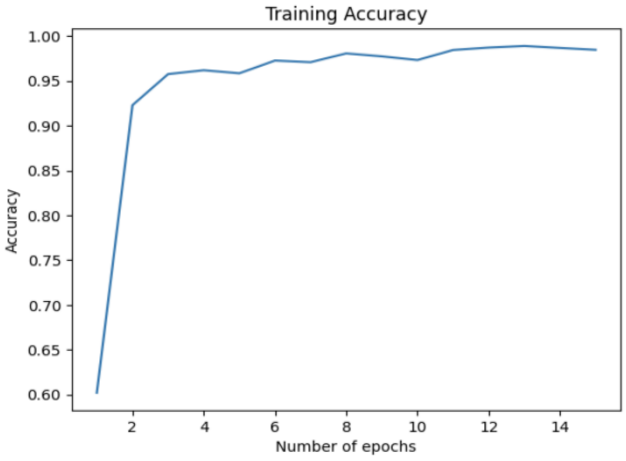

Fig. 6. Model's accuracy on the training set

Fig. 7. Model's accuracy on the testing set

B. Loss Function

A loss of 3.72 % has been achieved on the training data and a loss of 9.14 % has been achieved on the validation data. The graphs representing the loss function corresponding to the number of epochs are given below in Figure 8 and Figure 9.

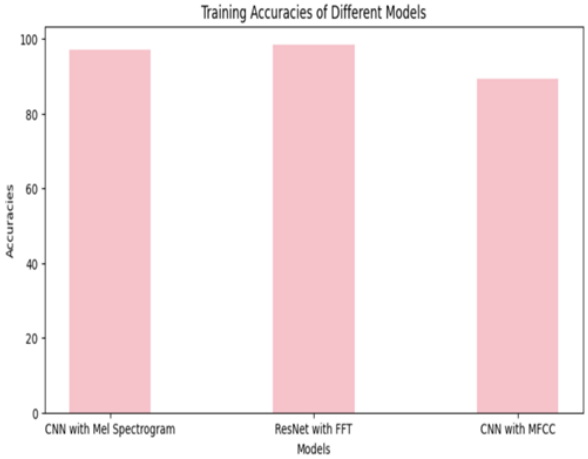

Fig. 8. Loss Due to Training Dataset

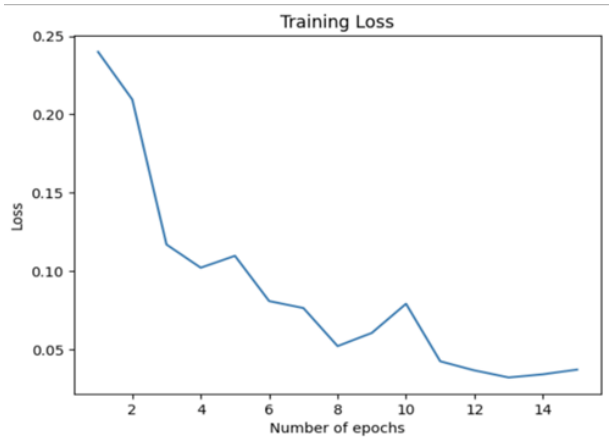

Fig. 9. Validation Loss

C. Comparison with Existing Models

The existing model has been compared with two other models using different feature extraction techniques. Their accuracy on the training and testing set has been given in Table 2 below. Figures 10 and 11 represent the training and validation accuracies of the different models via a bar graph.

Table 2. Comparative Study

Model Type	Feature Extraction Technique	Training Set Accuracy	Validation Set Accuracy
2-D CNN	Mel Spectrogram	97.03 %	93.20 %
1-D CNN ResNet	Fast Fourier Transform (FFT)	98.46 %	98 %
2-D CNN	Mel-Frequency Cepstral Coefficients (MFCC)	89.24 %	90.94 %

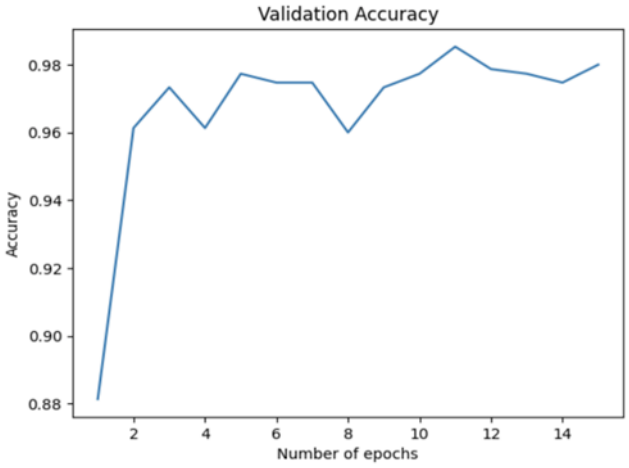

Fig. 10. Accuracy of the different models on the training data.

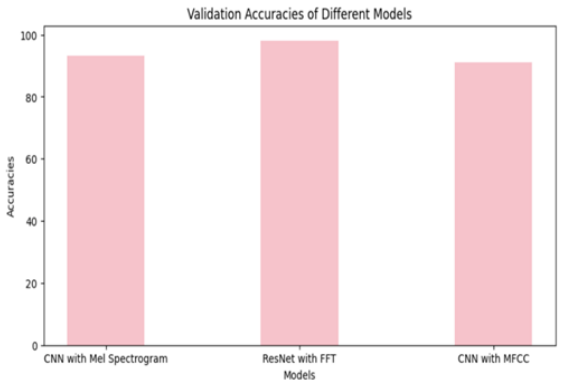

Fig. 11. Accuracy of the different models on the validation set.

The model has demonstrated an impressive performance, which shows its effectiveness in identifying voices accurately across diverse scenarios. The high accuracy achieved demonstrates the reliability of the model and shows its potential for real-world applications. The proposed integration with UPI adds an extra layer of security, thus demonstrating advancements in security and user experience in financial transactions. The model's precision can therefore contribute to a seamless and trustworthy transaction environment. It's paramount to authenticate the drawbacks, like variations in voice quality and external environmental factors, that might impact the model's performance. Continuous refinement is advised to overcome these challenges so that the model can adjust to the constantly changing real-world conditions.

5. Conclusion

This research presents a robust approach to improving the security of the UPI framework by using a deep-learning-based voice authentication system. The model built shows an impressive accuracy of 98% in analyzing the person based on audio biometrics and the proposed integration of the model into the UPI ecosystem improves the security as well as the user experience. The voice authentication system provides a successful defense mechanism against unauthorized access and identity fraud in financial transactions. As more and more people are shifting towards UPI, the article offers an advanced solution to safeguard digital transactions against malicious users. Even though this research shows promising results, the authors acknowledge the rise of challenges in cybersecurity and suggest adding more layers of security as a future scope. In conclusion, this paper contributes to the ongoing research and improvement of a safe and secure digital ecosystem.

Author contributions

Purav Nirav Doshi: Conceptualization, Methodology, Software, Field study, Data curation.

Ganesh Khekare: Writing-Original draft preparation, Validation, Writing-Reviewing and Editing.

Uddhav Khetan: Visualization, Investigation, Field study.

Conflicts of interest

The authors declare no conflicts of interest.

References

[1] Singh, Nilu. Automatic speaker recognition: current approaches and progress in the last six decades. Global Journal of Enterprise Information System 9.3, 2017, pp. 45-52.

[2] Shayamunda, C., Ramotsoela, T. D., and Hancke, G. P. Biometric Authentication System for Industrial Applications using Speaker Recognition. The 46th Annual Conference of the IEEE Industrial Electronics Society, Industrial Electronics Society (IECON), 2020, pp. 4459–4464.

[3] Mohd Hanifa, R., Isa, K., and Mohamad, S. A review on speaker recognition: Technology and challenges. Computers and Electrical Engineering, 2021.

[4] Lakshmi, K. K., Gupta, H., and Ranjan, J. UPI Based Mobile Banking Applications – Security Analysis and Enhancements. 2019 Amity International Conference on Artificial Intelligence (AICAI), 2019, pp. 1– 6.

[5] Madwanna, Y., Khadse, M., and Chandavarkar, B. R. Security Issues of Unified Payments Interface and Challenges: Case Study. ICSCCC 2021 - International Conference on Secure Cyber Computing and Communications, 2021, pp. 150-154.

[6] Thimmaraja Yadava, G., Nagaraja, B.G. and Raghudathesh, G.P. Real-Time Automatic Continuous Speech Recognition System for Kannada Language/Dialects. Wireless Pers Commun, 2024.

[7] Gupta, P., Patil, H.A. and Guido, R.C. Vulnerability issues in Automatic Speaker Verification (ASV) systems. J AUDIO SPEECH MUSIC PROC. 2024, 10, 2024.

[8] Balakrishnan, M. The Unified Payment Interface and the growth of digital payments in India: An analysis. Journal of Payments Strategy & Systems, 17(3), 2023, pp. 250–270.

[9] Popa, D., & Simion, E. Enhancing security by combining biometrics and cryptography. 2017 9th International Conference on Electronics, Computers and Artificial Intelligence (ECAI), Electronics, Computers and Artificial Intelligence (ECAI), 2017, pp. 1–7.

[10] Khan, S. A., & Naaz, S. Comparative Analysis of Finger Vein, Iris and Human Body Odor as Biometric Approach in Cyber Security System. 2020 2nd International Conference on Innovative Mechanisms for Industry Applications (ICIMIA), 2020, pp. 525–530.

[11] Vassilev, V. et al. Two-factor authentication for voice assistance in digital banking using public cloud services. Proceedings of the Confluence 2020 - 10th International Conference on Cloud Computing, Data Science and Engineering, 2020, pp. 404-409.

[12] Yerramreddy, D. R. et al. Speaker Identification Using MFCC Feature Extraction: A Comparative Study Using GMM, CNN, RNN, KNN and Random Forest Classifier. 2023 Second International Conference on Trends in Electrical, Electronics, and Computer Engineering (TEECCON), 2023, pp. 287–292.

[13] Prachi, N. N. et al. Deep Learning Based Speaker Recognition System with CNN and LSTM Techniques. 2022 Interdisciplinary Research in Technology and Management (IRTM), Technology and Management (IRTM), 2022 Interdisciplinary Research In, 2022, pp. 1–6.

[14] Barhoush, M., Hallawa, A., and Schmeink, A. Robust Automatic Speaker Identification System Using Shuffled MFCC Features. 2021 IEEE International Conference on Machine Learning and Applied Network Technologies (ICMLANT), 2021, pp.1–6.

[15] Jaffino, G., Raman, R., and Jose, J. P. Improved Speaker Identification System Based on MFCC and DMFCC Feature Extraction Technique. 2021 Fourth International Conference on Electrical, Computer and Communication Technologies (ICECCT), 2021, pp. 1– 5.

[16] Singh, M. K. A text independent speaker identification system using ANN, RNN, and CNN classification technique. Multimedia Tools and Applications: An International Journal, 2023, pp. 1–13.

[17] Shihab, M. S. H. et al. A Hybrid GRU-CNN Feature Extraction Technique for Speaker Identification. 2020 23rd International Conference on Computer and Information Technology (ICCIT), 2020, pp. 1–6.

[18] Khekare, Ganesh, Pushpneel Verma, and Seema Raut. "The Smart Accident Predictor System using Internet of Things." Cloud IoT. Chapman and Hall/CRC, 2022. 163-175.

[19] Ganesh Khekare, K. Pavan Kumar, Kundeti Naga Prasanthi, Sanjiv Rao Godla, Venubabu Rachapudi, Mohammed Saleh Al Ansari and Yousef A. Baker El-Ebiary, “Optimizing Network Security and Performance Through the Integration of Hybrid GAN-RNN Models in SDN-based Access Control and Traffic Engineering” International Journal of Advanced Computer Science and Applications(IJACSA), 14(12), 2023. http://dx.doi.org/10.14569/IJACSA.2023.0141262

[20] G. Khekare, S. Gambhir, I. S. Abdulrahman, C. M. S. Kumar and V. Tripathi, "D2D Network: Implementation of Blockchain Based Equitable Cognitive Resource Sharing System," 2023 3rd International Conference on Advance Computing and Innovative Technologies in Engineering (ICACITE), Greater Noida, India, 2023, pp. 908-912, doi: 10.1109/ICACITE57410.2023.10182834.

[21] G. Khekare and Midhunchakkravarthy, "Smart Image Recognition System for The Visually Impaired People," 2023 International Conference on Energy, Materials and Communication Engineering (ICEMCE), Madurai, India, 2023, pp. 1-6, doi: 10.1109/ICEMCE57940.2023.10434130.