Voice Based Secure Transaction System With Deep Fake Detection

Ms.Yogasree M[1], Ms.S.Sabaria[2]
[1, 2] Dept of Computer Applications
[1, 2] B.S. Abdur Rahman Crescent Institute Of Science And Technology

Abstract- *The development and growth of digital payment systems have created a need for secure authentication mechanisms. The use of passwords, PINs and one-time passwords is no longer secure against cyber attacks and identity theft. The paper suggests a Voice Based Secure Transaction System with Deep fake Detection. The suggested system will use machine learning algorithms for speaker verification and deep fake detection. The MFCC method will be used for feature extraction. The user's voice command will be converted to text by a speech recognition system. The text will be compared to the saved voice samples for verification. The deep fake detection module will check whether the voice is genuine or not. If the voice is genuine, the transaction will proceed otherwise, the transaction will be denied. The suggested system is efficient and secure for digital payment systems.*

Keywords: Voice Authentication, Deep fake Detection, Secure Transactions, Voice Biometrics, MFCC, Machine Learning, Speech Recognition, Digital Payment Security.

I. INTRODUCTION

With the rapid advancement of digital payment systems and online banking, the use of electronic transactions is increasing significantly. With the rise of mobile banking, digital wallets and online payment systems, the security of the system with robust authentication is a major concern. Although traditional methods of authentication, including passwords, PINs and OTPs are used for electronic transactions, they are often threatened by cybercrime, phishing, identity theft and unauthorized access. Hence, the need for more robust and efficient methods of authentication is felt. Voice-based authentication is a revolutionary method for the security of the system, as users can access the system with the aid of voice commands. Voice biometrics identify the unique features of the user's voice, thereby identifying the user without the use of complex systems. However, with the latest advancements in artificial intelligence, the security of the system is threatened by voice spoofing, which involves the use of deep fake voices for unauthorized access. In order to address these challenges, a Voice Based Secure Transaction System with Deep fake Detection is proposed. In the proposed system, speech recognition technology is used for converting voice commands into text format. In addition, voice features are extracted using Mel Frequency Cepstral Coefficients (MFCC) for speaker verification. Moreover, a deep fake detection module is integrated into the proposed voice-based digital payment system for detecting deep fake voice signals. Overall, the proposed voice-based digital payment system improves the security and usability of digital payment systems.

II. LITERATURE REVIEW

The rapid growth of Artificial Intelligence in financial technology has enhanced the security and efficiency of digital payment systems. Conventional digital payment systems employ password, PIN and OTP-based authentication. These conventional digital payment systems are prone to cyber attacks, phishing and identity theft. In order to avoid these challenges, biometric-based authentication systems, such as voice recognition have been employed to ensure the security and reliability of digital payment systems. In voice recognition, speech recognition and voice feature extraction techniques are employed to authenticate and recognize users. With the help of machine learning and deep learning algorithms, voice recognition systems have become more reliable and accurate. However, recent advancements in deep fake and voice synthesis technologies have posed a significant security risk to digital payment systems. In deep fake technology, attackers can easily create a fake voice that can be used to trick a conventional voice recognition system. Conventional voice recognition systems are not able to recognize fake voices. Therefore, it is necessary to employ deep fake detection technology along with voice recognition technology to ensure security and reliability. The proposed system will help to resolve these challenges and ensure the security and reliability of digital payment system.

III. PROBLEM DEFINITION

The conventional digital transaction system mainly uses authentication techniques like passwords, PINs and OTPs to authenticate users in the process of financial transactions. However, these authentication techniques are exposed to

various cyber attacks like phishing, identity theft, hacking and unauthorized access. The conventional digital transaction system has become a major challenge with the development of digital banking and online payment systems. The voice authentication system has become an effective and convenient system for user authentication in digital transactions. However, this system has become exposed to various security attacks with the development of deep fake and voice spoofing techniques. The attackers can create new voices to impersonate genuine users and bypass the conventional voice authentication system. In addition, factors like noise, voice changes and recording device changes are also creating difficulties in speaker verification systems. Therefore, there is a need to develop an advanced system that integrates voice biometric authentication and deep fake detection techniques to ensure secure, reliable and efficient voice authentication in financial transactions.

OBJECTIVE

The primary aim of the proposed research is the creation of a secure transaction system through vocal recognition. The aim of the proposed research are as follows:

- The utilization of voice recognition technology for securing transactions.
- Detection of deepfake voice using machine learning techniques.
- Extraction of voice features such as pitch and frequency for user identification
- Real-time authentication process for transactions.
- Prevention of fraud and any other illegal activities during payments.

IV. PROPOSED SYSTEM

The proposed system is an AI-based Voice Based Secure Transaction System with Deep fake Detection. The system is intended for secure voice-based transactions. In this system, a user can make a transaction by giving a voice command through the application. The command is captured and processed by speech recognition technology. The captured command is converted to text format. The captured voice command is subjected to preprocessing techniques such as noise reduction, normalization and segmentation. The quality of the captured command is enhanced by these techniques. The quality-enhanced command is subjected to feature extraction techniques such as Mel Frequency Cepstral Coefficients (MFCC) to identify the key features of the command. The key features identified from the command are compared with the available command in the database for speaker verification using machine learning algorithms. The command is also subjected to deep fake detection techniques to identify whether the command is genuine or deep fake. If the command is identified as genuine and no deep fake is detected, the transaction is processed securely using a payment gateway such as a digital wallet or UPI. The transaction is denied if the command is identified as deep fake. The proposed system is intended for secure and efficient voice-based digital transactions.

V. SYSTEM ARCHITECTURE

The architecture proposed for the system is a modular pipeline architecture for secure voice-based financial transactions with deep fake detection. The architecture has several main stages in the system, including voice input acquisition, audio preprocessing, feature extraction, deep fake detection, speaker authentication and transaction processing with decision output. The architecture has been designed in a modular manner to ensure secure authentication, voice verification, and to prevent fraudulent transactions with synthetic voice inputs.

FIG 1. ARCHITECTURE DIAGRAM

The architecture begins with the user giving a voice command to the system through a microphone, which acts as the input to the system. The voice input is then sent to an audio preprocessing stage, in which various operations are performed on the voice signal to enhance its quality. The voice signal is then sent to the feature extraction stage, in which Mel Frequency Cepstral Coefficients (MFCC) are used to extract the unique voice characteristics of the user. The extracted MFCC features are used for two parallel operations. Firstly, the features are passed through a deep fake detection module, where it is analyzed whether the voice is real or a deep fake created by deep fake technology or voice cloning. Secondly, the features are passed through a voice authentication module,

where the extracted features are compared with the existing voice samples stored in the system's database to authenticate the identity of the user. The results obtained from the above two operations are passed through a transaction processing system, where verification and security operations are performed. If the voice is successfully authenticated, and no deep fake is detected, the transaction is approved and the details are recorded in the database. If a fake voice or unauthorized user is detected, the transaction request is rejected to avoid any fraudulence. This architecture is used for secure voice authentication, deep fake detection and safe digital transaction processing.

VI. METHODOLOGY

The proposed Voice Based Secure Transaction System with Deep fake Detection uses a structured AI-based processing pipeline to ensure the security of voice authentication and transaction systems.

1. **Data Collection:** Authorized user voice samples are collected during the registration phase. The collected voice samples are stored in the system database.
2. **Preprocessing:** The collected voice input is preprocessed to enhance the quality of the voice. Noise reduction, normalization and segmentation techniques are employed to clean the voice.
3. **Feature Extraction:** Important voice features are identified from the voice signal. Mel Frequency Cepstral Coefficients (MFCC) are employed to identify the voice features.
4. **Speaker Verification:** The identified voice features are compared with the voice templates collected earlier. Machine learning algorithms are employed to compare the voice features.
5. **Deep fake Detection:** The voice is analyzed to identify the presence of deepfake.AI-based deep fake detection algorithms are employed to analyze the voice.
6. **Transaction Processing and Response:** If the voice is identified as genuine and not a deep fake, the system processes the transaction through a payment interface and logs the transaction details. If the voice is identified as fake and unauthorized, the system rejects the transaction request to avoid fraudulent activities.

VII. IMPLEMENTATION

The Voice Based Secure Transaction System with Deep fake Detection is a module-based system that brings together voice processing, machine learning models and a secure transaction system. The system is designed to facilitate a web interface through which a user can carry out financial transactions such as sending money and checking their account details through voice commands. The voice of the user is picked up through a microphone and transmitted securely to the backend server for processing. The backend system is responsible for processing the voice signals received from the frontend and carrying out the required transactions. The voice signals are first preprocessed to enhance their quality through techniques such as noise reduction and normalization. After the preprocessing stage, the voice features are extracted using Mel Frequency Cepstral Coefficients (MFCC), which is used to identify the voice of the user. These features are then used for speaker verification by comparing them with the voice template stored in the database to verify the identity of the user. At the same time, the voice used for the transaction is analyzed by a deep fake detection model to ascertain whether the voice is original or has been created by voice cloning technologies. The system database contains information about the user, voice and transactions, which are used for the authentication and verification process. If the voice is identified as authentic and the deep fake detection model does not flag the voice, the transaction request is processed securely and the transaction is recorded in the database. If the system detects a fake voice, the transaction request is declined to avoid any fraud transactions. APIs are used by the system components to communicate with each other, enabling voice-based transaction services securely.

VIII. EXPERIMENTAL RESULTS

In the current research, the performance of the Voice Based Secure Transaction System with Deep fake Detection was assessed based on the authentication accuracy, the ability to detect deep fake and the transaction processing efficiency. In the assessment, the system was subjected to various voice samples of the users, as well as manipulated voice samples, to ascertain the performance of the authentication module and the deep fake detection module. From the experimental results, it was evident that the system was able to successfully extract voice features based on Mel Frequency Cepstral Coefficients (MFCC). Additionally, the system was able to successfully verify the users based on speaker verification. During the testing process, the system was able to correctly authenticate genuine users and detect synthetic voices generated using deep fake or voice cloning techniques. The preprocessing and feature extraction stages improved the quality of the voice signal and helped the machine learning models perform accurate classification. The deep fake detection module successfully identified manipulated audio signals, preventing unauthorized users from performing fraudulent transactions. In addition, the transaction processing module was able to complete secure financial transactions with low response time after successful authentication. The

integration of voice recognition, speaker verification and deep fake detection ensured that only legitimate users could perform transactions. Overall, the proposed system demonstrated reliable performance in voice authentication, deep fake detection and secure transaction processing, making it suitable for modern voice-based financial applications.

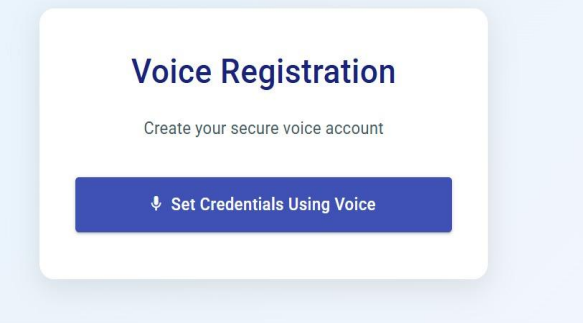

FIG 2. REGISTRATION PAGE

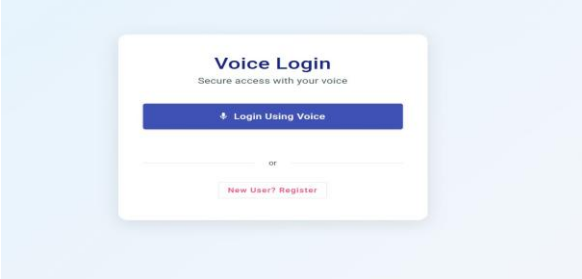

FIG 3. LOGIN PAGE

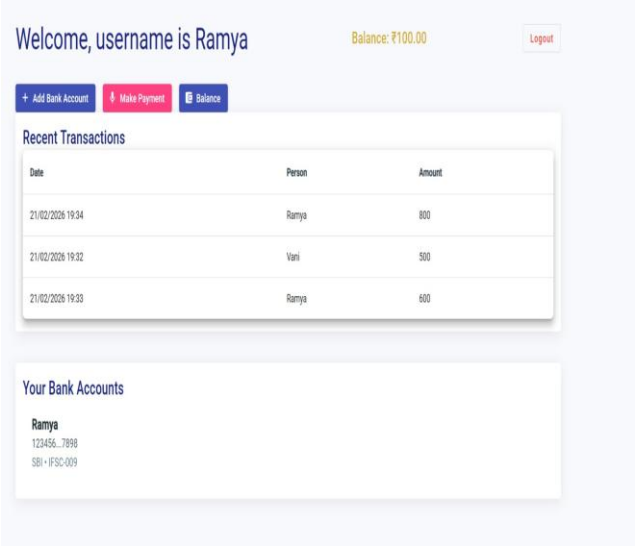

FIG 4. DASHBOARD PAGE

FIG 5. LOGOUT PAGE

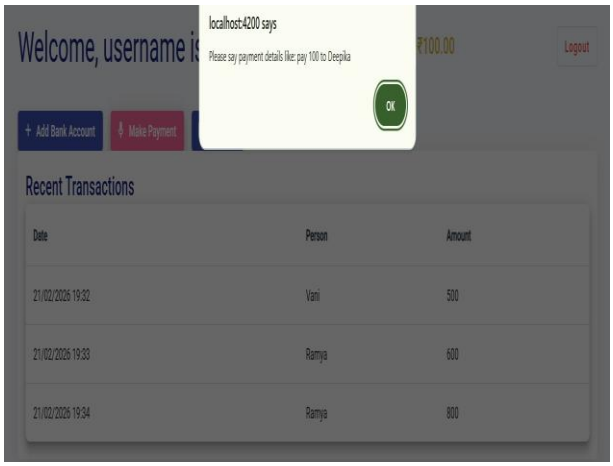

FIG 6. ACTIVITY PAGE

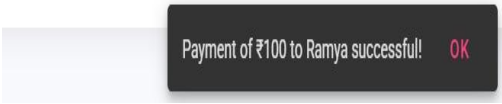

FIG 7. OUTPUT PAGE

IX. DISCUSSION

The implementation of the Voice Based Secure Transaction System with Deep fake Detection ensures enhanced security and reliability of voice-based financial transactions through the integration of voice biometrics, machine learning and deep fake detection technology. The layered architecture of the system ensures a modular approach, which makes it a scalable, flexible and maintainable system. The implementation of MFCC-based voice feature extraction and machine learning algorithms ensures accurate identification of the user's identity through their voice characteristics. Moreover, the integration of a deep fake detection module ensures enhanced security of the system, which prevents fraud transactions through voice spoofing and voice cloning attacks. The system also ensures enhanced accessibility and ease of use through voice-based transactions. The implementation of a voice-based transaction system ensures that only authenticated users can conduct transactions. However, it is clear that the performance of the system

depends on various factors, such as environmental noise, voice quality and server capabilities. Therefore, it is recommended that further optimization and advanced AI algorithms be used to enhance the performance of the system.

ADVANTAGES

The proposed Voice Based Secure Transaction System has the following advantages:

- Secure voice-based authentication of financial transactions
- Detection of deep fake voice attacks or synthetic voice attacks
- Hands-free operation through voice commands
- Reduces the need to rely on passwords, PINs, and OTPs
- Faster and user-friendly transaction system
- Increases security against identity theft and fraud
- Architecture is scalable to suit digital payment systems

LIMITATIONS

However, the system has some limitations:

- Voice recognition might be affected if there is background noise
- Variations in voice because of illness or other environmental factors might affect the system
- It requires updates for the deep fake detection system because of the emergence of new types of attacks
- It needs a quality microphone for the system to process the sound correctly
- It might depend on the network and the server for its performance

FUTURE ENHANCEMENT

There are a few possible improvements to be made to the system in the future to improve its functionalities:

- Integration of advanced deep learning models to improve the accuracy of deep fake detection
- Support for multiple language voice commands
- Development of mobile applications to provide voice commands for banking services
- Integration of banking APIs and digital payment gateways
- Implementation of real-time fraud detection systems
- Cloud deployment of the system to improve its scalability and performance
- Improved noise filtering and voice processing techniques to improve

X. CONCLUSION

In this study, a Voice Based Secure Transaction System with Deep fake Detection is proposed, which can enhance the security of voice-based digital financial transactions. In this system, voice commands are used as a means of authentication for initiating digital transactions. The voice signal is processed, and important voice features are extracted using Mel Frequency Cepstral Coefficients (MFCC) for speaker verification through machine learning algorithms. This helps in identifying the authorized user for initiating digital transactions. In addition, a deep fake detection mechanism is also integrated into this system, which helps in detecting synthetic or manipulated voices generated through artificial intelligence technologies. This helps in preventing voice spoofing attacks during digital financial transactions. The proposed system can be effectively used in digital financial transaction systems, mobile banking applications.where voice-based digital financial transactions are performed. The proposed system is secure, reliable and convenient for users, thereby enhancing the overall security of voice-based digital financial transaction systems.

REFERENCES

[1] R. K. Das, T. Kinnunen, and H. Li, “A survey on spoofing and countermeasures for automatic speaker verification,” Speech Communication, vol. 132, pp. 1–28, 2022.

[2] Y. Wu, P. L. De Leon, and M. P. Singh, “Deep fake audio detection using machine learning techniques,” IEEE Access, vol. 10, pp. 102345–102356, 2022.

[3] S. Todisco, H. Delgado, and N. Evans, “ASV spoof challenge: Automatic speaker verification spoofing and countermeasures,” IEEE Journal of Selected Topics in Signal Processing, vol. 11, no. 4, pp. 588–604, 2023.

[4] J. Villalba, N. Brummer, and E. Lleida, “State-of-the-art speaker recognition with deep neural networks,” IEEE Transactions on Audio, Speech, and Language Processing, vol. 31, pp. 1254–1266, 2023.

[5] Z. Wu and H. Li, “Voice conversion and spoofing detection using deep learning methods,” IEEE/ACM Transactions on Audio, Speech, and Language Processing, vol. 32, 2024

[6] L. Liu, Y. Wang, and S. Zhang, “Speech recognition based secure voice command systems for digital applications,” *IEEE Access*, vol. 12, pp. 45678–45689, 2024.

[7] A. K. Jain, A. Ross, and S. Prabhakar, “An introduction to biometric recognition,” *IEEE Transactions on Circuits and Systems for Video Technology*, vol. 14, no. 1, pp. 4–20, 2022.

[8] R. K. Das, X. Tian, and T. Kinnunen, “Generalization of spoofing countermeasures for automatic speaker verification,” *IEEE/ACM Transactions on Audio, Speech, and Language Processing*, vol. 30, pp. 120–132, 2024.
[9] Y. Zhang, H. Chen, and Q. Liu, “Deep learning based voice authentication for secure financial transactions,” *IEEE Access*, vol. 13, pp. 78543–78555, 2025.
[10] M. Patel and S. Gupta, “Deep fake voice detection using spectrogram analysis and machine learning,” *Journal of Artificial Intelligence Research*, vol. 79, pp. 345–360, 2025.
[11] K. J. Piczak, “Environmental sound classification with convolutional neural networks,” IEEE International Workshop on Machine Learning for Signal Processing, 2015.
[12] J. Salamon and J. P. Bello, “Deep convolutional neural networks and data augmentation for environmental sound classification,” IEEE Signal Processing Letters, vol. 24, no. 3, pp. 279–283, 2017.
[13] Y. LeCun, Y. Bengio, and G. Hinton, “Deep learning,” Nature, vol. 521, pp. 436–444, 2015
[14] B. Logan, “Mel frequency cepstral coefficients for music modeling,” International Symposium on Music Information Retrieval (ISMIR), 2000.
[15] P. W. Ellis, “PLP and RASTA (and MFCC, and inversion) in MATLAB,” Columbia University, 2005.