Audio Spoof Detection Integrated with a Home Automation System Using IoT

Dr.M.Mohanapriya[#1], Gourav Gopal[#2], Nalin Suriya S[#3], Vishal Karthik S[#4], Yokesh R S[#5]

Associate Professor, Computer Science and Engineering Department, Coimbatore Institute of Technology[#1].

Fourth year B.E, Computer Science and Engineering Department, Coimbatore Institute of Technology[#2,#3,#4,#5].

Coimbatore, TamilNadu, India

[1]mohanapriya.m@cit.edu.in, [2]1905015cse@cit.edu.in,
[3]1905031cse@cit.edu.in, [4]1905060cse@cit.edu.in, [5]1905062cse@cit.edu.in

***Abstract*—Authentication has become an essential aspect of our daily lives, with various authentication systems in place, ranging from traditional lock screens to biometric authentication systems. Among these systems, audio-based authentication has gained popularity, where users use specific words or phrases to unlock their devices and objects such as doors and mobile phones. However, the current audio authentication systems face a significant issue, as they only verify and extract the features of words and voices without classifying human voice and recorded human voices, leading to audio spoof attacks. To address this issue, the proposed system aims at using advanced machine learning models such as RNN and LSTM to classify human voice and recorded human voices, overcoming the problem of audio spoof attacks and recognizing the genuineness of the voice. The proposed system can be further integrated into any IoT system or home automation system, adding an additional layer of security to the accessibility of the device. The system's integration with IoT and home automation systems further enhances its security capabilities, making it a reliable authentication system for everyday use.**

***Keywords*—-RNN, LTSM, GMM, IoT**

1 INTRODUCTION

Development of VCD's, have boosted the realization of smart homes, voice-controlled authentication systems etc.,.These VCDs are vulnerable to different spoofing attacks. Audio Authentication is becoming very essential part of our lives. Audio Authentication spoofing is becoming an issue. High-quality audio recorders enable bypassing this audio authentication system by just recording the human voice and reusing them for accessing the same system. Thus, there exists a need to develop a voice anti-spoofing framework capable of detecting multiple audio spoofing attacks.

2 RELATED WORK

The literature survey was carried out in great detail using existing Audio Spoof Detection Methods.

Ivan Rakhmanenko [1] Bidirectional Long Short Term Memory (BiLSTM) networks with constant Q cepstral coefficients (CQCC) are used to classify real audio from fake audio in anti-spoofing systems. By fusing the BiLSTM and GMM-UBM systems, a fusion mechanism is used to increase the variability of the systems' decision-making processes and their accuracy.Over the baseline systems, these proposed systems significantly improved performance.

Kishor Kumar Sethy [2] *S*peech recognition based IoT system is carried out.For IoT system Raspberry pi board is used and for wifi communication ESP-32 module is used. For speech recognition Support vector machine model is used for training the speech recognition model.

Md Mahadi Hasan Nahid[3]Speaker identification (SI) is the system to identify the person by the signal pattern of their voices. With many speaker identification models have been proposed, but till now speaker identification technology do not reach their full potential. This paper presents a comprehensive comparative study of VQ and GMM to identify the speaker who speaks in Bengali accent. We consider the problem of text-independent speaker identification. We compare the performance/accuracy of VQ and GMM based Speaker Identification System (SIS). They've used Mel Frequency Cepstral Coefficients (MFCC) and Liner Predictive Coding Coefficients (LPCC) for feature extraction.These extracted features are then sent to the Convolutional Neural Network as input and then are classified as either synthetic or replay attacks.

Diqun Yan [4] says feature extraction methods like LFCC and MFCC are used to extract audio features and then these are sent as input to the Recurrent Neural Network (RNN) frame with two-layer LSTM to detect four common audio forgery operations.These are experimented mainly on TIMIT and UME databases and various evaluations like intra-database evaluation as well as cross-database evaluation are done and the detection accuracies of each of the above are identified.

Shilpa Lunagaria[5]presents deep fake audio forgery isidentified using Deep Learning algorithms. Audio files are taken as input and model is trained to uniquely identify features for voice creation and voice detection. The model could then classify between whether the audio is real or fake.The accuracy obtained for this model during training and validation phases are pretty high but the testing accuracy could be improved more by extracting more features and using different algorithms.

Alexander Shelupanov [6] and his colleagues describes Simple machine learning algorithms are used for voice Identification of 150 speakers. Dataset contains around 3000 samples. Mel Frequency cepstral coefficients(MFCC) is used for extracting features from audio samples. Machine learning algorithms like Support vector machine, Random forest algorithm are used for training the system

Jincheng Zhou **[7]** discusses about the immense usage of Automatic Speaker Verification (ASV) system which verifies users with their voices and it's susceptibility to voice spoofing attacks - logical and physical access attacks.A secured voice spoofing countermeasure to detect voice replay attacks is proposed. This has enhanced the ASV system security by building a spoofing countermeasure dependent on the decomposed signals that consist of prominent information. It uses two main features— the Gammatone Cepstral Coefficients (GCC) and Mel-Frequency Cepstral Coefficients (MFCC) — for the audio representation. For the classification of the features, Bi-directional Long-Short Term Memory Network in the cloud, a deep learning classifier.Numerous audio features and respective feature's capability to obtain the most vital details from the audio for it to be labelled genuine or a spoof speech is examined. Furthermore, it uses various machine learning algorithms to illustrate the superiority of the system compared to the traditional classifiers. The results of the experiments were classified according to the parameters of accuracy, precision rate, recall, F1-score, and Equal Error Rate (EER). The results were 97%, 100%, 90.19% and 94.84%, and 2.95%, respectively.

Mohit Dua [8] the system that is proposed tries to address the problem of classifying legitimate user and the malicious attacks using deep learning (DL) methods and ensemble of different neural networks. The first model that is discussed is a combination of time-distributed dense layers and long short-term memory (LSTM) layers.The other two deep neural networks (DNNs) are based on temporal convolution (TC) and spatial convolution (SC). Finally, an ensemble model comprising of these three DNNs has also been analysed. All these models are analysed with Mel frequency cepstral coefficients (MFCC), inverse Mel frequency cepstral coefficients (IMFCC) and constant Q cepstral coefficients (CQCC) at the frontend, where the proposed ensemble performs best with CQCC features. The proposed work uses ASVspoof 2015 and ASVspoof 2019 datasets for training and testing, with the evaluation set having speech synthesis (SS) and voice conversion (VC) attacked utterances. Performance of proposed system trained with ASVspoof 2015 dataset degrades with evaluation set of ASVspoof 2019 dataset, whereas performance of the same system improves when training is also done with the ASVspoof 2019 dataset.

3 PROPOSED METHODOLOGY

An Iot system which acts as an home automation system is proposed which uses a Bidirectional LSTM model to classify the audio samples and GMM model to identify the specific voice of the speaker. If the given audio input is found to be an spoofed audio or an invalid user , then the system denies the access to the system. Features fed into both the bidirectional LSTM model as well as GMM model are generated using MFCC methodologies. The proposed system performs binary classification of audio data which are mapped to two classes

a) Authentic/Bonafide
b) Spoofed

And identifies the speaker voice using the GMM model and performs the necessary action in the Home automation system.

4 DATASET USED

1.Audio Spoof Detection Model

The Dataset used in the model is AutomaticSpeakerVerification(ASV) 2019. ASV Spoof dataset contains two types of Audio files Physical Access - Bonafide utterances are made in a real, physical space in which spoofing attacks are captured and then replayed within the same physical space using replay devices of varyinquality.Logical Access - Bonafide and spoofed utterances generated using text-to-speech (TTS) and voice conversion (VC) algorithms are communicated across telephony and VoIP networks with various coding and transmission effectsThe dataset includes genuine and spoofed speech from 20 speakers (8 male, 12 female).Each spoofed utterance is generated according to one of 2 voice conversion and 3 speech synthesis algorithms. The voice conversion systems include those based on (i) neural-network-based and (ii)transfer-function-based methods.The speech synthesis systems were implemented with (i) waveform concatenation, (ii) neural-network-based parametric speech synthesis using source-filter vocoders and (iii) neural-network-based parametric speech synthesis using Wavenet.

2. Speaker Identification Model

The model receives 5 audio samples from the user for training and creates an GMM modelfor that specific speaker. When an audio file is given as input it compares the scores of all theGMM model and finds the model that produces the highest score value. The speaker is then identified with the label for whom the it is made.

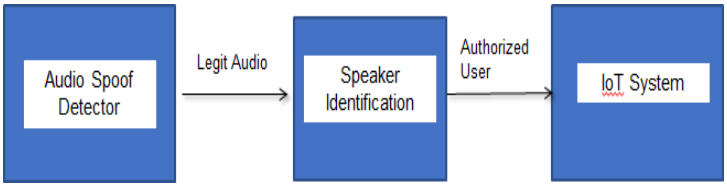

Fig 4.1

5 RESULT & DISCUSSION

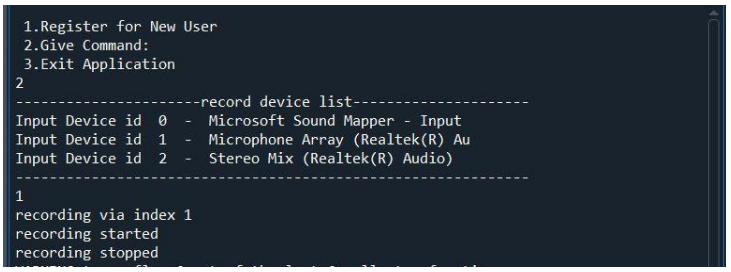

Fig 5.1

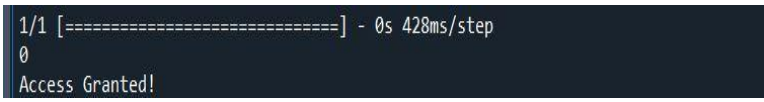

Fig 5.2

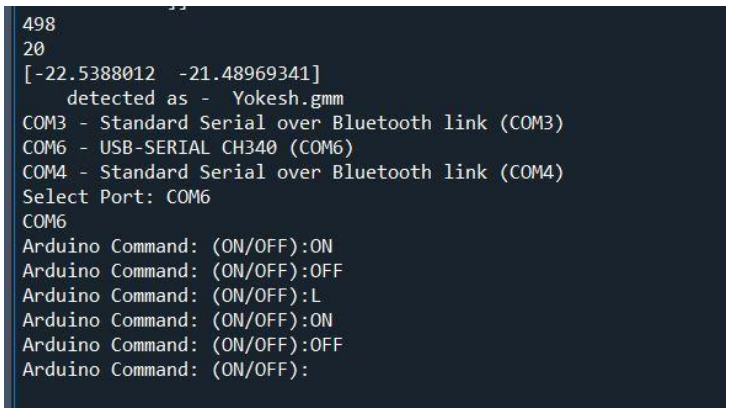

Fig 5.3

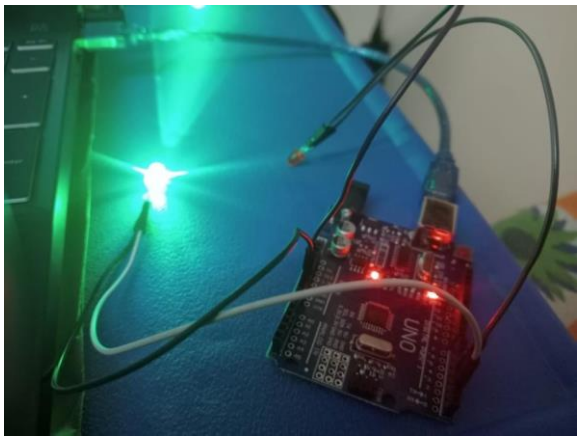

Fig 5.4

6 CONCLUSION

The Home automation system developed in this project is a reliable and efficient solution for safeguarding against audio spoofing attacks. The integration of an audio spoof detector and speaker identification system ensures that only authorized users can access the automated devices in the home. The accuracy obtained for the audio spoof detection model is 85%, which is a significant improvement over existing methods.The proposed speaker identification system, which utilizes MFCC feature extraction and GMM model, has been tested against both known and unknown users. The system performed as expected, effectively identifying the known user while denying access to the unknown users. This ensures that only authorized users can control the home automation system, thereby increasing security.The use of Arduino UNO board, LED lights, servo motor,connecting wires, and breadboard for IoT integration provides a cost-effective and reliable solution for home automation. The integration of IoT devices with the audio spoof detector and speaker identification system ensures that users can control their home environment with ease and convenience.The System demonstrates the effectiveness of integrating an audio spoof detector and speaker identification system with IoT devices for home automation. The proposed system provides reliable protection against audio spoofing attacks while ensuring that only authorized users can access the automated devices in the home. The accuracy obtained for the proposed system is promising and shows great potential for future research in this field.

7 REFERENCES

- Dua, M., Jain, C. & Kumar, S. LSTM and CNN based ensemble approach for spoof detection task in automatic speaker verification systems. *J Ambient Intell Human Comput* **13**, 1985–2000 (2022). https://doi.org/10.1007/s12652-021-02960-0
- Yamagishi, Junichi; Todisco, Massimiliano; Sahidullah, Md; Delgado, Héctor; Wang, Xin; Evans, Nicolas; Kinnunen, Tomi; Lee, Kong Aik; Vestman, Ville; Nautsch, Andreas. (2019). ASVspoof 2019: The 3rd Automatic Speaker Verification Spoofing and Countermeasures Challenge database, [sound]. University of Edinburgh. The Centre for Speech Technology Research (CSTR). https://doi.org/10.7488/ds/2555.
- https://www.simplilearn.com/tutorials/deep-learning-tutorial/rnn
- https://colah.github.io/posts/2015-08-Understanding-LSTMs/
- Sak, H., Senior, A.W., & Beaufays, F. (2014). Long short-term memory recurrent neural network architectures for large scale acoustic modeling. *INTERSPEECH*.
- Ankur, Tanjemoon & Kundu, Bipasha & Foysal, Md & Ortiz, Bengie & Chong, Jo. (2022). LSTM-Based COVID-19 Detection Method Using Coughing. 10.21203/rs.3.rs-2106413/v1.
- Akyol K, Şen B. Automatic Detection of Covid-19 with Bidirectional LSTM Network Using Deep Features Extracted from Chest X-ray Images. Interdiscip Sci. 2022 Mar; 14(1):89-100. doi: 10.1007/s12539-021-00463-2. Epub 2021 Jul 27. PMID: 34313974; PMCID: PMC8313418.
- Ivan Rakhmanenko Fusion of BiLSTM and GMM-UBM Systems for Audio Spoofing Detection August 2019 International Journal of Advanced Trends in Computer Science and Engineering 6(4):1741-1746.
- Jichen Yang, Rohan Kumar Das,Improving anti-spoofing with octave spectrum and short-term spectral statistics information,Applied Acoustics, Volume 157,2020,107017,ISSN 0003-682X,

- https://ijirt.org/master/publishedpaper/IJIRT149877_PAPER.pdf
- Mittal, Aakshi & Dua, Mohit. (2022). Automatic speaker verification systems and spoof detection techniques: review and analysis. International Journal of Speech Technology. 25. 10.1007/s10772-021-09876-2.
- https://irojournals.com/aicn/article/pdf/4/3/4
- https://journals.pan.pl/dlibra/publication/141648/edition/123487/content/archives-of-acoustics-2022-vol-47-no-2-spoofed-speech-detection-with-weighted-phase-features-and-convolutional-networks-br-disken-gokay?language=en
- Zhou, J., Hai, T., Jawawi, D.N.A. *et al*. Voice spoofing countermeasure for voice replay attacks using deep learning. *J Cloud Comp* **11**, 51 (2022). https://doi.org/10.1186/s13677-022-00306-5