Journal of Next Generation Technology (ISSN: 2583-021X)
Vol. 5, Issue 2, April 2025

Voice Spoofing Detection Using One Class Learning

G.Priyadharshini[1], Mrs.P.Rohini[2], Ms.S.Durga[3]

[1]Student, Department of CSE, Chendhuran College of Engineering&Technology, Pudukkottai. India

[2,3]Assistant Professor, Department of CSE, Chendhuran College of Engineering&Technology, Pudukkottai. India

Email id : priyagopal1903@gmail.com[1], rohini.ccet@gmail.com[2], durga.prec@gmail.com[3]

Citation

G.Priyadharshini [1], Mrs.P.Rohini [2], Ms.S.Durga i, "Voice Spoofing Detection Using One Class Learning", Journal of Next Generation Technology (ISSN: 2583-021X), vol. 5, no. 2, pp. 58-65. April 2025. DOI: 10.5281/zenodo.15617704

ABSTRACT: Spoofing detection is increasingly critical as voice recognition systems are widely used in security applications. Spoofing occurs when an individual attempts to replicate another person's voice to gain unauthorized access, presenting substantial risks in sectors like finance and telecommunications. Traditional detection methods require large datasets of both genuine and spoofed voice samples, which can be difficult to obtain, especially as new spoofing techniques emerge. This project addresses these challenges by implementing a one-class learning approach that uses only authentic voice samples to train the detection model. Through an emphasis on the distinctive features of real voices, the system recognizes spoof sounds as anomalies or "out of place." To build a reliable and flexible detection system, sophisticated methods like autoencoders and One-Class Support Vector Machines (SVMs) are used. This creative method offers a versatile solution that is simple to incorporate into current voice security systems, in addition to increasing the accuracy of voice spoofing detection. The ultimate goal of this effort is to increase security for users who depend on voice recognition technology and strengthen defenses against voice-based fraud.

Keywords: support vector machine, Artificial Intelligence, Machine Learning, Voice Spoofing

I. INTRODUCTION

Voice-based authentication systems have become increasingly popular, being implemented in a variety of applications ranging from personal devices to secure access control in businesses. These technologies give a quick and safe way to use their voice to verify their identification. However, they face a significant vulnerability voice spoofing. In this type of attack, a malicious individual attempts to mimic a legitimate user's voice in order to bypass security measures and gain unauthorized access to sensitive information or systems. Traditional methods for detecting voice spoofing typically rely on extensive datasets containing both real and spoofed voice samples. The collection of vast quantities of spoof data, however, can be difficult, particularly as more advanced spoofing methods keep appearing. Voice spoofing

techniques are becoming increasingly varied as voice synthesis technology advance, resulting in a constantly changing threat scenario that calls for One-class learning, which trains the detection model using only real sound samples, is a promising approach to this problem.

By using this method, the system may concentrate on learning the distinctive qualities of real voices, which will help it spot possible spoofing attempts as anomalies when an incoming voice deviates from these recognized patterns. This approach offers a more scalable and adaptable solution and does away with the requirement for labeled spoof data. By employing this novel approach, the project seeks to improve the security of voice-based systems against changing threats and offer a more dependable and effective way to protect voice authentication technologies by investigating different one-class techniques learning to create an efficient and flexible model for voice spoofing detection.

II. LITERATURE REVIEW

Sahidullah et al. (2016) looked into a variety of feature extraction techniques, including spectral characteristics, Mel-Frequency Cepstral Coefficients (MFCCs), and Linear Predictive Coding (LPC), with the purpose of identifying spoof voices. The study discovered that improving the accuracy of spoofing detection models requires careful feature selection. Extensive research demonstrated that certain features are more successful than others in differentiating between real and fake audio, which can greatly increase detection rates, particularly when paired with machine learning classifiers[1]. Constant Q Cepstral Coefficients (CQCCs) are a unique feature set that Todisco et al. (2017) introduced with the goal of enhancing replay attack detection performance. The study indicated that CQCCs outperform established features like FCCs, particularly in tough noise settings, making them desirable addition to the future extraction toolset for spoofing detection[2]. Long Short-Term Memory (LSTM) networks, which are used in Automatic Speaker Verification (ASV) systems to identify spoofing. Their study shown that deep learning models are highly effective at identifying intricate patterns in voice data that differentiate real samples from fakes.This study demonstrates how deep neural networks can increase detection accuracy, particularly when more conventional approaches are unable to adapt to novel spoofing strategies[3]. The study focuses on creating solutions that can function effectively on devices with limited processing power, like smartphones and Internet of Things devices. 4. paired with Support Vector Machine (SVM) classifiers for voice spoofing detection in resourceconstrained environments. The suggested system showed competitive performance in detecting replay attacks, making it a workable solution for real-world applications where computational resources are limited[4].Wu et al. (2021) talked about how data augmentation methods affect how reliable voice spoofing detection systems are. The study sought to improve model generalization and recreate a variety of real-world scenarios by implementing transformations such noise addition, temporal stretching, and pitch shifting. The findings suggested that by exposing machine learning models to a wider variety of spoofing scenarios, data augmentation could increase the detection accuracy of these models and strengthen systems against new threats[5]. Luo et al. (2021) looked on the creation of artificially spoofing audio samples for detection model training using Generative Adversarial Networks (GANs). This method improves the resilience of the model by producing realistic voice attacks, which solves the problem of the restricted supply of spoof data. According to the study, generative models can be a useful tool in anti spoofing research because they improved detection systems' capacity to identify hidden forms of spoofing when trained on a combination of real and GAN-generated spoofing data[6]. A thorough assessment of the state of voice anti spoofing research was given by Kinnunen et al. (2022), who also talked about potential future paths for the discipline. To address the dynamic nature of spoofing tactics, the

study underlined the significance of creating universal and adaptive detection strategies, such as one-class learning approaches. In order to ensure their efficacy in realworld applications where threats are ever-evolving, the authors emphasized that strong detection systems must be able to adjust to new types of spoofing without requiring lengthy retraining[7]-[11].

III. PROPOSED SYSTEM

The "Voice Spoofing Detection Using One-Class Learning" project's suggested methodology focuses on creating a system that can detect new spoofing techniques without requiring large datasets of fake audio because it only needs to be trained on real speech samples. First, real voice recordings will be gathered and preprocessed, which will entail lowering background noise and normalizing audio levels. To capture crucial speech characteristics, key features such as Mel-Frequency Cepstral Coefficients (MFCCs) and Constant Q Cepstral Coefficients (CQCCs) will be retrieved.

Autoencoders and One-Class Support Vector Machines (OC-SVM) will be the two primary models utilized. While the autoencoder will reconstruct audio features and identify spoof sounds based on high reconstruction errors, the OC-SVM will learn the typical patterns of real voices and flag abnormalities. A dataset of authentic and spoof audio will be used to train and validate the models, and performance will be assessed using metrics like accuracy and Equal Error Rate (EER). Implementing a real-time detection system that can effectively and quickly identify fake voices is the ultimate goal in order to provide strong security for voice authentication applications.

ALGORITHM

RESNET ALGORITHM

- Every succeeding CNN-based architecture since the first CNN-based architecture (AlexNet) won the ImageNet 2012 competition uses more layers in a deep neural network to lower the error rate. This works for fewer layers, but as the number of layers increases, a common issue in deep learning is known as the Vanishing.

VGC16 MODEL

- VGG16 is a convolutional neural network model that's used for image recognition. It unique in that it has only 16 layer that have weight as opposed to relying on a huge number of hyper parameters.

METHODOLOGY

ONE CLASS LEARNING

- A subset of multi-class classification known as One-Class Classification (OCC) uses training data from a single positive class.
- Learning a representation and/or a classifier that allows for the recognition of favorably labeled questions during inference is the aim of OCC.8

FLASK FRAMEWORK

- Known for its scalability and ease of use, Flask is a lightweight and adaptable Python microframework for web application development that enables programmers to produce web applications rapidly and effectively.

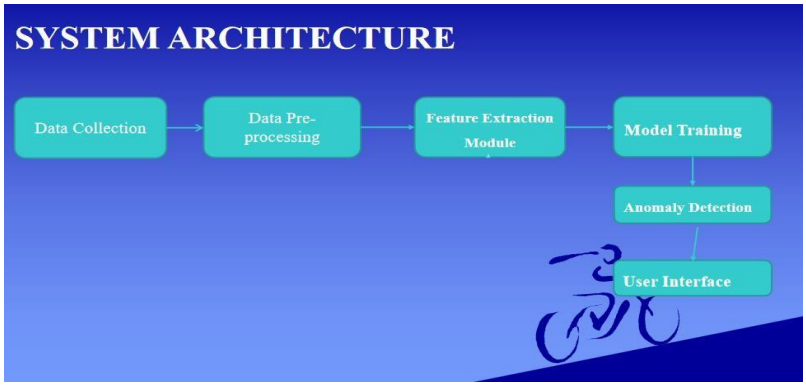

Fig.1 Architecture diagram

DATA COLLECTION:

The first module involves gathering and preparing the dataset for training the spoofing detection model. Since this project focuses on one-class learning, the dataset will consist solely of genuine voice samples, representing normal, authentic speech. Once the dataset is collected, preprocessing is a crucial step. This includes standardizing the audio format (e.g., converting to 16 kHz, mono), removing background noise, and normalizing the volume levels across samples to ensure consistency.

FETURE EXTRACTION:

Feature extraction is a key step that transforms raw audio data into meaningful representations that can be used by machine learning models. In this module, relevant audio features are extracted from the preprocessed voice samples. Common features used for spoofing detection include Mel-Frequency Cepstral Coefficients (MFCCs), which capture the spectral properties of speech, and Constant Q Cepstral Coefficients (CQCCs), known for their effectiveness in replay attack detection.

MODEL TRAINING:

One-Class Support Vector Machine (OC-SVM) and autoencoders. OCSVM works by learning a boundary around the data points that represent the normal voice characteristics. During testing, any input that falls outside this boundary is flagged as anomalous or spoofed. Autoencoders, on the other hand, are neural networks trained to compress and reconstruct input data. When the model encounters a spoofed voice, the reconstruction error will be high, indicating that the input is not similar to the genuine voices the model was trained on.

ANOMALY DETECTION:

In the anomaly detection module, the trained models are used to identify deviations from the normal voice patterns that were learned in the previous step. When a new voice sample is input into the system, the model compares the features of the sample to the distribution of genuine voice features it has learned.

FINAL PREDICTION:

The final prediction module is responsible for providing the output classification, determining whether a given voice sample is genuine or spoofed. After the anomaly detection step, this module takes the results from the anomaly detection models (OC-SVM or autoencoder) and generates a final decision based on the features and reconstruction error. If the sample is flagged as an anomaly, it is classified as spoofed; otherwise, it is considered genuine. The final prediction step integrates the results into a usable form, such as a binary classification output ("genuine" or "spoofed") or a confidence score indicating the likelihood of spoofing.

IV SOFTWATRE SPECIFICATION

PYTHON

The Google Brain Team created the open-source machine learning framework TensorFlow. It is among the most widely used libraries for creating and refining deep neural networks and other machine learning models. TensorFlow makes it simple for developers to create sophisticated models for natural language processing, picture and audio recognition, and other applications. TensorFlow's capacity to manage intricate calculations and big datasets is one of its primary characteristics, which makes it appropriate for deep neural network training. Faster training times are made possible by the parallelization of computations across several CPUs or GPUs. Additionally, TensorFlow offers Keras, a high-level API that streamlines the model-building and training process.

Integration with other Python libraries and frameworks is made simple by TensorFlow's extensive collection of tools and libraries. Preparing data for training and analyzing model performance is made simple by its integrated support for data preprocessing and visualization. TensorFlow's ability to distribute models across multiple platforms, including as mobile devices and the web, is one of its main features. For deploying models on Android, iOS, and other mobile platforms, TensorFlow Lite is a mobile-optimized version of TensorFlow. TensorFlow.js is a JavaScript package that enables for training and deployment of models directly in the browser. To construct and train machine learning models, TensorFlow offers a variety of features and tools. TensorFlow's salient characteristics include:

COMPUTING BASED ON GRAPHS:

TensorFlow employs a computing model based on graphs, which enables effective computation across a number of devices and CPUs/GPUs.

AUTOMATIC DIFFERENTIATION:

TensorFlow's automatic differentiation feature makes it possible to compute gradients for backpropagation methods in an effective manner. **HIGH-LEVEL APIS**: TensorFlow offers high-level APIs, like Keras, that let programmers create and train intricate models rapidly and with little code.

PREPROCESSING AND DATA AUGMENTATION: TensorFlow offers a variety of preprocessing and data augmentation techniques, such as data normalization and picture and text preprocessing TensorFlow facilitates distributed training across a number of devices, CPUs, and GPUs, which enables quicker training periods and more economical resource.

DECISION MAKING STATEMENTS

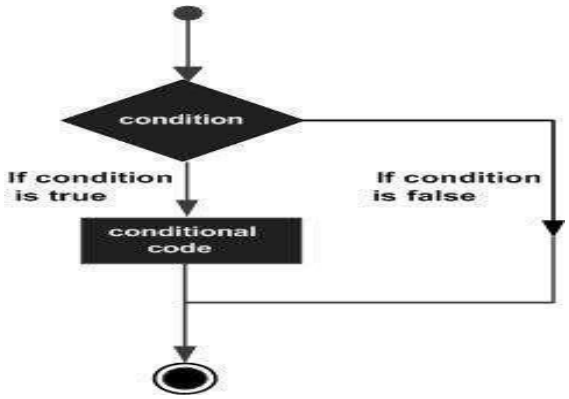

Fig.2 flowchart for if conditions

V. FUTURE ENHANCEMENTS

For future work in the "Voice Spoofing Detection Using One-Class Learning" project, several directions can be explored to enhance the system's robustness, accuracy, and adaptability to emerging threats.

Fig.3 home page

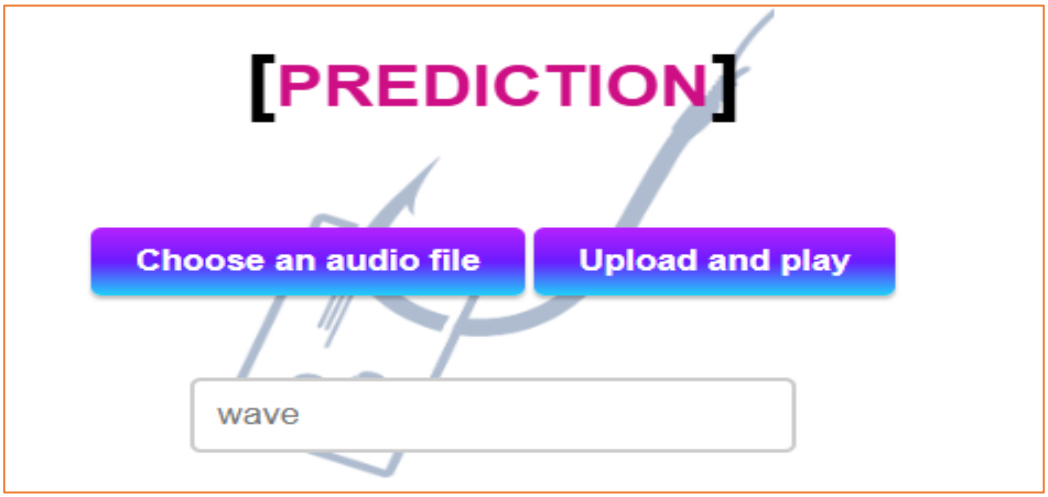

Fig.4 upload audio files

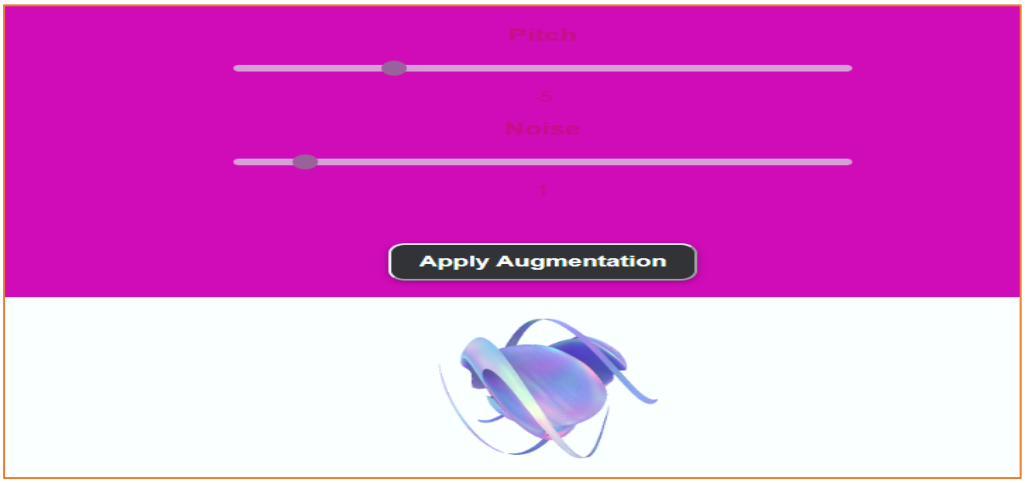

Fig.5 final Augmentation

VI. CONCLUSION

In summary, the "Voice Spoofing Detection Using One-Class Learning" initiative offers a novel solution to the expanding problem of speech spoofing in security systems. The challenge of acquiring complete and varied spoof data frequently limits the use of traditional techniques, which depend on big datasets of both real and spoof voices. The suggested system is more flexible to new spoofing tactics because it just needs real voice samples for training, eliminating the requirement for large amounts of labeled data, thanks to the use of one-class learning techniques.

This method lessens reliance on constantly updated datasets of phony sounds while simultaneously streamlining the data collection procedure. A thorough and effective system is guaranteed by the suggested methodology, which consists of dataset collection and preprocessing, feature extraction, model training, anomaly detection, and final prediction. The

system can recognize spoof sounds by identifying deviations from the learnt regular patterns of real speech through the use of autoencoders and One-Class SVM. Since the models are only trained on real voice data, they are more adaptable and capable of identifying spoofing attempts that haven't been seen before. This is important for real-world applications where spoofing methods are always changing.

REFERENCES

[1] Machine intelligence is another name for artificial intelligence (AI).Artificial Intelligence: https://en.wikipedia.org/wiki/Artificial_intelligence

[2] "Comparing Speech Recognition Systems (Microsoft API, Google API, and CMU Sphinx)" by Bohouta.G and Këpuska V.Z, International Journal of Engineering Research and Application, 2017.

[3] CMUSphnixFundamental ideas of speech: The process of voice recognition. http://tutorialconcepts.sourceforge.net/cmusphinx

[4] Apple Siri, Google Assistant, and Cortana Intelligence. Carpenter, R. and Fryer, L.K. (2006). Bots as language-learning aids Technology and Language Learning

[5] Ford, W.R., Hill, J., and Farreras, I.G. (2015). a study of real-world human-to-human and human-to-chatbot interactions with artificial intelligence. Human Conduct in Relation to Computers.

[6] Huang, J., Zhou, M. and Yang, D., 2007, January. Extracting Chatbot Online Discussions Forums Provide Information. In IJCAI(Vol. 7, pp. 423-428).

[7] Marr. B, The Amazing Ways Google Uses Deep Learning AI.

[8] Mohasi, L. and Mashao, D., 2006. Text-to-Speech Technology in Human-Computer Interaction. In 5th Conference on Human Computer Interaction in Southern Africa, South Africa (CHISA 2006, ACM SIGHI) (pp. 79-84).

[9] Noda.K, Arie.H, Suga.Y, Ogata.T, Multimodal integration learning of robot behavior using deep neural networks, Elsevier: Robotics and Autonomous Systems, 2014.

[10] Thakur, N., Hiwrale, A., Selote, S., Shinde, A. and Mahakalkar, N., Artificially Intelligent Intelligent

[11] A. N. Akkewar, “Statistical analysis of hybrid renewable energy systems by using artificial intelligence,” Journal of Next Generation Technology (ISSN: 2583–021X), vol. 1, no. 2, 2021.