A Review Paper on Deepfake Voice Detection

Adityanarayan Chaudhary[1], Jainam Chheda[2], Bhavik Panchal[3], Lavanya Deole[4]

Department of Information Technology, Shah and Anchor Kutchhi Engineering College Mumbai, Maharashtra, India

Dr. Bhavesh Patel

Principal, Shah and Anchor Kutchhi Engineering College

Mr. Manish Bhelande

Assistant Professor, Information Technology Department, Shah and Anchor Kutchhi Engineering College

Abstract-**Voice-based authentication and identification systems have gained considerable popularity due to their convenience and security. However, the proliferation of voice-based technologies has also led to an increase in fake voice fraud, which is a significant threat to the integrity of such systems. This review paper provides an in-depth overview and analysis of the voice deepfakes system. We explore the various techniques and methods used to detect fake voices, their advantages, limitations, and possible future developments.**

Keywords- Voice-based authentication, fake voice fraud, deep fake voice recognition, advantages, limitations, data protection, conclusion.

I. INTRODUCTION

The speech recognition and identification system is an essential part of modern technology, from personal device security to financial transaction authentication. The ease of use and convenience of voice-based systems have become increasingly popular, but they have also created a new challenge i.e. fake voice attacks. Fraudsters can use various techniques such as identity theft and unauthorized access to manipulate or imitate voices for malicious purposes. [5] False language detection systems have been developed to counter these threats and are continuing to evolve.

II. TYPES OF SPOOFED AUDIO ATTACKS

A. Sound shaping

Voice changes include changing personality and voice so that it sounds like someone else. Attackers can change the tone and other audio features with software or hardware.

B. Repeated attacks

Repeated attacks include recording the actual audio type and replaying the real speaker. These attacks are relatively simple but can be effective if they are not detected.

C. Synthetic sound production

Repeated attacks include recording the actual audio type and replaying the real speaker. These attacks are relatively simple but can be effective if they are not detected.

FAKE VOICE DETECTION TECHNIQUE

In our system, we use Tortoise models to detect voices. Tortoise-TTS is a multilingual text-to-speech model that produces high-quality voices with realistic prosody and tones. [9] It is customizable and allows users to create their unique sound by providing reference clips of the desired speakers. Turtle TTS can be used for a variety of applications, including audiobooks, podcasts, video game characters, personal assistants, tutorials, and video editing software. [9]

Here are some technical details about how deep fake voice recognition models work:

A. Select audio features:

The first step is to extract audio features from the audio clip that is entered. This can be done in several ways, including: MFCC and Waveform analysis. [9]

B. Mel-frequency cepstral coefficients :

MFCCs, or multifrequency color-coded representations of sound waves, are useful tools for differentiating between speakers and speech traits. A visual depiction of a sound's frequency and time components is called a spectrogram. They can be used to identify sound patterns, like irregular pitch or intonation, that might point to deep distortion. [9]

C. *Waveform Analysis:*

Waveform analysis can be used to identify minute variations in the waveform between a genuine sound and a deep false sound, such as variations in the background noise or syllable timing. Put the deep learning model into practice. After the audio features are extracted, a deep learning model is trained with them to identify between deep and real sounds. Several deep learning architectures, such as recurrent or convolutional neural networks, can be used for this. [9]

D. *Detect deep fake voices:*

The model can be trained to identify fake deep sounds in fresh audio clips. To achieve this, a new clip's audio properties are fed into the model, which then determines whether the clip is real or a deepfake. These are some specific instances of how deep fakes can be identified using deep fake voice recognition models. [9]

E. *Recognizing unnatural pitch or intonation:*

The voices of deepfakes are frequently artificial or robotic. This is because deepfake models are still being developed and have difficulty accurately mimicking human speech. By searching for patterns in voice characteristics that are different from those of real voices, deep-fake voice recognition models can identify intonations or pitches that are not natural. [9]

F. *Detecting audio waveform inconsistencies:*

Inconsistencies in the sound waveform, such as variations in the timing of individual syllables or background noise, can also be present in deep false sounds. By comparing the audio waveform of the input clip with a database of real and deepfake audio waveforms, deep fake voice recognition models can identify these discrepancies. [9]

It's crucial to remember that deep fake voice recognition algorithms are not flawless. [5] A sophisticated deep spoofing algorithm can still fool them. But as deepfake voice recognition technology advances quickly, it gets harder and harder to produce deep fakes that sound just like real voices.

III. FIGURES

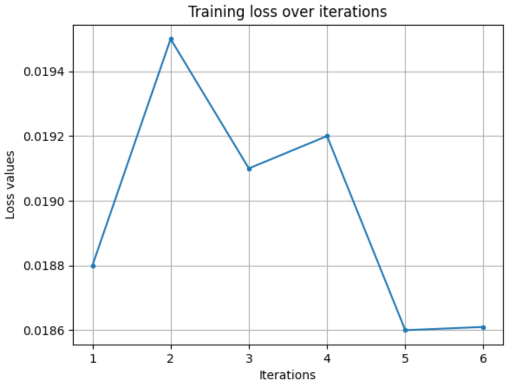

Fig. 1: This line plot depicts the training loss of our custom RNN over iterations, which was trained on an unlabeled dataset of 48 deepfake voice files with varying sampling rates and lengths.

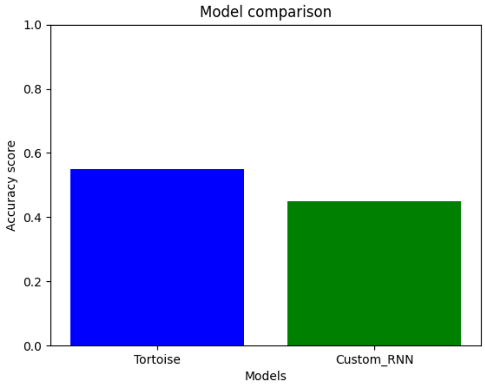

Fig. 2: This bar chart compares Tortoise - TTS prediction scores to our custom RNN (which was trained on an unlabeled dataset of 48 deepfake voice files with varying sampling rates and lengths).

IV. ADVANTAGES OF FAKE VOICE RECOGNITION SYSTEMS

A. *Advanced protection*

Counterfeit systems significantly improve the security of voice-based authentication and identification systems by identifying and blocking fraudulent companies.

B. *Ease of use*

Users can enjoy the convenience of voice-based systems without worrying about unauthorized use or customization.

C. *Adaptability*

Spoofing systems can adapt to evolving speech-processing technologies, making them resilient to new threats.

V. LIMITATIONS AND CHALLENGES

A. False positives and negatives

Fake voice recognition systems can sometimes produce false positives or false negatives, affecting the user experience. [5]

B. Resource intensity

Some detection methods can be resource-intensive, leading to higher computational requirements and possible delays. [5]

C. Data protection issues

Collecting audio data for authentication can raise privacy concerns, especially in situations where the data is stored centrally. [1]

VI. FURTHER DEVELOPMENTS

A. Multimodal authentication

Integrating multiple authentication factors such as voice, facial recognition, and behavioral biometrics can improve security. [2] [6]

B. Behavioral analysis for Voice Recognition:

Behavioral analysis may become more integral in voice recognition systems. Analyzing patterns in how individuals speak, their intonation, and other behavioral aspects could contribute to more accurate and personalized authentication.

C. Improved Deep Learning Model:

Further developments in artificial intelligence and deep learning will probably result in more complex and precise fake voice detection models. As a result, there may be fewer false positives and negatives as these models improve at differentiating between real and artificial voices. [4]

D. Hybrid Approaches

The combination of different detection technologies and technologies forms a hybrid approach. For example, incorporating traditional signal processing methods with advanced deep learning models to create a more complete and robust false voice detection system.

VII. CONCLUSION

As speech-based technologies become more common, the need for robust fake voice detection systems becomes increasingly critical. These systems play a key role in securing personal information and ensuring the reliability of voice-based authentication and identification. Continued research and development are essential to keep up with the evolving techniques used by attackers. The future of voice security is in the hands of innovative solutions that can provide effective privacy protection against spoofed voice attacks.

ACKNOWLEDGEMENT

We extend our heartfelt appreciation to Dr Bhavesh Patil, our dedicated project guide and the Principal of Shah and Anchor Kutchhi Engineering College. His invaluable guidance and unwavering support have played a pivotal role in the successful completion of this research paper. Additionally, we would like to acknowledge Mr. Manish Bhelande, our co-guide, for his expertise and assistance, which have significantly contributed to the refinement and shaping of our work. We are truly grateful for their mentorship and support throughout this research endeavour.

REFERENCES

[1] Dora M. Ballesteros, Yohanna Rodriguez, Diego Renza, “A dataset of histograms of original and fake voice recordings”, 2022

[2] Jayaram, M. V. Gopalachari, S. Rakesh, J. S. Sai, and G. K. Kumar, “Fake face image detection using feature network: A deep learning framework for detecting fabricated parts of images in social media”, 2022

[3] Darius Afchar, Vincent Nozick, Junichi Yamagishi, Isao Echizen, “Mesonet: a compact facial video forgery detection network”, 2018

[4] Thanh Thi Nguyen, Cuong M. Nguyen, Tien Dung Nguyen, Saeid Nahavandi, and Thanh Tam Nguyen, “Deep learning for deepfakes creation and detection: a survey”, 2019

[5] Santosh Kolagati, Thenuga Priyadharshini, V. Mary Anita Rajam, “Exposing deep fakes using a deep multilayer perceptron – convolutional neural network model”, 2022

[6] Jayaram, M. Venu Gopalachari, S. Rakesh, J. Shiva Sai, and G. Kiran Kumar, “Fake face image detection using feature network”, 2022

[7] Ruben Tolosana, Sergio Romero-tapiador, Ruben vera-Rodriguez, Ester Gonzalez-Sosa, Julian Fierrez, “Deepfakes detection across generations: analysis of facial regions, fusion, and performance evaluation”, 2022

[8] Siddharth Solaiyappan, Yuxin Wen, “Machine learning based medical image deep fake detection: A Comparative Study”, 2022

[9] D Bertero, Pascale Fung, “A First Look into a Convolutional Neural Network for Speech Emotion Detection”, 2017