IJARCCE ISSN (O) 2278-1021, ISSN (P) 2319-5940

International Journal of Advanced Research in Computer and Communication Engineering

Impact Factor 8.102 ⁜ Peer-reviewed & Refereed journal ⁜ Vol. 13, Issue 3, March 2024

DOI: 10.17148/IJARCCE.2024.13344

Deepfake Audio Detection using Deep Learning

Ankith Shetty[1], Hanzala Karani [2], Shreya K H[3], Raheeza Khan[4], Mr. Amruth A G[5]

Student, Dept. of Artificial Intelligence & Machine Learning, Mangalore Institute of Technology & Engineering, Moodabidre, India[1-4]

Professor, Dept. of Artificial Intelligence & Machine Learning, Mangalore Institute of Technology & Engineering, Moodabidre, India[5]

Abstract: The rise of deepfake technology poses a significant threat to the authenticity and integrity of multimedia content, including audio recordings. In response to this challenge, this project proposes a deep learning-based approach for detecting deepfake audio. Leveraging advancements in machine learning and signal processing, the proposed system aims to distinguish between genuine and manipulated audio recordings with high accuracy.The project begins with a comprehensive exploration of existing deepfake detection techniques, focusing on their limitations and strengths, particularly in the context of audio manipulation. Subsequently, a novel deep learning architecture is designed and implemented to effectively capture the subtle cues and patterns indicative of audio manipulation.Key components of the proposed system include feature extraction modules tailored to the unique characteristics of audio data, as well as deep neural network models trained on large-scale datasets of both genuine and deepfake audio samples. Through extensive experimentation and evaluation, the effectiveness and robustness of the developed system are assessed across various types of audio manipulation techniques and levels of sophistication.

Keywords: Deepfake, Audio manipulation, Deep learning, Detection, Feature extraction, Neural networks

I. INTRODUCTION

The advent of deepfake technology has ushered in a new era of multimedia manipulation, posing unprecedented challenges to the authenticity and integrity of audiovisual content. Deepfakes, which are highly realistic synthetic media generated using artificial intelligence algorithms, have raised concerns regarding their potential misuse for malicious purposes such as spreading misinformation, impersonation, and undermining trust in audio recordings.

Amidst this landscape, the detection of deepfake audio has emerged as a critical area of research and development. Unlike traditional methods of audio manipulation, deepfake techniques employ advanced machine learning algorithms to seamlessly alter speech, intonation, and other acoustic attributes, making it increasingly difficult to distinguish between genuine and manipulated audio recordings. Consequently, there is a pressing need for robust and reliable detection mechanisms capable of identifying deepfake audio with high accuracy and efficiency.

This project endeavors to address this challenge by leveraging the power of deep learning—a subset of machine learning that utilizes neural networks with multiple layers of abstraction—to develop an effective deepfake audio detection system. By harnessing the inherent complexity and non-linear relationships within audio data, deep learning models offer the potential to discern subtle patterns and anomalies indicative of audio manipulation, thereby enabling the automated detection of deepfake content.

The primary objective of this project is to design, implement, and evaluate a deep learning-based approach for detecting deepfake audio. This involves the development of novel architectures and methodologies tailored to the unique characteristics of audio data, as well as the collection and curation of large-scale datasets comprising both genuine and manipulated audio samples. Through rigorous experimentation and evaluation, the performance and robustness of the proposed system will be assessed across various types of audio manipulation techniques and levels of sophistication.

Furthermore, this project aims to contribute to the broader research efforts aimed at combating the proliferation of deepfake content and safeguarding the trustworthiness and reliability of multimedia communication channels. By advancing the state-of-the-art in deepfake audio detection, this work seeks to empower individuals, organizations, and technology platforms with the tools and insights needed to mitigate the potential risks associated with audio manipulation in the digital age.

DOI: 10.17148/IJARCCE.2024.13344

II. LITERATURE SURVEY

In [1]Khochare, Janavi, et al. "A deep learning framework for audio deepfake detection." Arabian Journal for Science and Engineering (2021): 1-12.

In [2]Wani, Taiba Majid, and Irene Amerini. "Deepfakes audio detection leveraging audio spectrogram and convolutional neural networks." International Conference on Image Analysis and Processing. Cham: Springer Nature Switzerland, 2023.

In [3]Suratkar, Shraddha, et al. "Deep-fake video detection approaches using convolutional–recurrent neural networks." Journal of Control and Decision 10.2 (2023): 198-214.

In [4]Mcuba, Mvelo, et al. "The effect of deep learning methods on deepfake audio detection for digital investigation." Procedia Computer Science 219 (2023): 211-219.

In [5]Wijethunga, R. L. M. A. P. C., et al. "Deepfake audio detection: a deep learning based solution for group conversations." 2020 2nd International conference on advancements in computing (ICAC). Vol. 1. IEEE, 2020.

III. SCOPE AND METHODOLOGY

Aim of the project

The aim of this project is to employ deep learning techniques to develop an accurate and robust system for detecting deepfake audio. By utilizing deep neural networks, the project seeks to differentiate between genuine and manipulated audio recordings with high precision.

Through the exploration and optimization of deep learning architectures, the project aims to effectively capture subtle patterns and anomalies indicative of audio manipulation. The primary objective is to contribute to the advancement of deep learning-based methods for audio forensics, thereby enhancing the reliability and trustworthiness of multimedia content in the face of emerging threats posed by deepfake technology.

Existing system

The existing system for deepfake audio detection relies solely on conventional machine learning (ML) techniques. It involves feature engineering and the utilization of standard classifiers. However, this approach may struggle to capture complex patterns in audio data effectively.

Additionally, manual feature selection can be labor-intensive and may not fully exploit the richness of the data. Consequently, there is a need to explore more advanced methods such as deep learning to improve accuracy and robustness in detecting deepfake audio.

Proposed system

The proposed system for deepfake audio detection leverages deep learning methodologies exclusively. It aims to replace traditional machine learning techniques with neural network architectures capable of learning intricate patterns directly from raw audio data.

By utilizing deep learning models such as Convolutional Neural Networks (CNNs) or Recurrent Neural Networks (RNNs), the system aims to achieve superior performance in distinguishing between genuine and manipulated audio recordings.

Transfer learning from pre-trained models and techniques like adversarial training may be incorporated to enhance robustness and adaptability. Ultimately, the proposed system seeks to advance the accuracy and reliability of deepfake audio detection by harnessing the power of deep learning.

DOI: 10.17148/IJARCCE.2024.13344

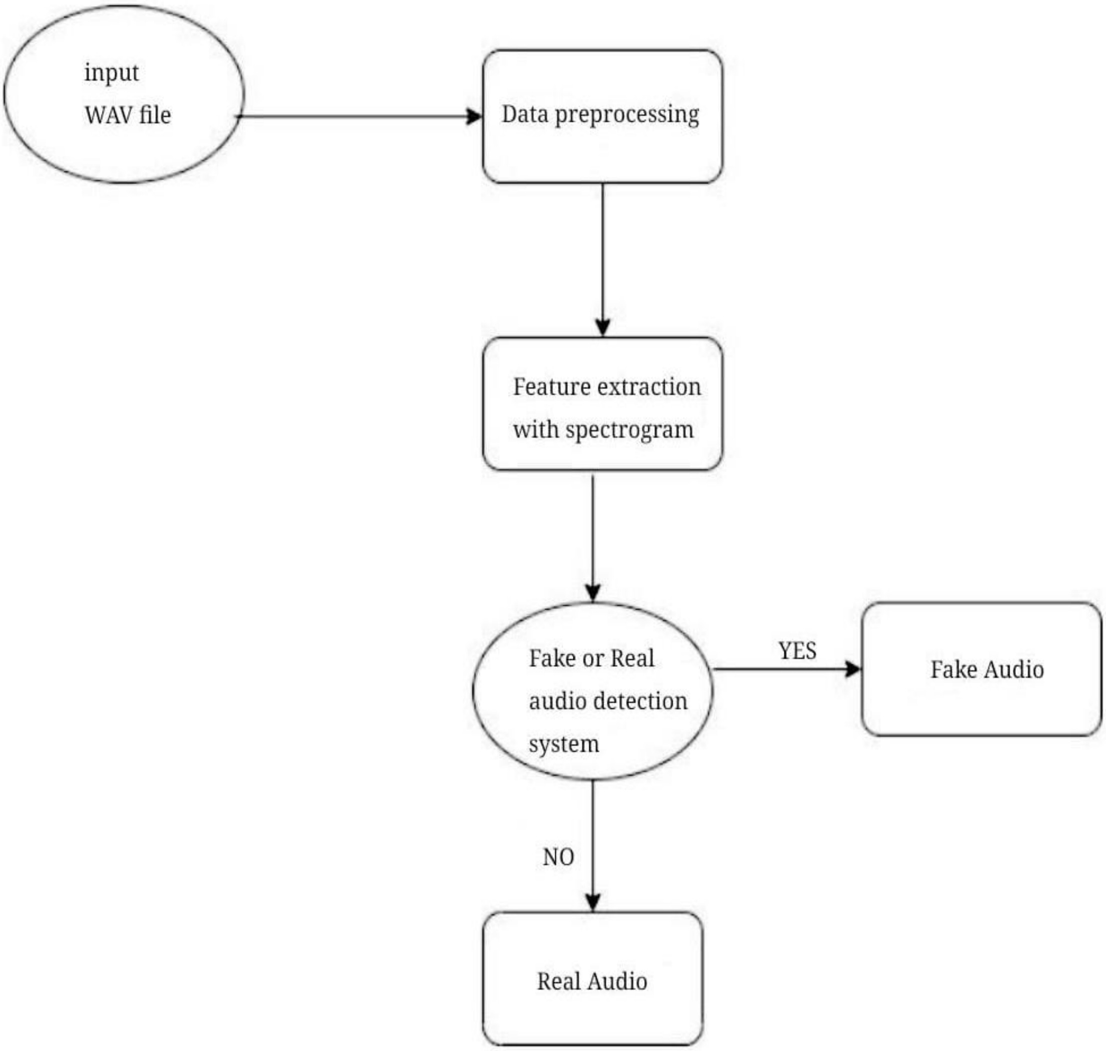

Fig 1. Data flow diagram

IV. CONCLUSION

In conclusion, the adoption of deep learning techniques for deepfake audio detection presents a promising avenue for enhancing the accuracy and robustness of detection systems. By leveraging neural network architectures and learning directly from raw audio data, the proposed system shows significant potential in effectively distinguishing between genuine and manipulated audio recordings.

Transfer learning and adversarial training further bolster the system's performance and adaptability to emerging threats. Through these advancements, the proposed system contributes to the ongoing efforts in combating the proliferation of deepfake content, safeguarding the integrity of multimedia communication channels in the digital age.

REFERENCES

[1]. Hamza, Ameer, et al. "Deepfake audio detection via MFCC features using machine learning." IEEE Access 10 (2022): 134018-134028.
[2]. Abbasi, Ahmed, et al. "A large-scale benchmark dataset for anomaly detection and rare event classification for audio forensics." IEEE Access 10 (2022): 38885-38894.
[3]. Javed, Abdul Rehman, et al. "A comprehensive survey on digital video forensics: Taxonomy, challenges, and future directions." Engineering Applications of Artificial Intelligence 106 (2021): 104456.

DOI: 10.17148/IJARCCE.2024.13344

[4]. Almutairi, Zaynab, and Hebah Elgibreen. "A review of modern audio deepfake detection methods: challenges and future directions." Algorithms 15.5 (2022): 155.

[5]. Reimao, Ricardo, and Vassilios Tzerpos. "For: A dataset for synthetic speech detection." 2019 International Conference on Speech Technology and Human-Computer Dialogue (SpeD). IEEE, 2019

[6]. Oord, Aaron van den, et al. "Wavenet: A generative model for raw audio." arXiv preprint arXiv:1609.03499 (2016).

[7]. Kapka, Sławomir. "ID-conditioned auto-encoder for unsupervised anomaly detection." arXiv preprint arXiv:2007.05314 (2020).

[8]. Wu, Zhizheng, et al. "ASVspoof 2015: the first automatic speaker verification spoofing and countermeasures challenge." Sixteenth annual conference of the international speech communication association. 2015.

[9]. Tom, Francis, Mohit Jain, and Prasenjit Dey. "End-To-End Audio Replay Attack Detection Using Deep Convolutional Networks with Attention." Interspeech. 2018.

[10]. Abbood, Zainab Ali, et al. "Speaker identification model based on deep neural networks." Iraqi Journal For Computer Science and Mathematics 3.1 (2022): 108-114.

[11]. Ping, Wei, et al. "Deep voice 3: Scaling text-to-speech with convolutional sequence learning." arXiv preprint arXiv:1710.07654 (2017).

[12]. Khanjani, Zahra, Gabrielle Watson, and Vandana P. Janeja. "How deep are the fakes? focusing on audio deepfake: A survey." arXiv preprint arXiv:2111.14203 (2021)