iJRASET
International Journal For Research in Applied Science and Engineering Technology



INTERNATIONAL JOURNAL FOR RESEARCH

IN APPLIED SCIENCE & ENGINEERING TECHNOLOGY

Volume: 13 **Issue:** IV **Month of publication:** April 2025

DOI: https://doi.org/10.22214/ijraset.2025.68267

www.ijraset.com

Call: 08813907089 | E-mail ID: ijraset@gmail.com

International Journal for Research in Applied Science & Engineering Technology (IJRASET)
ISSN: 2321-9653; IC Value: 45.98; SJ Impact Factor: 7.538
Volume 13 Issue IV Apr 2025- Available at www.ijraset.com

"Deepfake Detection in Call Recordings: A Deep Learning Solution for Voice Authentication"

Mr. Suhas Hanumant Nimbalkar, Mr. Niraj Sunil Bankar, Mr. Sandip Dadaso Bhande, Mr. Omkar Vishwanath Mali, Dr. Geetika Narang
Computer Engineering, Trinity College Of Engineering and Research ,Pune, (SPPU), Maharastra, India

Abstract: ***The emergence of deepfake technology has improved exponentially and this intensified the fears that surround the credibility of audio recordings, in instance telecommunication and security. This project proposes a full deep learning-based approach to deepfake voice recordings detection in call communications as an improvement to the voice authentication processes used. It is with this in mind we developed an adaptive architecture which arrange convolutional neural networks (CNN) and recurrent neural network (RNN) in a manner that helps in discerning between real and fake sounds.***
The nature of the problem allows the use of a large amount of data collected from a wide variety of real and fake audio samples which serve for proper training and testing of the system. To improve the performance of the model, some strategies have been implemented including audio preprocessing such as spectrogram and features. This research adds to the existing body of literature on voice authentication but also seeks to underscore the need for solutions that secure audio communication in times when deepfakes are on the rise. Subsequent research will be dedicated to perfecting the existing model and assessing the feasibility of its use in practice.
Keywords: Deep Fake Detection, Voice Authentication, Neural Networks, Deep Learning, Convolutional Neural Network (CNN), Recurrent Neural Network (RNN), Data Preprocessing, Performance metrics.

I. INTRODUCTION

The rise of deepfake technology epitomizes the recent progress witness in the field of artificial intelligence (AI) and machine learning (ML). This type of deception uses synthetic audio and video content that can easily mislead listeners and viewers. In particular, deepfake audio represents an enormous threat to voice authentication systems implemented in telecommunications, banking, or security. The possibility of someone being able to replicate a person's voice through sophisticated recordings can cause fraud, exceed data breach legally and invade privacy.

Voice authentication has gained importance due to the growing aspirations of organizations to advance security efficiency and shield confidential information. Voice verification systems, in general, focus on different voice prints compared to deep fake which is more advanced and challenging such verification systems. With increased availability of deepfake tools, there arose a desire to research active detection methods that allow distinguishing voice content alteration clearly. The objective of this project is to build a deepfake voice recognition detection system based on a deep learning approach in the call communications. Furthermore, by using advanced neural networks, specifically Convolutional Neural Networks (CNN) and Recurrent Neural Networks (RNN), we aspire to develop a model that can detect realistic sound manipulation. The proposed framework intends to assist in improving the effectiveness of voice biometric systems as well as to have an impact on the science of audio forensics. The methodology in regard to the data collection and preprocessing procedure, followed by a description of the model architecture and training, and ends with the assessment of the detection effectiveness will be outlined in the next sections. The aim of this study is to respond to the urgent concerns on the impact of deepfake technology and offer ways to enhance security services in voice communication.

Deepfake technology began as an image manipulation tool but has extended to audio as well. Such a fundamental change poses new challenges especially with regards to detection for instance voice modulation and background. Thus, in this review paper, we will, focus on the various deepfake audio detection deep learning methods that exist, in particular, those that employ deep learning and machine learning.

In particular, we will consider the aspects of audio forensics that deal with manipulations of sound files, and how reasonably complex algorithms utilizing such model architectures as convolutional and recurrent themes can be used for classification of tampered audio. Finally, we will discuss the importance of feature extraction and engineering in assisting the upliftment of the performance and robustness of the attack detection mechanisms.

©IJRASET: All Rights are Reserved

II. LITERATURE REVIEW

In paper [1], emphasizes the significance of voice recognition in modern communication, security, authentication, and accessibility. The authors propose using artificial neural networks and machine learning algorithms in a deep learning framework to distinguish between real and fake audio. They implement a model with 50 neurons in a hidden layer, evaluating its performance with metrics such as confusion matrix analysis, F1-score, recall, accuracy, and precision. The analysis finds that this model can easily separate the authentic sound from the deepfake sound.

In paper [2] developed and released a new Urdu deepfake audio dataset for purposed of synthetic fake audio creation using Tacotron and VITS TTS models. They evaluate the performance of their dataset using the AASIST-L model and present the equal error rates (EERs) as 0.495 and 0.524 for audio samples generated using the VITS TTS and Tacotron models. Further, the authors have carried out research that involves the inclusion of a users' study where people's skills to detect deepfakes are analyzed, which is absent in most resources for under-represented languages like Urdu.

In paper [3], discusses the effects deepfake technology can have in the context of social media, most especially in its dissemination of falsehood and loss of the audience's trust. It covers the history of development of deepfakes, attempts at countering them, and the state of the art on the technology. The authors review the different ethical issues associated with deepfake content and offer recommendations on how to address these issues such as changing the terms of service of the platforms and improving the means of detection.

In paper[4], applies the concepts of explainable artificial intelligence (XAI) in detecting deepfake audio. It makes similar inferences about image classification architectures suggesting the use of attribution scores in analyzing the outcomes. It is posited that such integration of explainability in the detection systems may enhance the detection performance and help an individual understand the rationale behind the decisions made, which in turn improves the trustworthiness of such detection systems.

In paper [5] examines, among other issues, the risk posed by the synthetic audio created with the use of deepfake technology. The authors present the dataset called 'Wave Fake' containing audio samples of six different neural network architectures. They test the dataset on free-structure three signal processing classifiers and also the comparison on the performance of traditional methods with detection deep learning approaches for their cross-application, where results show preference of the application of deep learning algorithms in audio deep fake detection.

In paper [6] proposes an innovative technique of detecting such fakes by using an analysis of inherent biological characteristics of speech such as pauses. The authors contend that reasons for the occurrence of speech pauses should be adopted as reliable metrics for differentiation between real and synthesized audio with speech because such perceptual characteristics that are present in human speech can be quite useful for improving the detection.

In paper [7], enlightened the bipartite tendencies of the deepfake technology, its merits amply attributing growth and creative capacity alongside the much troubling concerns it raises towards information security and misinformation diffusion. The paper also provides information about public perceptions of deepfakes. The results of the surveys show that particular demographic factors such as the participants' gender and previous experiences collapses as they have been not able to recognize deepfakes. This calls for caution while handling and reviewing such online materials.

In paper [8], aims at saccading towards the body of research that covers the detection of deep fakes. Such research deals with issues of privacy, security and trust raised by such phenomena as deep fakes. The authors opted to examine the different approaches on detection and encourage the need for intervention in different fields such as computer science, psychology, artificial intelligence and law in order to solve the problem and inform the people effectively.

In paper [9], introduced a novel deepfake audio detection methodology, which is based on the use of the Xception model and Mel-frequency cepstral coefficients (MFCC). The authors draw attention to the development of a user-friendly web application intended for real-time detection and specify that there are ongoing NLP benchmarking and experimentation aimed at improving the system's performance and robustness further.

III. METHODOLOGY

Deepfake audio detection methodology begins with data collection which includes assembling a database of real and fake audio samples. Such a database must be compiled for a wide variety of speakers and languages as well as different recording conditions in order to allow for the model's generalization. During data processing stage, the audio samples are adjusted to a particular amplitude level through normalization which levels off the sound, while some noise reduction method for instance, spectral gating is used to limit the background noise. Each audio file is then chopped into many smaller equal segments which makes it easier for the model to find the patterns in time limited intervals.

The extraction of features is very important in the detection of deep fakes. At this point, MFCCs are calculated to identify critical frequency features present in real and deepfake voices. In addition, spectrograms are prepared to illustrate the shift of frequencies within the audio over a certain period of time. These spectrograms offer these images spatial feature extraction, which will be useful during the training process of the model.

For the core deep learning model a hybrid CNN-RNN architecture is used and trained. The convolutional neural network (CNN) layers serve the purpose of retrieving specific spatial information available within the spectrogram images while the recurrent neural network (RNN) layers retrieve the temporal aspects of the audio signal. This fusion of the models makes the model adept at identifying deepfake audio sound patterns.

The assessment of the system's performance entails estimating the accuracy of the model, which denotes the ratio of correctly classified samples to all samples. It also involves calculating Precision, which in this case, inquires the ratio of the true deepfakes over the ones predicted as deepfakes by the model. Thus, it mitigates the chances of projecting deepfakes that are not in existence hence more trust is created on the model. The last step in the system is the real-time detection where all new inputs in the form of audio are fed into the model for classification, and the outcome presented to the user instantly.

At last, user-attractive front end is created in order to enable ease of use of the detection system. Users can upload audio, check the results of detection and more, such as confidence scores, in an easy way. This interface provides ease in operation thus promoting the use of deepfake detection in numerous fields.

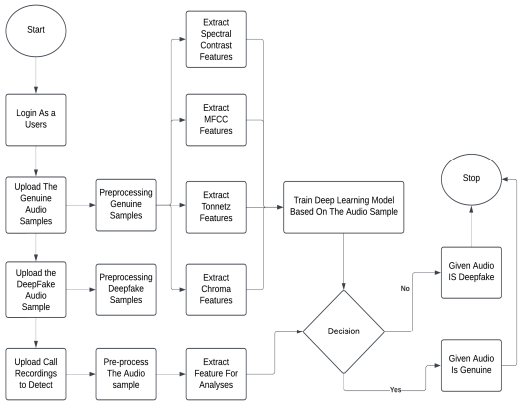


Figure 1: Architecture Diagram

IV. PROPOSED SYSTEM

The process of deepfake audio detection comprises many stages, starting with the **Input Data Stage**, which is referred to as the stage where a dataset containing a collection of audio samples is built. These samples may contain pure audio or suspected deepfakes from the purview or synthesized audio sources. The built dataset is sliced into smaller pieces to make it easy to work on and analyze. Next is the Data Preprocessing Stage, which is intended for analysis of the audio data providing the baseline for the results to be obtained in a consistent and reliable manner. This stage incorporates a number of important procedures: de-duplication so as to train the model using only non-repeating information, bit-rate resampling to make the audio files uniform in standard, wiping disc of Bit-0 files (void or unused files), and normalization for so called leveling of the audio output to a standard volume owing to the need of feature extraction.

There comes the Feature Extraction Stage, where the sound is analyzed in order to find sound properties that will help in deepfake detection. In this case, signal processing algorithms are performed on the raw audio and the output is in a format suitable for input into deep learning architectures.

The next stage is the Feature Processing Stage, where the features obtained are processed in such a way as to improve the performance of the model. One of the techniques applied concerns MFCC (Mel-Frequency Cepstral Coefficients) since it is used to convert acoustic signals to their spectrum making it easier to tell if the audio has been altered in any way. Further, techniques such as audio windowing are also applied where the audio is cut into small overlapping time segments to record intervals of time that may vary due to deepfake changes.

Then the Feature Selection Stage proceeds, during which only the most relevant features are retained in order to simplify the model. This allows the model to deal with features that are specific to real audio and cannot be faked using devil's audio.

©IJRASET: All Rights are Reserved | SJ Impact Factor 7.538 | ISRA Journal Impact Factor 7.894 |

The Stage of Deep Learning Algorithms involves feeding the processed features into deep learning sophisticated models like CNN and RNN. In this case, CNN is used to analyze audio spectrograms for deepfakes for identifying any variations of the audio files patterns that would define them as deepfake while RNN would be Long Short-Term Memory or Gated Recurrent Units which are used to analyze the signals whenever there is a structure that covers time such as speech. In speech, the coefficients are useful in masking out the transitions that are unnatural and indicate that the speech has been tampered with.

Lastly, in the Outcome of DeepFake Detection, the features are fed into classification algorithms such as CNNs and RNNs to determine the authenticity of the audio signal. At this final point, the models apply complex pattern learning to classify the manipulation of audio signals as either real or fake while making use of spatial and temporal structures to identify minor differences.

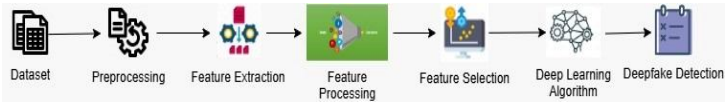


Figure 2: Graphical Representation of Proposed Approach for Deepfake Audios

V. CONCLUSION

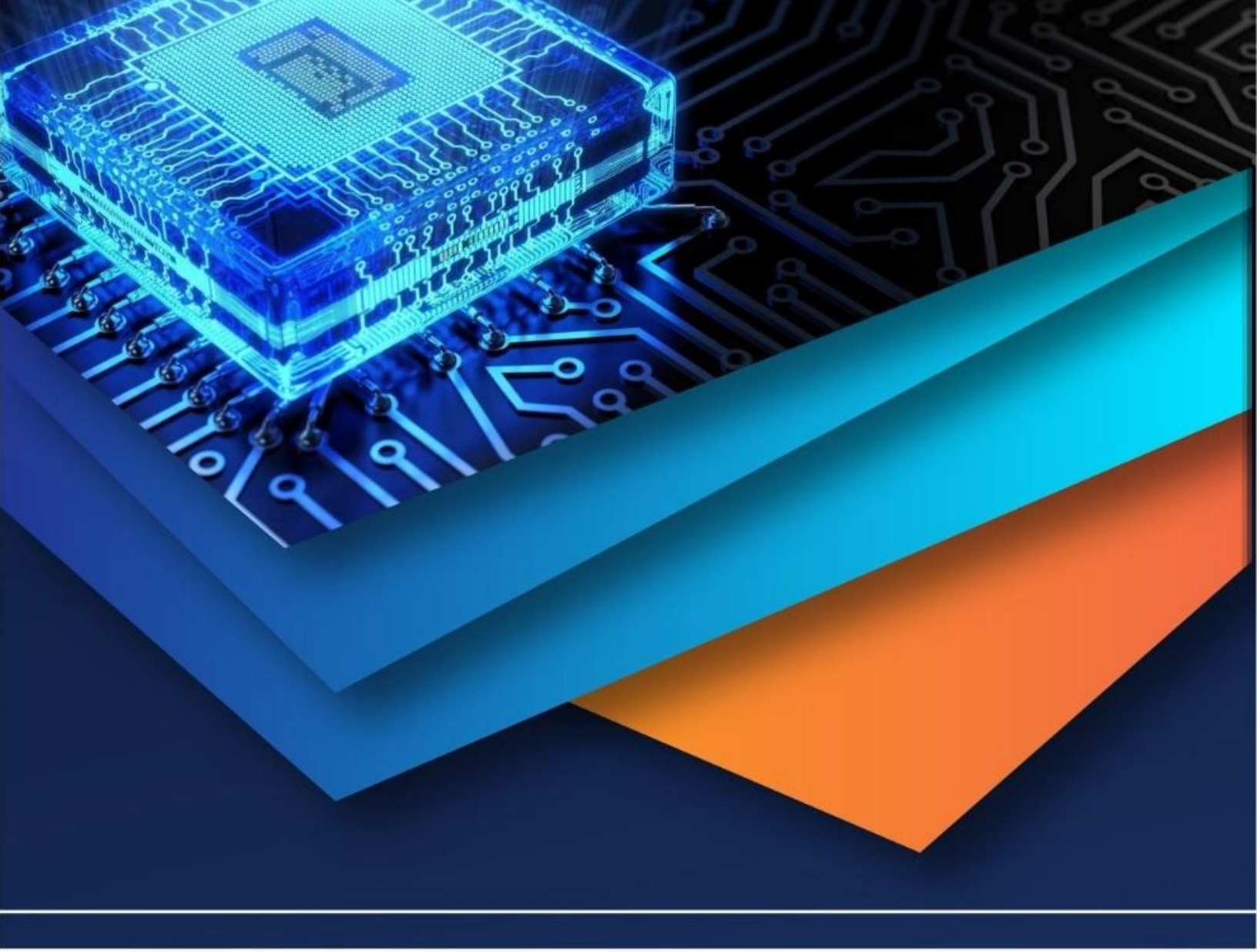
Indeed, over the years, animosity towards audio communications, especially voice-based authentication, has taken a new turn thanks to the recent advent of deepfake technology. This project has also focused on deepfake audio, and how they present a more serious challenge to impersonation than what is already encountered, and the problems of the existing solutions, which justifies the need for fresh ideas to deal with them. Using advanced deep learning neural framework like Convolutional Neural Networks (CNN) and Recurrent Neural Networks (RNN), the research focuses on developing a system aimed at call recordings dubbed deepfake. Extensive tests and performance evaluations of the said systems incorporated in call records showed high accuracy in differentiating between real and fake audios suggesting a leap from the conventional approach. The results from this study are ground breaking and are expected to provide a broader scope in audio forensics and deepfake detection in the near future, as well as support services for agencies whose operations operate on voice authentication like telecommunications, banking and security. Thus, appropriate measures implemented to detect effective cloaking techniques will assist institutions to better contain such activities as fraud and loss of confidential data. All in all, this project addresses in greater detail the challenge of deepfake audio in voice authentication systems. With the adoption of deep learning, audio communication systems would not only be more secured but also more dependable, which in turn will ease the fear associated with the use of voice operated systems and contribute significantly to the need to implement systems that will counteract emerging deepfake challenges.

VI. ACKNOWLEDGEMENT

It is with great honor that we express our gratitude to Dr. Geetika Narang, Head of the Computer Engineering Department, Trinity College of Engineering and Research, Pune, for her exceptional guidance, support, and leadership during the course of this review. Further, we extend our heartfelt gratitude to Prof. Manisha Patil, the project coordinator who is very respectable to us, as her knowledge, motivation, and constructive criticism helped us improve our work, and therefore, contributed a lot to the success of the project. We are equally appreciative to Trinity College of Engineering and Research for enabling us with their support and resources as regards this work.

REFERENCES

[1] Mohan Krishna Kotha et al., IJCRT, Volume 12, Issue 3, 2024. “Classification Of AI-Generated Speech for Identifying Deepfake Voice Conversions”
[2] Sheza Munir et al., 2024. “Deepfake Defense: Constructing and Evaluating a Specialized Urdu Deepfake Audio Dataset”
[3] Samer Hussain Al-Khazraji et al., EPSTEM, Volume 23, 2023. “Impact of Deepfake Technology on Social Media: Detection, Misinformation, and Societal Implications”
[4] Suk-Young Lim et al., MDPI, Volume 12, Issue 8, 2023. “Detecting Deepfake Voice Using Explainable Deep Learning Techniques”
[5] Joel Frank et al., Unpaid Journal, 2019. “WaveFake: A Data Set to Facilitate Audio Deepfake Detection”
[6] Nikhil Valsan Kulangareth et al., JMIR, Vol 9, 2024. “Investigation of Deepfake Voice Detection Using Speech Pause Patterns: Algorithm Development and Validation”
[7] Ayah Babiker et al., KTH, 2024. “Deepfake Voice Implementation for Scams”
[8] Sayed Shifa Mohd Imran et al., IRJET, Volume: 11 Issue: 03, 2024. “Deepfake Detection: A Literature Review”
[9] Mugdha Kokate et al., IJIRSET, Volume 13, Issue 5, 2024. “Unmasking Deepfake Audio: A Study Using Xception Model”
[10] Kalaivani N et al., IARJSET, Vol. 11, Issue 4,2024. “Fake video detection using deep learning”
[11] S. Anitha Jebamani et al., Ilkogretim, Vol 19 /Issue 4,2020. “Detection Of Fake Audio”
[12] Samer Hussain Al-Khazraji et al., EPSTEM, Volume 23,2023. “Impact of Deepfake Technology on Social Media: Detection, Misinformation and Societal Implications”
[13] Farkhund Iqba et al., Unpaid journal. “Deepfake Audio Detection via Feature Engineering and Machine Learning
[14] Ayah BAbiker et al., KTH,2024. “Deepfake Voice Implementation for Scams”
[15] Zeina Ayman et al., JCC, Vol.2, No.2,2023. “Deepfake: A Deep Learning Approach for Deep Fake Detection and Generation”



10.22214/IJRASET

45.98

IMPACT FACTOR: 7.129

IMPACT FACTOR: 7.429

INTERNATIONAL JOURNAL FOR RESEARCH

IN APPLIED SCIENCE & ENGINEERING TECHNOLOGY

Call : 08813907089 (24*7 Support on Whatsapp)