IJARSCT
International Journal of Advanced Research in Science, Communication and Technology
International Open-Access, Double-Blind, Peer-Reviewed, Refereed, Multidisciplinary Online Journal
ISSN: 2581-9429
Volume 5, Issue 8, May 2025
Impact Factor: 7.67

Predictive Techniques for Recognizing and Classifying Real and AI-Generated Voices

Manali Shukla[1,2*], Anjali Saraswat[2,3†], Soumik Mallick[1,2†], Maharaja Sagar Mishra[2,3†]
[1*]Computer Science and Engineering Technology
[2,3,4]Computer Science and Engineering Technology
ITM University, Turari, Gwalior, Madhya Pradesh, India.
*Corresponding author(s). E-mail(s): saraswatanjali455@gmail.com;
Contributing authors: shukla.manali2014@gmail.com ; mallicksoumik1711@gmail.com;
maharajasagarmishra@gmail.com;
[†]These authors contributed equally to this work

Abstract: *Rapid growth in the technology specially in the field of artificial intelligence has transformed this world with the advent of generative artificial intelligence. Numerous applications which are easily accessible over the internet based on gen- erative artificial intelligence models presenting challenges in front of researchers and security experts. In the present era of time , a variety of AI-generated tools are utilized to generate fake audio , image , video etc leading to a new challenge known as deep fake detection . However, the motives behind the use of generative artificial intelligence are not destructive but unfortunately misused. Tremendous growths in the cyber-attacks employing fake audio have been reported in the present era of time that raising the concern of security experts and researchers. Therefore, this is evolving as challenge in field of research to address recognition of fake voice and which motivates this research work. This research paper pro- poses an effective methodology to detect real and ai voice using artifact score and which shows promising results and comprehensive review of the research work done to address such type's deep fake detection.*

Keywords: Artificial Intelligence, Artifact, Cyber-Attack, Deepfake, Generative AI

I. INTRODUCTION

Currently, artificial intelligence is prevalent in almost every area of our daily lives, whether personal or professional. Voice technology has come a long way since its incep- tion. One area that is still seeing fresh and advances in the context of AI is the voice over industry. Nowadays, we can't seem to get away from robotic voices and virtual assistants that sound like they came directly out of a science fiction film. But how do AI-generated voices compare to a human's? Is there any discernible difference between the two? Generative AI is subfield of AI technology that generate the new text, audio, video , image, document , music, programming code , data , poetry ,conversational dialogues etc. all are things are deep fake its generated by AI. The word "deepfake" is derived from the underlying technology, "deep learning," which is a type of AI. Deep learning algorithms, which train themselves how to solve problems when given enormous amounts of data (audio, video, text, voice), are used to create the most realistic speech with AI audio using data and different languages for making video, animation, scenes and digital content, producing realistic-looking false media. More- over there is a lot of enormous volumes of voice recordings are shared online, making it difficult to distinguish fake content among them. Audio deepfake is employed in a variety of applications, one of which being financial scams. Audio deepfakes have previously been used to clone voices, deceive people into thinking they are speaking with someone trustworthy, and defraud them. Earlier this year, scammers attempted to persuade an employee of a technology company to send money to the scammer's account by using a deepfake of the CEO's voice. This is not the first time scammers have used the exact same tactic to defraud a firm of 240, 000. [1] This has led to mass public concern with the negative impacts of deep fakes in cyber security. However, this

IJARSCT
International Journal of Advanced Research in Science, Communication and Technology
International Open-Access, Double-Blind, Peer-Reviewed, Refereed, Multidisciplinary Online Journal
ISSN: 2581-9429
Volume 5, Issue 8, May 2025
Impact Factor: 7.67

technology has been proven to be efficient hence improving audio deep fakes as compared to simple text, emails, and email link. This makes it possible for someone to use this logical access audio-spoofing method [2], This opens the door to propa- ganda, slander, and even terrorism as methods of swaying public opinion. Detecting fakeness in massive amounts of audio recordings published online every day is tough. [3]. However, leaders and governments are not safe to deep-fake attacks. [4] To this end, it is becoming more challenging to detect fake audio. Deep fakes have emerged in three major types: which is based on artificial data, fake voices, and repeated data
. One of the types of deepfake is human audio data that is generated by AI . There are various techniques which assist in isolating real speech being found in an audio recording from some other sounds. Different methods have been applied in establish- ing DL and ML models for false audios recognition. Since then, there are still many gaps in those proposed algorithms. We all have come across the term Deepfake all over world, recently methods that are synthesized using AI have been developed to produce voices that are real. But although these technologies were created to assist people, they have also been utilized to use audio to disseminate false information globally.
[5], Fear of the ”Audio Deepfake” has been stoked by its malicious use. Simple mobile devices or personal computers are increasingly becoming capable of producing audio deep fakes [6], also known as audio manipulations or voice cloning at the beginning of early 2000's. There exists three types of deepfake namely deepake vision, deepfake audio and deepfake(vision + Audio) that is product of AI, which is a blooming tech- nology nowadays. In the real world, AI is used in many field like transportation, retail, energy, government, agriculture, healthcare, manufacturing and production and so on. Cyber Security is the practice of protecting user's computer systems and their personal data from malicious attacks. Therefore , It has become a World-wide important for data security and privacy. Many researches had already provided various work done in providing security to individuals data. One of the type cyber-attack phishing that has been rapidly increasing as technology advances, and to a great degree, phishing is driven by the advancement of social networking technologies [7]. In this cyberse- curity attack, the fraudster deceptively obtains sensitive information by imitating a reliable or trustworthy third party. Attackers can use such information for criminal activity like identity theft. There is one a type of phishing is Voice Phishing which is also referred as Vishing. It is an attack where the attacker tries to access the creden- tials details of the user using audio. These attacks are conducted using text to speech systems where artificial intelligence is used to convert the text into speech. Figure 1 illustrates the diagrammatic view of general classification model used in ai voice and real voice detection which involved preprocessing and training phase to build the classification model in order to classify input into real and fake voices.

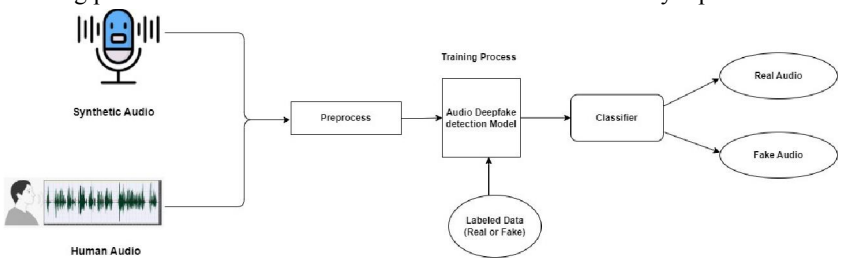

Fig. 1 General Classification approach for fake voice detection

II. LITERATURE REVIEW

AI trends in the present generation lead to deep fakes that are perilous to society, pol- itics, and data authenticity [8]. These manipulations can be performed on any content that is either in the form of image, video, audio, or text form, hence raising the need for effective detection. It has been noted that with deep learning methods especially in deep learning, deep fake detection has been shown to be more effective than tradi- tional methods [9]. For example, in VGG19 architecture, they got the 95% accuracy in face detection operations [10]. For audio deepfakes that threaten the use of voice inter- faces and speaker verification systems, large margin cosine loss function, and online frequency masking augmentation improve the detection of deep fake by lowering the equal error rate to 1.26% [11]. Due to the advancement in deepfake technology, there is emerging demand for integrated, timely, and flexible technologies for detection with more awareness and legal measures among the general population [8][9][11]. Voice con- tent deepfakes

IJARSCT
International Journal of Advanced Research in Science, Communication and Technology
International Open-Access, Double-Blind, Peer-Reviewed, Refereed, Multidisciplinary Online Journal
ISSN: 2581-9429
Volume 5, Issue 8, May 2025
Impact Factor: 7.67

are challenging to detect or generalize when such AI techniques are used to synthesize or modify audio [12]. Several works has been done in recent years and to specifically study the possible ways of improving detection of audio deepfakes. For instance, estimating Constant-Q Transform (CQT) or using log-spectral features instead of Mel-spectral features can enhance the performance by 37 in terms of Equal Error Rate (EER) [13]. Yet, it has been pointed out that most of the offered methods are inefficient in certain real-world circumstances and therefore require more potent approaches. To overcome such challenges, as proposed in one of the methods, a large margin cosine loss was used with online frequency masking augmentation, resulting, into equal error rate of 1.26% on the ASVspoof 2019 database. In one of their recent creative work, Almutairi and Elgibreen 2023 proposed a new self- supervised learning known as Arabic-AD, which was far much more effective than the above mentioned benchmarks providing 0.027% EER and 97% ASR [15]. In addition, there has been a systematic analysis of different detection architectures with studies identifying that some preprocessing can be beneficial [16]. However, some of the current methods faced difficulties in working with real data, which creates questions about overfitting to ref- erence datasets. A global classification of adversarial attacks on Automatic Speech Recognition systems has also been created alongside an evaluation of current threat detection measures [17]. Another work incorporated a self-supervised learning method to detect fake and imitated Arabic speech that has high accuracy without relying on large datasets [18]. Additionally, a comprehensive tutorial for laypeople on how to detect audio deepfakes appeared recently, discussing handcrafted features, from end-to-end models and deep learning by themselves [19]. In this context, authors in
[20] presented a study on ”Audio Deepfake Detection, their work implies adversarial attacks both on AI-based audio authentication systems, and specially the Deep4SNet classifier, using generative adversarial networks (GANs). However using classifiers for deepfake audio samples detection, has high classification accuracy, but these classifiers are vulnerable to adversarial attacks that can significantly decrease the classifica- tion accuracy to almost zero [20]. Features like MFCC have been utilized along with SVM and VGG-16 models; The architectures provided encouraging outcomes with nearly every database [21]. Moreover, spectrograms and chromograms have been used in order to distinguish near-real and altered voice [22]. Audio deepfakes, especially at modifier-based, increases efficiency in applications such as audiobooks; however, they pose enormous security threats and subsequently, the need to detect both imita- tion and synthetic-based deepfake effects. Other recent research also works on further developing the methods to detect deepfakes from the audio domain. For example, timbre and shimmer features in a time-domain were introduced by authors who sug- gested they are informative for the discrimination between synthetic and human voices [23]. Another study introduced ABC-CapsNet, which combines Mel spectrograms with VGG18 and cascaded capsule networks, achieving impressive equal error rates on mul- tiple datasets [24]. Moreover, a 34-layer ResNet with multi-head attention pooling and neural stitching has shown strong performance in the ADD 2022 challenge [25], while wav2vec 2.0-base features with ECAPA-TDNN have been utilized for deepfake algo- rithm recognition, incorporating data augmentation to enhance model generalization [26]. These studies collectively illustrate the diverse approaches being explored in audio deepfake detection, ranging from acoustic feature analysis to advanced neural network architectures, all aimed at improving the accuracy and reliability of synthetic speech identification systems. Authors in [27] proposed a method with use of deep learning which trains a convolutional neural network (CNN) that analyses the visual represen- tation of sound frequency known as spectrogram. The approach suggested describes that the cloned samples generated by any model shows some continuous patterns that are not present in the samples. These patterns are the vertical line that is visible via spectrogram. Convolutional neural network in deep learning are mostly used for tasks which requires image classification, it widely uses 2D or 3D image samples.[27]. Another study Author in [28] suggested a model named deepsonar to overcome this issue. The model proposed by authors in [28] is deepsonar which works on the move- ment of neurons (artificial) in deep neural network (DNN) during speech recognition. This model works on the pattern of input at different layers of DNN. Authors also discussed about other methods like Hidden Markov Model, Gaussian mixture model which learns and replicates the features of speech. Another study focuses on the chal- lenges produced by these artificially generated voices in the society which is similar to humans and difficult to differentiate which further could use to spread false informa- tion. To handle this issue Authors in [29] developed a method to recognize this fake synthesized voices by searching for specific signs. In this paper , authors conducted test on existing model named rawnet2 by training it to understand vocoder artifacts. Authors in [30] used

IJARSCT
International Journal of Advanced Research in Science, Communication and Technology
International Open-Access, Double-Blind, Peer-Reviewed, Refereed, Multidisciplinary Online Journal
ISSN: 2581-9429 Volume 5, Issue 8, May 2025 Impact Factor: 7.67

natural biological features to tell the differences in between real and artificially generated voice. According to the research, it's really not possible for artificial intelligence to completely clone a voice there are some features that are really difficult to clone like natural pauses for thinking or breathing and taking break in between sentences etc. the authors collected 49 audio samples of means with different accent. These audio samples are then used to train 3 voice cloning models to pro- duce synthetic voice. Authors compared the generated cloned voice and real voice by identifying features like pauses in between speech and tone of voice. By using this data 5 machine learning models were trained to predict and analyze the real and fake voice.[29][30]. The authors in [31] explained in a detailed manner about a model which is built to differentiate between fake and real speeches and to identify who the real speaker is by their voice samples. Authors also submitted their model in a compe- tition called ASVspoof5 which has primarily two tasks first is to spot fake and real voices and another is to identify the person from their voice samples. Researches in [32] developed a model that works on special module called SLS (sensitive layer selection). This module helps the model to extract the important features from an audio sample by analyzing different layers in the provided sample. Model is based on a pre-trained system XLS-R which is used to extract different features from an audio sample. By using SLS module, authors in [32] provided evidence that this method performs better than existing ones and provide a good clarity in separating fake voices with real ones. All the research work done so far emphasizes for the necessity to develop an effective methodology to determine the difference between ai generates voice and human voices . In continuing efforts , the next section of this research paper proposes a methodology to differentiate between human and ai generated voice. Next section discussed the distinctions between audio, video, and hybrid deepfakes.

III. TYPES OF DEEP FAKE CATEGORIES

Deepfake generation is widely classified into four major study areas: face swap, which focuses on the complete replacement of one person's face with another; reenactment, which involves altering facial expressions to mimic specific actions or gestures; gesture synthesis, which emphasizes replicating physical behaviors; and lip-syncing, where mouth movements are synchronized with the accompanying text or audio content. Deepfake content can be categorized into image, video, and audio deepfakes. Although image and video deepfakes are distinct, video content fundamentally consists of a sequence of images. This paper focuses specifically on managing audio deepfake content, with an emphasis on applications such as speech synthesis and voice swapping [34], [35].

IV. PROPOSED METHODOLOGY

Numerous researchers have presented various types of methodologies to detect and dif- ferentiate between AI generated voice and human voice. There are various distinctions between the voices generated by artificial intelligence (AI) and a normal human voice, most notably in the methods of generation and innate characteristics. Some of the key features that could differentiate between AI voice and human voice include flexibil- ity or adaptability .Given the right circumstances and intentions, the pitch, tone, and emotion of a human voice can change dramatically. Even while artificial intelligence voices have come a long way in recent years, they might not be as expressive and natural-sounding as human voices. Another aspect of differentiation can be emotional intent in voice. Human voices contain emotional nuances and clues that express atti- tudes, intentions, and sentiments. Although AI-generated voices can mimic emotions to some degree, their expressions could not be as rich or authentic as those made by humans. There also exists imperfection in human voices which is natural and makes it different from AI generated voice Natural voices are distinctive and human-like because of their flaws, which include stutters, hesitations, and breath noises. Artificial intelligence (AI) voices are typically more polished, which can occasionally make them sound less real. One more significant aspect of human voices is adaptability. Human voices may effortlessly transform between different languages, accents, or dialects, and adapt to varied social circumstances. Although artificial intelligence (AI) voices can be trained to mimic several languages and dialects, they might not be as flexible or aware of context as human speakers. Human are creative in their voice generation. Human voices are capable of imaginative wordplay, such as puns and jokes, as well as the use of comparisons and metaphors to explain difficult concepts. Although certain comedy and wordplay can be taught to AI-generated voices, their ability to be cre- ative may be constrained by the algorithms and training data they use. Since every human voice is different, with a distinct tone, pitch, and manner of

IJARSCT
International Journal of Advanced Research in Science, Communication and Technology
International Open-Access, Double-Blind, Peer-Reviewed, Refereed, Multidisciplinary Online Journal
ISSN: 2581-9429
Volume 5, Issue 8, May 2025
Impact Factor: 7.67

speaking. Artificial intelligence (AI) voices can be made to imitate a large variety of vocal traits, although they could not be as unique as human-voice. Notwithstanding these variations, the quality and naturalness of AI-generated voices is rising, closing the gap with that of real human voices. AI voices will probably start to sound more and more like human voices in the future due to advancements in AI research and technology. Therefore there exists a need to explore methodology that could help in detecting the difference between AI generated voice and natural voice. Process of fake voice detection involves the following steps as illustrated in the figure 2 given below.

In the data collection phase of fake voice detection , AI voice generated system uses significant amount of audio data with human voices is gathered in order to construct an AI voice. In order to provide diversity and richness in the training dataset, this data comprises a variety of speakers, languages, accents, and speaking styles. To eliminate any noise or unnecessary information, the audio data is processed and cleansed once

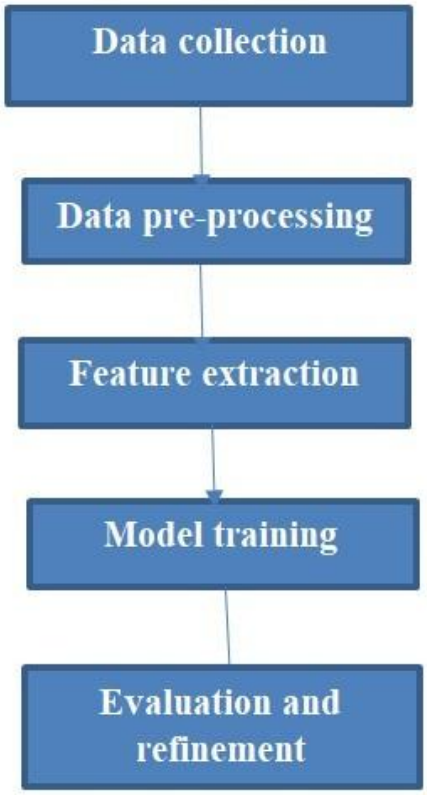

Fig. 2 Process of fake voice detection

it has been gathered. After that, the audio is divided into smaller segments—typically phonemes or words—and matched with the associated transcriptions. Pitch, timbre, and other spectral characteristics are taken from the pre-processed audio data and used to represent the fundamental aspects of the human voice. The processed audio data and attributes that are retrieved are used to train a machine learning model, which is often a deep neural network. Through the identification of patterns and relationships in the dataset, the model gains the ability to produce speech. Some popular models for AI voice generation are WaveNet, Tacotron, and FastSpeech. The AI-generated voice is evaluated for naturalness, intelligibility, and expressiveness using objective measures, subjective listening tests, or both.

This research paper presents a methodology that employs artefact- score .[33] This score evaluates the degree to which manipulation has altered an audio sample from typical human speech. Lower ratings can imply the voice is more natural or less affected, while higher scores might indicate notable deviations or abnormalities typical of synthetic voices. Pseudo code of proposed methodology is given below in Figure 3.

IJARSCT
International Journal of Advanced Research in Science, Communication and Technology
International Open-Access, Double-Blind, Peer-Reviewed, Refereed, Multidisciplinary Online Journal
ISSN: 2581-9429
Volume 5, Issue 8, May 2025
Impact Factor: 7.67

Algorithm 1 Detection of Fake and Real Voice Using Artifact Score
1: Input: Dataset containing fake and real voice audio data
2: Output: Label as real or fake voice
3: Step 1: Compute the mean (mean value) and standard deviation (std dev) of the audio data.
4: Step 2: Calculate skewness (skewness) and kurtosis (kurt) to capture non- normality and outliers.
5: Step 3: Calculate artifact score using a combination of statistics: 6: artifact score = calculate artifact score(audio data) 7: function CALCULATE ARTIFACT SCORE(audio data)
8: Read the audio data using read wav(audio data)
9: Compute: artifact score = std dev + skewness + kurt
10: return artifact score
11: end function
12: Step 4: Predict if the input voice is fake or real:
13: label = predict voice(audio data, artifact score)
14: function PREDICT VOICE(audio data, artifact score)
15: if artifact score is high then
16: label = "AI voice"
17: else
18: label = "real voice"
19: end if
20: return label
21: end function

Figure 3 : Pseudo code for fake audio detection

This methodology employs a dataset of 200 voices, consisting of 100 real voices and 100 fake voices. The real voices are collected from a university using in person interaction, with a quarter of them sourced from the standard Kaggle repository, whereas the fake voices are generated using AI tools. The input to the proposed system is provided in .wav format. Pre-processing is applied to the dataset to remove noise.

To calculate the artifact score, which is the sum of skewness, standard deviation, and kurtosis, the processed data is analyzed

Statistical Measures for Audio Analysis

I. Skewness

Skewness measures the asymmetry of a dataset's distribution around its center. A symmetrical distribution appears balanced on both sides of the central point. Positive skewness indicates a distribution with a longer or heavier tail on the right side, while negative skewness indicates a longer or heavier tail on the left side. In the context of audio features, skewness can reveal deviations in signal amplitude or feature distributions, such as irregular energy levels in AI-generated voices. High skewness in certain features may indicate potential anomalies or artifacts in the audio data.

II. Kurtosis

Kurtosis quantifies the "tailedness" of a dataset's distribution compared to a normal distribution. A higher kurtosis value signifies the presence of heavier tails, which correspond to more frequent extreme values or outliers. In audio processing, kurtosis can highlight the presence of unusual spikes or abrupt changes in acoustic features that are common in synthetic or manipulated audio. These characteristics help identify unnatural patterns in AI-generated voices, which often deviate from the smoother distributions found in real human speech.

III. Standard Deviation

Standard deviation measures the spread or variability of data points around the mean. A low standard deviation suggests that data points are tightly clustered, whereas a high standard deviation indicates greater dispersion. In audio analysis, standard deviation can capture variations in pitch, amplitude, or other features, providing insight into the consistency or irregularity of a voice sample.

IJARSCT
International Journal of Advanced Research in Science, Communication and Technology
International Open-Access, Double-Blind, Peer-Reviewed, Refereed, Multidisciplinary Online Journal
ISSN: 2581-9429
Volume 5, Issue 8, May 2025
Impact Factor: 7.67

Count_artifact_score() function extracts audio measurement results by evaluating skew, kurtosis, and standard deviation levels. The artifact score system detects AI-generated audio because synthesized speech shows distinct errors in its output. Real human voices show reduced artifacts in their performance and display optimal distribution patterns. When the calculated artifact score of the original audio goes beyond a certain level the model marks the content as AI-produced but labels it real if the score remains below the specified threshold. Our system platform uses Flask to let users upload audio files that the method then processes for prediction output.

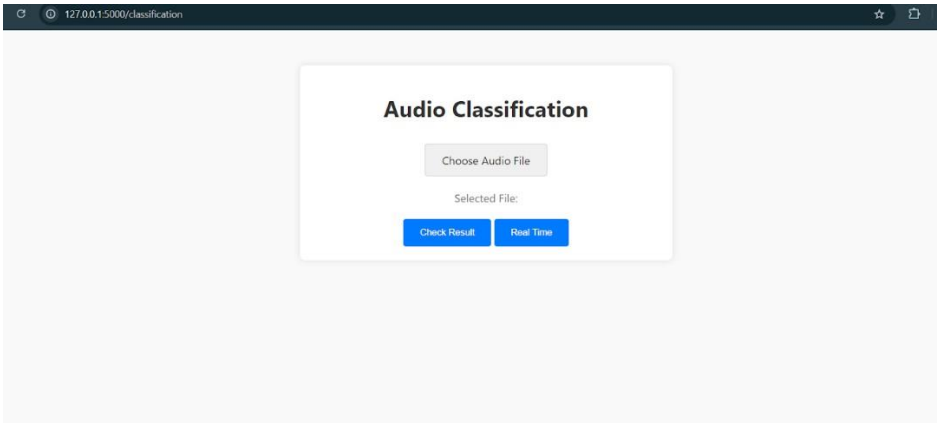

Fig. 3 Audio Classification

During the testing phase , the proposed methodology have shown promising results.Accuracy of the proposed methodology is calculated and turn out 90 % which makes this proposed approach simple yet effective.

V. CONCLUSION

Development of robust fake voice Detection system that identifies the difference between real and ai voice in the present era of generative ai is evolved as an emerging field of research due to its malicious use. Although this research has demonstrated the

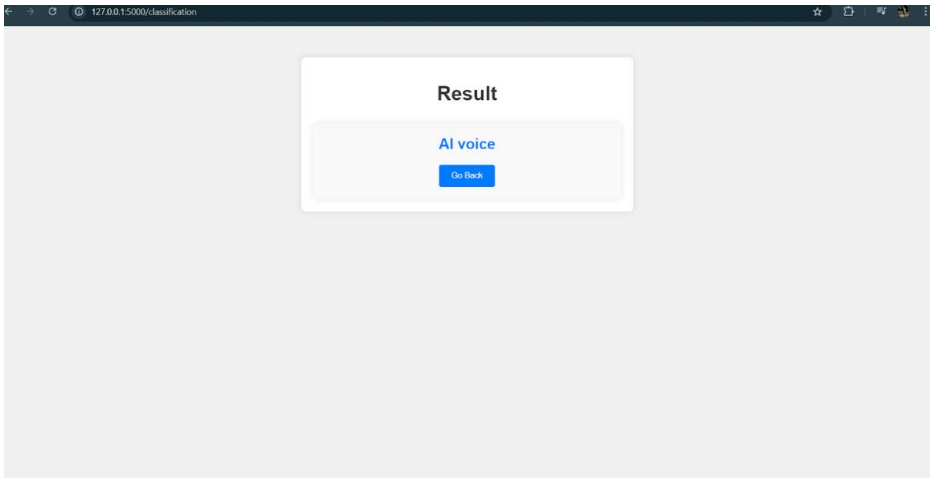

Fig. 4 Result: Real Or Fake

potential for accurately identifying the difference between ai audio and human voice using simple yet effective proposed methodology which employs artifact score to com- bat against the malicious implications of generative ai models . The findings highlight the significance of continuously adapting and improving detection technology to stay up with emerging manipulation strategies. This study not only establishes a platform for future breakthroughs in audio authenticity verification, but it also presented a com- prehensive review that illustrates the importance of cross-disciplinary collaboration in addressing the issues posed by artificially generated voice detection .

IJARSCT
International Journal of Advanced Research in Science, Communication and Technology
International Open-Access, Double-Blind, Peer-Reviewed, Refereed, Multidisciplinary Online Journal
ISSN: 2581-9429
Volume 5, Issue 8, May 2025
Impact Factor: 7.67

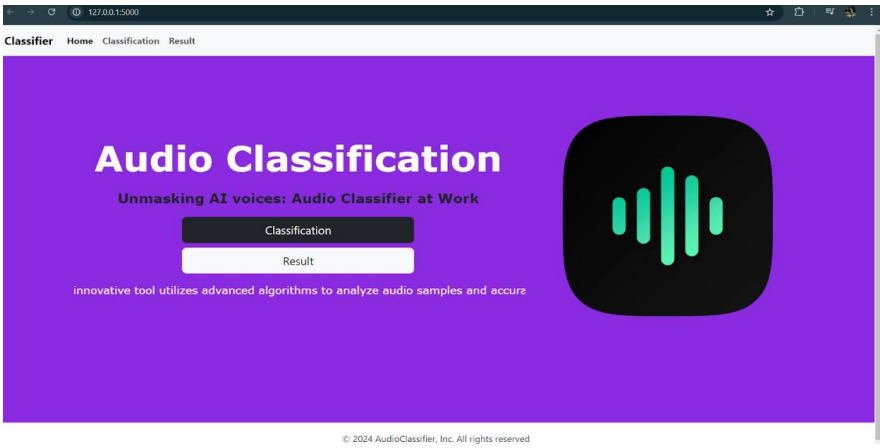

Fig. 6 Proposed Fake audio detection System

REFERENCES

[1]. Stupp, Catherine. ”Fraudsters used AI to mimic CEO’s voice in unusual cybercrime case.” The Wall Street Journal 30.08 (2019).

[2]. Chen, T., Kumar, A., Nagarsheth, P., Sivaraman, G., Khoury, E. (2020, Novem- ber). Generalization of Audio Deepfake Detection. In Odyssey (pp. 132-137).

[3]. Rodr´ıguez-Ortega, Yohanna, Dora Mar´ıa Ballesteros, and Diego Renza. ”A machine learning model to detect fake voice.” International Conference on Applied Infor- matics. Cham: Springer International Publishing, 2020.

[4]. Ballesteros, D. M., Rodriguez-Ortega, Y., Renza, D., Arce, G. (2021). Deep4SNet: deep learning for fake speech classification. Expert Systems with Applications, 184, 115465.

[5]. Lyu, Siwei. ”Deepfake detection: Current challenges and next steps.” 2020 IEEE international conference on multimedia expo workshops (ICMEW). IEEE, 2020.

[6]. Shen, S., Li, W., Huang, X., Zhu, Z., Zhou, J., Lu, J. (2023). Sd-nerf: Towards lifelike talking head animation via spatially-adaptive dual-driven nerfs. IEEE Transactions on Multimedia.

[7]. Ujjwal Saini. ”Voice phishing attack.” International Research Journal of engineering and technology (IRJECT), Volume -7, Page -1, July 2020.

[8]. Mubarak, R., Alsboui, T., Alshaikh, O., Inuwa-Dute, I., Khan, S., Parkinson, S. (2023). A survey on the detection and impacts of deepfakes in visual, audio, and textual formats. IEEE Access.

[9]. Rana, M. S., Nobi, M. N., Murali, B., Sung, A. H. (2022). Deepfake detection: A systematic literature review. IEEE access, 10, 25494-25513.

[10]. Taeb, M., Chi, H. (2022). Comparison of deepfake detection techniques through deep learning. Journal of Cybersecurity and Privacy, 2(1), 89-106.

[11]. Chen, T., Kumar, A., Nagarsheth, P., Sivaraman, G., Khoury, E. (2020, Novem- ber). Generalization of Audio Deepfake Detection. In Odyssey (pp. 132-137).

[12]. Khanjani, Z., Watson, G., Janeja, V. P. (2023). Audio deepfakes: A survey. Frontiers in Big Data, 5, 1001063.

[13]. Mu¨ller, N. M., Czempin, P., Dieckmann, F., Froghyar, A., B¨ottinger, K. (2022). Does audio deepfake detection generalize?. arXiv preprint arXiv:2203.16263.

[14]. Chakraborty, T., KS, U. R., Naik, S. M., Panja, M., Manvitha, B. (2024). Ten years of generative adversarial nets (GANs): a survey of the state-of-the-art. Machine Learning: Science and Technology, 5(1), 011001.

[15]. Bhanushali, A. R., Mun, H., Yun, J. (2024). Adversarial Attacks on Automatic Speech Recognition (ASR): A Survey. IEEE Access.

IJARSCT
International Journal of Advanced Research in Science, Communication and Technology
International Open-Access, Double-Blind, Peer-Reviewed, Refereed, Multidisciplinary Online Journal
ISSN: 2581-9429
Volume 5, Issue 8, May 2025
Impact Factor: 7.67

[16]. Almutairi, Z. M., Elgibreen, H. (2023). Detecting fake audio of Arabic speakers using self-supervised deep learning. IEEE Access, 11, 72134-72147.

[17]. Sch¨afer, K., Choi, J. E., Zmudzinski, S. (2024, June). Explore the World of Audio Deepfakes: A Guide to Detection Techniques for Non-Experts. In MAD@ ICMR (pp. 13-22).

[18]. Rabhi, Mouna, Spiridon Bakiras, and Roberto Di Pietro. "Audio-deepfake detec- tion: Adversarial attacks and countermeasures." Expert Systems with Applications 250 (2024): 123941.

[19]. Hamza, A., Javed, A. R. R., Iqbal, F., Kryvinska, N., Almadhor, A. S., Jalil, Z., Borghol, R. (2022). Deepfake audio detection via MFCC features using machine learning. IEEE Access, 10, 134018-134028.

[20]. Mcuba, M., Singh, A., Ikuesan, R. A., Venter, H. (2023). The effect of deep learning methods on deepfake audio detection for digital investigation. Procedia Computer Science, 219, 211-219.

[21]. Almutairi, Z., Elgibreen, H. (2022). A review of modern audio deepfake detection methods: challenges and future directions. Algorithms, 15(5), 155.

[22]. Ezer Osei Yeboah-Boating and Priscilla Mateko Amanor, Phishing, Smishing and Vishing: An Assesment of Threats against Mobile Device. "Journal of emerging Trends in Computing and Information science". Volume -5, Page-1, 4 April 2014.

[23]. Chaiwongyen, A., Songsriboonsit, N., Duangpummet, S., Karnjana, J., Kong- prawechnon, W., Unoki, M. (2022, November). Contribution of timbre and shimmer features to deepfake speech detection. In 2022 Asia-Pacific Signal and Informa- tion Processing Association Annual Summit and Conference (APSIPA ASC) (pp. 97-103). IEEE.

[24]. Wani, T. M., Gulzar, R., Amerini, I. (2024). ABC-CapsNet: Attention based Cascaded Capsule Network for Audio Deepfake Detection. In Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition (pp. 2464-2472).

[25]. Yan, R., Wen, C., Zhou, S., Guo, T., Zou, W., Li, X. (2022, May). Audio deepfake detection system with neural stitching for add 2022. In ICASSP 2022-2022 IEEE International Conference on Acoustics, Speech and Signal Processing (ICASSP) (pp. 9226-9230). IEEE.

[26]. Zeng, X. M., Zhang, J. T., Li, K., Liu, Z. L., Xie, W. L., Song, Y. (2023). Deepfake Algorithm Recognition System with Augmented Data for ADD 2023 Challenge. In DADA@ IJCAI (pp. 31-36).

[27]. Malik, H. (2019, June). Fighting AI with AI: fake speech detection using deep learn- ing. In 2019 AES INTERNATIONAL CONFERENCE ON AUDIO FORENSICS (June 2019).

[28]. Wang, R., Juefei-Xu, F., Huang, Y., Guo, Q., Xie, X., Ma, L., Liu, Y. (2020, October). Deepsonar: Towards effective and robust detection of ai-synthesized fake voices. In Proceedings of the 28th ACM international conference on multimedia (pp. 1207-1216).

[29]. Sun, C., Jia, S., Hou, S., Lyu, S. (2023). Ai-synthesized voice detection using neural vocoder artifacts. In Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition (pp. 904-912).

[30]. Kulangareth, N. V., Kaufman, J., Oreskovic, J., Fossat, Y. (2024). Investigation of Deepfake Voice Detection Using Speech Pause Patterns: Algorithm Development and Validation. JMIR biomedical engineering, 9, e56245.

[31]. Chen, Y., Wu, H., Jiang, N., Xia, X., Gu, Q., Hao, Y., ... Xu, M. (2024).

[32]. USTC-KXDIGIT System Description for ASVspoof5 Challenge. arXiv preprint arXiv:2409.01695.

[33]. Zhang, Q., Wen, S., Hu, T. (2024). Audio deepfake detection with self-supervised XLS-R and SLS classifier. In ACM Multimedia 2024.

[34]. T. -H. Shih, C. -Y. Yeh and M. -S. Chen, "Does Audio Deepfake Detection Rely on Artifacts?," ICASSP 2024 - 2024 IEEE International Conference on Acous- tics, Speech and Signal Processing (ICASSP), Seoul, Korea, Republic of, 2024, pp. 12446-12450, doi: 10.1109/ICASSP48485.2024.10446558.

[35]. Dehghani, A., & Saberi, H. (2025). Generating and Detecting Various Types of Fake Image and Audio Content: A Review of Modern Deep Learning Technologies and Tools. arXiv preprint arXiv:2501.06227.

[36]. A. Kaur, A. Noori Hoshyar, V. Saikrishna, S. Firmin, and F. Xia, "Deepfake video detection: challenges and opportunities," Artificial Intelligence Review, vol. 57, no. 6, pp. 1-47, 2024.