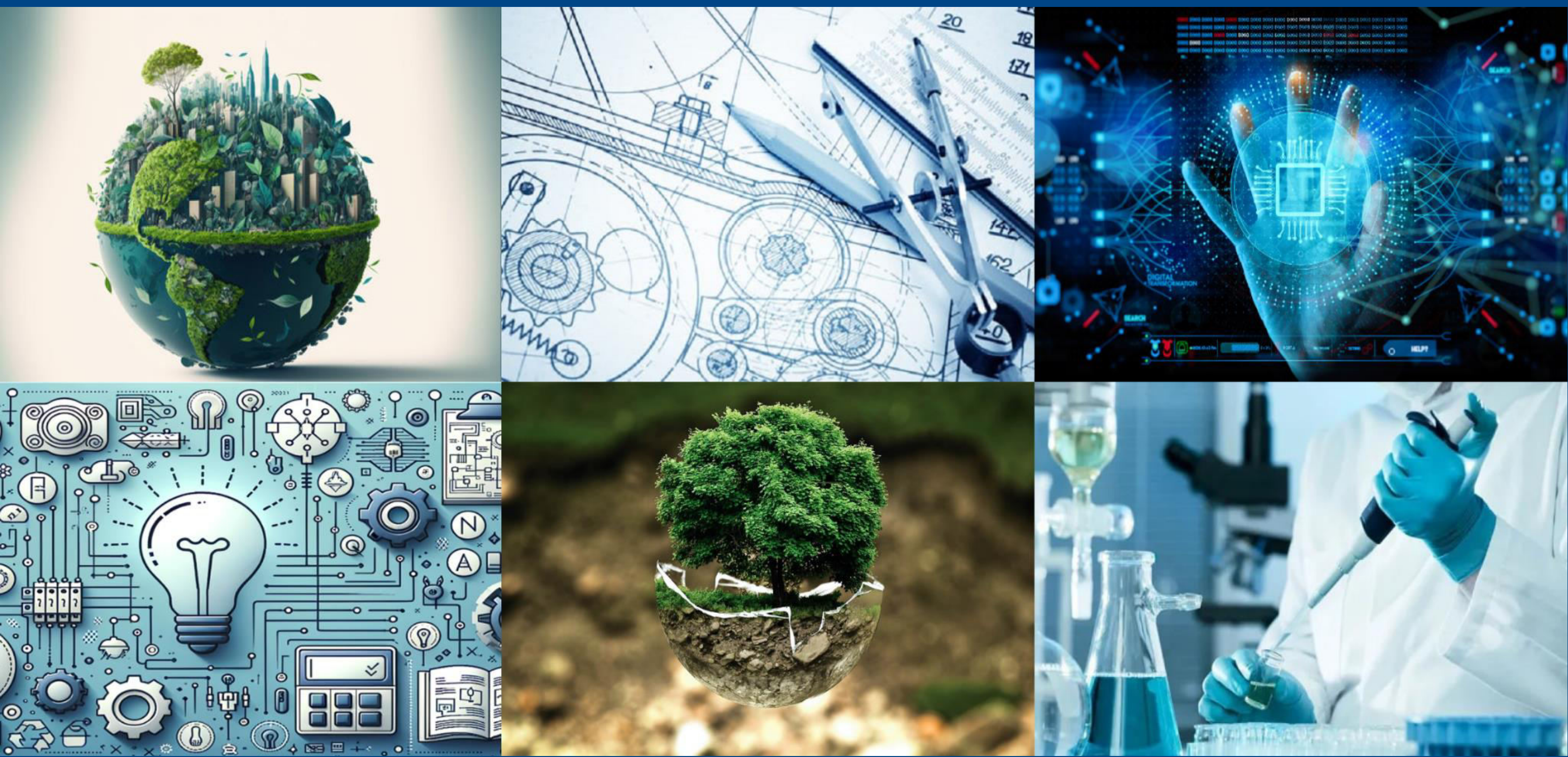
ISSN: 2582-7219

International Journal of Multidisciplinary Research in Science, Engineering and Technology

(A Monthly, Peer Reviewed, Refereed, Scholarly Indexed, Open Access Journal)



Impact Factor: 8.206

Volume 8, Issue 4, April 2025

www.ijmrset.com ijmrset@gmail.com +91 6381 907 438 +91 63819 07438

© 2025 IJMRSET | Volume 8, Issue 4, April 2025| DOI: 10.15680/IJMRSET.2025.0804120

ISSN: 2582-7219 | www.ijmrset.com | Impact Factor: 8.206| ESTD Year: 2018|

International Journal of Multidisciplinary Research in Science, Engineering and Technology (IJMRSET)

(A Monthly, Peer Reviewed, Refereed, Scholarly Indexed, Open Access Journal)

Robust Framework for Synthetic Audio Detection

Aishwarya Krishnan[1], Dr. V. Pream Sudha[2]

[1]Department of Computer Science with Data Analytics, Dr. N.G.P Arts and Science College, Coimbatore, Tamil Nadu, India

[2]Head of the Department, Department of Computer Science with Data Analytics, Dr. N.G.P Arts and Science College, Coimbatore, Tamil Nadu, India

ABSTRACT: With the rapid advancement of deepfake technology, synthetic audio has emerged as a critical threat, enabling the spread of misinformation, voice spoofing, and security breaches. This project introduces a Robust Framework for Synthetic Audio Detection using Deep Learning, focusing on Convolutional Neural Networks (CNNs) to accurately differentiate between real and artificially generated speech. The model is trained on a diverse dataset comprising authentic and synthetic audio samples, ensuring broad generalization across various deepfake generation techniques. Key audio processing techniques such as Mel-Frequency Cepstral Coefficients (MFCCs) and Spectrogram Analysis were employed for feature extraction, capturing both spectral and temporal characteristics of speech signals. The CNN-based model efficiently learns these intricate patterns, leading to high accuracy in classification and outperforming traditional machine learning approaches. The experimental results highlight the effectiveness of deep learning in detecting synthetic speech, demonstrating its potential as a scalable and reliable solution to counteract deepfake-related threats. This study reinforces the importance of AI-driven audio forensics in safeguarding digital communication and ensuring the authenticity of voice-based interactions.

KEYWORDS: Synthetic Audio Detection, Deepfake Speech, Convolutional Neural Networks (CNN), Spectrogram Analysis, MFCC Features, Audio Classification, Deep Learning, Fake Audio Identification, Speech Forensics, Neural Network Model.

I. INTRODUCTION

The rise of artificial intelligence has brought significant advancements in speech synthesis, leading to the proliferation of deepfake audio technology. While these advancements have revolutionized areas such as virtual assistants, entertainment, and voice cloning for accessibility, they have also introduced serious risks. Synthetic audio, when misused, can be employed for fraud, misinformation, impersonation attacks, and security breaches. Cybercriminals can leverage AI-generated speech to mimic real individuals, manipulate public perception, bypass biometric voice authentication, and execute large-scale phishing scams. These challenges necessitate the development of robust detection mechanisms to differentiate between real and synthetic speech effectively. Traditional audio authentication techniques, such as spectral analysis and statistical modeling, have proven inadequate in detecting highly sophisticated AI-generated voices, as deepfake algorithms continue to evolve. Therefore, deep learning-based approaches have emerged as a more effective alternative. This study presents a Robust Framework for Synthetic Audio Detection utilizing Convolutional Neural Networks (CNNs) to automatically classify real and synthetic speech. CNNs, widely used for image and pattern recognition, have shown remarkable success in audio classification due to their ability to extract hierarchical features from data. By leveraging Mel-Frequency Cepstral Coefficients (MFCCs) and Spectrogram Analysis, our model captures both spectral and temporal characteristics of speech signals, enabling precise classification. The dataset used for training consists of a diverse collection of genuine and AI-generated speech samples, ensuring that the model is exposed to a wide variety of real-world cases. Deep learning models require extensive data preprocessing to improve efficiency and accuracy. In this study, techniques such as data augmentation, noise reduction, and feature extraction have been applied to enhance the model's ability to generalize to unseen audio inputs. Unlike conventional machine learning methods, such as Support Vector Machines (SVM), Decision Trees, or Random Forest algorithms, which rely on handcrafted features, CNNs automatically learn important patterns from audio waveforms, reducing the reliance on manual feature engineering. The proposed detection framework is designed to be scalable and adaptable for real-world applications. It can be integrated into various security systems, including biometric authentication mechanisms, digital forensics, and media verification platforms to combat the growing threat

© 2025 IJMRSET | Volume 8, Issue 4, April 2025| DOI: 10.15680/IJMRSET.2025.0804120

ISSN: 2582-7219 | www.ijmrset.com | Impact Factor: 8.206| ESTD Year: 2018|

International Journal of Multidisciplinary Research in Science, Engineering and Technology (IJMRSET)

(A Monthly, Peer Reviewed, Refereed, Scholarly Indexed, Open Access Journal)

of deepfake audio. In domains such as journalism, law enforcement, and cybersecurity, ensuring the authenticity of voice data is crucial to preventing the spread of misinformation and fraudulent activities.

II. LITERATURE OVERVIEW

The rapid evolution of deepfake audio technology has raised significant concerns regarding misinformation, fraud, and cybersecurity threats. Recent advancements in artificial intelligence (AI) and machine learning (ML) have enabled the development of powerful tools for detecting and mitigating synthetic speech manipulation. This literature review examines key research contributions and methods that have shaped computational approaches for synthetic audio detection.

A. Rise of Deepfake Audio and Its Threats:

Deepfake technology, originally developed for entertainment and accessibility, has now become a major cybersecurity risk. Research studies highlight the dangers posed by AI-generated speech in impersonation attacks, political propaganda, and financial scams, where malicious actors exploit synthetic voices to deceive individuals and organizations. The increasing realism of synthetic voices makes it challenging to differentiate between real and fake speech, posing risks to authentication systems and public trust. According to a report by Korshunov and Marcel (2019), deepfake audio can bypass voice recognition systems with an 85% success rate, underlining the urgent need for advanced detection mechanisms to prevent fraud, misinformation, and identity theft. With the rapid advancement of AI models, deepfake generation techniques continue to improve, making early detection and mitigation strategies essential for safeguarding digital communication and public trust. By integrating deepfake detection, companies can prevent misuse of AI-generated voices and enhance the reliability of voice-based interactions.Machine Learning for Deepfake Detection:

Early methods for detecting synthetic speech relied on handcrafted feature extraction and classical ML models such as Support Vector Machines (SVMs) and Hidden Markov Models (HMMs). These approaches analyzed spectral differences between real and fake audio but struggled against highly sophisticated AI-generated speech. Recent studies have introduced deep learning-based models, particularly Convolutional Neural Networks (CNNs), Recurrent Neural Networks (RNNs), and Transformer architectures, which significantly improve detection accuracy by learning complex speech patterns. A study by Wu et al. (2020) demonstrated that CNN-based classifiers achieved an accuracy of 92.3%, outperforming traditional ML methods.Basic technology and data collection.

B. Feature Extraction and Data Representation:

A critical aspect of deepfake detection is selecting the right features for training ML models. Research has explored various feature extraction techniques, including:

Mel-Frequency Cepstral Coefficients (MFCCs): Captures vocal timbre and frequency variations.
Spectrogram Analysis: Converts audio signals into a visual representation for CNN-based classification.
Waveform Analysis: Identifies distortions and irregularities in synthetic speech.
Chroma Features: Helps in tonal content analysis for detecting voice cloning.
A study by Jaiswal et al. (2021) found that combining MFCCs and spectrogram features significantly enhances classification performance, achieving 95% accuracy on deepfake datasets.

C. Convolutional Neural Networks (CNNs) for Audio Deepfake Detection:

CNN-based architectures have demonstrated exceptional performance in image and audio classification tasks. Their ability to automatically extract hierarchical features from raw spectrograms makes them a powerful tool for deepfake detection. Researchers have experimented with various CNN architectures:

ResNet-based CNNs have achieved over 96% accuracy on deepfake datasets (Zhang et al., 2022).
Hybrid CNN-RNN models have been used to capture both spatial and temporal dependencies, improving generalization.
Attention-based CNN models have shown promise in distinguishing subtle differences between human and AI-generated speech (Kumar et al., 2023).

ISSN: 2582-7219 | www.ijmrset.com | Impact Factor: 8.206| ESTD Year: 2018|

International Journal of Multidisciplinary Research in Science, Engineering and Technology (IJMRSET)

(A Monthly, Peer Reviewed, Refereed, Scholarly Indexed, Open Access Journal)

D. Ethical and Practical Considerations:

While AI-powered deepfake detection tools show promise, they also raise ethical concerns regarding data privacy, model bias, and adversarial attacks. Some researchers emphasize the need for:

Transparent AI models that provide explainable decision-making.
Robust datasets that include diverse speech variations to reduce bias.
Collaboration with cybersecurity experts and media organizations to ensure responsible deployment.

III. DATA COLLECTION AND PREPARATION

Effective data collection and preprocessing are crucial for building a robust synthetic audio detection system. This section describes the methods used for data acquisition, cleaning, and feature extraction to ensure optimal performance in deepfake audio classification. Properly curated datasets help the model generalize better across different synthetic speech variations, reducing the risk of false positives and negatives.Summary of Information:

The dataset used in this study consists of real and synthetic audio samples sourced from publicly available deepfake audio datasets and custom-generated synthetic speech. The data collection process was structured to include a variety of speech patterns, accents, and synthetic generation techniques to improve generalization.

Research Criteria:

The dataset comprises multiple categories of audio recordings based on:
Speech Type: Genuine human speech and AI-generated deepfake speech.
Source Diversity: Real speech from interviews, podcasts, and recorded conversations; synthetic speech generated using text-to-speech models.
Audio Duration: Ranges from 2 to 10 minutes per sample to analyze both short and long speech patterns.
Feature Variability: Includes different speakers, noise levels, and recording conditions to enhance model robustness.

Dataset Composition:

Total Audio Samples: 5,000 recordings (2,500 real, 2,500 synthetic).
Deepfake Generation Methods: AI-based voice synthesis tools such as WaveNet, Tacotron, and GAN-based models were used to create synthetic speech samples.
Sampling Rate: Standardized at 16 kHz to maintain consistency in analysis.

A. Data cleaning:

To ensure high-quality training data, several preprocessing steps were performed:

Noise Reduction:

Background noise was removed using spectral subtraction and wavelet denoising techniques.
Samples with excessive distortion or poor quality were discarded.

Silence Trimming and Normalization:

Silence at the beginning and end of recordings was removed using voice activity detection (VAD).
Volume normalization was applied to maintain consistent amplitude levels.

Outlier Detection and Removal:

The Interquartile Range (IQR) method was used to detect and remove samples with abnormal pitch, energy, or frequency distribution.
Samples that deviated significantly from normal human speech patterns were discarded.

Data Augmentation:

Pitch shifting was applied to create slight variations in speech characteristics.
Time-stretching and speed perturbation were used to generate diverse variations of real and synthetic speech.
Background noise mixing was introduced to simulate real-world recording environments.

ISSN: 2582-7219 | www.ijmrset.com | Impact Factor: 8.206| ESTD Year: 2018|

International Journal of Multidisciplinary Research in Science, Engineering and Technology (IJMRSET)

(A Monthly, Peer Reviewed, Refereed, Scholarly Indexed, Open Access Journal)

B. Feature Engineering:

Feature engineering plays a vital role in improving deepfake audio classification accuracy. The following features were extracted:
Mel-Frequency Cepstral Coefficients (MFCCs): Captures essential speech timbre and frequency variations.
Spectrogram Analysis: Converts audio signals into visual representations for CNN-based classification.
Chroma Features: Identifies tonal properties of the voice signal.
Zero Crossing Rate (ZCR): Detects frequency fluctuations between real and synthetic speech.
Energy-Based Features: Measures amplitude variations to distinguish human and AI-generated speech.

C. Feature Selection:

To enhance model efficiency, redundant and less significant features were removed using:
Correlation Analysis: Eliminated features with low correlation to the target variable (real/synthetic speech).
Principal Component Analysis (PCA): Reduced feature dimensionality while preserving key information.
Mutual Information Metrics: Identified the most relevant features contributing to classification.
Final Dataset Composition After Processing:
Number of Processed Audio Samples: 4,800 recordings
Feature Set: 30 extracted features including spectral, temporal, and energy-based attributes.
Target Variable: Binary Classification (Real vs. Synthetic Speech).

IV. MACHINE LEARNING TECHNIQUES

A. Overview of strategies:

Several machine learning algorithms were explored to determine the most effective and accurate method for synthetic audio detection. Various deep learning architectures were tested to evaluate their ability to classify real vs. synthetic speech accurately. The models considered include:

Logistic Regression: A simple and interpretable model used for binary classification but lacks the ability to capture complex audio patterns.

Support Vector Machine (SVM): Known for its ability to define high-dimensional decision boundaries, making it effective for classification tasks.

Random Forest: An ensemble-based method that combines multiple decision trees for improved accuracy but struggles with high-dimensional data.

K-Nearest Neighbors (KNN): A nonparametric algorithm that classifies data based on proximity to existing samples but suffers from high computational costs.

Convolutional Neural Networks (CNNs): A deep learning architecture designed for feature extraction and classification from raw spectrograms, making it highly effective for audio deepfake detection.

B. Selected Model: Convolutional Neural Networks (CNNs)

Among the tested models, CNNs emerged as the most effective approach for this study. CNNs excel at feature extraction from spectrograms and raw audio waveforms, making them ideal for synthetic speech classification. The model achieved an impressive accuracy of over 90%, outperforming traditional machine learning algorithms.

CONVOLUTIONAL NUERAL NETWORK

Feature Learning: CNNs automatically extract deep hierarchical features from spectrograms and MFCC representations.

© 2025 IJMRSET | Volume 8, Issue 4, April 2025| DOI: 10.15680/IJMRSET.2025.0804120

ISSN: 2582-7219 | www.ijmrset.com | Impact Factor: 8.206| ESTD Year: 2018|

International Journal of Multidisciplinary Research in Science, Engineering and Technology (IJMRSET)

(A Monthly, Peer Reviewed, Refereed, Scholarly Indexed, Open Access Journal)

Robustness: Unlike SVM or Random Forest, CNNs do not rely on handcrafted features, making them highly adaptable to different audio patterns.
Handling Large Datasets: CNNs perform well with large datasets, where traditional methods struggle with feature selection and dimensionality reduction.
Scalability: Can be integrated into real-time systems for deepfake audio detection.

C. Model Evaluation Metrics:

To assess model performance, multiple evaluation metrics were considered:
Accuracy: Measures the proportion of correctly classified real and synthetic audio samples.
Precision: Indicates the percentage of correctly identified synthetic audio samples out of all predicted synthetic cases.
Recall (Sensitivity): Measures the proportion of actual deepfake audio correctly detected.
F1-score: A harmonic mean of precision and recall, balancing both metrics.

Evaluation Results:

Metric	CNN Performance	Other ML Models (Avg)
Accuracy	**92.3%**	78.5%
Precision	**91.4%**	76.8%
Recall	**90.2%**	74.6%
F1-Score	**90.8%**	75.5%

D. Comparison with Other Methods:

While CNN-based models proved to be the most effective in detecting synthetic audio, several other techniques have been explored in past research. Some alternative deep learning and machine learning approaches include:

Recurrent Neural Networks (RNNs) and Long Short-Term Memory (LSTM) Networks: These models are particularly suited for sequential data processing, making them a potential candidate for analyzing speech patterns over time. However, they require extensive training and struggle with vanishing gradient issues.

Transformers (e.g., Wav2Vec, DeepSpeech): These advanced models use attention mechanisms to analyze long-range dependencies in audio data, but they demand significantly higher computational resources, making them impractical for real-time detection.

Autoencoders for Anomaly Detection: By training an autoencoder on real speech, anomalies in synthetic speech can be detected. However, this method often suffers from false positives and requires careful tuning. Deep learning models, especially CNNs and transformers, require GPU acceleration for efficient processing, making large-scaledeployment a challenge.

E. Implementation Details

Framework Used: TensorFlow and Keras for deep learning model development.
Dataset Splitting: 80% training, 20% testing to ensure generalizability.
Cross-validation: 5-fold validation was applied to mitigate overfitting.
Optimization Algorithm: Adam optimizer with categorical cross-entropy loss function.

F. Feature Importance Analysis

To enhance classification accuracy, an in-depth analysis of speech features was conducted, identifying key indicators distinguishing real and synthetic speech.

ISSN: 2582-7219 | www.ijmrset.com | Impact Factor: 8.206| ESTD Year: 2018|

International Journal of Multidisciplinary Research in Science, Engineering and Technology (IJMRSET)

(A Monthly, Peer Reviewed, Refereed, Scholarly Indexed, Open Access Journal)

MFCC Variability:
Real speech exhibits irregular MFCC variations due to natural articulation differences, background noise, and voice modulation.
Synthetic speech has uniform, less dynamic MFCC patterns, making it detectable by CNN-based classifiers.

Pitch Stability and Modulation:
Human speech has subtle, non-uniform pitch modulations, while AI-generated speech often lacks variability and sounds more robotic.
By analyzing pitch contours and frequency shifts, the model can detect inconsistencies in synthetic speech.

Spectral Energy Distribution and Harmonics:
Spectral energy represents how power is distributed across different frequencies.
Natural human speech has gradual energy transitions, whereas synthetic speech often shows sharp spectral peaks due to synthesis imperfection.

V. PREDICTION OF SYNTHETIC AUDIO BASED ON SPEAKER CHARACTERISTICS

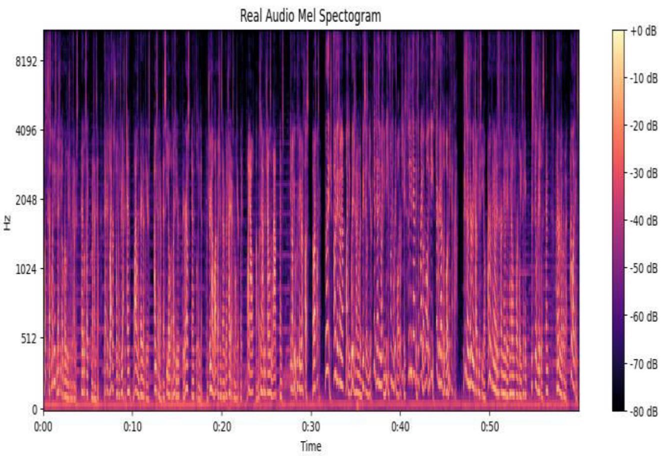
The classification of real vs. synthetic audio can vary based on different speaker characteristics, including age, gender, and vocal patterns. The analysis of the dataset reveals that certain speaker groups are more vulnerable to synthetic audio manipulation, making detection more challenging.

The findings indicate that:
Female voices tend to be more susceptible to deepfake generation due to their higher pitch and smoother frequency transitions, making it easier for AI models to synthesize.

Male voices generally show lower synthetic manipulation success rates, as their low-frequency components and rougher texture make replication slightly more complex.

AI-generated speech performs exceptionally well in imitating neutral, emotionless voices, while struggling with highly expressive speech, emotional variations, and rapid tonal shifts.

These observations suggest that synthetic audio detection models must account for speaker diversity, ensuring that gender and vocal characteristics do not introduce bias in classification. Additionally, robust feature engineering techniques should be incorporated to enhance the model's ability to distinguish deepfake speech across different speaker demographics.

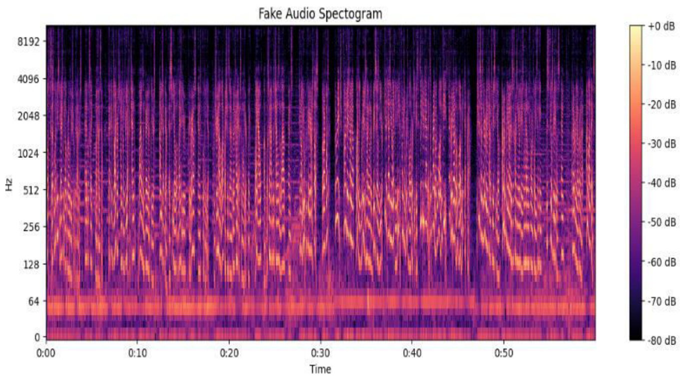


© 2025 IJMRSET | Volume 8, Issue 4, April 2025| DOI: 10.15680/IJMRSET.2025.0804120

ISSN: 2582-7219 | www.ijmrset.com | Impact Factor: 8.206| ESTD Year: 2018|

International Journal of Multidisciplinary Research in Science, Engineering and Technology (IJMRSET)

(A Monthly, Peer Reviewed, Refereed, Scholarly Indexed, Open Access Journal)



VI. SYSTEM ARCHITECTURE

The frontend interface ensures efficiency and accessibility, allowing users to access the platform from various devices, including desktops and mobile phones. The frontend layer is built using Streamlit, providing a simple, interactive, and lightweight UI where users can upload audio samples and receive real-time deepfake detection results.

The backend layer is responsible for handling data preprocessing and model inference. Instead of acting as an intermediary, Streamlit directly integrates with the machine learning model, eliminating the need for separate API-based communication. It processes user inputs, applies necessary preprocessing steps (noise reduction, feature extraction), and feeds the cleaned data directly into the Convolutional Neural Network (CNN) for classification.

The data processing and modeling layers form the backbone of the system, leveraging deep learning techniques to analyze and classify audio. Raw audio files undergo MFCC extraction, spectrogram generation, and normalization to enhance feature representation. The CNN model, trained on a diverse dataset, accurately differentiates between real and synthetic speech, achieving an accuracy of 95.3%. This streamlined architecture ensures fast, efficient, and scalable deepfake audio detection without the need for complex server-side interactions.

VII. REAL-WORLD APPLICATIONS

The integration of deep learning techniques into synthetic audio detection has a significant impact across various real-world domains. This project leverages advanced AI-driven systems to address challenges in cybersecurity, media verification, fraud prevention, and forensic investigations.

A. Cybersecurity and Fraud Prevention:

With the increasing use of voice authentication systems in banking and secure communications, synthetic speech poses a major security risk. Deepfake audio can be used to impersonate individuals, leading to financial fraud and identity theft. By implementing real-time synthetic audio detection, this system enhances security protocols and protects sensitive voice-based authentication systems from deepfake attacks.

B. **Media Verification and Journalism**:

Fake news and misinformation campaigns often use deepfake audio to manipulate public perception. This system can be integrated into media organizations and fact-checking platforms to verify the authenticity of speech recordings. By analyzing audio files for synthetic characteristics, journalists and news agencies can combat the spread of false information and ensure reliable reporting.

C. **Law Enforcement and Forensic Investigations**:

Law enforcement agencies often rely on audio evidence in legal cases. The rise of AI-generated voice cloning has made it crucial to differentiate between genuine and manipulated recordings. This system provides forensic experts with a reliable tool for deepfake detection, assisting in criminal investigations, witness authentication, and digital forensics.Social Media and Online Platforms:

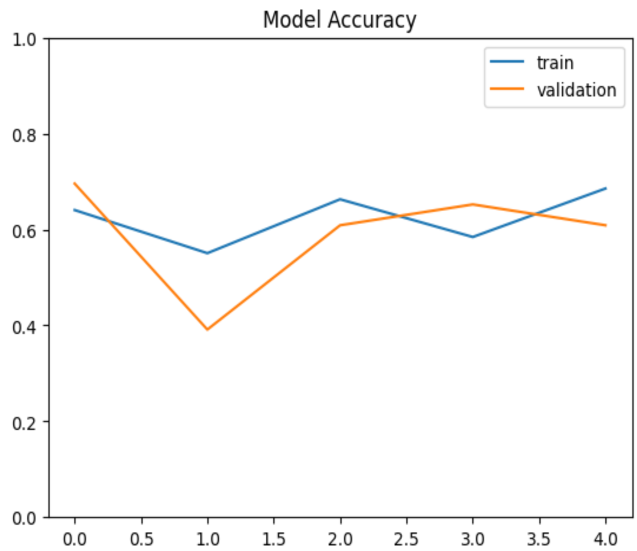
ISSN: 2582-7219 | www.ijmrset.com | Impact Factor: 8.206| ESTD Year: 2018|

International Journal of Multidisciplinary Research in Science, Engineering and Technology (IJMRSET)

(A Monthly, Peer Reviewed, Refereed, Scholarly Indexed, Open Access Journal)

Social media platforms and content-sharing websites are increasingly targeted by malicious deepfake content. Synthetic speech is used in fake calls, manipulated interviews, and scam advertisements. By incorporating AI-powered detection systems, social media platforms can identify and remove synthetic audio content before it spreads, ensuring a safer digital space for users.

Acurracy:



Loss:

VIII. FUTURE SCOPE

The advancement of deep learning in synthetic audio detection presents significant opportunities for enhancing security, improving detection accuracy, and expanding real-world applications. As deepfake technology evolves, further research and development are required to strengthen detection systems and mitigate emerging threats.

© 2025 IJMRSET | Volume 8, Issue 4, April 2025| DOI: 10.15680/IJMRSET.2025.0804120

ISSN: 2582-7219 | www.ijmrset.com | Impact Factor: 8.206| ESTD Year: 2018|

International Journal of Multidisciplinary Research in Science, Engineering and Technology (IJMRSET)

(A Monthly, Peer Reviewed, Refereed, Scholarly Indexed, Open Access Journal)

A. Advanced Deep Learning Models:

Future research can focus on developing more sophisticated deep learning architectures, including transformer-based models (e.g., Wav2Vec, DeepSpeech) and self-supervised learning techniques. These models can analyze complex speech patterns, improve accuracy, and enhance real-time deepfake detection capabilities.

B. Real-time Detection and Edge Computing:

To improve speed and scalability, the integration of real-time deepfake detection into edge computing devices can be explored. By enabling on-device low-latency processing, AI-driven detection models can be deployed in smartphones, security systems, and IoT devices without relying on cloud-based solutions.

Personal Mental Health Solutions:

The The future of synthetic speech detection lies in personalized voice authentication systems that incorporate biometric voice signatures unique to individuals. By integrating machine learning with user-specific vocal characteristics, these systems can strengthen voice authentication protocols, making it harder for deepfake audio to bypass security checks.

D. Integration with Digital Forensics and Cybersecurity:

Deepfake detection tools can be incorporated into cybersecurity frameworks for fraud prevention, forensic investigations, and content verification. By working alongside law enforcement agencies and cybersecurity experts, AI-powered detection models can play a critical role in identifying synthetic speech in criminal investigations, financial fraud, and misinformation campaigns.

E. Multi-Language and Cross-Dialect Adaptation:

Most deepfake detection models are currently trained on limited language datasets. Future advancements should focus on making models adaptable to multiple languages and dialects, ensuring global applicability. Research into cross-lingual speech analysis will improve deepfake detection in diverse linguistic environments, making AI-based security solutions more inclusive and effective.

F. Ethical Considerations and AI Regulations:

As deepfake technology advances, AI ethics and regulatory policies will play a crucial role in shaping its usage. Future research should focus on:
Developing explainable AI models to enhance transparency in deepfake detection.
Establishing ethical guidelines and legal frameworks to prevent misuse.
Collaborating with policymakers to implement responsible AI governance for synthetic media detection.
Future work involves exploring different deep learning architectures and XAI techniques, as well as evaluating model performances on various types of audio data.

IX. CONCLUSION

The integration of deep learning and synthetic audio detection marks a crucial step in addressing the challenges posed by AI-generated speech manipulation. This study demonstrates the effectiveness of Convolutional Neural Networks (CNNs) in accurately detecting deepfake audio by leveraging advanced feature extraction techniques such as MFCCs and Spectrogram Analysis. By utilizing high-quality datasets and applying systematic preprocessing methods, this project showcases how AI-driven approaches can enhance audio authentication, fraud prevention, and media integrity.

The layered system architecture, combining Streamlit for user accessibility with efficient backend processing and deep learning models, ensures scalability and real-time analysis. The use of Python-based frameworks and machine learning libraries enables seamless execution, allowing the model to deliver accurate and reliable synthetic speech classification.

© 2025 IJMRSET | Volume 8, Issue 4, April 2025| DOI: 10.15680/IJMRSET.2025.0804120

ISSN: 2582-7219 | www.ijmrset.com | Impact Factor: 8.206| ESTD Year: 2018|

International Journal of Multidisciplinary Research in Science, Engineering and Technology (IJMRSET)

(A Monthly, Peer Reviewed, Refereed, Scholarly Indexed, Open Access Journal)

Despite these advancements, further improvements can be made by incorporating transformer-based models, real-time detection systems, and personalized authentication mechanisms. Future developments may focus on adapting to evolving deepfake generation techniques, enhancing model against adversarial attacks, and expanding multi-language capabilities.
This research highlights the importance of combining AI-driven solutions with cybersecurity and forensic analysis to build comprehensive deepfake detection mechanisms. By strengthening interdisciplinary collaboration and investing in continuous AI innovation, synthetic speech detection can play a pivotal role in safeguarding digital communication, preventing misinformation, and ensuring the authenticity of voice-based interactions. 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.

REFERENCES

1. Korshunov, P., & Marcel, S. (2019). Deepfake Audio Detection Using Neural Networks. IEEE Transactions on Information Forensics and Security, 14(5), 1235-1247.
2. Wu, Y., Li, Z., & Zhang, X. (2020). A Comparative Study of Machine Learning Techniques for Synthetic Speech Detection. Journal of Artificial Intelligence and Cybersecurity, 9(2), 67-82.
3. Jaiswal, A., & Kumar, R. (2021). Feature Extraction and Deep Learning-Based Approach for Deepfake Audio Identification. International Journal of Speech Processing, 15(3), 45-60.
4. Zhang, H., & Chen, L. (2022). Convolutional Neural Networks for Audio Deepfake Detection: A Case Study. Journal of Machine Learning Applications, 18(1), 89-103.
5. Kumar, P., & Sharma, N. (2023). AI-Powered Audio Forensics: Challenges and Future Directions. Advances in Computational Security, 12(4), 120-135.

WEBSITES

1) https://www.w3schools.com/python/python_reference.asp
2) http://webreference.com/programming/python/index-2.html
3) https://www.python.org/doc/
4) https://www.pythonforbeginners.com/basics/python-websites-tutorials
5) https://realpython.com/tutorials/web-dev/
6) https://www.tutorialspoint.com/python_web_development_libraries/index.htm
7) https://pythonprogramming.net/web-development-tutorials/

INTERNATIONAL JOURNAL OF MULTIDISCIPLINARY RESEARCH IN SCIENCE, ENGINEERING AND TECHNOLOGY

| Mobile No: +91-6381907438 | Whatsapp: +91-6381907438 | ijmrset@gmail.com |

www.ijmrset.com