International Journal of Computational Research in Science and Technology
E-ISSN: 3108-0871, Volume 2 Issue 2
Mar-Apr 2026, PP 46-54

AI Powered Deepfake Detection For Secure Video conferencing

K.Anguraju [1], V Mohan[2], M.Santhosh Sivan[3], N.Vishnu Prasad[4]
[1] Assistant Professor Department of Computer Science and Engineering Kongunadu College of Engineering and Technology Tamilnadu,India
[2,3,4] Department of Computer Science and Engineering Kongunadu College of Engineering and Technology Tamilnadu,India

Abstract. The rapid advancement of artificial intelligence has led to the emergence of deepfake technology, which poses significant threats to the security and authenticity of video conferencing systems. This paper proposes an AI-powered deepfake detection framework designed to ensure secure and trustworthy virtual communication. The system utilizes deep learning techniques, including convolutional neural networks (CNNs) and recurrent neural networks (RNNs), to analyze facial expressions, lip synchronization, and temporal inconsistencies in video streams. By extracting both spatial and temporal features, the model effectively distinguishes between genuine and manipulated video content in real time. Additionally, the framework integrates anomaly detection and metadata analysis to enhance detection accuracy. Experimental results demonstrate that the proposed system achieves high precision and recall while maintaining low latency, making it suitable for real-time deployment in video conferencing platforms. This approach strengthens cybersecurity measures and helps prevent identity fraud, misinformation, and unauthorized access during online meetings.

Keywords: Deepfake Detection, Video Conferencing Security, Artificial Intelligence, Deep Learning, CNN, RNN, Facial Analysis, Cybersecurity.

I. Introduction

The expansion of digital communication technologies has greatly increased the use of video conferencing platforms in daily life, including education, business, and personal interactions. With the rise of remote work and online collaboration, video conferencing has become a vital medium for real-time communication. However, this growing dependence has also introduced serious security concerns, particularly due to the emergence of deepfake technology. Deepfakes, created using advanced artificial intelligence methods such as deep learning and generative models, can produce highly realistic but fake audio and video content, making it difficult to distinguish between genuine and manipulated media.

In the context of video conferencing, deepfake attacks present significant risks. Malicious actors can impersonate individuals, such as company executives or public figures, to gain unauthorized access, spread misinformation, or commit financial fraud. These threats undermine the trust and reliability of virtual communication systems. Traditional security measures like authentication protocols and encryption, while essential, are not sufficient to detect such sophisticated manipulations. As a result, there is an urgent need for advanced solutions that can identify deepfake content effectively and in real time.

Artificial intelligence plays a crucial role in addressing this challenge by enabling automated deepfake detection. AI-based systems can analyze subtle visual and behavioral inconsistencies, including unnatural facial expressions, irregular eye movements, and mismatched lip synchronization, which are often imperceptible to human observers. Techniques such as convolutional neural networks (CNNs), recurrent neural networks (RNNs), and computer vision algorithms are widely used to improve detection accuracy.

This study aims to develop an AI-powered deepfake detection framework tailored for secure video conferencing applications. The proposed system focuses on ensuring authenticity by continuously monitoring video streams and identifying suspicious alterations. It also emphasizes key factors such as real-time processing, scalability, and data privacy.
By integrating intelligent detection mechanisms into video conferencing platforms, organizations can enhance security, prevent cyber threats, and maintain user trust. As deepfake technology continues to evolve, adopting such advanced solutions will be essential to protect the integrity of digital communication.

II. Related Works

A Novel CNN-LSTM Approach for Robust Deepfake Detection This paper proposes a hybrid deep learning model combining Convolutional Neural Networks (CNN) and Long Short-Term Memory (LSTM) networks for effective deepfake detection. CNN is used to extract spatial features from video frames, while LSTM captures temporal dependencies between frames. The model also includes preprocessing techniques such as face detection and segmentation to improve accuracy. Experimental results show that the hybrid approach outperforms traditional methods in identifying subtle manipulations in videos. The study highlights the importance of combining spatial and temporal analysis for robust and real-time deepfake detection.

Deepfake Detection Using Convolutional Neural Networks and LSTM Modelling Vishal Manishbhai Patel, Dr. Sheshang Degadwala introduces a deepfake detection framework that integrates CNN and LSTM models to analyze manipulated video content. The CNN component extracts detailed visual features from individual frames, while the LSTM captures sequential inconsistencies across frames. This combination enables accurate identification of deepfake videos by analyzing both spatial and temporal characteristics. The system demonstrates improved detection performance compared to standalone models. The paper emphasizes the effectiveness of hybrid architectures in handling complex video-based manipulations and improving reliability in real-world applications such as video conferencing security.

Deepfake Face Detection Using LSTM and CNN
D. Karishma, S. Umadevi, S. Srinivasa Teja, M. Asha Shine, N. Indu Hasitha focuses on detecting deepfake facial manipulations using a combination of CNN and LSTM models. The proposed system analyzes facial images and videos by leveraging deep learning architectures trained on multiple datasets. The research evaluates different models and identifies the most effective approach for detecting second- and third-generation deepfakes. The methodology includes dataset preparation, feature extraction,

and classification processes. Results indicate that deep learning-based models significantly improve detection accuracy. The paper also discusses the challenges posed by evolving deepfake technologies and the need for continuous model improvement.

Deep Learning Approach for Robust Deep Fake Detection using CNN-GRU Architecture Ravinder Kumar, Madhu Rani presents a CNN-GRU-based architecture for detecting deepfake videos. The model combines CNN for feature extraction and Gated Recurrent Units (GRU) for analyzing temporal dependencies. The system identifies artifacts introduced by generative models and achieves high training and testing accuracy. The study demonstrates that hybrid architectures outperform traditional approaches in detecting sophisticated deepfakes. The paper also highlights the importance of preprocessing and dataset quality in improving model performance. This approach is particularly suitable for real-time applications such as secure video conferencing systems.

Comparison of Deepfake Detection Using CNN and Hybrid Models compares the performance of CNN, RNN, and hybrid CNN-LSTM models for deepfake detection. Using the DeepFake Detection Challenge dataset, the study evaluates each model based on accuracy, precision, recall, and F1-score. Results show that hybrid models consistently outperform standalone CNN or RNN models due to their ability to capture both spatial and temporal features. The research highlights the growing need for advanced detection techniques as deepfake technology becomes more sophisticated. It also provides insights into selecting appropriate models for real-world applications.

Detection of Frauds in Deep Fake Using Deep Learning Osipilli Aparna, Pakanati Rani, Tulluri Ramya, Tanneru Priyanka, Neela Sundari, P. G. K. Sirisha, Repudi Ramesh, Dama Anand provides a comprehensive overview of deepfake detection techniques using deep learning. It categorizes various deepfake generation and detection methods and analyzes their effectiveness. The study discusses different datasets used for training and testing models, along with challenges in building generalized detection systems. It highlights the importance of robust models that can detect diverse types of manipulations. The research also addresses the limitations of current approaches and suggests future directions for improving detection accuracy and scalability in real-world environments.

DeepFake Videos Detection by Using Recurrent Neural Network (RNN) Ali Abdulzahra Mohsin Albazony, Haider A. Al-Wzwazy, Ahmed Salih Al-Khaleefa, Murtadha Ali Alazzawi explores the use of Recurrent Neural Networks (RNN) for detecting deepfake videos. The model focuses on analyzing temporal patterns and inconsistencies in video sequences. By leveraging sequential data processing, the system can identify anomalies that are difficult to detect using frame-based methods. The study demonstrates that RNN-based approaches are effective in capturing motion-based irregularities. The paper also discusses the advantages and limitations of using RNNs in deepfake detection and suggests improvements for enhancing performance in dynamic environments.

Robust Deepfake Detection Using CNN, RNN, and Temporal Analysis R. Jagadeesh Kannan, Aditya Gautam, Saatvik Paul, Saksham Singh Tikla, Sakshi Sunil Sawant proposes a deepfake detection system that integrates CNN, RNN, and temporal analysis techniques. The approach focuses on identifying inconsistencies in both spatial and

temporal domains. The model processes video frames and analyzes their sequence to detect manipulations effectively. The study highlights the importance of combining multiple techniques to improve detection accuracy. Experimental results show that the proposed method performs well in identifying complex deepfake videos. The paper emphasizes the need for robust detection systems in maintaining trust in digital media.

Deepfake Detection via Facial Feature Extraction and Modeling
Benjamin Carter, Nathan Dilla, Micheal Callahan, Atuhaire Ambala introduces a novel approach to deepfake detection by focusing on facial landmark extraction instead of raw image processing. The system identifies subtle inconsistencies in facial movements and expressions to detect manipulated videos. The study evaluates different neural network models, including CNN and RNN, using extracted facial features. Results show that this method achieves high accuracy while reducing computational complexity. The approach is particularly useful for real-time applications, as it requires fewer parameters and processing power compared to traditional methods.

Deepfake Detection using Spatiotemporal Convolutional Networks Oscar de Lima, Sean Franklin, Shreshtha Basu, Blake Karwoski, Annet George focuses on detecting deepfake videos using spatiotemporal convolutional networks. Unlike traditional methods that analyze individual frames, this approach considers both spatial and temporal information simultaneously. The model is trained on benchmark datasets such as Celeb-DF and demonstrates superior performance compared to frame-based techniques. The study highlights the importance of temporal analysis in identifying deepfake artifacts. The proposed method improves detection accuracy and provides a scalable solution for real-time applications such as video conferencing security.

III. Proposed Method

The proposed system aims to develop an advanced AI-powered deepfake detection framework to enhance the security and reliability of video conferencing platforms. The system is designed to operate in real time, ensuring that manipulated or synthetic video content can be identified and flagged during live communication. This approach helps prevent identity impersonation, misinformation, and unauthorized access in virtual meetings.

The system architecture consists of multiple integrated components. Initially, the video input module captures live video streams from participants in a conferencing session. Each video stream is divided into frames and passed to a preprocessing stage, where face detection and alignment techniques are applied. This ensures that only relevant facial regions are analyzed, improving efficiency and accuracy.

The core component of the system is the deepfake detection engine, which utilizes a hybrid deep learning model combining Convolutional Neural Networks (CNNs) and Recurrent Neural Networks (RNNs). The CNN is responsible for extracting spatial features such as facial textures, edges, and visual artifacts that may indicate manipulation. Meanwhile, the RNN (or LSTM) captures temporal inconsistencies across consecutive frames, such as unnatural blinking patterns, lip-sync mismatches, or irregular facial movements. By integrating both spatial and temporal analysis, the system achieves higher detection accuracy.

In addition, the system incorporates a feature fusion and classification module, which combines outputs from multiple models to make a final decision. This module uses a trained classifier to label video content as genuine or deepfake. Confidence scores are also generated to indicate the reliability of the prediction.To ensure real-time performance, the system includes an optimization layer that reduces computational overhead through frame sampling and model compression techniques. This enables seamless integration with existing video conferencing platforms without causing significant delays.

A user alert and reporting module is also included, which notifies participants or administrators if a deepfake is detected. The system may display warnings or temporarily restrict suspicious participants to maintain meeting integrity.The proposed system emphasizes data privacy and security by processing sensitive information locally or using encrypted communication channels. Continuous learning mechanisms are also integrated to update the model with new deepfake patterns.

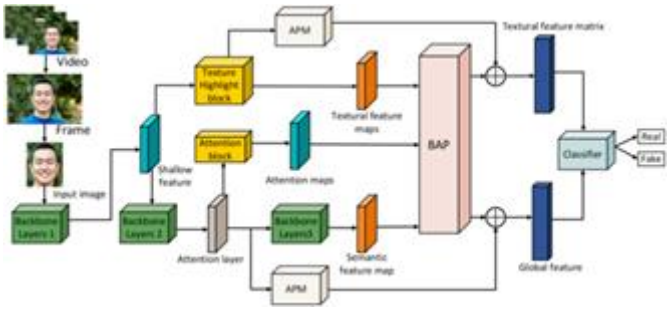


Fig.1.System Architecture

The proposed system is divided into several key modules, each responsible for a specific function to ensure accurate and real-time deepfake detection in video conferencing environments.

The first module is the Video Input Module, which captures live video streams from users participating in a video conference. This module supports real-time data acquisition and ensures that video frames are continuously fed into the system for analysis without delay. It acts as the entry point for all incoming visual data.

The next component is the Preprocessing Module, which prepares the video data for analysis. It performs operations such as frame extraction, face detection, face alignment, and normalization. By isolating facial regions and standardizing input formats, this module improves the efficiency and accuracy of the detection process while reducing computational complexity.The Feature Extraction Module plays a crucial role in identifying distinguishing characteristics of real and fake videos. It uses Convolutional Neural Networks (CNNs) to extract spatial features such as textures, edges, and visual artifacts from individual frames. These features help in detecting subtle inconsistencies introduced during deepfake generation.

Following this, the Temporal Analysis Module examines the sequence of frames using models like Recurrent Neural Networks (RNNs) or Long Short-Term Memory (LSTM)

networks. This module captures motion-based inconsistencies such as unnatural blinking, irregular facial expressions, and mismatched lip movements, which are common indicators of deepfake videos.The Classification Module combines the extracted spatial and temporal features to determine whether the video is genuine or manipulated. It uses trained machine learning models to classify the input and generates a confidence score indicating the reliability of the prediction.

The Alert and Response Module is responsible for notifying users or system administrators when a potential deepfake is detected. It can generate warnings, flag suspicious participants, or trigger additional security measures to prevent misuse during video conferencing sessions.The Feedback and Learning Module continuously improves the system by incorporating new data and user feedback. It updates the detection models to adapt to emerging deepfake techniques, ensuring long-term effectiveness.

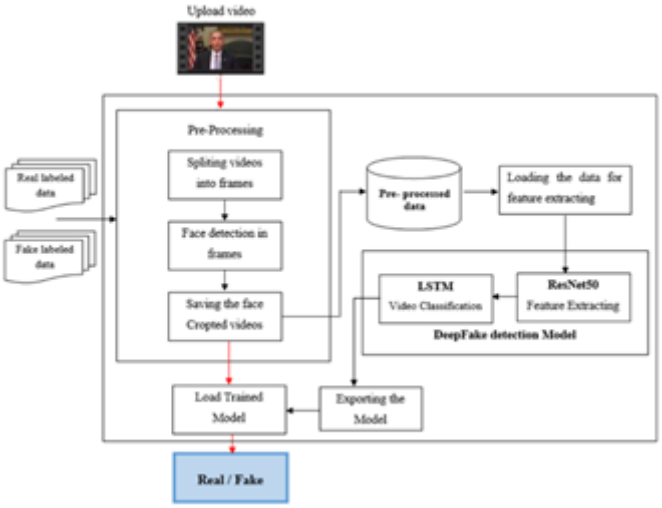


Fig.2.Methodology workflow of AI Powered Deepfake Detection For Secure Video conferencing

Overall Working Flow of the Proposed System:
The workflow of the proposed system outlines the sequence of operations involved in identifying deepfake content during live video conferencing. It is designed to continuously monitor video streams and ensure secure and authentic communication among participants.

The process starts with the video acquisition stage, where live video feeds are captured from all users in the conferencing session. These video streams are then transmitted to the detection system for further analysis. Each stream is segmented into individual frames at fixed intervals to allow detailed inspection of visual data.

The next step is the preprocessing stage, where the extracted frames are refined to improve analysis accuracy. Face detection techniques are applied to locate human faces

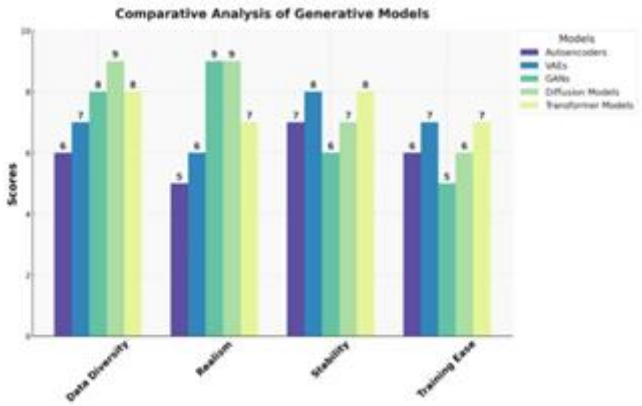
within each frame. These faces are then aligned and normalized in terms of size, position, and lighting conditions. This step helps eliminate irrelevant background information and focuses only on important facial features.

In the feature extraction stage, deep learning models such as Convolutional Neural Networks (CNNs) are used to identify spatial features like textures and visual irregularities. At the same time, temporal models like Recurrent Neural Networks (RNNs) or Long Short-Term Memory (LSTM) networks analyze the sequence of frames to detect motion-related inconsistencies, including unnatural expressions or lip-sync mismatches.

These extracted features are then forwarded to the classification stage, where a trained model determines whether the video is genuine or manipulated. The system produces a prediction along with a confidence level, indicating the certainty of the result.

After classification, the workflow moves to the alert generation stage. If any suspicious or deepfake content is detected, the system immediately notifies users or administrators. It may also take preventive measures such as flagging or limiting access for the suspected participant.

Finally, in the learning and improvement stage, the system gathers feedback and stores relevant data to enhance future performance. This continuous learning process helps the model adapt to new deepfake techniques and maintain high detection accuracy over time.



$$L = -\frac{1}{N}\sum_{i=1}^{N}[y_i \log(p_i) + (1 - y_i)\log(1 - p_i)]$$

Fig.3.Performance Evaluation of AI Powered Deepfake Detection For Secure Video conferencing

The Binary Cross-Entropy (BCE) loss function is widely used in deepfake detection models for binary classification tasks, where the output is either real or fake. It measures the difference between the predicted probability and the actual class label. A lower loss value indicates better model performance. In deepfake detection, this function helps train neural networks by penalizing incorrect predictions more strongly. It ensures that the model learns to distinguish subtle differences between authentic and manipulated

videos, improving classification accuracy over time through optimization techniques like gradient descent.

$$\sigma(z_i) = \frac{e^{z_i}}{\sum_{j=1}^{K} e^{z_j}}$$

The convolution operation is the core mathematical function in Convolutional Neural Networks (CNNs), which are widely used in deepfake detection. It involves applying a filter (kernel) over an input image to extract important features such as edges, textures, and patterns. In deepfake detection, convolution helps identify visual artifacts and inconsistencies in manipulated frames. By stacking multiple convolutional layers, the model can learn complex hierarchical features. This operation is essential for analyzing spatial information in video frames and plays a critical role in detecting fake content.

$$C_t = f_t \cdot C_{t-1} + i_t \cdot \tilde{C}_t$$

This equation represents the update of the cell state in a Long Short-Term Memory (LSTM) network, which is used for analyzing temporal sequences in videos. The forget gate (f_t) decides what information to discard from the previous state, while the input gate (i_t) determines what new information to add. In deepfake detection, LSTM models analyze sequences of frames to identify inconsistencies over time, such as unnatural facial movements. This equation allows the model to retain important temporal features and discard irrelevant data, improving the detection of dynamic manipulations.

V. Conclusion

In conclusion, the rapid advancement of deepfake technology has introduced serious security challenges, particularly in video conferencing environments where authenticity and trust are essential. This project presented an AI-powered deepfake detection system designed to identify manipulated video content in real time and ensure secure communication. By leveraging advanced deep learning techniques such as Convolutional Neural Networks (CNNs) for spatial analysis and Recurrent Neural Networks (RNNs) or LSTM for temporal analysis, the system effectively detects inconsistencies in facial features and motion patterns.

The proposed system integrates multiple modules, including preprocessing, feature extraction, classification, and alert generation, to provide a comprehensive and efficient detection framework. Its ability to operate in real time makes it suitable for practical deployment in modern communication platforms. Additionally, the inclusion of continuous learning mechanisms allows the system to adapt to evolving deepfake techniques, ensuring long-term reliability.

Overall, this approach enhances cybersecurity by preventing identity fraud, misinformation, and unauthorized access during virtual meetings. As digital communication continues to grow, implementing intelligent deepfake detection systems will be crucial in maintaining the integrity, privacy, and trustworthiness of online interactions.

VI. Future Work

Although the proposed AI-based deepfake detection system is effective in improving video conferencing security, there are several opportunities for further enhancement. One key area for future work is strengthening the system's capability to detect more advanced and realistic deepfakes created using emerging generative technologies. As deepfake methods continue to evolve, detection models must be regularly updated with new datasets and improved algorithms to maintain accuracy.

Another important direction is enhancing the system's efficiency and scalability. Future improvements should focus on supporting high-quality video streams and managing multiple users simultaneously without affecting real-time performance. Incorporating multimodal analysis, which combines audio, video, and text data, can further improve detection accuracy by identifying cross-modal inconsistencies.

Additionally, developing lightweight and efficient models that can operate on edge devices like mobile phones and personal computers will reduce reliance on cloud computing. Strengthening privacy measures through techniques such as secure data handling and decentralized learning approaches will also be essential.

Future work can focus on seamless integration with existing video conferencing platforms and improving user experience, making the system more practical and widely usable in real-world applications.

References

1. M. S. Rana, M. N. Nobi, B. Murali and A. H. Sung, "Deepfake Detection: A Systematic Literature Review," IEEE Access, vol. 10, pp. 1–1, 2022.
2. D. Güera and E. J. Delp, "Deepfake Video Detection Using Recurrent Neural Networks," in Proc. IEEE AVSS, 2018, pp. 1–6.
3. S. Hussain, P. Neekhara, M. Jere, F. Koushanfar and J. McAuley, "Adversarial Deepfake," in Proc. IEEE WACV, 2021, pp. 3347–3356.
4. F. Matern, C. Riess and M. Stamminger, "Exploiting Visual Artifacts to Expose Deepfakes," in Proc. IEEE WACV Workshops, 2019.
5. D. Pan, L. Sun, R. Wang and X. Zhang, "Deepfake Detection through Deep Learning," in Proc. IEEE BDCAT, 2020, pp. 134–143.
6. Z. Tolosana et al., "DeepFakes Detection Across Generations," Engineering Applications of AI, 2022.
7. A. K. Singh and P. Singh, "Detection of AI-Synthesized Speech," in Proc. IEEE MIPR, 2021.
8. H. Khalid et al., "Evaluation of Audio-Visual Deepfake Dataset," in Proc. ACM Workshop, 2021.
9. A. Hamza et al., "Deepfake Audio Detection Using MFCC," IEEE Access, vol. 10, 2022.
10. M. Maksutov et al., "Deepfake Detection Based on Machine Learning," in Proc. IEEE EIConRus, 2020.
11. T. Nguyen et al., "Deep Learning for Deepfakes Creation and Detection," 2019.
12. L. A. Passos et al., "Deep Learning-Based Deepfake Detection: A Review," 2022.
13. P. Liu, Q. Tao and J. Zhou, "Multi-modal Deepfake Detection Survey," 2024.

14. J. Wang et al., “M2TR: Multi-modal Transformers for Deepfake Detection,” 2021.
15. Y. Zhang et al., “Global Multimedia Deepfake Detection,” 2024.
16. Y. Ju et al., “Improving Fairness in Deepfake Detection,” 2023.
17. R. Mubarak et al., “Survey on Detection and Impacts of Deepfakes,” IEEE Access, 2023.
18. K. Narayan et al., “Deepfake Source Identifier,” in Proc. IEEE CVPR, 2022.
19. P. Neekhara et al., “Adversarial Threats to Deepfake Detection,” in Proc. IEEE CVPR, 2021.
20. A. Mitra et al., “Machine Learning Approach for Deepfake Detection,” SN Computer Science, 2021.
21. Y. Mirsky and W. Lee, “The Creation and Detection of Deepfakes,” ACM Computing Surveys, 2021.
22. R. Tolosana et al., “Deepfake Detection Benchmark Analysis,” 2020.
23. J. Yang et al., “Detecting Fake Images Using Texture Differences,” 2021.
24. S. Suratkar and F. Kazi, “Deepfake Detection Using Transfer Learning,” 2023.

International Journal of Computational Research in Science and Technology
E-ISSN: 3108-0871, Volume 2 Issue 2
Mar-Apr 2026, PP 46-62