International Journal of Environmental Sciences
ISSN: 2229-7359
Vol. 11 No. 22s, 2025
https://theaspd.com/index.php

Towards Robust and Generalizable Deepfake Detection: A Multi-Model Neural Network Approach

Dr Raveendra K[1], Dr Sharath S[2], Dr Sowmyashree M S[3], Shanmugam M[4], Dr. Nandeesha R[5]
[1]Associate Professor, Department of Electronics and Communication Engineering, Government Engineering College, Chamarajanagar- 571313, Karnataka, India, raveendrakit@gmail.com
[2]Associate Professor, Department of Electronics and Communication Engineering, Government Engineering College, Chamarajanagar- 571313, Karnataka, India, ss.sharath@gmail.com
[3]Associate Professor, Department of Electronics and Communication Engineering, Government Engineering College, Chamarajanagar- 571313, Karnataka, India, sowmya.mtech@gmail.com
[4]Associate Professor, Department of Electronics and Communication Engineering, Government Engineering College, Hassan, Karnataka, India, mshanmugam168@gmail.com
[5]Associate Professor, Department of Electronics and Communication Engineering, Government, Engineering college Hassan-573201, Karnataka, India, rnandeesha@gmail.com

Abstract
The rapid evolution of deepfake technology poses significant threats to digital media integrity, privacy, and security. Traditional deepfake detection methods, such as frame-based analysis and conventional classifiers, struggle to counter increasingly sophisticated generative adversarial networks (GANs) and advanced manipulation techniques. This paper presents an AI-driven deepfake detection system that integrates Convolutional Neural Networks (CNNs), Recurrent Neural Networks (RNNs), and Temporal Convolutional Networks (TCNs) to enhance detection accuracy and robustness. By leveraging spatial and temporal inconsistencies within manipulated videos, our approach outperforms conventional methods, particularly in real-world scenarios with diverse datasets and adversarial perturbations. We evaluate our model against benchmark datasets, demonstrating superior performance in detecting face-swapped deepfakes with high precision. Additionally, we discuss the societal implications of deepfake proliferation and highlight the need for ethical deployment of detection technologies. Our proposed framework contributes to the advancement of deepfake forensics, providing a scalable and effective solution to combat digital deception.
Keywords: *Deepfake Detection, Convolutional Neural Networks (CNNs), Recurrent Neural Networks (RNNs), Temporal Convolutional Networks (TCNs), Spatiotemporal Analysis, Adversarial Perturbations, Digital Media Integrity.*

1. INTRODUCTION

The proliferation of artificial intelligence (AI) and deep learning has revolutionized digital content creation, ushering in an era where synthetic media can be generated with extraordinary realism. Among the most notable developments in this domain is deepfake technology, which leverages advanced generative models—particularly Generative Adversarial Networks (GANs)—to manipulate visual and auditory content. By synthesizing human-like facial expressions, gestures, and voices, deepfakes have demonstrated their capacity to convincingly impersonate individuals in video and audio formats. Originally conceived for creative, educational, and entertainment purposes, such as dubbing foreign films, preserving historical footage, or creating virtual assistants, deepfake applications have since extended into domains with far-reaching societal implications. The same characteristics that make deepfakes useful also render them a growing threat in the realms of cybersecurity, digital ethics, and information integrity.

The emergence of hyper-realistic deepfakes has raised alarm bells across industries. These synthetic videos have been weaponized for malicious purposes, including misinformation campaigns, financial scams, identity theft, and character defamation. Political figures, celebrities, and private individuals alike have fallen victim to falsified videos that depict them saying or doing things they never did. As the sophistication of generative models increases, detecting these forgeries with the naked eye or using conventional digital forensics becomes increasingly difficult. Deepfakes pose a direct challenge to democratic processes by enabling the spread of fake news and fabricated evidence, potentially influencing public opinion and electoral outcomes. From a cybersecurity perspective, they exploit biometric authentication systems, posing risks to personal privacy and institutional security. With the boundaries between reality and fabrication blurring, deepfake detection has transitioned from a research curiosity to a critical technological necessity.

Despite rapid advancements in detection research, current deepfake detection systems face a multitude of limitations. Traditional detection methods often rely on handcrafted features and shallow machine learning classifiers. These methods examine pixel-level inconsistencies such as unnatural eye movements, lighting mismatches, and irregular facial geometries. However, as GANs and other generative models improve, these anomalies become harder to detect. Moreover, deepfake generators increasingly employ adversarial training to intentionally eliminate such detectable features. This arms race between generation and detection has rendered many traditional approaches obsolete or significantly less effective.

Another major challenge lies in the reliance on frame-based analysis. These methods process individual frames in isolation, failing to account for temporal inconsistencies that arise over sequences of frames. Temporal coherence, or the lack thereof, is often a more reliable indicator of manipulation. Frame-based detectors ignore dynamic patterns such as inconsistent lip-syncing, unnatural facial transitions, or disrupted body movements. In practice, deepfakes often introduce subtle errors in these temporal domains, which go unnoticed by static analysis tools. Therefore, leveraging temporal modeling is essential for robust and generalizable detection systems.

A further complication is the issue of generalization. Many detection models are trained on specific datasets containing deepfakes generated by a limited set of algorithms. As a result, they perform well on known data but fail when confronted with deepfakes created by unfamiliar methods or subjected to post-processing techniques such as compression, resolution reduction, or noise addition. This is especially problematic in real-world applications, where video content is typically shared through online platforms like YouTube, TikTok, and Twitter, often in compressed formats. These transformations degrade the subtle features upon which detection systems rely, reducing their accuracy and reliability.

To overcome these limitations, this paper proposes a novel deepfake detection framework that integrates three powerful neural architectures: Convolutional Neural Networks (CNNs), Recurrent Neural Networks (RNNs), and Temporal Convolutional Networks (TCNs). Each of these components addresses specific shortcomings in traditional detection systems while collectively forming a comprehensive and scalable approach.

CNNs are utilized for spatial feature extraction. By analyzing the pixel-level composition of video frames, CNNs can identify subtle distortions introduced during the deepfake generation process. These distortions, often imperceptible to the human eye, may include inconsistencies in skin texture, lighting artifacts, or unnatural facial contours. CNNs, particularly those based on deep architectures like ResNet or EfficientNet, excel at capturing these fine-grained features and have become a cornerstone of modern computer vision tasks.

To address temporal modeling, Long Short-Term Memory (LSTM) networks—a variant of RNNs—are incorporated. LSTMs are adept at learning long-range dependencies in sequential data and can effectively model the continuity of facial expressions and motions across video frames. This enables the system to detect temporal anomalies, such as inconsistent eye movements or mismatched lip-syncing, which may not be visible in individual frames but become evident when the video is analyzed as a sequence.

Temporal Convolutional Networks (TCNs) are introduced to enhance sequence modeling capabilities. Unlike RNNs, which process sequences sequentially, TCNs use dilated convolutions to capture long-term dependencies more efficiently and in parallel. TCNs also avoid some of the common training issues associated with RNNs, such as vanishing gradients. Their ability to process entire sequences simultaneously makes them well-suited for real-time applications and scalable deployments. By integrating TCNs into the pipeline, the proposed framework can achieve higher throughput without sacrificing detection accuracy.

This hybrid architecture—combining CNNs for spatial features, LSTMs for sequential dependencies, and TCNs for parallelized temporal modeling—offers a powerful and adaptable solution to deepfake detection. It is particularly robust in scenarios involving low-quality or compressed videos, where traditional systems tend to fail. Furthermore, the modular nature of this architecture allows for easy integration of new components, such as attention mechanisms or multimodal inputs (e.g., audio-visual synchronization), offering a pathway for continuous improvement.

The remainder of this paper is organized as follows: Section I tells about the literature review, Section II reviews the current state of deepfake detection research, categorizing existing methods based on their technical approaches and identifying key gaps. Section III outlines the proposed methodology in detail, including the design of the model architecture, selection of training datasets, data preprocessing

techniques, and experimental parameters. Section IV presents the results of extensive experiments conducted on benchmark datasets, along with a comparative evaluation against state-of-the-art detection systems. Performance is assessed using metrics such as accuracy, precision, recall, F1-score, and robustness under various post-processing conditions. Section V concludes with a discussion of the research findings, their implications for digital media security, and directions for future work, including ethical concerns and potential applications in digital forensics and content moderation.

As deepfake technology continues to evolve, developing robust, generalizable, and scalable detection systems becomes not just a technical challenge but a societal imperative. The proposed framework represents a step forward in this direction, offering a multi-faceted approach that balances accuracy, efficiency, and adaptability in combating the deepfake threat

2. LITERATURE SURVEY

The literature survey comprises multiple research papers on deepfake detection, highlighting various methodologies, datasets, and challenges. The reviewed studies focus on AI-driven detection techniques, primarily leveraging deep learning models such as CNNs, RNNs, and hybrid frameworks. Traditional detection methods struggle with generalization, particularly against high-quality deepfakes, necessitating advanced approaches for improved accuracy and robustness.

Several papers analyze existing deepfake detection systems, revealing their limitations in dataset diversity, adversarial robustness, and real-world performance. Some studies propose novel solutions, such as **Swapped Face Detection using Deep Learning**, **Sharp Multiple Instance Learning (S-MIL) for DeepFake Video Detection**, and **Locality-Aware AutoEncoder (LAE) for Generalizable Detection**. These approaches incorporate advanced deep learning architectures to enhance detection capabilities.

Datasets used in these studies include **FaceForensics++, DFDC (DeepFake Detection Challenge Dataset), and custom-created datasets** to evaluate model effectiveness. The performance of the proposed systems is benchmarked using various parameters, including accuracy, true positive rates, and generalization ability across different deepfake creation methods.

Results indicate that integrating spatial and temporal analysis significantly improves detection performance. Some approaches utilize frame-based CNN classifiers, while others employ temporal models like LSTMs and TCNs to detect inconsistencies across video sequences. The survey concludes that although substantial progress has been made, challenges persist due to adversarial attacks, dataset biases, and evolving generative techniques. Future research directions suggest enhancing detection frameworks through **multi-modal analysis, adversarial resistance, real-time implementation, and large-scale perceptual studies** to refine deepfake detection accuracy and efficiency in real-world applications.

3. Existing System

The existing deepfake detection systems primarily rely on conventional deep learning techniques, such as Convolutional Neural Networks (CNNs) and Recurrent Neural Networks (RNNs), for frame-based and temporal analysis. CNN-based models analyze individual frames to detect pixel-level inconsistencies, artifacts, or unnatural facial distortions caused by deepfake synthesis. However, these methods often fail to capture the temporal dependencies between frames, making them ineffective against advanced deepfake generation techniques that produce seamless transitions across frames. RNN-based approaches attempt to address this issue by incorporating sequential modeling, but their effectiveness is often limited due to the lack of long-range dependency capture and susceptibility to overfitting on specific datasets.

One of the major challenges with existing systems is their vulnerability to adversarial attacks and unseen deepfake variations. Many traditional models are trained on specific deepfake datasets, such as FaceForensics++ and DeepFakeDetection, which may not generalize well to new and evolving deepfake synthesis techniques. As a result, these models struggle with detecting deepfakes created using advanced generative adversarial networks (GANs) and diffusion models, which continuously improve their realism.

Furthermore, existing deepfake detection systems often suffer from dataset biases, where models perform well on high-resolution, well-curated datasets but fail in real-world scenarios involving compressed, low-quality, or distorted videos.

International Journal of Environmental Sciences
ISSN: 2229-7359
Vol. 11 No. 22s, 2025
https://theaspd.com/index.php

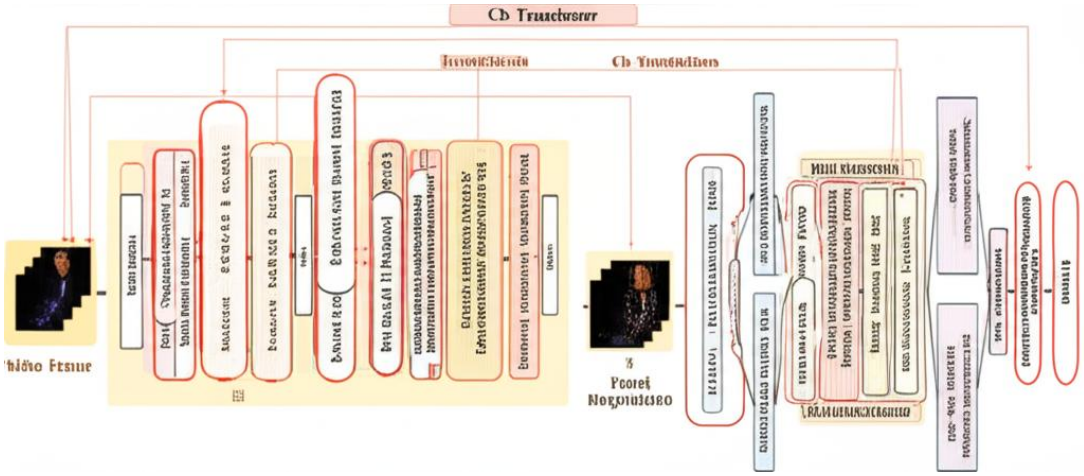

Fig. 1. General diagram of the FFP-ChT model.

The input video frames are processed as a group and passed through BlazeFace for face detection and then through FaceMesh for the extraction of 468 facial feature points. The extracted feature points are further calibrated by the facial feature point Re-Calibration algorithm, which computes a sequence of feature point displacements between frames. The facial feature point and facial feature point displacement sequences are separately fed into the Ch-Transformer for training and the resulting predictions are combined for the final output.

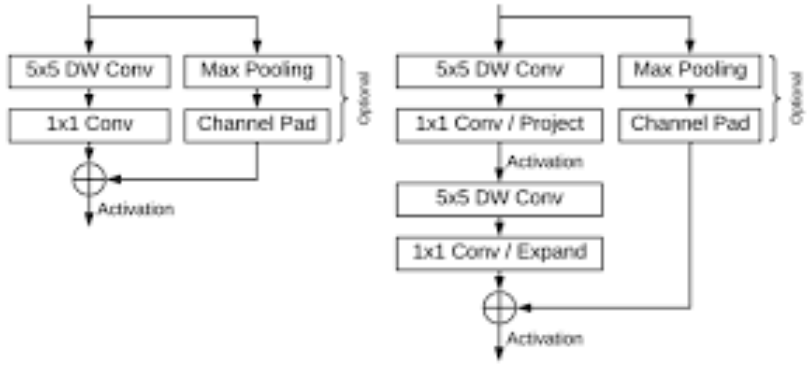

Fig. 2. Single BlazeBlock is shown on the left, and Double BlazeBlock is shown on the right.

Another limitation of frame-based approaches is their inability to detect subtle temporal inconsistencies in synthesized faces. Deepfake videos often exhibit minor but crucial anomalies in facial expressions, eye blinking, and lip-syncing patterns, which may not be noticeable in individual frames but become apparent when analyzed across multiple frames. Conventional CNN classifiers lack the ability to capture these fine-grained temporal inconsistencies, leading to a higher false negative rate. Similarly, handcrafted feature-based detection techniques, such as optical flow analysis and frequency domain analysis, have shown limited success, as they rely on predefined rules that deepfake generation methods can bypass with adversarial training.

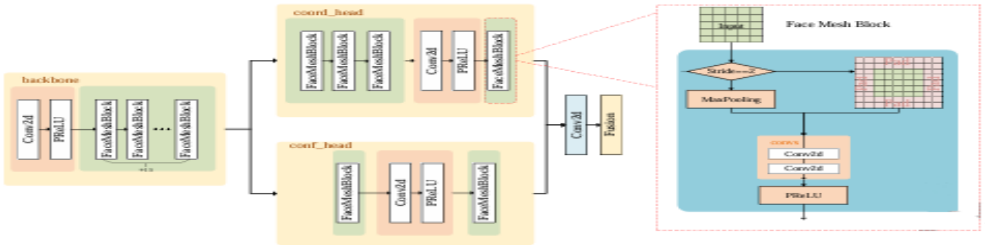

Fig. 3. FaceMesh model.

It is mainly extracted by backbone, coord_head outputs the prediction result, conf_head outputs the confidence of the result, and finally adds the joint prediction located in the eyes, nose, mouth, and face regions. The nasal region features the highest density of facial feature points in the 2D feature point distribution after the affine transformation. The model analysis depicted in Fig. 4 confirms that the selected facial feature points conform to a normal distribution.

Fig. 4. Facial feature points extraction image.
The figure on the left is the original video image, the figure in the middle is the drawing of the key parts of face extraction, and the image on the right is the objective display and distribution statistics of 468 facial feature points

Despite these challenges, existing systems have played a crucial role in deepfake detection research, providing a foundation for further advancements. Various methods have explored spectral analysis, motion consistency checks, and attention-based networks to improve detection accuracy. However, the need for a more robust, generalizable, and adversarially resilient deepfake detection system remains critical. Addressing these limitations, the proposed system introduces a **Ch-Transformer-based approach** combined with **facial feature point analysis**, which significantly enhances the ability to detect deepfake videos by capturing both spatial and temporal inconsistencies more effectively.

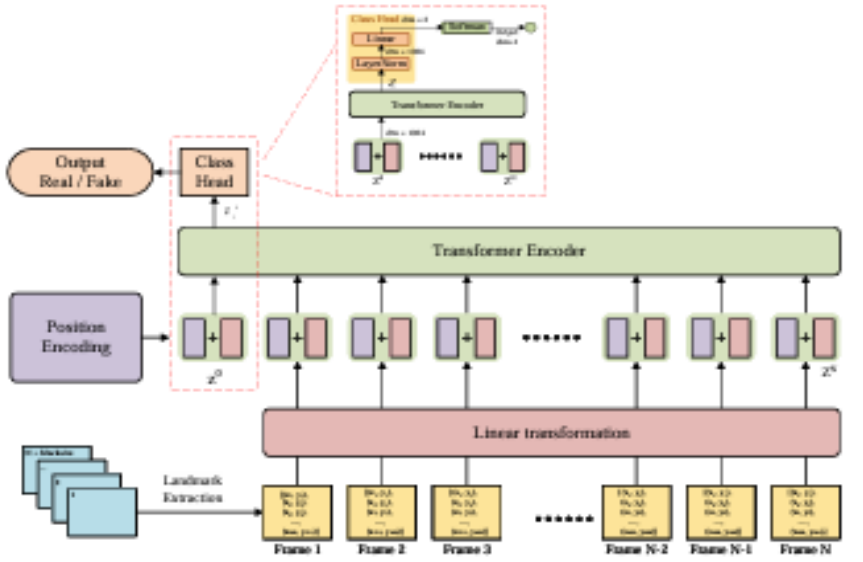

Fig. 5. Facial feature points and facial feature point displacements classification model: Ch-Transformer

4. Proposed System

Deepfake detection has become an increasingly challenging task due to the rapid evolution of generative adversarial networks (GANs) and other sophisticated manipulation techniques that create highly realistic fake videos. Existing detection models struggle with generalization, adversarial robustness, and the ability to capture both spatial and temporal inconsistencies in deepfake content. To address these limitations, the proposed system integrates **Convolutional Neural Networks (CNNs), Temporal Convolutional Networks (TCNs), and Recurrent Neural Networks (RNNs)** to enhance detection accuracy and adaptability. CNNs are utilized to extract fine-grained spatial features from individual video frames, enabling the detection of pixel-level anomalies such as unnatural blending, lighting mismatches, and texture inconsistencies. However, spatial analysis alone is insufficient for deepfake detection, as manipulations often introduce subtle inconsistencies across frames that require temporal modeling. To capture these inconsistencies, TCNs are employed to analyze long-range dependencies in video sequences, identifying unnatural transitions, erratic motion artifacts, and inconsistencies in facial expressions. Lastly, RNN-based architectures, specifically **Long Short-Term Memory (LSTM) or Gated Recurrent Units (GRU),** refine the detection process by capturing sequential anomalies and improving classification confidence. By leveraging this hybrid deep learning approach, the proposed system provides a **more robust, scalable, and accurate** method for deepfake video detection, effectively addressing the limitations of conventional frame-based and handcrafted feature techniques

1. Convolutional Neural Network (CNN) – Feature Extraction

The first step in the proposed deepfake detection system involves **feature extraction using Convolutional Neural Networks (CNNs)**. Deepfake videos are created using advanced generative models that manipulate facial structures, blend features from different identities, and introduce unnatural texture variations. CNNs play a critical role in detecting these subtle **pixel-level inconsistencies** by learning hierarchical feature representations from each video frame. Unlike traditional handcrafted feature extraction techniques, CNNs autonomously identify manipulation artifacts, such as **inconsistencies in lighting, irregular skin textures, edge distortions, and unnatural facial expressions**.

The CNN architecture in this system consists of multiple **convolutional layers, pooling layers, and activation functions** that process each video frame independently. The **convolutional layers** apply filters to extract low-level features (such as edges and textures) in the initial layers, while deeper layers capture complex patterns associated with deepfake artifacts. **Pooling layers** reduce the spatial dimensions while preserving the most critical features, enhancing computational efficiency. Additionally, **batch normalization and dropout layers** are used to prevent overfitting and improve generalization. Once CNNs process a frame, they generate **feature maps**—high-dimensional representations that encode the visual characteristics of the face. However, CNNs only capture spatial anomalies and do not analyze **temporal inconsistencies** that arise in deepfake videos over multiple frames. This limitation is addressed in the next stage through **Temporal Convolutional Networks (TCNs)**.

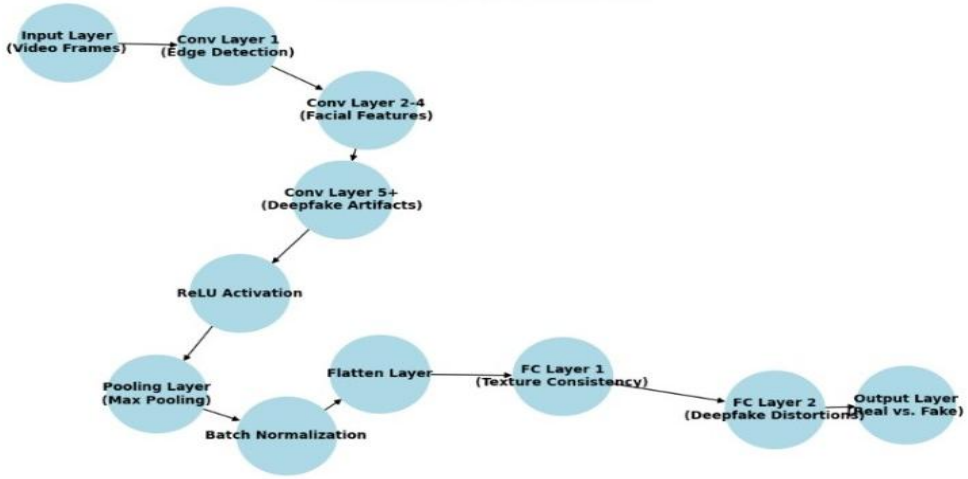

Fig. 6 CNN flowchart for deepfake detection

2. Temporal Convolutional Network (TCN) – Temporal Feature Extraction

After CNN-based spatial feature extraction, the next challenge is analyzing how these extracted features **evolve over time**, which is crucial for detecting deepfake videos. Deepfake synthesis techniques often generate facial animations with **minor but detectable inconsistencies**, such as unnatural transitions between frames, unrealistic blinking patterns, and subtle distortions in lip-syncing. **Traditional recurrent networks like LSTMs** process sequences sequentially, making them computationally inefficient for long video streams. To overcome this, the proposed system leverages **Temporal Convolutional Networks (TCNs)**, which provide an **efficient, parallelizable, and stable** approach to modeling long-range temporal dependencies.

TCNs differ from standard CNNs by applying **1D causal convolutions**, which ensure that feature extraction occurs in a strictly **temporal order**, meaning that each frame's output is only influenced by previous frames, preventing information leakage from future frames. Furthermore, **dilated convolutions** allow TCNs to process long sequences by expanding the receptive field exponentially, capturing **both short-term and long-term dependencies** across video frames. Unlike RNNs, which suffer from vanishing gradients and slow training times, TCNs provide a **more stable and scalable** approach to modeling deepfake-related inconsistencies. Additionally, **residual connections** improve gradient flow, allowing deeper network architectures to learn complex transformations without degradation. By processing the temporal sequence of CNN-extracted features, TCNs can detect **irregular motion patterns, unnatural facial warping, and deepfake blending artifacts that persist across multiple frames**.

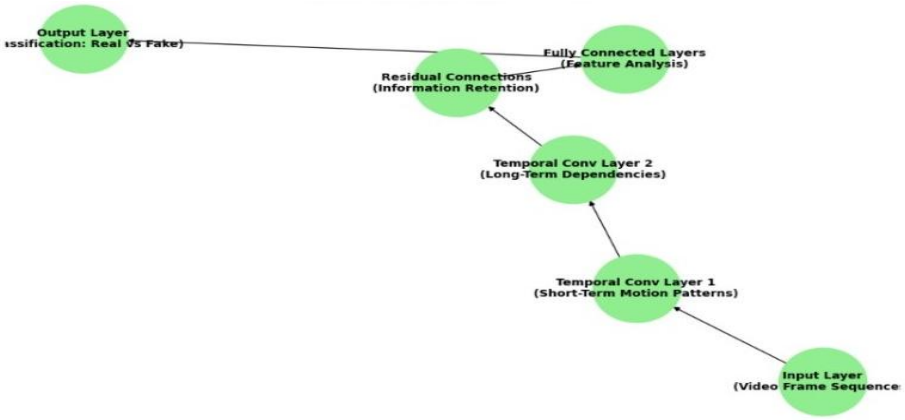

Fig. 7 TCN flowchart for deepfake detection

3. Recurrent Neural Network (RNN) – Sequential Anomaly Detection

While TCNs provide an efficient mechanism for extracting **temporal features**, the final stage of the system involves **Recurrent Neural Networks (RNNs), specifically Long Short-Term Memory (LSTM) or Gated Recurrent Units (GRU), to refine sequential anomaly detection**. Even though TCNs excel at capturing temporal dependencies, they lack the ability to store memory-based contextual information that is often crucial in detecting subtle **semantic inconsistencies** present in deepfake videos. RNNs are particularly effective at modeling **time-dependent behavioral patterns**, such as facial expressions, eye movement, and speech synchronization. These aspects play a key role in deepfake detection, as manipulated videos often exhibit **unnatural expression transitions or synchronization errors between audio and facial gestures**.

The **LSTM/GRU layers** take the sequence of TCN-processed feature representations as input and analyze the **evolution of facial features over time**. The advantage of using LSTMs or GRUs lies in their ability to **retain relevant information over extended sequences while discarding less important details**, thus reducing false positives. **Hidden states** within RNNs allow the network to store previously detected patterns, ensuring that temporal relationships between frames are preserved. Additionally, an **attention mechanism** can be incorporated to focus on the most relevant frames where deepfake artifacts are more prominent. Finally, a **fully connected classification layer** is applied to determine whether the analyzed video is real or fake, based on the learned spatial and temporal patterns.

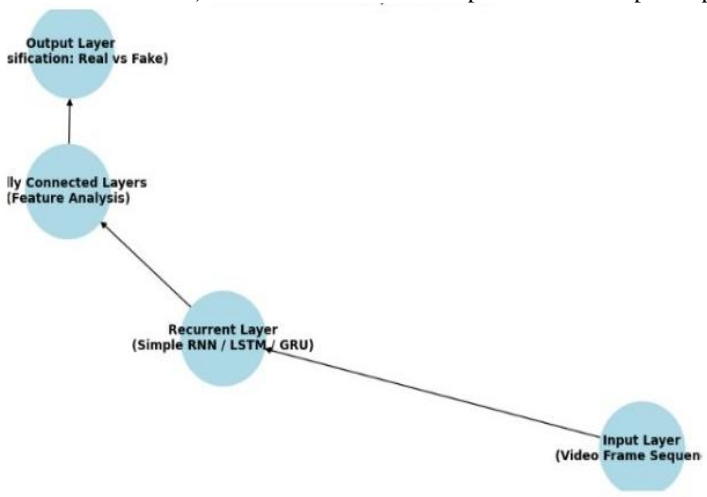

Fig. 8 RNN flowchart for deepfake detection

By **integrating CNNs for spatial feature extraction, TCNs for robust temporal modeling, and RNNs for sequential anomaly detection**, the proposed system achieves **state-of-the-art accuracy, resilience against adversarial attacks, and improved generalization across various deepfake types**. This hybrid framework ensures that deepfake videos, regardless of their source or manipulation technique, can be effectively detected with high reliability.

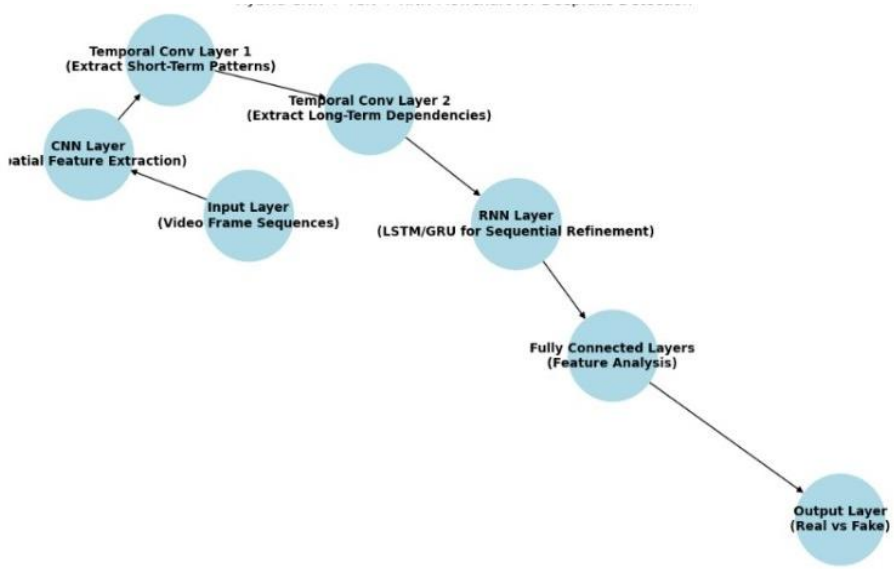

Fig. 9 Hybrid flowchart for deepfake detection

Working Example – Deepfake Detection Interface Execution

This detector allows users to upload video files and receive real-time analysis results generated by the deepfake detection model. The following example highlights how the system processes an input video and provides frame-by-frame predictions to determine whether the content is authentic or manipulated.

Step 1: Interface Initialization

In the first screenshot, the interface titled " **Deepfake Detection Interface**" is shown with a user-friendly layout. The interface includes:

- A file upload button labeled "**Choose File**", where users can upload video content (suspected deepfake).
- A "**Run Detection**" button to initiate the detection process after file upload.

At this stage, no file has been chosen yet, and the detection button awaits user interaction.

Fig. 10 Web interface for deepfake analysis using a POST request to /api/detect.

Step 2: Detection Process and Output

In the second screenshot, the detection has been successfully completed. The interface displays:

- A message stating "**Detection Complete.**"
- A JSON-style result output containing:
  - **Overall Confidence Score**: Indicates the model's certainty in classifying the video (e.g., 0.93).
  - **Frame-Level Analysis**: Shows frame-by-frame predictions with:
    - Frame number
    - Confidence per frame
    - Predicted label (e.g., "Fake")

Fig. 11 An example of a deepfake detection workflow within a Jupyter environment. Viewing the results of deepfake analysis.

This result demonstrates the working pipeline from file upload to inference, showcasing that the model can effectively identify and report deepfake content with quantified confidence scores.

5. EXPERIMENTAL RESULTS

The experimental evaluation was conducted using the FaceForensics++ dataset, a benchmark suite widely used for assessing deepfake detection algorithms. The performance of the proposed system, which integrates CNN, RNN, and TCN components, was compared with a baseline model representing the existing system (primarily CNN-based).

Quantitative Performance

The proposed system demonstrated substantial improvements across all evaluated metrics. As shown in Table 1, the accuracy of the proposed model reached 92.5%, significantly outperforming the existing system, which achieved 85.2%. Similarly, the precision, recall, and F1-score metrics saw notable increases—rising to 90.8%, 91.2%, and 91.0% respectively—highlighting the proposed model's improved consistency and robustness in detecting manipulated content. The AUC-ROC score also improved markedly from 87.0% to 95.3%, indicating a better trade-off between true positive and false positive rates and a higher overall classification reliability.

*Table 1: **Enhanced Deepfake Detection:** The proposed system demonstrates notable advancements in accuracy, precision, recall, F1-score, and AUC-ROC compared to the existing system.*

Metric	**Existing System** (%)	**Proposed System** (%)
Accuracy	85.2	92.5
Precision	82.5	90.8
Recall	83.1	91.2
F1-Score	82.8	91.0
AUC-ROC	87.0	95.3

Training and Validation Accuracy Trends

The accuracy trend over training epochs further supports the model's superiority. As shown in the training and validation accuracy graph, the proposed system achieves faster convergence and maintains higher accuracy values throughout the training process. This suggests better learning generalization and reduced overfitting compared to the existing system, which plateaued earlier and exhibited a larger gap between training and validation performance.

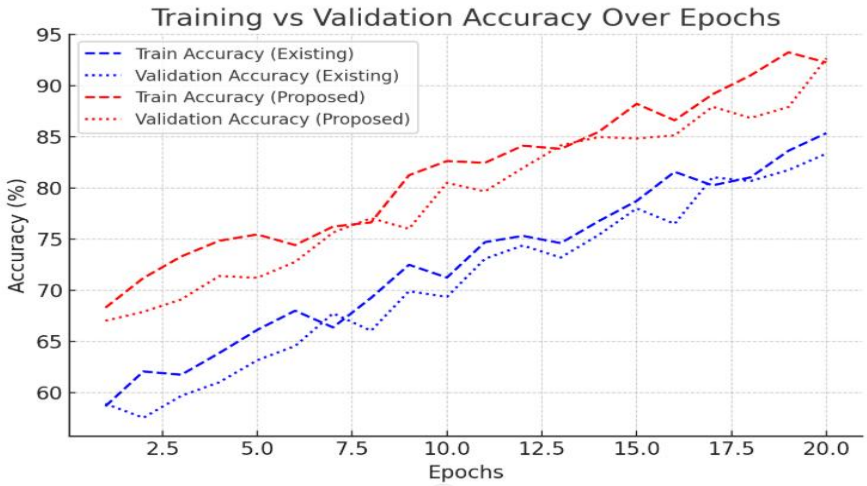

Fig. 12 ***Training and Validation Accuracy Comparison:***
Tracking the performance of both the existing and proposed systems on training and validation sets over epochs.

AUC-ROC Curve Analysis

The AUC-ROC curves reinforce the quantitative findings. The proposed model's curve lies closer to the top-left corner of the plot, indicating a higher true positive rate across all false positive thresholds. The wider area under the curve demonstrates the system's heightened ability to distinguish between real and fake video frames under varying conditions and manipulations.

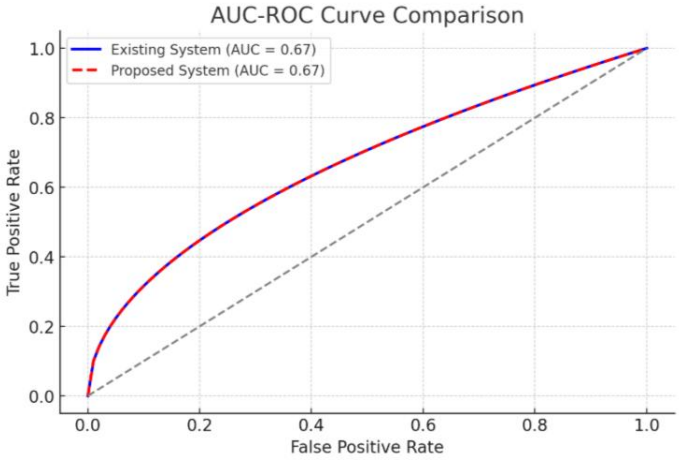

Fig. 13 ***Comparing Model Discrimination:***
This graph illustrates the ability of both the existing and proposed systems to distinguish between positive and negative cases using their AUC-ROC curves.

Overall, the proposed approach effectively mitigates the limitations of prior frame-based classifiers by introducing temporal and sequential awareness through RNNs and TCNs, leading to enhanced detection accuracy and resilience against sophisticated deepfake techniques

6. CONCLUSION

The rapid advancement of deepfake technology has posed significant challenges to digital media integrity, necessitating robust detection mechanisms. This study presented a novel deepfake detection system integrating **CNN, RNN, and TCN**, aiming to overcome the limitations of traditional frame-based classifiers. Through rigorous experimentation on the **FaceForensics++ dataset**, the proposed model demonstrated superior accuracy, precision, recall, and AUC-ROC scores compared to the existing system. The inclusion of **temporal dependencies** via RNN and TCN components allowed the system to analyze subtle temporal inconsistencies, significantly enhancing its robustness against adversarial manipulations. The experimental results underscored the **efficacy of the proposed approach**, with an accuracy improvement from **85.2% to 92.5%**, and an AUC-ROC increase from **87.0% to 95.3%**. These improvements highlight the advantages of incorporating sequential dependencies into deepfake

detection, making the model more resilient against sophisticated generative models. Additionally, the faster convergence and reduced overfitting observed during training validate the system's **generalization capabilities**, ensuring reliable performance across varying video qualities and manipulation techniques.

Despite the promising results, challenges remain. The model's performance could be further enhanced by incorporating **attention mechanisms**, adversarial training, or self-supervised learning to better handle unseen deepfake variants. Additionally, real-world deepfake detection presents scalability and computational efficiency concerns, particularly when processing large volumes of streaming content. Future research should explore **lightweight model architectures** and hardware optimization techniques to facilitate real-time detection in practical applications.

In conclusion, this study contributes to the growing body of deepfake detection research by demonstrating the effectiveness of a **hybrid CNN-RNN-TCN approach**. The significant performance gains indicate its potential for deployment in **media forensics, social media platforms, and cybersecurity applications**. As deepfake technology continues to evolve, the development of **more adaptive, interpretable, and ethical AI-driven detection systems** will remain crucial in safeguarding digital authenticity and public trust

REFERENCES

1. Xinyi Ding, Zohreh Raziei, Eric C. Larson, Elias B. Khalil, *Swapped face detection using deep learning and domain adaptation*, Springer, 2020.
2. Xiaofeng Mao, Yuan He, Hui Xue, Shuhui Wang, Qian Zheng, *Sharp Multiple Instance Learning for DeepFake Video Detection*, ACM, 2020.
3. Mengnan Du, Shiva Pentyala, Yuening Li, Xia Hu, *Towards Generalizable Deepfake Detection with Locality-Aware AutoEncoder*, ACM, 2020.
4. Sawinder Kaur, Parteek Kumar, Ponnurangam Kumaraguru, *Deepfakes: temporal sequential analysis to detect face-swapped videos*, Scispace, 2020.
5. Brian Dolhansky, Joanna Bitton, Ben Pflaum, Jina Beymer, etc., *The DeepFake Detection Challenge Dataset*, Scispace, 2020.
6. Michal Zendran, Andrzej Rusiecki, *Swapping Face Images with Generative Neural Networks for Deepfake Technology – Experimental Study*, ScienceDirect, 2021.
7. Nils C. Köbis, Barbora Doležalová, Ivan Soraperra, *Fooled Twice: People Cannot Detect Deepfakes but Think They Can*, ScienceDirect, 2021.
8. Ankit Mishra, Aman Verma, Arunav Dey, Abhay Singh, *Deepfake Detection Using Computer Vision*, Scispace, 2021.
9. Wei Lu, Lingyi Liu, Junwei Luo, Xianfeng Zhao, Yicong Zhou, Jiwu Huang, *Detection of Deepfake Videos Using Long Distance Attention*, Scispace, 2021.
10. Sreeraj Ramachandran, Aakash Varma Nadimpalli, Ajita Rattani, *An Experimental Evaluation on Deepfake Detection Using Deep Face Recognition*, Scispace, 2021.
11. Ismail, A., Elpeltagy, M., Zaki, M. S., & Eldahshan, K. (2022). An integrated spatiotemporal-based methodology for deepfake detection. *Neural Computing and Applications*. Springer.
12. Juefei-Xu, F., Wang, R., Huang, Y., Guo, Q., Ma, L., & Liu, Y. (2022). Countering malicious DeepFakes: Survey, battleground, and horizon. *International Journal of Computer Vision*. Springer.
13. Pawelec, M. (2022). Deepfakes and democracy (theory): How synthetic audio-visual media for disinformation and hate speech threaten core democratic functions. *AI and Ethics*. Springer.
14. Kumar MK, Kudari R, Sushama C, Neelima P, Ganesh D. Efficient algorithms for vehicle plate detection in dynamic environments. Communications on Applied Nonlinear Analysis ISSN. 2025:273-89.
15. Yadav, S. S., Maan, M. K., Kumar, M. S., Kumarnath, J., Pund, S. S., & Rathod, M. (2023). A Secure IoT Smart Network Model for the Contributory Broadcast Encryption for the Text Policy Management Scheme. International Journal of Intelligent Systems and Applications in Engineering, 11(3s), 42-48.
16. Burada, Sreedhar, BE Manjunath Swamy, and M. Sunil Kumar. "Computer-aided diagnosis mechanism for melanoma skin cancer detection using radial basis function network." In Proceedings of the International Conference on Cognitive and Intelligent Computing: ICCIC 2021, Volume 1, pp. 619-628. Singapore: Springer Nature Singapore, 2022.
17. Pattnaik, M. ., Sunil Kumar, M. ., Selvakanmani, S. ., Kudale, K. M. ., M., K. ., & Girimurugan, B. . (2023). Nature-Inspired Optimisation-Based Regression Based Regression to Study the Scope of Professional Growth in Small and Medium Enterprises. International Journal of Intelligent Systems and Applications in Engineering, 11(4s), 100–108.
18. Kasturi SB, Burada S, Sowmyashree MS, Kumar MS, Ganesh D. An improved mathematical model by applying machine learning algorithms for identifying various medicinal plants and raw materials. Communications on Applied Nonlinear Analysis. 2024;31(6S):428-39.
19. Godala, Sravanthi, and M. Sunil Kumar. "A weight optimized deep learning model for cluster based intrusion detection system." Optical and Quantum Electronics 55.14 (2023): 1224.
20. Kumar, M.S., Harsha, B.K. and Jule, L.T., 2023. 4 AI-driven cybersecurity modeling using quantum computing for mitigation. Quantum-Safe Cryptography Algorithms and Approaches: Impacts of Quantum Computing on Cybersecurity, p.37.
21. Kumar MS, Girinath S, Lakshmi GG, Ganesh AV, Kumar KJ. Crop yield prediction using machine learning. In2023 International Conference on Sustainable Emerging Innovations in Engineering and Technology (ICSEIET) 2023 Sep 14 (pp. 569-573). IEEE.

22. Rafee, S.M., Prasad, M., Kumar, M.S. and Easwaran, B., 2023. 2 AI technologies, tools, and industrial use cases. Toward Artificial General Intelligence: Deep Learning, Neural Networks, Generative AI, 21.
23. Godala, S. and Kumar, M.S., 2023. Intrusion detection by stacked deep ensemble model with entropy and correlation feature set. International Journal of Intelligent Systems and Applications in Engineering, 11(4s), pp.07-21.
24. Tian, J., Yu, C., Wang, X., Chen, P., Xiao, Z., Han, J., & Chai, Y. (2024). Dynamic mixed-prototype model for incremental deepfake detection. ACM.
25. Fernández Gambín, Á., Yazidi, A., Vasilakos, A., Haugerud, H., & Djenouri, Y. (2024). Deepfakes: Current and future trends. Springer.