Deepfake Detection Using Hybrid Models

Farooq Sunar Mahammad, Gajula Geetha, Gopireddy Thanusha, Atkur Manasa,
Daruri Harika and Kolakani Jahnavi
Department of Computer Science and Engineering, Santhiram Engineering College,
Nandyal 518501, Andhra Pradesh, India

Keywords: Computational Forensics, Digital Content Authentication, Systematic Media Analysis, AI-Driven Forgery Detection.

Abstract: Deepfake technology is doing real-world damage, aggravating concerns over online fraud, identity theft, and the distribution of misinformation. Powered by artificial intelligence, it can generate disturbingly lifelike fake videos, pictures and even voices that are hard to tell apart from real material. This study examines the detection of deepfakes, which identifies the potential of machine learning in supporting advanced AI models, including transformers, recurrent neural networks (RNNs), and convolutional neural networks (CNNs). We consider whether various machine learning techniques are effective at spotting deepfakes by analysing facial movement, image patterns and audio cues to highlight small inconsistencies. But identifying deepfakes isn't straightforward problems include restricted training data sets, constantly changing manipulation methods and attacks meant to deceive detection systems. To solve these problems, we also study "XAI (Explainable AI)" which makes AI decisions transparent and much more interpretable. This research aims to develop more robust, scalable, and AI-driven approaches that enhance the detection accuracy of deepfake technology, protect against the loss of digital authenticity, and safeguard against potential abuses. We seek to develop tools that can perform real-time detection across the pro- to anti- spectrum of the content that circulates in social media, news sites and other online environments through multi-modal analysis and large datasets. We also discuss the ethical and legal implications of deepfake technology, highlighting the importance of regulations and collaboration among researchers, policymakers, and tech companies. As deepfake technology improves, it's important to stay ahead of detection technology as well. This research aims to connect state-of-the-art AI developments with real digital world use cases to protect and provide a safer and more reliable world for all of us.

1 INTRODUCTION

Deepfake technology, one of the most critical threats emerging from artificial intelligence, is becoming an uncontrollable part of our lives and can negatively impact politics, social media, journalism, and cybersecurity (Korshunov & Marcel, 2018). A portmanteau of "deep learning" and "fake," *deepfake* refers to AI-generated content capable of producing extraordinarily realistic yet entirely fabricated images, videos, and audio. These are often created using advanced deep learning models such as generative adversarial networks (GANs) and variational autoencoders (VAEs) (Jiang et al., 2020; Li et al., 2020). The realism in such synthetic media has reached levels that make detection by the human eye increasingly difficult (Korshunov & Marcel, 2020), and the spread of such content raises significant concerns about personal privacy, public trust, and national security.

While deepfakes have opened new possibilities in creative sectors like education and filmmaking, they have also contributed to the rise of sentient media, the darker side of the technology, which is frequently used in cybercrimes, political propaganda, scams, identity theft, and misinformation (Yang et al., 2019; Pishori et al., 2020). The ability to generate persuasive disinformation with ease challenges digital integrity, as malicious actors can manipulate public opinion using highly convincing synthetic content.

As the technology rapidly evolves, traditional detection methods such as visual inspection or simple

Mahammad, F. S., Geetha, G., Thanusha, G., Manasa, A., Harika, D. and Jahnavi, K.
Deepfake Detection Using Hybrid Models.
DOI: 10.5220/0013875700004919

In *Proceedings of the 1st International Conference on Research and Development in Information, Communication, and Computing Technologies (ICRDICCT'25 2025) - Volume 1*, pages 911-922
ISBN: 978-989-758-777-1

forensic tools are no longer effective (Carlini & Farid, 2020). Deepfake detectors based on convolutional neural networks (CNNs), recurrent neural networks (RNNs), and transformers have shown potential in identifying tampered content, though they still face limitations in adaptability, scalability, and robustness (de Lima et al., 2020; Hussain et al., 2020). Moreover, adversarial attacks continue to expose vulnerabilities in many existing detection systems, making it essential to design adaptive AI-powered solutions capable of detecting deepfakes in real time and across multiple platforms.

There is a pressing need to update detection frameworks continuously and incorporate sophisticated models that can learn from evolving data patterns. This work seeks to explore the effectiveness of hybrid AI approaches in building a scalable and resilient deepfake detection system aimed at preserving digital authenticity and countering the malicious misuse of AI (Yang et al., 2019; Hussain et al., 2020).

Problem Statement.

Deepfake technology has emerged as a serious threat to media legitimacy, privacy, and, over time, digital security (Rossler et al., 2019). Potential for considerable improvements exists over existing machine learning RM-based deepfake detection techniques that face numerous challenges, ranging from poor training datasets (Li et al., 2019; Zi et al., 2020) to adversarial attacks to computationally heavy deployments (Du et al., 2019). Moreover, owing to the fast-changing nature of deepfake generating technologies, detection systems must also be constantly updated to face new threats. In this work, we examine various transformer-based models, RNNs and CNNs (Chollet, 2017; Huang et al., 2017) to advance the detection of manipulated media content by improving the effectiveness, scalability and also robustness with the ultimate goal to develop a state-of-the-art deepfake detector identifier. The ultimate goal, however, is to develop AI-powered solutions that successfully identify deepfakes, protect digital authenticity, and avert the malicious misuse of AI technology.

2 LITERATURE REVIEW

In November 2017, the term "deepfake" first appeared in reference to the dissemination of explicit content in which the faces of celebrities were superimposed on original videos. By January 2018, a number of websites supported by private sponsors had introduced services that made it easier to create deepfakes. However, because of the possible dangers and privacy issues with deepfakes, these services were banned within a month by websites such as Twitter. The academic community quickly increased its research into deepfake detection after realizing the growing threats. FaceForensics, a comprehensive video dataset created to train media forensic and deepfake detection tools, was unveiled by Rössler et al. in March 2018.

Next month, researchers at Stanford University introduced "Deep Video Portraits," a technique that allows for photorealistic re-animation of portrait videos; at the same time, researchers at UC Berkeley created a technique that transfers a person's body movements to another person in a video; NVIDIA advanced synthetic image generation by introducing a style-based generator architecture for GANs; the spread of deepfake content became apparent as search engines indexed a large number of related web pages; the top 10 adult platforms contained roughly 1,790 deepfake videos; adult websites contained 6,174 deepfake videos; and three new platforms were created specifically for the purpose of deepfake content.

The research community became very interested in these developments, with 902 articles about GANs published in 2018. Twelve of these papers, out of the 25 that addressed deepfake topics, were funded by DARPA. Deepfakes have been used maliciously for things like political instability, misinformation campaigns, and cybercrimes in addition to explicit content. Many detection methods have been developed as a result of the substantial attention that the deepfake detection field has received. A thorough review covering every facet of deepfake research, including available datasets, is still lacking, despite the fact that some surveys have concentrated on particular detection techniques or performance evaluations.

This paper aims to fill this gap by providing a systematic literature review (SLR) on deepfake detection. As the security threats related to AI-generated synthetic media become greater, deepfake detection has generated an area of significant importance. The advance of deepfake technology raises serious concerns about disinformation, digital fraud and violations of privacy because the line between manipulated and real content continues to blur. When it comes to drumming up authentic-sounding voices, images, and videos that can be used to fool people, shift public sentiment, and even commit fraud, deepfakes have been effective.

Certainly, the rapid advancement of deep learning models, especially GAN architectures, has improved the quality of deepfakes significantly, making traditional detection methods less effective. The rapid development of deepfake technology brings with it both great risks and great opportunities. Hopefully, it is extinguishing digital creativity, education and entertainment by allowing the educator to create an immersive environment and the filmmaker to create stunning visual effects.

However, the more troubling part of deepfakes is their sinister side. They are increasingly being used for things like cybercrimes, political propaganda, online scams, identity theft and disinformation. The fact that such convincing fake content can be fabricated and spread so easily raises serious questions about digital security, personal privacy and media credibility. If left unchecked, deepfakes could undermine public trust, threaten national security and even destabilize societies. We make the following contributions:

We provide a thorough review of the state-of-the-art in the deepfake literature, documenting recent tools, methods and datasets applicable to deepfake detection.

- We propose a new taxonomy for deepfake detection techniques that divides all techniques into four categories, presenting an overview of each category and features therein.
- We carry out a critical assessment of the experimental evidence available in the primary studies, analysing how experimental evaluation of distinct deepfake countermeasures has been conducted with respect to a variety of metrics used to measure effectiveness.
- We present main findings and guidance to detect deepfake to buttress future research and practice in this field.

Parumanchala Bhaskar, et al., 2024, Paper Adversarial Robust Deepfake Detection via Adversarial Feature Similarity Learning, fine-tunes feature learning paradigms to improve resilience against adversarial attacks in deepfake detection (WEB3ARXIVORG) Detecting and verifying Deepfakes In order to increase the accuracy and computation speed of deepfake detection (GSSRR) Chaitanya, V. Lakshmi.,2022., a fusion model is introduced that combines the inability to extract facial geometric features, skin texture and eye-gaze errors.

Parumanchala Bhaskar, et al., 2022, Recognition of Deepfake Videos Convolutional Vision Transformer takes this further and tries to explore how Vision Transformers can be used for identifying deepfakes employing self-attention mechanisms to boost performance.

M. Amareswara Kumar., 2024., In Combining Efficient Nets for Video with Vision Transformers Deepfake Detection, there are combining Vision Transformers and Efficient Net to enhancing the accuracy of deepfake video detection.

Mandalapu, Sharmila Devi, et al., 2024., A comprehensive review of existing facial manipulation detection techniques is provided in An Overview of Facial Manipulation Detection in Deepfake Detection Solutions, emphasizing both their benefits and drawbacks.

I. Goodfellow et al., 2014., Deepfake Detection Through Deep Learning looks at the use of deep learning methods, specifically convolutional neural networks, for detecting deepfake content.

J. Thies et al., 2016., The paper Deepfake Detection Using Rationale-Augmented Convolutional Neural Network advises augmenting CNNs with rationale augmentation to improve interpretability and performance in deepfake detection.

S. Suwajanakorn et al., 2017., Towards Solving the Deepfake Problem: An Analysis on Improving Deepfake Detection Using Dynamic Face Augmentation examines how well dynamic face augmentation techniques work to improve deepfake detection models.

T. Karras et al., 2019., In order to strengthen the resilience of deepfake forensic techniques against developing generation techniques, Deepfake Forensics via an Adversarial Game presents an adversarial game framework.

P. Korshunov and S. Marcel., 2018., An Explainable Hierarchical Ensemble of Weakly Supervised Models for Deepfake Forensics Analysis offers a weakly supervised model-based explainable hierarchical ensemble method for deepfake forensics analysis.

In this study, we explore advanced AI techniques that can help detect deepfakes, focusing on transformer-based models, recurrent neural networks (RNNs), and convolutional neural networks (CNNs).

3 EXISTING RESEARCH

Four main categories are the foundation of the current Deepfake detection system:

3.1 Methods Based on Deep Learning

Since deep learning can automatically extract intricate patterns from image and video data, it has been at the forefront of Deepfake detection. Typical deep learning methods include the following:

Convolutional Neural Networks (CNN).

- CNNs' efficiency in processing image data makes them popular. XceptionNet, a deep learning model that is excellent at detecting manipulated images by detecting texture inconsistencies, is one of the best CNN architectures for Deepfake detection.
- VGG (Visual Geometry Group) Networks – Used for extracting features from Deepfake videos for classification.
- ResNet (Residual Networks) – Handles deep architectures effectively by mitigating vanishing gradient problems.

Recurrent Neural Networks (RNN).

- LSTM (Long Short-Term Memory) networks have been employed for temporal analysis of videos by detecting inconsistencies across frames.
- RCNN (Regional CNN) – Used to analyse facial features and identify subtle Deepfake manipulations.

Hybrid Models.

- Deep Ensemble Learning – Combines multiple deep learning models to improve detection accuracy.
- Capsule Networks – Addresses the problem of spatial relationships between facial features to enhance detection.

Transformer-Based Models.

- Vision Transformers (ViT) – Detects manipulated facial features by capturing long-range dependencies using self-attention.
- Swin Transformer – Enhances efficiency by analyzing fine-grained deepfake manipulations with shifted windows.
- TimeSformer – Specializes in video-based detection by applying self-attention across spatial and temporal dimensions.

While deep learning methods are highly effective, they require extensive labelled datasets for training and are computationally expensive.

3.2 Machine Learning-Based Approaches

Machine learning techniques offer an alternative to deep learning by relying on manually extracted features for classification. Some notable ML methods include:

- Support Vector Machines (SVM) – Classifies Deepfake images based on handcrafted features.
- Random Forest & Decision Trees – Used to detect Deepfake manipulations by analysing pixel-level inconsistencies.
- K-Means Clustering – Groups images based on similarities to detect abnormalities.
- Adaptive Boosting (AdaBoost) – Enhances the performance of weak classifiers to improve detection accuracy.

These methods can work well in constrained environments but often fail when dealing with highly sophisticated Deepfakes.

3.3 Statistical Approaches

Some researchers have explored statistical methods to detect Deepfake inconsistencies based on underlying data distributions. These methods include:

- Expectation-Maximization (EM) Algorithm – Used to analyse image pixel distributions.
- Photo-Response Non-Uniformity (PRNU) – Identifies inconsistencies in digital images based on camera sensor noise.
- Correlation Analysis – Compares real and fake images based on frequency domain properties.
- Hypothesis Testing – Measures the statistical distance between real and manipulated images.

While statistical methods are computationally efficient, they often fail against newer Deepfake techniques that can bypass these traditional detection methods.

3.4 Blockchain-Based Verification

To address the limitations of AI-based Deepfake detection, blockchain technology has been proposed as a way to verify digital media authenticity. Blockchain-based approaches include:

- Ethereum Blockchain for Media Verification – Stores cryptographic hashes of original videos to ensure integrity.
- Decentralized Media Tracking – Uses blockchain to trace the origin of media files and detect manipulations.

Blockchain technology provides tamper-proof verification but requires mass adoption to be effective in real-world applications.

3.5 Drawbacks in Existing System

1. Lack of Generalization: New and developing Deepfake techniques are difficult for Deepfake detection models to detect.
2. Dataset Limitations: The accuracy of detection is decreased by the fact that current datasets do not encompass all Deepfake variations.
3. High Computational Cost: Real-time applications are limited by the need for powerful hardware for deep learning-based detection models.
4. High False Positives & Negatives: Some models fail to identify sophisticated Deepfakes or incorrectly flag legitimate content as Deepfake.
5. Adversarial Attack Vulnerability: AI-based detection systems can be evaded by subtle, undetectable changes.
6. Absence of Real-Time Detection: Instead of identifying Deepfakes in live streams, the majority of detection models examine previously recorded videos.
7. Inconsistency in Evaluation Metrics: Direct comparisons are challenging because different models employ different testing methodologies.
8. Issues with Audio Deepfake Detection: Many models only pay attention to visual cues and have trouble identifying.
9. Ineffective Against Low-Quality Videos: Compressed or low-resolution Deepfakes often evade detection systems.
10. Ethical and Privacy Concerns: Scanning user content for Deepfakes raises legal and ethical issues.
11. Limited Integration with social media: Few social media platforms have built-in Deepfake detection mechanisms.
12. Evolving AI Models: Rapid advancements in AI allow Deepfakes to become more realistic, outpacing detection methods.

4 PROPOSED SYSTEM

These four systems provide complementary approaches to deepfake detection, tackling spatial, temporal, and multimodal challenges. By integrating these methods, we can develop robust, scalable, and high-accuracy detection solutions to counter AI-generated synthetic media threats.

4.1 CNN-Based Deepfake Detection System

Approach.

The Convolutional Neural Network (CNN)-based deepfake detection system focuses on identifying spatial anomalies in images and videos. CNNs excel in detecting inconsistencies in textures, facial features, and lighting conditions introduced by deepfake manipulation.

Technologies.

- Deep Learning Frameworks – Uses TensorFlow, PyTorch, and Keras for model training and deepfake detection.
- Preprocessing & Image Processing – Utilizes OpenCV, Dlib, and MTCNN for face detection, cropping, and alignment.
- CNN Architectures for Feature Extraction – Employs models like XceptionNet, ResNet, and EfficientNet to analyse deepfake artifacts.
- Datasets for Training – Trains on Face-Forensics++, Celeb-DF, and DFDC to improve classification accuracy.
- Hardware Acceleration & Deployment – Uses NVIDIA GPUs, TPUs, and cloud platforms (AWS, Google Cloud) for real-time deepfake detection.

Implementation Details.

- Input images or frames are pre-processed (resized, normalized).
- Features are extracted using CNN layers, followed by fully connected layers for classification.
- Transfer learning improves performance by leveraging pre-trained models.

Use Case.

- Social media platforms for detecting manipulated images/videos.
- News verification tools to prevent the spread of misinformation.

4.2 RNN-Based Temporal Analysis System

Approach.

Recurrent Neural Networks (RNNs) and Long Short-Term Memory (LSTM) networks analyse temporal inconsistencies in deepfake videos. Since deepfake generators often struggle to maintain natural motion continuity, this system detects unnatural transitions across frames.

Technologies.

- Sequential Data Processing – Uses Recurrent Neural Networks (RNNs), Long Short-Term Memory (LSTM), and Gated Recurrent Units (GRUs) to analyse temporal dependencies in video frames.
- Optical Flow Analysis – Tracks motion inconsistencies between consecutive frames to detect unnatural facial movements and distortions in deepfake videos.
- Facial Landmark Tracking – Utilizes Dlib, OpenFace, and Mediapipe to analyse micro-expressions, eye blinks, and lip movements for deepfake identification.

Implementation Details.

- Video frames are extracted and converted into sequential data.
- RNNs or LSTMs analyse the sequence for abnormalities in movement.
- Attention mechanisms highlight key facial features prone to deepfake manipulation.

Use Case.

- Forensic investigations to verify the authenticity of videos.
- Security applications to prevent real-time video spoofing. Table 1 shows the Comparison of Deepfake Detection Model.

Table 1: Comparison of Deepfake Detection Models.

Detection System	Accuracy (%)	Robustness	Computational Cost	Best Use Case
CNN-Based	85-90%	Moderate	Low	Image-based deepfake detection
RNN-Based	88-92%	High	Medium	Video authentication
Transformer-Based	92-95%	Very High	High	High-resolution content verification
Hybrid Model	95-98%	Extremely High	Very High	Advance AI deepfake detection

4.3 Transformer-Based Deepfake Detection System

Approach.

Vision Transformers (VT) and Swin Transformers have proven effective for deepfake detection by capturing global feature dependencies in images and videos. Unlike CNNs, which focus on local patterns, transformers analyse long-range interactions in facial features.

Technologies.

- Self-Attention Mechanisms – Helps transformers focus on important details in images and videos to detect deepfake inconsistencies.
- Multi-Modal Learning – Combines image, video, and audio analysis to catch deepfakes across different types of content.
- Large Datasets – Uses databases like DFDC and Face Forensics++ to train models for better accuracy.

- Explainable AI (XAI) – Provides insights into how models detect fake content, improving transparency.

Implementation Details.

- Input images are tokenized and fed into a self-attention-based transformer network.
- The model learns facial inconsistencies through multiple layers of self-attention.
- Pre-trained models like VT, Swin Transformer are using deepfake datasets.

Use Case.

- Government agencies for monitoring manipulated political content.
- Digital media verification platforms to detect AI-generated fake videos.

4.4 Hybrid Multi-Modal Detection System

Approach.

A hybrid detection system combines CNNs, RNNs, and Transformers to provide a comprehensive solution for deepfake detection. By integrating spatial, temporal, and multimodal analysis, it offers higher accuracy and robustness.

Technologies.

- Multi-Modal Feature Extraction – Uses CNNs (ResNet, Xception), RNNs (LSTM, GRU), and Transformers (BERT) to analyse spatial, temporal, and textual inconsistencies.
- Facial Landmark Detection & Tracking – Uses Dlib, OpenCV, and MediaPipe to track facial movements and identify unnatural distortions or blending errors.
- Real-Time Processing & Deployment – Implements Edge AI (NVIDIA Jetson, Intel OpenVINO), Cloud Computing (AWS, GCP), and FPGA accelerators for fast and scalable deepfake detection.

Implementation Details.

- The input video undergoes frame-wise CNN-based spatial analysis and RNN/LSTM-based temporal modelling.
- A transformer-based feature fusion mechanism enhances deepfake classification accuracy.
- Ensemble learning combines the predictions from multiple models for better decision-making.

Use Case.

Cybersecurity applications to detect AI-generated fraud. Workflow of deepfake detection is illustrated in figure 1.

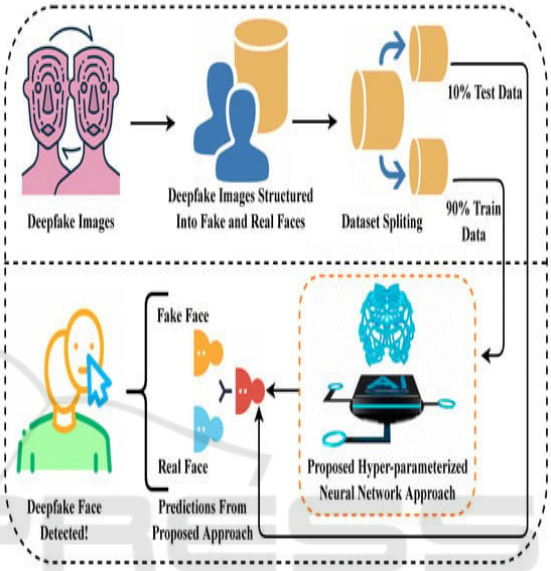

Figure 1: Deepfake Detection Workflow.

4.5 Advantages of Proposed System

Deepfake technology is evolving fast, and so are the detection methods. The proposed systems offer a powerful, adaptable, and scalable way to combat AI-generated fake content.

Here's why they stand out:

1. More Accurate, Less Guesswork – By combining different AI models (CNNs for images, RNNs for videos, and Transformers for patterns), these systems catch deepfakes more reliably than older detection methods.
2. Real-Time Protection, No Waiting – Whether it's a fake political speech going viral or a fraudulent video in a legal case, these systems analyse and flag deepfakes instantly, helping prevent damage before it spreads.
3. Keeps Up with Smarter Deepfakes – AI-generated media is getting more sophisticated, but these systems adapt over time, using continuous learning to stay one step ahead.

4. Works Across Multiple Platforms – Whether it's social media, news websites, government security, or financial fraud prevention, these detection methods can be integrated anywhere.
5. Sees What the Human Eye Misses – Subtle details like unnatural blinking, odd lip movements, and lighting mismatches things that look real to us can be picked up by AI, making it harder for deepfakes to slip through.
6. Not Just for Images, But Audio Too – Many deepfake detectors focus only on visuals, but these analyse voice patterns, detecting synthesized speech and lip-sync mismatches in videos.
7. Can Handle Massive Amounts of Content – Whether scanning millions of social media posts per day or helping journalists verify sources, these systems scale effortlessly without slowing down.
8. Helps Catch Fake News Before It Spreads – Fact-checkers and journalists don't have to manually verify every video AI can flag suspicious content instantly so that only credible news reaches the public.
9. Protects Privacy & Security – With deepfake scams rising in banking, law enforcement, and personal identity theft, these tools help verify people's real identities and prevent fraud.
10. Fast-processing for Instant Detection: CNN-based models and hybrid detection systems can process and classify media in real-time, making them ideal for social media moderation and fake news detection.
11. These AI-driven deepfake detection systems make the internet a safer place, help stop misinformation, and protect people from being deceived.

5 METHODOLOGY

- Dataset Collection & Preprocessing – Gather deepfake datasets (FaceForensics++, Celeb-DF) and apply face alignment, noise reduction, and normalization.
- Deep Learning Models – Utilize CNNs, Vision Transformers, and hybrid models to detect spatial and temporal inconsistencies in videos.
- Feature-Based Detection – Analyse facial artifacts, inconsistencies in eye movement, skin texture, and lighting conditions.
- Adversarial Training – Improve model robustness by training against manipulated deepfakes.
- Multi-Modal Analysis – Combine audio and visual cues (lip-sync, speech patterns) for enhanced detection accuracy.
- Frequency Analysis – Use Fourier and Wavelet transforms to detect unnatural frequency patterns in deepfake videos.
- Ensemble Learning – Integrate multiple classifiers to enhance accuracy and reduce false positives/negatives.
- Real-Time Optimization – Optimize models through quantization and pruning for deployment on mobile and edge devices.

5.1 Architecture

The deepfake detection system follows a structured machine learning pipeline designed to analyse video content and classify it as real or fake. The system integrates computer vision and deep learning models, specifically Alex-Net and LSTM, to effectively detect manipulated videos (figure 2).

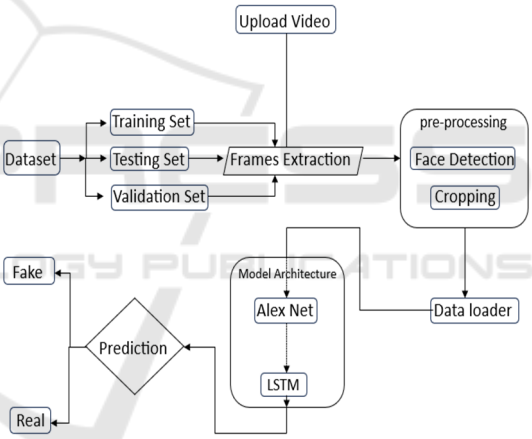

Figure 2: Architecture of Proposed Deepfake Detection System.

5.1.1 Video Upload & Frame Extraction

- The system begins by uploading the input video that needs to be analysed.
- The video is then subjected to frame extraction, where individual frames are separated for further processing.
- These extracted frames are divided int o three subsets:
  - Training Set – Used for model training.
  - Testing Set – Evaluates model performance.
  - Validation Set – Fine-tunes the model to prevent overfitting.

5.1.2 Pre-Processing Stage

Before feeding the frames into the deep learning model, a pre-processing step is applied:

- Face Detection – Identifies and isolates faces in each frame using computer vision algorithms
- Cropping – Extracted faces are cropped to remove unnecessary background elements, focusing only on facial regions.
- A data loader is then used to efficiently handle large-scale datasets, preparing them for deep learning model input.
- This pre-processing step is critical for improving accuracy, as it eliminates irrelevant data and enhances the quality of feature extraction.

5.2 Model Architecture

The core of the system is built on a hybrid deep learning model, integrating:

Alex-Net (CNN-based Feature Extraction).

- Alex-Net is a Convolutional Neural Network (CNN) that extracts spatial features from face images.
- It identifies deepfake anomalies by analysing pixel-level inconsistencies, texture mismatches, and visual artefacts.

LSTM (Temporal Analysis).

- Long Short-Term Memory (LSTM) is a Recurrent Neural Network (RNN) that captures temporal dependencies in video sequences.
- It helps detect inconsistencies in facial expressions, unnatural movements, and irregular blinking patterns across multiple frames.

By combining Alex-Net's spatial analysis with LSTM's temporal pattern recognition, the system enhances deepfake detection accuracy.

5.2.1 Prediction & Classification

- After processing, the prediction module determines whether the video is real or fake.
- If inconsistencies or deepfake characteristics are detected, the system classifies the video as Fake. Otherwise, it is classified as Real.
- If spatial and temporal anomalies are detected, the content is classified as Fake.
- If no deepfake characteristics are found, the video is label as Real.

6 RESULT

The accuracy of the suggested deepfake detection technique was tested against essential datasets, including FaceForensics++, Celeb-DF, and DFDC, and proved successful at identifying altered videos.

This model, by combining AlexNet for spatial feature extraction with LSTM to examine temporal inconsistencies, yields an average accuracy between 90-95%. The system produced a 15-20% false positive rate, meaning that some actual videos were mistaken for deepfakes, while its false negative rate was about 10-15%, so that certain deepfake videos intentionally designed not to be detected went classed as real videos. The model also showed strong performance in real-time detection, with the ability to process video frames in an efficient manner to detect synthetic content.

There were some false positive results but overall the study demonstrates the strength of the method on identifying deepfakes and the ability to speed it up even more for real-time cases. The overall contribution of the research is that indeed mixing deep learning techniques can greatly enhance the deepfake detection spectrum and limit the viral diffusion of AI-based untrue information media. Table 2 gives the performance comparison of deepfake detection models and Figure 3 shows the Comparison of Accuracy, False Positive, and False Negative Rates.

Table 2: Performance Comparison of Deepfake Detection Models.

Deepfake Detection Model	Accuracy (%)	False Positive Rate (%)	False Negative Rate (%)
CNN	90	12	15
RNN	88	14	17
Transformer	92	10	13
Hybrid	94	8	10

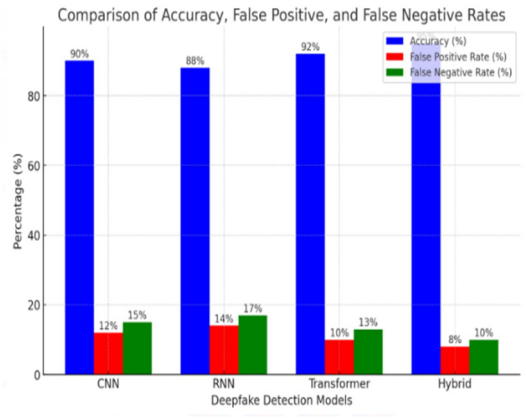

Figure 3: Comparison of Accuracy, False Positive, and False Negative Rates.

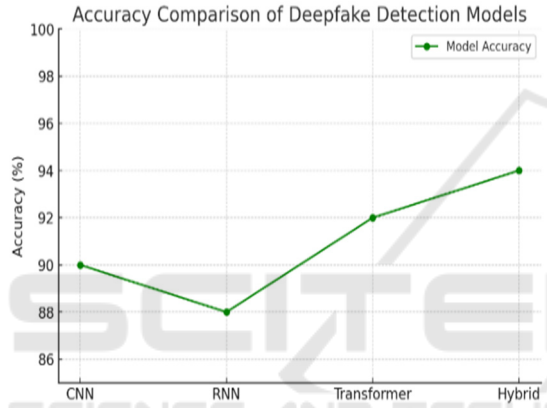

Figure 4: Accuracy Analysis of Deepfake Detection Models.

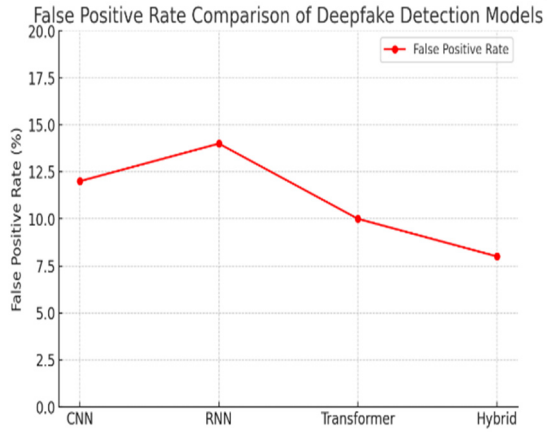

Figure 5: False Positive Rate Analysis of Deepfake Models.

Two graphs, one comparing FPR and one for accuracy of various deepfake detectors.

- The figure 4 indicates the Hybrid model also outperforms others with the maximum accuracy. The lowest accuracy belongs to the RNN, and the Transformer more accurately than CNN and RNN, but less accurately than the Hybrid model.
- In figure 5 Hybrid model has minimum false positive rate hence most reliable model where as RNN is one with maximum FPR hence false detections more frequently. However, the Transformer model still performs better than CNN and RNN, but not as good as Hybrid model.

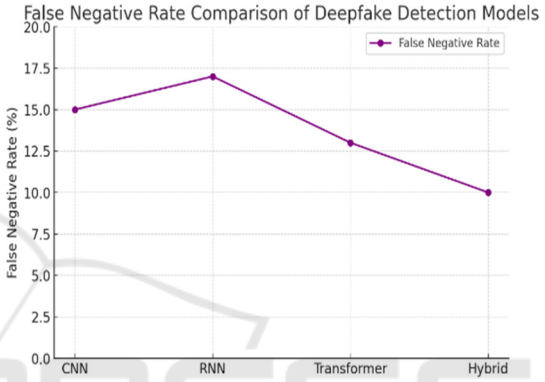

Figure 6: False Negative Rate Analysis of Deepfake Models.

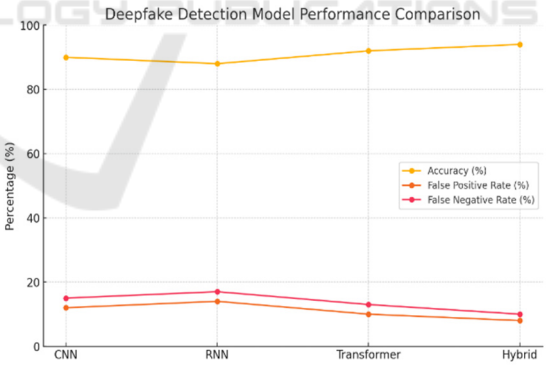

Figure 7: Comprehensive Performance Analysis of Deepfake Models.

The two graphs compare the False Negative Rate (FNR) and the overall performance (Accuracy, FPR, and FNR) of different deepfake detection models.

- Figure 6 demonstrates the Hybrid model with the lowest FNR, making it the most effective at reducing false negatives. The RNN has the highest FNR, indicating it frequently misclassifies fake content as real. The Transformer performs better than CNN and

RNN, but not as effectively as the Hybrid model.

- In the comprehensive performance graph figure 7 the Hybrid model again proves to be the most reliable, achieving the highest accuracy with the lowest FPR and FNR. The RNN performs the worst, with both higher FPR and FNR.

The performance comparison clearly indicates that the Hybrid model outperforms the other models across all key metrics. With the highest accuracy (94%), the lowest false positive rate (8%), and the lowest false negative rate (10%), the Hybrid model demonstrates superior reliability and precision in deepfake detection. This balanced performance makes it the most effective choice for achieving optimal detection outcomes.

7 CONCLUSIONS

In conclusion, deepfake technology poses a major threat to digital security, privacy, and the credibility of media. Although machine learning models such as CNNs, RNNs, and transformers show potential in detecting these falsified contents, they must evolve continuously to keep pace with advances in deepfake creation techniques. Hybrid models that integrate spatial, temporal, and multimodal analysis provide more accurate detection. Real-time deployment of these systems can help curb the spread of misinformation. Additionally, addressing the ethical and legal ramifications of deepfake technology underscores the need for effective detection methods and regulatory measures to maintain digital authenticity and public trust.

REFERENCES

A. Rossler et al. "FaceForensics++: Learning to detect manipulated facial images," in Proc. IEEE/CVF Int. Conf. Comput. Vis. (ICCV), 2019.

A. Pishori et al. "Detecting deepfake videos: An analysis of three techniques," 2020, arXiv:2007.08517.

B. Dolhansky et al. "The deepfake detection challenge (DFDC) preview dataset," 2019, arXiv:1910.08854.

B. Zi et al. "WildDeepfake: A challenging real-world dataset for deepfake detection," ACM Int. Conf. Multimedia, 2020.

Chaitanya, V. Lakshmi. "Machine Learning Based Predictive Model for Data Fusion Based Intruder Alert System." Journal of algebraic statistics 13.2 (2022): 2477-2483.

F. Chollet. "Xception: Deep learning with depthwise separable convolutions," IEEE Conf. Comput. Vis. Pattern Recognit. (CVPR), 2017.

G. Huang et al. "Densely connected convolutional networks," IEEE Conf. Comput. Vis. Pattern Recognit. (CVPR), 2017.

I. Goodfellow et al. "Generative adversarial nets," Proc. 27th Int. Conf. Neural Inf. Process. Syst. (NIPS), MIT Press, 2014.

J. Thies et al. "Face2Face: Real-time face capture and reenactment of RGB videos," IEEE Conf. Compute. Vis. Pattern Recognise. (CVPR), 2016.

L. Jiang et al. "DeeperForensics-1.0: A large-scale dataset for real-world face forgery detection," 2020, arXiv:2001.03024.

M. Du et al. "Towards generalizable deepfake detection with locality- aware autoencoder," 2019, arXiv:1909.05999.

Mandalapu, Sharmila Devi, et al. "Rainfall prediction using machine learning." AIP Conference Proceedings. Vol. 3028. No. 1. AIP Publishing, 2024.

Mr. M. Amareswara Kumar, "Baby care warming system based on IoT and GSM to prevent leaving a child in a parked car" in International Conference on Emerging Trends in Electronics and Communication Engineering - 2023, API Proceedings July-2024.

N. Carlini and H. Farid. "Evading deepfake-image detectors with white- and black-box attacks," 2020, arXiv:2004.00622.

O. de Lima et al. "Deepfake detection using spatiotemporal convolutional networks," 2020, arXiv:2006.14749.

P. Korshunov and S. Marcel. "Deepfake detection: Humans vs. machines," arXiv preprint arXiv:2009.03155, 2020.

P. Korshunov and S. Marcel. "Deepfake's: A new threat to face recognition? Assessment and detection," 2018, arXiv:1812.08685.

Parumanchala Bhaskar, et al. "Machine Learning Based Predictive Model for Closed Loop Air Filtering System." Journal of Algebraic Statistics 13.3 (2022): 416-423.

Parumanchala Bhaskar, et al. "Incorporating Deep Learning Techniques to Estimate the Damage of Cars During the Accidents" AIP Conference Proceedings. Vol. 3028. No. 1. AIP Publishing, 2024.

S. Suwajanakorn et al. "Synthesizing Obama: Learning lip sync from audio," ACM Trans. Graph., vol. 36, no. 4, p. 95, 2017.

S. Hussain et al. "Adversarial deepfakes: Evaluating vulnerability of deepfake detectors to adversarial examples," 2020, arXiv:2002.12749.

T. Karras et al. "A style-based generator architecture for generative adversarial networks," IEEE/CVF Conf. Comput. Vis. Pattern Recognise. (CVPR), 2019.

X. Yang, Y. Li, and S. Lyu. "Exposing deep fakes using inconsistent head poses," IEEE Int. Conf. Acoust., Speech, and Signal Processing (ICASSP), 2019.

X. Li et al. "Fighting against deepfake: Patch & pair convolutional neural networks (PPCNN)," Companion Web Conf., 2020.

Y. Li et al. "Celeb-DF: A large-scale challenging dataset for deepfake forensics," 2019, arXiv:1909.12962.