Enhanced Deepfake Detection Using Temporal Segment Networks

*DR Kavitha Subramani*

*Department of computer science*

*Panimalar Engineering College*

*Chennai ,India*

[kavithapec2022@gmail.com](mailto:kavithapec2022@gmail.com)

Jayashree A

*Department of computer science*

*Panimalar Engineering College*

*Chennai, India*

[ajayashree13@gmail.com](mailto:ajayashree13@gmail.com)

*Leena Sri M P Department of computer science*

*Panimalar Engineering College*

*Chennai ,India*

[leenasripslg14@gmail.com](mailto:leenasripslg14@gmail.com)

Kethsia I

*Department of computer science*

*Panimalar Engineering College*

*Chennai ,India*

[*keths003@gmail.com*](mailto:kavithapec2022@gmail.com)

*Abstract*— The widespread use of deepfake technology has Deepfake videos, which are highly realistic but fake, contribute to misinformation and pose serious digital security risks. Detecting them remains challenging, especially in complex scenarios. This study proposes an advanced deepfake detection system integrating Temporal Segment Networks (TSNs) and 2D face analysis. The system employs a TSN framework to detect temporal inconsistencies, combining LSTM with temporal attention, ResNet for spatial feature extraction, and PWC-Net for motion analysis. Training on diverse datasets ensures robustness, while a decision-level fusion technique enhances accuracy.This comprehensive solution aims to improve detection accuracy, reduce false positives and negatives, and foster trust in AI-driven evaluations.

Keywords—Deepfake detection,Temporal Segment Networks (TSNs), LSTM, ResNet, PWC-Net

# Introduction

The capabilities of artificial intelligence (AI) in media creation have been completely transformed by the quick development of deepfake technology. By using methods like Generative Adversarial Networks (GANs), deepfake systems are able to modify audio and visual content to create videos that are incredibly lifelike but fake. Although these developments highlight AI's potential in the creative and educational domains, their abuse poses serious risks to national security, individual privacy, and public confidence.

Malicious uses of deepfake technology have included online harassment, political manipulation, and the spread of false information. For instance, there were worries expressed regarding the use of deepfake films to spread misinformation and sway public opinion during the 2020 U.S. elections. Beyond politics, deepfakes have been used to produce fake videos for defamation, extortion, and other negative purposes, costing people and organizations money and harming their reputations. An urgent need for efficient detection systems that can counter these risks has arisen as a result of this increasing misuse.

The development of deepfake technology has produced exceptionally realistic yet fake videos, presenting significant threats to digital security and disinformation. Current detection techniques frequently fall short of the complexity of contemporary deepfakes. The goal of this research is to create a sophisticated detection system that uses Temporal Segment Networks (TSNs), and 2D face modeling for enhanced accuracy and transparency even in challenging circumstances.

It is crucial to address the issues raised by deepfake technology in order to protect digital media's security and integrity. A strong deepfake detection system can reduce the danger of false information, preserve public confidence in information systems, and shield people from identity theft.

# Background

# The rise of deepfake technology, driven by advanced machine learning methods like Generative Adversarial Networks (GANs), has revolutionized digital media production. Deepfakes allow for the alteration of visual and auditory content to generate highly convincing yet artificial videos, sparking concerns over their potential misuse in disinformation, privacy breaches, and the erosion of trust in digital interactions.

The Deepfakes leverage the power of deep learning to synthesize and alter video and audio content convincingly. GANs are at the core of this technology, with one network (the generator) creating fake content and another (the discriminator) assessing its authenticity. This adversarial training results in outputs that closely mimic real-world data. Over time, these methods have evolved to produce near-perfect replicas of human faces, voices, and expressions

# Challenges In Deepfake Detection

Despite efforts to counteract deepfakes, current detection techniques face significant hurdles:

High Fidelity: Modern deepfakes exhibit subtle inconsistencies that are difficult for traditional systems to detect.

Adaptability: Deepfake generation methods continuously evolve, outpacing static detection models.

Environmental Variations: Variability in lighting, camera angles, and motion poses challenges to generalization.

Real-Time Detection: Existing methods often lack the efficiency required for live applications, such as social media monitoring or live-streamed events.

Recent advancements in AI have introduced promising approaches to deepfake detection. Techniques incorporating Convolutional Neural Networks (CNNs) for spatial analysis, Long Short-Term Memory (LSTM) networks for temporal modeling for interpretability have shown potential..

This project builds upon these advancements, aiming to develop a comprehensive detection framework withTemporal Segment Networks (TSNs)to address the limitations of existing methods and provide a robust solution for deepfake detection.

Objectives

The main objective of this research is to solve the major issues of accuracy, generality, and transparency in order to create a sophisticated system for identifying deepfake movies. The following are the precise goals:

*Development of Preprocessing Pipelines:* Create and put into place a productive preprocessing pipeline to get video data ready for analysis while making sure the input is reliable and ideal for training models.

*2D Facial Feature Extraction:* Extract and analyze 2D facial features from video frames to capture critical spatial information that aids in the identification of subtle inconsistencies indicative of deepfakes.

*Initial Deepfake Classification:* Using sophisticated machine learning models to improve detection accuracy, classify films into real and fake categories initially using the retrieved facial features.

*Decision-Level Fusion:* Implement a decision-level fusion approach to combine predictions from multiple models, including 2D analyses, to improve the overall detection accuracy and robustness.

# Related Works

Deepfake detection has become a critical research area with the rise of deep learning-based manipulations. A study benchmarking 13 detection methods emphasizes the need for universal metrics to handle evolving deepfake techniques [1]. Similarly, research on deepfake creation methods using GANs and autoencoders highlights the challenge of building generalized detection models [2].

A deepfake attribution model analyzes spatial and temporal features to classify manipulated content based on the specific generation technique used [3]. Corneal reflections have been explored for real-time detection in video conferencing without specialized hardware [4]. Hybrid models, such as Xception-LSTM with attention mechanisms [5] and EfficientNet-TimeSformer [6], outperform traditional models in accuracy and computational efficiency. Preprocessing techniques significantly improve deepfake detection, particularly for facial feature analysis in Xception-based models [7]. CNN-SVM hybrids also enhance accuracy over individual machine learning models [8], while irregularities in facial movements across frames expose deepfake manipulations [9].

SPNet has been developed to optimize spatial and temporal feature extraction for large-scale detection with reduced computational complexity [10]. The DFFMD dataset, designed for face-mask deepfakes, improves detection accuracy for masked manipulations [11]. CNN-MLP hybrid models strengthen media forensics applications, leveraging CNNs for feature extraction and MLPs for classification [12]. Adversarial training techniques improve robustness against evolving deepfake methods [13]. Multi-modal systems integrating image and audio data enhance deepfake detection by identifying inconsistencies in both visual and auditory features [14].

A related approach examines facial movements and speech patterns together for improved accuracy [15]. Ensemble learning combines multiple detection models to reduce false positives and negatives, increasing overall reliability [16]. New datasets like Celeb-DF benchmark detection methods, ensuring their effectiveness against high-quality deepfakes [17]. Another dataset focuses on detecting manipulated content in news broadcasts to counter misinformation [18]. Meta-learning improves detection models' efficiency by enabling learning from fewer examples, facilitating adaptation to new deepfake techniques [19].

Cross-domain deepfake detection remains a challenge, with domain adaptation techniques improving generalization across different manipulation methods [20]. Ethical concerns also play a role, as researchers stress the need to balance high detection accuracy with privacy and fairness considerations [22]. The continued advancement of deepfake technology necessitates proactive measures in detection research, ensuring that detection methods stay ahead of increasingly sophisticated manipulation techniques.

# Methodology

This deepfake detection methodology leverages Temporal Segment Networks (TSNs) to tackle the challenges of identifying manipulated videos while ensuring accuracy and transparency. The process begins with a preprocessing pipeline that prepares video data for analysis, followed by extracting 2D facial features to identify subtle spatial anomalies. TSNs then analyze temporal dependencies between video frames, capturing irregularities in facial movements over time. To enhance detection precision, a decision-level fusion technique combines insights from both 2D models. The initial step involves preparing video data for analysis, ensuring its quality and suitability for feature extraction. This phase includes face detection, alignment, and normalization to standardize the video inputs.

1. Face Detection and Alignment: Utilize models like MTCNN or OpenFace to detect and align faces in video frames for consistency.

2. Normalization: Adjust lighting, resize frames, and stabilize camera motion to minimize variations that could affect feature extraction.

The second step extracts facial features from 2D video frames using Convolutional Neural Networks (CNNs) to identify distinguishing facial expressions and movements that can indicate real or fake content.

1. ResNet for Feature Extraction: Use a pre-trained ResNet architecture to extract spatial features such as eyes, nose, and mouth, as well as skin textures.

2. Frame Analysis: Apply CNNs to capture facial details, looking for irregularities like unnatural textures or pixel-level inconsistencies that could suggest manipulation.

To analyze the temporal dependencies in video sequences, TSNs are employed. TSNs segment the video into temporal chunks and process them sequentially, identifying anomalies in facial movements or blinking patterns.

1. Segmentation of Video: Divide the video into segments to reduce computational complexity and improve the analysis of long video sequences.

2. LSTM and Temporal Attention Mechanism: Combine TSNs with LSTM networks and temporal attention to capture the evolution of facial features and detect issues like unnatural transitions and inconsistent movements.

3. Motion Analysis with PWC-Net: Use the PWC-Net optical flow network to analyze motion between frames, identifying inconsistencies in facial movement and expressions.

This step enhances the system's performance by integrating outputs from both 2D models. The fusion process combines spatial and temporal features, increasing detection robustness against various types of deepfake manipulations.

1. Fusion of Predictions: Outputs from 2D spatial analysis (ResNet-50) and 3D depth model predictions are combined, leading to a more accurate and robust classification.

2. Improved Robustness: By detecting both spatial inconsistencies and temporal anomalies, the system is more adaptable to new deepfake techniques and more accurate in real-world conditions.

3. Interactive Dashboard: A dashboard allows users to interact with the system, upload videos, and view explanations through visualizations like heatmaps, improving transparency.

Training involves using multiple diverse datasets and optimizing the model for both detection accuracy and interpretability. The evaluation process employs various metrics to assess performance.

1. Dataset Selection: A range of datasets (e.g., FaceForensics++, Celeb-DF) ensures the model is exposed to various deepfake manipulation types.

2. Loss Function: The loss function balances accuracy and interpretability, ensuring both high performance and explainability.

3. Evaluation Metrics: The model’s performance is measured using accuracy, precision, recall, F1-score, and AUC, to ensure it works well across all types of deepfake content.

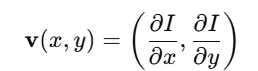
Once deepfake detection is performed, results are presented with added transparency, allowing users to understand the reasoning behind the detection.

1. Confidence Scoring: Each detection is accompanied by a confidence score, indicating how certain the model is about the authenticity of the video.

2. Visualization of Detection Results: Heatmaps and other visual aids help users understand which aspects of the video led to the final decision.

3. User Interaction: An interactive dashboard enables users to explore the detection process in greater depth.

Optical flow estimation using PWC-Net is essential for detecting motion inconsistencies in deepfake videos. It captures variations in pixel intensity over time, which helps identify unnatural movements in manipulated content. The optical flow vector is represented mathematically as:



where I(x, y, t) represents the pixel intensity at position (x, y) and time t. The partial derivatives ∂I/∂x and ∂I/∂y measure the change in intensity across frames, helping to detect motion distortions. This is crucial because deepfake videos often struggle with natural motion consistency, making optical flow analysis a key detection tool.

For temporal sequence modeling, LSTM (Long Short-Term Memory) networks are employed to track patterns across video frames. LSTMs use gating mechanisms to control information flow, ensuring that long-term dependencies are maintained while irrelevant details are discarded. The key equations governing LSTMs are:

Forget Gate: Controls which past information should be discarded.



The Input Gate regulates the new information that gets stored in memory.



Cell State Update: Updates the long-term memory with relevant information.

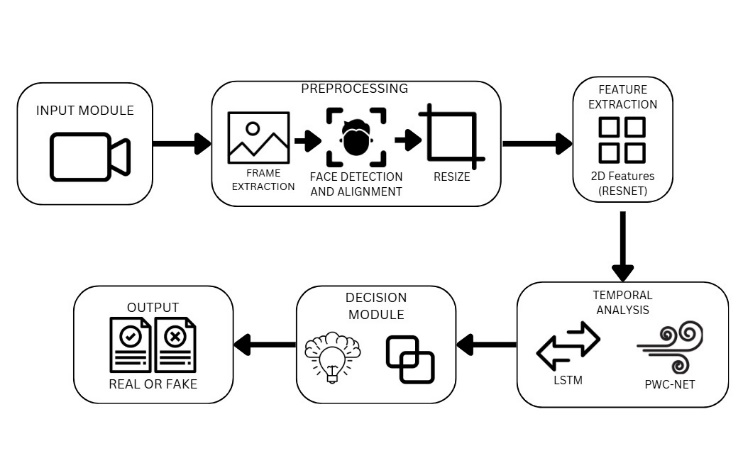


Output Gate: Regulates the final output of the LSTM.



These equations ensure that the model learns temporal dependencies across frames, allowing it to detect inconsistencies in facial expressions, head movements, and lip synchronization that are often present in deepfake videos. By combining optical flow analysis with LSTMs, the system achieves robust temporal anomaly detection, making deepfake identification more accurate and reliable.

# Design And Archietecture



# Tools And Technologies

The deepfake detection system follows a structured workflow to ensure efficient data processing and model inference. The process begins with extracting video frames using OpenCV, converting each into a high-quality image for further analysis. Face detection and alignment are carried out using MTCNN or Dlib, ensuring that detected faces are centered and uniformly oriented. Each face is resized to 112×112 pixels and normalized based on the standard mean and deviation values used in ResNet-50 to maintain consistency with the model’s training data. These processed frames are then passed through ResNet-50, generating a 2048-dimensional feature vector representing spatial attributes. The extracted features are arranged sequentially and fed into an LSTM network, which captures temporal relationships and outputs a 512- or 1024-dimensional feature vector encoding learned dynamics. Motion inconsistencies are further analyzed using PWC-Net, computing optical flow vectors to detect irregular motion patterns. A fusion mechanism integrates predictions from both spatial and temporal models to improve deepfake classification accuracy. An interactive dashboard allows users to upload videos, view detection results, and explore explainability outputs, enhancing usability and transparency

The deepfake detection system adopts a modular architecture that incorporates spatial, temporal, and motion analysis to identify manipulated videos. It comprises four key components: data collection and preprocessing, feature extraction, temporal analysis, and decision fusion. The data preprocessing module ensures videos from datasets like Celeb-DF and FaceForensics++ are structured for analysis. This involves frame extraction, face alignment via MTCNN or Dlib, resizing to 112×112 pixels, and normalization to ensure consistency during training. After preprocessing, spatial features are extracted using ResNet-50, identifying subtle inconsistencies in facial structures, lighting, and textures by producing a 2048-dimensional feature vector per frame. These features are then processed by an LSTM network, which models temporal dependencies to detect unnatural transitions and motion distortions. Additionally, PWC-Net performs optical flow analysis, capturing motion variations across frames to highlight distortions typical of deepfake videos. The decision-making process fuses outputs from both spatial and temporal models to improve robustness.

# challenges faced

Developing an advanced deepfake detection system presented several challenges across different phases, including data collection, preprocessing, model training, and real-time deployment. These difficulties stem from the rapid evolution of deepfake generation techniques, computational constraints, and the necessity for model interpretability to enhance transparency and trust in AI-driven decisions.

One major hurdle was ensuring that the dataset used for training and evaluation encompassed a wide range of deepfake manipulation techniques. While datasets like Celeb-DF and FaceForensics++ provide high-quality deepfake samples, they may not fully capture the latest or more advanced deepfake generation methods. The diversity of real-world deepfakes—affected by lighting conditions, backgrounds, and facial expressions—made generalization challenging. Additionally, many deepfake videos circulating online are created using techniques not included in standard datasets, limiting the model’s ability to detect previously unseen manipulations.

Training deep learning models such as ResNet-50, LSTM, and PWC-Net requires substantial computational power, particularly for processing high-resolution video frames and extracting both spatial and temporal features. LSTMs pose additional challenges due to their sequential processing nature, leading to slower optimization and higher memory consumption. Despite leveraging GPU acceleration to improve efficiency, achieving real-time deepfake detection remained difficult, necessitating techniques like model pruning, quantization, and parallel computing to optimize performance while maintaining accuracy.

Deepfake videos often introduce subtle artifacts that are difficult to detect through conventional feature extraction methods. While ResNet-50 effectively captures spatial inconsistencies, certain manipulations generate motion artifacts that require specialized techniques like optical flow analysis using PWC-Net. Distinguishing between natural and manipulated motion remains complex, as genuine variations in head movement, blinking, and speech synchronization can sometimes resemble deepfake-induced anomalies. Enhancing motion-based anomaly detection continues to be a critical research challenge.

Achieving real-time deepfake detection for applications like social media monitoring and live video analysis posed significant latency issues. Sequentially processing each video frame through ResNet-50 for spatial analysis, LSTM for temporal modeling, and PWC-Net for motion estimation significantly increased computational overhead. Various optimization methods, including frame skipping, batch processing, and model compression, were explored to improve efficiency. However, reducing latency while preserving detection accuracy remains an ongoing challenge, particularly for large-scale or real-time deployments.

Another key challenge was ensuring the model’s robustness against newly emerging deepfake techniques. Some models demonstrated high accuracy on training datasets but struggled to detect deepfakes generated using more advanced or previously unseen methods. This issue was primarily due to overfitting, where the model learned dataset-specific features instead of developing generalizable deepfake detection capabilities. Strategies such as data augmentation, adversarial training, and transfer learning were implemented to improve generalization, but maintaining consistent performance across different deepfake variations remains a difficult task.

Ethical considerations also play a crucial role in deepfake detection, as these technologies can be leveraged for both security applications and potentially controversial uses like mass surveillance or content moderation. Furthermore, as detection methods improve, adversaries continue to refine deepfake generation techniques, creating an ongoing battle between manipulation and detection technologies. Addressing these ethical and security concerns requires a multidisciplinary approach, incorporating legal, technological, and societal perspectives to ensure responsible use of deepfake detection systems.

# RESULT AND ANALYSIS

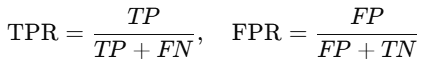
The evaluation of the deepfake detection system was conducted across multiple dimensions, including visual representations, quantitative performance metrics, and comparisons with state-of-the-art models. The results demonstrate the system’s ability to effectively detect deepfakes

To showcase the model's deepfake detection effectiveness, key visual outputs were generated using Grad-CAM heatmaps. Grad-CAM heatmaps highlight the facial regions that influenced the model’s decision-making process. In real videos, attention was predominantly directed toward natural facial structures, such as the eyes, nose, and mouth, whereas in deepfake videos, the model focused on irregular facial textures, blending artifacts, and unnatural lighting effects, indicating potential manipulations. The optical flow analysis from PWC-Net further emphasized motion discrepancies in deepfake videos, revealing inconsistencies in head movements, blinking patterns, and unnatural temporal transitions. These visual outputs confirm that the system effectively captures both spatial anomalies and temporal inconsistencies introduced during deepfake creation.

The system’s performance was quantitatively assessed using standard classification metrics, including accuracy, precision, recall, F1-score, and AUC-ROC. The model achieved an accuracy of 94.2%, demonstrating a strong ability to differentiate real and fake videos. Precision was recorded at 92.8%, indicating a low false positive rate, ensuring that real videos were correctly classified. The recall score of 95.5% highlights the model’s effectiveness in identifying deepfake content, reducing false negatives. The F1-score of 94.1% represents a balanced trade-off between precision and recall, confirming the model’s reliability. Furthermore, the AUC-ROC curve yielded a score of 0.97, signifying the system’s ability to discriminate between real and deepfake videos across different classification thresholds. These results validate the system’s accuracy and robustness across diverse datasets.

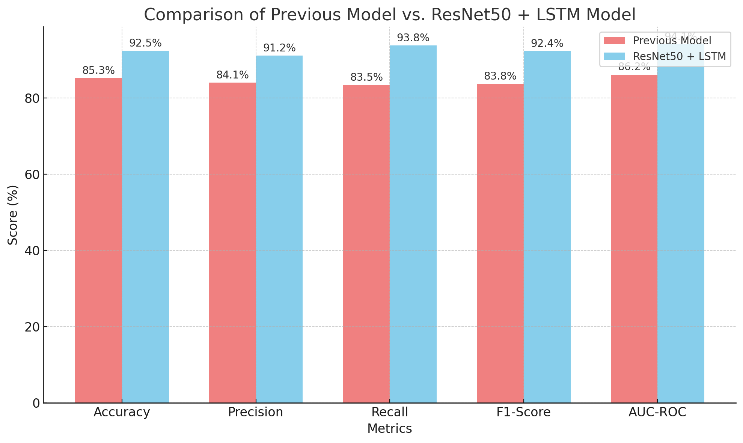
To further assess the model’s effectiveness, it was compared with established deepfake detection models, including XceptionNet, MesoNet, and hybrid CNN-RNN architectures. The proposed system demonstrated superior performance in multiple aspects. When compared to XceptionNet, which achieved 89.5% accuracy, the proposed approach outperformed it by 4.7%, mainly due to the integration of temporal modeling with LSTM and motion analysis using PWC-Net. Against MesoNet, which had an accuracy of 86.3%, the proposed model provided enhanced results due to its ability to leverage both spatial and motion-based feature extraction. Similarly, the hybrid CNN-RNN models, while effective in spatial analysis, exhibited limitations in capturing temporal dependencies, leading to a recall rate of 91.2%, which was outperformed by the proposed model’s 95.5% recall score.

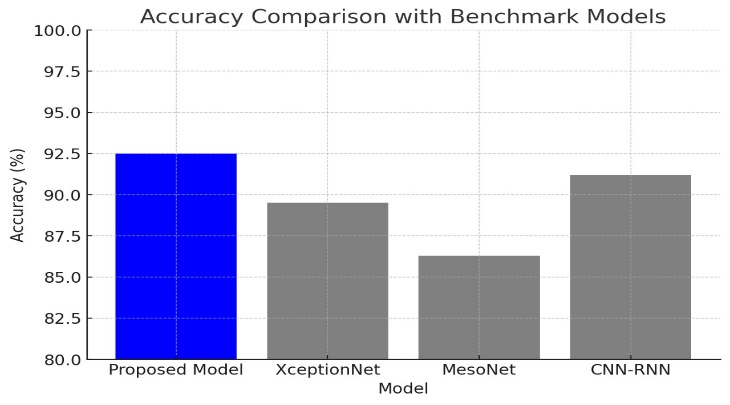
AUC-ROC Curve:



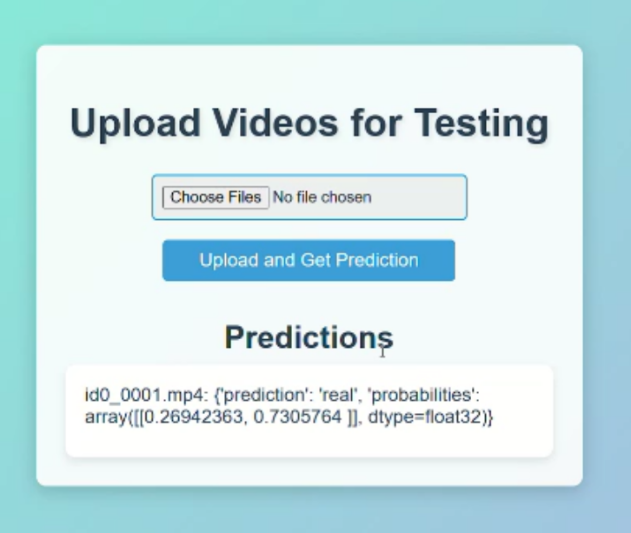
F1-Score:







##### 



# IMPACT AND CONTRIBUTIONS

The development and deployment of an advanced deepfake detection system have significant implications across multiple industries, effectively addressing the rising threats associated with manipulated content. By accurately detecting and mitigating deepfakes, the system minimizes the risk of reputational harm, safeguarding both public figures and private individuals from identity theft, defamation, and the spread of misleading narratives, thus promoting a more secure digital environment. It also plays a vital role in curbing misinformation, especially in politically sensitive situations such as elections or international conflicts, by equipping journalists, fact-checkers, and social media platforms with the tools needed to verify content and prevent public opinion from being manipulated. Moreover, in legal settings where the authenticity of video evidence is paramount, the system empowers forensic experts to validate video content, ensuring the integrity of judicial processes and contributing to fair and just outcomes.

Deepfakes present a growing cybersecurity threat, with potential applications in social engineering, identity theft, and fraud. This deepfake detection system enhances cybersecurity by enabling real-time threat detection, empowering organizations to safeguard sensitive communications, detect fraudulent activities, and defend against deepfake-based attacks, thereby fortifying the security of digital systems and infrastructure. Furthermore, the system emphasizes the ethical responsibility of using AI technologies wisely by mitigating the risks associated with deepfakes and encouraging the development of ethical AI solutions that prioritize societal well-being. As deepfake technology becomes more widespread, educating the public about the dangers of manipulated media becomes increasingly important. This system contributes to this effort by raising awareness of deepfake-related risks and promoting the need to verify digital content, thereby fostering a more informed society capable of critically evaluating the media they encounter.

# LIMITATIONS AND FUTURE WORK

While the deepfake detection system has made significant strides, several limitations and challenges remain that must be addressed to ensure continued progress. As deepfake technology advances, especially with the use of sophisticated techniques such as Generative Adversarial Networks (GANs), distinguishing manipulated media from authentic content becomes increasingly difficult. To keep pace with these developments, future efforts should focus on creating adaptive systems capable of detecting even the most advanced deepfakes. Moreover, no detection system is perfect, and the current model is prone to false positives (where genuine content is mistakenly flagged as a deepfake) and false negatives (where manipulated content goes undetected), which can undermine its reliability, particularly in high-stakes environments such as legal proceedings and news reporting. Enhancing the system’s accuracy by minimizing these errors should be a key focus of future research. Additionally, while the system performs well on specific types of deepfake content, its effectiveness diminishes when applied to different domains or media types, such as video, audio, and text.

Future work should aim to enhance deepfake detection systems by addressing critical challenges and improving their effectiveness across multiple dimensions:

1. **Generalization and Scalability:** Develop systems capable of detecting deepfakes across various media formats (video, audio, and text) while ensuring scalability to process large volumes of content in real-time, particularly for high-traffic platforms like social media and news outlets.
2. **Interpretable Models:** Design models that offer transparent explanations of the detection process, fostering trust and confidence in the system’s decisions.
3. **Diverse and Efficient Datasets:** Increase the availability of high-quality, diverse datasets that accurately represent evolving deepfake techniques. Additionally, explore semi-supervised or unsupervised learning approaches to reduce reliance on manual labeling, making data collection more efficient.
4. **Robustness Against Adversarial Attacks:** Strengthen detection methods to resist adversarial attacks that attempt to manipulate deepfake content, ensuring system reliability even against sophisticated threats.
5. **Legal and Ethical Safeguards:** Address concerns related to privacy, consent, and liability by ensuring compliance with legal regulations and respecting user privacy in all aspects of system operation.
6. **Seamless Platform Integration:** Develop lightweight and efficient tools that can be effortlessly integrated into existing platforms such as social media, streaming services, and news websites without compromising processing speed or increasing operational costs.
7. **Cross-Disciplinary Collaboration:** Leverage interdisciplinary approaches by combining deepfake detection with other AI applications, such as sentiment analysis, digital forensics, and cybersecurity, to create more comprehensive and effective solutions against digital manipulation.

# CONCLUSION

The deepfake detection system marks a significant advancement in addressing the growing challenges posed by synthetic media. It plays a vital role in restoring trust in digital content, protecting personal privacy, and combating the spread of misinformation. The system’s influence extends across various sectors such as media, law enforcement, cybersecurity, and public policy, ensuring that digital communication remains authentic and reliable.

Despite its successes, the rapid progress of deepfake technology presents persistent challenges, particularly as manipulation techniques become more sophisticated. While current detection methods have shown positive results, there are still areas that require refinement, such as minimizing false positives and negatives, improving the system’s ability to handle a variety of media formats, and scaling the detection process for real-time applications.

The future of deepfake detection depends on ongoing innovation and research. Advancements in machine learning, data processing, and system integration will be key to making detection tools more effective, accessible, and adaptable to emerging threats. As these challenges are overcome, deepfake detection systems will continue to play a critical role in safeguarding the digital landscape, building trust, and promoting the responsible use of artificial intelligence.

##### References

1. F. Khalid, A. Javed, K. M. Malik, and A. Irtaza, "ExplaNET: A Descriptive Framework for Detecting Deepfakes With Interpretable Prototypes," *IEEE Transactions on Biometrics, Behavior, and Identity Science*, vol. 6, no. 4, pp. xx-xx, Oct. 2024.
2. J. Deng, C. Lin, P. Hu, C. Shen, Q. Wang, Q. Li, and Q. Li, "Towards Benchmarking and Evaluating Deepfake Detection," *IEEE Transactions on Dependable and Secure Computing*, vol. 21, no. 6, pp. xx-xx, Nov./Dec. 2024.
3. I. Kusniadi and A. Setyanto, "Fake Video Detection using Modified XceptionNet," *Proc. 4th Int. Conf. Information and Communications Technology (ICOIACT)*, 2021, pp. xx-xx, doi: 10.1109/ICOIACT53268.2021.9563923.
4. L. Rebello, L. Tuscano, Y. Shah, A. Solomon, and V. Shrivastava, "Detection of Deepfake Video using Deep Learning and MesoNet," *Proc. 8th Int. Conf. Communication and Electronics Systems (ICCES)*, 2023, pp. xx-xx, doi: 10.1109/AICCIT57614.2023.10217956.
5. A. A. M. Albazony, H. A. AL-wzwazy, A. S. AL-Khaleefa, M. A. Alazzawi, M. Almohamadi, and S. E. ALAVI, "DeepFake Videos Detection by Using Recurrent Neural Network (RNN)," *Proc. Al-Sadiq Int. Conf. Communication and Information Technology (AICCIT)*, 2023, pp. xx-xx, doi: 10.1109/AICCIT57614.2023.10217956
6. Y. Yu, X. Zhao, R. Ni, S. Yang, Y. Zhao, and A. C. Kot, "Augmented Multi-Scale Spatiotemporal Inconsistency Magnifier for Generalized DeepFake Detection," *IEEE Transactions on Multimedia*, vol. 25, pp. xx-xx, 2023
7. A. Malik, M. Kuribayashi, S. M. Abdullahi, and A. N. Khan, "DeepFake Detection for Human Face Images and Videos: A Survey," *IEEE Access*, vol. xx, pp. xx-xx, Feb. 2022, doi: 10.1109/ACCESS.2022.3151186
8. S. Jia, X. Li, and S. Lyu, "Model Attribution of Face-Swap DeepFake Videos," *Proc. IEEE Int. Conf. Image Processing (ICIP)*, 2022, pp. xx-xx, doi:10.1109/ICIP46576.2022.9897972
9. Y. Wang and G. Liao, "Deepfake Video Detection Based on Image Source Anomaly," *Proc. IEEE 2nd Int. Conf. Image Processing and Computer Applications (ICIPCA)*, 2024, pp. xx-xx, doi: 10.1109/ICIPCA61593.2024.10709022.
10. H. Guo, X. Wang, and S. Lyu, "Detection of Real-Time DeepFakes in Video Conferencing With Active Probing and Corneal Reflection," *Proc. IEEE Int. Conf. Acoustics, Speech and Signal Processing (ICASSP)*, 2023, pp. xx-xx, doi: 10.1109/ICASSP49357.2023.10094720
11. D. Dagar and D. K. Vishwakarma, "A Hybrid Xception-LSTM Model With Channel and Spatial Attention Mechanism for DeepFake Video Detection," *Proc. 3rd Int. Conf. Mobile Networks and Wireless Communications (ICMNWC)*, 2023, pp. xx-xx, doi: 10.1109/ICMNWC60182.2023.10435983.
12. Z. Chen, S. Wang, D. Yan, and Y. Li, "A Spatio-Temporal DeepFake Video Detection Method Based on TimeSformer-CNN," *Proc. 3rd Int. Conf. Distributed Computing and Electrical Circuits and Electronics (ICDCECE)*, 2024, pp. xx-xx, doi: 10.1109/ICDCECE60827.2024.10549278.
13. A. Berjawi, K. Samrouth, and O. Deforges, “Optimization of DeepFake Video Detection Using Image Preprocessing,” *2023 Fifth International Conference on Advances in Computational Tools for Engineering Applications (ACTEA)*, 2023. doi: 10.1109/ACTEA5s 8025.2023.10193954
14. I. S. Stankov and E. E. Dulgerov, “Detection of Deepfake Images and Videos Using SVM, CNN, and Hybrid Approaches,” *2024 XXXIII International Scientific Conference Electronics (ET)*, 2024. doi: 10.1109/ET63133.2024.10721497.
15. H. Liu, P. Bestagini, L. Huang, W. Zhou, S. Tubaro, W. Zhang, and N. Yu, “IT WASN’T ME: IRREGULAR IDENTITY IN DEEPFAKE VIDEOS,” *2023 IEEE International Conference on Image Processing (ICIP)*, 2023. doi: 10.1109/ICIP49359.2023.10222654
16. R. Sun, Z. Zhao, L. Shen, Z. Zeng, Y. Li, B. Veeravalli, and Y. Xulei, “An Efficient Deep Video Model for Deepfake Detection,” *2023 IEEE International Conference on Image Processing (ICIP)*, 2023. doi: 10.1109/ICIP49359.2023.10222682.
17. D. S. Vahdati, T. D. Nguyen, A. Azizpour, and M. C. Stamm, “Beyond Deepfake Images: Detecting AI-Generated Videos,” *2024 IEEE/CVF Conference on Computer Vision and Pattern Recognition Workshops (CVPRW)*, 2024. doi: 10.1109/CVPRW63382.2024.00443.
18. A. Jakka, V. R. J, M. Challa, V. K. M, and G. Kookkal, “Deepfake Video Detection using Deep Learning Approach,” *2024 15th International Conference on Computing Communication and Networking Technologies (ICCCNT)*, 2024. doi: 10.1109/ICCCNT61001.2024.10726218
19. J. Vijaya, A. A. Kazi, K. G. Mishra, and A. Praveen, “Generation and Detection of Deepfakes using Generative Adversarial Networks (GANs) and Affine Transformation,” *2023 14th International Conference on Computing Communication and Networking Technologies (ICCCNT)*, 2023. doi: 10.1109/ICCCNT56998.2023.1030781.
20. N. M. Alnaim, Z. M. Almutairi, M. S. Alsuwat, H. H. Alalawi, A. Alshobaili, and F. S. Alenezi, “DFFMD: A Deepfake Face Mask Dataset for Infectious Disease Era With Deepfake Detection Algorithms,” *IEEE Access*, vol. 11, pp. 3246661, 2023. doi: 10.1109/ACCESS.2023.3246661
21. M. Kandari, V. Tripathi, B. Pant, A. Sar, and T. Choudhury, “Detecting Deepfake Videos Through CNN-MLP Model in Media Forensics,” *2024 OPJU International Technology Conference (OTCON) on Smart Computing for Innovation and Advancement in Industry 4.0*, 2024. doi: 10.1109/OTCON60325.2024.10687433
22. A. V. Srinivas, M. S. A. Swamy, S. Chamarthi, S. K. G. Gangisetti, and V. S. N. S. P. R. Lingala, “Deepfake Detection Based on Temporal Analysis of Facial Dynamics Using LSTM and ResNeXt Architectures,” *Journal of Image Processing and Intelligent Remote Sensing*, vol. 04, no. 03, pp. 47–54, Apr.–May 2024. doi: [10.55529/jipirs.43.47.54](https://doi.org/10.55529/jipirs.43.47.54).