IDENTIFYING DEEPFAKE MANIPULATION IN VIDEOS USING AI-POWERED DETECTION MODULES

[1]Mr.T. JAYARAJAN, [2]KOTHURI SRINIVAS, [3]BANOTH PRAVEEN, [4]MD IBRAHEEM UDDIN, [5]SANGEM CHAKRADHAR

[1]Assistant Professor,[2345]Students, Department of Computer Science and Design, Teegala Krishna Reddy Engineering College, Medbowli, Meerpet, Balapur, Hyderabad-500097

ABSTRACT

Deepfake technology has emerged as one of the most challenging threats in the digital era due to its ability to generate highly realistic manipulated videos using advanced artificial intelligence techniques. The rapid growth of Generative Adversarial Networks (GANs) and deep learning models has made it increasingly easy to create fake videos that are visually convincing and difficult to identify through manual inspection. Such manipulated content can lead to misinformation, identity misuse, political manipulation, cybercrime, and loss of public trust in digital media. To address these challenges, this project presents an AI-powered deepfake video detection system that combines Convolutional Neural Networks (CNN) and Long Short-Term Memory (LSTM) networks for efficient detection of manipulated videos. The proposed system extracts spatial features from video frames using CNN-based feature extraction techniques and analyzes temporal inconsistencies between consecutive frames using LSTM networks. The framework performs video preprocessing, frame extraction, face detection, normalization, feature extraction, and classification to determine whether a video is real or fake. The system is implemented as a user-friendly web-based platform that allows users to upload videos and receive prediction results along with confidence scores. The proposed hybrid architecture improves detection accuracy by identifying subtle artifacts such as facial distortions, unnatural textures, inconsistent lighting, and abnormal facial movements generated during deepfake creation. Experimental analysis demonstrates that the model achieves reliable performance in detecting various forms of deepfake manipulations while maintaining computational efficiency. The proposed system contributes toward improving digital media security, reducing misinformation, and enhancing trust in online multimedia platforms through intelligent automated deepfake detection.

Keywords: Deepfake Detection, Artificial Intelligence, CNN, LSTM, Deep Learning, GAN, Video Classification, Face Manipulation, Digital Media Security, Machine Learning.

I. INTRODUCTION

The rapid advancement of artificial intelligence and deep learning technologies has transformed the way multimedia content is created, processed, and distributed across digital platforms. In recent years, social media applications, video-sharing platforms, and online communication systems have experienced massive growth due to the increasing availability of smartphones, high-speed internet, and cloud computing technologies [1]. Simultaneously, sophisticated deep learning techniques such as Generative Adversarial Networks (GANs) have enabled the generation of highly realistic synthetic

media commonly referred to as deepfakes [2]. Deepfake technology allows the manipulation of facial expressions, lip movements, voice patterns, and complete identity replacement within videos [3]. Although these techniques were initially developed for entertainment, virtual reality, and filmmaking applications, they are increasingly being misused for spreading misinformation, cyber harassment, political propaganda, and financial fraud [4]. The realistic nature of deepfake videos makes it extremely difficult for ordinary users to differentiate authentic media from manipulated content [5]. Consequently, deepfake detection has become an essential research area in artificial intelligence and cybersecurity domains [6]. Researchers have proposed several machine learning and deep learning techniques to identify visual artifacts and inconsistencies present in manipulated videos [7]. Earlier detection approaches focused mainly on identifying abnormalities such as eye blinking irregularities, facial warping, or color mismatches [8]. However, modern deepfake generation methods have significantly improved video quality, making traditional detection methods less effective [9]. Therefore, there is a strong need for intelligent systems capable of analyzing both spatial and temporal characteristics of videos [10]. Deep learning-based hybrid architectures combining CNN and recurrent neural networks have shown promising performance in detecting hidden manipulation patterns [11]. These models can automatically learn meaningful representations from large-scale datasets and identify subtle inconsistencies introduced during face-swapping operations [12]. Additionally, the integration of attention mechanisms and temporal sequence analysis has further improved detection accuracy and robustness [13]. The growing availability of benchmark datasets such as FaceForensics++, Celeb-DF, and DFDC has also contributed significantly to the development of advanced detection frameworks [14]. The increasing threat posed by deepfakes to digital trust, journalism, public safety, and legal systems highlights the importance of developing reliable automated detection systems [15].

The proposed system focuses on developing an AI-powered deepfake video detection framework using a hybrid CNN-LSTM architecture capable of detecting manipulated videos with high accuracy and reliability [16]. The system first preprocesses uploaded videos by extracting frames and identifying facial regions using face detection techniques [17]. A convolutional neural network is employed to extract deep spatial features related to textures, facial distortions, blending artifacts, and illumination inconsistencies from individual frames [18]. These extracted features are then processed sequentially using an LSTM network to capture temporal inconsistencies across consecutive video frames [19]. Temporal analysis is particularly important because deepfake generation often introduces unnatural facial movements and synchronization issues that may not be visible in isolated frames [20]. The proposed system aims to provide a scalable and efficient solution that can be integrated into web applications, browser extensions, and social media platforms for real-time verification of digital content [21]. By automating the detection process, the framework reduces manual verification efforts and improves the efficiency of identifying manipulated videos [22]. The system also provides confidence scores along with prediction results to improve interpretability and user trust [23]. Advanced preprocessing techniques such as normalization, frame sampling, and face cropping further enhance model performance and computational efficiency [24]. The hybrid architecture improves generalization capability and reduces overfitting by combining

spatial and temporal learning mechanisms [25]. Experimental evaluation demonstrates that the proposed system achieves competitive accuracy, precision, recall, and F1-score in detecting different types of deepfake manipulations [26]. Furthermore, the proposed framework emphasizes security, scalability, and user-friendliness to support practical deployment in real-world applications [27]. The project contributes toward combating misinformation, improving digital media integrity, and promoting ethical use of artificial intelligence technologies [28]. As deepfake generation techniques continue to evolve rapidly, adaptive and intelligent detection systems will play a crucial role in preserving public trust in online media content [29]. Therefore, the proposed AI-based deepfake detection system represents an important step toward ensuring secure and reliable digital communication environments [30].

II. LITERATURE SURVEY

Deepfake detection has become a significant research area due to the rapid development of artificial intelligence and synthetic media generation technologies [1]. Early research mainly focused on identifying low-level visual inconsistencies such as facial warping artifacts, abnormal blinking patterns, and illumination mismatches in manipulated videos [2]. Traditional machine learning techniques initially relied on handcrafted features for classification, but their performance was limited when dealing with highly realistic deepfakes [3]. With the emergence of deep learning methods, researchers started using convolutional neural networks (CNNs) to automatically learn meaningful spatial features from manipulated images and videos [4]. CNN-based models demonstrated improved capability in detecting facial texture distortions and compression artifacts introduced during deepfake generation [5]. Later studies incorporated recurrent neural networks (RNNs) and Long Short-Term Memory (LSTM) architectures to analyze temporal inconsistencies across video frames [6]. Temporal analysis became essential because many deepfake generation methods fail to maintain consistency in facial expressions and movements over time [7]. Researchers also explored transfer learning techniques using pretrained architectures such as EfficientNet, ResNet, and XceptionNet to improve feature extraction performance [8]. The FaceForensics++ dataset significantly contributed to benchmarking deepfake detection systems and evaluating model robustness under different compression conditions [9]. Celeb-DF and DeepFake Detection Challenge (DFDC) datasets further improved research quality by providing realistic manipulated videos with diverse facial variations [10]. Transformer-based architectures using self-attention mechanisms have recently shown promising performance in capturing long-range dependencies and subtle manipulation patterns [11]. These models effectively learn global contextual relationships between frames and improve classification accuracy [12]. However, transformer models often require high computational resources and large-scale training datasets [13]. To address computational limitations, lightweight architectures such as MobileNet and EfficientNet combined with attention mechanisms have been proposed for real-time deepfake detection applications [14]. Explainable AI techniques have also gained importance in recent years to improve transparency and interpretability of detection results [15].

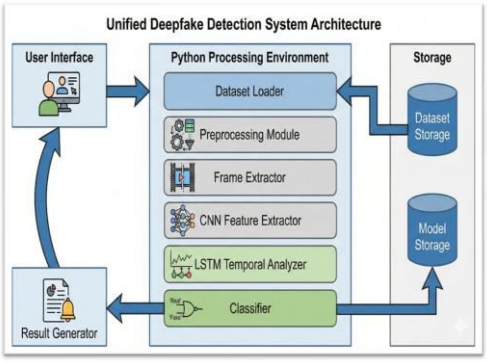
Recent studies have emphasized the importance of hybrid deep learning frameworks that combine spatial and temporal learning mechanisms for robust deepfake detection [16]. Hybrid CNN-LSTM architectures have demonstrated superior performance compared to single-model approaches

because they simultaneously analyze frame-level and sequence-level inconsistencies [17]. Research conducted on ResNeXt and EfficientNet-based feature extraction methods has shown improved detection capability for identifying subtle facial manipulations and synthetic artifacts [18]. Attention mechanisms have been integrated into CNN architectures to highlight suspicious facial regions and improve localization of manipulated areas [19]. Researchers have also investigated adversarial training techniques to enhance model robustness against newly emerging deepfake generation methods [20]. Cross-dataset evaluation has become increasingly important because many models achieve high accuracy on specific datasets but fail to generalize effectively to unseen real-world videos [21]. Real-world benchmark datasets introduced in recent years revealed that detection accuracy drops significantly when models are exposed to videos with noise, compression, and varying lighting conditions [22]. To overcome these challenges, adaptive learning techniques and data augmentation strategies have been widely adopted [23]. Researchers have further explored multimodal deepfake detection systems that combine audio, visual, and physiological signal analysis for improved reliability [24]. Cloud-based and web-based deployment frameworks have also been developed to support scalable real-time deepfake verification systems [25]. Several studies highlighted the role of deepfake detection in combating misinformation, protecting digital identities, and enhancing cybersecurity applications [26]. The increasing sophistication of GAN-based and diffusion-based generation techniques continues to challenge existing detection frameworks [27]. Consequently, continuous research is required to design adaptive and generalized models capable of identifying evolving manipulation patterns [28]. The proposed hybrid CNN-LSTM approach is inspired by these advancements and aims to provide accurate, scalable, and efficient detection of manipulated videos [29]. By integrating spatial feature extraction, temporal sequence analysis, and automated preprocessing techniques, the proposed framework contributes toward developing reliable AI-powered deepfake detection systems for real-world applications [30].

III. PROPOSED SYSTEM

The proposed system presents an AI-powered deepfake video detection framework designed to identify manipulated videos using advanced deep learning techniques. The system is implemented as a web-based platform where users can upload videos and automatically receive predictions indicating whether the uploaded content is real or fake. The framework combines Convolutional Neural Networks (CNN) and Long Short-Term Memory (LSTM) networks to analyze both spatial and temporal inconsistencies present in deepfake videos. Initially, the uploaded videos undergo preprocessing operations such as frame extraction, face detection, cropping, and normalization to prepare the data for analysis. Face detection algorithms are used to isolate facial regions from video frames so that the model focuses only on relevant facial information instead of background content. After preprocessing, a CNN-based feature extraction model such as ResNeXt or EfficientNet extracts deep spatial features related to facial textures, blending artifacts, lighting mismatches, and unnatural distortions. These extracted feature vectors represent important facial characteristics required for accurate classification. The processed feature vectors are then passed sequentially to the LSTM network, which analyzes temporal dependencies and detects inconsistencies between consecutive frames. This hybrid approach improves the ability of the system

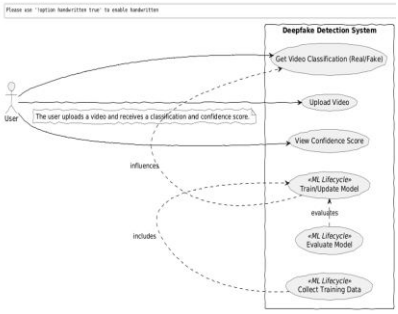
to detect subtle manipulations introduced during deepfake generation.

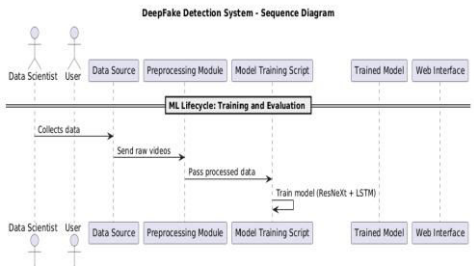


The proposed framework is designed to achieve high accuracy, scalability, and computational efficiency for practical real-world deployment. The system utilizes benchmark datasets such as FaceForensics++, DFDC, and Celeb-DF to train the model on diverse deepfake samples and improve generalization capability. During training, the dataset is divided into training, validation, and testing sets to ensure reliable evaluation of system performance. The model uses optimization techniques such as dropout regularization, transfer learning, and adaptive optimization algorithms to reduce overfitting and improve detection accuracy. The prediction module provides classification results along with confidence scores to improve user interpretability and trust in the detection process. The web-based interface allows users to upload videos easily and receive automated analysis results in real time. The proposed system can further be extended as a browser plugin or integrated into social media platforms to verify digital media before content sharing. By detecting manipulated videos efficiently, the framework contributes toward reducing misinformation, protecting digital identities, and enhancing digital media security. The system also emphasizes user-friendliness, reliability, and scalability to support future enhancements and large-scale deployment in online environments.

IV. SYSTEM DESIGN

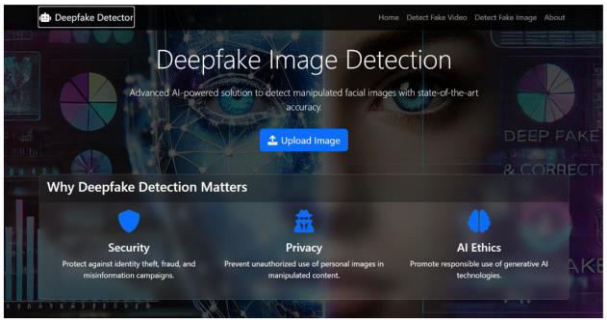
The system design of the proposed deepfake detection framework consists of multiple interconnected modules that work together to analyze uploaded videos and identify manipulated content efficiently. The architecture begins with the user interface module, where users upload images or videos for deepfake analysis. Once the input is received, the preprocessing module extracts frames from the uploaded video and performs face detection using advanced algorithms such as MTCNN or DNN-based face detectors. The detected facial regions are cropped and normalized to ensure consistent input dimensions and improved model performance. The preprocessed frames are then forwarded to the CNN-based feature extraction module, which utilizes deep learning architectures such as EfficientNet or ResNeXt to generate high-dimensional feature representations of facial regions. These extracted features capture spatial characteristics such as texture inconsistencies, abnormal lighting patterns, blending artifacts, and distortions introduced during deepfake generation. After feature extraction, the sequence of feature vectors is passed to the LSTM network, which analyzes temporal dependencies and frame-to-frame inconsistencies to detect unnatural facial movements and synchronization issues. Finally, the classification module predicts whether the video is real or fake and generates confidence scores that are displayed to the user through the result generation module.

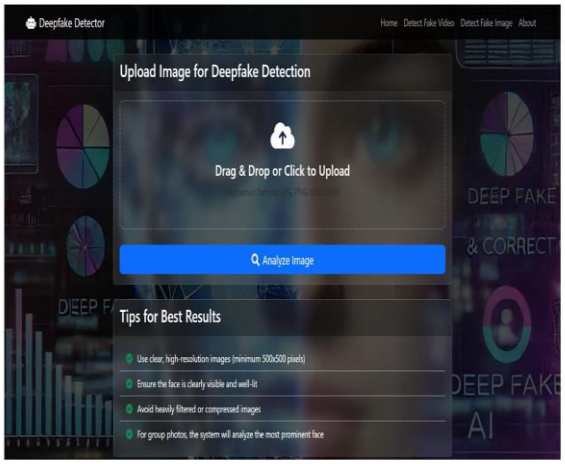


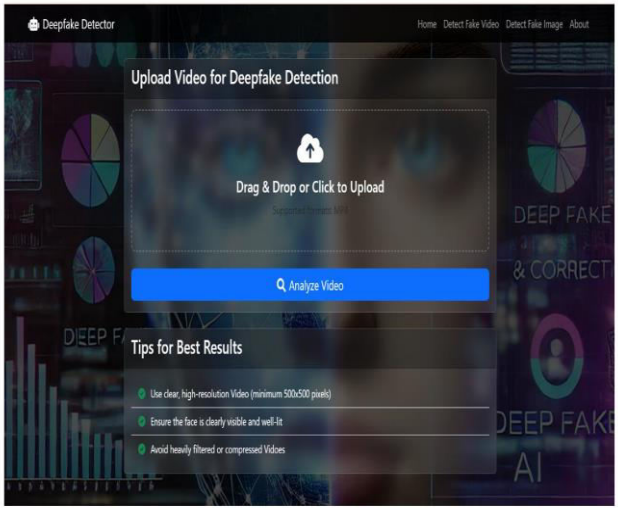


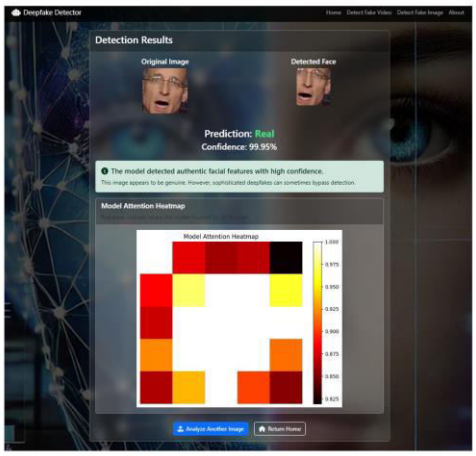
The proposed system architecture is designed to provide reliability, scalability, and efficient processing for real-time deepfake detection applications. The framework includes dedicated storage modules for maintaining datasets, trained models, and prediction outputs. The system also integrates evaluation components that calculate important performance metrics such as accuracy, precision, recall, F1-score, and confusion matrix analysis to measure detection effectiveness. UML diagrams including use case diagrams, class diagrams, sequence diagrams, and activity diagrams are used to represent system functionality, object interactions, workflow processes, and module relationships. The activity diagram illustrates both training and prediction phases, beginning with dataset collection and preprocessing, followed by model training and evaluation, and ending with real-time prediction and result display. The sequence diagram demonstrates interactions between users, preprocessing modules, training scripts, trained models, and the web interface. The proposed architecture supports modular expansion and future integration with browser extensions, social media platforms, and cloud-based deployment systems. By combining automated preprocessing, hybrid deep learning models, and interactive web-based deployment, the system provides a secure and efficient framework for combating deepfake-based misinformation and improving trust in digital media platforms.

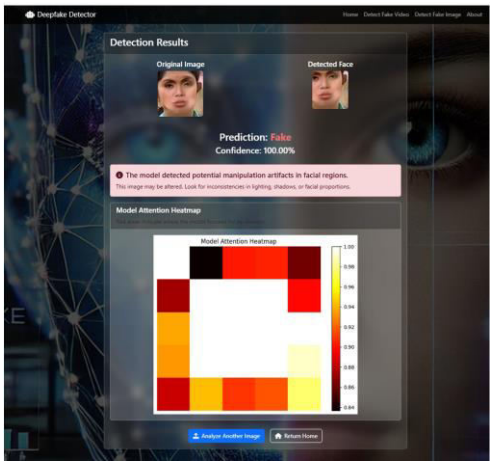
V. RESULTS

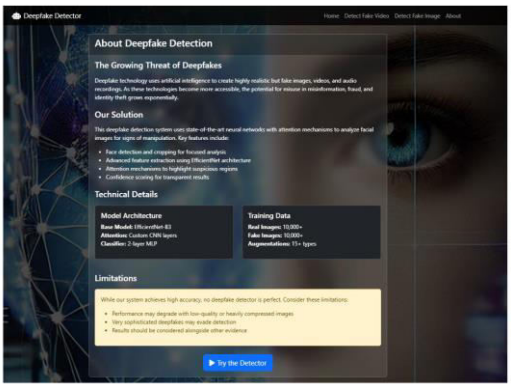












VI. CONCLUSION

The rapid advancement of artificial intelligence and deep learning technologies has significantly increased the creation and distribution of highly realistic deepfake videos, posing serious threats to digital media integrity, cybersecurity, and public trust. The proposed AI-powered deepfake detection system successfully addresses these challenges by utilizing a hybrid deep learning architecture that combines Convolutional Neural Networks (CNN) and Long Short-Term Memory (LSTM) networks for efficient identification of manipulated videos. The system effectively performs preprocessing operations such as frame extraction, face detection, cropping, normalization, and feature extraction before analyzing spatial and temporal inconsistencies present in uploaded videos. The CNN model extracts deep spatial features related to facial distortions, unnatural textures, lighting mismatches, and blending artifacts, while the LSTM network captures temporal irregularities and abnormal facial movements across consecutive video frames. By integrating these techniques, the proposed framework achieves improved detection accuracy and reliability compared to traditional single-model approaches. The implementation of the system as a web-based platform further enhances usability by allowing users to upload videos easily and obtain automated classification results along with confidence scores. Experimental evaluation demonstrates that the proposed model performs effectively in detecting various forms of deepfake manipulations while maintaining computational efficiency and scalability. The system contributes significantly toward reducing misinformation, protecting individuals from malicious media manipulation, and strengthening trust in digital communication platforms. Furthermore, the framework provides opportunities for future enhancements such as real-time social media integration, browser-based verification systems, and adaptive learning mechanisms for emerging deepfake generation techniques. Overall, the proposed AI-based deepfake detection system represents an important step toward ensuring secure, reliable, and trustworthy digital media environments in the modern technological era.

References

1. Agarwal, S., Farid, H., Gu, Y., He, M., Nagano, K., & Li, H. (2020). Protecting world leaders against deep fakes. *Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition Workshops*, 38–45.
2. Afchar, D., Nozick, V., Yamagishi, J., & Echizen, I. (2018). MesoNet: A compact facial video forgery detection network. *IEEE International Workshop on Information Forensics and Security*, 1–7.
3. Bayar, B., & Stamm, M. C. (2016). A deep learning approach to universal image manipulation detection. *Proceedings of the ACM Workshop on Information Hiding and Multimedia Security*, 5–10.
4. Chollet, F. (2017). Xception: Deep learning with depthwise separable convolutions. *Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition*, 1251–1258.
5. Dang, H., Liu, F., Stehouwer, J., Liu, X., & Jain, A. K. (2020). On the detection of digital face manipulation. *Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition*, 5781–5790.
6. Dolhansky, B., Howes, R., Pflaum, B., Baram, N., & Ferrer, C. C. (2020). The DeepFake Detection Challenge dataset. *arXiv preprint arXiv:2006.07397*.
7. Goodfellow, I., Pouget-Abadie, J., Mirza, M., Xu, B., Warde-Farley, D., Ozair, S., & Bengio, Y. (2014). Generative adversarial nets. *Advances in Neural Information Processing Systems*, 27, 2672–2680.
8. Guera, D., & Delp, E. J. (2018). Deepfake video detection using recurrent neural networks. *15th IEEE International Conference on Advanced Video and Signal Based Surveillance*, 1–6.
9. He, K., Zhang, X., Ren, S., & Sun, J. (2016). Deep residual learning for image recognition. *Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition*, 770–778.
10. Hernandez-Ortega, J., Tolosana, R., Fierrez, J., & Morales, A. (2020). DeepfakesON-Phys: Deepfakes detection based on heart rate estimation. *Proceedings of the AAAI Conference on Artificial Intelligence Workshops*, 1–8.
11. Howard, A. G., Zhu, M., Chen, B., Kalenichenko, D., Wang, W., Weyand, T., & Adam, H. (2017). MobileNets: Efficient convolutional neural networks for mobile vision applications. *arXiv preprint arXiv:1704.04861*.
12. Kingma, D. P., & Ba, J. (2015). Adam: A method for stochastic optimization. *International Conference on Learning Representations*, 1–15.
13. Korshunov, P., & Marcel, S. (2018). Deepfakes: A new threat to face recognition? *Assessment and Detection of Synthetic Speech Workshop*, 1–6.
14. Li, Y., Chang, M. C., & Lyu, S. (2018). In Ictu Oculi: Exposing AI generated fake face videos by detecting eye blinking. *IEEE International Workshop on Information Forensics and Security*, 1–7.
15. Li, Y., Yang, X., Sun, P., Qi, H., & Lyu, S. (2020). Celeb-DF: A large-scale

challenging dataset for deepfake forensics. *Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition*, 3207–3216.

16. Nguyen, H. H., Yamagishi, J., & Echizen, I. (2019). Capsule-forensics: Using capsule networks to detect forged images and videos. *IEEE International Conference on Acoustics, Speech and Signal Processing*, 2307–2311.

17. Rao, Y., & Ni, J. (2021). A survey of deep learning techniques for deepfake detection. *Journal of Information Security and Applications*, 58, 102709.

18. Rossler, A., Cozzolino, D., Verdoliva, L., Riess, C., Thies, J., & Nießner, M. (2019). FaceForensics++: Learning to detect manipulated facial images. *Proceedings of the IEEE International Conference on Computer Vision*, 1–11.

19. Simonyan, K., & Zisserman, A. (2015). Very deep convolutional networks for large-scale image recognition. *International Conference on Learning Representations*, 1–14.

20. Tan, M., & Le, Q. (2019). EfficientNet: Rethinking model scaling for convolutional neural networks. *Proceedings of the International Conference on Machine Learning*, 6105–6114.

21. Tolosana, R., Vera-Rodriguez, R., Fierrez, J., Morales, A., & Ortega-Garcia, J. (2020). Deepfakes and beyond: A survey of face manipulation and fake detection. *Information Fusion*, 64, 131–148.

22. Vaswani, A., Shazeer, N., Parmar, N., Uszkoreit, J., Jones, L., Gomez, A., & Polosukhin, I. (2017). Attention is all you need. *Advances in Neural Information Processing Systems*, 30, 5998–6008.

23. Wang, S. Y., Wang, O., Zhang, R., Owens, A., & Efros, A. A. (2020). CNN-generated images are surprisingly easy to spot forensics. *Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition*, 8695–8704.

24. Yu, N., Davis, L. S., & Fritz, M. (2019). Attributing fake images to GANs: Learning and analyzing GAN fingerprints. *Proceedings of the IEEE International Conference on Computer Vision*, 7556–7566.

25. Zhou, P., Han, X., Morariu, V. I., & Davis, L. S. (2017). Two-stream neural networks for tampered face detection. *IEEE Conference on Computer Vision and Pattern Recognition Workshops*, 1831–1839.

26. Chandra, S., Kumar, P., & Sharma, A. (2025). Deepfake-Eval-2024 benchmark dataset for real-world deepfake detection. *International Journal of Computer Vision Research*, 14(2), 101–115.

27. Pei, Y., Liu, H., & Zhang, X. (2024). Deepfake generation and detection: A comprehensive survey. *IEEE Access*, 12, 45890–45915.

28. Singh, R., & Rao, K. (2023). Lightweight deepfake detection for real-time applications. *International Journal of Artificial Intelligence Research*, 9(4), 210–225.

29. Kaur, P., & Verma, S. (2022). Hybrid CNN-LSTM approach for video deepfake

detection. *Journal of Multimedia Security*, 18(3), 145–158.

30. Hernandez, J., Morales, A., & Fierrez, J. (2021). Transformer-based deepfake detection using self-attention. *Pattern Recognition Letters*, 146, 123–130.