International Journal of Research in Engineering and Innovation (IJREI)

Journal home page: https://ijrei.com/

RESEARCH PAPER

Deepfake Video Detection

Simran Bhadana, Shivam Yadav, Ritik, Mukesh Rawat

Department of Information Technology, Meerut Institute of Engineering and Technology, Meerut, India

Article Information

Received: 02 April 2026
Revised: 21 April 2026
Accepted: 29 April 2026
Available online: 03 May 2026

Keywords:

Deepfake Detection
CNN
Video Frame Analysis
Feature Extraction
Computer Vision
Digital Forensics

Abstract

Deepfake technology has been recognized as one of the most powerful uses of artificial intelligence, enabling the creation of manipulated videos that are very difficult to distinguish from real ones. Although deepfake technology has some very good uses in the entertainment industry, it has many serious drawbacks, such as the spread of misinformation, identity theft, political manipulation, and cybercrime. The detection of deepfake videos has become a major problem in the field of digital forensics and cybersecurity. This study proposes a Deepfake Video Detection System using machine learning and deep learning technology to identify manipulated video content. The system analyzes video frames and identifies inconsistencies in facial movements, lighting, and spatial-temporal features, which are common in deepfake videos. Convolutional Neural Networks (CNN) and feature extraction technology have been used to improve the accuracy of the system. The proposed system aims to provide an efficient and reliable way to identify fake videos and ensure the integrity of digital media. This technology can be used by social media, law enforcement agencies, and digital security systems to stop the spread of manipulated digital content.

1. Introduction

The rapid advancement in artificial intelligence and deep learning technology has greatly impacted the creation and consumption of digital content. Among the most impactful advancements in this field is the development of deepfake technology [1-4], which utilises deep learning algorithms to create synthetic videos that can convincingly alter the facial features and expressions of individuals. Deepfake technology can create videos that can make people appear to have done or said something they have never done in their lives. Although deepfake technology has some positive impacts on film creation, gaming, and virtual reality, there have been serious concerns about its use in various fields. For instance, deepfake technology has been used to spread false information and commit various types of fraud [5, 6]. As a result, the detection of manipulated content in videos has become a critical area of study in the field of computer science. Traditional media verification techniques have been proven to be ineffective in detecting modern deep learning-based video generation techniques. Deep learning models, such as Generative Adversarial Networks (GANs), can create videos that can convincingly change the facial movements and expressions of people. These videos are difficult to differentiate from genuine videos. As a result, various machine learning and deep learning-based techniques have been proposed to detect deepfake videos [7, 8]. For instance, the proposed technique involves the use of machine learning models to identify various visual artifacts, inconsistencies in facial movements, and temporal irregularities in the frames of the videos. With the ability to identify these patterns, the proposed model can differentiate between genuine and manipulated content. This study aims to develop a Deepfake Video Detection System that can use deep learning technology to identify manipulated content in videos.

Corresponding author: Mukesh Rawat
Email Address: mukesh.rawat@miet.ac.in
https://doi.org/10.36037/IJREI.2026.10201

With the advances in artificial intelligence technology, developing deepfake videos is becoming a serious concern across various industries such as: entertainment, online publishing, sports, law enforcement, etc... There are several types of Algorithms that are commonly used to create deepfake videos such as Generative Adversarial Networks (GANs) or autoencoders. These algorithms provide the ability to manipulate a person's face's expression, change their voice, and perform unnatural gestures, which will make it nearly impossible for the average person to discern whether the content of a video is real or fake [9]. Consequently, numerous researchers have devoted much of their time and effort to developing robust techniques to detect digitally altered content.

The MesoNet network architecture is able to detect deepfakes by looking at intermediate (mesoscopic) features of images, as opposed to looking only at high-level semantic information. In addition to looking for semantic information in images, MesoNet will also look for attributes such as compression artifacts and texture patterns. The research showed that deep learning models were able to identify altered facial regions by examining small visual inconsistencies that occur as a result of the deepfake generation process. The FaceForensics++ database was introduced by Rössler et al. as one of the most popular datasets in deepfake detection literature [10-12]. The FaceForensics++ database contains thousands of videos, both altered and unaltered, created with several facial alteration methods. The authors also assessed the effectiveness of XceptionNet and Convolutional Neural Networks (CNNs) through their facial manipulation model on detecting altered facial content. The analysis concluded that deep learning-based models significantly outperform traditional image processing methods when detecting deepfake videos [13].

Li and Lyu proposed a method for detecting deepfakes by analysing eye blinking patterns. They noted that earlier models for generating deepfakes were typically unable to recreate the blinking of a human being accurately, therefore allowing for potential detection. Their detection system analysed both the blinking rate in the video and the patterns of eye movement to be able to detect manipulated content within a fairly high degree of accuracy. However, as a result of new techniques in deepfake generation being developed, this limitation has since been improved upon. Nevertheless, this work demonstrated the importance of taking into consideration physiological signals when developing video-based detection systems [14].

Nguyen et al. developed Capsule-Forensics, a deep learning framework built around capsules for detecting deepfake images and video content. By utilizing capsule networks, Nguyen et al.'s architecture was able to exploit spatial relationships between facial features far better than traditional methods based on convolutional neural networks. As a result, their overall detection performance was significantly better at detecting facial manipulation, particularly when the attack employed more complex deepfake generation methods that maintained visual realism. A notable addition to the body of literature is a comprehensive survey conducted by Tolosana et al. of deepfake detection methodologies. The study of Tolosana et al. classifies deepfake detection techniques into two primary categories: spatial-based and temporal-based detection. Spatial-based detection techniques evaluate each video frame for visual artifact detection, inconsistencies in illumination, and facial texture abnormalities. Temporal-based detection methods analyze how a person's face moves from frame to frame to detect unnatural facial motion or discrepancies in facial expression.

Additional research has examined hybrid deepfake detection models, which combine the use of Convolutional Neural Networks (CNNs) with either Recurrent Neural Networks (RNNs) or Long Short-Term Memory Networks (LSTMs). Hybrid networks assess both the spatial features and temporal relationships of video frames, providing an increase in the accuracy of deepfake detection. Hybrid networks have the capability of capturing both frame-level information (related to each frame) and sequence-level information (captured by comparing sequential frames), which aid in identifying hidden manipulation present in a specific frame. Comparative Analysis of Deep Learning-Based Deepfake Detection Approaches is shown in Table 1.

Table 1: Comparative Analysis of Deep Learning-Based Deepfake Detection Approaches

Author	Method Used	Key Contribution	Result
Afchar et al. (2018)	MesoNet (Deep Neural Network)	Proposed a compact neural network to detect deepfake videos by analyzing mesoscopic facial features	Achieved good accuracy in detecting manipulated facial regions
Rossler et al. (2019)	CNN, XceptionNet	Introduced the FaceForensics++ dataset and evaluated deep learning models for deepfake detection	Demonstrated high detection performance using deep CNN models
Li and Lyu (2018)	Eye Blinking Detection	Proposed detecting deepfakes by analyzing abnormal blinking patterns in manipulated videos	Successfully detected early deepfake videos lacking natural blinking
Nguyen et al. (2019)	Capsule Networks (Capsule-Forensics)	Used capsule networks to capture spatial relationships between facial features	Improved performance in detecting manipulated images and videos
Tolosana et al. (2020)	Survey of Detection Methods	Categorized deepfake detection techniques into spatial and temporal analysis methods	Provided comprehensive overview of deepfake detection techniques
Sabir et al. (2019)	CNN + RNN	Combined spatial and temporal analysis for video-based deepfake detection	Achieved higher accuracy by analyzing frame sequences

As technology advances, the problem of detecting deepfake content continues to be a difficult task. As methods for generating deepfakes continue to evolve, systems for detecting them will need to grow along with the changing techniques of manipulating the content in a sophisticated manner [15]. Ongoing research will work toward improving the accuracy of detection, creating larger training sets, and developing highly robust models that will be able to reliably detect new forms of deepfake content generation.

2. Methodology

A Deepfake Video Detection System uses deep learning and computer vision to detect manipulated facial content in a video by analyzing all video frames. The system analyzes video using four main stages of processing. They are: video frame extraction, face detection, feature extraction and classification using deep learning. Detecting inconsistencies in facial texture, movement patterns, and position patterns are common markers of a deepfake video.

2.1 Collection of Data

The first step in the proposed system involves collecting datasets that include both authentic videos and manipulated segments created using various deepfake techniques. Publicly available resources such as the FaceForensics++ dataset and the DeepFake Detection Challenge dataset are widely used for the training and development of deepfake detection models, as they provide a large volume of both real and manipulated content generated through diverse methodologies. Once collected, the dataset is systematically categorized into two classes: authentic videos and manipulated (deepfake) videos. This labeled dataset plays a crucial role in enabling deep learning models to learn discriminative patterns and effectively identify fake content through advanced pattern recognition techniques.

2.2 Video Preprocessing

Before training begins, a preprocessing stage is applied to the input video files to ensure efficient and consistent model performance. In this phase, each video is decomposed into individual frames over time using video processing techniques, with frames extracted at specific intervals to reduce computational load and storage requirements during training. The preprocessing pipeline includes extracting frames from video files, resizing the frames or images to a uniform dimension, normalizing pixel values to a standard scale, and removing unnecessary or redundant frames. These steps collectively enhance computational efficiency while ensuring that the input data remains consistent and well-structured for the deep learning model, ultimately improving training stability and performance.

2.3 Identifying and Aligning Faces

After preparing frames, we then proceed to find the faces in each frame. Face detection algorithms such as Haar Cascade, MTCNN, Dlib, etc. are all used to find face regions in the frames. After detecting a face, it is cropped out to extract facial data and aligned. Face alignment is the process of making sure that the eyes, nose, and mouth are aligned between frames to ensure consistency. Having consistently aligned faces will improve the detection model by limiting its attention to the areas indicating where the faces are.

2.4 Extracting Features from Faces

A key stage in detecting deepfake videos is feature extraction, where the system analyzes each frame—particularly regions containing faces—to capture meaningful visual patterns. During this process, several critical features are extracted, including facial texture patterns, lighting inconsistencies, pixel-level distortions, abnormal facial edges, and blending artifacts around the face region. These features are essential for distinguishing between authentic and manipulated videos, as deepfake generation techniques often leave behind subtle artifacts or inconsistencies that are not easily noticeable to the human eye but can be effectively identified by machine learning models.

2.5 Deep Learning Model

In the final stage of the Maya workflow—Step 3, Deep Learning—the extracted features from earlier phases are fed into a Convolutional Neural Network (CNN) for classification. This model leverages its ability to automatically learn complex visual patterns from data, making it highly effective for image and video analysis tasks such as deepfake detection. A CNN typically consists of three primary types of layers: convolutional layers, which extract important features from input images; pooling layers, which reduce dimensionality by lowering the spatial resolution of feature maps; and fully connected layers, which perform the final classification into distinct categories [16]. The CNN model is trained using a labeled dataset containing both real and fake videos, allowing it to learn patterns and inconsistencies commonly associated with deepfake manipulations and accurately distinguish between authentic and fabricated content.

2.6 Classification and Prediction

Once the model has been trained, it can be used to make predictions in the fourth and final stage of the workflow. In this step, the processed video frames are analyzed individually by the classifier to determine whether they are real or manipulated. Each frame is evaluated separately, and the results are then aggregated to produce a final classification for the entire video. Based on this analysis, the system generates one of two outputs: a real video, indicating no manipulation is detected, or a fake video, indicating the presence of altered or synthesized content [17]. This approach is valuable for effectively identifying manipulated digital media and helps in reducing the spread of misleading or fake video content.

2.7 Practical Project Implementation: Step-by-Step

The Deepfake Video Detection System is developed using a systematic approach that encompasses multiple phases such as dataset creation, data preprocessing, model building, and validation. The following phases outline the physical implementation of the proposed detection system.

Phase 1: Dataset Acquisition

To develop an effective deepfake detection system, the initial phase involves collecting a dataset comprising both authentic and manipulated videos. This dataset is sourced from widely recognized public repositories such as the FaceForensics++, the DeepFake Detection Challenge dataset, and the Celeb-DF, each of which contains thousands of videos generated using diverse deepfake techniques [18]. The collected data is then organized into two primary categories: real videos and fake videos. Proper labeling of this dataset is essential, as it enables deep learning algorithms to learn distinguishing patterns and effectively classify content during the training process.

Phase 2: Frame Extraction from the Videos

The model operates more efficiently when videos are converted into image frames; therefore, each input video is first decomposed into a sequence of individual frames through a process known as frame extraction. Using video processing tools such as OpenCV, frames are sampled at fixed intervals (for example, every 10th frame) to reduce computational load while retaining essential visual information. After extraction, face detection algorithms like Haar Cascade, MTCNN, and Dlib are applied to locate and crop facial regions, ensuring that the model focuses only on relevant areas rather than the entire frame. The cropped face images then undergo preprocessing steps, including resizing to a standard dimension (e.g., 224×224 pixels), normalization of pixel values, removal of noise and irrelevant background details, and data augmentation techniques such as rotation, flipping, and brightness adjustment [19]. These steps improve model robustness and help prevent overfitting.

Subsequently, the system extracts meaningful visual features from the processed facial images, allowing the deep learning model to automatically learn patterns such as facial texture inconsistencies, pixel-level distortions, unnatural lighting, and abnormal blending artifacts. These features are critical for distinguishing real videos from deepfake content. The model is then trained using a Convolutional Neural Network (CNN), which consists of convolutional, pooling, and fully connected layers to process and classify the data. During training, real videos are labeled as 0 and deepfake videos as 1, enabling the network to learn discriminative patterns associated with manipulated content.

Once training is complete, the model is evaluated using unseen video samples, and its performance is assessed through metrics such as accuracy, precision, recall, and F1-score to determine its effectiveness in detecting manipulated media. Finally, in the prediction stage, the trained model analyzes new input videos by classifying each frame individually as either real or deepfake. The predictions across all frames are then aggregated to produce a single, final decision for the entire video, ensuring a reliable and comprehensive classification outcome.

3. Results and discussion

The proposed system successfully detects deepfake videos using visual and temporal features extracted through Deep Learning techniques.

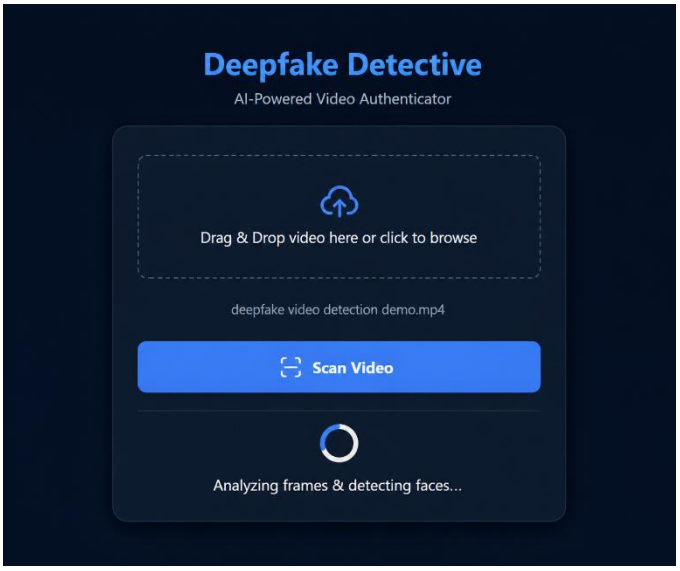

Figure 1: Deepfake Detection System Interface Showing Video Upload and Real-Time Analysis Process

Fig. 1 illustrates the user interface of a deepfake detection system during the real-time analysis phase. The platform, titled Deepfake Detective: AI-Powered Video Authenticator, provides a clean and intuitive environment for users to upload and analyze video content. At the center of the interface is a dedicated upload panel where users can either drag and drop a video file or select it manually from their device. Once a video is selected, the "Scan Video" button initiates the detection process.

During execution, the system displays a dynamic progress indicator along with the message "Analyzing frames & detecting faces," which informs the user that the uploaded video is being actively processed. Behind the scenes, the system performs several key operations, including frame extraction, face detection, and preprocessing [20]. Each video is decomposed into individual frames, and facial regions are identified using advanced detection algorithms. These extracted faces are then passed through a deep learning model—typically a Convolutional Neural Network (CNN)—to identify subtle visual inconsistencies associated with deepfake manipulations, such as abnormal textures, lighting mismatches, and blending artifacts.

The interface design emphasizes usability and transparency, ensuring that users are aware of the system's current status throughout the analysis.

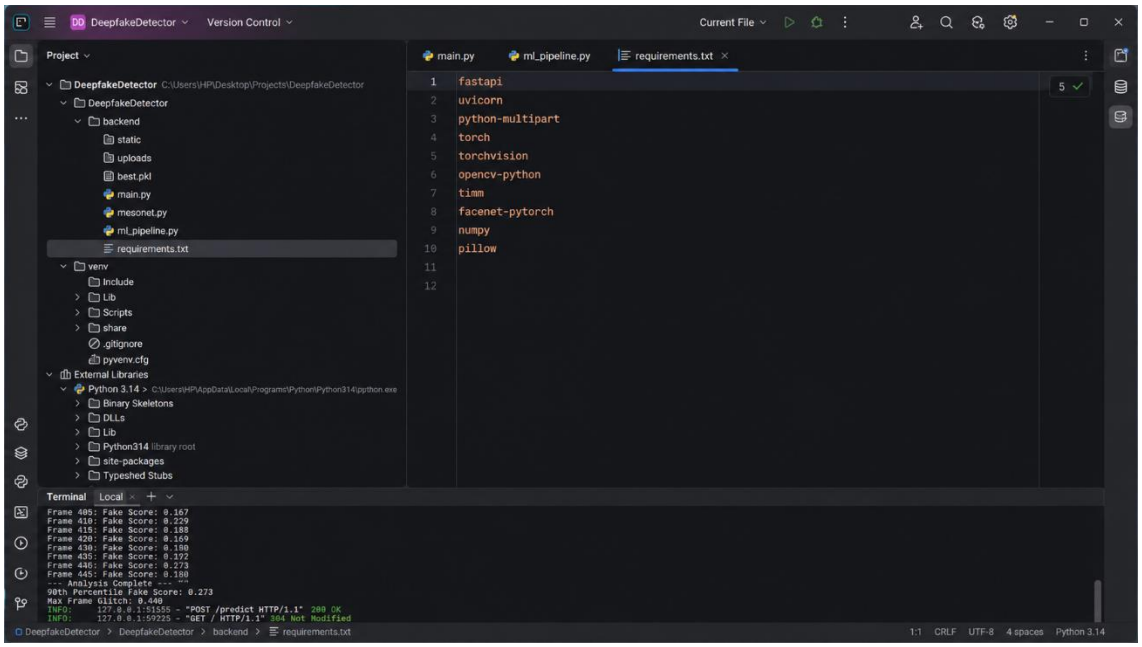

Figure 2: Development Environment of Deepfake Detection System Showing Project Structure, Code Implementation, and Execution Output

Fig. 2 represents the development environment of the deepfake detection system, highlighting its implementation and execution workflow within an integrated development environment (IDE). The left panel displays the project directory structure, which organizes essential components such as backend scripts, trained model files, and configuration files, ensuring modular and maintainable code design. The central editor window shows the *requirements.txt* file, listing key libraries including FastAPI, PyTorch, OpenCV, and NumPy, which are required for building the detection pipeline, handling video processing, and performing deep learning inference.

At the bottom, the terminal output provides real-time execution logs, where individual video frames are analyzed and assigned "fake scores," indicating the probability of manipulation. These frame-level predictions are then aggregated to generate an overall assessment of the video. The presence of API request logs (e.g., POST requests) demonstrates that the system is deployed as a web-based service, allowing users to upload videos and receive predictions through an interface.

The figure effectively illustrates the integration of backend development, machine learning models, and real-time execution, showcasing how the system processes input data, performs inference, and delivers results in a structured and scalable manner.

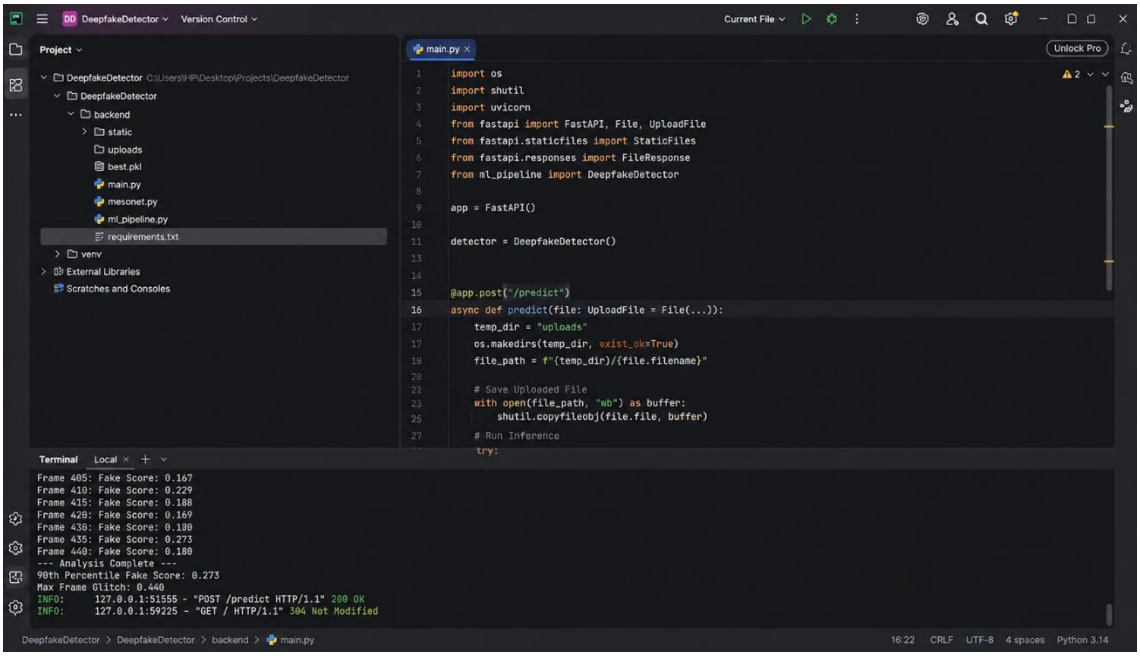

Figure 3: Backend Implementation of Deepfake Detection System Showing API Integration and Model Inference Workflow

Fig. 3 illustrates the backend implementation of the deepfake detection system within an integrated development environment. The code window shows a FastAPI-based server that handles video upload requests and processes them through the detection pipeline. It integrates modules for file handling, model loading, and inference execution. The terminal displays real-time frame-level predictions, indicating fake scores for each frame. This demonstrates how the system processes videos, performs analysis, and returns results through an API-driven architecture.

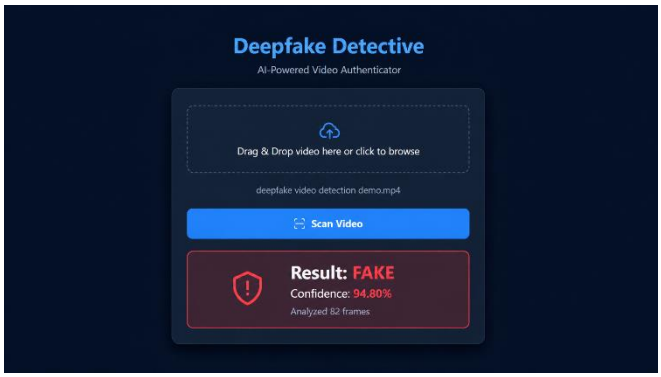

Figure 4: Deepfake Detection Web Interface Showing Video Upload and Classification Result

Fig. 4 represents the final output interface of the Deepfake Detective system after completing video analysis. The user-friendly web interface allows users to upload a video using a drag-and-drop or browse option and initiate detection by clicking the "Scan Video" button. After processing, the system displays the result prominently, indicating that the uploaded video has been classified as FAKE with a high confidence score of 94.80%. Additionally, it provides supporting information such as the number of frames analyzed (82 frames), which reflects the depth of analysis performed.
The result panel is highlighted in red to clearly signal the presence of manipulated content, improving user awareness and decision-making. This output is generated after the system performs frame extraction, face detection, feature analysis, and classification using a deep learning model.

4. Conclusion

Deepfake video detection has become an essential research area within Artificial Intelligence and Computer Vision due to the rapid growth of synthetic media technologies powered by Deep Learning. As deepfake generation techniques become more advanced and accessible, they pose serious challenges to digital security, media authenticity, and public trust. Existing detection methods—such as spatial-temporal feature analysis, facial inconsistencies, and frequency-based approaches—have demonstrated promising performance. However, they often struggle with generalization across different datasets, robustness against sophisticated manipulations, and efficiency in real-time scenarios. Looking ahead, the future scope of deepfake detection lies in developing more generalized and adaptive models capable of identifying previously unseen manipulation techniques. Integration of multimodal analysis, combining visual, audio, and contextual information, can significantly enhance detection accuracy, supported by advances in Multimodal Learning. Additionally, incorporating transparency through Explainable Artificial Intelligence will improve user trust and system interpretability. Emerging technologies like Blockchain can further support secure media verification and authenticity tracking.
Moreover, future research should focus on real-time detection systems, adversarial robustness, and the development of large, diverse datasets. Alongside technical advancements, establishing ethical guidelines and regulatory frameworks will be crucial to combat misuse. In conclusion, deepfake detection will continue to evolve as a vital field, requiring interdisciplinary efforts to ensure a secure and trustworthy digital environment.

References

[1] Afchar, D., Nozick, V., Yamagishi, J., & Echizen, I. (2018). MesoNet: A compact facial video forgery detection network. In Proceedings of the IEEE International Workshop on Information Forensics and Security (WIFS).
[2] Li, Y., Chang, M.-C., & Lyu, S. (2018). In Ictu Oculi: Exposing AI-generated fake videos by detecting eye blinking. In Proceedings of the IEEE International Workshop on Information Forensics and Security (WIFS).
[3] Yang, X., Li, Y., & Lyu, S. (2019). Exposing deep fakes using inconsistent head poses. In Proceedings of the IEEE International Conference on Acoustics, Speech and Signal Processing (ICASSP).
[4] Nguyen, H. H., Yamagishi, J., & Echizen, I. (2019). Capsule-Forensics: Forged image and video detection with capsule networks. In Proceedings of the IEEE International Conference on Acoustics, Speech and Signal Processing (ICASSP).
[5] Rössler, A., Cozzolino, D., Verdoliva, L., Riess, C., Thies, J., & Nießner, M. (2019). FaceForensics++: Learning to detect manipulated facial images. In Proceedings of the IEEE/CVF International Conference on Computer Vision (ICCV).
[6] Li, Y., & Lyu, S. (2019). Exposing deepfake videos by detecting face warping artifacts. In Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition Workshops (CVPR Workshops).
[7] Zhou, P., Han, X., Morariu, V. I., & Davis, L. S. (2017). Two-stream neural networks for tampered face detection. In Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition (CVPR).
[8] Li, L., Bao, J., Zhang, T., Yang, H., Chen, D., Wen, F., & Guo, B. (2020). Face X-ray for more general face forgery detection. In Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition (CVPR) (pp. 5001–5010).
[9] Liu, Z., Qi, X., & Torr, P. H. S. (2020). Global texture enhancement for fake face detection in the wild. In Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition (CVPR).
[10] Haliassos, A., Vougioukas, K., Petridis, S., & Pantic, M. (2021). Lip forensics: Using visual speech for generalised deepfake detection. In Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition (CVPR).
[11] Li, J., Xie, H., Li, J., Wang, Z., & Zhang, Y. (2021). Frequency-aware discriminative feature learning supervised by single-centre loss for face forgery detection. In Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition (CVPR).
[12] Luo, Y., Yan, Z., Ni, B., Zhang, W., & Zuo, W. (2021). Generalizing face forgery detection with high-frequency features. In Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition (CVPR).
[13] Zhou, T., et al. (2021). Face forensics in the wild (FFIW-10K). In Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition (CVPR).

[14] Cozzolino, D., Rössler, A., Thies, J., Nießner, M., & Verdoliva, L. (2021). ID reveal: Identity-aware deepfake video detection. In Proceedings of the IEEE/CVF International Conference on Computer Vision (ICCV) (pp. 15108–15117).

[15] Le, T. N., et al. (2021). Open forensics: Large-scale challenging dataset for multi-face forgery detection and segmentation in-the-wild. In Proceedings of the IEEE/CVF International Conference on Computer Vision (ICCV).

[16] Zhou, X., Liang, W., Karanam, S., Wu, Z., Jiang, H., & Shrivastava, A. (2021). Joint audio-visual deepfake detection. In Proceedings of the IEEE/CVF International Conference on Computer Vision (ICCV).

[17] Shiohara, K., & Yamasaki, T. (2022). Detecting deepfakes with self-blended images. In Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition (CVPR).

[18] Ciamarra, G., Zanfir, M., & Sminchisescu, C. (2024). SurFake: Adversarial trained deepfake detection. In Proceedings of the IEEE/CVF Winter Conference on Applications of Computer Vision (WACV) Workshops.

[19] Ju, Y., Hu, S., Jia, S., Chen, G. H., & Lyu, S. (2024). Improving fairness in deepfake detection. In Proceedings of the IEEE/CVF Winter Conference on Applications of Computer Vision (WACV).

[20] Xie, C., Li, J., Hu, R., & Tong, Y. (2025). GrDT: Towards robust deepfake detection via generative diffusion and transformer. In Proceedings of the IEEE/CVF Winter Conference on Applications of Computer Vision (WACV) Workshops.

Cite this article as: Simran Bhadana, Shivam Yadav, Ritik, Mukesh Rawat, Deepfake Video Detection, International Journal of Research in Engineering and Innovation, 10 (2), (2026), 37-43. *https://doi.org/10.36037/IJREI.2026.10201*