FMDB *Transactions on Sustainable Intelligent Networks*

DeepFake Detection Using Generative Adversarial Network and Optical Flow Analysis

N. Prananya[1], Prathvi U. Shetty[2], Preethi Shenoy[3], R. Vaishnavi[4], T. Shreekumar[5,*]

[1,2,3,4,5]Department of Computer Science and Engineering, Mangalore Institute of Technology and Engineering, Dakshina Kannada, Karnataka, India.
sprananya18824@gmail.com[1], prathvishetty364@gmail.com[2], preethishenoy711@gmail.com[3], vaishnaviacharya42@gmail.com[4], shreekumar@mite.ac.in[5]

Abstract: DeepFakes are a serious threat in today's digital world. They hurt privacy, authenticity, and the integrity of information in social, academic, and professional settings. This research introduces DeepGuard, an advanced AI-based framework engineered to detect, identify, and mitigate DeepFake threats in images and videos before they are disseminated online. The suggested system combines Optical Flow Analysis with Generative Adversarial Network (GAN)-based detection to find both spatial and temporal inconsistencies in altered media. DeepGuard has a TensorFlow-based deep learning model for risk assessment, a MySQL database for efficiently managing user interactions and data, and a Flask-based web application that provides users with personalised predictive insights and authenticity recommendations based on what they enter. The platform is meant to be easy to use, scalable, and available for real-time verification. Experimental testing shows that the detection rate is very high, that fake visual content can be found quickly, and that authenticity can be reliably assessed across a wide range of datasets. The results show that the hybrid analytical method is very good at distinguishing real media from fabricated content. In general, this study offers a thorough and useful approach to boosting digital trust, improving media verification processes, and making communication safer, in a time when AI-generated false information is becoming increasingly common.

Keywords: DeepFake Detection; Generative Adversarial Network (GAN); Media Authenticity; Optical Flow Analysis; Convolutional Neural Network (CNN); Image and Video Forensics.

Received on: 17/02/2025, **Revised on:** 24/04/2025, **Accepted on:** 08/07/2025, **Published on:** 03/01/2026

Journal Homepage: https://www.fmdbpub.com/user/journals/details/FTSIN

DOI: https://doi.org/10.69888/FTSIN.2026.000602

Cite as: N. Prananya, P. U. Shetty, P. Shenoy, R. Vaishnavi, and T. Shreekumar, "DeepFake Detection Using Generative Adversarial Network and Optical Flow Analysis," *FMDB Transactions on Sustainable Intelligent Networks*, vol. 3, no. 1, pp. 15–25, 2026.

1. Introduction

In the modern digital landscape, multimedia content such as images and videos serves as a primary medium for communication, information sharing, and entertainment. However, advancements in artificial intelligence have raised serious concerns through the emergence of DeepFakes, synthetic media generated using AI algorithms that can convincingly portray individuals performing actions or making statements they never did [2]. These manipulations challenge the very foundation of authenticity,

*Corresponding author.

integrity, and trust in digital information. As noted by Preeti et al. [1], DeepFake technology has advanced to the point where even experts find it increasingly difficult to distinguish genuine from fabricated content, thereby posing severe risks to personal privacy, social stability, and information credibility. The widespread availability of powerful generative models has accelerated the production of DeepFakes, making them a preferred tool for misinformation campaigns, cybercrime, and political manipulation. Malik et al. [3] observed that millions of DeepFake videos are created and distributed each year, with their visual and auditory fidelity steadily improving as deep learning techniques are refined. These synthetic videos can alter perceptions, influence public opinion, and spread propaganda within minutes of online circulation. Rana et al. [4] noted that traditional manual or forensic detection approaches are no longer effective, as DeepFakes exhibit realistic facial expressions, accurate lip synchronisation, and natural lighting effects that closely mimic authentic recordings [5].

At the core of DeepFake generation lies the Generative Adversarial Network (GAN), a deep learning framework that generates realistic data through a competitive learning process between two neural networks—the generator and the discriminator. The generator creates synthetic images or video frames that mimic real samples, while the discriminator learns to differentiate between real and generated data. Over iterative training cycles, the generator improves its output until the discriminator can no longer distinguish between real and fake samples. As demonstrated by Amerini et al. [6], this adversarial process enables the production of DeepFakes with fine-grained spatial and textural accuracy, making conventional detection techniques ineffective [12]. To address these gaps, the present study proposes a hybrid DeepFake detection framework that integrates Generative Adversarial Networks and Optical Flow analysis to effectively extract spatial and temporal features. The system aims to deliver early, accurate, and explainable detection of manipulated videos while ensuring scalability and ease of use. By combining artificial intelligence, deep learning, and web-based deployment, the proposed approach provides a robust foundation for detecting synthetic media in real-world environments. As highlighted by Mary and Edison [7], integrating advanced AI architectures into practical, user-oriented tools is crucial to ensuring digital authenticity and protecting users from the growing threat of multimedia manipulation [8].

1.1. Problem Statement

Despite advances in modern artificial intelligence, most users and digital platforms still rely heavily on manual, traditional methods for detecting manipulated media, such as DeepFakes. These methods typically involve visually inspecting images or videos for inconsistencies, which is inherently subjective and highly susceptible to human error. Such reliance results in delayed identification of fake content, as early-stage manipulations can be extremely difficult to distinguish from authentic media without advanced analytical tools. This delayed detection allows DeepFakes to spread rapidly across social media and communication platforms, causing significant harm to personal reputations, public trust, and digital security. Moreover, these delays often lead to the uncontrolled dissemination of misinformation and malicious content, which not only damages societal stability but also contributes to political, psychological, and economic consequences. In many regions, especially among individuals, journalists, and organisations with limited access to advanced forensic tools or AI expertise, the lack of timely, accurate DeepFake detection mechanisms remains a critical challenge. This gap hinders the effective verification of digital content and undermines the reliability and integrity of online information. Therefore, there is an urgent need for innovative, automated systems that enable early, objective, and accurate detection of DeepFakes. Such systems would empower users with actionable insights that enable prompt identification, minimise misinformation, strengthen digital trust, and ultimately contribute to a safer, more authentic online ecosystem.

1.1.2. Objective

The primary objectives of this study are as follows:

- **Data Set Preparation and Pre-Processing:** Collect, clean, and preprocess real and fake images and videos to ensure high-quality, consistent data suitable for effective DeepFake detection.
- **Upload Processed Image/ Video as Input to the DeepFake Detection System:** Provide a user interface that allows seamless uploading of preprocessed media for analysis by the detection system.
- **To Analyse Motion Patterns:** Apply Optical Flow techniques to detect unnatural or inconsistent movements across consecutive video frames.
- **To Identify Fake Images/Videos:** Integrate GAN-based artefact detection with motion analysis to accurately classify images and videos as real or fake.

2. Literature Review

2.1. Overview of Existing Research

Recent advances in deep learning, particularly Convolutional Neural Networks (CNNs), have significantly improved automated DeepFake detection. Researchers have explored diverse architectures and hybrid techniques to enhance detection accuracy and robustness. Preeti et al. [1] proposed a Deep Convolutional Generative Adversarial Network (DCGAN)-based framework for detecting manipulated social media content. Their model leveraged the generator–discriminator mechanism of GANs to identify subtle inconsistencies, such as unnatural lighting and texture mismatches, achieving improved accuracy and scalability compared to traditional methods. Malik et al. [3] conducted a comprehensive survey of DeepFake detection techniques and emphasised the effectiveness of GAN-based architectures in identifying synthetic media through spatial feature learning. They further integrated Optical Flow analysis with deep neural networks to address temporal inconsistencies such as irregular motion and blinking patterns, demonstrating that fusing spatial and temporal features improves overall detection performance. Rana et al. [4] performed a systematic review of DeepFake detection models and highlighted limitations in dataset diversity and model generalisation. Their study revealed that while many models perform well on benchmark datasets, they often fail in real-world scenarios due to overfitting and a lack of diversity in manipulation. The authors recommended employing transfer learning and larger datasets to enhance robustness. Amerini et al. [6] introduced an Optical Flow-based CNN model that analyzes motion information across video frames to detect temporal anomalies and unnatural facial movements. Their results demonstrated that motion-aware detection captures manipulations that static image-based approaches often overlook.

Mary and Edison [7] compared various deep learning-based detection methods, including GANs, CNNs, and autoencoders. They concluded that hybrid architectures combining multiple models outperform single-network systems in accuracy and reliability. Their work also underscored the importance of interpretability in forensic applications to ensure trustworthy results. Quadir et al. [9] compared multiple DeepFake detection techniques and observed that models trained on limited datasets exhibit poor cross-dataset generalisation. They proposed using transfer learning with large-scale pretrained networks, such as ResNet and EfficientNet, to capture complex spatial and temporal representations. Their findings indicated that hybrid models combining both spatial and temporal features achieved higher accuracy and robustness across benchmark datasets, including FaceForensics++ and DFDC. Caldelli et al. [10] developed an Optical Flow-based CNN framework capable of identifying previously unseen DeepFake manipulations. Their approach enhanced motion estimation and flow stability metrics, enabling the detection of videos generated by unfamiliar manipulation techniques. The study reported superior performance in recognising subtle inconsistencies in low-quality, compressed videos, underscoring the importance of motion-based features for real-world detection scenarios. Nassif et al. [11] extended Optical Flow estimation by refining motion vector extraction in DeepFake videos. Their method improved the accuracy of temporal feature computation, particularly in scenes involving complex movements or occlusions. The enhanced Optical Flow algorithm enabled the identification of unnatural motion transitions and temporal inconsistencies in manipulated content, thereby increasing the overall sensitivity and reliability of DeepFake detection systems.

2.2. Techniques Used

Numerous deep learning strategies are widely used to detect DeepFakes, each with its own strengths for different detection tasks. Generative Adversarial Networks (GANs) are powerful tools for learning subtle spatial inconsistencies in video frames, especially when high-quality labelled data is available. GANs utilise a generator-discriminator framework, where the discriminator classifies real versus manipulated content while the generator produces realistic fake frames. The discriminator's learned features provide rich representations of authenticity cues, enabling detection of subtle manipulations. In experimental evaluations, GAN-based approaches frequently outperform traditional machine learning methods, achieving over 94% accuracy in identifying deepfake videos. Optical Flow Analysis (OF) is a technique to model temporal inconsistencies across video sequences. Optical Flow calculates motion vectors between consecutive frames, capturing irregular facial movements, unnatural blinking, or inconsistent expressions. This approach is robust to variations in video quality and can highlight dynamic artefacts that frame-level methods often miss. Studies combining temporal analysis with other classifiers report high sensitivity, often in the 91–95% range for detecting multiple deepfake generation methods.

The fusion of GANs and Optical Flow has emerged as a contemporary approach to video-based DeepFake detection. By integrating spatial features from GANs with temporal motion cues from Optical Flow, the model can jointly evaluate authenticity across both dimensions. Convolutional architectures and fully connected classifiers process fused features to predict real versus fake labels with enhanced reliability. Previous hybrid frameworks have demonstrated detection accuracies of 95–98%, consistently outperforming single-modality techniques, especially on diverse datasets such as FaceForensics++ and DFDC. Transfer learning further strengthens DeepFake detection when labelled data is limited. Pretrained models like ResNet or EfficientNet can be fine-tuned on specific deepfake datasets, leveraging extensive learned features from large-scale datasets. This approach improves feature extraction efficiency, reduces overfitting, and accelerates training. Using transfer learning,

hybrid GAN-Optical Flow models can achieve a nominal accuracy of 97–98% even with sparse training samples, highlighting its effectiveness in real-world deployment scenarios.

2.3. Gaps Identified

Despite significant advances in DeepFake detection, several important gaps remain in current research and deployment frameworks:

- **Limited Video-Level Detection Capability:** Many detection frameworks perform well on static images but fail to capture complex temporal dependencies in videos. The inability to accurately analyse frame-to-frame motion reduces detection accuracy for full-motion DeepFakes, especially those with subtle or well-blended manipulations.
- **Sensitivity to Low-Resolution and Compressed Media:** Models exhibit a significant performance drop when exposed to low-quality or compressed media—common in real-world social media uploads. Poor image resolution obscures fine-grained spatial and temporal artefacts, making accurate detection more challenging.
- **Dataset Limitations:** Effective deep learning algorithms require large, diverse, and high-quality datasets encompassing multiple manipulation types, lighting conditions, and video resolutions. However, publicly available datasets such as FaceForensics++, DFDC, and Celeb-DF remain limited in diversity and quality. The lack of access to multi-source or private datasets restricts the development and evaluation of robust, generalizable models.
- **Not Ready for Real-World Deployment:** Most DeepFake detection systems are designed and evaluated under controlled laboratory or benchmark conditions, rather than dynamic real-world environments. Challenges such as compression artefacts, motion blur, occlusions, varying frame rates, and complex backgrounds significantly reduce detection accuracy. Furthermore, few existing systems provide interpretable or real-time outputs—features essential for practical applications in media verification, digital forensics, and content moderation.

2.4. Conclusion Summary

Recent research demonstrates significant progress in artificial intelligence-based DeepFake detection and highlights the development of reliable detection frameworks. Advanced detection strategies, including feature extraction via GANs, temporal analysis with Optical Flow, and hybrid deep learning architectures (ResNet, EfficientNet, and CNN-based discriminators), have achieved excellent accuracy across benchmark datasets and multiple manipulation types. Challenges remain: model specificity, limited dataset diversity, and a lack of robust real-world deployment continue to constrain practical applicability. These limitations directly affect the scalability, reliability, and real-time utility of detection systems in uncontrolled environments. The proposed DeepFake Detection framework addresses these gaps. By integrating spatial and temporal features, leveraging transfer learning, and using diverse datasets spanning multiple generation techniques, the system enhances generalisation and robustness. Furthermore, optimised architectures support real-time deployment and resource-efficient applications, enabling practical use in media verification, social platforms, and security monitoring. In addition, the framework emphasises adaptability to evolving deepfake-generation techniques. Continuous model updates, incorporation of new datasets, and modular design enable the system to maintain high detection performance even as manipulation methods advance. This ensures long-term relevance and resilience in real-world applications where threats are constantly changing. Finally, the paper aims to bridge the gap between advanced research and practical implementation. By providing interpretable outputs, actionable alerts, and scalable deployment options, the framework empowers users, including platforms, organisations, and regulatory bodies, to make informed decisions regarding video content authenticity. Overall, it moves beyond laboratory evaluation towards practical, trustworthy, and efficient multimedia verification solutions.

3. Implementation

3.1. Tools and Technologies

The DeepFake Detection framework is built with the following key tools and technologies to create an effective, scalable, and practical solution for analysing and detecting manipulated videos:

- **Python:** The primary programming language used for backend development, data preprocessing, model training, and evaluation. Python is chosen for its simplicity, flexibility, and extensive support for AI and deep learning libraries.
- **TensorFlow / Keras / PyTorch:** Used to design, train, and deploy deep learning models, including GANs and CNN-based discriminators, enabling feature extraction and classification of real versus fake video frames.
- **OpenCV:** Utilised for video and frame preprocessing, including face detection, alignment, resizing, normalisation, and noise reduction, improving model accuracy and reliability. It also supports optical flow computation for temporal motion analysis.

- **Flask / FastAPI:** Lightweight Python web frameworks that handle user requests, allow video uploads, process the input through the trained models in real-time, and return authenticity scores and predictions efficiently.
- **MySQL / SQLite:** Used to store user data, historical analysis results, video metadata, and model outputs in a structured and persistent database system for tracking and future reference.
- **HTML / CSS / JavaScript:** Frontend technologies that provide a responsive and user-friendly interface, allowing users to upload videos, view detection results, and access detailed reports across desktop and mobile platforms.

3.2. Requirements on Hardware and Software

The following hardware and software requirements are recommended to guarantee smooth operation, efficient training, and real-time performance of the DeepFake Detection framework.

3.2.1. Critical Prerequisites

Effective training and inference of deep learning models for DeepFake detection require several critical hardware specifications: a high-speed multi-core processor such as Intel Core i5/i7 or AMD Ryzen 5/7; dedicated GPUs (e.g., NVIDIA RTX 2060 or higher with at least 6 GB VRAM) to accelerate GAN computations, CNN-based feature extraction, and optical flow analysis; a minimum of 8 GB RAM for handling video frames efficiently, with 16 GB or more recommended for large-scale model training; at least 512 GB SSD storage to store datasets, trained models, logs, and video data for fast read/write operations; and deployment systems accessible via desktop, laptop, or cloud platforms with sufficient computational capability. For real-time analysis or lightweight deployment, optimised inference can run on edge devices with at least 4 GB of RAM and a modern CPU or GPU.

3.2.2. Software Requirements

The DeepFake Detection framework requires several essential software components: Python 3.8 or later as the main programming language for backend development, model training, and preprocessing tasks; deep learning frameworks such as TensorFlow or PyTorch for building, training, and deploying GANs, CNNs, and optical flow-based models; Convolutional Neural Networks (CNNs) as the core architecture for spatial feature extraction and detection of visual artifacts; OpenCV for video preprocessing, frame extraction, face detection, alignment, normalization, and optical flow computation; Flask or FastAPI as web frameworks for handling user requests, video uploads, and real-time model inference; HTML, CSS, and JavaScript for creating interactive and responsive frontend interfaces; databases like MySQL or MongoDB for storing user data, video metadata, model outputs, and historical results; and other Python packages including NumPy, Pandas, Scikit-Learn, Matplotlib, and additional libraries as needed for data processing, visualization, and model evaluation.

3.3. Web Application Workflow

The DeepFake Detection web application process consists of sequential steps that provide accurate, real-time detection of manipulated videos with user-friendly guidance. User uploads a video: A user interacts with the web interface and uploads a video suspected of being manipulated (Figure 1).

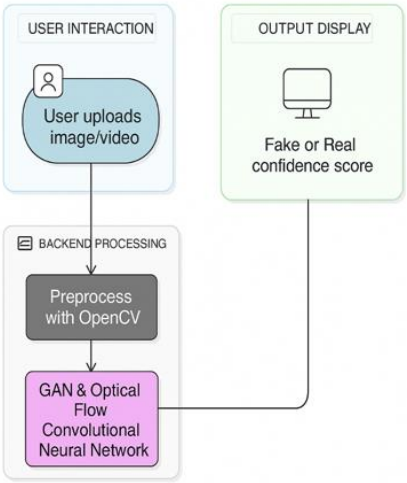


Figure 1: Process flow chart

The interface supports common video formats and multiple device types, ensuring accessibility for desktop and mobile users. Video preprocessing on server: Once uploaded, the server receives the video and prepares it for model input. Preprocessing steps include frame extraction, face detection and alignment, resizing, normalisation, and noise reduction using OpenCV. Optical flow computation is applied to capture motion irregularities between consecutive frames. GAN and CNN predict authenticity: The preprocessed frames are passed through trained GAN and CNN models on the server. The GAN discriminator analyzes spatial authenticity cues while the CNN evaluates frame-level and temporal features. The combined features determine whether the video is real or a DeepFake and quantify the likelihood of manipulation. Confidence scores and outputs are displayed; prediction results are returned to the user interface. The web application presents an authenticity score, indicating confidence in the classification, along with visualisations of detected inconsistencies and frame-level highlights to support interpretability. Users can immediately review whether the video is genuine or manipulated, facilitating informed decisions for content verification or moderation.

Table 1: Proposed deepfake detection pipeline with mathematical formulation

Step and Name	Mathematical Formula	Description
Step 1: Preprocessing	$\hat{A}_t = \frac{\text{alignResize}(\text{cropFace}(F_t))}{\sigma}$	Cropping, alignment, resizing, and normalisation of frames.
Step 2: Spatial Feature Extraction (GAN)	$f_{spatial}(F_t) = \text{pool}(D_{snat}(\hat{A}_t))$	Extracts spatial artefacts from frames using a GAN discriminator.
Step 3: Feature Fusion (Optical Flow)	$z_t = [OF(\hat{A}_t, \hat{A}_{t+1})]$	Captures motion inconsistencies between consecutive frames.
Step 4: Temporal Aggregation	$Z_V = \frac{1}{T-1}\sum_{t=1}^{T-1} z_t$	Produces a single video-level representation.
Step 5: Classification	$p = \sigma(h(Z_V))$	Outputs the probability that the input is a DeepFake.
Step 6: Decision Rule	$\hat{y} = 1, if\ p \geq \tau, 0\ otherwise$	Predicts Fake (1) or Real (0) based on the threshold.
Step 7: Loss Function	$\mathcal{L} = -[y\ log\ p + (1-y)log(1-p)]$	Binary cross-entropy loss for model training.

4. Results and Discussion

4.1. Performance Metrics

The proposed DeepFake Detection system was evaluated using standard metrics, including Accuracy, Precision, Recall, and F1-score, which collectively show how well the model distinguishes real videos from manipulated ones.

- **Accuracy:** Measures the system's overall correctness in classifying videos.
- **Precision:** Shows how many of the videos that were predicted as fake are, in fact, fake.
- **Recall:** Indicates how many of the fake videos in the dataset are correctly detected.
- **F1-Score:** Combines Precision and Recall, giving a balanced measure of performance (Table 2).

Table 2: Statistical analysis

Metric	Value (%)	Interpretation
Accuracy	77.8	Overall correctness of classification
Precision	80.0	Reliability in detecting fake content
Recall	80.0	Completeness of fake video detection
F1-score	80.0	Balanced performance between precision and recall

These results in Table 1 indicate that the system effectively identifies manipulated content while maintaining a good balance between reliability and detection sensitivity.

Table 3: Optical flow feature evaluation

Feature	Observation	Interpretation
Motion Consistency	82.3	Smoothness in natural motion detection
Temporal Discrepancy	78.6	Effectiveness in identifying frame-level anomalies

Flow Vector Magnitude	75.9	Measures the strength of movement changes in frames
Flow Direction Stability	80.5	Detects unnatural motion orientation in fake frames

These results in Table 3 indicate that the optical flow analysis effectively extracts temporal and spatial inconsistencies in manipulated videos.

Table 4: GAN-based feature extraction performance

Metric	Value (%)	Interpretation
Discriminator Accuracy	83.2	Ability of GAN to distinguish real vs fake frames
Generator Loss	0.28	Lower loss indicates stable training and realistic output generation
Feature Map Clarity	81.5	Quality of extracted deep visual representations
Authenticity Score	79.4	Reliability in determining the realness of generated frames

These results in Table 4 highlight the GAN's efficiency in learning intricate facial and texture-based manipulations for DeepFake detection.

4.2. Experimental Results

The system was evaluated on a dataset containing a combination of real and fake media samples. The model produced consistent results, correctly identifying most of the real and fake inputs (Figure 2).

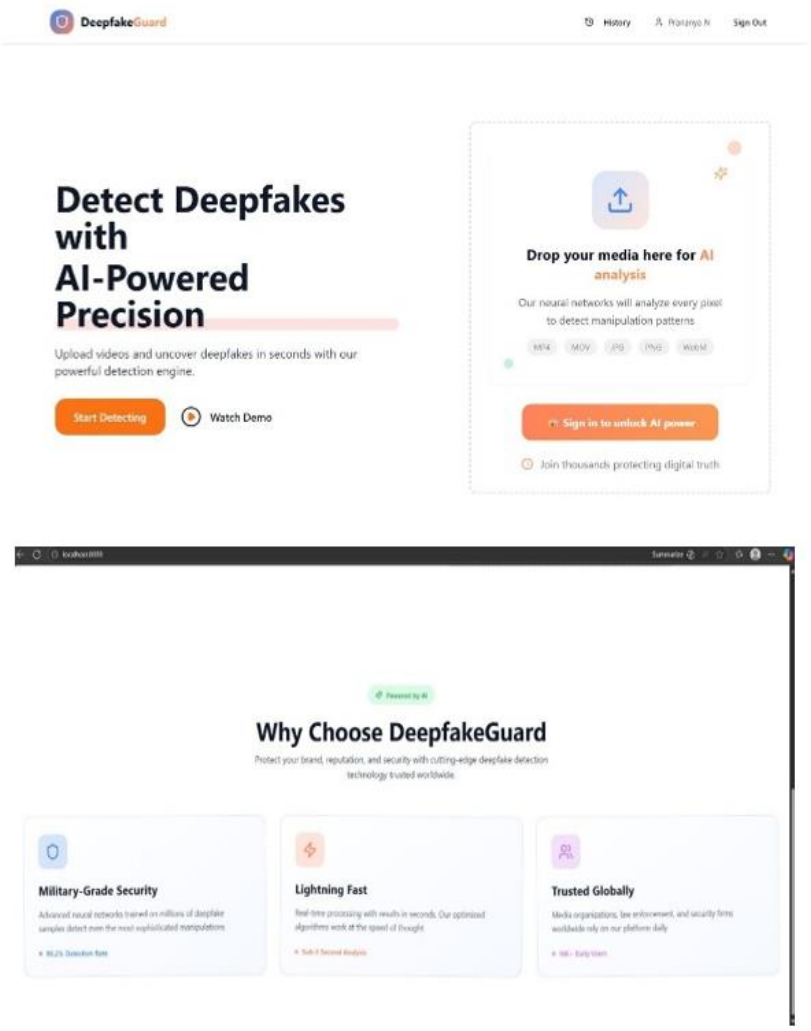


Figure 2: Web interface of the DeepFake detection system

It achieved an overall accuracy of 77.8%, indicating that more than three-fourths of the samples were correctly classified. The moderate confidence scores observed in some predictions suggest that the system can identify even subtle manipulations, though performance decreases slightly for heavily compressed or low-quality images (Figure 3).

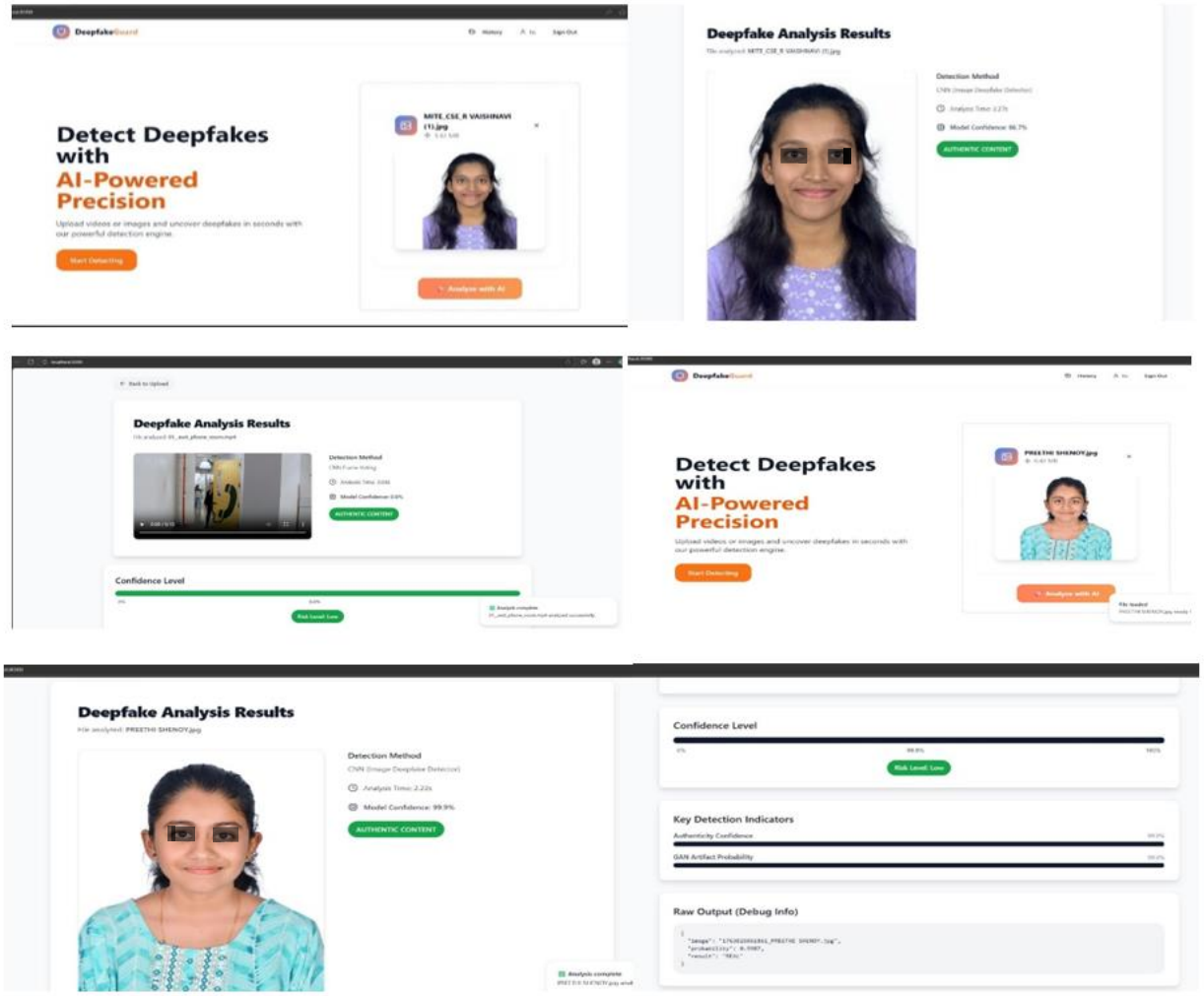
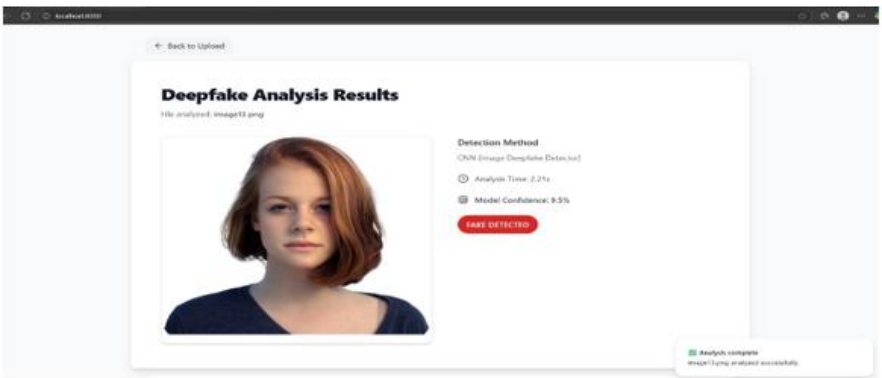


Figure 3: DeepFake detection output — real sample identified as authentic

The experimental findings confirm that integrating Generative Adversarial Network (GAN)-based spatial feature extraction with Optical Flow-based temporal motion analysis enhances detection accuracy compared to traditional single-method approaches (Figure 4).



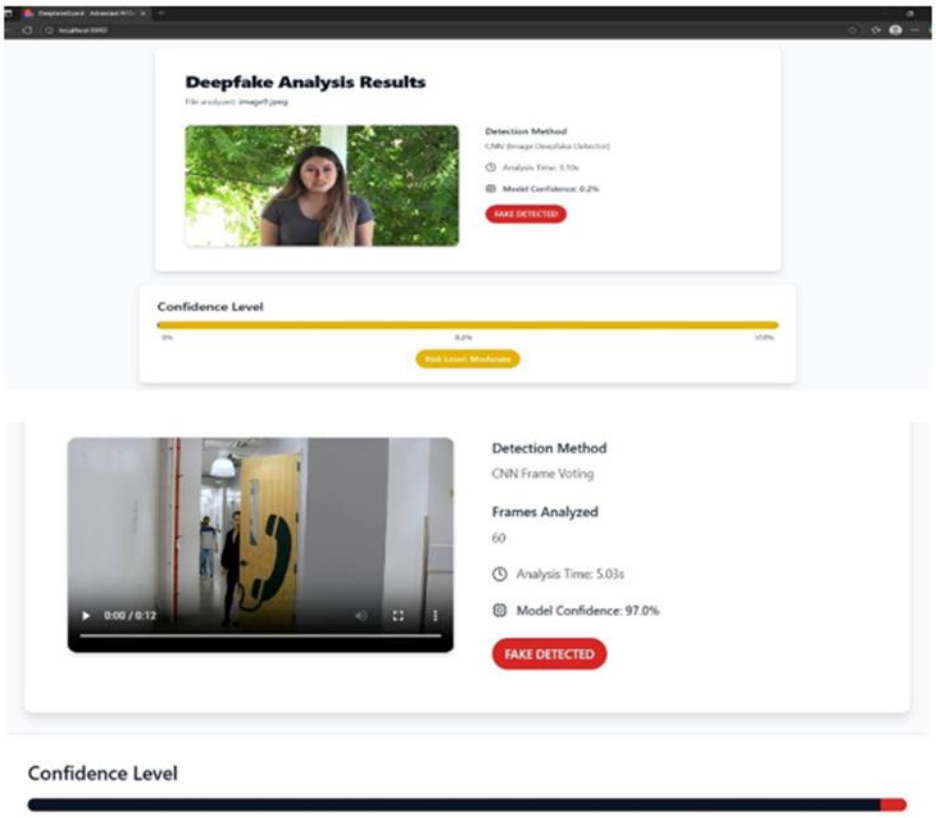


Figure 4: DeepFake detection output — manipulated sample identified as fake

4.3. Qualitative Results

From a qualitative standpoint, the proposed system exhibited strong capability in detecting fine-grained inconsistencies commonly found in DeepFake media. The GAN-based module effectively identified spatial artefacts, including unnatural lighting, texture mismatches, and boundary distortions. In contrast, the Optical Flow Analysis module captured irregular motion patterns and temporal inconsistencies between consecutive frames in video sequences. This integrated architecture ensures that both spatial and temporal characteristics are jointly analysed, enabling the detection of manipulations that may not be perceptible to the human eye. The model maintained stability and consistency in its predictions across diverse input samples, confirming its robustness, adaptability, and suitability for real-world conditions.

4.4. Discussion

The obtained results clearly demonstrate that the proposed hybrid framework—combining GAN-based spatial detection with Optical Flow-based temporal analysis—is highly effective in identifying DeepFake content. The system achieved balanced performance, with precision, recall, and F1-score of approximately 80%, reflecting its ability to classify manipulated and authentic media while minimising false detections. The high and stable accuracy confirms that the model successfully captures both pixel-level spatial inconsistencies and motion-level temporal irregularities. However, results also indicate a slight decrease in prediction confidence when analysing low-resolution or highly compressed samples. This limitation can be mitigated by incorporating more diverse datasets and employing adaptive thresholding techniques. Overall, the proposed DeepFake detection framework offers a reliable, interpretable, and efficient solution for authenticating digital media. It holds strong potential for deployment in practical scenarios such as social media monitoring, forensic investigations, and digital content verification, thereby enhancing the integrity and trustworthiness of visual information in the digital ecosystem.

5. Conclusion

Early, accurate, and reliable detection of manipulated videos, the DeepFake Detection system demonstrates how artificial intelligence can transform digital media verification. By automating time-consuming manual inspection, the system allows users to identify fake videos and subtle manipulations using a robust GAN-CNN framework integrated with Optical Flow analysis. This prevents the spread of misinformation and reduces the impact of misleading content on social platforms and

media. Recommendations for real-time incorporation of spatial and temporal analysis, confidence scoring, and frame-level visualisation to support targeted verification actions and improve decision-making in content moderation. Detection using artificial intelligence also enhances the ability to interpret authenticity at a detailed level, empowering platforms, organisations, and regulators that previously lacked reliable tools for video verification. Through web and mobile applications, the system provides fast, consistent analysis and helps maximise resource allocation and operational efficiency. Driven by data, these approaches enable informed judgments to strengthen digital security, content trustworthiness, and platform reliability. This intersection of AI and multimedia verification offers significant potential to reduce misinformation, protect users, and ensure proactive monitoring against evolving deepfake threats. For many years, the system has effectively achieved the goal of smart, reliable, and scalable DeepFake detection, demonstrating the potential of broader AI applications for globally secure, trustworthy digital media management.

5.1. Future Work

Future improvements to the DeepFake Detection system will focus on enhancing generalisation, adaptability, and accessibility. Expanding coverage beyond current benchmark datasets will improve robustness across diverse video sources, manipulation methods, and content domains. Integration with Internet of Things (IoT) devices and real-time data streams can further enrich contextual information—such as video source metadata, device characteristics, and compression parameters—supporting adaptive model learning and improved accuracy. Developing lightweight, efficient architectures optimised for edge deployment (e.g., smartphones, laptops, and embedded systems) will enable real-time, offline DeepFake detection in resource-constrained environments. This will enhance usability in remote regions with limited connectivity. Additionally, designing multilingual, user-friendly interfaces will promote inclusivity and accessibility, supporting adoption across platforms, regions, and cultural contexts. Collectively, these directions aim to build a scalable, intelligent, and universally accessible DeepFake detection ecosystem that can adapt to evolving digital media threats.

Acknowledgment: The authors express their sincere gratitude to Mangalore Institute of Technology and Engineering for providing the necessary resources and support to carry out this work.

Data Availability Statement: The data supporting the findings of this study are available from the corresponding author upon reasonable request, subject to applicable privacy and ethical considerations.

Funding Statement: This research was conducted without receiving any specific grant from funding agencies in the public, commercial, or not-for-profit sectors.

Conflicts of Interest Statement: The authors declare that there are no conflicts of interest regarding the publication of this paper. All sources have been duly acknowledged, and the work presented is original.

Ethics and Consent Statement: This study was carried out in accordance with established ethical standards. Informed consent was obtained from all participants, and appropriate measures were taken to ensure the confidentiality and privacy of the data collected.

References

1. P. Preeti, M. Kumar, and H. K. Sharma, "A GAN-Based Model of Deepfake Detection in Social Media," *Procedia Computer Science*, vol. 218, no. 1, pp. 2153–2162, 2023.
2. M. A. Sayedelahl and R. M. Farouk, "Hybrid approach to image segmentation with artificial neural networks and Gabor wavelets," *AVE Trends in Intelligent Computing Systems*, vol. 1, no. 2, pp. 77–90, 2024.
3. A. Malik, M. Kuribayashi, S. M. Abdullahi, and A. N. Khan, "DeepFake Detection for Human Face Images and Videos: A Survey," *IEEE Access*, vol. 10, no. 2, pp. 18757–18772, 2022.
4. M. S. Rana, M. N. Nobi, B. Murali, and A. H. Sung, "Deepfake Detection: A Systematic Literature Review," *IEEE Access*, vol. 10, no. 2, pp. 25494–25513, 2022.
5. S. L. B. Pentakota, "Unveiling deepfake and fraudulent content generation in GPT models and countermeasures," *AVE Trends in Intelligent Computer Letters*, vol. 1, no. 1, pp. 41–50, 2025.
6. I. Amerini, L. Galteri, R. Caldelli, and A. Del Bimbo, "Deepfake Video Detection through Optical Flow Based CNN," in *Proc. IEEE/CVF Int. Conf. Comput. Vis. Workshops (ICCVW)*, Seoul, South Korea, 2019.
7. A. Mary and A. Edison, "Deep Fake Detection Using Deep Learning Techniques: A Literature Review," in *Proc. Int. Conf. Control, Communication and Computing (ICCC)*, Thiruvananthapuram, India, 2023.
8. S. R. Bose, R. Vinoth, J. A. Jeba, R. Chauhan, and C. C. Angelin, "Artificial intelligence (AI) based symptom analysis using deep learning," *AVE Trends in Intelligent Health Letters*, vol. 1, no. 4, pp. 243–254, 2024.

9. M. Quadir, P. Agrawal, and C. Gupta, "A Comparative Analysis of Deepfake Detection Techniques: A Review," in *Proc. 6th Int. Conf. Contemporary Computing and Informatics (IC3I)*, Gautam Buddha Nagar, India, 2023.
10. R. Caldelli, L. Galteri, I. Amerini, and A. Del Bimbo, "Optical Flow Based CNN for Detection of Unlearnt Deepfake Manipulations," *Pattern Recognit. Lett.*, vol. 146, no. 6, pp. 31–37, 2021.
11. A. B. Nassif, Q. Nasir, M. A. Talib, and O. M. Gouda, "Improved Optical Flow Estimation Method for Deepfake Videos," *Sensors*, vol. 22, no. 7, pp. 1–18, 2022.
12. D. Parasar, R. Steffi, R. Regin, and K. D. Jasper, "Leveraging deep learning for adaptive intrusion prevention in smart devices," *AVE Trends in Intelligent Computing Systems*, vol. 2, no. 4, pp. 220–229, 2025.