International Research Journal of Innovations in Engineering and Technology (IRJIET)

ISSN (online): 2581-3048
Volume 10, Issue 5, pp 284-290, May-2026
https://doi.org/10.47001/IRJIET/2026.105038

DeepVision: A Hybrid Deepfake Detection Framework Using Deep Learning Approaches

[1]Dheeraj Shukla, [2]Dinesh Sonawane, [3]Jitendra Kulkarni, [4]Neha Agale, [5]Dakshita Pawar

[1,2,3,4,5]Department of Computer Engineering, P.S.G.V.P. Mandal's D.N. Patel College of Engineering, Shahada, India
E-mail: [1]dheerajbshukla@gmail.com, [2]dineshsonawanew2004@gmail.com, [3]jitendrakulkarni01@gmail.com, [4]nehaagle89@gmail.com, [5]dakshitathakur61@gmail.com

***Abstract* - Over the past decade, rapid progress in artificial intelligence (AI), machine learning, and deep learning has introduced sophisticated techniques for multimedia manipulation. Although such technologies have legitimate applications in entertainment and education, malicious actors increasingly exploit them for disinformation campaigns, political propaganda, identity fraud, and targeted harassment. High-quality synthetic videos and images commonly known as deepfakes pose a growing threat to digital security and public trust. This paper introduces DeepVision, a hybrid deepfake detection framework that fuses EfficientNet-B0 with a Vision Transformer (ViTB/16) to exploit both local texture features and global spatial dependencies simultaneously. The EfficientNet-B0 branch extracts fine-grained local texture and manipulation artefacts, while the Vision Transformer captures long range contextual relationships across facial regions using multi-head self-attention. The model is trained on a combined dataset derived from FaceForensics++ (FF++) and the DeepFake Detection Challenge (DFDC), comprising 120,000 labeled face images. Model performance is evaluated using accuracy, precision, recall, F1- score, confusion matrix, and ROC-AUC metrics. Experimental results demonstrate strong classification performance, achieving 98% accuracy and an AUC of 0.9973 on the combined dataset, representing competitive performance relative to recent state-of-the-art studies. The proposed framework supports both image-based and video-based deepfake detection and is suitable for real-world deployment in digital forensics and media authentication applications.**

Keywords: Deepfake Detection, Deep Learning, Convolutional Neural Network (CNN), Vision Transformer (ViT).

I. INTRODUCTION

In recent years, the rapid progress in artificial intelligence has given rise to deepfake technology, which uses Generative Adversarial Networks (GANs) and related techniques to generate realistic synthetic videos and images. Although deepfakes have some legitimate uses in entertainment and education, they are increasingly being misused for spreading disinformation, identity fraud, and political manipulation. Yu et al. noted that deepfake video detection remains a critical open problem, with generalization and robustness in real-world conditions being the most significant unsolved challenges [1].

As deepfake generation has grown more sophisticated, the research community has responded with a wide range of detection approaches. Wang et al. conducted a survey of fourteen Vision Transformer-based deepfake detection models, categorizing them into standalone, sequential, and parallel architectures, and found that while ViT-based detectors achieve AUC values up to 99.33%, limited local feature extraction and large data requirements remain key weaknesses [2]. Similarly, Rana et al. reviewed 112 deepfake detection papers from 2018 to 2020, covering deep learning, machine learning, statistical, and blockchain-based methods, and observed that while deep learning methods achieved up to 99% accuracy, overfitting under diverse real-world conditions was a consistent concern [3].

Beyond surveys, several detection systems have been proposed targeting different aspects of the problem. Alfraihi et al. combined ResNet50, MobileNetV3, and EfficientNetB7 with a heuristic optimizer for video forgery detection, achieving 95.26% accuracy, though the method is limited to copy-move forgery scenarios [4]. Soudy et al. built a three-model ensemble using a CNN for facial region features and a Convolutional Vision Transformer for full-face analysis, reaching 97% accuracy on FaceForensics++, but with reduced performance on cross-dataset evaluation [5]. Sar et al. proposed SachAI, a multimodal system combining video, audio, and image streams for deepfake detection across FaceForensics++ and Celeb-DF, achieving 97.76% accuracy on video and 99.13% on audio, demonstrating that fusing multiple modalities can substantially improve reliability [6].

Despite these advances, two key limitations remain. Zafar et al. showed that combining spatial and temporal features through EfficientNet B0 and Temporal Convolutional

International Research Journal of Innovations in Engineering and Technology (IRJIET)

ISSN (online): 2581-3048

Volume 10, Issue 5, pp 284-290, May-2026

https://doi.org/10.47001/IRJIET/2026.105038

Networks gives 92.45% accuracy at only 0.45 GFLOPs, but also highlighted that purely temporal models tend to miss the fine-grained spatial artefacts introduced by GAN-based face swaps [7]. Hussein and Mohamed demonstrated that ViT-based detectors using Deep-ViT and Cross-ViT can achieve 98% accuracy on FaceSwap, yet noted that such transformer architectures carry significant computational cost that limits their use in real-time applications [8].

To address these challenges, this paper proposes DeepVision, a hybrid deep learning framework for detecting deepfake images and videos. The proposed approach integrates EfficientNet-B0 with a Vision Transformer (ViT-B/16) architecture to capture both local texture features and global contextual information effectively. The system incorporates preprocessing steps including face detection, frame extraction, and normalization to enhance detection reliability. The proposed framework aims to provide an efficient and scalable solution for deepfake detection, contributing to improved digital security and public trust in multimedia content.

II. RELATED WORK

Deepfake detection research has evolved through several key methodological directions. This section reviews existing work organized by approach, summarizing the core techniques, datasets, and limitations of each study.

2.1 Survey and Review-Based Approaches

Yu et al. surveyed deepfake video detection methods, classifying them into five feature-based categories and concluding that generalization and robustness in real-world scenarios remain critical unsolved challenges [1]. Wang et al. provided a focused survey of fourteen ViT-based deepfake detection models organized into standalone, sequential, and parallel architectures, reporting AUC values up to 99.33%, while identifying data requirements and limited local feature extraction as primary limitations [2]. Rana et al. conducted a systematic literature review of 112 papers published between 2018 and 2020, categorizing techniques into deep learning, machine learning, statistical, and blockchain-based methods; deep learning-based methods achieved up to 99% accuracy but were noted to suffer from overfitting under diverse conditions [3].

2.2 CNN-Based and Transfer Learning Approaches

To incorporate temporal information, researchers began combining CNN spatial encoders with sequential models. Al-Zahrani et al. developed ECMVFD-FTLTDO, a fusion of ResNet50, MobileNetV3, and EfficientNetB7 transfer learning models with an enhanced recurrent neural network (ERNN) classifier optimized by the Tasmanian Devil Optimizer, achieving 95.26% and 92.67% accuracy on the GRIP and VTD datasets, respectively, though the method is limited to copy-move video forgery [4]. Soudy et al. proposed a three-model ensemble using a CNN for eye and nose feature extraction combined with a Convolutional Vision Transformer (CViT) for full-face detection, achieving 97% CNN accuracy and 85% CViT accuracy on FF++ and DFDC datasets using majority voting [5]. Sar et al. developed the SachAI framework, integrating Eulerian Video Magnification with a ResNext CNN for feature extraction, then feeding the resulting features into an LSTM classifier, achieving 97.76% accuracy on FaceForensics++ [6]. Zafar et al. proposed Enhanced EfficientNet B0 paired with Temporal Convolutional Networks (TempCNNs), obtaining 92.45% accuracy on the FFIW 10K dataset with a lightweight 0.45 GFLOPs computational footprint [7].

2.3 Vision Transformer (ViT)-Based Approaches

Hussein and Mohamed integrated Deep-ViT and Cross-ViT architectures to classify deepfakes on FF++, achieving 98% accuracy on FaceSwap manipulation using a re-attention mechanism that mitigates attention collapse [8]. Nguyen et al. proposed FakeFormer, which extends ViT with a Learning-based Local Attention (L2-Att) module targeting artifact-vulnerable patches, achieving a cross-dataset AUC exceeding 94% across six benchmarks with low computational cost [9]. Chen et al. introduced GFF, a frozen CLIP-ViT guided by a trainable Deepfake- Specific Feature Guidance Module (DFGM) and a multi-stage FuseFormer, achieving 99% accuracy on unseen GAN-generated images and 97% on diffusion model outputs after only five training epochs on ProGAN data [10]. Lamichhane fine-tuned a ViT model on 30,000 GAN-generated images from Kaggle, achieving 98.2% accuracy with a 16×16 patch size and 12-layer configuration, though the method is limited to static images [11].

2.4 Hybrid Spatial-Temporal Approaches

Sar et al. developed SachAI, a multimodal framework combining Eulerian Video Magnification with ResNext and LSTM for video (97.76%), a CNN for audio (99.13%), and DenseNet121 for image deepfake detection (93.64%) across FF++ and Celeb-DF datasets [6]. Zafar et al. combined Enhanced EfficientNet B0 with a Temporal Convolutional Network (TempCNN) and Feature Pyramid Network (FPN), achieving 92.45% accuracy on the FFIW 10K dataset at only 0.45 GFLOPs, making it suitable for resource-constrained real-world deployment [7]. Alrawahneh et al. proposed a hybrid ResNext50 and Bidirectional LSTM (BiLSTM) framework that captures both spatial and bidirectional temporal features, achieving 96.11% accuracy and 98.89%

International Research Journal of Innovations in Engineering and Technology (IRJIET)

ISSN (online): 2581-3048

Volume 10, Issue 5, pp 284-290, May-2026

https://doi.org/10.47001/IRJIET/2026.105038

AUC across FF++, DFDC, and Celeb-DF datasets [12]. Almestekawy et al. combined a 3D CNN with a Siamese architecture and spatiotemporal texture features to improve reproducibility, achieving 97.51% AUC on same-dataset evaluation and 95.44% AUC on cross-dataset evaluation across four benchmarks [13].

2.5 Biometric Feature-Based Approaches

Tchaptchet et al. proposed a two-level iris analysis method exploiting the biological symmetry of authentic human irises using gradient maps, achieving 0.979 accuracy and 0.921 sensitivity on FFHQ and StyleGAN2 datasets. The method is interpretable and computationally efficient, but is limited to GAN-synthesized face detection and does not generalize to face reenactment deepfakes [14].

Table 1: Summary of Related Work on Deepfake Detection

Author	Model	Dataset	Results (Acc)	Limitations
Rana et al.(2022)[3]	SLR: DL, ML, Statistical, Blockchain	FF++, Celeb-DF, DFDC	Up to 99% (DL)	Overfitting; limited dataset diversity
Wang et al.(2024)[2]	ViT Survey (14 models)	FF++, Celeb-DF, DFDC, DFD	AUC up to 99.33%	Data-hungry; weak local feature extraction
Hussein et al.(2024)[4]	Deep-ViT + Cross-ViT	FF++, CelebDF-V2	98% (FaceSwap)	High compute; limited cross-dataset testing
Nguyen et al. (2024)[5]	FakeFormer (ViT + L2-Att)	FF++, Celeb-DF, DFDC, DFD	AUC ~94% cross-dataset	Sensitive to deepfake quality level
Chen et al. (2024)[6]	GFF (CLIP-ViT + DFGM + FuseFormer)	ProGAN, unseen GANs, Diffusion	~99% (GANs); 97% (Diffusion)	Trained on ProGAN only; no text modality
Almestekawy et al. (2024)[7]	3D CNN + Siamese + Spatiotemporal Attention	Celeb-DF, FF++, DeepfakeTIMIT	AUC 97.51%; Acc 91.96%	High memory; limited manipulation types
Al-Dhaheri et al. (2024)[8]	CNN + CViT (3-model ensemble)	FF++, DFDC	CNN 97%; CViT 85%	Lower CViT accuracy; video only
Lamichhane(2025)[9]	Fine-tuned ViT (ImageNet pretrained)	Kaggle AI-Image Dataset (30K)	98.2%	Static images only; no video support
Tchaptchet et al.(2025)[10]	Iris Gradient Map + Pupil Shape Analysis	FFHQ, StyleGAN2	Acc 0.979; Sens 0.921	GAN faces only; eye conditions cause errors
Zafar et al.(2025)[11]	EfficientNet B0 + TempCNN + FPN	FFIW 10K	92.45% (0.45 GFLOPs)	Single dataset; limited generalization
Sar et al.(2025)[12]	Sach-AI: ResNext+LSTM+CNN+DenseNet121	FF++, Celeb-DF, ASVspoof 2019	Video 97.76%; Audio 99.13%	High compute; real-time scalability limited
Alrawahneh et al.(2025)[13]	ResNext50 + BiLSTM	FF++, DFDC, Celeb-DF	Acc 96.11%; AUC 98.89%	Expensive; cross-dataset not fully explored
Al-Zahrani et al.(2025)[14]	ResNet50+MobileNetV3+EfficientNetB7+ERNN+TDO	GRIP, VTD	GRIP 95.26%; VTD 92.67%	Copy-move only; not on face-manipulation benchmarks

III. METHODOLOGY

The DeepVision system follows a structured end-to-end pipeline for deepfake detection. Real and fake face image datasets are first collected and organized from publicly available benchmarks. A preprocessing stage performs face detection, frame extraction, resizing, and normalization to ensure uniform input quality. The hybrid model integrates EfficientNet-B0 for local texture feature extraction and ViT-B/16 for capturing global contextual relationships. The system is trained using supervised learning on labeled datasets and evaluated on held-out test splits.

3.1 Dataset Preparation

This study uses a combined deepfake image dataset created from two widely used benchmark sources: FaceForensics++ (FF++) [15] and the DeepFake Detection Challenge (DFDC) dataset [16]. The combined dataset contains 120,000 images, comprising both authentic and manipulated face samples extracted from real and synthetic videos. Images are organized into two binary classes Real and Fake and partitioned as follows: 80,000 images for training, 20,000 for validation, and 20,000 for testing. This partition ensures rigorous evaluation while maintaining a balanced class distribution throughout all splits.

ISSN (online): 2581-3048
Volume 10, Issue 5, pp 284-290, May-2026
https://doi.org/10.47001/IRJIET/2026.105038

3.2 Data Preprocessing

All images are resized to 224×224 pixels and normalized using channel-wise mean and standard deviation prior to model input. Data augmentation techniques including random horizontal flipping, rotation, and color jitter are applied during training to improve model generalization and reduce overfitting. For video-based deepfake detection, uploaded videos are first decoded into individual frames using frame extraction, after which each frame is processed independently through the same preprocessing pipeline.

3.3 Hybrid Deep Learning Architecture

The proposed DeepVision model employs a hybrid architecture fusing EfficientNet-B0 and Vision Transformer (ViT-B/16). EfficientNet-B0 serves as a local feature extractor, capturing fine-grained texture patterns, blending boundaries, and spatial manipulation artefacts at the pixel level. The ViT-B/16 branch divides each input image into non-overlapping 16×16 patches and applies multi-head self-attention to model long-range dependencies across facial regions. The feature representations produced by both branches are concatenated and passed through fully connected layers followed by a sigmoid/softmax output layer for binary classification into Real or Fake categories.

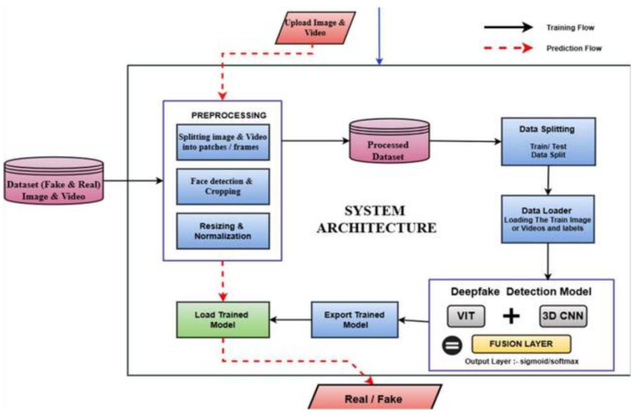


Figure 1: System Architecture DeepVision

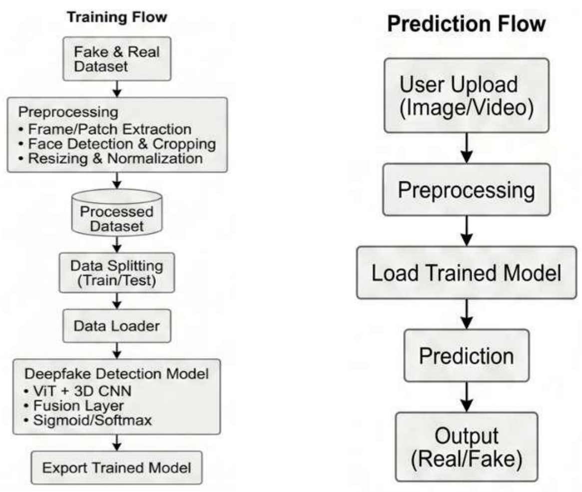
3.4 Image and Video Deepfake Detection

DeepVision supports both image-based and video-based deepfake detection within a unified inference pipeline. For image detection, an uploaded image is preprocessed and directly classified by the trained hybrid model. For video detection, the system first decodes the video into individual frames, preprocesses each frame, and analyzes it independently using the trained model. The video is classified as fake if manipulated features or inconsistencies are detected in any extracted frame. This frame-wise approach leverages visual inconsistencies introduced by deepfake generation processes and provides reliable detection across different video lengths and encoding formats.

3.5 Model Training and Evaluation

The model is implemented using the PyTorch deep learning framework and trained using the AdamW optimizer with a weighted cross-entropy loss function to handle potential class imbalance. Validation accuracy is monitored after each epoch to save the best-performing model checkpoint. The final system is evaluated using accuracy, precision, recall, F1-score, confusion matrix, and ROC-AUC score to comprehensively measure classification performance.

3.6 System Workflow



(a)Training Flow **(b) Prediction Flow**

Figure 2: Training and Prediction flow for DeepVision

IV. RESULTS AND DISCUSSION

The performance of the DeepVision framework was assessed using multiple evaluation metrics, including accuracy, precision, recall, F1-score, confusion matrix analysis, and ROC-AUC score. The system was tested on 20,000 held-out images comprising 10,000 real and 10,000 fake samples to evaluate the effectiveness of the proposed hybrid deep learning architecture under balanced class conditions.

4.1 Experimental Setup

The proposed DeepVision framework was implemented using the PyTorch deep learning framework with a hybrid EfficientNet-B0 and Vision Transformer (ViT-B/16) architecture for deepfake detection. The experimental dataset consisted of 120,000 labeled face images collected from the FaceForensics++ (FF++) and DeepFake Detection Challenge (DFDC) datasets, which were divided into training, validation, and testing sets containing 80,000, 20,000, and 20,000

International Research Journal of Innovations in Engineering and Technology (IRJIET)
ISSN (online): 2581-3048
Volume 10, Issue 5, pp 284-290, May-2026
https://doi.org/10.47001/IRJIET/2026.105038

samples respectively. All input images were resized to 224×224 pixels and normalized before training. To improve model robustness and generalization capability, data augmentation techniques such as random horizontal flipping, rotation, and color jittering were applied during preprocessing. The model was trained using the AdamW optimizer with a weighted cross-entropy loss function for 10 epochs using a batch size of 32 and an initial learning rate of 0.0001. Validation accuracy was monitored after each epoch, and model checkpointing was used to save the best-performing model and reduce overfitting. The experiments were conducted using GPU-accelerated computation within the Kaggle environment, and the final model performance was evaluated using accuracy, precision, recall, F1-score, confusion matrix analysis, and ROC-AUC metrics to comprehensively assess the effectiveness of the proposed deepfake detection system.

4.2 Training and Validation Analysis

The learning behavior of the model was monitored over 10 training epochs to assess convergence, stability, and generalization capability.

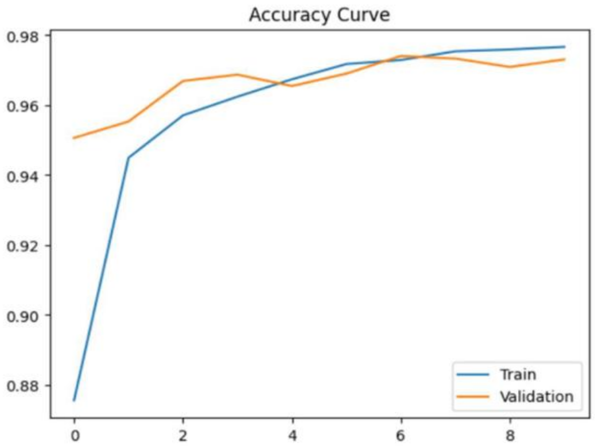


Figure 3: Training and Validation curve DeepVision

- **Initial Learning:** Training accuracy began at 87.5% and improved substantially by Epoch 2, reaching 95.8%, indicating rapid feature learning in the early training phase.
- **Generalization:** Validation accuracy remained consistently high throughout training, starting at 95.1% and converging at 97.3%, confirming stable learning without significant overfitting.
- **Overfitting Analysis:** The marginal gap between final training accuracy (97.6%) and validation accuracy (97.3%)—less than 0.3 percentage points—confirms that the model generalizes effectively to previously unseen data.

4.3 Testing Performance

4.3.1 Classification Report Analysis

Table 2: Classification Report of the DeepVision Model

Class	Precision	Recall	F1-Score	Support
0 (Real)	0.98	0.98	0.98	10,000
1 (Fake)	0.98	0.98	0.98	10,000
Overall Accuracy	**0.98**	**0.98**	**0.98**	**20,000**
Macro Avg	0.98	0.98	0.98	20,000
Weighted Avg	0.98	0.98	0.98	20,000

As shown in Table 2, the DeepVision model achieves balanced and high classification performance across both the Real and Fake classes. Both classes obtain precision, recall, and F1-score values of 0.98, confirming that the model does not exhibit a systematic bias toward either class. The overall accuracy of 98% on the 20,000-sample test set demonstrates the robustness of the hybrid EfficientNet-B0 and ViT fusion approach.

4.3.2 Confusion Matrix

The confusion matrix provides a detailed breakdown of classification outcomes across the 20,000-sample test set:

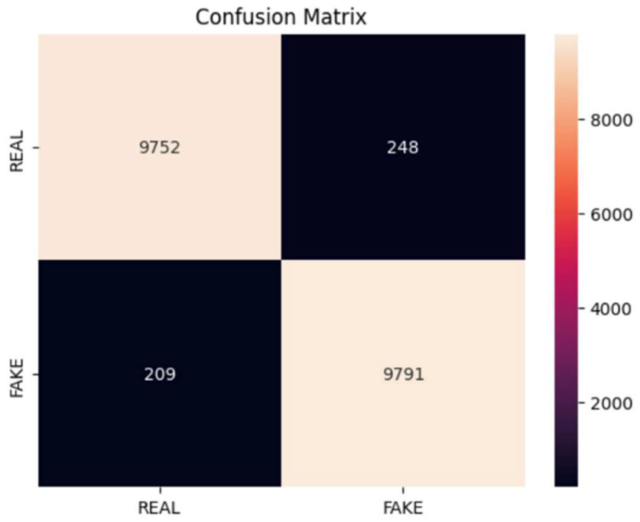


Figure 4: System Architecture DeepVision

- **True Negatives (TN):** 9,752 authentic images were correctly identified as real.
- **True Positives (TP):** 9,791 deepfake images were correctly detected as manipulated.
- **False Positives (FP):** 248 real images were incorrectly flagged as manipulated (Type I error, false alarm rate of 2.48%).
- **False Negatives (FN):** 209 deepfake images were missed by the system (Type II error, miss rate of 2.09%), maintaining a high security threshold for detection.

International Research Journal of Innovations in Engineering and Technology (IRJIET)

ISSN (online): 2581-3048

Volume 10, Issue 5, pp 284-290, May-2026

https://doi.org/10.47001/IRJIET/2026.105038

The recall of 97.9% for the Fake class is particularly significant for the DeepVision application, as it reflects the model's robustness in capturing the vast majority of manipulated content while keeping the false negative rate low—a critical requirement for security-sensitive deployment scenarios.

4.3.3 ROC-AUC Analysis

The ROC curve of the DeepVision model demonstrates excellent discriminative performance, yielding an AUC score of 0.9973. The curve rises sharply toward the upper-left corner of the ROC space, indicating a very high True Positive Rate (TPR) at a very low False Positive Rate (FPR).

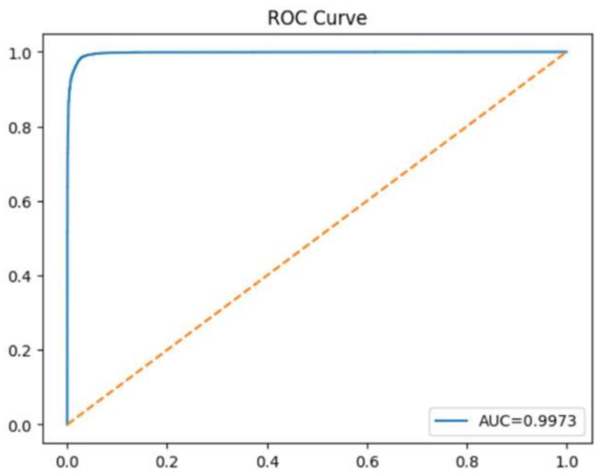


Figure 5: ROC Curve for DeepVision

This behavior confirms that the hybrid EfficientNet-B0 and ViT architecture can effectively separate real from fake samples across a wide range of decision thresholds. The curve remains significantly above the random classifier diagonal, confirming that the system performs substantially better than chance. The high AUC value attests to the model's strong robustness, reliability, and generalization capability for real-world deepfake detection tasks.

The results demonstrate that DeepVision achieves highly competitive performance compared to recent methods, particularly in terms of AUC score (0.9973), which is indicative of strong discriminative capability across varying decision thresholds. The hybrid fusion of EfficientNet-B0 and ViT effectively addresses the complementary limitations of purely spatial and purely transformer-based architectures, enabling robust detection of both local texture manipulation and global facial inconsistencies.

V. CONCLUSION

This paper presented an effective deepfake detection system using deep learning techniques to identify fake images and videos. The proposed model combines CNN and Vision Transformer methods to analyze both facial details and overall image patterns for better detection accuracy. Experimental results showed that the system can successfully detect manipulated media with high performance on combine datasets. The study highlights the importance of AI-based solutions in preventing the misuse of fake digital content and reducing the spread of misinformation. The proposed work can be further improved in the future for real-time detection and deployment in social media and security applications.

REFERENCES

[1] P. Yu, Z. Xia, J. Fei, and Y. Lu, "A survey on deepfake video detection," *IET Biometrics*, vol. 10, no. 6, pp. 607–624, 2021.

[2] Z. Wang, Z. Cheng, J. Xiong, X. Xu, T. Li, B. Veeravalli, and X. Yang, "A timely survey on vision transformer for deepfake detection," *arXiv preprint arXiv*:2405.08463, 2024.

[3] M. S. Rana, M. N. Nobi, B. Murali, and A. H. Sung, "Deepfake detection: A systematic literature review," *IEEE Access*, vol. 10, pp. 25494–25513, 2022.

[4] H. Alfraihi et al., "A multi-model feature fusion-based transfer learning with heuristic search for copy-move video forgery detection," *Scientific Reports*, vol. 15, no. 1, Art. no. 4738, 2025.

[5] A.H. Soudy et al., "Deepfake detection using convolutional vision transformers and convolutional neural networks," *Neural Computing and Applications*, vol. 36, pp. 19759–19775, 2024, doi: 10.1007/s00521-024-10181-7.

[6] A.Sar et al., "A unified neural framework for real-time deepfake detection across multimedia modalities to combat misleading content," *IEEE Access*, vol. 13, pp. 48683–48702, 2025, doi: 10.1109/ACCESS.2025.3550770.

[7] F. Zafar, T. A. Khan, S. Akbar, M. T. Ubaid, S. Javaid, and K. A. Kadir, "A hybrid deep learning framework for deepfake detection using temporal and spatial features," *IEEE Access*, vol. 13, pp. 79560–79570, 2025, doi: 10.1109/ACCESS.2025.3566008.

[8] S. A. Hussein and S. N. Mohamed, "Deepfake video detection using a vision transformer," *International Journal of Intelligent Computing and Information Sciences*, vol. 24, no. 1, pp. 55–68, 2024.

[9] D. Nguyen, M. Astrid, E. Ghorbel, and D. Aouada, "FakeFormer: Efficient vulnerability-driven transformers for generalisable deepfake detection," *arXiv preprint arXiv*:2410.21964v2, 2024.

[10] Y. Chen, L. Zhang, Y. Niu, P. Chen, L. Tan, and J. Zhou, "Guided and fused: Efficient frozen CLIP-ViT with feature guidance and multi-stage feature fusion for

ISSN (online): 2581-3048
Volume 10, Issue 5, pp 284-290, May-2026
https://doi.org/10.47001/IRJIET/2026.105038

generalizable deepfake detection," *arXiv preprint arXiv*:2408.13697v1, 2024.

[11] D. Lamichhane, "Advanced detection of AI-generated images through vision transformers," *IEEE Access*, vol. 13, pp. 3644–3652, 2025, doi: 10.1109/ACCESS.2024.3522759.

[12] A.A.-M. Alrawahneh et al., "Decision-aid framework for face authentication detection using ResNext50 and BiLSTM to enhance media integrity," *IEEE Access*, vol. 13, pp. 89858–89873, 2025, doi: 10.1109/ACCESS.2025.3569792.

[13] A.Almestekawy, H. H. Zayed, and A. Taha, "Deepfake detection: Enhancing performance with spatiotemporal texture and deep learning feature fusion," *Egyptian Informatics Journal*, vol. 27, Art. no. 100535, 2024.

[14] E. Tchaptchet et al., "Deepfakes detection by iris analysis," *IEEE Access*, vol. 13, pp. 8977–8987, 2025.

[15] A.Rössler, D. Cozzolino, L. Verdoliva, C. Riess, J. Thies, and M. Nießner, "FaceForensics++: Learning to detect manipulated facial images," *in Proceedings of the IEEE/CVF International Conference on Computer Vision (ICCV)*, 2019, pp. 1–11.

[16] B. Dolhansky, J. Bitton, B. Pflaum, J. Lu, R. Howes, M. Wang, and C. Canton Ferrer, "The DeepFake Detection Challenge (DFDC) dataset," *arXiv preprint arXiv*:2006.07397, 2020.

Citation of this Article:

Dheeraj Shukla, Dinesh Sonawane, Jitendra Kulkarni, Neha Agale, & Dakshita Pawar. (2026). DeepVision: A Hybrid Deepfake Detection Framework Using Deep Learning Approaches. *International Research Journal of Innovations in Engineering and Technology - IRJIET*, 10(5), 284-290. Article DOI https://doi.org/10.47001/IRJIET/2026.105038

****