International Journal of Engineering Inventions
e-ISSN: 2278-7461, p-ISSN: 2319-6491
Volume 15, Issue 2 [February 2026] PP: 54-58

Frame Level Deepfake Detection

Subramanian.E[1], Ajay.k[2], Ananthu.A.S[3], Gobi Krishnan.M[4,] Harish.S[5]

[1]*Assistant Professor, Computer Science Engineering, Sri Shakthi Institute of Engineering and Technology, Coimbatore, India*
[2]*Student, Computer Science Engineering, Sri Shakthi Institute of Engineering and Technology, Coimbatore, India*
[3]*Student, Computer Science Engineering, Sri Shakthi Institute of Engineering and Technology, Coimbatore, India*
[4]*Student, Computer Science Engineering, Sri Shakthi Institute of Engineering and Technology, Coimbatore, India*
[5]*Student, Computer Science Engineering, Sri Shakthi Institute of Engineering and Technology, Coimbatore, India*

Abstract
This work proposes a frame-level deepfake detection system using a fine-tuned EfficientNetB5 model to identify manipulated facial content. Video frames are extracted and augmented to improve robustness against real-world variations, while AdamW optimization and a cosine learning rate scheduler ensure stable training. Grad-CAM visualizations provide interpretability, and a web-based interface enables real-time detection. The system achieves strong performance across standard evaluation metrics, offering a reliable and user-friendly deepfake detection solution.

Keywords: *Deepfake Detection, Frame-Level Analysis, EfficientNetB5, Transfer Learning, Grad-CAM, Computer Vision*

Date of Submission: 22-01-2026 Date of acceptance: 04-02-2026

I. INTRODUCTION

Recent advances in artificial intelligence have enabled the creation of highly realistic synthetic media known as deepfakes. While these technologies have useful applications, they also pose serious risks to privacy, security, and trust in digital content by making fake images and videos difficult to distinguish from real ones.

The increasing misuse of deepfakes for misinformation, identity fraud, and media manipulation has made their detection a critical challenge in computer vision and cybersecurity. Reliable automated systems are essential to preserve the authenticity of online information.

This project, *Frame-Level Deepfake Detection*, focuses on analyzing individual frames using deep learning to determine media authenticity. Frame-level analysis allows faster, more accurate detection and supports real-time and limited-data scenarios.

II. LITERATURE SURVEY

The rapid growth of deepfake and synthetic media technologies has raised serious concerns regarding media authenticity, privacy, and public trust. As deepfake content becomes increasingly realistic, researchers have focused on developing automated detection techniques using deep learning. Recent studies highlight the need for robust and scalable systems capable of identifying manipulated facial content in images and videos.

Early deepfake detection approaches relied on hand-crafted features such as eye blinking and facial motion inconsistencies, but these methods proved ineffective against advanced generative models like GANs and autoencoders. Modern research has shifted toward deep convolutional neural networks and transformer-based models that can automatically learn discriminative features, significantly improving detection accuracy.

Transfer learning has emerged as an effective strategy in deepfake detection, especially when labeled data is limited. Pretrained models such as ResNet and EfficientNet, when fine-tuned on deepfake datasets, have demonstrated strong performance while reducing training time. This approach leverages rich feature representations learned from large-scale image datasets.

Recent literature also emphasizes the advantages of frame-level analysis over video-level classification. Frame-wise detection enables faster inference, higher flexibility, and real-time applicability, particularly in scenarios with limited video data. Combined with effective data preprocessing and augmentation techniques, frame-level models offer improved robustness and practicality for real-world deployment.

III. ARCHITECTURE OF FRAME-LEVEL DEEPFAKE DETECTION SYSTEM

The proposed system, **Frame-Level Deepfake Detection**, is designed to automatically identify manipulated facial content in images and videos through deep learning–based frame-wise analysis. The system leverages a modular and layered architecture that integrates deep neural networks, efficient data pipelines, explainable AI techniques, and interactive web-based interfaces. This architecture ensures scalability, interpretability, real-time inference, and ease of deployment.

The system follows a modular layered architecture, divided into four functional layers:

1. Input Acquisition and Frame Extraction Layer
2. Data Processing and Augmentation Layer
3. Deep Learning and Inference Layer
4. Explainability and Visualization Layer
5. User Interface and Deployment Layer

Each layer operates independently while interacting through well-defined data flows, ensuring flexibility, maintainability, and high performance.

1. Input Acquisition and Frame Extraction Layer

This layer is responsible for handling all input sources, including static images and video files.

- Users upload images or videos through the web interface.
- For video inputs, **OpenCV** is used to extract individual frames at fixed intervals
- Each extracted frame is treated as an independent sample, enabling fine-grained frame-level analysis.

Workflow:

1. Video file is read using OpenCV's VideoCapture
2. Frames are extracted, resized, and stored temporarily.
3. Each frame is forwarded to the preprocessing pipeline.

This approach allows detection even when only a few frames are available and supports real-time analysis scenarios.

2. Data Processing and Augmentation Layer

This layer prepares frames for deep learning inference and improves model generalization.

Key responsibilities include::

- Image resizing and normalization
- Color space conversion
- Data augmentation

Techniques Used:

- Albumentations and torchvision transforms perform random flipping, rotation, brightness adjustment, and normalization.
- Augmentation simulates real-world variations such as lighting changes, pose differences, and facial expressions.

By introducing controlled randomness, this layer reduces overfitting and enhances robustness across diverse datasets.

3. Deep Learning and Inference Layer

This is the core computational layer of the system.

- Implemented using **PyTorch 2.x**
- Uses **EfficientNet-B5** as the backbone CNN architecture
- Model is pretrained on ImageNet and fine-tuned using transfer learning

Training and Optimization:

- **AdamW optimizer** ensures stable weight updates
- **Cosine learning rate scheduler** improves convergence
- Mixed-precision training with **CUDA 11.8** accelerates computation on NVIDIA GPU

Each frame is classified independently as **Real** or **Fake**, along with a confidence score. This modular design allows easy replacement or extension with newer models in future iterations.

4. Explainability and Visualization Layer

To address the black-box nature of deep learning models, explainability is integrated into the system.

- **Grad-CAM (Gradient-weighted Class Activation Mapping)** is used to generate heatmaps
- Highlights facial regions influencing the model's decision

Functionality:

- Activations are extracted from the final convolutional layers.
- Gradient-weighted maps are overlaid on original frames.
- Helps users visually understand manipulation cues

This layer enhances transparency, trust, and interpretability, which is critical for forensic and security-related applications.

5. User Interface and Deployment Layer

The frontend layer enables user interaction and real-time testing.

- Built using **Gradio 4** and **Streamlit**
- Provides the web-based interface **"Deepfake Vision Studio"**

Features:

- Image and video frame upload.
- Real-time prediction results.
- Confidence score visualization
- Grad-CAM heatmap display
- Responsive UI with embedded HTML/CSS

This layer abstracts the technical complexity, making the system accessible to both technical and non-technical users.

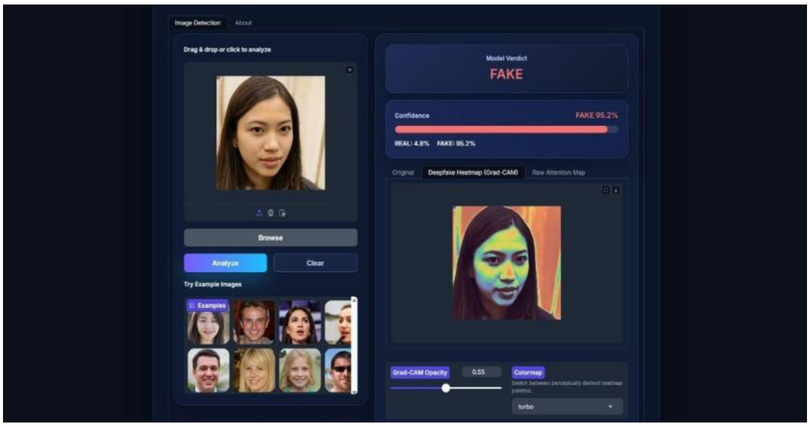
IV. WORKFLOW AND IMPLEMENTATION

When the system starts, the user uploads an image or video through the web interface. If a video is provided, it is decomposed into individual frames using OpenCV. Each frame passes through preprocessing and augmentation before being fed into the EfficientNet-B5 model for inference.

The model predicts whether each frame is real or fake, generating a confidence score. Simultaneously, Grad-CAM visualizations are computed to highlight influential facial regions. Results are streamed back to the interface in real time, allowing users to inspect predictions visually.

This end-to-end workflow enables fast, interpretable, and accurate deepfake detection at the frame level. The modular architecture supports future integration with cloud platforms, video surveillance systems, or large-scale forensic pipelines, making it a strong foundation for real-world deployment.

In addition, the system maintains detailed execution logs and prediction records during both training and inference phases to ensure traceability and performance monitoring. Each processed frame is associated with its prediction outcome, confidence score, and corresponding Grad-CAM visualization, enabling systematic analysis and debugging. The modular workflow allows seamless updates to model parameters, datasets, or preprocessing techniques without disrupting the overall pipeline. This extensible design supports continuous improvement, scalability, and future integration with real-time video streams or cloud-based forensic platforms.



Drag & drop or click to analyze

Model Verdict
FAKE

Confidence FAKE 95.2%

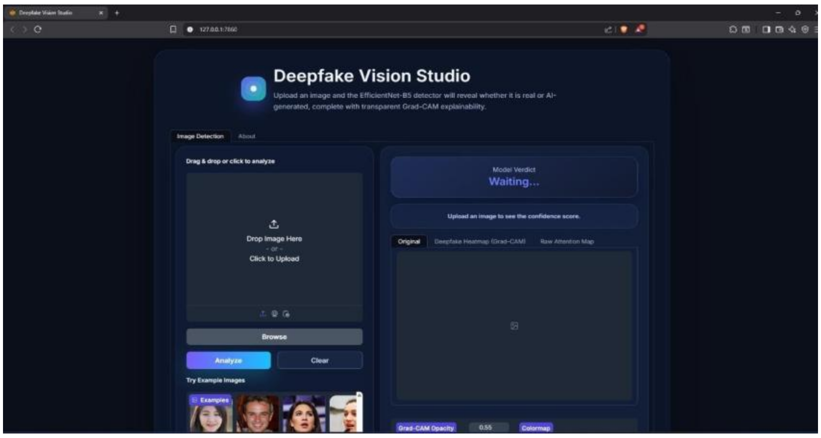
REAL: 4.8% FAKE: 95.2%

Grad-CAM Opacity 0.55

Colormap: Switch between perceptually distinct heatmap palettes.

turbo

Fig. 1: Deepfake Detection Page



Deepfake Vision Studio

Upload an image and the EfficientNet-B5 detector will reveal whether it is real or AI-generated, complete with transparent Grad-CAM explainability.

Drag & drop or click to analyze

Drop Image Here
- or -
Click to Upload

Model Verdict
Waiting...

Upload an image to see the confidence score.

Grad-CAM Opacity 0.55

Fig. 1: Gram Cam Result Page

V. RESULT AND DISCUSSION

The results show that the proposed **Frame-Level Deepfake Detection System** effectively distinguishes real facial content from manipulated data. Using EfficientNet-B5 with transfer learning, the model achieves high accuracy, precision, recall, and F1-score, demonstrating reliable performance in frame-level deepfake classification across diverse datasets.

Training stability is achieved through the AdamW optimizer and cosine learning rate scheduler, resulting in smooth convergence and minimal overfitting. Data augmentation further improves robustness against variations in lighting, pose, and facial expressions, enabling better generalization to real-world scenarios.

Explainability is enhanced through the integration of Grad-CAM, which highlights key facial regions influencing model predictions. These visual explanations help users understand the model's decisions and increase trust in the detection process, while also aiding forensic analysis.

The **Deepfake Vision Studio** interface provides a simple and interactive platform for real-time testing. Users can upload images or video frames, view prediction confidence scores, and analyze Grad-CAM heatmaps, confirming that the system successfully combines accuracy, interpretability, and usability for effective deepfake detection.

VI. CONCLUSION AND FUTURE ENHANCEMENTS

The **Frame-Level Deepfake Detection System** successfully demonstrates the effectiveness of deep learning and explainable AI in identifying manipulated facial content in images and videos. By utilizing EfficientNet-B5 with transfer learning in the PyTorch 2.x framework, the system achieves high accuracy and reliable frame-level classification. Optimization techniques such as AdamW, cosine learning rate scheduling, and extensive data augmentation contribute to stable training and strong generalization. The integration of Grad-CAM enhances interpretability by visually highlighting key facial regions influencing predictions, while the **Deepfake Vision Studio** interface provides a user-friendly platform for real-time analysis, confidence visualization, and transparent decision-making.

In the Future enhancements can focus on extending the system beyond frame-level analysis to incorporate temporal modeling for full video-based deepfake detection. Integrating multi-modal learning with audio-visual features, optimizing the model for real-time and edge-device deployment, and adopting advanced explainability techniques such as SHAP or LIME can further improve performance and trustworthiness. Additionally, developing mobile or browser-based tools and strengthening robustness against adversarial attacks will enable broader adoption, positioning the system as a scalable and comprehensive solution for combating deepfake-driven misinformation and ensuring digital media authenticity.

REFERENCES

[1] Tolosana, R., et al. (2020). Deepfakes and beyond: A survey of face manipulation and fake detection. *Information Fusion*, 64, 131–148.

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

[3] Sabir, E., et al. (2019). Recurrent convolutional strategies for face manipulation detection in videos. *IEEE Conference on Computer Vision and Pattern Recognition Workshops (CVPRW)*.

[4] Rossler, A., et al. (2019). FaceForensics++: Learning to detect manipulated facial images. *IEEE International Conference on Computer Vision (ICCV)*, 1–11.

[5] Afchar, D., et al. (2018). MesoNet: A compact facial video forgery detection network. *IEEE International Workshop on Information Forensics and Security (WIFS)*, 1–7.

[6] Tan, M., & Le, Q. (2019). EfficientNet: Rethinking model scaling for convolutional neural networks. *International Conference on Machine Learning (ICML)*, 6105–6114.

[7] Selvaraju, R. R., et al. (2017). Grad-CAM: Visual explanations from deep networks via gradient-based localization. *IEEE International Conference on Computer Vision (ICCV)*, 618–626.

[8] Paszke, A., et al. (2019). PyTorch: An imperative style, high-performance deep learning library. *Advances in Neural Information Processing Systems (NeurIPS)*, 8026–8037.

[9] Buslaev, A., et al. (2020). Albumentations: Fast and flexible image augmentations. *Information*, 11(2), 125.

[10] Pedregosa, F., et al. (2011). Scikit-learn: Machine learning in Python. *Journal of Machine Learning Research*, 12, 2825–2830.