International Journal of Engineering Research & Technology (IJERT)
ISSN: 2278-0181
Vol. 14, Issue 03, ICCT - 2026

DEEPFAKE DETECTOR FOR DIGITAL EVIDENCE IN COURT

Y. Sheela
Faculty
Computer Science and Engineering,
Jayaraj Annapackiam CSI College of Engineering, Nazareth, India
sheelajabez@gmail.com

J. Jackshiya
Student
Computer Science and Engineering,
Jayaraj Annapackiam CSI College of Engineering, Nazareth, India.
jackshiyajacob@gmail.com

M. Vanitha
Student
Computer Science and Engineering,
Jayaraj Annapackiam CSI College of Engineering, Nazareth, India.
vanitham2305@gmail.com

A. Cathrin Win Smilein
Student
Computer Science and Engineering,
Jayaraj Annapackiam CSI College of Engineering, Nazareth, India.
Smileincathrinwinsmilein@gmail.com

M. Jayashree
Student
Computer Science and Engineering,
Jayaraj Annapackiam CSI College of Engineering, Nazareth, India.
jeyashree4405@gmail.com

***Abstract* - Deepfake technology has rapidly evolved with the advancement of artificial intelligence and deep learning, making it increasingly difficult to distinguish between real and manipulated media. Deepfake videos and images can be used for misinformation, identity theft, and other malicious purposes. This project proposes a Deepfake Detection System that automatically identifies whether an image or video is real or fake using deep learning techniques.The system uses EfficientNet-B0 Convolutional Neural Network (CNN) for feature extraction and classification. The uploaded image or video is preprocessed by resizing, normalization, and noise simulation before being analyzed by the deep learning model. The trained model then predicts whether the content is real or fake along with a confidence score.**

The proposed system helps in preventing misinformation, improving digital security, and assisting media verification processes. It can be applied in social media platforms, journalism, and digital forensics to detect manipulated media effectively.

Keywords* - *Deepfake Detection, Artificial Intelligence, Deep Learning, Convolutional Neural Network, EfficientNet-B0, Image Processing, Video Analysis, Digital Forensics.

I. INTRODUCTION

In recent years, artificial intelligence has enabled the creation of highly realistic manipulated media known as deepfakes. These deepfakes are generated using deep learning techniques such as Generative Adversarial Networks (GANs), which can create realistic images and videos that appear authentic. While this technology has useful applications in entertainment and media production, it also poses serious threats when used maliciously.Deepfake media can be used to spread misinformation, manipulate public opinion, damage reputations, and create security risks. With the increasing availability of deepfake tools, detecting manipulated media has become a major challenge.

To address this issue, a Deepfake Detection System is proposed. The system uses deep learning models to analyze visual patterns and inconsistencies in images and videos to determine whether they are authentic or manipulated.

The system processes uploaded media through several stages including preprocessing, feature extraction, and classification. The EfficientNet-B0 model is used to extract important features from the input data, which are then used to classify the media as real or fake.

By automating the detection process, the system aims to improve the reliability of digital media and help prevent the spread of fake content.

II. LITERATURE REVIEW

Several researchers have worked on detecting deepfake media using machine learning and deep learning techniques.

Afchar et al. proposed a deep learning-based method known as MesoNet for detecting deepfake videos. The model analyzes facial inconsistencies and artifacts generated during manipulation. Although effective, it struggles with highly realistic deepfakes.

Rossler et al. introduced the FaceForensics++ dataset, which contains manipulated videos used to train deep learning models for deepfake detection. This dataset has significantly improved the performance of detection models.

Nguyen et al. developed a capsule network-based deepfake detection system that captures spatial relationships in images. The model improves detection accuracy but requires high computational resources.

Although existing systems achieve good performance, there is still a need for more efficient and scalable solutions. The proposed system addresses this by using the EfficientNet-B0 model, which provides high accuracy with lower computational cost.

The uploaded media undergoes preprocessing before being passed to the deep learning model. This stage includes resizing the images, normalization, and noise simulation to enhance feature detection.

Preprocessing helps improve the accuracy of the model by ensuring that all input data is in a consistent format.

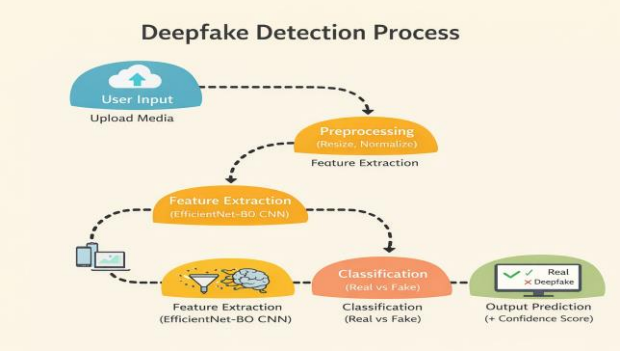
International Journal of Engineering Research & Technology (IJERT)
ISSN: 2278-0181
Vol. 14, Issue 03, ICCT - 2026

III. RESEARCH METHODOLOGY

The proposed deepfake detection system follows a deep learning-driven methodology designed to accurately classify media as real or manipulated. The approach integrates multiple stages including data acquisition, preprocessing, feature extraction, classification, and prediction.

The methodology is structured to ensure high accuracy, scalability, and robustness against modern deepfake generation techniques.

System Architecture



The architecture of the proposed system consists of the following pipeline:

User Input → Preprocessing → Feature Extraction → Classification → Output Prediction

The system is designed in a modular manner where each stage performs a specific function. The user uploads an image or video through the interface, which is then processed step-by-step to generate the final prediction.

The architecture ensures:

- Efficient handling of large media files
- High accuracy in detection
- Compatibility with real-world applications

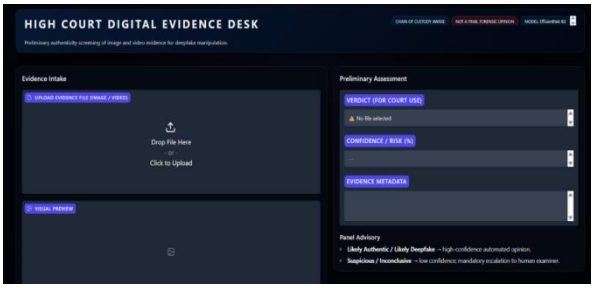
A. User Input Module

The User Input Module acts as the entry point of the system. It provides an intuitive interface where users can upload media files such as images and videos for analysis.

The system supports multiple formats including:

- Images: JPG, PNG, JPEG
- Videos: MP4, AVI

For video inputs, the system extracts key frames at regular intervals to analyze temporal inconsistencies.



This module ensures:

- Easy accessibility for users
- Seamless upload functionality
- Input validation to avoid corrupted or unsupported files

B. Preprocessing Module

The preprocessing stage is critical for improving model performance and ensuring uniformity across inputs.

Key preprocessing steps include:

Resizing: All images are resized to a fixed resolution (e.g., 224×224) to match model input requirements

Normalization: Pixel values are scaled to a standard range (0 to 1)

Frame Extraction (for videos): Important frames are extracted to reduce computational complexity

Noise Simulation: Helps the model generalize better by exposing it to variations

Face Detection (optional enhancement): Extracts facial regions for more focused analysis

This stage reduces noise, enhances important features, and ensures consistency in data representation.

C. Feature Extraction Module

Feature extraction is performed using the EfficientNet-B0 Convolutional Neural Network (CNN), which is known for its high performance and computational efficiency.

EfficientNet uses compound scaling, which balances:

- Network depth
- Width
- Resolution

The model extracts deep features such as:

- Facial landmark inconsistencies
- Texture distortions
- Blending artifacts
- Irregular lighting patterns

These features are crucial in identifying subtle manipulations present in deepfake media.

Advantages of using EfficientNet-B0:

- High accuracy with fewer parameters
- Faster training and inference

International Journal of Engineering Research & Technology (IJERT)
ISSN: 2278-0181
Vol. 14, Issue 03, ICCT - 2026

- Suitable for real-time applications

D. Classification Module

The classification module uses the features extracted by the CNN to determine whether the media is Real or Fake.

This stage typically includes:

- Fully connected (dense) layers
- Activation functions (ReLU, Softmax/Sigmoid)
- Dropout layers to prevent overfitting

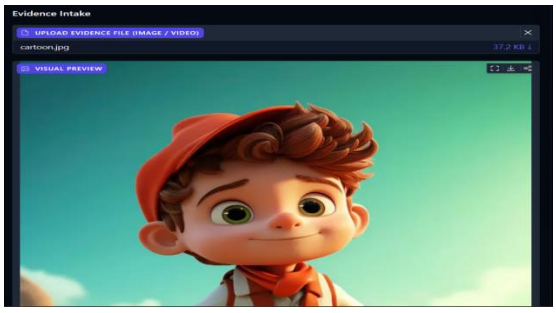
The model is trained on a labeled dataset containing:

- Authentic media samples
- Deepfake/manipulated media samples

The classification decision is based on learned patterns such as:

- Pixel-level inconsistencies
- Facial misalignments
- Temporal irregularities (in videos)

The output is a probability score indicating the likelihood of the media being fake.



E. Prediction Output Module

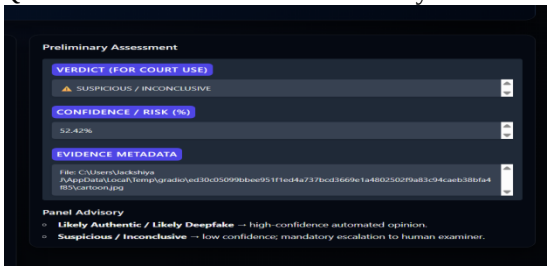
The final stage presents the results to the user in a clear and interpretable format.

The output includes:

- Prediction Label: Real / Deepfake
- Confidence Score: Probability value (e.g., 92% Fake)
- The system may also include:
- Visual indicators (green for real, red for fake)
- Confidence bars or graphs

This module ensures:

- Easy understanding for non-technical users
- Quick verification of media authenticity



IV. CONCLUSION

This project successfully demonstrates a deep learning-based approach for detecting deepfake media using advanced convolutional neural networks. By leveraging the EfficientNet-B0 architecture, the system effectively captures intricate visual patterns and inconsistencies that are difficult to detect manually.

The proposed system provides:

- High accuracy in classification
- Efficient processing of both images and videos
- Scalability for real-world deployment

With the increasing misuse of deepfake technology, this system plays a crucial role in:

- Preventing misinformation
- Enhancing cybersecurity
- Supporting digital forensics and media verification

Overall, the project highlights the importance of artificial intelligence in building trustworthy digital ecosystems and combating emerging threats in multimedia manipulation.

V. FUTURE WORK

The system can be further improved by incorporating advanced features and expanding its capabilities.

Future enhancements include:

A. Real-Time Detection

Implementing real-time deepfake detection for live video streams, which can be useful in video conferencing and surveillance systems.

B. Advanced Models

Integrating more powerful architectures such as:

- Vision Transformers (ViT)
- Hybrid CNN + LSTM models for temporal analysis

C. Larger and Diverse Datasets

Training the model on larger datasets like

- FaceForensics++
- Celeb-DF

This will improve generalization and robustness.

D. Browser Extension / API Integration

Deploying the system as:

- A browser extension
- A web API for social media platforms

This enables automatic detection of deepfake content online.

E. Explainable AI (XAI)

Adding explainability features such as:

- Heatmaps (Grad-CAM)
- Highlighting manipulated regions

International Journal of Engineering Research & Technology (IJERT)
ISSN: 2278-0181
Vol. 14, Issue 03, ICCT - 2026

This improves transparency and user trust.

F. Mobile Application

Developing a mobile app for real-time detection using smartphone cameras.

REFERENCES

[1] Afchar, D., et al., "MesoNet: A Compact Facial Video Forgery Detection Network," IEEE International Workshop on Information Forensics and Security, 2018.

[2] Rossler, A., et al., "FaceForensics++: Learning to Detect Manipulated Facial Images," IEEE International Conference on Computer Vision (ICCV), 2019.

[3] Nguyen, H., et al., "Capsule Networks for Deepfake Detection," IEEE International Conference on Multimedia and Expo, 2019.

[4] Chollet, F., Deep Learning with Python, Manning Publications, 2017.