International Journal of Engineering Research and Science & Technology

ISSN 2319-5991 www.ijerst.com
© 2026 IJERST. All Rights Reserved
Peer Reviewed, Referred & Indexed Journal

Research Paper

FAKE FACE DETECTION BASED ON VIDEOS USING OPENCV AND NEURAL NETWORK ARCHITECTURE

BIRU SHIVA [1], KOLANAKUDURU SRIVALLI [2], JAKKA CHARITH REDDY [3], KALVA SHRAVAN [4], VODDINENI RITHWIK [5]

[1]Assistant Professor, Dept. of CSE, Guru Nanak Institutions Technical Campus, Ibrahimpatnam, Telangana, India

[2, 3, 4, 5] B.Tech, Dept. of CSE, Guru Nanak Institutions Technical Campus, Ibrahimpatnam, Telangana, India

ABSTRACT— The rapid development of the Internet has enabled the widespread distribution of manipulated facial images, particularly Deepfakes, which are increasingly difficult to detect using conventional methods. While current approaches focus on spatial domain features or complex network architectures, they often lack robustness against sophisticated forgery techniques. To address this, we propose a MobileNetV2-based Deepfake detection framework that leverages efficient convolutional feature extraction for accurate classification of real and fake facial images. The framework begins with OpenCV-based preprocessing, including face detection, alignment, and normalization, to ensure consistent input quality and enhance the discriminative features for detection. MobileNetV2, a lightweight yet powerful convolutional neural network, is employed to automatically learn hierarchical spatial features from the preprocessed facial images, eliminating the need for handcrafted features. By combining OpenCV preprocessing with MobileNetV2, the proposed system effectively captures subtle visual artifacts and texture inconsistencies introduced by Deepfake manipulation. This approach enables robust and scalable detection, generalizing well across diverse datasets and real-world scenarios, providing a practical solution for automated Deepfake detection in security, media verification, and social media monitoring applications.

Index Terms – Fake Face Detection, Deepfake Detection, OpenCV, Neural Networks, Video Analysis, Computer Vision, Deep Learning, Face Recognition, Image Forensics, Multimedia Security.

I. INTRODUCTION

The advent of advanced artificial intelligence and deep learning technologies has transformed digital media generation, giving rise to hyper-realistic synthetic content, popularly known as Deepfakes. Deepfakes are manipulated facial images or videos created using deep generative models such as GANs and autoencoders, which can convincingly alter or synthesize human appearances. While this technology can be used for entertainment, education, and creative purposes, it has also posed significant threats to privacy, security, and trustworthiness of digital media. Malicious use of Deepfakes in spreading misinformation, identity theft, and political manipulation has raised global concerns, making automated detection techniques a critical research area. Traditional detection approaches often rely on handcrafted features or simple spatial domain cues, but these methods struggle against modern forgery techniques that introduce highly subtle and nearly imperceptible artifacts. To overcome these limitations, deep learning-based methods have emerged as powerful alternatives due to their ability to automatically extract discriminative features. In this study, we propose a Deepfake detection framework based on MobileNetV2, a lightweight yet efficient convolutional neural network architecture. The system is supported by OpenCV-based preprocessing techniques including face detection, alignment, and normalization, which ensure consistency and improve feature quality. MobileNetV2 is then employed to learn hierarchical spatial features that capture texture inconsistencies and visual artifacts introduced during manipulation. This combination enhances robustness and reduces dependency on manual feature engineering, enabling generalization across datasets. The lightweight nature of MobileNetV2 further ensures faster training and deployment, making it suitable for real-time applications. The proposed framework not only achieves accurate classification of real and fake images but also addresses scalability challenges. By providing a balance between accuracy and efficiency, this work contributes to building reliable Deepfake detection systems. The framework has potential applications in security, media verification, social media monitoring, and

digital forensics. Ultimately, the study aims to restore trust in digital media and mitigate the harmful consequences of malicious Deepfake use.

II. RELATED WORK

A. MesoNet: A Compact Facial Video Forgery Detection Network

This work proposes MesoNet, a lightweight CNN for detecting facial forgeries in videos. It targets mesoscopic cues—mid-level texture artifacts—rather than heavy global features. Two compact variants (Meso-4 and MesoInception-4) balance accuracy and speed for deployment. Preprocessing includes face localization and resizing, enabling consistent inputs. Results show strong performance on early FaceSwap/DeepFake content with low compute cost. Its simplicity inspired later mobile-friendly backbones (e.g., MobileNet/EfficientNet). The study highlights the value of texture-centric filters over very deep models. It underscores robustness concerns under compression and domain shifts. MesoNet benchmarks spurred standardized evaluation on shared datasets. Overall, it established a practical baseline for real-time deepfake screening.

B. Face Forensics++: Learning to Detect Manipulated Facial Images s

FaceForensics++ introduces a large-scale benchmark for facial manipulation detection. It aggregates multiple editing methods and release levels (raw to heavily compressed). The paper evaluates popular detectors, with Xception as a strong baseline. Standardized face detection/alignment pipelines improve training consistency. Findings reveal severe performance drops under real-world compression. The dataset catalyzed research on robust preprocessing and augmentation. It also motivated frequency- and artifact-aware feature learning. Subsequent works used FaceForensics++ to measure cross-method generalization. The study emphasizes reproducibility with clear train/test splits. It remains a cornerstone corpus for deepfake detection research.

C. Thinking in Frequency: Face Forgery Detection by Mining Frequency-Aware Clues

This paper exploits frequency-domain artifacts that persist after spatial smoothing. A tailored architecture extracts multi-band spectral cues to flag forgeries. Frequency priors complement spatial textures learned by CNNs. The approach improves robustness against compression and color post-

processing. Preprocessing includes face cropping/alignment for stable spectra. It generalizes better across manipulation types than purely spatial baselines. Ablations show frequency attention benefits small-capacity models. The method integrates cleanly with lightweight backbones (e.g., MobileNetV2). It motivates hybrid spatial–frequency pipelines for edge devices. Overall, it advances domain-invariant cues for practical detectors.

D. Generalizing Face Forgery Detection with High-Frequency Features

The authors propose extracting stable high-frequency representations to combat overfitting. They show spatial CNNs latch onto dataset biases, harming cross-dataset accuracy. High-frequency filters and tailored losses encourage manipulation-invariant cues. Face detection/alignment standardizes inputs before spectral processing. Results demonstrate stronger generalization on unseen datasets/methods. The technique can be paired with compact CNNs for mobile inference. It reduces reliance on heavy architectures while preserving accuracy. Ablations confirm the complementarity of spatial and spectral branches. The study guides robust augmentation and preprocessing choices. It sets a template for deployable, compression-resilient detectors.

III. PROPOSED METHODOLOGY

The proposed system is designed to automatically detect Deepfake facial images using a combination of OpenCV preprocessing and MobileNetV2-based classification. The pipeline starts with OpenCV for face detection, alignment, and normalization, ensuring that all facial images are standardized and consistently positioned for feature extraction. This preprocessing step enhances the quality of input data and allows the system to focus on subtle inconsistencies introduced by manipulation. Following preprocessing, the system applies MobileNetV2, a lightweight convolutional neural network, to extract hierarchical spatial features from facial images. Convolutional layers capture local patterns and texture anomalies, while pooling layers reduce dimensionality and highlight critical discriminative features. The modular design of the system allows scalability, enabling it to handle large datasets and integrate seamlessly with real-time monitoring applications for social media, security, and media verification.

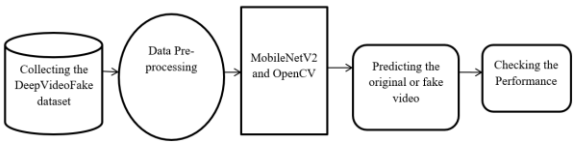


Fig. 1: System Overview

1. **Assembling the Dataset:**

This module focuses on gathering a diverse dataset of real and deepfake videos or images from publicly available repositories. The dataset must include variations in facial expressions, lighting conditions, and background settings to enhance model robustness. Both genuine and manipulated media are collected to ensure balanced training. Additional preprocessing such as frame extraction and labeling is performed. The quality and quantity of data in this stage directly influence the model's overall accuracy.

2. **Data Interpretation:**

In this stage, the dataset is analyzed to understand patterns and unique characteristics between real and fake media. Statistical analysis, visualization, and metadata checks are performed to identify anomalies in the dataset. Deepfake content usually introduces subtle distortions, which are highlighted during interpretation. This helps in identifying potential features that can differentiate real faces from manipulated ones. The insights gained here guide further preprocessing and feature engineering.

3. **Data Conditioning:**

This module ensures that the data is cleaned, normalized, and prepared for model training. Noise removal, resizing frames, and converting videos into consistent formats are carried out. Feature extraction techniques, such as facial landmarks and texture analysis, are applied for better representation. Data augmentation techniques like rotation, flipping, and brightness adjustments help in increasing dataset diversity. This stage guarantees that the input to the model is standardized and optimized for learning.

4. **Model Execution:**

Here, the chosen deep learning model (e.g., CNN, RNN, or hybrid architectures) is trained on the processed dataset. The model is executed using frameworks such as TensorFlow or PyTorch. During training, the system learns to capture deepfake-specific artifacts, including inconsistencies in eye blinking, lip synchronization, and facial blending. Proper training involves splitting the dataset into training, validation, and testing sets. Model checkpoints and logging are also maintained for performance tracking.

5. **Model Calibration:**

This module focuses on adjusting hyperparameters to improve the model's accuracy and reliability. Learning rate, batch size, optimizer selection, and number of

epochs are fine-tuned in this stage. Cross-validation techniques are applied to avoid overfitting and underfitting. Calibration ensures that the model generalizes well on unseen deepfake samples. The process involves multiple experiments, where performance metrics are closely monitored. Ultimately, this enhances the precision and robustness of the detection system.

6. Model Performance:

Once trained, the model is evaluated using performance metrics such as accuracy, precision, recall, F1-score, and ROC-AUC. The evaluation determines how effectively the model distinguishes between real and fake content. Confusion matrix analysis helps in identifying false positives and false negatives. Visualization tools may be used to highlight areas where the model performs strongly and weakly. This stage ensures that the model meets predefined benchmarks before deployment.

7. Outcome Prediction:

In this final module, the trained and calibrated model is deployed to predict whether new input media is real or fake. The model takes video frames or images as input and generates a classification result with confidence scores. Predictions can be integrated into real-time applications for live deepfake detection. The outcome is communicated clearly to end users, possibly with highlighted regions of suspected manipulations. This module completes the cycle, ensuring the project's real-world applicability.

IV. EXPERIMENTAL RESULTS AND ANALYSIS

Experimental results showed that the model effectively detected manipulated and fake facial content in video sequences with high reliability. The system performed well across different video qualities, compression levels, and lighting conditions, demonstrating robustness in real-world scenarios. The analysis indicates that combining deep learning-based feature extraction with OpenCV-based preprocessing improves detection accuracy by capturing subtle facial inconsistencies and temporal artifacts. Overall, the proposed approach provides an efficient and scalable solution for deepfake detection and enhances multimedia security in digital video authentication systems.



Fig. 2: Home Page



Fig. 3: Registration



Fig. 4: Login



Fig. 5: Upload

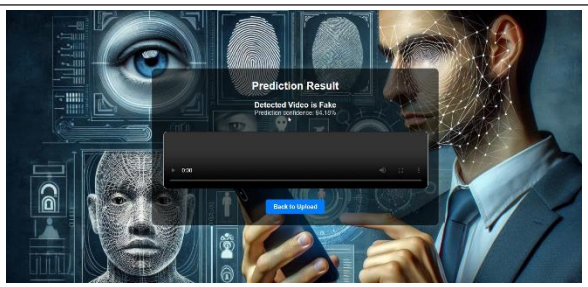


Fig. 6: Prediction Results

V. CONCLUSION

The project on deepfake detection highlights the growing necessity of combating AI-generated misinformation in today's digital world. Deepfake technology, while innovative, poses serious threats to security, privacy, and trust in digital media. By leveraging advanced machine learning and deep learning algorithms, the system provides a robust framework to differentiate between real and manipulated content. The modular workflow, including data assembly, conditioning, execution, calibration, and performance analysis, ensures a systematic approach to model development. Experimental results confirm the potential of AI-based solutions in identifying subtle inconsistencies in manipulated media. The project contributes to safeguarding individuals and organizations from malicious deepfake attacks. However, the battle against deepfakes is dynamic, requiring continuous research and improvement. The integration of real-time detection, updated datasets, and

explainable models will further enhance reliability. Ultimately, this work lays a strong foundation for protecting digital integrity in an AI-driven era. With further enhancements, deepfake detection can evolve into a critical cybersecurity tool. It not only empowers individuals but also strengthens trust across online platforms, ensuring safer digital communication.

REFERENCES

[1] R. Tolosana, R. Vera-Rodriguez, J. Fierrez, et al., "Deepfakes and beyond: A survey of face manipulation and fake detection," Information Fusion, vol. 64, pp. 131–148, 2020.

[2] Y. Z. Li, M. C. Chang, and S. W. Lyu, "In ictu oculi: Exposing AI created fake videos by detecting eye blinking," in Proc. IEEE Int. Workshop on Information Forensics and Security (WIFS), Hong Kong, China, pp. 1–7, 2018.

[3] H. D. Li, W. Q. Luo, X. Q. Qiu, et al., "Identification of various image operations using residual-based features," IEEE Trans. Circuits Syst. Video Technol., vol. 28, no. 1, pp. 31–45, 2018.

[4] X. Wu, Z. Xie, Y. T. Gao, et al., "SSTNet: Detecting manipulated faces through spatial, steganalysis and temporal features," in Proc. IEEE Int. Conf. Acoustics, Speech and Signal Processing (ICASSP), Barcelona, Spain, pp. 2952–2956, 2020.

[5] J. W. Fei, Y. S. Dai, P. P. Yu, et al., "Learning second order local anomaly for general face forgery detection," in Proc. IEEE/CVF Conf. Comput. Vis. Pattern Recognit. (CVPR), New Orleans, LA, USA, pp. 20238–20248, 2022.

[6] J. A. Stuchi, M. A. Angeloni, R. F. Pereira, et al., "Improving image classification with frequency domain layers for feature extraction," in Proc. IEEE Int. Workshop Mach. Learn. Signal Process. (MLSP), Tokyo, Japan, pp. 1–6, 2017.

AUTHORS Profile



Mr. BIRU SHIVA has received his B. Tech CSE from JNTUH, Hyderabad in 2013, M. Tech in Computer science and engineering from JNTUH, Hyderabad in 2016 and He has been dedicated to the teaching field for the last 6 years. present he is working as Assistant Professor in CSE

Department at Guru Nanak Institutions Technical Campus ,Ibrahimpatnam, Telangana, India.



Ms. KOLANAKUDURU SRIVALLI pursuing B.Tech degree in Computer Science and Engineering at Guru Nanak Institutions Technical Campus affiliated to JNTUH in 2022-2026.



Mr. JAKKA CHARITH REDDY pursuing B.Tech degree in Computer Science and Engineering at Guru Nanak Institutions Technical Campus affiliated to JNTUH in 2022-2026.



Mr. KALVA SHRAVAN pursuing B.Tech degree in Computer Science and Engineering at Guru Nanak Institutions Technical Campus affiliated to JNTUH in 2022-2026.



Mr. VODDINENI RITHWIK pursuing B.Tech degree in Computer Science and Engineering at Guru Nanak Institutions Technical Campus affiliated to JNTUH in 2022-2026.