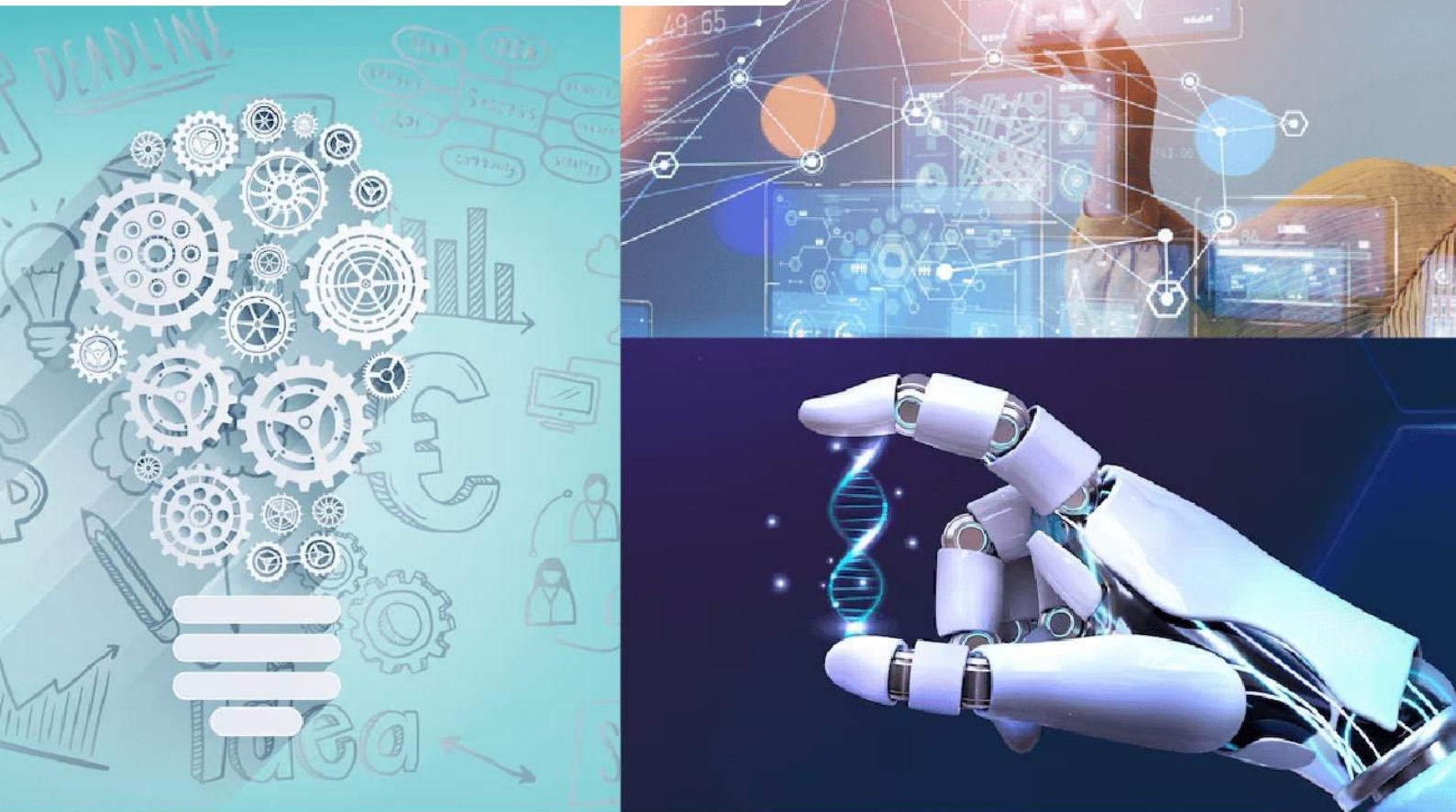




International Journal of Advanced Research in Education and Technology (IJARETY)

Volume 13, Issue 2, March - April 2026

Impact Factor: 8.152



INTERNATIONAL
STANDARD
SERIAL
NUMBER
INDIA



Fake Face Detection Based on Videos using OpenCV and Neural Network Architecture

A. Laksmipathi rao¹, A. Sai Ayush², A. Lokesh³, A. Rahul⁴

Assistant Professor, Department of CSE, Guru Nanak Institute of Technology, Hyderabad, Telangana, India ¹

Students, Department of CSE, Guru Nanak Institute of Technology, Hyderabad, Telangana, India ^{2,3,4}

ABSTRACT: The rapid development of the Internet has led to the large-scale 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 advanced manipulation techniques. To address this, we propose a MobileNetV2-based Deepfake detection framework that utilizes 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 used 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.

KEYWORDS: Deepfake detection, manipulated facial images, MobileNetV2, OpenCV preprocessing, face detection, face alignment, normalization, convolutional neural network, spatial features, feature extraction, classification, visual artifacts, texture inconsistencies, security, media verification, social media monitoring

I. INTRODUCTION

Not long ago, spotting a fake video was fairly easy — blurry edges, unnatural skin tones, or slightly off lip movements served as obvious clues for detecting fake videos. Today, that is no longer the case. Thanks to deep learning tools like Generative Adversarial Networks (GANs), anyone with enough computing power can create a video of a person saying or doing something they never did, and it can look completely genuine to the untrained eye. These are what researchers call Deepfakes, and they have grown into a serious problem — used for spreading false information, harassing individuals, and even influencing elections.

Traditional detection tools were built for an earlier era of digital manipulation. They looked for simple spatial clues in images — things like abnormal blinking or irregular skin patterns — and worked effectively well on lower-quality fake images or videos. But as Deepfake technology improved, these methods started to fall behind. The manipulated content became too good, and the old approaches could not keep up, especially when videos were compressed or resized after creation. This paper proposes a detection framework that aims to achieve an effective balance between accuracy and efficiency to catch well-made Deepfakes, but lightweight enough to run in real-world conditions. We use MobileNetV2 as the core detection network because it is fast, does not require expensive hardware, and still produces strong results. It is combined with an OpenCV preprocessing step that normalizes each face before analysis, the system can pick up on the tiny inconsistencies that Deepfake generation leaves behind — even when those signs are difficult for humans to see. The objective is to develop a tool that can be practically deployed on social media platforms, news verification systems, or security applications.

II. LITERATURE SURVEY

Title: MesoNet: A Compact Facial Video Forgery Detection Network

Authors: D. Afchar, V. Nozick, J. Yamagishi, and I. Echizen

Year: 2018

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

Title: Face Forensics++: Learning to Detect Manipulated Facial Images

Authors: A. Rössler, D. Cozzolino, L. Verdoliva, C. Riess, J. Thies, and M. Nießner

Year: 2019

Description: FaceForensics++ presents a large-scale benchmark for facial manipulation detection. It combines multiple manipulation techniques and release levels (raw to heavily compressed). The paper assesses popular detectors, with Xception as a strong baseline. Standardized face detection/alignment pipelines improve training consistency. Results indicate severe performance drops under real-world compression. The dataset promoted research on robust preprocessing and augmentation. It also motivated frequency- and artifact-aware feature learning. Later studies used FaceForensics++ to measure cross-method generalization. The study highlights reproducibility with clear train/test splits. It remains a foundational dataset for deepfake detection research.

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

Authors: Y. Qian, G. Yin, L. Sheng, Z. Chen, and J. Shao

Year: 2020

Description: This paper utilizes frequency-domain artifacts that remain after spatial filtering. A specialized architecture extracts multi-band spectral cues to detect manipulated content. Frequency-based features enhance spatial textures learned by CNNs. The approach improves robustness against compression and color post-processing. Preprocessing includes face cropping/alignment for consistent spectral features. It generalizes better across manipulation types than spatial-only baseline methods. Ablation studies indicate frequency attention improves small-capacity models. The method integrates effectively with lightweight backbones (e.g., MobileNetV2). It encourages hybrid spatial–frequency pipelines for edge devices. Overall, it enhances generalizable features for real-world detection systems.

Title: Generalizing Face Forgery Detection with High-Frequency Features

Authors: Y. Luo, Y. Zhang, J. Yan, W. Liu, and D. Wang

Year: 2021

Description: The authors present the extraction of stable high-frequency representations to reduce overfitting. They show spatial CNNs tend to rely on dataset biases, reducing cross-dataset accuracy. High-frequency filters and some loss functions promote manipulation-invariant cues. Face detection/alignment standardizes inputs before spectral processing. Results show improved generalization on unseen datasets/methods. The technique can be paired with compact CNNs for mobile inference. It reduces dependency on heavy architectures while maintaining accuracy. Ablations confirm the complementary nature of spatial and spectral branches. The study guides robust augmentation and preprocessing choices. It sets a template for deployable, compression-resilient detectors.

Title: Protecting Celebrities from Deepfake with Identity Consistency Transformer

Authors: X. Dong, J. Bao, D. Chen, N. Yu, and D. Chen

Year: 2022

Description: This work presents an identity-consistency transformer for manipulated content detection. It analyzes cross-frame identity coherence, revealing subtle temporal inconsistencies. The process begins with face alignment and quality filtering of frames. Transformer attention captures long-range dependencies beyond CNN local context fields. Improvements appear on high-quality, low-artifact deepfake samples where cues are subtle. It enhances lightweight CNN backbones through late fusion or knowledge distillation. Temporal modeling improves resilience to compression and occlusions. The method highlights benefits of semantics (who) plus artifacts (how). It informs hybrid designs combining MobileNetV2 spatial features with temporal cues. Overall, it enhances identity-aware, real-world deepfake detection.

III. EXISTING SYSTEM

The existing Deepfake detection systems largely rely on spatial domain analysis using Convolutional Neural Networks (CNNs) and sometimes utilizes hybrid architectures that utilize transformer-based architecture. These models aim to detect artifacts or inconsistencies in facial regions, such as unnatural blinking, distorted facial expressions, or inconsistencies in texture. Some methods attempt to improve feature extraction by converting images into frequency or residual features, while others use attention mechanisms or multi-scale feature extraction. Some approaches also consider metadata and image quality inconsistencies for improved classification performance. However, these existing systems face challenges when Deepfake generation techniques continue to evolve and produce more photorealistic content. Spatial-only methods may fail to generalize across different Deepfake datasets and often suffer from overfitting. Transformer-based models, while powerful, are computationally intensive and require large datasets to train effectively. These systems may also face challenges in real-world scenarios with low-resolution inputs, occlusions, or background clutter, thus highlighting the need for more robust and adaptable approaches.

EXISTING SYSTEM DISADVANTAGES

- Limited to spatial features
- Poor generalization.
- High computational cost
- Insensitive to subtle artifacts
- Static learning

IV. PROPOSED SYSTEM

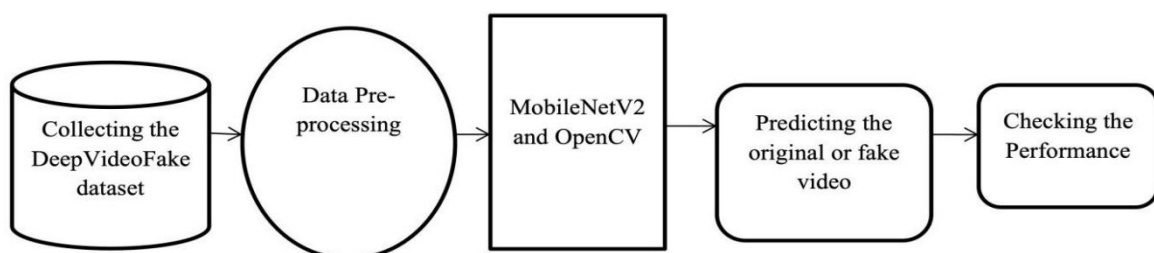
The proposed system is designed to automatically detect Deepfake facial images using a combination of OpenCV preprocessing and MobileNetV2-based classification. The process starts with OpenCV for face detection, alignment, and normalization, ensuring that all facial images are normalized and consistently 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 inconsistencies, while pooling layers reduce dimensionality and retain the most relevant information. The modular design of the system allows scalability, enabling it to handle large datasets and integrate effectively with real-time monitoring applications for social media, security, and media verification.

PROPOSED SYSTEM ADVANTAGES:

- Adaptive learning through Adversarial Training
- Enhanced via High-Frequency analysis
- Improved accuracy and robustness
- Effective use of attention mechanisms
- Better generalization to unseen Data

V. SYSTEM ARCHITECTURE

Fig 1: System Architecture



The diagram above shows how the system flows from raw video input through to a final real-or-fake decision. Individual frames are extracted from the uploaded video and passed into the OpenCV face detection module. Once a face is found, it is aligned and normalised before being sent to MobileNetV2. The network produces a confidence

score for each frame, and those scores are combined across all frames to give a final result for the whole video. If more than half the frames are flagged as fake, the system classifies the video as fake.

VI. METHODOLOGIES

MODULES NAME:

- Assembling the Dataset
- Data Interpretation
- Data Conditioning
- Model Execution
- Model Calibration
- Model Performance
- Outcome Prediction

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 blurring. 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.

VII. ALGORITHM USED

EXISTING TECHNIQUE

The existing Deepfake detection algorithms largely rely on Convolutional Neural Networks (CNNs) to extract spatial features from facial images. These CNN-based models are trained to recognize visual inconsistencies such as unnatural facial expressions, improper blending effects, and inconsistent lighting that often appear in manipulated content. In recent developments, CNNs are often combined with transformers, which are capable of capturing long-range dependencies and context-aware features within the image. These hybrid models use self-attention mechanisms to focus on important facial regions, improving the performance over traditional CNN-only approaches. Despite their effectiveness, these algorithms face several limitations. First, they often require large datasets for training, and their performance drops when applied to Deepfakes from unseen sources. Additionally, these models mainly focus on spatial features, neglecting other informative domains like frequency or noise patterns. Their dependence on manually designed architectures and static training sets also limits their ability to adapt when new types of Deepfake manipulation techniques appear. They are prone to high-quality, realistic manipulated contents that do not display clear visual artifacts.

PROPOSED ALGORITHM

The proposed algorithm first uses OpenCV to detect and extract facial regions from raw images, performing alignment and normalization to standardize inputs. These steps ensure that features learned by the neural network are consistent and focused on relevant facial areas, improving detection accuracy without the need for manually designed feature extraction. Next, the MobileNetV2 architecture processes the preprocessed facial images. Its convolutional layers automatically learn hierarchical spatial features that show subtle artifacts introduced by Deepfake generation. The network is trained using supervised learning, reducing classification errors between real and fake images. By combining OpenCV preprocessing with MobileNetV2 feature extraction, the algorithm gives a robust, scalable, and efficient approach for automated Deepfake detection in real-world applications.

VIII. EXPERIMENTAL RESULTS

Home Page: The home page is the first thing users see when they open the application. It introduces the project and gives two clear options — register as a new user or log in with an existing account. The interface is well-organized, and its purpose is immediately clear, as part of a well-planned effort to make the system accessible to non-technical users.



Fig 2: Home Page

User Registration Page: New users can create an account by entering a username and password. The registration step is straightforward — The form ensures user inputs before saving and confirms successful registration. Linking user data to user accounts allows the system to maintain upload history for auditing purposes.



Fig 3: User Registration Page

User Login Page: Returning users to log in through this screen. The system checks the user details against stored records before giving access. This layer of authentication make sure that only verified users can run the detection process.

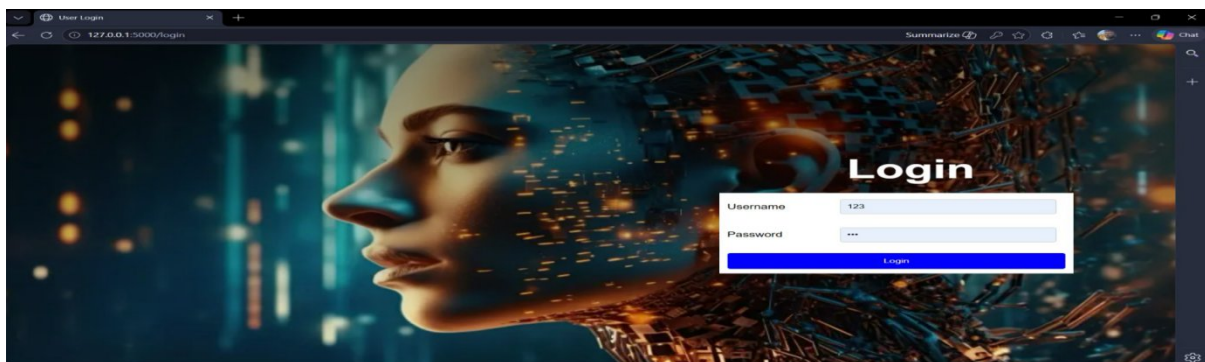


Fig 4: User Login Page

Upload File Page: This is where the main action happens. After logging in, users goes to the upload page, where they can select a video file from their device. Once a file is chosen, clicking the Predict button starts the detection process — OpenCV extracts and preprocesses the faces, MobileNetV2 analyses them, and the result appears on screen.

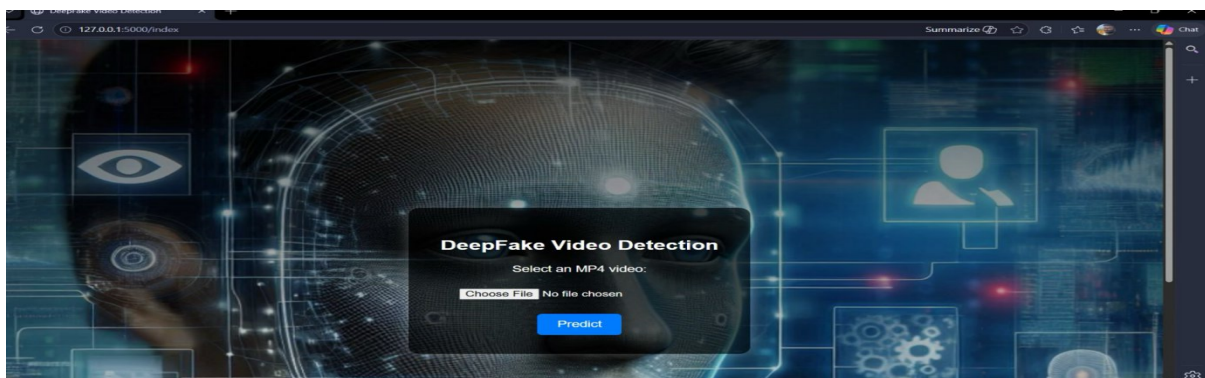


Fig 5: Upload File Page

Celeb_Synthesis Folder: To evaluate the system, we utilized the Celeb-DF benchmark dataset, which organizes video clips into two categories. The celeb_synthesis folder contains AI-generated video samples — these are the fakes the system needs to correctly detect. A well-organized test set like this enables consistent and reliable evaluation of model performance.



Fig 6: Celeb_Synthesis Folder

Result — Synthetic Video Detected: When a video from the celeb_synthesis folder is chosen, the system correctly identifies it as fake. The confidence score reflects some uncertainty, which is expected given how photo-realistic modern Deepfakes can be, but the classification itself is accurate. This result shows that the system is capable of catching high-quality manipulations, not just low-effort ones.

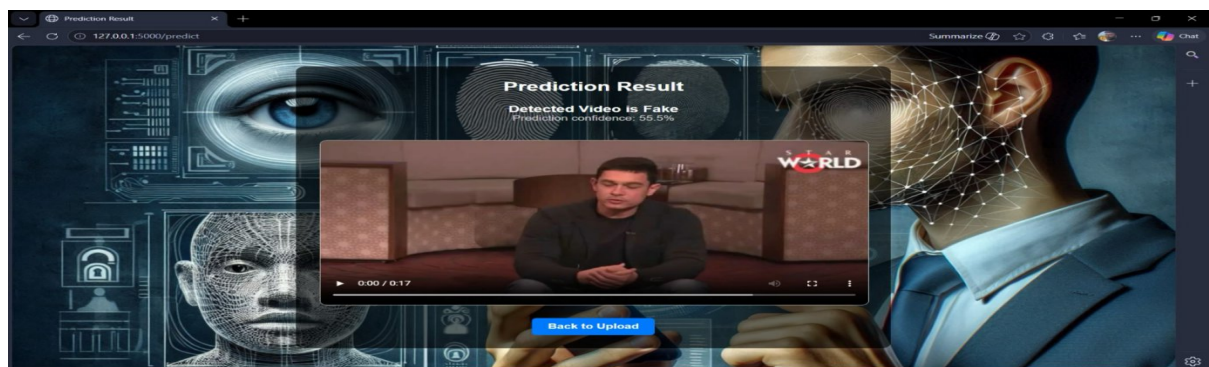


Fig 7: Result — Synthetic Video Detected

Celeb_Real Folder: The celeb_real folder holds genuine and real video clips. The system should classify these as real with high confidence. Including both classes in the test set ensures we are measuring balanced performance rather than just checking how well the system catches fakes.



Fig 8: Celeb_Real Folder

Result — Genuine Video Confirmed: Videos from the celeb_real folder are correctly classified as real, and with a very high confidence compared to the synthetic samples. This makes sense — real videos have natural, consistent

facial patterns that are easier for the model to confirm as real. The strong confidence on real samples, combined with correct detection on fake ones, shows the system is working as intended on both sides of the classification.

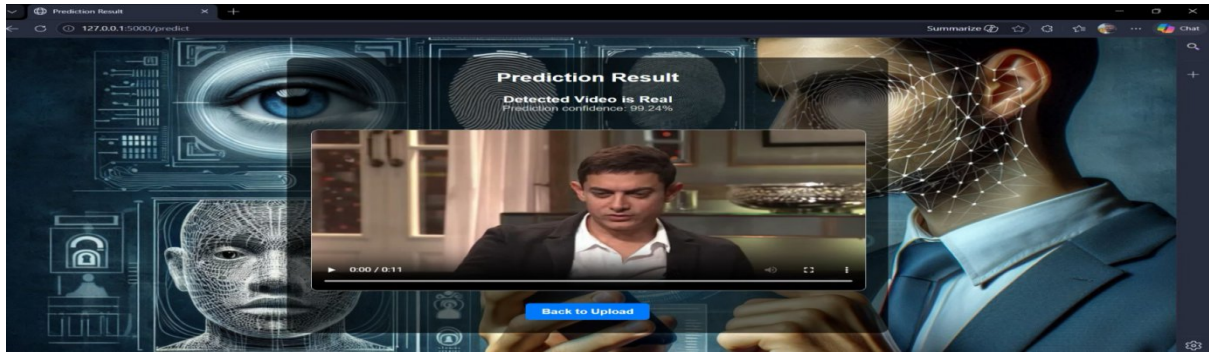


Fig 9: Result — Genuine Video Confirmed

IX. CONCLUSION

The proposed deepfake detection system shows the increasing need to address AI-generated misinformation in today's digital world. Deepfake technology, while innovative, poses serious threats to security, privacy, and trust in digital media. By using machine learning and deep learning techniques, the system gives a reliable approach for distinguishing between real and manipulated media. The structured workflow, including data assembly, conditioning, execution, calibration, and performance analysis, ensures a systematic approach to model development. Results from the experiments shows that AI-based approaches can successfully identify subtle irregularities in manipulated content. This project contributes toward protecting individuals and organizations from potential deepfake attacks. However, the fight against deepfakes continues to evolve, requiring continuous research and improvement. The integration of real-time detection, updated datasets, and explainable models will further improves reliability. Ultimately, this work provides a solid foundation for protecting digital integrity in an AI-driven era. With further enhancements, deepfake detection can evolve into a critical cybersecurity tool.

X. FUTURE ENHANCEMENT

Future enhancements may include the integration of multimodal analysis, where audio, visual, and textual information are combined and analyzed to improve detection accuracy. Real-time detection mechanisms can be implemented into social media platforms to automatically detect or block harmful content. Collaboration with government and media organizations can improve or strengthen regulatory frameworks. Continuous dataset updates will help the model adapt to new types of deepfakes. Adversarial training can improve robustness against evolving manipulation techniques.

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.
- [7] J. Fridrich and J. Kodovsky, "Rich models for steganalysis of digital images," *IEEE Trans. Inf. Forensics Security*, vol. 7, no. 3, pp. 868–882, 2012.

- [8] F. Matern, C. Riess, and M. Stamminger, “Exploiting visual artifacts to expose deepfakes and face manipulations,” in Proc. IEEE Winter Conf. Appl. Comput. Vis. Workshops (WACVW), Waikoloa, HI, USA, pp. 83–92, 2019.
- [9] A. Rössler, D. Cozzolino, L. Verdoliva, et al., “Faceforensics++: Learning to detect manipulated facial images,” in Proc. IEEE/CVF Int. Conf. Comput. Vis. (ICCV), Seoul, South Korea, pp. 1–11, 2019.
- [10] H. Dang, F. Liu, J. Stehouwer, et al., “On the detection of digital face manipulation,” in Proc. IEEE/CVF Conf. Comput. Vis. Pattern Recognit. (CVPR), Seattle, WA, USA, pp. 5780–5789, 2020.
- [11] H. Q. Zhao, T. Y. Wei, W. B. Zhou, et al., “Multi-attentional deepfake detection,” in Proc. IEEE/CVF Conf. Comput. Vis. Pattern Recognit. (CVPR), Nashville, TN, USA, pp. 2185–2194, 2021.
- [12] L. Chen, Y. Zhang, Y. B. Song, et al., “Self-supervised learning of adversarial example: Towards good generalizations for deepfake detection,” in Proc. IEEE/CVF Conf. Comput. Vis. Pattern Recognit. (CVPR), New Orleans, LA, USA, pp. 18689–18698, 2022.
- [13] X. Y. Dong, J. M. Bao, D. D. Chen, et al., “Protecting celebrities from deepfake with identity consistency transformer,” in Proc. IEEE/CVF Conf. Comput. Vis. Pattern Recognit. (CVPR), New Orleans, LA, USA, pp. 9458–9468, 2022.
- [14] D. Cozzolino, G. Poggi, and L. Verdoliva, “Splicebuster: A new blind image splicing detector,” in Proc. IEEE Int. Workshop on Information Forensics and Security (WIFS), Rome, Italy, pp. 1–6, 2015.
- [15] Y. Y. Qian, G. J. Yin, L. Sheng, et al., “Thinking in frequency: Face forgery detection by mining frequency-aware clues,” in Proc. Eur. Conf. Comput. Vis. (ECCV), Glasgow, UK, pp. 86–103, 2020.
- [16] Y. C. Luo, Y. Zhang, J. C. Yan, et al., “Generalizing face forgery detection with high-frequency features,” in Proc. IEEE/CVF Conf. Comput. Vis. Pattern Recognit. (CVPR), Nashville, TN, USA, pp. 16312–16321, 2021.
- [17] A. Vaswani, N. Shazeer, N. Parmar, et al., “Attention is all you need,” in Advances in Neural Information Processing Systems (NeurIPS), Long Beach, CA, USA, pp. 6000–6010, 2017.
- [18] B. V. Salim, Chyntia, J. O. Indrawan, et al., “Face shape classification using swin transformer model,” *Procedia Comput. Sci.*, vol. 227, pp. 557–562, 2023.
- [19] Y. Z. Li, X. Yang, P. Sun, et al., “Celeb-DF: A large-scale challenging dataset for DeepFake forensics,” in Proc. IEEE/CVF Conf. Comput. Vis. Pattern Recognit. (CVPR), Seattle, WA, USA, pp. 3204–3213, 2020.
- [20] B. Dolhansky, J. Bitton, B. Pflaum, et al., “The DeepFake detection challenge (DFDC) dataset,” arXiv preprint, arXiv:2006.07397, 2020.

International Journal of Advanced Research in Education and Technology

ISSN: 2394-2975

Impact Factor: 8.152