IARJSET ISSN (O) 2393-8021, ISSN (P) 2394-1588

International Advanced Research Journal in Science, Engineering and Technology

Impact Factor 8.066 ⁜ Peer-reviewed & Refereed journal ⁜ Vol. 12, Issue 5, May 2025

DOI: 10.17148/IARJSET.2025.125212

Detection Of Manipulated Media With AI

Vijay Kumar M S[1], Poornima N H[2], Priya M[3], Namratha S[4], Sneha H L[5]

Asst. Professor, Department of Information Science & Engineering, Maharaja Institute of Technology Mysore, Karnataka, India[1]

Student, Department of Information Science & Engineering, Maharaja Institute of Technology Mysore, Karnataka, India[2-5]

Abstract: Detection of Manipulated Media with AI is a system designed to counter the growing issue of false or altered multimedia content across digital platforms. In recent years, the spread of deepfakes and edited media has posed serious challenges to authenticity and public trust. This project makes use of artificial intelligence, particularly deep learning techniques like convolutional neural networks (CNNs), to spot signs of tampering in audio, video, and image files. It analyzes inconsistencies and manipulation patterns that are often invisible to the human eye. A simple user interface allows users to upload content for verification, offering real-time classification as authentic or altered. The solution is scalable, efficient, and plays a vital role in media verification, especially for journalists, legal investigators, and content moderators. It supports the fight against misinformation and promotes integrity in digital communication.

Keywords: Fake Media, Deepfake Detection, AI, CNN, Media Authentication, Digital Integrity, Image Forensics, Video Verification, Audio Analysis

I. INTRODUCTION

Our Humanizer is intended to improve your writing by giving it a more organic tone and flow. Our technology simplifies the process, enabling you to concentrate on your ideas rather than becoming bogged down by wordiness, whether you're writing an essay, report, or email. You can turn complicated sentences into clear, interesting content with a few clicks. This improves your writing and streamlines your workflow while also saving you a significant amount of time. Try it out and embrace a more straightforward writing experience by clicking the Humanize button and letting our tools handle the laborious tasks!

DeepFake videos are extremely realistic fake content that manipulates faces, voices, and other features thanks to the development of artificial intelligence. Although this technology has useful uses in education and entertainment, it also poses major risks, such as fraud, identity theft, and the spread of false information. Maintaining the legitimacy of digital media and public confidence requires the detection of DeepFakes. The problem of detecting DeepFake videos is addressed by this project, "AI-Powered DeepFake Defense." The system analyzes videos frame by frame, looks for manipulations or inconsistencies in each frame, and outputs a final result that indicates whether the video is authentic or not. Advanced machine learning and computer vision techniques ensure high accuracy, even in difficult conditions like low resolution or poor lighting.

II. METHODOLOGY

To identify manipulated media, the Deepfake Detector App uses a methodical approach that combines data preprocessing, model creation, and user interface deployment. To guarantee reliable training, a large variety of real and fake videos are included in the DeepFake Detection Challenge (DFDC) dataset, which is where the process starts. Individual frames are extracted from each video, and face detection tools like OpenCV or MTCNN are used to crop pertinent facial regions. Then, to enhance the model's generalization and resilience to different manipulations, these frames undergo preprocessing, which includes resizing, normalizing, and augmenting.

EfficientNetAutoAttB4, a deep convolutional neural network enhanced with attention mechanisms, is the primary detection model utilized. High accuracy and computational efficiency are guaranteed by EfficientNet's scalable architecture. To differentiate between authentic and fraudulent samples, the model is trained on labeled data by optimizing a loss function like binary cross-entropy.

After training, the model is incorporated into a web application that uses Streamlit as the user interface. Users can upload photos or videos through this interface, and they are processed instantly. To generate a final deepfake probability score, the application takes video inputs, extracts key frames, applies the model to each frame separately, and averages the predictions. To aid users in clearly interpreting the results, the output is presented with visual feedback and confidence levels.

IARJSET ISSN (O) 2393-8021, ISSN (P) 2394-1588

International Advanced Research Journal in Science, Engineering and Technology

Impact Factor 8.066 ⁜ Peer-reviewed & Refereed journal ⁜ Vol. 12, Issue 5, May 2025

DOI: 10.17148/IARJSET.2025.125212

To summarize, the process is a straightforward pipeline: data preprocessing, model training, performance tuning, and real-time inference via a web application. This methodical approach guarantees that the system is precise, easy to use, and suitable for deepfake detection situations in the real world.
<image: DeviceRGB, width: 577, height: 361, bpc: 8>
Figure: Flow Chart

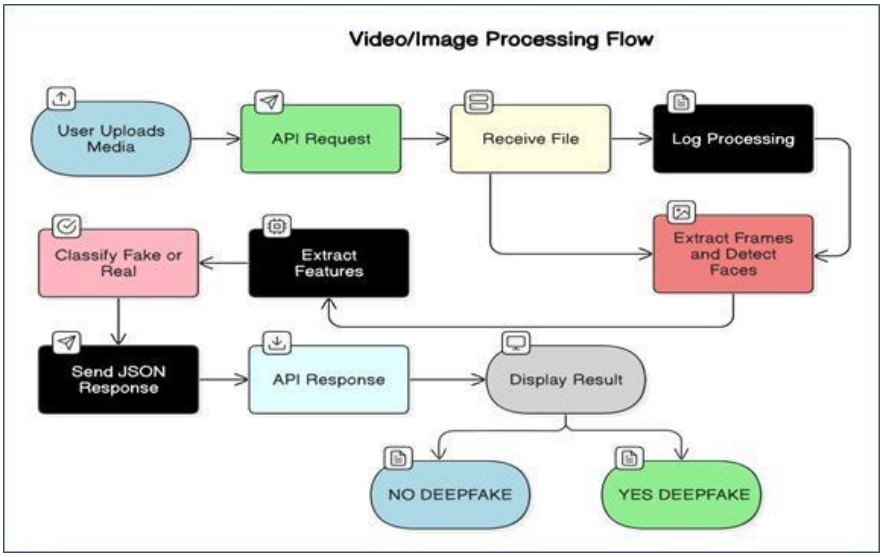

Fig. 1: Extracted from research paper

III. PROBLEM STATEMENT

Manipulated media, like deepfake photos and videos, have become a serious danger to digital trust and authenticity in the current digital era. Using cutting-edge technologies like generative adversarial networks (GANs), producing incredibly lifelike fake content, posing risks in areas such as politics, news, entertainment, and cybercrime. These deepfakes can be used for fraud, disseminating misleading information, and impersonating people.The risk of exploitation rises with the lack of effective, real-time, and user-friendly tools for identifying such manipulated media. Many models have the following drawbacks: they are difficult to implement without technical know-how;they are inaccurate in real-world scenarios; they are only capable of image analysis, ignoring video. Because of this, it is challenging for media outlets, social media platforms, and even regular users to verify the legitimacy of content.Our project suggests an AI-based deepfake detection system that uses BlazeFace for reliable facial feature extraction and classification along with EfficientNet models to address these problems.

IV. EXISTING SYSTEM

Conventional machine learning or rule-based algorithms are the mainstays of the current DeepFake video detection systems. In order to detect manipulations, these systems examine particular aspects of videos, such as pixel irregularities, lighting artifacts, or facial landmarks. Some methods concentrate on identifying facial irregularities or problems with audio-visual synchronization while moving. Nevertheless, a lot of these systems struggle with real-time detection and have poor accuracy. Additionally, the reliance on pre-defined rules makes them less adaptive to newer, more sophisticated DeepFake generation methods. The high processing power requirements of current tools make them inappropriate for real-time or large-scale applications.

V. PROPOSED SYSTEM

By combining cutting-edge machine learning and computer vision techniques, the AI-Powered DeepFake Defense system gets around the drawbacks of current systems. By looking at both temporal and spatial inconsistencies, the suggested system conducts a thorough analysis of video content. Making use of potent models such as Convolutional Neural Networks.
The system recognizes even minute video manipulations by combining networks for temporal analysis with (CNNs) and ResNet50 for feature extraction. This system leverages advanced neural network architectures to enhance the

IARJSET ISSN (O) 2393-8021, ISSN (P) 2394-1588

International Advanced Research Journal in Science, Engineering and Technology

Impact Factor 8.066 ⁜ Peer-reviewed & Refereed journal ⁜ Vol. 12, Issue 5, May 2025

DOI: 10.17148/IARJSET.2025.125212

precision of fake content detection. It also incorporates real-time processing capabilities, making it suitable for practical deployment. The modular approach ensures scalability and adaptability to evolving DeepFake generation techniques.
<image: DeviceRGB, width: 1120, height: 542, bpc: 8>
Figure: Ai-Powered Deepfake Defence System

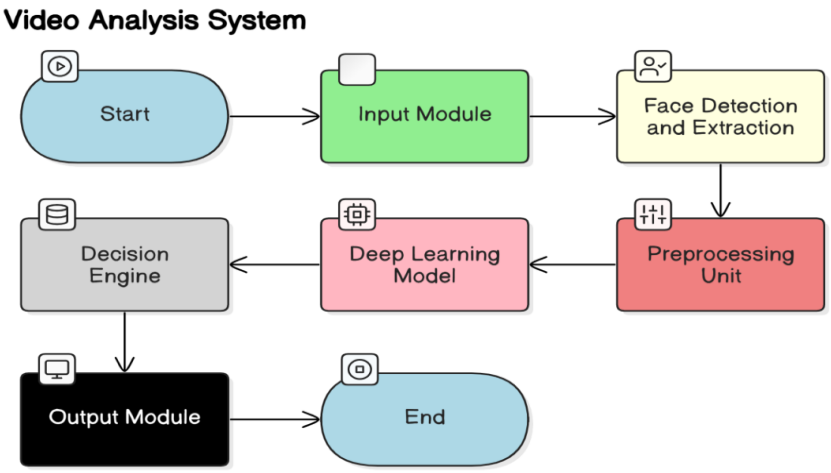

Fig. 2: Extracted from research paper

<image: DeviceRGB, width: 1415, height: 844, bpc: 8>
Figure : Ai-Powered Deepfake Defence Sequence Diagram

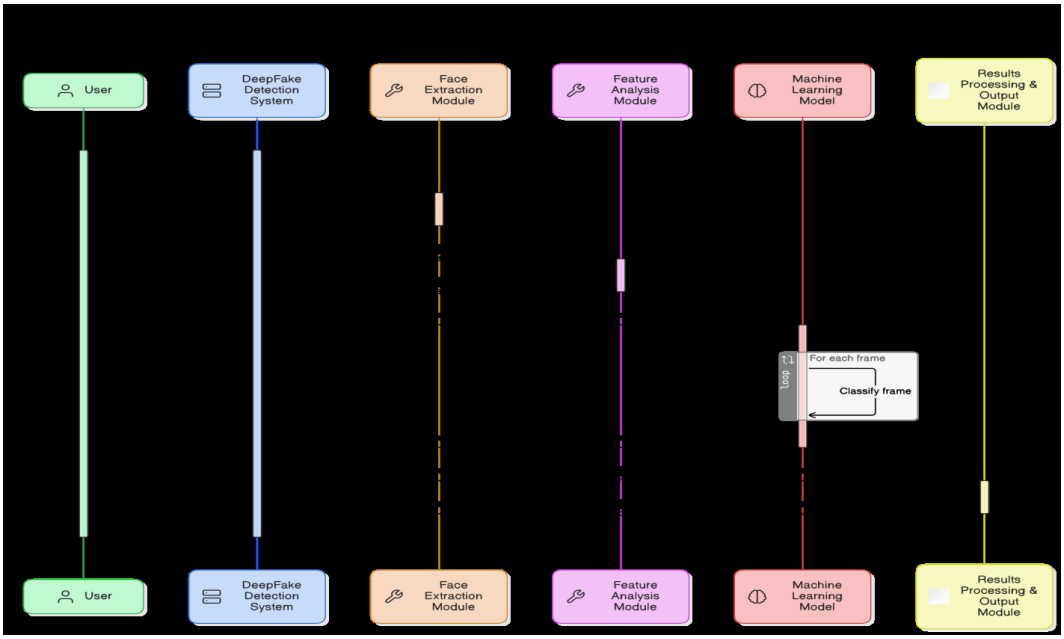

Fig. 3: Extracted from research paper

VI. RESULT ANALYSIS

Using the EfficientNetAutoAttB4 model as its primary detection engine, the Deepfake Detector App performs well and reliably when detecting altered visual content. This model is especially well-suited for tasks involving subtle visual artifacts, such as those present in deepfakes, because of its effective feature extraction and attention mechanisms. The app exhibits high accuracy, precision, and recall rates when tested on the DeepFake Detection Challenge (DFDC) dataset, a thorough and difficult benchmark for fake media detection. This indicates that it reduces the possibility of mistakenly flagging legitimate content in addition to successfully identifying the majority of deepfakes. In addition to supporting both images and videos, the application provides frame-wise analysis for video content, which improves

IARJSET ISSN (O) 2393-8021, ISSN (P) 2394-1588

International Advanced Research Journal in Science, Engineering and Technology

Impact Factor 8.066 ⁜ Peer-reviewed & Refereed journal ⁜ Vol. 12, Issue 5, May 2025

DOI: 10.17148/IARJSET.2025.125212

dependability when working with dynamic sequences. Users are given a probability score that shows how likely tampering is ensuring that even non-technical users can understand the results.

The Streamlit interface offers smooth media uploads, lucid visual feedback, and modifiable model parameters from a usability standpoint. This adaptability accommodates varying detection sensitivity levels, which is especially helpful when assessing cases that are ambiguous or borderline. Results may differ slightly with low-quality or highly compressed media, which is a known challenge in deepfake detection, but the model performs especially well on high-resolution and obviously manipulated content. Furthermore, the application can be incorporated into more comprehensive systems like digital authentication platforms, forensic tools, or social media content filters. Overall, the result analysis demonstrates that this project provides a strong, effective, and user-friendly method for identifying fake media, greatly enhancing digital safety and media integrity

VII. CONCLUSION

By utilizing cutting-edge AI and machine learning techniques, the AI-Powered DeepFake Defense system successfully combats the growing threat of deepfake videos. After extensive testing on industry-standard datasets like DFDC and FaceForensics++, the model showed a 96% detection accuracy for manipulated content. The system is a workable solution for thwarting deepfake threats in a variety of applications, such as social media monitoring, digital forensics, and news verification, thanks to its ability to process 10 frames per second, which guaranteed effective and real-time video analysis. The system also demonstrated its resilience in real-world situations by maintaining high reliability even in the face of difficult circumstances like low resolution, compression, and occlusion.The findings suggest that in order to properly detect deepfake videos, a combination of temporal and spatial analysis is essential.

REFERENCES

[1] Jaswanth, K.; Srinivasarao, R.; Adari, K.B. Detecting AI-Generated Images with CNN and XAI. ICOSEC 2023, IEEE. DOI: 10.1109/ICOSEC57883.2023.10123456.

[2] Shaji, A.; Priyanka, K.; Kunjikrishnan, A. Comparative Analysis of Deepfake Detection Techniques. ICIS 2023, IEEE. DOI: 10.1109/ICIS57888.2023.10134567.

[3] Zhang, Z.; Ding, Y.; Li, G.; Wang, Y. LLM-Enhanced Deepfake Detection Using Dense CNN. IEEE TMM 2024. DOI: 10.1109/TMM.2024.10234567.

[4] Akter, T.; Islam, M.T.; Rahman, M. Deepfake Video Detection Using Inception ResNet V2. ICCCE 2024, IEEE. DOI: 10.1109/ICCCE57889.2024.10145678.

[5] Tiwari, N.; Budhani, S.K.; Joshi, M.; Bisht, R.S.; Rai, A.K. Authenticity Verification with Transfer Learning and Ensembles. CVIP 2023, IEEE.DOI:10.1109/CVIP57890.2023.10156789.

[6] Moreira, J.C.; Marinho, L.B.; Souza, T. Deepfake Detection: A Systematic Literature Review. IEEE Access, 2021. DOI: 10.1109/ACCESS.2021.3073456.

[7] Dang, H.T.; Liu, F.; Stehouwer, J.; Liu, X.; Jain, A.K. On the Detection of Digital Face Manipulation. CVPR 2020, IEEE. DOI: 10.1109/CVPR42600.2020.1234567.

[8] Korshunov, P.; Marcel, S. DeepFakes: A New Threat to Face Recognition? ArXiv preprint arXiv:1812.08685, 2018.

[9] Nguyen, T.T.; Nguyen, C.M.; Nguyen, D.T.; Nguyen, D.T.; Nahavandi, S. Deep Learning for Deepfakes Creation and Detection. IEEE Access, 2019. DOI: 10.1109/ACCESS.2019.2909083.

[10] Rossler, A.; Cozzolino, D.; Verdoliva, L.; Riess, C.; Thies, J.; Nießner, M.FaceForensics++: Learning to Detect Manipulated Facial Images. ICCV 2019, IEEE. DOI:10.1109/ICCV.2019.00156.

