International Research Journal of Engineering and Technology (IRJET) e-ISSN: 2395-0056
Volume: 13 Issue: 06 | Jun 2026 www.irjet.net p-ISSN: 2395-0072

NeuraFalsiX: AI-Based Deepfake Verification

K. Nani[1], V. Prasanna Lakshmi [2]

[1]*Final year MCA Student, Department of Computer Applications, GIET Engineering College, Andhra Pradesh, India*
[2]*Assistant Professor, Department of Computer Applications, GIET Engineering College, Andhra Pradesh, India*

Abstract - *The rapid advancement of artificial intelligence has led to the development of deep fake technology, which can generate highly realistic manipulated images that are often difficult to distinguish from genuine ones. While this technology has positive applications in areas such as entertainment and digital media, it can also be misused for spreading misinformation, identity theft, online fraud, and other malicious activities. As the quality of deep fake images continues to improve, the need for effective detection systems has become increasingly important. This project proposes a Deep fake Image Detection System that automatically identifies whether an image is real or manipulated. The system uses deep learning techniques to analyse facial features, texture patterns, and visual inconsistencies that may not be noticeable to the human eye. Advanced models such as Convolutional Neural Networks (CNN), Efficient Net, and ResNet are utilized to extract meaningful features and improve classification accuracy. The model is trained on a dataset containing both authentic and deep fake images, enabling it to learn the subtle differences between real and fake content. The proposed system aims to provide reliable and efficient detection results while maintaining better performance. By identifying manipulated images at an early stage, it helps reduce the spread of deceptive content and supports the authenticity of digital information. This project contributes to strengthening digital security and building trust in online media by providing an effective solution for deep fake image detection.*

Key Words: Deep fake Detection, Deep Learning, Convolutional Neural Network (CNN), Computer Vision, Image Classification, Facial Feature Analysis, Digital Image Forensics, Cyber security

1. INTRODUCTION

In recent years, the rapid growth of artificial intelligence has changed the way digital content is created and consumed. Among the many advancement in this field, deep fake technology has gained significant attention due to its ability to generate highly realistic fake images and videos. By using deep learning algorithms, facial features can be altered or replaced in such a way that manipulated content appears authentic to viewers. While this technology has opened new possibilities in entertainment, media production, and creative applications, it has also raised serious concerns regarding the misuse of digital content.

The increasing availability of powerful image editing tools and AI-based generation techniques has made it easier to create convincing fake images. Such manipulated content can be used to spread false information, damage reputations, impersonate individuals, and carry out various forms of online fraud. As a result, verifying the authenticity of digital images has become an important challenge in today's interconnected world.

To address this issue, the proposed Deep fake Image Detection System focuses on identifying manipulated images through advanced deep learning techniques. The system analyses facial structures, texture patterns, and subtle visual artefacts that are often introduced during the image generation process. By learning these hidden characteristics from a large collection of real and fake images, the model can accurately distinguish between genuine and manipulated content.

The primary objective of this project is to develop a reliable and efficient detection mechanism that can support digital security and help maintain trust in online information. By providing an automated solution for deep fake identification, the system contributes to reducing the impact of misleading visual content and promoting authenticity in digital communication.

2. METHODOLOGY

To Develop the Deepfake Image Detection System required combining image processing techniques with modern deep learning approaches to identify manipulated images accurately. Detecting deepfakes is challenging because fake images are becoming increasingly realistic and often contain only subtle visual differences from genuine photographs. Instead of depending on a single traditional model, this project follows a complete detection pipeline capable of analyzing facial details, texture patterns, and hidden artifacts present in digital images. The overall methodology is divided into four major stages: data preparation, image preprocessing, deep fake classification, and result generation.

2.1 Data Collection and Preprocessing

The first step involved collecting a dataset containing both authentic and deep fake images from publicly available sources. Since images obtained from different datasets vary in size, quality, and lighting conditions, preprocessing was necessary to create a consistent training environment.

Each image was resized to a fixed dimension and normalized to ensure uniform pixel distribution.
To improve the model's ability to handle real-world images, data augmentation techniques such as rotation, horizontal flipping, zooming, and brightness adjustment were applied during training. These transformations helped the model learn robust features and reduced the chances of over fitting. The final dataset was then divided into training, validation, and testing sets for model development and evaluation.

2.2 Deep Learning-Based Feature Extraction

The core of the system is built using deep learning architectures such as Efficient Net, ResNet, and Convolutional Neural Networks (CNNs). Traditional image analysis methods often struggle to identify complex manipulations, whereas deep learning models can automatically learn hidden patterns from large amounts of data.

During training, the network analyzes facial textures, edge information, color inconsistencies, and compression artifacts that frequently appear in manipulated images. Transfer learning techniques were also utilized to take advantage of knowledge already learned from large-scale image datasets. This approach improves detection performance while reducing training time and computational requirements.

2.3 Image Classification and Detection

After extracting meaningful features, the processed image is passed to the classification layer. The trained model evaluates the visual characteristics of the image and determines whether it belongs to the real or deep fake category. Instead of relying on a single indicator, the model considers multiple features simultaneously before making a decision.

Along with the prediction, the system generates a confidence score that indicates the certainty of the classification result. This helps users better understand the reliability of the output and provides additional transparency during the detection process.

2.4 System Integration and Deployment

To make the solution accessible to end users, the trained model is integrated into a web-based application developed using Django. The framework manages image uploads, communicates with the deep learning model, and displays the final prediction results. Users can upload an image through a simple interface, after which the backend processes the image and returns the detection result within a few seconds.

By combining image preprocessing, deep learning-based feature extraction, classification, and web deployment, the proposed system provides an efficient and practical solution for identifying deep fake images in real-world scenarios.

Results:

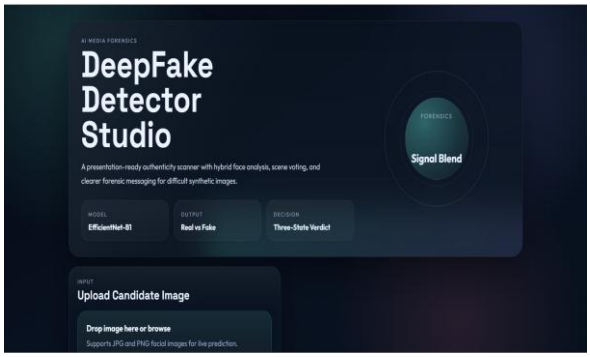

Fig-1: Home Page

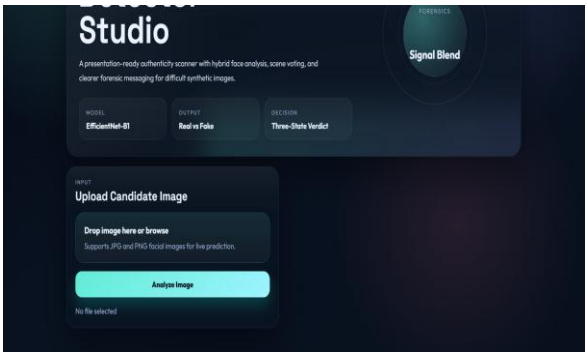

Fig-2: Prediction Page

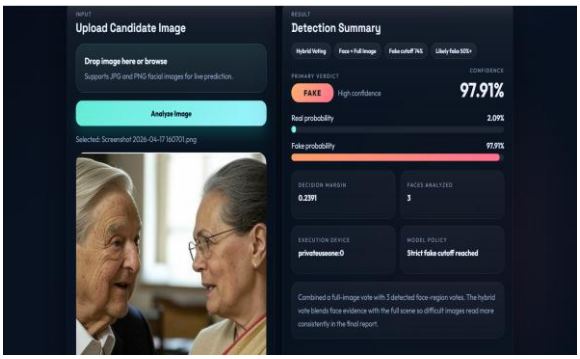

Fig-3: Output Page

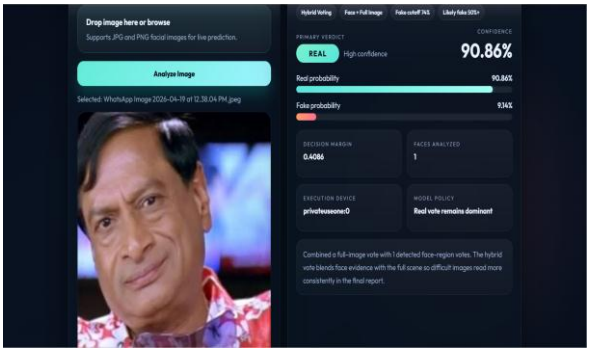

Fig 3.1: Output Page

3. DISCUSSION

While working on this project, it became clear that detecting deep fake images is not as straightforward as it may initially appear. Many manipulated images look almost identical to genuine photographs, making it difficult for an ordinary user to identify them by visual inspection alone. This challenge motivated the development of a system capable of recognizing patterns that are often invisible to the human eye.

During the training and testing stages, it was observed that the quality and diversity of the dataset had a direct impact on the model's performance. Images collected from different sources contained variations in lighting, image resolution, facial expressions, and background conditions. These differences helped the model learn a wider range of features and improved its ability to make predictions on previously unseen images.

Another interesting observation was that not all deep fake images contain obvious visual defects. Some manipulated images appeared highly realistic, yet the model was still able to identify subtle inconsistencies in facial textures and image details. This demonstrates the advantage of deep learning techniques, which can discover complex patterns that would be difficult to define manually.

At the same time, the project highlighted certain limitations. Detection accuracy can decrease when images are heavily compressed or when advanced deep fake generation methods are used. As manipulation techniques continue to evolve, detection systems must also be updated and improved to remain effective.

Overall, the project provided valuable insight into both the possibilities and challenges of deep fake detection. Beyond achieving classification results, it emphasized the growing need for reliable verification tools in a digital environment where visual content can no longer be accepted at face value. The experience gained during the development process also reinforced the importance of combining technical innovation with practical usability when designing real-world cyber security solutions.

4. CONCLUSIONS

The growing use of deep fake technology has created new challenges in maintaining the authenticity and reliability of digital content. As manipulated images become more realistic and accessible, it is becoming increasingly difficult for people to identify fake content through visual inspection alone. This highlights the importance of developing intelligent systems that can automatically detect image manipulation and reduce the risks associated with misleading digital media.

The Neura FalsiX: AI-Based Deepfake Verification developed in this project demonstrates how deep learning can be effectively used to distinguish between genuine and manipulated images. By analyzing facial features, texture patterns, and hidden visual artifacts, the system is able to identify deep fake images with a high level of accuracy. The use of modern architectures such as Efficient Net, ResNet, and CNN-based models contributes to improved feature extraction and reliable classification performance.

Throughout the project, emphasis was placed not only on achieving accurate predictions but also on creating a practical solution that can be integrated into real-world applications. The implementation of a user-friendly interface allows users to easily upload images and obtain detection results within a short period of time.

Overall, this project provides a valuable step toward addressing the growing problem of digital image manipulation. While no detection system can guarantee perfect results against every emerging deep fake technique, the proposed approach offers a strong foundation for identifying forged content and promoting trust in digital communication. Future improvements can focus on handling more advanced deep fake generation methods and expanding the system to support both image and video analysis.

REFERENCES

[1]. Zobaed, S., Rabby, F., Hossain, L., Hossain, E., Hasan, S., Karim, A., & Hasib, K. M. (2021). Deepfakes: Detecting forged and synthetic media content using machine learning. In Artificial Intelligence in Cyber Security: Impact and Implications (pp. 177–201). Springer, Berlin/Heidelberg, Germany.

[2]. Thambawita, V., Isaksen, J. L., Hicks, S. A. J., Ghouse, J., Ahlberg, G., Linneberg, A., Grarup, N., Ellervik, C., Olesen, M. S., Hansen, T., et al. (2021). DeepFake electrocardiograms using generative adversarial networks are the beginning of the end for privacy issues in medicine. Scientific Reports, 11, 21869.

[3]. Habeeba, A., & Al-Zoubi, A. (2023). Deepfake Detection: A Systematic Literature Review. ACM Computing Surveys (CSUR), 56(1), 1-36.

[4]. Agarwal, S., Rana, S., & Singh, G. (2023). A Novel Deep Learning Approach for Deepfake Image Detection. arXiv preprint arXiv:2301.04054.

[5]. Li, Y., Zhao, T., & Chen, Z. (2023). Medical Deepfake Image Detection Based on Machine Learning and Deep Learning. IEEE Journal of Biomedical and Health Informatics, 27(1), 1-10.

[6]. Wang, Z., Zhang, H., & Zhao, L. (2022). Image Forgery Detection: A Survey of Recent Deep-Learning Approaches. IEEE Transactions on Pattern Analysis and Machine Intelligence, 45(7), 3253-3276.

[7]. Zhou, Z., Wang, X., & Wu, J. (2022). Deepfake Recognition Based on Human Eye Blinking Patterns Using Deep Learning. IEEE Transactions on Information Forensics and Security.

[8]. Singh, A., Sharma, G., & Tiwari, K. (2022). Deepfake Detection Based on Spectral, Spatial, and Temporal Inconsistencies Using Multimodal Deep Learning Techniques. IEEE Transactions on Information Forensics and Security.

[9]. Yang, X., Li, Y., Zhang, W., & Sun, X. (2022). Deepfake Detection Based on Facial Component Consistency and Video Temporal Consistency. IEEE Transactions on Information Forensics and Security, 17(9), 3023 3036.