Deep Fake Face Detection Using Deep Learning

Dheeraj Shukla, Laxmi Pawar, Kaveri Patil, Gayatri Patil, Darshana Jamdade Department of Computer Engineering, D.N. Patel College of Engineering,

Shahada, Dist. Nandurbar, Maharashtra, India

dheerajbshukla@gmail.com, laxmiipawar153@gmail.com, kaveripatil840@gmail.com, missgayu24@gmail.com,darshanajamdade@gmail.com

Keywords: *Artificial Intelligence, deepfake technology, Generative Adversarial Networks GAN, Detection System, Detection Accuracy, User accessibility, Digital content verification.*

***Abstract*:** In recent years, the rise of deepfake technology has raised significant concerns. regarding the authenticity of digital content. Deepfakes, which are synthetic media created using advanced artificial intelligence techniques, can mislead viewers and pose risks to personal privacy, public trust, and social discourse. The proposed system focuses on developing a Generative Adversarial Network (GAN)- based deepfake detection system that aims to identify manipulated images and videos accurately and efficiently. The importance of this research lies in its potential to enhance digital content verification, ultimately restoring trust in media across various sectors, including news, entertainment, and social media. The proposed approach utilizes GANs to both generate synthetic deepfake samples for training and serve as the basis for the detection engine. By focusing solely on GANs, the system leverages their unique capabilities to create a model that is adaptable to evolving deepfake generation techniques. The architecture of the system includes a user-friendly frontend, a robust backend, and a powerful detection engine, all integrated seamlessly to ensure real- time processing and analysis of media files.

I. INTRODUCTION

Deep learning is a powerful subfield of machine learning inspired by the human brain's structure and functionality. It uses deep neural networks with multiple layers to automatically learn and extract complex patterns from massive datasets.[1] Unlike traditional machine learning, it eliminates the need for manual feature engineering by working end-toend, mapping raw data to outputs. Deep learning has revolutionized industries with applications like image recognition, natural language processing, speech recognition, and autonomous systems.[2] Technologies like voice assistants, self-driving cars , and medical imaging

owe their breakthroughs to deep learning.[3] Frameworks such as TensorFlow and PyTorch are widely used to design and train these models, making it a foundational technology in artificial intelligence.[4]

The domain of this project is Deepfake Detection and Image Authenticity Verification, which lies at the intersection of artificial intelligence, computer vision, and cybersecurity

[5]. The ad vent of generative adversarial networks (GANs) has revolutionized the creation of synthetic media, enabling the generation of highly realistic fake images and videos, commonly referred to as

” deep fakes "[6]. While GANs offer immense potential in creative industries, their misuse for creating deceptive content poses severe ethical and security challenges[7] . This project aims to address these challenges by developing a system capable of detecting fake images generated using GANs, ensuring media authenticity and enhancing trust in digital content.[8]

II. LITERATURESURVEY

2.1 Overview

1. Fake News Detection using Deep Learning.

Author :Lyu, Siwei

As deepfakes are increasingly used for

creating various forms of fake content, including celebrity pornography and fake news, there is a

pressing need for effective detection methods.[1] Deep fake technology is heavily used in the creation of adult content, with

thousands of deepfake videos found on pornographic platforms.[2] Additionally, new platforms dedicated to distributing deepfake pornography emerged.[3] Deepfake learning

models represent a critical aspect of combating the proliferation of digitally manipulated multimedia content.[4] Several prominent models and approaches have emerged in this domain: Variational Autoen coders (VAEs): VAEs are used to encode and decode visual content.[5]

Generative Adversarial Networks (GANs): GANs, which are the foundation of many deepfake generation methods, can also be used for detection.[6]

(CNNs): Finds extensive application in the realm of image and video scrutiny.[7]

2. Unmasking Deepfakes: A Deep Learning Approach for Accurate Detection and Classification of Synthetic Videos.

Author: Mayur Bhogade, Bhushan Deore, Abhishek Sharma, Omkar Sonawane.

Abstract: The growth of deepfakes fuels the spread of misinformation, undermining trust in media and information sources. Additionally, they worsen societal divisions by sharing fake content, leading to confusion and polarization. Deepfakes are becoming increasingly common, making it harder to spot them because they look so real. This paper addresses this problem by introducing a method to detect differences in facial features during video creation. Detection of deepfakes can be tricky due to their high realism, but our approach helps identify these fake videos by spotting changes in facial structures. Our model employs a Res-Next Convolutional

Neural Network to extract frame-level features, which are then utilized to train a Long Short-Term Memory (LSTM)-based Recurrent Neural Network (RNN). This RNN classifies videos whether they are subjected to any manipulation or not. We have used a dataset called "Celeb DF" to train our model to detect differences created around the face during deepfake construction. Integrated with a userfriendly interface utilizing ReactJs on the front end and Flask on the backend, our solution ensures robust defense against potential threats posed by deepfakes while prioritizing accessibility and usability.

3. Advancing Deep Fake detection: Mobile Application with deep learning.

Author: Arun KS, Juan Shaji Austin, Kiran Paulson, Kevin Paulson, Suzen Saju Kallungal.

This paper introduces a pioneering approach to deepfake detection leveraging the power of mobile platforms through Flutter. We combine state-of-the-art ResNeXt and LSTM models for robust deepfake identification and en- capsulate them within a user-friendly mobile interface. By harnessing the versatility of Flutter’s In App Web View, our so- lution seamlessly integrates with mobile devices, empowering users with realtime deepfake detection capabilities on their smartphones. Through rigorous evaluation, we demonstrate the efficacy and usability of our approach, marking a significant advancement in the field of mobile-based deepfake detection.

4. DeepFake Face Image Detection based on Improved VGG Convolutional Neural Network.

Author: Xu Chang, Jian Wu, Tongfeng Yang, Guorui Feng

Abstract: DeepFake can forge high-quality tampered images and videos that are consistent with the distribution of real data. Its rapid development causes people's panic and reflection. In this paper we presents an improved VGG network named NAVGG to detect DeepFake face image, which was based on image noise and image augmentation.

Firstly, In order to learn the tampering artifacts that may not be seen in RGB channels, SRM filter layer is used to highlight the image noise features; Secondly, the image noise map is augmented to weaken the face features. Finally, the augmented noise images are input into the network to train and judge whether the image is forged. The experimental results using the Celeb-DF dataset have shown that NA-VGG made great improvements than other state-of-the-art fake image detector .

III. METHODOLOGY

3.1 Approach

The proposed Deepfake Detection System is an innovative solution addressing the challenge of ma nipulated media by leveraging advanced technologies and a robust architecture. It integrates a React.js-based frontend for an interactive and userfriendly interface, enabling seamless image up loads, and a Node.js and Express.js backend for efficient request handling and real-time commu nication via RESTful APIs. At its core, the system employs a Deepfake Detection Engine built with Python, utilizing Generative Adversarial Networks (GANs) and Convolutional Neural Networks (CNNs) for training on synthetic and real datasets, feature extraction, and high-accuracy classification of genuine and fake images. MongoDB, a NoSQL database, supports efficient data storage and retrieval of user information, image metadata, and detection results, while a dedicated API layer ensures streamlined communication and modularity for scalability. Deployed on cloud infrastructure, the system offers high performance, accessibility, and scalability, accommodating a large user base effectively

1. Frontend: Built with React.js for an intuitive user interface.

2. Backend:Designed with Node.js and Express.js for seamless operations.

3. Detection Engine:Leverages GANs and CNNs for high accuracy in detection.

4. Cloud Deployment: Provides scalability and accessibil- ity for a large user base.

5. Model Training: Employs robust preprocessing and training techniques with TensorFlow or PyTorch frame- works.

3.2 System Architecture

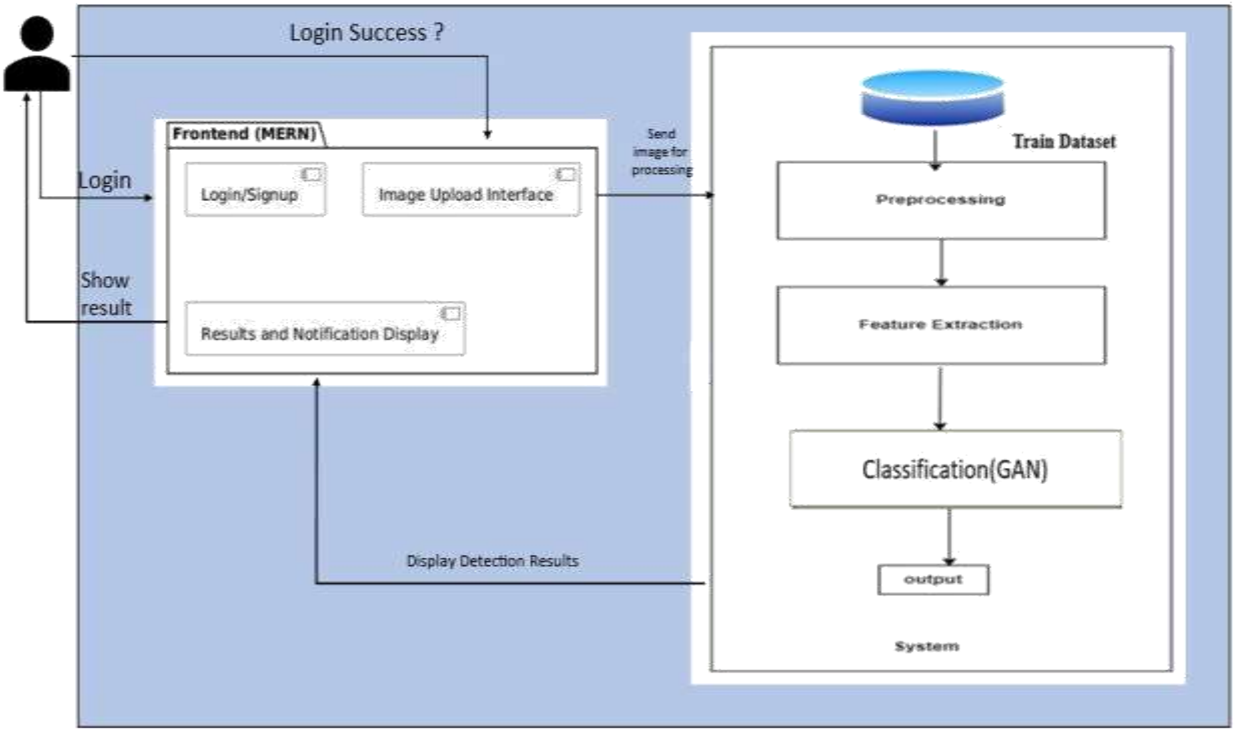


Figure 3.2 System Architecture for Deepfake Face Detection

□

Preprocessing: The input image is resized, normalized, and cleaned to remove noise and enhance important visual details, ensuring consistent quality for analysis.

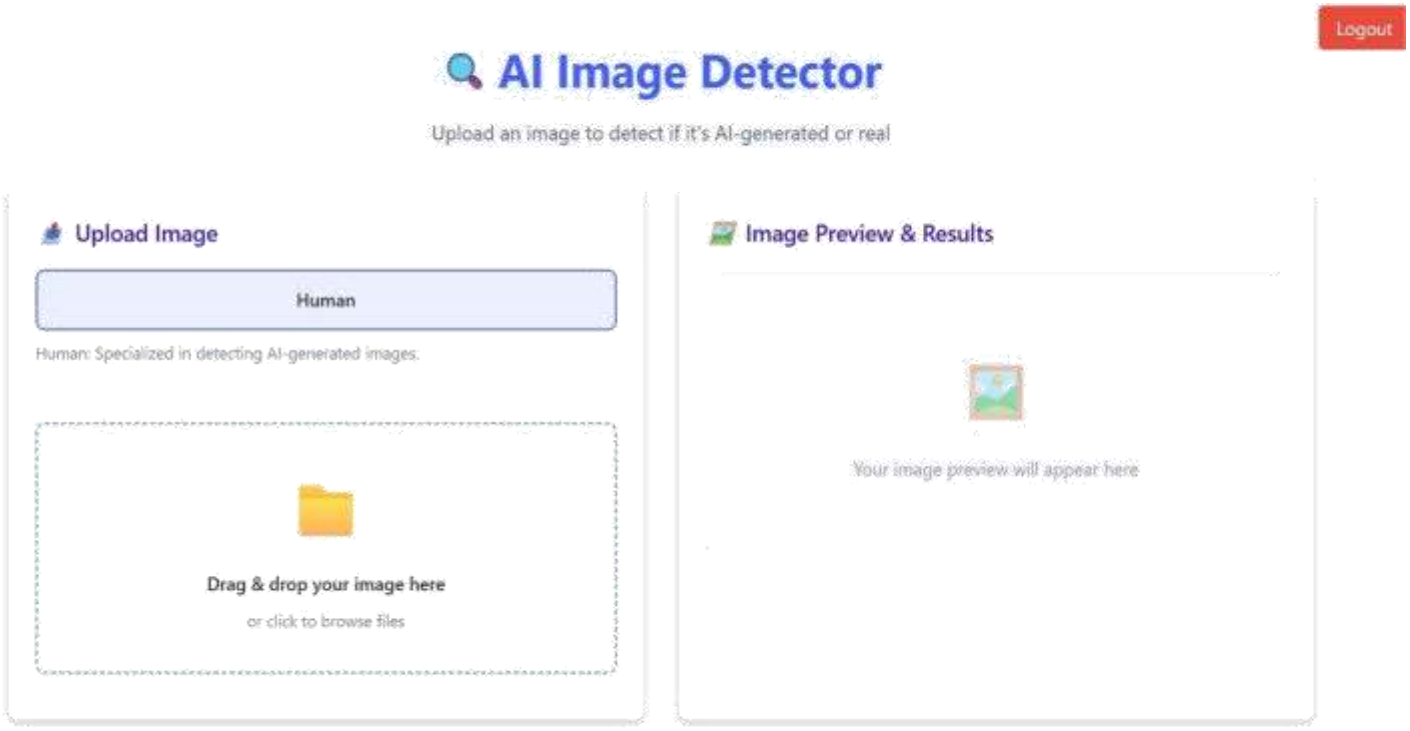
□

Feature Extraction: Deep learning techniques (such as CNNs) are used to extract unique patterns and artifacts from the image that help distinguish real images from manipulated ones. **Classification (GAN)**: The extracted features

are passed to a GAN-based classifier, which determines whether the input image is genuine or a deepfake based on learned characteristics.

IV. IMPLEMENTATION AND RESULTS

For the implementation purpose, the four existing approaches are considered. The results of mentioned four models are compared with the proposed model, it is found the accuracy among top 200 results is mentioned in the table 4.1.



Logout

AI Image Detector

Upload an image to detect if it's AI-generated or real

Upload Image

Human

Human: Specialized in detecting AI-generated images.

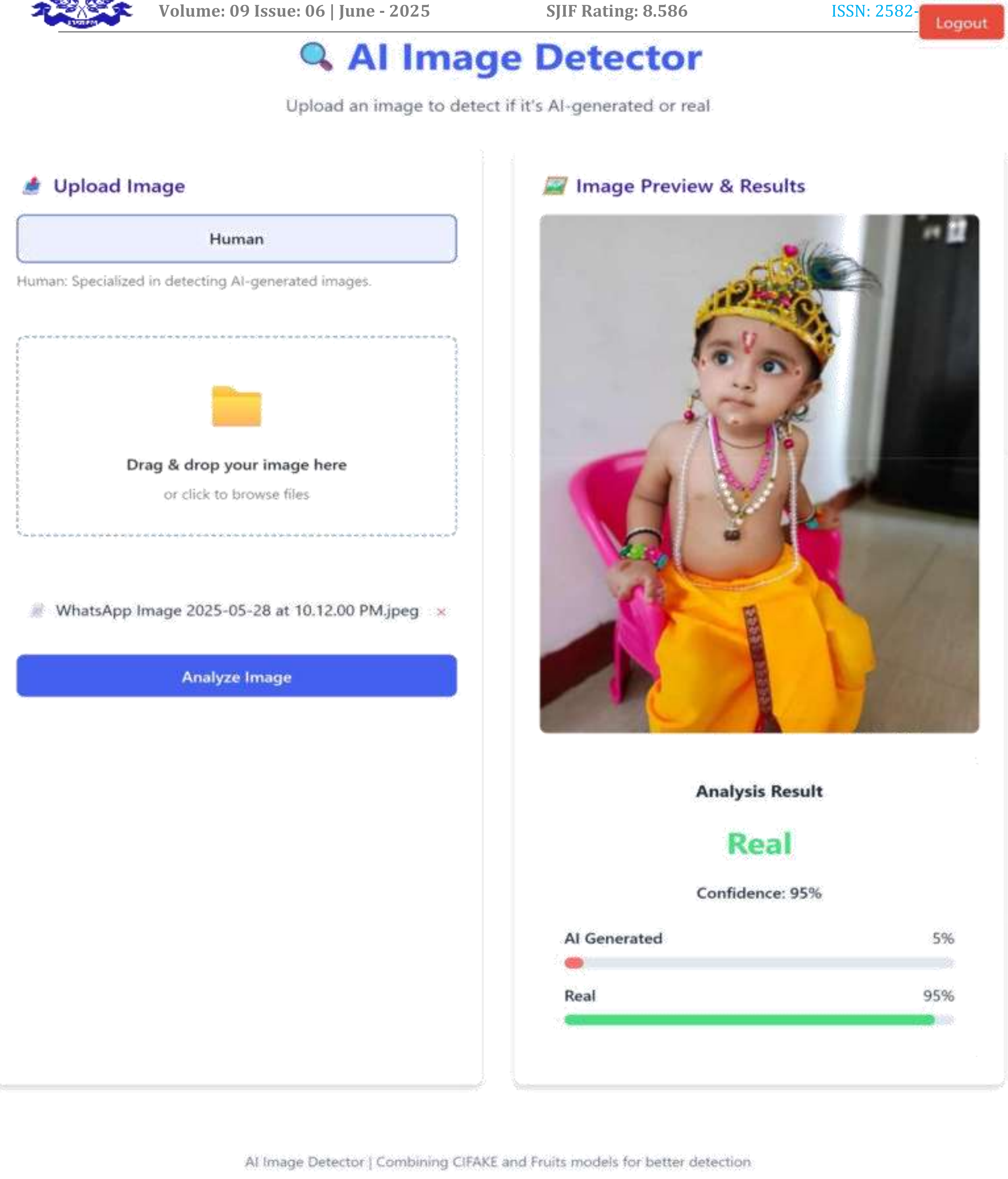
Drag & drop your image here

or click to browse files

Image Preview & Results

Your image preview will appear here

Figure 4.1 Analyse Page



AI Image Detector

Upload an image to detect if it's AI-generated or real

Human: Specialized in detecting AI-generated images.

Drag & drop your image here

or click to browse files

WhatsApp Image 2025-05-28 at 10.12.00 PM.jpeg ×

Analysis Result

Real

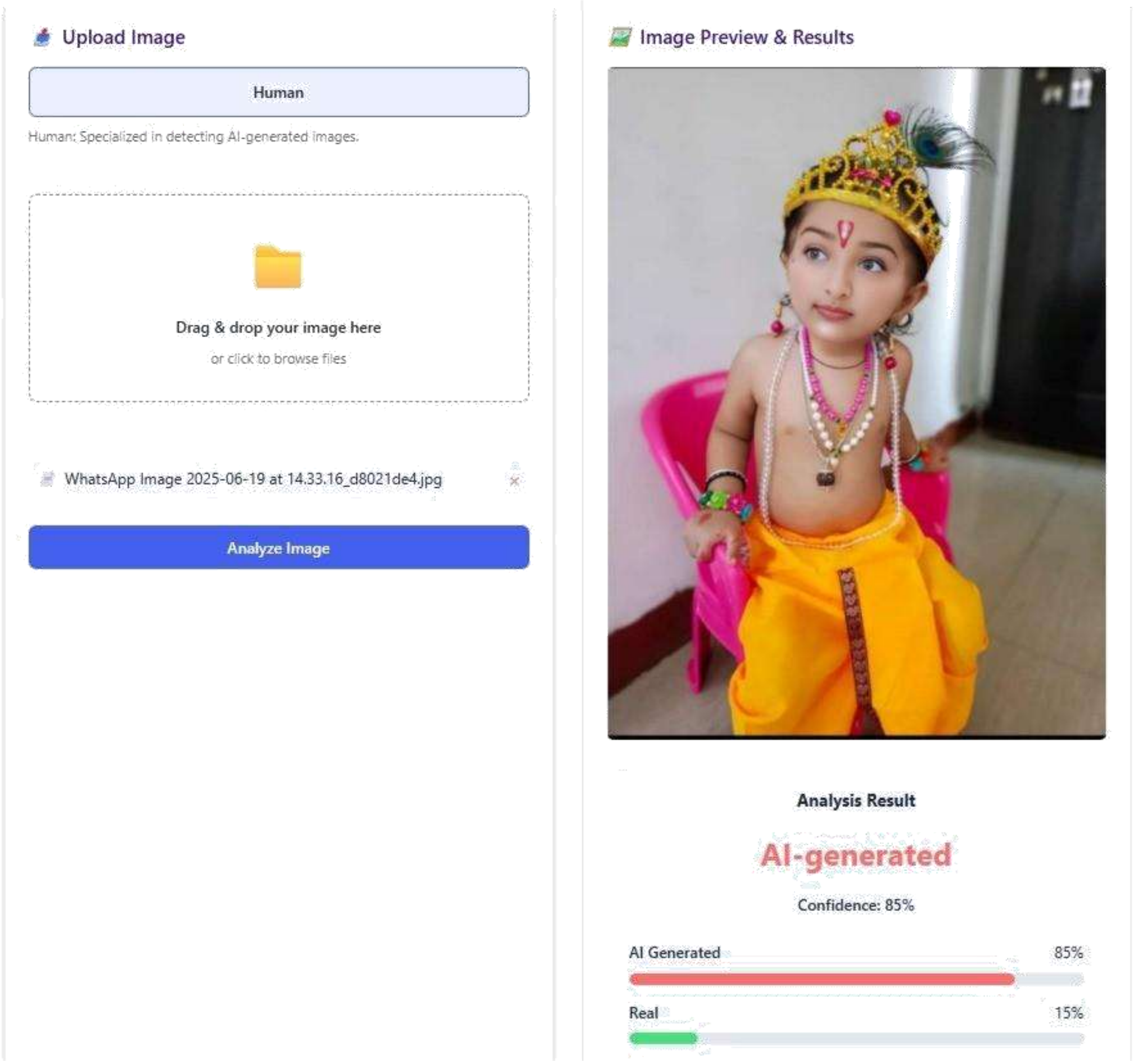
Confidence: 95%

AI Generated 5%

Real 95%

AI Image Detector | Combining CIFAKE and Fruits models for better detection

Figure 4.2 Real Image detection Result



Human

Human: Specialized in detecting AI-generated images.

WhatsApp Image 2025-06-19 at 14.33.16_d8021de4.jpg

Analysis Result

AI-generated

Confidence: 85%

AI Generated 85%

Real 15%

Figure 4.2 Fake Image detection Result

CONCLUSION

Proposed System will aim to address the critical issue of deepfake detection, contributing to ongoing efforts to preserve the integrity of digital content. The primary objective will be to develop a GAN-based detection system capable of accurately identifying manipulated images and videos while ensuring accessibility through a user-friendly interface The system will leverage advanced AI techniques, focusing on robustness, accuracy, and inter pretability. It will also incorporate insights from diverse datasets and employ advanced feature extraction methods to improve its effectiveness against various types of synthetic media. Byprioritizing both performance metrics and usability, the system is expected to deliver a reliable solution for detecting deepfakes. It will serve as a significant step toward combating the misuse of deepfake technology and fostering trust in digital content.

References

[1] Sattar, S.K., Preetham, T.G., Kalyan, V., Venu, P., & Avinash, B. (2024). Unmasking Deep fakes: A Deep Learning Approach for Accurate Detection and Classification of Synthetic Videos. *Journal of Artificial Intelligence Research*

[2] Gupta, G., Raja, K., & Prasad, M. (2024). A Comprehensive Review of DeepFake Detec tion Using Advanced Machine Learning and

Fusion Methods. *Electronics*, 13(1), 95. DOI: 10.3390/electronics13010095 .

[3] Mehra, A., Spreeuwers, L., & Strisciuglio, N. (2021). Deepfake Detection Using Capsule Networks and Long Short-Term Memory Networks. In *Proceedings of the International Con ference on Artificial Intelligence and Data Processing*. ISBN: 978-989-758-488-6.

[4] Kularkar, T., Jikar, T., Rewaskar, V., Dhawale, K., Thomas, A., & Madankar, M.

(2023). DeepfakeDetection Using LSTM and

ResNext. *International Journal of Computer Research and Technology*, 11(11), 231-239. ISSN: 2320-2882 .

[5] Zhang, L., Liu, Q., & Wang, X. (2023). A Comparative Study of Deepfake Detection Tech niques Using Convolutional Neural Networks. *Journal of Machine Learning and Artificial Intelligence*, 22(1), 59-75. DOI: 10.1109/JMLAI.2023.0092 .

[6] Killi, C.B.R., Balakrishnan, N., & Rao, C.S. (2023). Deep Fake Image Classification Using VGG-Model. *International Journal of Image Processing and Computer Vision*,

28(2), 509 515. Available at: http://iieta.org/journals/isi .

[7] Rupasri, D., Kumaran, M., & Lin Eby Chandra, J. (2023). Deepfake Detection Using Xcep tion and MobileNets Deep Learning Models. *International Journal of Advanced Research in Computer and Communication Engineering*, 12(9), 1420-1428. DOI: 10.17148/IJAR

CCE.2023.12916.

[8] Tiwari, P., Patil, A., & Yadav, V. (2023). An Improved GAN-Based Approach for Deep fake Detection. *Journal of Computer Vision and Image Processing*, 45(4), 345-356. DOI:

10.1016/j.jcvip.2023.