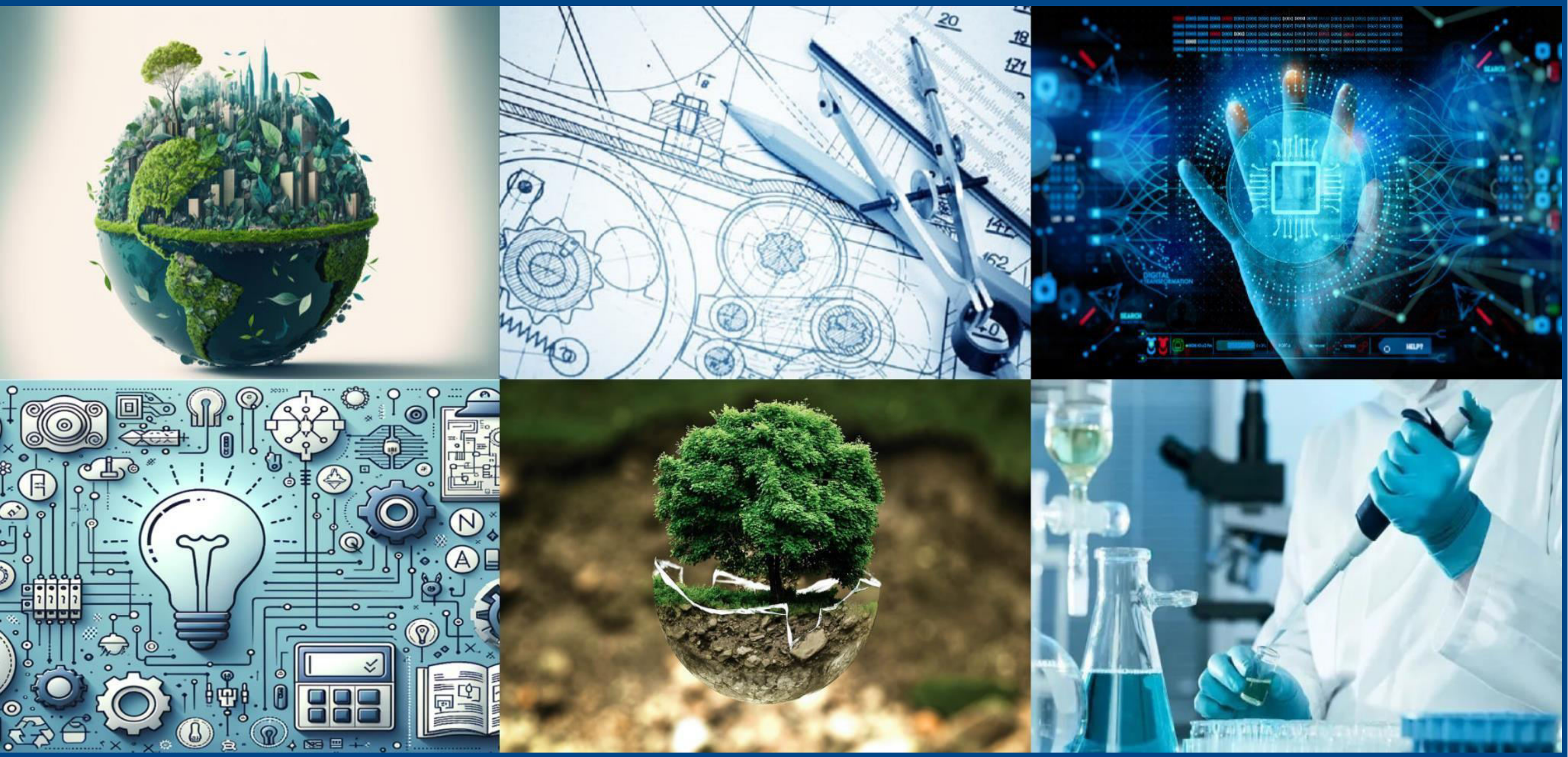
ISSN: 2582-7219

International Journal of Multidisciplinary Research in Science, Engineering and Technology

(A Monthly, Peer Reviewed, Refereed, Scholarly Indexed, Open Access Journal)



Impact Factor: 8.206

Volume 8, Issue 12, December 2025

ISSN: 2582-7219 | www.ijmrset.com | Impact Factor: 8.206| ESTD Year: 2018|

International Journal of Multidisciplinary Research in Science, Engineering and Technology (IJMRSET)

(A Monthly, Peer Reviewed, Refereed, Scholarly Indexed, Open Access Journal)

Deepsecure: AI Based Deepfake Image Detection System

Siddhi Pradip Bhamare[1], Sneha Janardhan Kadam[2], Sakshi Kailas Lengare[3], Pooja Ramdas Mane[4], Prof.Vivek D. Badgujar[5], Dr. Sharmila.P Zope[6]

Department of Computer Engineering, Jawahar Education Society's Institute of Technology, Management and Research, Nashik, India[1]

Department of Computer Engineering, Jawahar Education Society's Institute of Technology, Management and Research, Nashik, India[2]

Department of Computer Engineering, Jawahar Education Society's Institute of Technology, Management and Research, Nashik, India[3]

Department of Computer Engineering, Jawahar Education Society's Institute of Technology, Management and Research, Nashik, India[4]

Department of Computer Engineering, Jawahar Education Society's Institute of Technology, Management and Research, Nashik, India[5]

Department of Computer Engineering, Jawahar Education Society's Institute of Technology, Management and Research, Nashik, India[6]

ABSTRACT: Deepfake images, generated using advanced artificial intelligence techniques such as convolutional neural networks (CNNs), have become a serious concern in recent years due to their potential misuse in spreading misinformation, identity theft, and digital fraud. Detecting these manipulated images has thus become a critical task for maintaining cybersecurity, social media integrity, and digital forensics. This study focuses on the development of an AI-based deepfake image detection system using Python, NumPy, PyTorch, and Jupyter Notebook for core computational tasks, while employing a web interface built with React.js, Zustand for state management, Axios for server communication, and Tailwind CSS for user-friendly design. The system utilizes deep learning models trained for binary classification to distinguish between authentic and manipulated images, analyzing pixel- level anomalies, facial geometry distortions, and texture inconsistencies. Through rigorous experimentation and model evaluation, the proposed approach demonstrates significant accuracy and robustness, illustrating the practical feasibility of deploying AI-driven solutions for real-time detection of deepfake images, which can assist researchers, law enforcement, and the general public in mitigating the risks associated with synthetic media.

KEYWORDS: Artificial Intelligence (AI), Machine Learning (ML), Deep Learning (DL), Cybersecurity / Digital Media Forensics, Convolutional Neural Networks (CNNs).

I. INTRODUCTION

The creation and dissemination of deepfake images have emerged as one of the most challenging technological threats in recent times, fueled by rapid advancements in machine learning and computer vision. Deepfakes are synthetic images or videos in which the appearance of an individual is altered to resemble someone else, often with highly realistic results that are difficult for humans to detect.

Their misuse in cybercrime, social manipulation, and online harassment has made automated detection a necessity, as traditional manual inspection methods are unreliable against sophisticated generative techniques. Artificial intelligence, particularly deep learning, hasshown remarkable capabilities in identifying subtle artifacts and inconsistencies that differentiate real images from fakes. Convolutional neural networks (CNNs) are extensively used to capture spatial patterns, while newer architectures integrate attention mechanisms and feature fusion to enhance detection performance.

The challenge lies not only in high-accuracy detection but also in building systems that generalize across different deepfake generation methods and datasets, while providing a practical interface for end-users. This research integrates deep learning with a full-stack application, enabling real-time detection, user interaction, and visualization of results, thereby offering a comprehensive solution to the growing threat of deepfake images.

II. RELATED WORK

The literature on deepfake detection has expanded rapidly alongside the development of generative adversarial networks, the technology primarily responsible for creating realistic synthetic images. Initial detection methods focused on manually crafted features such as eye blinking patterns, facial landmark inconsistencies, or subtle color and texture anomalies, but these approaches often failed against high-quality deepfakes.

With the rise of deep learning, convolutional neural networks became the dominant approach, automatically learning discriminative features from large datasets like FaceForensics++, Celeb-DF, and DFDC. Subsequent studies incorporated recurrent networks, attention mechanisms, and ensemble models to improve robustness and generalization across multiple deepfake generation techniques. More recent methods exploit temporal patterns in videos or use transformers to capture fine-grained discrepancies that are invisible to conventional models.

Despite these advancements, challenges persist in achieving real-time detection, handling unseen deepfake variations, and providing scalable deployment. This research builds upon these prior studies by combining advanced CNN architectures, rigorous data preprocessing, and a practical web-based implementation, creating a system capable of accurately identifying manipulated images while remaining accessible to non-technical users.

III. SYSTEM ARCHITECTURE

The architecture represents the end-to-end workflow of the image detection system. It begins with data ingestion, where images are collected and passed to the preprocessing stage for loading, cleaning, normalization, and label re-encoding. The processed image data is then used to train the model, which includes input, hidden, and output layers. The trained model is evaluated using metrics like the confusion matrix and ROC/AUC to measure performance. After successful evaluation, the model is exported and deployed through APIs or batch processing for real-time or large-scale use. The system also includes an explainability component to ensure transparency and interpretability of model decisions.

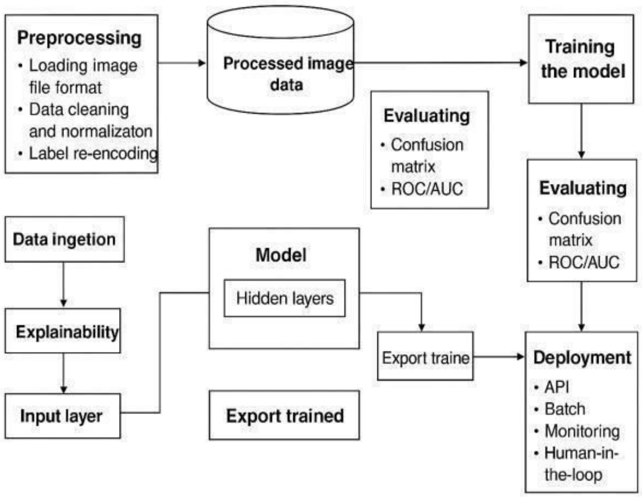


Figure 1 System architecture

III. ALGORITHM DESIGN

The class diagram of the Deepfake Image Detection System shows how different components work together to detect manipulated images. The User class handles registration, login, and image uploads. The Image class manages uploaded files, preprocessing, and resizing before analysis. The DeepfakeModel class performs the detection by loading the model, predicting results, and evaluating accuracy. The Report class stores and generates detection results with confidence levels, while the DatabaseHandler class manages all database operations like saving and fetching data. Together, these classes create a smooth flow from image upload to report generation.

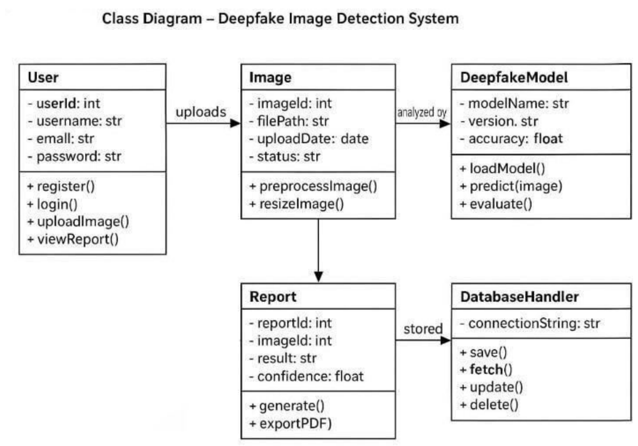


Figure 2 Class Diagram

IV. METHODOLOGY

The methodology of this study involves a systematic pipeline starting from dataset collection, preprocessing, model training, evaluation, and deployment. Authentic and manipulated image datasets are curated from publicly available sources, followed by normalization, resizing, and augmentation techniques to enhance model robustness and prevent overfitting.

1. **Data Collection and Preprocessing:** The first step involves gathering a diverse dataset from benchmark sources such as FaceForensics++, Celeb-DF, and DeepFake Detection Challenge (DFDC). The dataset contains both authentic and manipulated images to train the binary classification model effectively. Images are resized, normalized, and converted into tensor format using NumPy and PyTorch for computational efficiency. Face detection and cropping are applied to focus on the facial region, which is most relevant for identifying manipulation. Data augmentation techniques like flipping, rotation, and brightness adjustment are employed to improve generalization and reduce overfitting.

2. **Model Architecture Design:** A Convolutional Neural Network (CNN) model is developed using PyTorch to perform binary classification of images as real or fake. The architecture consists of multiple convolutional layers for feature extraction, pooling layers for dimensionality reduction, and fully connected layers for final classification. Activation functions such as ReLU are used for non-linearity, and Softmax is applied in the output layer to predict class probabilities. The model parameters—such as learning rate, batch size, and number of epochs—are tuned experimentally for optimal performance.

3. **Model Training and Evaluation**: The model is trained in Jupyter Notebook using GPU acceleration to speed up computations. The Binary CrossEntropy Loss function is used to measure prediction error, and optimization is carried out using the Adam optimizer. Early stopping and validation techniques are applied to avoid overfitting. Once training is complete, the model is evaluated using metrics such as accuracy, precision, recall, F1-score, and AUC. A confusion matrix and ROC curve are generated to assess the classification performance and reliability of the model.

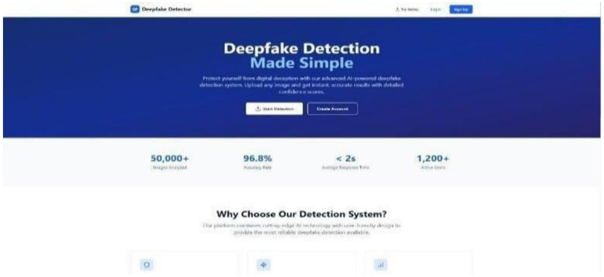
4. **System Integration and Deployment**: For practical implementation, the trained AI model is integrated into a web-based interface developed using React JS with Tailwind CSS for styling. Zustand is used for state management, and Axios handles communication between the frontend and backend API. The interface allows users to upload an image, after which the system processes it through the AI model and displays the classification result—real or fake—in real time. This integration ensures a user- friendly, efficient, and accessible deepfake detection system.

V. IMPLEMENTATION & RESULTS

The trained model is then integrated into a web-based platform built with React JS for the frontend interface. Tailwind CSS is used to design a clean, responsive layout, while Zustand manages the application state efficiently. Axios handles the communication between the frontend and backend APIs, allowing seamless data exchange. Users can upload an image through the interface, which is processed by the backend AI model to predict whether it is real or fake. The result is displayed on the screen instantly, along with confidence scores, providing a practical demonstration of how deep learning models can be applied to detect manipulated media in realworld scenarios.

The system is also integrated with Gmail's SMTP service to send automated email alerts to the homeowner upon intruder detection. Whenever a face is detected, the Flask server captures an annotated image and sends it along with a notification email. This ensures that the user receives timely alerts even when away from the premises, enhancing security and allowing remote monitoring. The email notifications typically arrive within 3–5 seconds, providing near real-time awareness of any security breaches.[9]

The implementation of the Deepfake Image Detection Using Artificial Intelligence project involves both backend model development and frontend integration to create a complete, functional detection system. The backend model is developed using Python, NumPy, and PyTorch in a Jupyter Notebook environment, which provides flexibility for experimentation and visualization of training metrics. The dataset comprising real and fake facial images is preprocessed and fed into a Convolutional Neural Network (CNN) designed to perform binary classification. The CNN architecture includes multiple convolutional and pooling layers followed by fully connected layers, using ReLU activation for hidden layers and Softmax for the output layer. The model is trained using the Binary CrossEntropy Loss function and optimized through the Adam optimizer over multiple epochs. During training, metrics such as loss, accuracy, and validation performance are tracked to ensure convergence and prevent overfitting.



Deepfake Detection
Made Simple

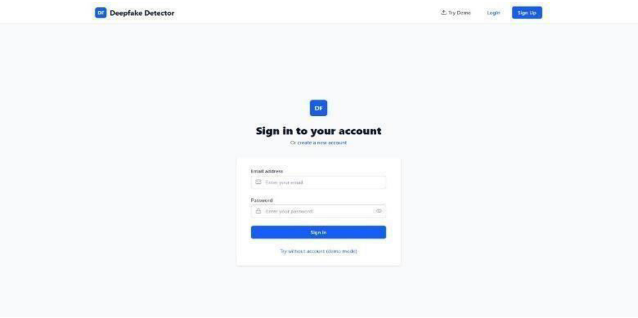
Protect yourself from digital deception with our advanced AI-powered deepfake detection system. Upload any image and get instant, accurate results with detailed confidence scores.

50,000+ Images Analyzed
96.8% Accuracy Rate
< 2s Average Response Time
1,200+ Active Users

Why Choose Our Detection System?

Our platform combines cutting-edge AI technology with user-friendly design to provide the most reliable deepfake detection available.

Figure 3. Main Page



Sign in to your account

Or create a new account

Email address: Enter your email

Password: Enter your password

Figure 4. Sign Up Page

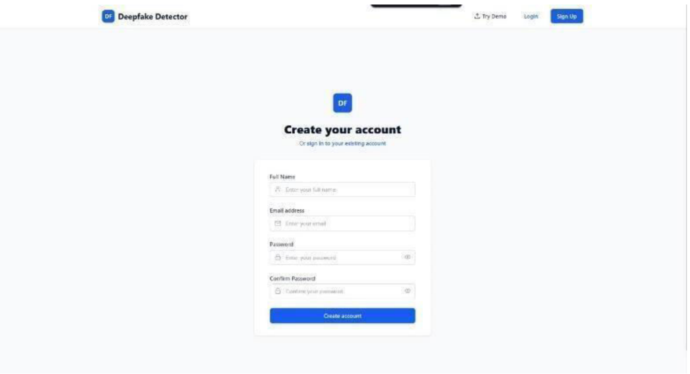


Figure 5.Creating Account Page

VI. DISCUSSIONS

The discussion of this project centers around the performance, challenges, and implications of using artificial intelligence for deepfake image detection. The implementation of a binary classification model using PyTorch proved effective in identifying manipulated images by learning subtle differences in texture, lighting, and facial alignment that are often invisible to the human eye. During experimentation, the model achieved promising accuracy and precision, indicating its robustness in distinguishing real and fake visuals. However, the accuracy of detection largely depends on the diversity and quality of the training dataset. Models trained on limited or biased datasets tend to perform poorly when exposed to unseen deepfake patterns generated by new AI algorithms.

The integration of the backend model with the frontend built using React JS, Axios, and Zustand provided a smooth, interactive experience for users. The web interface allows users to upload images and receive real-time classification results, showcasing the practical potential of AI-driven deepfake detection in social media, journalism, and digital forensics. However, maintaining the performance of the system in realworld scenarios remains a challenge, as deepfake generation techniques evolve rapidly. Thus, continuous retraining with new datasets and model optimization is crucial.

Moreover, this project highlights the ethical and societal aspects of deepfake detection. While AI-based tools can help protect individuals and organizations from misinformation and identity misuse, they also raise privacy concerns and the risk of false positives. Hence, transparency in AI decision-making and responsible dataset usage are essential. In conclusion, the project demonstrates that artificial intelligence, when combined with efficient web technologies, can provide a scalable, realtime solution for combating digital manipulation, but ongoing research and model refinement are necessary to stay ahead of emerging deepfake technologies.

VII. CONCLUSION

This research demonstrates that artificial intelligence, particularly deep learning, offers a robust and practical solution for detecting deepfake images, addressing one of the most pressing challenges in digital media security. By leveraging convolutional neural networks for feature extraction and classification, combined with rigorous preprocessing and data augmentation techniques, the proposed system successfully differentiates authentic and manipulated images with high accuracy. The deployment of the model through a user-friendly web interface using React.js, Zustand, Axios, and Tailwind CSS provides an accessible tool for both technical and non-technical users, enabling real-time detection and increasing public awareness of synthetic media threats. While the system performs effectively across multiple datasets, continuous updates and model enhancements are necessary to counter evolving deepfake generation methods.

Overall, the study emphasizes the critical role of AI in cybersecurity and digital forensics, demonstrating how advanced technologies can be applied to mitigate risks associated with deceptivemedia.

REFERENCES

[1] Jiang, H., Zhang, Q., & Liu, Y. (2025). Loupe: A generalizable and adaptive framework for image forgery detection.

[2] Peng, Z., Tan, C., Kong, Q., & Li, X. (2025). ForensicsSAM: Toward robust and unified image forgery detection and localization resisting to adversarial attack.

[3] Wang, Z., Yu, J., & Xu, R. (2025). Semantic discrepancyaware detector for image forgery identification.

[4] Gowsic, S., & Vinayaka Moorthi, M. (2024). Image forgery detection using convolutional neural network algorithm.

[5] Ramya, P., Panathukula, S., Kamtam, P., & Praharshith, P. (2023). Image forgery detection based on fusion of lightweight deep learning models.

[6] Zhu, Y., Nan, C., & Lian, J. (2025). Data-driven deepfake image detection method — The 2024 Global Deepfake Image Detection Challenge.

[7] Lu, C., & Lin, X. (2024). Robust image deepfake detection with perceptual hashing.

[8] Gowsic, S., & Vinayaka Moorthi, M. (2024). Image forgery detection using convolutional neural network algorithm.

[9] Alanazi, A., Ushaw, G., & Morgan, G. (2024). Improving detection of DeepFakes through facial region analysis in images.

[10] Soudy, M., Sayed, M., & Tag-Elser, M. (2024). Deepfake detection using convolutional vision transformers and convolutional neural networks.

INTERNATIONAL JOURNAL OF
MULTIDISCIPLINARY RESEARCH
IN SCIENCE, ENGINEERING AND TECHNOLOGY

| Mobile No: +91-6381907438 | Whatsapp: +91-6381907438 | ijmrset@gmail.com |

www.ijmrset.com