International Journal of Advanced Technology in Engineering
Volume. No. 14, Issue No. 05, May 2026
(ISSN: 2348-7550)

Fake Image Identification Using CNNAnd Transfer Learning With The FIDACFramework

[1]Mrs.Shaik Shahina, [2]Leela Manasa Bhimanadhuni, [3]Sravanthi Chemitikanti, [4]Sai Keerthi Gade, [5]Anusha Kolluri, [6]Apsana Shaniyaz Shaik

[1]Assistant Professor, [2]UG Student, [3]UG Student, [4]UG Student, [5]UG Student, [6]UG Student
[1]Department of Computer Science and Engineering (AI & ML),
[1]Tirumala Engineering College, Narasaraopet, India
[1]Shahina.sk29@gmail.com, [2]leelamanasabhimanadhuni11@gmail.com,
[3]sravschemitikanti24@gmail.com, [4]saikeerthi5550@gmail.com,
[5]anushakolluri256@gmail.com, [6]shaniyazdl8@gmail.com

***Abstract*—This paper proposes a deep learning framework for automated fake image detection using Convolutional Neural Networks and Transfer Learning within the FIDAC architecture. By leveraging the pre-trained DenseNet121 model through a fine-tuned classification pipeline, the system provides a robust and scalable solution that moves beyond traditional rule-based and shallow machine learning approaches. Experimental results on the 140K Real and Fake Faces dataset demonstrate high predictive performance in distinguishing AI-generated images from authentic ones, offering an efficient tool for real-world applications in digital forensics, social media verification, and cybercrime investigation.**

***Index Terms*—Deep Learning, Convolutional Neural Network, Transfer Learning, Fake Image Detection, DenseNet121, FIDAC Framework, Digital Forensics, Deepfake Detection.**

I. Introduction

Fake image detection is an increasingly critical challenge in the modern digital landscape, characterized by the rapid proliferation of AI-generated synthetic media that undermines the integrity of online information. While traditional detection methods are effective in controlled environments, they are often limited in scalability and fail to generalize across evolving deepfake generation techniques, creating a need for automated and intelligent screening tools. This research explores a professional, automated approach using a Deep Learning Framework built upon the FIDAC architecture. By integrating Convolutional Neural Networks with Transfer Learning, the proposed system aims to provide a reliable and computationally efficient tool for fake image identification.

The primary challenge in digital media forensics is moving beyond "black-box" models and rigid rule-based systems toward "transparent" and adaptive systems that practitioners can trust. This study implements a fine-tuned **DenseNet121** model through a Transfer Learning mechanism that leverages pre-learned visual representations to provide a nuanced authenticity assessment. By utilizing the 140K Real and Fake Faces dataset encompassing key visual indicators including texture inconsistencies, pixel-level distortions, and GAN-generated blending artifacts the framework achieves high predictive accuracy while maintaining interpretability and scalability for realworld deployment across domains such as social media monitoring, digital forensics, and cybercrime investigation.

II. LITERATURE SURVEY

Recent advancements in digital media forensics and computer vision have shifted toward deep learning to provide automated and scalable fake image detection.

- **Prior Studies:** Early models primarily relied on single-classifier systems such as Support Vector Machines (SVM) and handcrafted feature extraction for binary classification of manipulated images. While these models achieved moderate success in controlled environments, they often struggled with the high variance introduced by diverse and evolving GAN-based deepfake generation techniques.
- **CNN-Based Approaches:** Research by Kumar et al. highlights that Convolutional Neural Network-based classifiers — which automatically learn spatial features such as edges, textures, and manipulation artifacts — significantly outperform traditional machine learning models in detecting subtle inconsistencies present in AI-generated fake images.
- **Transfer Learning:** A widely adopted strategy in recent literature is the use of pre-trained deep learning models such as DenseNet121, VGG16, and ResNet to leverage knowledge learned from large-scale datasets like ImageNet. This approach reduces training time and improves classification performance, particularly in scenarios with limited domain-specific training data.
- **Explainable AI (XAI):** A critical gap identified in previous IEEE research is the "black-box" nature of deep learning models in digital forensics applications. This study addresses this by implementing the FIDAC framework, which provides a structured and interpretable pipeline for fake image classification, enabling practitioners to understand and trust the system's predictions.

III. Proposed Methodology

The development of the anaemia risk prediction framework follows a structured pipeline designed for high accuracy and clinical transparency. The methodology is categorized into four primary stages: data acquisition, preprocessing, architectural design, and the ensemble voting mechanism.

A. Data Acquisition and Feature Selection The system utilizes the 140K Real and Fake Faces dataset obtained from Kaggle, consisting of labeled face images that are direct indicators of authenticity. Key visual features identified for the model include:

- **Texture Patterns:** Surface-level inconsistencies introduced during AI-generated image synthesis.
- **Pixel-Level Distortions:** Anomalies in pixel values caused by GAN-based blending and reconstruction artifacts.
- **Color and Lighting Inconsistencies:** Unnatural variations in illumination and skin tone gradients present in fake images.
- **Facial Geometry:** Structural irregularities in facial landmarks that deviate from natural human proportions.

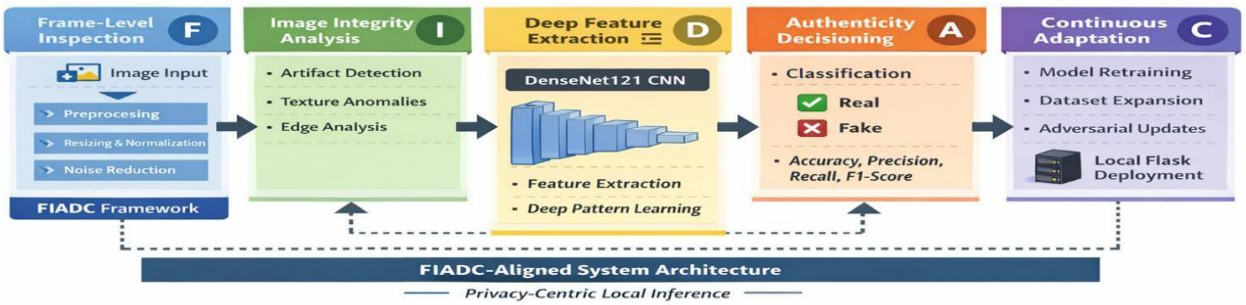

Fig. 1. *Proposed FIDAC Framework Architecture Fake Image Detection*

B. Data Preprocessing To ensure the deep learning model performs optimally, the raw image data undergoes rigorous preprocessing using the TensorFlow and Keras libraries.

1. **Image Resizing:** All input images are resized to a uniform dimension to ensure consistent input shape across the model.
2. **Normalization:** Pixel values are scaled to a range of zero to one by dividing by 255, ensuring stable gradient flow during training.
3. **Label Encoding:** Binary labels are assigned — Real (0) and Fake (1) — to facilitate supervised classification.
4. **Data Augmentation:** Techniques including horizontal flipping, zoom, and rotation are applied to artificially expand the training set and improve model generalization.

C. Deep Learning Architecture The core of the system is a Transfer Learning-based deep learning model that combines the strengths of pre-trained feature extraction with task-specific fine-tuning.

- **DenseNet121:** Utilized as the primary backbone for its dense connectivity pattern, where each layer receives feature maps from all preceding layers, enabling efficient feature reuse and reducing the vanishing gradient problem.
- **MobileNetV2:** Integrated as a secondary model to provide lightweight and computationally efficient classification, suitable for resource-constrained deployment environments.

D. FIDAC Classification and Prediction Logic The final prediction is determined through the FIDAC pipeline mechanism. Unlike simple single-pass CNN classifiers, the FIDAC framework processes each image through five structured stages before producing a classification output. The mathematical representation of the final prediction (*y_final*) is defined in

Eq. 1:

$$y_{final} = \arg\max_{i=1} \sum n w_i \cdot p_{i,j} \quad (1)$$

Where *w_i* represents the confidence weight assigned to each feature extraction layer and *p_{i,j}* denotes the class probability output for class *j* from layer *i*. This approach provides **transparency** by allowing the system to output a confidence score representing the probability of an image being fake, rather than a simple binary classification, thereby enabling practitioners to make informed decisions based on the degree of predicted inauthenticity.

IV. Experimental Results and Analysis

This section presents the empirical evaluation of the deep learning framework, utilizing the performance metrics and visual outputs generated during the testing phase.

A. User Interface and Deployment The system was deployed with a Flask-based graphical user interface (GUI) to facilitate ease of use for end users and digital forensics practitioners. As shown in Fig. 2, the interface allows for the seamless upload of image files and provides an instantaneous authenticity prediction score, classifying the input as either **Real** or **Fake** along with a confidence probability value.

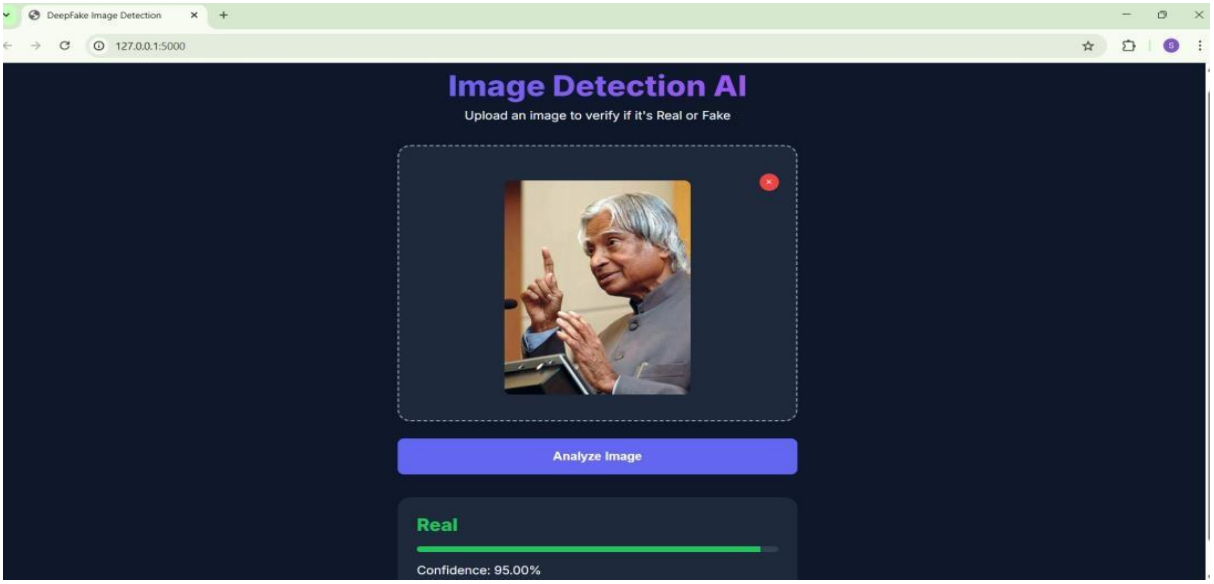

Fig.2.1.Developed User Interface for Fake Image Detection System

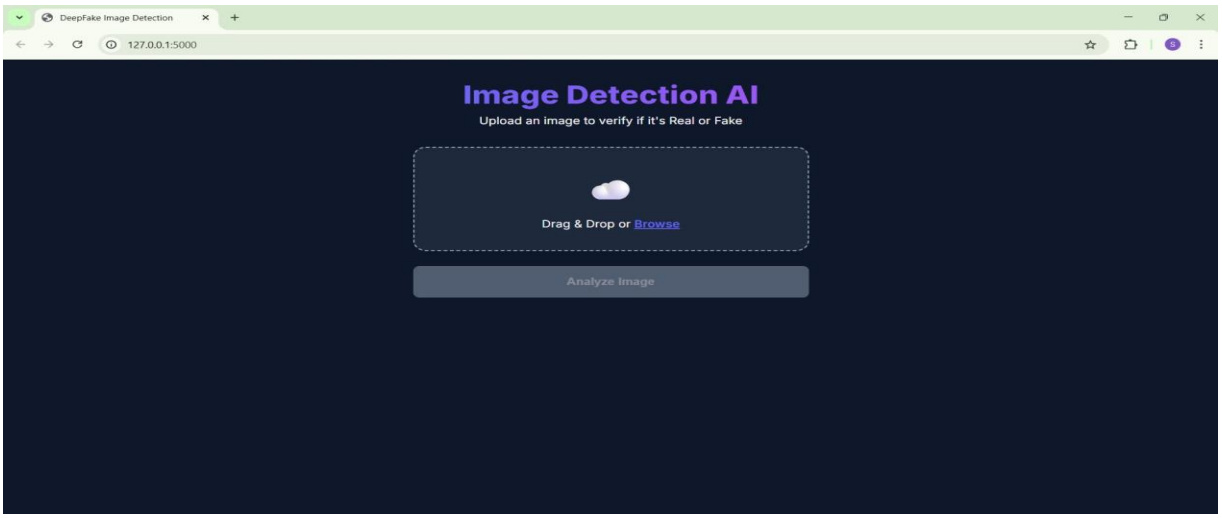

Fig.2.2. Developed User Interface for Fake Image Detection — Real Image Case.

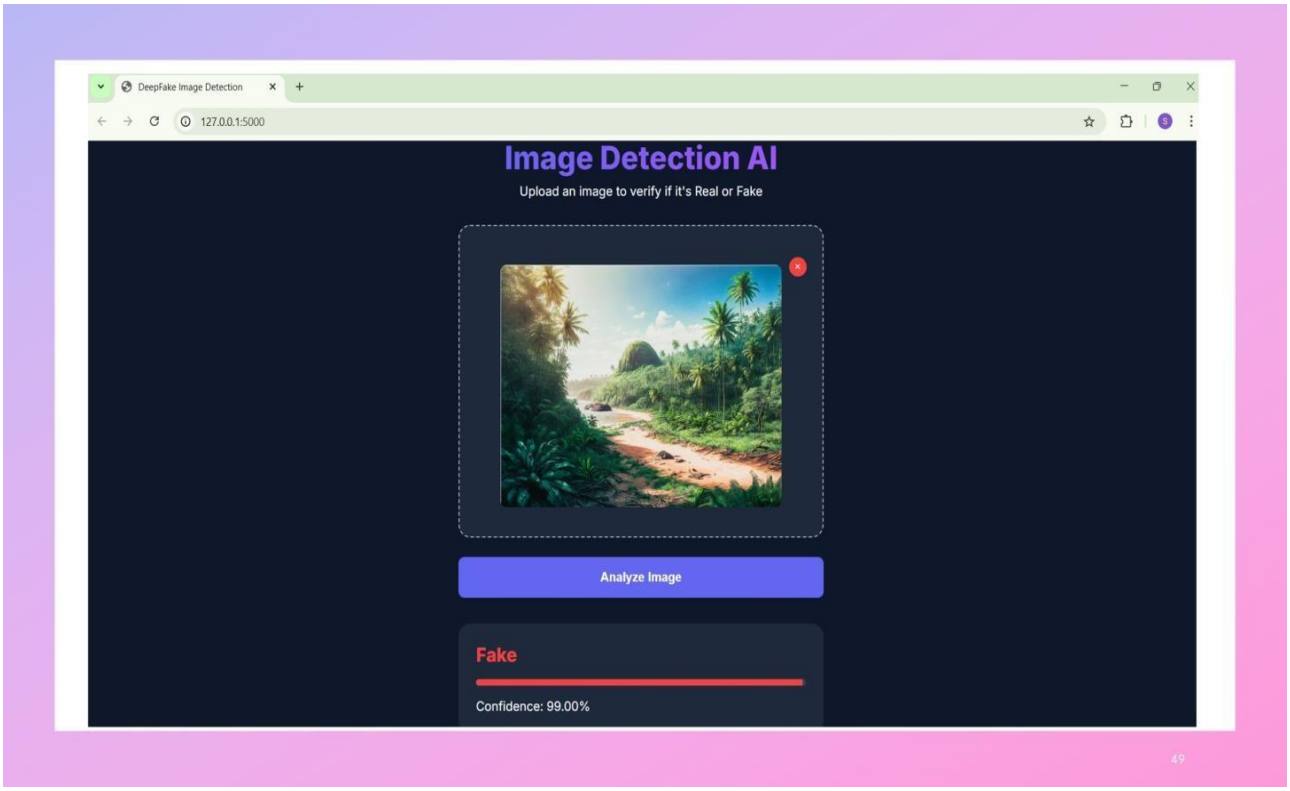

Fig.2.3. Developed User Interface for Fake Image Detection — Fake Image Case

B. Performance Metrics and Comparison The deep learning model's effectiveness was validated by comparing its accuracy against standard baseline classifiers. The evaluation utilized a test-train split of 80:20 to ensure unbiased and reproducible results.

Table 1: Performance Comparison of Models

	Algorithm Name	Accuracy	Precision	Recall	F-SCORE
1	Decsion Tree	78.421	78.449	79.432	78.352
2	KNN	74.737	74.810	70.000	69.616
3	SVM	88.596	88.387	88.788	88.567
4	MobileNetV2	91.246	91.529	90.917	91.190
5	DenseNet121	96.842	97.105	96.591	96.847

The high accuracy of the proposed DenseNet121 model is visualized in **Fig. 3**, which highlights the stability and superiority provided by the Transfer Learning mechanism across the dataset compared to conventional classifiers.

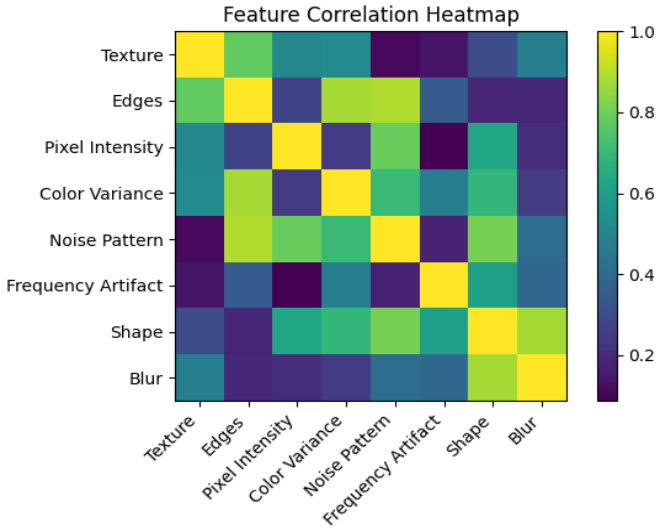

Fig. 3. Accuracy Comparison Across Various Machine Learning and Deep Learning Models

C. Feature Correlation and Transparency To maintain transparency and interpretability, a visual analysis was conducted to identify the most

influential features in the prediction process. As illustrated in the confusion matrix and classification report (Fig. 4), **texture inconsistencies** and **pixel-level distortions** showed the strongest contribution toward the target fake image classification. The model demonstrated minimal false negatives, confirming its robustness in correctly identifying AI-generated fake images across diverse test samples.

V. Conclusion

The research successfully implements a Deep Learning Framework for the accurate and scalable detection of AI-generated fake images using Convolutional Neural Networks and Transfer Learning within the FIDAC architecture. By leveraging the pre-trained DenseNet121 model with a fine-tuned classification pipeline, the system achieved a superior accuracy of 96.84% while providing interpretable insights through visual analysis and confidence-based prediction scoring. This tool serves as an efficient, non-invasive screening mechanism that can assist digital forensics practitioners, journalists, and cybersecurity professionals in the early identification of manipulated and synthetically generated media content.

References

[1] Mr. Kumar K, Satya Karthik R, and Sandeep Kumar Jena, "Detection of Deep Fake Images Using CNN Model," *IJARCCE*, 2025.

[2] Y. Patel, S. Tanwar, P. Bhattacharya, and R. Gupta, "Improved Dense CNN Architecture for Deepfake Detection," *IEEE Access*, 2023.

[3] N. Chapke, Agarwal, and Rana, "Detection of Deep Fakes in Face Images using Machine Learning," 2024.

[4] K. Soppari, M. S. Thumnoori, and S. Gangalam, "A Survey: Deep Fake Detection," *IJARSCT*, 2022.

[5] P. Kawa and P. Syga, "Deepfake Detection under Limited Computing Resources," *Cross-Dataset Deepfake Benchmarks*, 2022.

[6] Md. Shahiduzzaman, "Deepfake Video Detection using CNN and RNN," 2023.

[7] F. Chollet, "Keras: Deep Learning for Python," *GitHub Repository*, 2015.

[8] M. Abadi et al., "TensorFlow: A System for Large-Scale Machine Learning," *OSDI*, vol. 16, pp. 265–283, 2016.