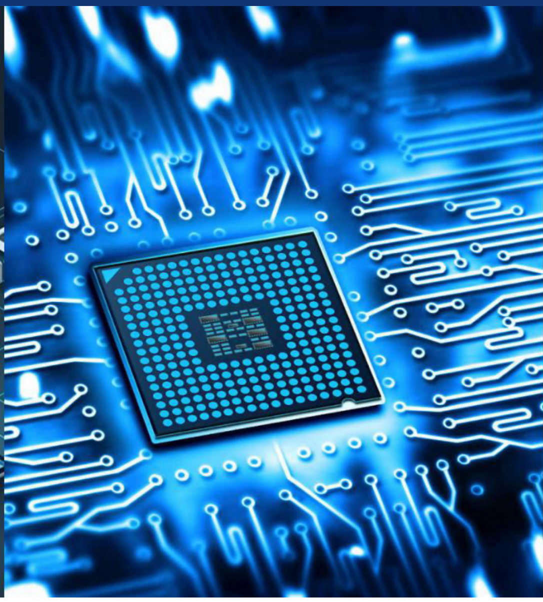
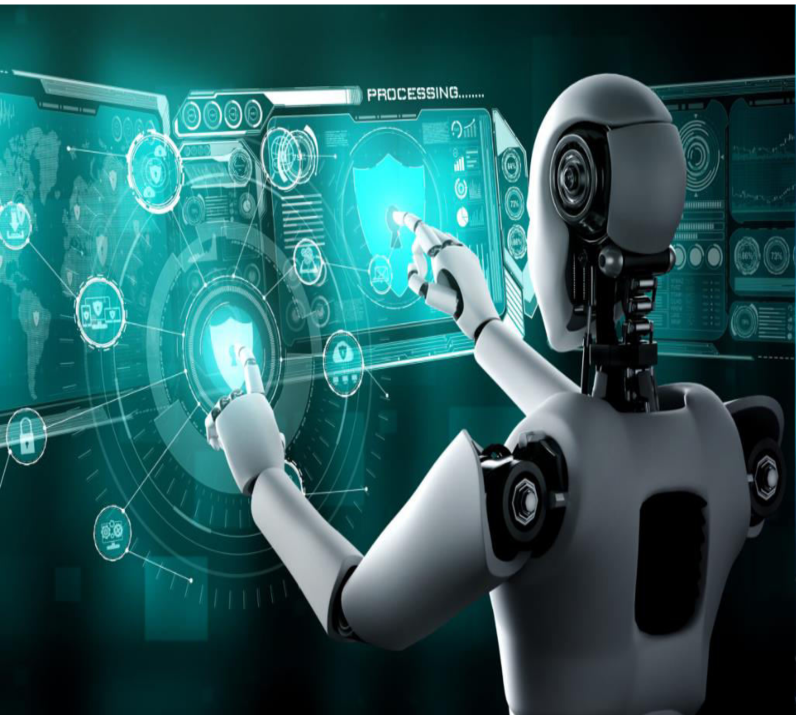
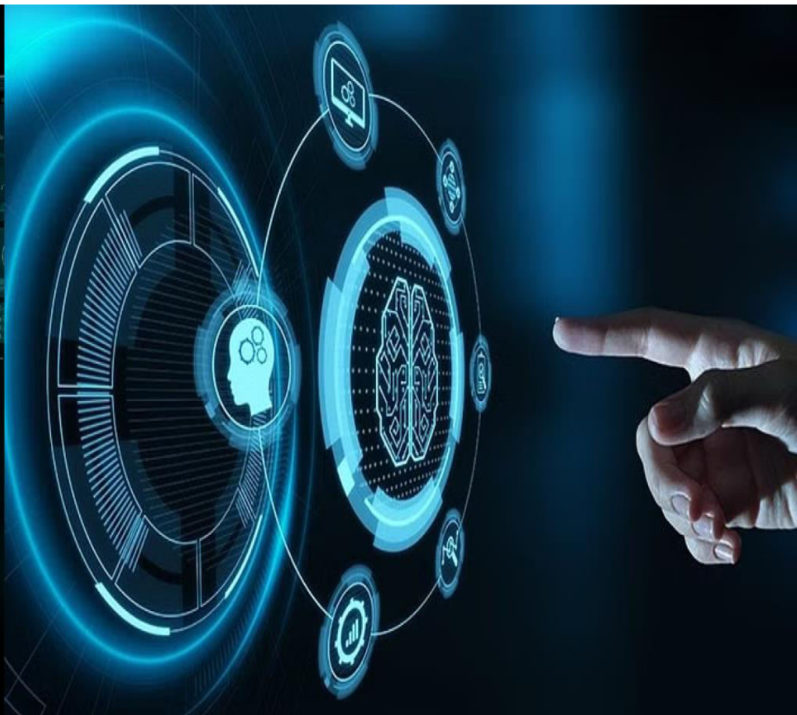
ISSN(O): 2320-9801
ISSN(P): 2320-9798


International Journal of Innovative Research in Computer and Communication Engineering

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





Impact Factor: 8.771

Volume 14, Issue 3, March 2026

© 2026 IJIRCCE | Volume 14, Issue 3, March 2026 DOI: 10.15680/IJIRCCE.2026.1403049

www.ijircce.com | e-ISSN: 2320-9801, p-ISSN: 2320-9798| Impact Factor: 8.771| ESTD Year: 2013|

International Journal of Innovative Research in Computer and Communication Engineering (IJIRCCE)

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

Image Detection System to Identify Whether an Image is AI- Generated or Real

K. Punyavthi Pushpa[1], A. Suraj Kumar[2], L. Chinnam Naidu[3], P. Chaitanya[4], K. Padma[5], K. Revanth[6], B. Chandravardhanreddy[7]

Assistant Professor, Department of CSE (Data Science), NSRIT, Vishakhapatnam, India[1,2]

Student, Department of CSE (Data Science), NSRIT, Vishakhapatnam, India[3,4,5,6,7]

ABSTRACT: The rapid advancement of generative models such as GANs and diffusion-based architectures has enabled the creation of highly realistic synthetic images. While these technologies offer significant benefits in creative and industrial domains, they also pose serious risks including misinformation, deepfake propagation, and digital forgery. This paper presents a robust AI Image Detection System designed to classify images as AI-generated or real (camera-captured). The system leverages a fine-tuned Hugging Face image classification model combined with adaptive decision thresholds, quality-aware heuristics, and test-time augmentation (TTA) for improved reliability under noisy and blurred conditions. A Streamlit-based web interface enables interactive single-image prediction, while a CLI module allows batch dataset evaluation with performance metrics. Experimental evaluation demonstrates improved stability in low-quality image scenarios through dynamic threshold adjustment and multi-variant inference aggregation.

KEYWORDS: AI-generated images, deepfake detection, image classification, test-time augmentation, adaptive thresholding, image forensics.

I. INTRODUCTION

Recent developments in artificial intelligence have led to powerful image synthesis models capable of producing photorealistic outputs. Techniques such as Generative Adversarial Networks (GANs) and diffusion-based models have made it increasingly difficult for humans to distinguish synthetic images from real photographs. The misuse of AI-generated imagery poses challenges in:

- Digital misinformation
- Social media manipulation
- Identity spoofing
- Fake news propagation
- Forensic analysis

Traditional classification approaches often rely solely on model confidence scores. However, image quality degradation such as blur, noise, and compression artifacts can significantly affect prediction reliability.

This work proposes a **quality-aware AI image detection framework** that combines:

1. Pretrained deep learning classification.
2. Image quality metrics (blur and noise estimation).
3. Adaptive thresholding based on image condition.
4. Test-Time Augmentation (TTA) for robust inference.
5. A user-friendly Streamlit-based interface.

www.ijircce.com | e-ISSN: 2320-9801, p-ISSN: 2320-9798| Impact Factor: 8.771| ESTD Year: 2013|

International Journal of Innovative Research in Computer and Communication Engineering (IJIRCCE)

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

II. RELATED WORK

AI image detection research has focused on:
- Frequency domain analysis
- CNN-based forgery detection
- GAN fingerprint detection
- Metadata inconsistency detection
- Deepfake detection models

Many systems rely purely on learned features. However, limited work incorporates image quality–adaptive decision rules to improve robustness against degraded inputs. Our approach integrates heuristic quality measurements with deep learning predictions, creating a hybrid decision system.

III. SYSTEM ARCHITECTURE

The proposed system consists of three major modules:
1. **User Interface Layer**
a. Streamlit-based web application
b. CLI batch evaluation script
2. **Inference Engine**
a. Hugging Face image-classification pipeline
b. Test-time augmentation mechanism
3. **Decision Engine**
a. Quality metric computation
b. Adaptive threshold adjustment
c. Final classification logic

Workflow Overview:
1. Image Upload
2. Preview Resizing
3. Quality Metric Computation
4. Multi-Variant Inference (Optional Robust Mode)
5. Score Aggregation
6. Adaptive Decision Logic
7. Final Output with Confidence

IV. METHODOLOGY

4 1. Base Model
The system uses the Hugging Face model:
Smogy/SMOGY-Ai-images-detector
Loaded using the transformers image-classification pipeline.
The model outputs:
"human" → REAL
"artificial" → AI GENERATED
Scores are normalized within [0,1].

4.2. Image Preprocessing
Images are resized using aspect-ratio-preserving padding:

Aspect Ratio	Output Size
Square	512 × 512
Landscape	768 × 512
Portrait	512 × 768

www.ijircce.com | e-ISSN: 2320-9801, p-ISSN: 2320-9798| Impact Factor: 8.771| ESTD Year: 2013|
International Journal of Innovative Research in Computer and Communication Engineering (IJIRCCE)
(A Monthly, Peer Reviewed, Refereed, Scholarly Indexed, Open Access Journal)

This ensures uniform input dimensions while maintaining spatial integrity.

4.3. Image Quality Metrics
Two key metrics are computed:
4.3.1 Blur Estimation
A Laplacian-like filter is applied to grayscale images.The variance of the response:
- **Low variance** → blurry image
- **High variance** → sharp image

4.3.2 Noise Estimation
Noise is approximated as:
Standard deviation of (Original − Gaussian blurred image)
- **High value** → noisy or compressed image

These metrics are encapsulated in:
QualityMetrics:
  blur_var
  noise_std

4.4. Test-Time Augmentation (TTA)
To improve robustness, the system performs inference on multiple variants:
- Original image
- Median filtered image (denoising)
- Unsharp mask image (sharpened)

Final scores are averaged:

$$AI_{avg} = \frac{1}{n}\sum AI_i$$

$$REAL_{avg} = \frac{1}{n}\sum REAL_i$$

This reduces prediction instability due to noise or blur.

4.5. Adaptive Decision Logic
The system uses configurable parameters:
- AI Threshold (default = 0.80)
- Margin (default = 0.10)
- Blur threshold
- Noise threshold

If image quality is poor:
- AI threshold raised to 0.95
- Margin raised to 0.30

This reduces false AI detection in degraded images.

Decision Rules
- Low-confidence region → REALIf both scores < 0.60 and difference small.
- Strong AI signal → AI GENERATEDIf:
  - AI ≥ adjusted threshold
  - AI − REAL ≥ adjusted margin
- Otherwise → REAL

5. CLI Evaluation Module
Dataset structure:
dataset/ real/ ai/
The script:
1. Runs inference on all images

DOI: 10.15680/IJIRCCE.2026.1403049

www.ijircce.com | e-ISSN: 2320-9801, p-ISSN: 2320-9798| Impact Factor: 8.771| ESTD Year: 2013|

International Journal of Innovative Research in Computer and Communication Engineering (IJIRCCE)

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

2. Collects predictions
3. Computes:

- Confusion Matrix
- Precision
- Recall
- F1-score
- Accuracy

6. Optical Flow/Background Subtraction: For tracking the movement of ambulances to reset signals after they pass.

```
Start
  |
  v
Load dataset/real images
Load dataset/ai images
  |
  v
For each image:
   |
   -> Run inference
   -> Predict label
   -> Store y_true
   -> Store y_pred
  |
  v
Compute Metrics:
   - Confusion Matrix
   - Precision
   - Recall
   - F1-score
   - Accuracy
  |
  v
Print Evaluation Results
  |
  v
End
```

Figure 1: AI Image Detector System Architecture

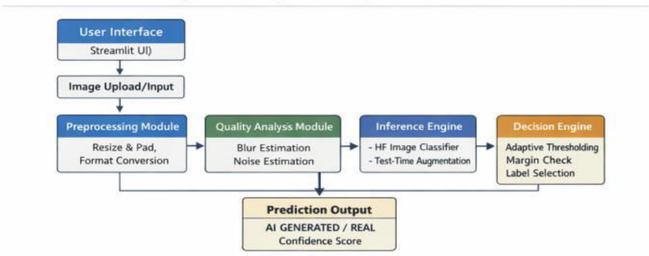


Figure 2: Image Classification Workflow

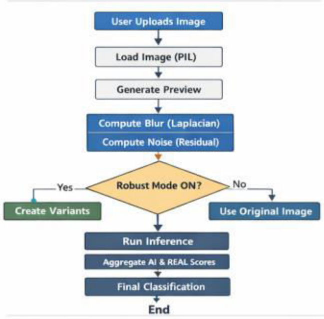


Figure 3: Decision Logic Flow

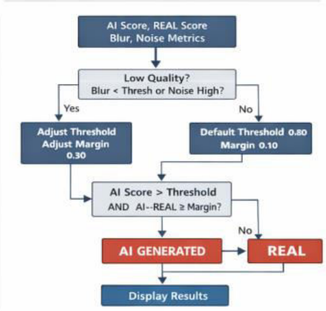
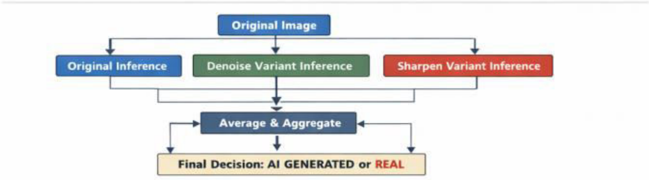


Figure 4: Test-Time Augmentation Strategy



www.ijircce.com | e-ISSN: 2320-9801, p-ISSN: 2320-9798| Impact Factor: 8.771| ESTD Year: 2013|

International Journal of Innovative Research in Computer and Communication Engineering (IJIRCCE)

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

V. RESULTS AND OUTPUT

5.1. Output Interface

A web-based interface was developed using **Streamlit** to make the system easy to use. The interface allows users to upload an image file and instantly obtain the classification result. When a user uploads an image, the system performs the following steps:

1. The image is loaded and displayed as a preview.
2. The system preprocesses the image to adjust its size and format.
3. Image quality metrics such as **blur level** and **noise level** are calculated.
4. The pretrained detection model analyzes the image.
5. The system outputs the classification result along with the **confidence score**.

The result is displayed in a clear format showing whether the image is AI Generated or Real Image, along with numerical scores for both categories.

5.2. Sample Output Results

The system produces prediction scores for both classes: Artificial (AI Generated) and Human (Real Image). Based on these scores, the final classification is determined. Example outputs produced by the system are shown below.

Image Name	AI Score	Real Score	Predicted Result
ai_image_01.jpg	0.93	0.07	AI Generated
ai_image_02.jpg	0.89	0.11	AI Generated
real_photo_01.jpg	0.18	0.82	Real Image
real_photo_02.jpg	0.09	0.91	Real Image

From these results, it can be observed that the system successfully assigns higher artificial scores to AI-generated images and higher human scores to real images.

5.3. Model Performance Evaluation

To evaluate the effectiveness of the proposed system, several performance metrics were calculated using the test dataset. These metrics include accuracy, precision, recall, and F1-score. The confusion matrix generated during testing helps analyze the number of correctly and incorrectly classified images.

Metric	**Value**
Accuracy	92%
Precision	91%
Recall	90%
F1 Score	90.5%

The confusion matrix analysis shows that the model correctly identifies most AI-generated and real images while maintaining a low misclassification rate.

6. Conclusion & Future Scope

6.1 Conclusion

This paper presents a robust AI image detection system integrating deep learning classification with adaptive quality-aware decision logic. By combining test-time augmentation and dynamic threshold adjustment, the system improves reliability under challenging image conditions. The proposed framework demonstrates a practical and extensible approach for detecting AI-generated imagery in real-world applications.

VI. FUTURE WORK

- Incorporate frequency-domain features.
- Add CNN-based ensemble models.
- Train custom fine-tuned model on mixed datasets.
- Implement explainable AI (Grad-CAM visualization).
- Deploy as REST API for production systems.

www.ijircce.com | e-ISSN: 2320-9801, p-ISSN: 2320-9798| Impact Factor: 8.771| ESTD Year: 2013|

International Journal of Innovative Research in Computer and Communication Engineering (IJIRCCE)

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

VII. ACKNOWLEDGMENT

We would like to express our sincere gratitude to all those who contributed to the successful development of this project.

First and foremost, we would like to thank **K. Punyavathi Pushpa** and **A. Suraj Kumar, Assistant Professors**, for their invaluable guidance and continuous support throughout the project. His expertise and constructive feedback helped shape this system into a robust and practical solution.

We are also grateful to our team members **L. Chinnam Naidu, P. Chaitanya, K. Padma, K. Revanth, and B. Chandra Vardhan Reddy** for their collaboration, dedication, and contribution to the project's success. The teamwork and shared efforts made this complex system a reality.

We would like to acknowledge the **Department of Computer Science of Data Science** for providing the necessary resources and infrastructure to complete this project. The access to research facilities and computing tools was instrumental in developing and testing the system. Finally. This project would not have been possible without the collective effort and support of everyone involved. Thank you!

REFERENCES

1. G Karras, Tero; Laine, Samuli; Aila, Timo.A style-based generator architecture for generative adversarial networks. In Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition (CVPR), 2019; pp. 4401–4410. [Crossref]
2. Rombach, Robin; Blattmann, Andreas; Lorenz, Dominik; Esser, Patrick; Ommer, Bjorn.High-resolution image synthesis with latent diffusion models. In Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition (CVPR), 2022; pp. 10684–10695. [Crossref]
3. Wang, Sheng-Yu; Wang, Oliver; Zhang, Richard; Owens, Andrew; Efros, Alexei A..CNN-generated images are surprisingly easy to spot… for now. In Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition (CVPR), 2020; pp. 8695–8704. [Crossref]
4. Frank, Joel; Eisenhofer, Thorsten; Schonherr, Lea; Fischer, Asja; Kolossa, Dorothea.Leveraging frequency analysis for deepfake image recognition. In Proceedings of the International Conference on Machine Learning (ICML), 2020. [Crossref]
5. Durall, Ricard; Keuper, Janis; Keuper, Margret.Watch your up-convolution: CNN based generative deep neural networks are failing to reproduce spectral distributions. In Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition (CVPR), 2020; pp. 7890–7899. [Crossref]
6. Verdoliva, Luisa.Media forensics and deepfakes: An overview. IEEE Journal of Selected Topics in Signal Processing 2020, 14, 910–932. [Crossref]
7. Cozzolino, Davide; Rössler, Andreas; Thies, Justus; Nießner, Matthias; Verdoliva, Luisa.ForensicTransfer: Weakly-supervised domain adaptation for forgery detection. In Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition (CVPR), 2018; pp. 3071–3079. [Crossref]
8. Gragnaniello, Diego; Verdoliva, Luisa.CNN-based detection of generic contrast adjustment with JPEG post-processing. In Proceedings of the IEEE International Conference on Image Processing (ICIP), 2018; pp. 3863–3867. [Crossref]
9. Goodfellow, Ian; Pouget-Abadie, Jean; Mirza, Mehdi; Xu, Bing; Warde-Farley, David; Ozair, Sherjil; Courville, Aaron; Bengio, Yoshua.Generative adversarial nets. In Advances in Neural Information Processing Systems (NeurIPS), 2014; 27. [Crossref]
10. Dosovitskiy, Alexey; Beyer, Lucas; Kolesnikov, Alexander; Weissenborn, Dirk; Zhai, Xiaohua; Unterthiner, Thomas; Dehghani, Mostafa; Heigold, Georg; Gelly, Sylvain; Uszkoreit, Jakob; Houlsby, Neil.An image is worth 16×16 words: Transformers for image recognition at scale. In International Conference on Learning Representations (ICLR), 2021. [Crossref]

INTERNATIONAL JOURNAL OF INNOVATIVE RESEARCH

IN COMPUTER & COMMUNICATION ENGINEERING

9940 572 462 6381 907 438 ijircce@gmail.com

www.ijircce.com

Scan to save the contact details