iJRASET
International Journal For Research in Applied Science and Engineering Technology



INTERNATIONAL JOURNAL FOR RESEARCH

IN APPLIED SCIENCE & ENGINEERING TECHNOLOGY

Volume: 14 **Issue:** X **Month of publication:** October 2026

DOI: https://doi.org/10.22214/ijraset.2026.85559

www.ijraset.com

Call: 08813907089 | E-mail ID: ijraset@gmail.com

International Journal for Research in Applied Science & Engineering Technology (IJRASET)
ISSN: 2321-9653; IC Value: 45.98; SJ Impact Factor: 7.538
Volume 14 Issue X Oct 2026- Available at www.ijraset.com

AI-Generated Image Detection Using a Multi-Model Ensemble and Forensic Analysis

R. Vijayalakshmi[1], Lakshmi K[2], Mohana D[3], Nivetha R[4]
Department of Computer Science and Engineering, Adhiyamaan College of Engineering, Hosur, India

Abstract: ***The rapid growth of generative AI has made it increasingly difficult to tell AI-generated images from real photographs, and a single detection model often fails on images from generators it was not trained on. This paper presents the AI Image Authenticator, a web-based system that combines a six-model neural ensemble with a five-layer forensic suite to decide whether an uploaded image is AI-generated or real. Six pretrained Vision Transformer and Swin Transformer classifiers each output a probability, and a weighted democratic consensus with uncertainty tracking merges them into one verdict. In parallel, five forensic layers—C2PA Content Credentials, EXIF hardware verification, Error Level Analysis, 2D Fast Fourier Transform spectral analysis and sensor noise residual analysis—supply evidence that does not depend on any trained classifier. The system uses React.js on the front end and Flask with Hugging Face Transformers on the back end, and all models run locally on a CPU, so no paid API is needed. For every image it reports a verdict, a confidence score, each model's vote and five forensic metric cards, so the decision is explainable rather than a single opaque number. Tests on a DALL-E 3 image shared through WhatsApp, a Gemini image carrying C2PA credentials and real smartphone photographs, including WhatsApp-compressed copies, were all classified correctly. The results show that combining diverse classifiers with forensic evidence removes the blind spots of any single detector.***

Keywords: AI-Generated Image Detection, Deepfake Detection, Ensemble Learning, Vision Transformer, Error Level Analysis, Frequency Analysis, C2PA Content Credentials, Digital Image Forensics

I. INTRODUCTION

Image generators such as DALL-E, Midjourney, Stable Diffusion, Flux and Gemini now produce pictures that look like photographs. These images are useful for design and education, but they are also used for misinformation, identity fraud, fake news and non-consensual deepfakes. A viewer can no longer judge authenticity by eye, so an automated way of checking an image is needed.

Most existing detectors rely on one trained classifier. Such a classifier learns the artefacts of the generators present in its training data, so it fails on newer or different generators, and it sometimes flags genuine photographs that have been compressed by messaging applications. Both failures were observed during this work: a detector tuned for Stable Diffusion XL (SDXL) missed DALL-E 3 images, compression by WhatsApp produced false positives on real smartphone photographs, and even a three-model ensemble was fooled by a Gemini-generated image because none of its models had seen that generator.

This paper presents the AI Image Authenticator, which addresses these problems with two ideas. First, the single classifier is replaced by an ensemble of six specialised pretrained models whose votes are combined by weighted democratic consensus. Second, a five-layer forensic suite examines the file itself: its cryptographic provenance (C2PA), camera metadata (EXIF), compression behaviour (Error Level Analysis), frequency spectrum (2D-FFT) and sensor noise. The user uploads an image on a React web page and receives a verdict, a confidence percentage, every model's vote and every forensic metric, all computed locally on a CPU without any paid service. The main contributions are:

1) A six-model Vision Transformer and Swin Transformer ensemble with label normalisation, weighted voting and uncertainty tracking.
2) A five-layer forensic suite in which a C2PA manifest declaring AI generation acts as an authoritative override.
3) An explainable interface that shows every model vote and forensic metric instead of one opaque score.
4) A free, privacy-preserving implementation that runs on ordinary CPU hardware with locally cached models and is tested on images from different generators and on compressed real photographs.

The rest of the paper is organised as follows. Section II reviews related work, Section III describes the proposed system, Section IV explains the implementation, Section V presents the results and limitations, and Section VI concludes the paper.

©IJRASET: All Rights are Reserved

II. LITERATURE REVIEW

A. Watermark and Provenance Approaches

Several studies address provenance by embedding marks in generated images. Müller et al. [1] show that semantic watermarks of diffusion models can be forged by an attacker who can only query the system, so a watermark alone cannot be trusted as proof of origin. Duan et al. [2] survey watermarking for diffusion models and note the open problems of robustness to editing and the lack of standard schemes. InvisMark [4] and dynamic watermarks [5] propose invisible, robust and changing marks for tracking generated images. All of these methods identify only images whose generator cooperated by embedding a mark; images from other sources remain unchecked.

B. Passive Detection

Passive detectors analyse the image itself. Zhang et al. [3] improve the generalisation of deepfake detectors to unseen forgeries by widening the forgeries seen in training and by judging the confidence of predictions, but the method still depends on a single detector. Corvi et al. [8] show that detectors trained on GAN images transfer poorly to diffusion-model images, which confirms that a detector tied to one generator family misses others. The reviews in [6] and [7] compare passive and watermark-based approaches and conclude that no single method is sufficient and that detectors must keep pace with new generators.

C. Building Blocks

The classifiers used here are built on the Vision Transformer [9], which processes image patches with a transformer encoder, and the Swin Transformer [10], which computes attention inside shifted local windows. Frank et al. [11] and Durall et al. [12] show that the up-sampling layers of generative networks leave periodic artefacts in the frequency domain, which motivates the 2D-FFT layer. Error Level Analysis, introduced by Krawetz [13], re-saves an image at a known JPEG quality and examines the difference. The C2PA specification [14] defines a signed manifest that records how content was created; a manifest declaring AI generation is cryptographic evidence rather than a statistical estimate, but it is lost when an image is screenshotted or re-encoded by a platform that strips metadata.

TABLE I
SUMMARY OF THE SURVEYED DETECTION WORK

Ref.	Approach	Limitation for This Work
[1]	Attack study on semantic watermarks	Watermarks can be forged; no classifier for unmarked images
[2]	Survey of diffusion watermarking	Works only when the generator embeds a mark
[3]	Passive detector, expansive learning	Relies on one trained detector
[4]	Invisible provenance watermark	Identifies only marked images
[5]	Dynamic watermarks	No help when no watermark is present
[6]	Review of synthetic-image verification	Concludes that no single method is enough
[7]	Passive versus watermark comparison	Does not combine an ensemble with forensics
[8]	Diffusion-image detection	Detector tied to one generator family transfers poorly

©IJRASET: All Rights are Reserved | SJ Impact Factor 7.538 | ISRA Journal Impact Factor 7.894 |

ISSN: 2321-9653; IC Value: 45.98; SJ Impact Factor: 7.538
Volume 14 Issue X Oct 2026- Available at www.ijraset.com

D. Research Gap

The survey shows three gaps. First, most passive detectors are evaluated as a single trained model, so their accuracy depends on how closely a test image resembles the training data and new generators cause missed detections. Second, watermark and provenance schemes give strong evidence only for images that carry a mark. Third, detectors usually return one score with no explanation. The proposed system replaces the single detector with six models that differ in architecture and training data, keeps C2PA as an authoritative signal when it exists without depending on it, and shows every vote and every metric so the user can judge how far to trust the verdict.

III. PROPOSED SYSTEM

A. System Architecture

Fig. 1 shows the architecture. The front end (React.js with Vite) lets the user upload an image, preview it and view the result. The back end (Flask REST API) receives the image on the /api/scan route, validates it, runs the ensemble and the forensic layers, fuses the evidence and returns a JSON result. The neural ensemble holds six pre-trained models that run on a CPU from a cached copy, and the forensic suite holds five layers that analyse the file. The fusion step produces the verdict: a C2PA declaration of AI generation overrides everything, otherwise the consensus vote and the forensic evidence are combined.

B. Neural Ensemble

Six pre-trained image-classification models from the Hugging Face Hub are used (Table II). They differ in architecture and in training data, which is why they complement one another. Because the models name their classes differently (for example artificial, human, real, fake, or three classes), a normalisation step maps every output into one uniform distribution of AI and real before voting. The NYUAD model separates DALL-E and Stable Diffusion images from authentic photographs, and its two AI classes are merged into one AI probability.

TABLE II
PRE-TRAINED MODELS IN THE ENSEMBLE

No.	Model (Hugging Face)	Architecture and Specialty
1	Smogy/SMOGY-Ai-images-detector	Modern generative AI, strong on DALL-E
2	NYUAD-ComNets/NYUAD_AI-generated_images_detector	ViT; DALL-E vs. Stable Diffusion vs. real
3	haywoodsloan/ai-image-detector-deploy	SwinV2; camera-sensor realism
4	dima806/deepfake_vs_real_image_detection	ViT; facial realism
5	umm-maybe/AI-image-detector	Swin; Midjourney, StyleGAN, diffusion

©IJRASET: All Rights are Reserved | SJ Impact Factor 7.538 | ISRA Journal Impact Factor 7.894 |

No.	Model (Hugging Face)	Architecture and Specialty
6	Organika/sdxl-detector	SDXL latent-space artefacts

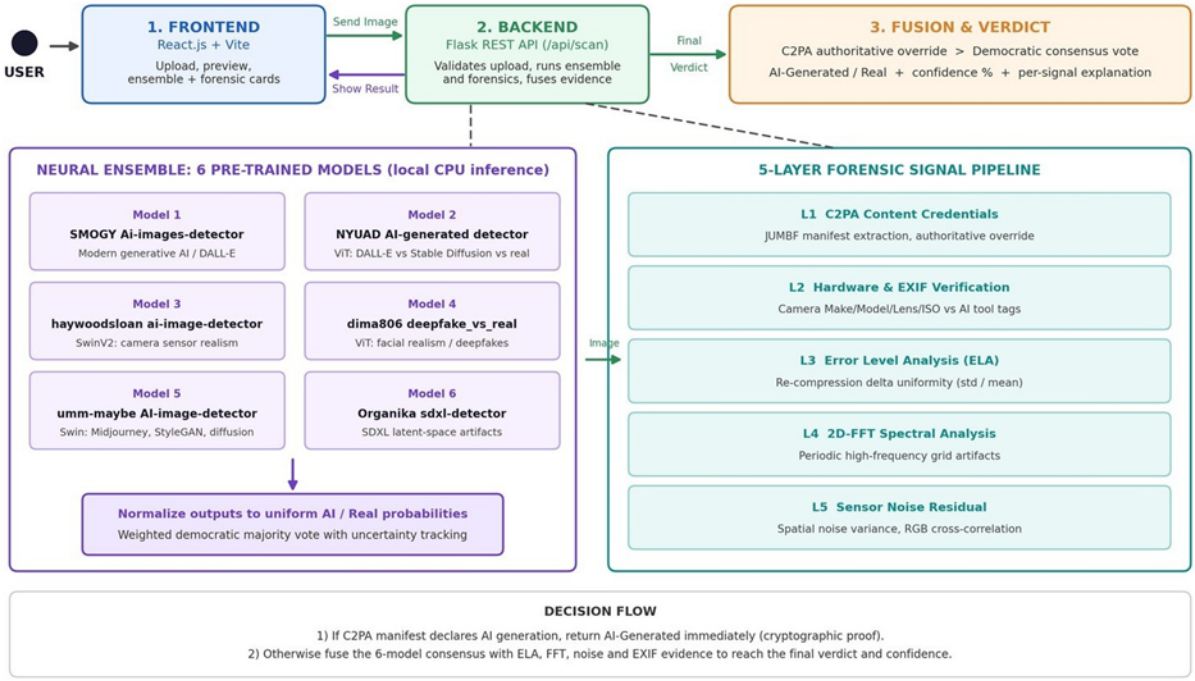


Fig. 1 Architecture of the AI Image Authenticator: front end, back end, six-model neural ensemble, five-layer forensic pipeline and fusion of the verdict.

C. Probability Normalisation and Weighted Consensus

Each classifier produces logits $z_1, \ldots, z_C$ for its C classes, which are converted to probabilities by the softmax function:

$$p_i = \exp(z_i) / \Sigma_j \exp(z_j) \qquad (1)$$

The class labels are grouped into an AI set and a real set. If *a* and *r* are the summed probabilities of the AI classes and of the real classes, the uniform AI probability of a model is

$$P_{ai} = a / (a + r) \qquad (2)$$

Model *k* has weight w_k, AI probability P_k and casts the vote $v_k = 1$ if $P_k \geq 0.5$ and 0 otherwise. The weighted AI share *S*, the weighted mean probability *M* and the margin *m* are

$$S = \Sigma\, w_k v_k / \Sigma\, w_k \qquad (3)$$

$$M = \Sigma\, w_k P_k / \Sigma\, w_k \qquad (4)$$

$$m = |2S - 1| \qquad (5)$$

The margin is 0 for a split vote and 1 for a unanimous vote, and the vote is marked uncertain when $m < 0.25$. An uncertain vote lowers the confidence instead of forcing a certain answer, so one over-confident model cannot decide the result.

D. Forensic Suite

Five forensic layers read the file itself and do not rely on how any classifier was trained.

1) *C2PA Content Credentials:* The layer extracts the cryptographic JUMBF manifest embedded in the file. If the manifest declares AI generation, the system returns AI-Generated at once, because this is signed evidence and not a statistical estimate.

2) *Hardware and EXIF Verification:* The layer reads camera make, model, lens, ISO and shutter speed and compares them with the tags AI tools leave behind, such as prompts, steps and parameters written by Automatic1111 or ComfyUI. Messaging applications remove metadata, so missing EXIF is not treated as proof of anything.
3) *Error Level Analysis (ELA):* The image is re-saved as JPEG at quality 90 and the absolute difference E from the original is measured. The uniformity of the compression error is

$$U = \sigma_E / \mu_E \qquad (6)$$

where μ_E and σ_E are the mean and standard deviation of E. A low U means suspiciously uniform error, which is typical of generated images.
4) *2D-FFT Spectral Analysis:* The grayscale image is resized, transformed with a two-dimensional FFT and the low-frequency centre is masked. In the remaining high-frequency region H of the log-magnitude spectrum the peak ratio is

$$R = \max(H) / \text{mean}(H) \qquad (7)$$

and isolated peaks indicate the periodic lattice left by the up-sampling layers of generative networks [11], [12].
5) *Sensor Noise Residual Analysis: The noise residual is obtained as*

$$N = I - G(I) \qquad (8)$$

where I is the image and G a Gaussian blur. The residual standard deviation and the correlation between the red, green and blue channels are measured. Real sensors produce characteristic noise, whereas generated images are often unnaturally clean.

Table III summarises how the layers are read. These readings are heuristics that heavy re-compression, resizing or editing can change, which is why the forensic layers support the consensus and do not replace it.

TABLE III
INTERPRETATION OF THE FORENSIC LAYERS

Layer	Typical of Real Photographs	Typical of Generated Images
C2PA	Absent, or capture credentials	Manifest declaring AI generation
EXIF	Camera make, model, lens, ISO	No camera tags; generator parameters
ELA	Higher, more varied error levels	Low, uniform error levels
2D-FFT	Smooth decay, no isolated peaks	Isolated peaks from up-sampling
Noise	Camera-like noise present	Very clean or unnatural noise

E. Evidence Fusion

The consensus score is C = 0.5S + 0.5M. The forensic score F, a weighted mean of the ELA, FFT and noise scores in the range 0 to 1, is combined with it as

$$A = 0.80C + 0.20F \qquad (9)$$

The verdict is AI-Generated if $A \geq 0.5$ and Real / Authentic otherwise, with confidence 100A or 100(1 − A) respectively. EXIF evidence adjusts A: generator tags raise it to at least 0.85 and camera hardware tags lower it slightly. A C2PA declaration of AI generation overrides the calculation. The weights and thresholds are tunable constants that should be calibrated on a labelled image set. Algorithm 1 summarises the pipeline and Fig. 2 shows the data flow.

Algorithm 1: Scan pipeline

Input: uploaded file *f*. Output: verdict, confidence, evidence.

1. Validate *f* (type, size, readable image); otherwise return an error.
2. Read the C2PA manifest of *f*; compute the EXIF, ELA, FFT and noise metrics.
3. Run the six models; normalise each output to P_k; compute S, M and m.
4. If the manifest declares AI generation, return AI-Generated (99.9%) with source C2PA.

©IJRASET: All Rights are Reserved | SJ Impact Factor 7.538 | ISRA Journal Impact Factor 7.894 |

5. Otherwise compute A = 0.80C + 0.20F and adjust it with the EXIF evidence.
6. Return the verdict and confidence with every model vote and forensic metric.

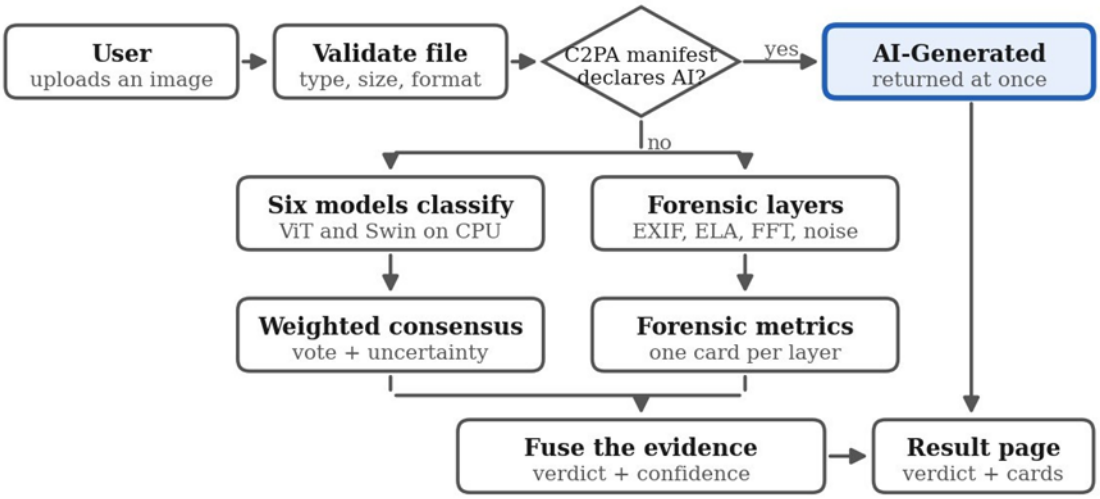


Fig. 2 Data flow of a scan: validation, C2PA check, parallel ensemble and forensic analysis, fusion and result display.

IV. IMPLEMENTATION

A. Front End

The front end is a React.js application built with Vite. The upload module offers a drag-and-drop zone and a file-browse button, checks that the file is JPG, JPEG, PNG or WEBP and not larger than 10 MB, and shows a preview. The interface moves through the states idle, image selected, scanning, result shown and reset. The result module shows a large verdict badge (green for real, red for AI-generated), the confidence percentage with a bar, an ensemble breakdown card that lists every model with its vote and confidence, and five forensic metric cards for the EXIF hardware check, the C2PA signature, ELA, the FFT spectrum and sensor noise.

B. Back End and API

The back end is written in Python with Flask and Flask-CORS. It exposes three JSON endpoints: GET /api/health reports the server status and the number of loaded models, GET /api/models lists the registered models with their specialty and weight, and POST /api/scan accepts an image as multipart form data and returns the verdict, the confidence, the per-model results, the forensic metrics and the elapsed time. All six models are loaded once at start-up, so a scan does not pay the loading delay, and a model that fails to load is skipped so that the system still works with the rest. The back end repeats the client checks and returns HTTP 400 for a missing image, 413 for a file larger than 10 MB, 415 for an unsupported type, 422 for a corrupted file and 500 for an unexpected error. Every module call is wrapped in error handling so that a single failure gives a clear message instead of a crash.

C. Technology Stack and Requirements

PyTorch and Hugging Face Transformers run the models, Pillow and NumPy implement the image forensics, and Git and GitHub are used for version control. The server needs a quad-core CPU at 2.5 GHz or better, at least 8 GB of RAM (16 GB recommended because six models are held in memory) and about 10 GB of free disk space for the cached models; no GPU is required. Python 3.10 or higher and Node.js 18 or higher are needed. An internet connection is required only for the first download of the models, after which the system works fully offline.

©IJRASET: All Rights are Reserved | SJ Impact Factor 7.538 | ISRA Journal Impact Factor 7.894 |

D. Security and Privacy

Every upload is treated as untrusted input. Files are validated by extension, by size and by actually decoding the image, so renamed or corrupted files are rejected. A maximum pixel count protects against decompression bombs, file names are displayed only as text to avoid cross-site scripting, and temporary data is discarded after each scan. All analysis runs on the local machine, so no image is sent to an external service. In a public deployment the CORS origin should be restricted to the front end, a rate limit added and the model revision pinned.

V. RESULTS AND DISCUSSION

A. Test Cases

The system was tested with images from different sources to check that the ensemble and the forensic suite remove the failures of a single model (Table IV).

TABLE IV
TEST IMAGES AND RESULTS

Test Image (Expected Class)	Result	Deciding Evidence
DALL-E 3 image shared via WhatsApp (AI)	AI-Generated, 99.7%	Six-model consensus
Real smartphone photographs, 12 MP and WhatsApp-compressed (Real)	Real / Authentic, 87.5%	Realism majority of the models
Gemini image with C2PA credentials (AI)	AI-Generated, 99.9%	C2PA cryptographic provenance

1) *DALL-E 3 Image Shared Through WhatsApp:* WhatsApp re-compresses an image and removes its metadata, so compression hides some generator artefacts and no provenance data remains. The ensemble nevertheless classified the image as AI-Generated with 99.7% confidence, whereas a single SDXL-specialised model had classified the same kind of image as real.
2) *Real Smartphone Photographs:* Photographs taken with a 12 MP smartphone were tested as original files and after WhatsApp compression, the case that earlier produced false positives. The system classified them as Real / Authentic with 87.5% confidence because the majority of the models voted real.
3) *Gemini Image with C2PA Credentials:* An image generated by Google Gemini and saved directly from the source kept its C2PA manifest. In earlier experiments the classifiers alone had labelled a Gemini image as real, because none of their training data contained that generator. In the final system the C2PA layer read the manifest, found the AI declaration and returned AI-Generated with 99.9% confidence, naming C2PA as the deciding evidence.

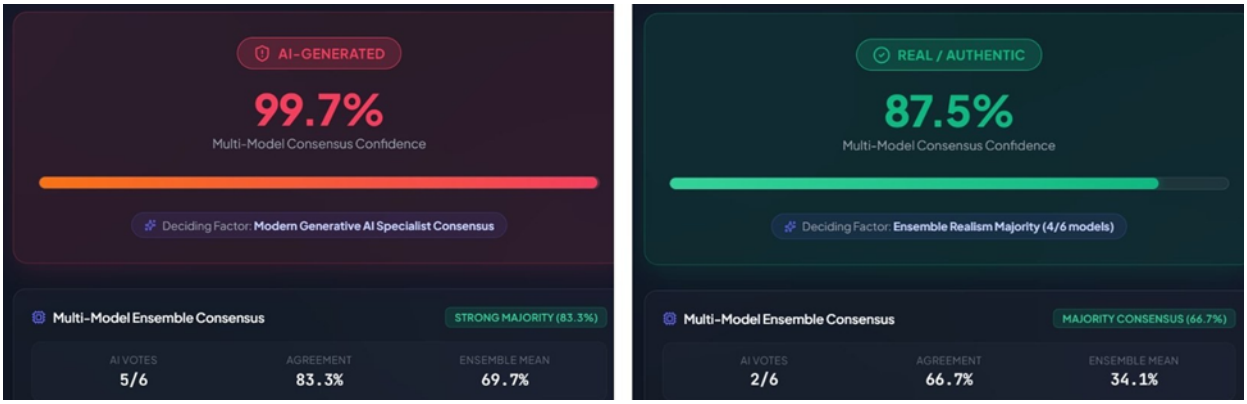


Fig. 3 Result screens of the interface: (a) AI-generated image classified with 99.7% confidence; (b) real photograph classified as Real / Authentic with 87.5% confidence.

©IJRASET: All Rights are Reserved

B. Behaviour of the Ensemble

Table V lists the AI probability returned by each model for the two images in Fig. 3. On the AI-generated image five of six models voted AI and the ensemble mean was 69.7%, while the face-specialised Dima806 model gave only 3.4% and was outvoted. On the real photograph the opposite happened: NYUAD (63.3%) and HaywoodSloan (78.3%) leaned towards AI, yet four models voted real, so the consensus, with an ensemble mean of 34.1%, classified the photograph as real. These two cases show how the democratic vote prevents the error of a single model, in either direction, from deciding the result.

TABLE V
AI PROBABILITY (%) RETURNED BY EACH MODEL

Model	AI-Generated Image	Real Photograph
Smogy Swin	99.7 (AI)	6.8 (Real)
NYUAD multi-class ViT	79.2 (AI)	63.3 (AI)
HaywoodSloan SwinV2	90.6 (AI)	78.3 (AI)
Dima806 deepfake ViT	3.4 (Real)	3.0 (Real)
Umm-Maybe Swin	55.2 (AI)	45.2 (Real)
Organika SDXL	90.3 (AI)	8.3 (Real)
AI votes / ensemble mean	5/6 / 69.7	2/6 / 34.1

C. Forensic Evidence

Fig. 4 shows the forensic cards for the real photograph. No C2PA manifest and no EXIF tags were present, so those two layers were neutral; the system does not treat missing metadata as evidence of AI generation. The ELA ratio was 1.51 (mean error 1.386), the FFT peak ratio was 1.827 with a smooth spectral decay and no isolated peaks, and the noise residual variance was 23.224 with an RGB channel correlation of 0.967. These values agree with the readings expected for a naturally captured image in Table III.

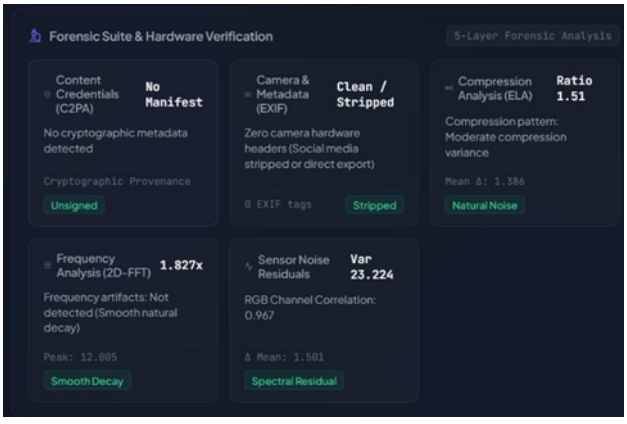


Fig. 4 Forensic metric cards for the real photograph.

D. Functional Testing

The back end was exercised with an automated self-test that runs without the neural models, and the complete flow from upload to result was tested manually in the browser. Valid JPG, PNG and WEBP files were accepted, previewed and analysed. A GIF or PDF was rejected with HTTP 415, a file larger than 10 MB with HTTP 413, and a text file renamed to .jpg with HTTP 422. Pressing Analyze with no file sent no request, and stopping the back end during a scan produced a clear connection error.

©IJRASET: All Rights are Reserved | SJ Impact Factor 7.538 | ISRA Journal Impact Factor 7.894 |

An image with an AI declaration in its C2PA manifest was returned as AI-Generated immediately, a split vote was marked uncertain with capped confidence, the server still started when one model failed to load, and Scan Another Image returned the interface to its idle state.

E. Comparison with Existing Approaches

Table VI compares the proposed system with a single online detector and with a watermark check. Only the proposed system works on unmarked images, combines several independent models, uses provenance data when it exists, explains its verdict, and keeps the image on the local machine.

TABLE VI
COMPARISON WITH EXISTING APPROACHES

Feature	Single Online Detector	Watermark Check	This Work
Works on unmarked images	Yes	No	Yes
Several independent models	No	No	Yes (six)
Explains the verdict	Rarely	No	Yes
Uses C2PA provenance	No	Yes	Yes
Free to run	Often limited	Depends	Yes
Images stay local	No	Depends	Yes

F. Limitations

The system has the following limitations. 1) Detectors recognise only generators similar to those in their training data, so a brand-new generator may be classified as real until the models are updated. 2) C2PA and EXIF data are removed by screenshots and by many messaging applications, so a missing manifest says nothing about an image. 3) ELA, FFT and noise analysis are heuristics that heavy re-compression, resizing or editing can change, so they act only as supporting evidence. 4) The fusion weights and thresholds are tuned constants, and the evaluation reported here is qualitative on a small set of images; accuracy, precision, recall and F1-score on a larger labelled set, including WhatsApp-compressed copies, remain to be measured. 5) All inference runs on a CPU, which is slower than a GPU when many users scan images at the same time.

VI. CONCLUSION AND FUTURE SCOPE

This paper showed that reliable detection of AI-generated images needs more than one trained classifier. By combining six specialised Vision Transformer and Swin models through weighted democratic consensus, and by adding a five-layer forensic suite of C2PA Content Credentials, EXIF hardware verification, Error Level Analysis, 2D-FFT spectral analysis and sensor noise residual analysis, the AI Image Authenticator removed the failures observed with a single model: DALL-E 3 images that a single detector missed are now detected, real smartphone photographs compressed by WhatsApp are classified as real, and images carrying C2PA credentials are identified with cryptographic certainty. Every result shows the verdict, the confidence, each model's vote and the five forensic metrics, and the application runs on ordinary CPU hardware with locally cached models, needs no paid API and keeps the user's images on the local machine.

©IJRASET: All Rights are Reserved | SJ Impact Factor 7.538 | ISRA Journal Impact Factor 7.894 |

Future work includes calibrating the model weights and forensic thresholds on a larger labelled collection and reporting accuracy, precision, recall and F1-score; batch scanning with a summary table; exportable PDF audit reports for each scan; additional forensic layers such as double-JPEG detection, copy-move analysis and lighting-consistency analysis; extension of the pipeline to video by analysing frames; optional GPU and cloud deployment; and a browser extension and documented API so that images can be checked where they appear.

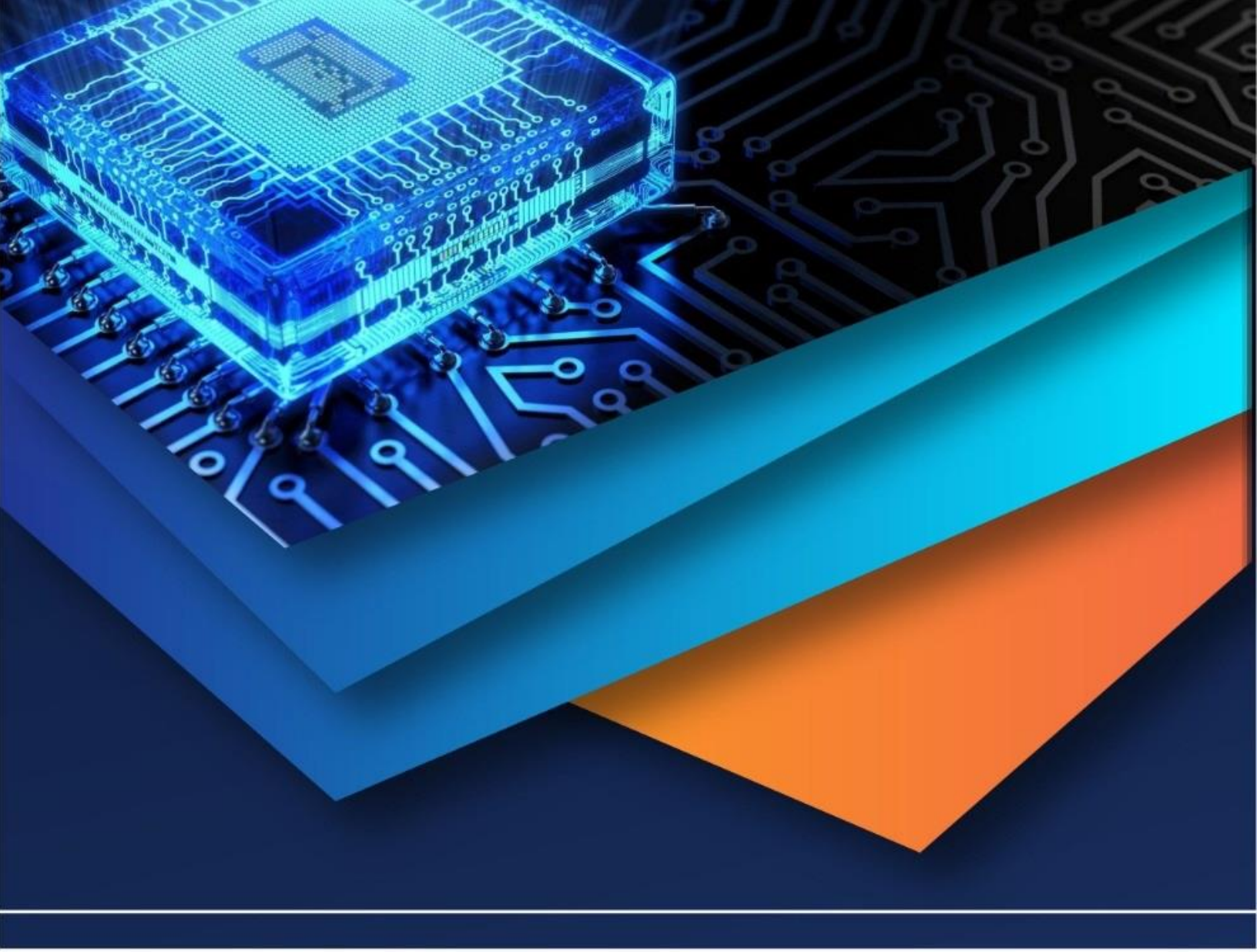
VII. ACKNOWLEDGMENT

The authors thank the Principal, Dr. R. Radhakrishnan, and the Head of the Department of Computer Science and Engineering, Dr. G. Fathima, Adhiyamaan College of Engineering, Hosur, for providing the opportunity and for their guidance and encouragement. The authors are also grateful to the project supervisor, the Project Coordinator and all staff members for their support in completing this work.

REFERENCES

[1] A. Müller, D. Lukovnikov, J. Thietke, A. Fischer, and E. Quiring, “Black-box forgery attacks on semantic watermarks for diffusion models,” in Proc. IEEE/CVF Conf. Computer Vision and Pattern Recognition (CVPR), 2025, pp. 20937–20946.
[2] J. Duan, J. Guang, W. Yang, and R. He, “Visual watermarking in the era of diffusion models: Advances and challenges,” 2025.
[3] K. Zhang, Z. Hou, Z. Hua, Y. Zheng, and L. Y. Zhang, “Boosting deepfake detection generalizability via expansive learning and confidence judgement,” IEEE Trans. Circuits Syst. Video Technol., 2025.
[4] R. Xu et al., “InvisMark: Invisible and robust watermarking for AI-generated image provenance,” in Proc. IEEE/CVF Winter Conf. Applications of Computer Vision (WACV), 2025, pp. 909–918.
[5] Y. Chen, J. Vice, N. Akhtar, N. Haldar, and A. Mian, “Dynamic watermarks in images generated by diffusion models,” in Proc. IEEE/CVF Conf. Computer Vision and Pattern Recognition Workshops (CVPRW), 2025, pp. 5271–5277, doi: 10.1109/CVPRW67362.2025.00523.
[6] D. Tariang et al., “Synthetic image verification in the era of generative artificial intelligence: What works and what isn’t there yet,” IEEE Security & Privacy, 2024.
[7] M. Guo et al., “AI-generated image detection: Passive or watermark?,” 2024.
[8] R. Corvi, D. Cozzolino, G. Zingarini, G. Poggi, K. Nagano, and L. Verdoliva, “On the detection of synthetic images generated by diffusion models,” in Proc. IEEE Int. Conf. Acoustics, Speech and Signal Processing (ICASSP), 2023.
[9] A. Dosovitskiy et al., “An image is worth 16x16 words: Transformers for image recognition at scale,” in Proc. Int. Conf. Learning Representations (ICLR), 2021.
[10] Z. Liu et al., “Swin Transformer: Hierarchical vision transformer using shifted windows,” in Proc. IEEE/CVF Int. Conf. Computer Vision (ICCV), 2021.
[11] J. Frank, T. Eisenhofer, L. Schönherr, A. Fischer, D. Kolossa, and T. Holz, “Leveraging frequency analysis for deep fake image recognition,” in Proc. Int. Conf. Machine Learning (ICML), 2020.
[12] R. Durall, M. Keuper, and J. Keuper, “Watch your up-convolution: CNN based generative deep neural networks are failing to reproduce spectral distributions,” in Proc. IEEE/CVF Conf. Computer Vision and Pattern Recognition (CVPR), 2020.
[13] N. Krawetz, “A picture’s worth: Digital image analysis and forensics,” Black Hat Briefings, 2007.
[14] Coalition for Content Provenance and Authenticity (C2PA), Content Credentials Technical Specification.

©IJRASET: All Rights are Reserved | SJ Impact Factor 7.538 | ISRA Journal Impact Factor 7.894 |



10.22214/IJRASET

45.98

IMPACT FACTOR: 7.129

IMPACT FACTOR: 7.429

INTERNATIONAL JOURNAL FOR RESEARCH

IN APPLIED SCIENCE & ENGINEERING TECHNOLOGY

Call : 08813907089 (24*7 Support on Whatsapp)