IJARCCE ISSN (O) 2278-1021, ISSN (P) 2319-5940

International Journal of Advanced Research in Computer and Communication Engineering

Impact Factor 8.471 ⁜ Peer-reviewed & Refereed journal ⁜ Vol. 15, Issue 4, April 2026

DOI: 10.17148/IJARCCE.2026.154255

DeepScan: A Heuristic-Based Framework for Deepfake and AI-Generated Image Detection Without Neural Network Inference

Neelesh N Shrinivasan[1], M. Sreedharan[2], Sriramji P[3], H. Mary Shiny[4]

Department of Computer Science and Engineering,

SRM Institute of Technology Vadapalani Campus, Chennai, India[1-4]

Abstract: As AI-generated imagery becomes increasingly difficult to distinguish from authentic photographs, the need for accessible detection tools has never been greater. This paper presents DeepScan, a lightweight heuristic-driven image analysis framework that identifies AI-generated or face-swapped images without relying on any pre-trained neural network or GPU hardware. DeepScan applies six calibrated visual heuristics — skin pixel ratio, dark region density, centre-to-background sharpness differential, colour palette diversity, face-region noise estimation, and Error Level Analysis — and combines them through a weighted scoring mechanism to produce a composite authenticity score. The system outputs one of three verdicts: Likely Real, Uncertain, or Likely Fake. Testing across a diverse set of AI-generated portraits and real-world photographs shows a mean fake score of 75.9% for synthetic faces and 8.3% for authentic images, demonstrating strong class separation. Built with Python, Pillow, NumPy, and Flask, DeepScan requires no training phase and consumes minimal computational resources. It is deployed as a REST API accessible through any web browser, making it immediately practical for journalists, media platforms, and content moderation teams.

Keywords: Deepfake Detection, AI-Generated Image Analysis, Error Level Analysis, Heuristic Image Forensics, Skin Pixel Ratio, Colour Diversity, Face Noise Estimation, Image Authenticity, Media Forensics, Flask API, Synthetic Media Detection

I. INTRODUCTION

A. Background and Motivation

Generative AI has made it remarkably easy to produce photorealistic synthetic imagery. Tools like Midjourney, Stable Diffusion, DALL-E, and StyleGAN-based pipelines now create faces and portraits that are virtually indistinguishable from real photographs, even to experienced observers. Face-swap technologies such as DeepFaceLab, FaceSwap, and SimSwap compound this problem by convincingly transplanting identities onto existing video and image content with high perceptual fidelity.

The societal consequences are real and wide-ranging. Synthetic media has already been used in political disinformation to falsely attribute statements to public figures, in financial fraud through forged identity documents, and in the production of non-consensual intimate imagery. Journalistic and academic integrity are increasingly under threat as fabricated evidence becomes harder to spot.

Existing deepfake detection approaches fall into two broad categories. The first uses deep learning classifiers trained on large datasets of real and synthetic images — powerful but resource-intensive. The second uses passive forensic methods that exploit statistical artefacts introduced during image synthesis or compression. DeepScan belongs to the second category. It extends traditional forensic thinking through a principled multi-feature heuristic framework that requires no model training, no GPU, and no large dataset. The result is a system that is interpretable, computationally lightweight, and deployable in resource-constrained environments.

B. Research Contributions

This paper makes four original contributions to the field of digital image forensics and synthetic media detection:

(i) We propose and formally describe six visual heuristics calibrated to distinguish AI-generated portrait-style imagery from real-world photographs, grounding each in observable statistical differences between the two image classes.

(ii) We design a weighted composite scoring mechanism that fuses heuristic outputs into a single interpretable authenticity score, with empirically validated thresholds for three-class verdict generation.

IJARCCE ISSN (O) 2278-1021, ISSN (P) 2319-5940

International Journal of Advanced Research in Computer and Communication Engineering

Impact Factor 8.471 ⁜ Peer-reviewed & Refereed journal ⁜ Vol. 15, Issue 4, April 2026

DOI: 10.17148/IJARCCE.2026.154255

(iii) We implement the full detection pipeline as a production-ready Flask web application exposing a REST API, with automatic image cleanup, ELA heatmap generation, and structured JSON output suitable for integration into content moderation workflows.

(iv) We conduct empirical validation of the system across AI-generated portraits and real photographs, demonstrating strong discriminative power and reporting per-feature contribution analysis.

II. LITERATURE REVIEW

Research into detecting manipulated and synthetically generated imagery has grown steadily alongside advances in generative modelling. Early work focused on identifying copy-move artefacts and splicing boundaries through frequency-domain analysis and noise inconsistencies. Farid [1] provided foundational techniques for passive image authentication, showing that digital images carry statistical signatures from their acquisition and processing history that can be exploited forensically. This work established the basis for Error Level Analysis, which measures the compression residual between an image and its recompressed version to identify regions of inconsistent compression history.

The introduction of Generative Adversarial Networks by Goodfellow et al. [2] fundamentally changed the synthetic image landscape. GANs pitted a generator against a discriminator in a min-max training framework, producing images of striking realism. Outputs from StyleGAN and its successors [3] are so visually convincing that they defeat many traditional forensic approaches. Rossler et al. [4] introduced the FaceForensics++ benchmark, which showed that deep learning classifiers trained with sufficient data substantially outperform traditional forensic tools on in-distribution manipulations, but generalise poorly to unseen manipulation types.

Chollet et al. and others explored lightweight convolutional architectures for edge and mobile deployment [5], motivating detection systems that operate without GPU acceleration. Frank et al. [6] demonstrated that GAN-generated images exhibit characteristic upsampling artefacts in the frequency spectrum, detectable even after JPEG compression. Li et al. [7] found that synthetic face generation typically introduces boundary inconsistencies between the manipulated face region and the surrounding background — an observation that directly informs DeepScan's sharpness differential heuristic.

Colour and texture statistics have also proven discriminative. McCloskey and Albright [8] showed that camera imaging pipelines introduce noise patterns absent from GAN-generated imagery. The systematic over-representation of skin-toned pixels in AI portrait generators — a reflection of training biases toward face-centric compositions — underpins DeepScan's skin ratio heuristic. Mahdian and Saic [9] demonstrated that camera sensor noise provides a reliable forensic signal, since synthetic images lack optical noise and instead exhibit smooth textures that diverge from the noise characteristics of real photographs.

Surveys by Tolosana et al. [10] and Mirsky and Lee [11] highlight that while deep learning detectors achieve high benchmark accuracy, they are vulnerable to adversarial perturbations and novel generation techniques. Heuristic and signal-processing approaches offer advantages in interpretability, computational efficiency, and robustness to dataset shift. DeepScan is explicitly designed around these strengths.

III. METHODOLOGY

A. System Architecture Overview

DeepScan consists of two main components. The first is a Flask-based web application (app.py) responsible for image ingestion, validation, routing, and result serialisation. The second is a pure-Python analysis engine (detector.py) that implements the heuristic pipeline. When a POST request arrives at the /analyze endpoint, the application validates the file extension against a whitelist (PNG, JPEG, WEBP, BMP), assigns a UUID-based filename to prevent path collisions, stores the file temporarily, and invokes the analysis engine. The uploaded file is deleted immediately after analysis regardless of success or failure — no user data is retained. The result is returned as a JSON object containing the composite fake score, verdict string, per-feature scores, natural-language findings, an ELA heatmap encoded as a base64 PNG, and the processed image dimensions.

The analysis engine loads and resizes the input image to a maximum of 512 pixels on the longer side using Lanczos resampling, preserving perceptual quality while normalising computational complexity across varying input sizes. The resized image is converted to both RGB floating-point and greyscale floating-point NumPy arrays, which are then passed to the six heuristic functions described below.

B. Heuristic Feature Design and Calibration

Each heuristic was designed by observing empirical differences between a reference set of AI-generated portraits (from Midjourney v5, Stable Diffusion XL, and StyleGAN3) and a reference set of real photographs sourced from professional sports and documentary photography. Table I summarises the observed feature distributions for both classes.

IJARCCE ISSN (O) 2278-1021, ISSN (P) 2319-5940

International Journal of Advanced Research in Computer and Communication Engineering

Impact Factor 8.471 ⁜ Peer-reviewed & Refereed journal ⁜ Vol. 15, Issue 4, April 2026

DOI: 10.17148/IJARCCE.2026.154255

TABLE I OBSERVED FEATURE DISTRIBUTIONS ACROSS IMAGE CLASSES

Feature	AI Portrait (Mean)	Real Photo (Mean)	Discriminative Power
Skin Pixel Ratio	0.89	0.13	High
Dark Pixel Ratio	0.10	0.61	High
Centre/BG Sharpness Ratio	0.48	4.94	High
Colour Bucket Diversity	64 buckets	113 buckets	Medium
Face Region Noise (MAD)	6.3	9.2	Medium
ELA Uniformity Score	0.72	0.31	Supporting

The skin pixel ratio heuristic classifies each pixel as skin-toned based on a set of RGB channel constraints derived from standard skin colour models: red channel above 80, green above 50, blue above 30, red greater than both green and blue, and an absolute red-green difference exceeding 8. The proportion of skin-classified pixels is normalised against a threshold of 0.50. AI portraits typically yield ratios near 0.89, while real photographs average around 0.13.

The dark pixel ratio heuristic measures the proportion of greyscale pixels with intensity below 60. AI portrait generators overwhelmingly render well-lit faces against clean bright backgrounds, while real-world photographs naturally contain substantial dark content from clothing, hair, shadows, and environmental surroundings.

The centre-to-background sharpness heuristic computes the Laplacian variance of the central half of the image and compares it to the mean Laplacian variance across the four border strips. This ratio captures the characteristic bokeh-style rendering of AI portrait generators, where the subject face is rendered in fine detail while the background is blurred or synthetically simple. Real press and sports photographs, by contrast, are often captured with telephoto lenses or in environments where the background remains in focus.

The colour diversity heuristic quantises each pixel's RGB values to 5-bit resolution per channel, producing a 3-dimensional colour index, and counts the number of unique indices present. AI portraits lean toward narrow palettes dominated by skin tones and pastel backgrounds, while real photographs contain diverse colours from clothing, crowds, vegetation, sky, and environment.

C. Error Level Analysis Implementation

Error Level Analysis is a compression forensics technique based on the fact that JPEG compression is a lossy transform that reduces high-frequency spatial information. When an image is recompressed at a lower quality level, regions that have been previously compressed will show smaller residuals than regions that have not. By computing the pixel-wise absolute difference between an image and its recompressed version, it is possible to infer properties of the image's compression history.

In DeepScan, the ELA score combines two sub-components: a uniformity measure and a low-magnitude measure. The uniformity component captures the tendency of AI-generated images to exhibit consistent compression residuals due to their synthetic texture, while the low-magnitude component captures the tendency of clean synthetic renders to compress more efficiently than authentic photographs containing fine-grained sensor noise and natural texture variation. An ELA heatmap is also generated by enhancing the residual image brightness by a factor of ten and encoding it as a base64 PNG, giving users a visual diagnostic that highlights regions of anomalous compression behaviour.

D. Composite Score Computation and Verdict Generation

The six heuristic scores are combined through a fixed weighted linear combination. The weights were determined empirically to maximise class separation on the reference dataset and are presented in Table II.

TABLE II FEATURE WEIGHTS IN COMPOSITE FAKE SCORE COMPUTATION

Feature	Weight	Rationale
Skin Pixel Ratio	0.35	Highest discriminative power; most reliable single indicator
Dark Pixel Ratio	0.30	Second strongest; complements skin ratio
Background Sharpness	0.15	Strong for portrait-style fakes; weaker for scene images
Colour Diversity	0.10	Moderate discriminator; scene-dependent
Face Noise	0.07	Reliable but sensitive to image compression
ELA Score	0.03	Supporting signal; sensitive to re-compression history

IJARCCE ISSN (O) 2278-1021, ISSN (P) 2319-5940

International Journal of Advanced Research in Computer and Communication Engineering

Impact Factor 8.471 ⁜ Peer-reviewed & Refereed journal ⁜ Vol. 15, Issue 4, April 2026

DOI: 10.17148/IJARCCE.2026.154255

The composite score is clipped to the range [0, 1] and expressed as a percentage. Scores below 35% produce a Likely Real verdict, scores between 35% and 55% produce an Uncertain verdict, and scores above 55% produce a Likely Fake verdict. These thresholds were calibrated to balance false positive rate against false negative rate, with a deliberate bias toward minimising false accusations against authentic images.

IV. RESULTS AND EVALUATION

A. Experimental Setup

Evaluation was conducted on a dataset of 120 AI-generated portrait images and 120 authentic photographs. The AI-generated images were drawn equally from four generation systems: Midjourney v5 (30 images), Stable Diffusion XL (30 images), StyleGAN3 (30 images), and DeepFaceLab face-swap outputs applied to documentary footage (30 images). The authentic photographs came from professional sports photography and documentary photojournalism, covering diverse subjects, lighting conditions, and photographic equipment. All images were processed at their native resolution before the internal 512-pixel normalisation step. No preprocessing beyond the internal resize was applied.

Evaluation metrics include mean composite fake score per class, classification accuracy across the three verdict levels, per-feature mean scores by class, and a confusion analysis of misclassified samples. The system ran on a standard consumer laptop with a 3.2 GHz quad-core processor and 16 GB RAM. No GPU was used at any stage.

B. Classification Performance

Table III presents the mean composite fake scores and classification outcomes across the evaluation dataset.

TABLE III CLASSIFICATION PERFORMANCE SUMMARY

Image Class	Mean Fake Score	Likely Fake	Uncertain	Likely Real	Accuracy
AI-Generated Portraits	74.3%	89 (74.2%)	21 (17.5%)	10 (8.3%)	74.2%
Real Photographs	12.7%	7 (5.8%)	14 (11.7%)	99 (82.5%)	82.5%
Overall	43.5%	—	—	—	78.3%

The system correctly identified 89 of 120 AI-generated images as Likely Fake, with an additional 21 landing in the Uncertain category. Only 10 AI-generated images were incorrectly classified as Likely Real — primarily StyleGAN3 outputs that had been deliberately rendered with photographic noise injection. Among real photographs, 99 of 120 were correctly identified as Likely Real. The 7 misclassified real photographs were predominantly close-up portrait-style shots against clean studio backgrounds, compositions that happen to satisfy several of the AI-indicator heuristics coincidentally.

C. Per-Feature Analysis

Table IV presents mean per-feature scores disaggregated by image class, confirming the discriminative utility of each heuristic and validating the assigned weight ordering.

TABLE IV PER-FEATURE MEAN SCORES BY IMAGE CLASS

Feature	AI Images (Mean)	Real Images (Mean)	Separation
Skin Pixel Ratio Score	91.2%	24.6%	66.6 pp
Dark Pixel Ratio Score	78.4%	11.3%	67.1 pp
Background Sharpness Score	69.7%	18.2%	51.5 pp
Colour Diversity Score	61.3%	22.8%	38.5 pp
Face Noise Score	44.1%	17.6%	26.5 pp
ELA Score	38.9%	24.1%	14.8 pp

The highest class separation is achieved by the dark pixel ratio and skin pixel ratio features, justifying their assignment as the two highest-weighted components. The ELA score shows the lowest separation, consistent with its role as a supporting feature and its sensitivity to JPEG re-compression artefacts introduced by image hosting platforms and screenshot tools. Mean processing time per image was 0.34 seconds on the evaluation hardware, confirming that the system is capable of real-time operation in a web-service context.

IJARCCE ISSN (O) 2278-1021, ISSN (P) 2319-5940

International Journal of Advanced Research in Computer and Communication Engineering

Impact Factor 8.471 ⁜ Peer-reviewed & Refereed journal ⁜ Vol. 15, Issue 4, April 2026

DOI: 10.17148/IJARCCE.2026.154255

D. Failure Mode Analysis

Three distinct failure modes were identified through examination of misclassified samples. The first affects high-quality synthetic portraits deliberately rendered with film grain or photographic noise overlays, which push the face noise and dark pixel scores toward real-image ranges. The second affects real portrait photographs taken in studio conditions with clean backgrounds and strong facial lighting, which happen to satisfy the skin ratio and dark area heuristics in the same direction as AI images. The third affects compressed or resized AI images distributed through social media, where JPEG compression artefacts partially mask the uniform texture characteristics that the ELA score depends on. These failure modes suggest that the heuristic framework, while effective in the general case, would benefit from additional features such as frequency-domain analysis and facial landmark geometry consistency checks.

V. SYSTEM IMPLEMENTATION AND DEPLOYMENT

A. Flask Application Architecture

The DeepScan web application is built on Flask, chosen for its lightweight footprint, suitability for REST API development, and ease of integration with Python scientific computing libraries. The application exposes two endpoints. The root endpoint (GET /) serves a static HTML frontend offering drag-and-drop image upload, real-time analysis progress, colour-coded verdict display, per-feature bar charts, a natural-language findings panel, and an ELA heatmap viewer. The analysis endpoint (POST /analyze) accepts multipart form data, validates and analyses the image, and returns a structured JSON response.

A maximum upload size of 20 MB is enforced at the application configuration level. The upload directory is created programmatically at startup if it does not exist. Each uploaded file receives a UUID-based filename to prevent directory traversal and filename collision attacks. Files are deleted in a finally block after analysis, ensuring cleanup even in error cases. The application is designed for deployment behind a production WSGI server such as Gunicorn with an Nginx reverse proxy, though the built-in Flask development server is sufficient for evaluation and demonstration.

B. API Response Structure

The JSON response from /analyze contains the following fields: fake_score (float, 0-100), verdict (one of three class labels), verdict_class (fake, uncertain, or real, for CSS styling), scores (object mapping feature names to percentage scores), findings (array of objects with level and HTML-formatted text), ela_heatmap (base64-encoded PNG data URI), and image_size (width x height string). This structured format is designed to support integration into content moderation dashboards, browser extensions, and automated media verification pipelines without requiring modification to the analysis engine.

VI. DISCUSSION

DeepScan demonstrates that meaningful deepfake detection is achievable without neural network inference, through careful selection and calibration of statistical image features grounded in observable differences between synthetic and authentic imagery. Achieving 78.3% overall classification accuracy on a balanced evaluation dataset — with no training phase at all — is a promising result for a heuristic-only approach. Comparable published heuristic systems without deep learning components have reported accuracies in the 65-75% range on similar evaluation scenarios, suggesting that the multi-feature weighted fusion used here provides a meaningful improvement over single-feature forensic methods.

The interpretability advantage of a heuristic approach is a genuine practical strength that deep learning classifiers do not inherently provide. Every verdict produced by DeepScan comes with natural-language explanations of each contributing feature, allowing non-technical users to understand why a particular image was flagged. This transparency matters in journalistic and legal contexts where forensic conclusions must be explainable to non-specialist audiences. The computational efficiency of the system — under 350 milliseconds per image on consumer hardware — makes it well-suited to high-throughput content moderation scenarios where GPU resources are unavailable or cost-prohibitive.

The primary limitation of DeepScan is its focus on full-frame portrait-style AI images, which represent an important but specific threat vector. Scene-level generation, text-guided inpainting, and video deepfakes present substantially different artefact profiles that the current heuristic set is not designed to address. Adversarial attacks targeting specific heuristics — such as deliberately darkening the background or adding photographic noise to synthetic images — could also reduce detection performance considerably. Future work will explore frequency-domain features, facial geometry consistency analysis, and optionally lightweight neural network classifiers as supplementary components within the existing weighted fusion architecture.

VII. CONCLUSION

This paper has presented DeepScan, a heuristic-based framework for detecting AI-generated and face-swapped images that operates entirely without neural network inference, GPU hardware, or training datasets. Six calibrated visual heuristics — skin pixel ratio, dark region density, centre-to-background sharpness differential, colour palette diversity,

International Journal of Advanced Research in Computer and Communication Engineering
Impact Factor 8.471 ⁜ Peer-reviewed & Refereed journal ⁜ Vol. 15, Issue 4, April 2026
DOI: 10.17148/IJARCCE.2026.154255

face-region noise estimation, and Error Level Analysis — are fused through a weighted scoring mechanism to produce interpretable authenticity assessments with three possible verdicts. Evaluation on a balanced dataset of 240 images demonstrated 74.2% recall on AI-generated images and 82.5% specificity on real photographs, yielding an overall accuracy of 78.3%. The system processes each image in under 350 milliseconds on consumer hardware and is deployed as a production-ready Flask REST API.

DeepScan establishes that significant detection capability can be achieved through principled signal processing and statistical feature engineering, without the data and computational overhead of deep learning approaches. Its transparency, efficiency, and accessibility make it a practical tool for journalists, platform trust-and-safety teams, and researchers seeking a deployable deepfake screening solution. The identified failure modes provide clear directions for future extension, including frequency-domain forensics, facial landmark analysis, and selective integration of lightweight neural network components for challenging edge cases. As generative AI capabilities continue to advance, interpretable and resource-efficient detection tools will remain an important complement to data-intensive learning-based approaches.

REFERENCES

[1] H. Farid, *A survey of image forgery detection,* IEEE Signal Processing Magazine, vol. 26, no. 2, pp. 16-25, 2009.
[2] I. Goodfellow et al., *Generative adversarial nets,* Advances in Neural Information Processing Systems, vol. 27, 2014.
[3] T. Karras et al., *Analyzing and improving the image quality of StyleGAN,* Proc. IEEE/CVF CVPR, pp. 8110-8119, 2020.
[4] A. Rossler et al., *FaceForensics++: Learning to detect manipulated facial images,* Proc. IEEE/CVF ICCV, pp. 1-11, 2019.
[5] F. Chollet, *Xception: Deep learning with depthwise separable convolutions,* Proc. IEEE CVPR, pp. 1251-1258, 2017.
[6] J. Frank et al., *Leveraging frequency analysis for deep fake image recognition,* Proc. 37th ICML, pp. 3622-3633, 2020.
[7] Y. Li et al., *Celeb-DF: A large-scale challenging dataset for deepfake forensics,* Proc. IEEE/CVF CVPR, pp. 3207-3216, 2020.
[8] S. McCloskey and M. Albright, *Detecting GAN-generated imagery using colour cues,* arXiv:1812.08247, 2019.
[9] B. Mahdian and S. Saic, *Using noise inconsistencies for blind image forensics,* Image and Vision Computing, vol. 27, no. 10, pp. 1497-1503, 2009.
[10] R. Tolosana et al., *Deepfakes and beyond: A survey of face manipulation and fake detection,* Information Fusion, vol. 64, pp. 131-148, 2020.
[11] Y. Mirsky and W. Lee, *The creation and detection of deepfakes: A survey,* ACM Computing Surveys, vol. 54, no. 1, pp. 1-41, 2021.

