Journal of Applied Technology and Innovation
Vol. 10, Issue 3, 2026, pp. 49-60
e-ISSN:2600-7304-ISSN
doi: https://doi.org/10.65136/jati.v10i3.359

Handcrafted Feature Extraction and Classification for Distinguishing Real and AI-Generated Images

Mahadeva Prasad M [1*], Akash K R [2], Jithendra S R [2], Arundhathi H M [2]

[1] Department of Studies in Electronics, University of Mysore, Hassan, Karnataka, India
[2, 3, 4] Department of Studies in Electronics, Hassan University, Hassan, Karnataka, India
*Corresponding author: prasada9@gmail.com

Received: 2026-09-26; Accepted: 2026-10-23; Published: 2026-10-23

Abstract

The generative machine-learning approaches are extensively used to create artificial images. These images are called AI-generated images. The AI-generated images find applications like rapid content creation, creative design and visualization, rapid prototyping, etc. In spite of many applications, the AI-generated images are leading to issues like bias, copyright issues, deepfakes, etc. Owing to the importance of discriminating between synthetic and real images, many research approaches have been proposed. Among these methods, deep learning approaches are popular. However, the design of common deep learning detection approaches to detect the synthetic images that are generated by different generative models is a challenging task. There are issues with these methods to discriminate images that are subjected to various pre-processing steps. To overcome these issues, machine learning approaches with the handcrafted features are gaining importance. In this context, five types of handcrafted feature extraction and random forest-based classification is proposed to distinguish between AI-generated and real images. The simulation experiments are carried out with AI-generated and real facial images, with each group having 70,000 samples. Out of 70,000 samples from each group, 50,000, 10,000, and 10,000 samples are respectively used for training, validating, and testing the classifier. The performance of the classifier is evaluated for accuracy, precision, recall, and F1 score. The recognition accuracy, precision, recall, and F1 scores of 86.03%, 86.15%, 85.87%, and 86.01%, respectively, are achieved for three feature combinations.

Keywords: *AI-generated verses real images; AI image generation; handcrafted features; feature selection and combination.*

1. Introduction

AI-image generation is a generative machine-learning process in which neural networks learn statistical patterns and relationships from large image datasets (Goodfellow, I. et al., 2014). In text-to-image systems, a textual prompt is transformed into a semantic representation that guides the synthesis of an image. Earlier methods include Generative Adversarial Networks (GANs) (Goodfellow, I. et al., 2014) and Variational Autoencoders (VAEs) (Kingma, D. P., & Welling, M., 2019), while contemporary systems increasingly incorporate Transformer-based architectures. Modern image-generation systems predominantly employ diffusion models (Ho, J., Jain, A., & Abbeel, P., 2020), (Rombach, R., et al., 2022), which generate images by progressively denoising random noise under the guidance of the input

prompt (Saharia, C., 2022), (Radford, A., et al., 2021). These developments have enabled the generation of highly realistic and diverse images for applications such as creative design, visualization, image editing, rapid prototyping, and synthetic-data generation. The increasing quality and accessibility of AI-generated images provide significant benefits in terms of rapid content creation, customization, and automation. However, their increasing realism also introduces several concerns, including visual inaccuracies, bias, copyright, and ownership issues, misinformation, privacy, and malicious applications such as deepfakes. Moreover, AI-generated images may contain artefacts or statistical characteristics that differ from naturally captured images. These differences can arise from the image-generation and reconstruction processes and may be reflected in texture, pixel statistics, frequency characteristics, noise patterns, and local structural properties. Consequently, reliable methods for distinguishing AI-generated images from real images have become increasingly important.

Distinguishing real images from AI-generated images is generally formulated as an image-forensics and classification problem. Existing approaches investigate characteristics in both the spatial and frequency domains. Spatial-domain methods analyze properties such as texture, edges, noise, intensity, and local structural consistency, whereas frequency-domain methods exploit differences in spectral characteristics introduced during image synthesis and processing (Frank, J., et al., 2020). Another important approach is the identification of artificial fingerprints associated with generative models. GANs and other generative architectures can leave characteristic patterns that can be exploited to detect synthetic images or attribute them to particular generating models (Marra, F., et al., 2019), (Yu, N., Davis, L. S., & Fritz, M., 2019).

With the advances of deep learning, CNN-based detectors have demonstrated strong capabilities in learning discriminative representations directly from image data. Wang et al. showed that the CNN-based models can effectively distinguish generated images from real photographs (Wang, S, Y., et al. 2020). However, their performance can decrease when applied to images generated by previously unseen models. More recent studies have consequently focused on universal detectors capable of generalizing across different generative models, including diffusion-based image generators (Wang, S, Y., et al. 2020), (Verdoliva, l., 2020). In parallel, metadata, watermarking, and image-provenance techniques provide complementary approaches by attempting to verify the origin of digital content rather than inferring authenticity solely from visual characteristics.

Despite these developments, reliable AI-generated image detection remains challenging. A detector trained on images from a particular generator may learn generator-specific characteristics and consequently perform poorly on images produced by other models. Detection performance may also be affected by common pre-processing operations such as resizing, compression, cropping, filtering, and image editing. Therefore, effective detection methods should identify characteristics that are sufficiently discriminative across different image-generation models and robust to common image-processing operations (Verdoliva, l. 2020), (Ojha, U., Li, Y., & Lee, Y. J. 2023).

In this context, handcrafted features provide an interpretable approach for representing measurable characteristics of images. Unlike deep-learning methods that learn representations automatically, handcrafted methods explicitly characterize properties such as texture, statistical distribution, frequency content, noise, color, and local structural patterns. These features can subsequently be provided to machine-learning classifiers to distinguish real and AI-generated images. Their explicit representation also enables the contribution of individual feature types to be analysed, providing greater interpretability and potentially reducing the computational requirements associated with training large deep-learning models. Motivated by these considerations, this study investigates a handcrafted-feature-based framework for distinguishing real images from AI-generated images using machine-learning techniques. The proposed approach extracts complementary image characteristics and evaluates their discriminative capability for binary classification. The study particularly focuses on determining whether appropriately selected and combined handcrafted features can provide reliable discrimination between real and AI-generated images.

The remainder of this paper is organized as follows. Section 2 presents the related work on AI-generated image detection, with emphasis on handcrafted, frequency-domain, fingerprint-based, and deep-learning approaches. Section 3 describes the proposed methodology, including dataset preparation, preprocessing, handcrafted feature extraction, and feature representation, classification, and evaluation measures. Section 4 reports and discusses the experimental results. Finally, Section 5 concludes the study and presents directions for future research.

2. Literature review

Research on distinguishing real images from AI-generated images has evolved considerably with the rapid development of generative models. Early studies primarily investigated image-forensic artifacts and generator-specific fingerprints, whereas recent research increasingly employs deep learning, hybrid feature representations, large-scale datasets, and generalized detection frameworks. The existing literature can broadly be categorized into handcrafted and forensic feature-based methods, deep-learning approaches, hybrid methods, diffusion-image detection, and approaches addressing generalization and robustness. Brief overview on different approaches used in the discrimination between AI-generated and real image is discussed in the following sections.

2.1. Handcrafted and Forensic Feature-Based Detection

Traditional image-forensic approaches attempt to distinguish synthetic images by identifying measurable differences in image statistics, texture, noise, frequency characteristics, and structural properties. Such characteristics can be represented explicitly using handcrafted features and subsequently classified using machine-learning algorithms. Frequency-domain analysis demonstrated that the generated images may contain distinctive spectral characteristics that can be exploited for detection (Frank, J., et al., 2020). Similarly, studies on artificial fingerprints showed that generative models can leave systematic patterns in synthesized images, which can be used to distinguish generated content from naturally captured images (Marra, F., et al, 2019), (Yu, N., Davis, L. S., & Fritz, M., 2019).

Recent research has continued to investigate low-level image characteristics for AI-generated image detection. Martin-Rodriguez et al investigated forensic characteristics based on Photo Response Non-Uniformity (PRNU) and Error Level Analysis (ELA), demonstrating the potential of combining complementary image-level evidence for distinguishing real and AI-generated images (Martin-Rodriguez, F., et al., 2023). These approaches are more relevant to handcrafted-feature-based detection because the characteristics are explicitly extracted from the image rather than learned entirely through a large neural network. Yu et al. further investigated a lightweight detection strategy based on image noise entropy (Yu, N., Davis, L. S., & Fritz, M., 2019). Their approach transforms images into an entropy representation of Non-Local Means (NLM) noise before classification. The large-scale dataset containing more than 730000 real and AI-generated images produced using multiple generation techniques, including text-to-image generation, inpainting, image refinement, and face swapping. The study reports that models trained on this dataset achieve strong generalization, indicating the potential of noise-related characteristics for synthetic-image detection (Yu, et al., 2025). These studies indicate that low-level characteristics such as noise, texture, statistical distributions, and frequency patterns can provide useful evidence for distinguishing AI-generated images. They also provide motivation for systematically investigating complementary handcrafted features within a machine-learning framework.

2.2. GAN Fingerprints and Synthetic-Image Artifacts

The development of Generative Adversarial Networks (GANs) introduced the concept of generator-specific fingerprints. Marra et al. demonstrated that GAN-generated images can contain artificial fingerprints associated with the generation process (Marra, F., et al, 2019). Yu et al. subsequently investigated the attribution of fake images to their generating GAN models by learning generator-specific fingerprints (Yu, N., Davis, L. S., & Fritz, M., 2019. These studies established that synthetic-image detection can exploit artifacts introduced by the generative architecture itself. However, generator-

specific characteristics can restrict generalization. A detector trained to identify the fingerprint of a particular GAN may not perform equally well when applied to images produced by another GAN or by a fundamentally different generation mechanism. This limitation has motivated subsequent research toward generator-independent and universal detection.

2.3. Deep-Learning based Detection

The increasing quality of AI-generated images has encouraged the use of CNN-based approaches that automatically learn discriminative representations. Wang et al. demonstrated that CNN-generated images can be detected effectively using convolutional neural networks, while also highlighting the difficulty of generalizing to images produced by previously unseen generators (Saharia, C., 2022). Bird and Lotfi investigated real-versus-AI image classification using a CNN trained on a synthetic dataset generated from CIFAR-10 classes using latent diffusion (Bird, J. J., & Lotfi, A., 2024). Their optimal network achieved 92.98% accuracy, and Grad-CAM analysis indicated that the classifier relied on small visual imperfections, particularly in background regions, rather than solely on the semantic objects within the images. The resulting CIFAKE dataset was also made publicly available for further research. Manish et al. have investigated different neural-network architectures for binary classification of real and AI-generated images. A comparative study reported the evaluation of nine neural-network architectures and emphasized the importance of regularization and data augmentation for improving classification stability and reducing overfitting (Raut, & Manish., 2026). ResNet50 has also been investigated for detecting subtle artifacts and inconsistencies, including background distortions, lighting anomalies, and unnatural textures that may not be readily identifiable through human visual inspection.

More recently, Transformer-based approaches have been investigated for AI-generated image detection. Lamichhane et al. applied a Vision Transformer (ViT) (Lamichhane, D., 2024) to a balanced real-versus-AI-generated image dataset and investigated the effects of patch size, attention heads, model depth, and data augmentation. The study reported strong classification performance and highlighted the capability of Transformer-based models to capture long-range visual dependencies. Although CNNs and Transformers provide powerful automatically learned representations, their performance is often strongly influenced by the training dataset and the generative models represented during training. Therefore, high performance on a particular benchmark does not necessarily imply robust performance on newly developed image generators.

2.4. Hybrid Handcrafted and Deep Features

To exploit the complementary advantages of explicit image descriptors and automatically learned representations, recent studies have investigated hybrid approaches. Handcrafted texture descriptors such as Local Binary Patterns (LBP) and Gray-Level Co-occurrence Matrix (GLCM) features have been combined with deep representations extracted from networks such as EfficientNet and ResNet (Martin-Rodriguez, F., et al., 2023), (Tabassum, S., et al., 2025). These approaches demonstrate that low-level handcrafted information can complement high-level deep representations. Similarly, studies combining spatial, frequency, histogram, color, texture, and CNN-derived features indicate that multiple representations can provide complementary information for synthetic-image classification (Tiwari, M. R., & Patil, S. S. 2026). Such approaches are particularly relevant to the present research because they demonstrate that AI-generated images may be distinguished using characteristics from multiple image domains. However, hybrid approaches generally increase model complexity and computational requirements when large, pre-trained networks are included. In addition, the individual contribution of handcrafted feature categories is often not systematically examined. This creates an opportunity to investigate whether carefully selected handcrafted features alone can provide effective and interpretable discrimination.

2.5. Generalization and Large-Scale Datasets

Generalization to unseen generative models is one of the central challenges in AI-generated image detection. Ju et al. investigated global and local feature representations to improve generalization across

multiple generation and manipulation models. Subsequent research further examined the effects of pre-processing operations and unseen generators (Ju, C., Bibaut, A., & van der Laan, M., 2018). The development of large and diverse datasets has become an important component of this research. Hong et al. introduced WildFake (Hong, Y., et al., 2025), a dataset designed to improve the diversity of AI-generated images by incorporating images from a broad range of contemporary generators, generator categories, applications, image classes, and styles. The hierarchical organization of the dataset was designed to facilitate evaluation of detection generalizability and robustness in practical conditions. Yu et al. similarly addressed limitations of existing datasets, including limited image diversity, low image quality, simple prompts, and restricted generalization. The proposed dataset contains more than 7,30,000 real and AI-generated images covering multiple categories and generation processes (Yu, et al., 2025). In addition to text-to-image synthesis, it includes image inpainting, refinement, and face-swapping operations, with additional annotations such as generation method and inpainting masks. These developments demonstrate that the quality and diversity of datasets are critical factors in evaluating AI-image detectors. A detector trained using a narrow set of generators may learn dataset or generator specific characteristics rather than general properties of synthetic images.

2.6. Robustness under Real-World Conditions

In addition to generator generalization, robustness under real-world image transformations is an important research problem. Images distributed through social-media platforms and other online services may undergo compression, resizing, repeated transmission, cropping, filtering, or re-digitization. Such operations can modify the subtle characteristics used by forensic detectors. Li et al. introduced the Real-World Robustness Dataset (RRDataset) to evaluate AI-image detectors under three dimensions: scenario generalization, internet-transmission robustness, and re-digitization robustness (Li, C., et al., 2025). The study benchmarked 17 detection systems and 10 vision-language models and included a human study involving 192 participants. The results highlighted limitations of existing detection methods under practical real-world conditions and emphasized the importance of robustness beyond conventional benchmark datasets. Pellegrini et al. addressed a related problem through AI-GenBench, a benchmark that evaluates detectors temporally according to the evolution of generative models (Pellegrini, l., et al., 2025). Rather than relying only on conventional random train-test splits, the benchmark evaluates whether detectors trained on earlier generators can generalize to subsequently developed models. This is particularly important for assessing the transition from GAN-based generation to diffusion-based generation and for obtaining a more realistic measure of detector generalization.

2.7. Diffusion-Generated Image Detection

The emergence of diffusion models has substantially changed the landscape of AI image generation. Diffusion-based systems can generate highly realistic and diverse images, creating new challenges for detection methods originally developed for GAN-generated content. Corvi et al. investigated the detection of synthetic images generated by diffusion models and demonstrated the need to examine characteristics specific to these newer generation mechanisms (Corvi, R., et al., 2023). Recent research has consequently shifted toward low-level, semantic, and hybrid representations that can generalize across different generation architectures. Studies combining high- and low-level features have attempted to capture both semantic information and subtle generation artifacts. This transition indicates that a detector designed specifically around a single generator family may become ineffective as new generation technologies emerge.

2.8. Universal, Multimodal, and Explainable Detection

Ojha et al. proposed a universal fake-image detection approach designed to generalize across different generative models (Ojha, U., Li, Y., & Lee, Y. J., 2023). This work represents an important transition from generator-specific detection toward generator-independent representations. More recent approaches have extended this concept through multimodal models and combinations of semantic and pixel-level information. Wang et al. introduced DFBench, a large-scale benchmark containing approximately 540000 images covering real, AI-edited, and AI-generated content from 12 state-of-the-

art generation models. The study also proposed a Mixture-of-Agents framework using multiple large multimodal models for deepfake detection (Wang, J., et al., 2025). Explainability has also become increasingly important. AI image detection system combining CNN prediction with handcrafted characteristics such as noise score, edge density, brightness, contrast, and color variance is discussed in (Tabassum, S., et al., 2025). The system additionally generates explanations and automated reports containing classification results, confidence values, feature analysis, and explanations. This integration of handcrafted characteristics with deep learning demonstrates the potential of explicitly measurable image properties for improving interpretability in AI-image detection. Tiwari and Patil reviewed machine-learning-based deepfake detection methods, including classical machine-learning classifiers, CNNs, attention mechanisms, multimodal systems, and hybrid feature-fusion approaches (Tiwari, M. R., & Patil, S. S., 2026). Their review identified cross-dataset generalization, explainability, adversarial susceptibility, and multimodal deepfake attacks as important unresolved challenges.

The reviewed literature demonstrates a clear progression from traditional forensic and handcrafted features to deep-learning, hybrid, Transformer, multimodal and universal detection approaches. Existing studies confirm that AI-generated images contain discriminative characteristics in their texture, noise, frequency distribution, statistical properties, local structures, and semantic representations. Handcrafted descriptors have demonstrated the ability to capture such characteristics explicitly, while deep-learning models provide powerful automatically learned representations. Nevertheless, several research gaps remain. First, many recent detection approaches rely heavily on deep neural networks and large-scale datasets, resulting in increased computational and training requirements. Second, handcrafted features are frequently incorporated as supplementary components within hybrid frameworks rather than being systematically evaluated as an independent and interpretable detection representation. Third, the contribution of different handcrafted feature categories and their complementary relationships has not been sufficiently explored. Fourth, detector performance can deteriorate when images originate from unseen generators or undergo compression, resizing, transmission, re-digitization, and other post-processing operations. Finally, the rapid development of diffusion and other generative models requires detection methods that are less dependent on the artifacts of a particular generator.

Therefore, there is a need for an interpretable and computationally efficient framework that systematically investigates complementary handcrafted characteristics for distinguishing real and AI-generated images. Motivated by these observations, present study focuses on extracting, and combining the handcrafted features that represent spatial, texture, statistical, frequency, noise, and related low-level properties, followed by machine-learning-based classification. The study aims to evaluate the discriminative contribution of these features and determine whether their combination can provide reliable detection of AI-generated images while maintaining interpretability and reduced dependence on computationally intensive deep architectures.

3. Methodology

The proposed method uses handcrafted image features to distinguish AI-generated images from real images. Five complementary feature groups are extracted from each image. Fast Fourier Transform (FFT), Photo Response Non-Uniformity (PRNU)-based noise features, Local Binary Pattern (LBP), Gray-Level Co-occurrence Matrix (GLCM), and Error Level Analysis (ELA) are the five features extracted from the images. The extracted features are normalized, combined into a single feature vector, and classified using a Random Forest classifier. The overall processing flow is shown in Figure 1.

3.1. Dataset Preparation

The primary dataset contains two classes, AI-generated and Real images. The authentic images are high-quality human-face photographs derived from the Flickr-Faces-HQ (FFHQ) collection (Karras, T., Laine, S., & Aila, T., 2019), while the AI-generated images are produced using a StyleGAN generator. The dataset was obtained from Kaggle (Kaggle) and is class-balanced, with predefined and disjoint training,

validation, and testing partitions. These partitions are used directly in the experiments to maintain a consistent and reproducible evaluation setup. The images are loaded using MATLAB's image Datastore, with class labels assigned according to their corresponding folder names. AI-generated images are considered the positive class, while real images are considered the negative class. The dataset was selected because of its large scale, balanced class distribution, standardized data partitions, and focus on human faces, which is an important application domain for AI-generated image detection.

3.2. Image Preprocessing

Each input image is read and resized into a spatial size of 224×224 pixels. The RGB image is represented using normalized double-precision values in the range [0, 1]. A grayscale version of the image is then generated for the frequency and texture feature extraction methods. The RGB image is retained for ELA because ELA is calculated from the differences between the original RGB image and its JPEG-recompressed version.

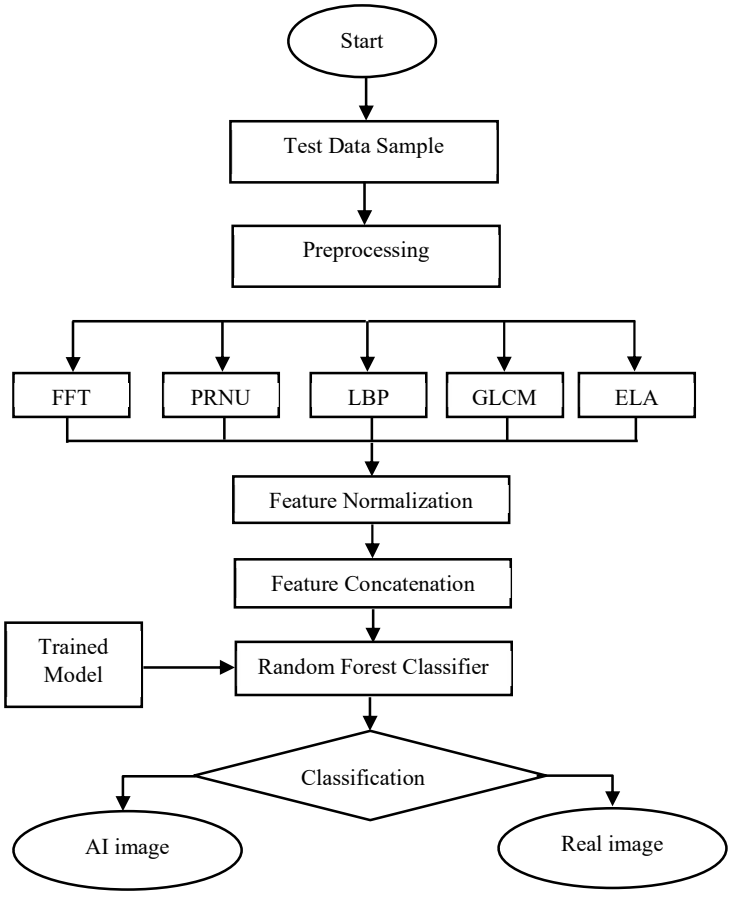


Figure 1. AI-image detection system. architecture.

3.3 Handcrafted Feature Extraction

Five feature groups are extracted to represent different level characteristics of the images. These features are selected to capture texture, spatial statistics, frequency, noise, and compression-related information. The overview of these feature extraction methods is given below.

3.3.1. FFT-Based Frequency Features

The two-dimensional Fourier transform of the grayscale image $I(x,y)$ is calculated using Eq. 1. The magnitude spectrum is shifted so that the zero-frequency component is located at the center. A logarithmic transformation is then applied using Eq. 2.

$$F(u,v) = F\{I(x,y)\} \quad (1)$$

$$M(u,v) = \log(1 + |F(u,v)|) \quad (2)$$

To obtain a compact representation of the frequency distribution, the spectrum is divided into 32 radial frequency bins. The mean magnitude within each radial bin is calculated. In addition, four global statistics are obtained from the spectrum. They are, mean, standard deviation, skewness, and kurtosis. Thus, the FFT feature group contains (32+4=36) features.

3.3.2. PRNU-Based Noise Features

PRNU-related characteristics are extracted from the grayscale image by estimating a high-frequency noise residual. A Gaussian filter with a standard deviation of 1 is used to obtain a smoothed version of the image. The residual is calculated as shown in Eq. 3.

$$R = I - I_{denoised} \quad (3)$$

Statistical characteristics of the residual are then calculated, including mean, standard deviation, skewness, kurtosis, energy, and median absolute residual. A 16-bin normalized histogram of the residual is also calculated. The PRNU feature group therefore contains (6+16=22) features.

3.3.3. LBP-Based Texture Features

The local texture characteristics of an image are represented by LBP features. The LBP descriptor is extracted from the grayscale image using MATLAB's uniform LBP implementation with upright orientation and L2 normalization. The resulting uniform LBP histogram contains 59 features. These features describe local intensity patterns and texture structures within the image.

3.3.4. GLCM-based Texture Features

Gray-Level Co-occurrence Matrix (GLCM) features are extracted to describe spatial relationships between gray-level pixels. Four spatial directions are considered with the orientation of $(0°, 45°, 90°, 135°)$. The grayscale image is quantized into eight-gray levels. For each direction, four GLCM properties of contrast, correlation, energy, and homogeneity are calculated. Hence, the GLCM feature vector has (4 x 4 = 16) features.

3.3.5. ELA-Based Compression Features

The ELA feature is used to characterize differences in JPEG recompression behavior. The input RGB image is saved as a JPEG image with a quality factor of 90 and then read again. The absolute difference between the original and recompressed images is calculated using Eq. 4.

$$E(x,y,c) = |I(x,y,c) - I_R(x,y,c)| \quad (4)$$

The three colour-channel error maps are combined to obtain a grayscale ELA map. From this map, seven global statistics are extracted. They are; mean, standard deviation, maximum, median, skewness, kurtosis, and squared-error energy. The mean error of the R, G, and B channels is also calculated. In addition, a 16-bin normalized histogram of the ELA values is obtained. The ELA feature group therefore contains (7+3+16=26) features.

3.3.6. Feature Normalization and Fusion

Each feature group is independently Z-score normalized before feature concatenation. For a feature$\mathbf{f}$, the normalized feature is calculated by Eq. 5.

$$F_i = \frac{f_i - \mu}{\sigma} \quad (5)$$

Where, μ and σ are the mean and standard deviation of the elements within that feature groups. The normalized feature groups are then concatenated to form a single feature vector leading a feature vector size of 159. Hence, each image is represented by a 159-dimensional handcrafted feature vector. After feature fusion, column-wise standardization is performed before Random Forest training. The mean and standard deviation for this operation are calculated only from the training data. The same training statistics are then applied to the validation and test sets.

4. Experiments and results

The simulation experiments are carried out under the MATLAB environment in a computer with the specifications of 12th Gen Intel(R) Core(TM) i5-12450HX, 2.40 GHz, 16GB RAM. The image dataset collected from Kaggle consists of 140000 facial image samples, with the two classes equally balanced, comprising of 70000 AI-generated images and 70000 Real images. Out of 70000 image samples from each group, 50000 image samples are used for training, 10000 image samples are used for validation, and 10000 image samples are used for testing. The image samples are subjected to preprocessing, and feature extraction. The extracted features are normalized using Z-score normalization technique.

The fused handcrafted feature vectors are classified using a Random Forest implemented with MATLAB's TreeBagger. A minimum leaf size of 5 is used to control the growth of individual trees. The number of trees is selected using the validation set. Four candidate values are evaluated such that, $T \in 50,100,200,300$. For each candidate value, a Random Forest is trained using the training set and its classification accuracy is measured using the validation samples. The number of trees giving the highest validation accuracy is selected. If two candidate values provide the same accuracy, the smaller forest is selected. The final Random Forest is then trained using the selected number of trees. Out-of-bag prediction is enabled for the final model to obtain an additional model diagnostic. The test set is not used during feature standardization, tree-count selection, or model training. It is used only for the final performance evaluation. The simulation is repeated for the fused feature vector of size 159, and also for individual features.

The performance of the proposed handcrafted-feature classifier is evaluated using accuracy, precision, recall, F1-score and area under the receiver operating characteristic curve (AUC-ROC). The AI-generated class is treated as the positive class. The ROC curve is generated using the Random Forest score for the AI-generated class, and the area under this curve is reported as AUC-ROC. A confusion matrix is generated to show the classification performance for the AI-generated and real classes. The recognition accuracy of 85.21% is achieved for five feature combinations.

The performance of the classifier for individual feature is shown in Table 1. The recognition accuracies listed in Table 1 reveals that the accuracy of the classifier is better for FFT, LBP, and ELA features when compared to PRNU, and GLCM features. The simulation experiment is repeated to validate the performance of the classier for the fused FFT, LBP, and ELA features. The performance of the classifier for three feature combinations is listed in Table 2. The recognition accuracy of 86.03% is achieved for the three-feature combinations, which is better than 85.21% achieved for the five feature combination. The confusion matrix and AUC-ROC for the fused feature vector, FFT, LBP and ELA are shown in Figure

2. It is clear from Figure 2 that the model is generalized and not biased toward single class of images in the dataset. Which means it has learnt the features from both AI generated and Real images.

Table 1. Classifier accuracy for individual features.

Feature	Accuracy
FFT	83.4%
PRNU	72.5%
LBP	84.1%
GLCM	71.9%
ELA	79.2%

Table 2. Classifier performance for three feature combination.

Feature Group	Accuracy	Precision	Recall	F1 score
FFT+LBP+ ELA	86.03%	86.15%	85.87%	86.01%

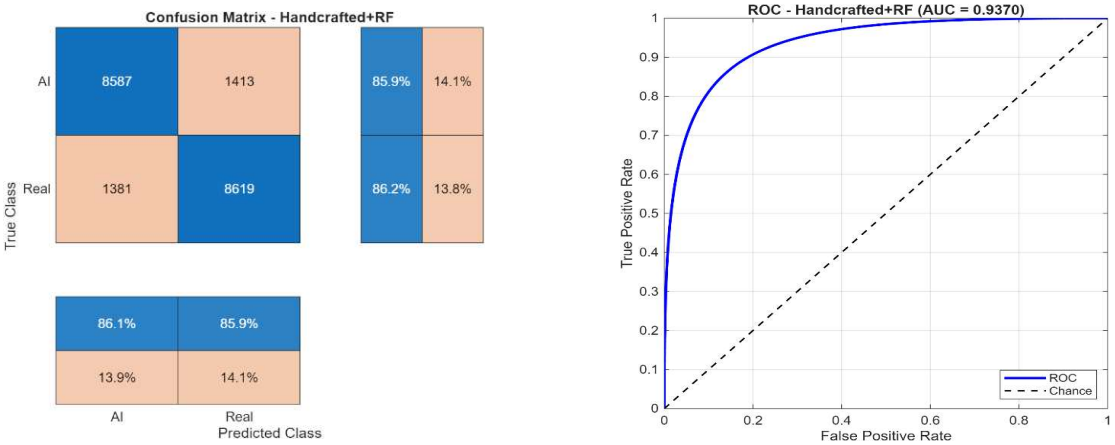


Figure 2. Confusion matrix and AUC-ROC curve.

5. Conclusion

This study presented a handcrafted feature-based approach for distinguishing AI-generated images from real images using a Random Forest classifier. Performance of the classifier is evaluated for five feature sets, FFT, PRNU, LBP, GLCM, and ELA. Among the individual features, LBP achieved the highest accuracy of 84.1%, followed by FFT with 83.4% and ELA with 79.2%, while PRNU and GLCM achieved comparatively lower accuracies of 72.5% and 71.9%, respectively. The combination of FFT, LBP, and ELA provided the best classification performance, achieving an accuracy of 86.03%. In comparison,

combining all five feature groups resulted in an accuracy of 85.21%, indicating that adding PRNU and GLCM did not improve the overall performance for the considered dataset. These results show that carefully selected handcrafted features can effectively capture frequency, texture, and compression related differences between real and AI-generated images. Overall, the proposed approach provides a simple and interpretable method for AI-generated image detection without relying on computationally intensive deep-learning models. Future work focuses on evaluating the selected features on images generated by different generative models and under common image-processing operations to study their generalization and robustness.

Declarations

Funding: This research received no external funding.

Conflict of Interest: The authors declare no conflict of interest.

Ethical Approval and Informed Consent: Not applicable.

Data Availability: Not applicable.

References

Bird, J. J., & Lotfi, A., (2024). CIFAKE: Image classification and explainable identification of AI-generated synthetic images. *IEEE Access*, 12, 15642–15650. https://doi.org/10.1109/ACCESS.2024.3356122.

Corvi, R., Cozzolino, D., Zingarini, G., Poggi, G., Nagano, K., & Verdoliva, L. (2023). On the detection of synthetic images generated by diffusion models. *In ICASSP IEEE Int. Con. on Acoustics, Speech and Signal Processing* (pp. 1-5). IEEE. https://doi.org/10.1109/ICASSP49357.2023.10095167

Frank, J., Eisenhofer, T., Schönherr, L., Fischer, A., Kolossa, A., & Holz, T. (2020). Leveraging frequency analysis for deep fake image recognition. *In Proc. 37th Int. Conf. Machine Learning* (pp. 3247–3258). https://doi.org/10.48550/arXiv.2003.08685

Goodfellow, I., Pouget-Abadie, J., Mirza, M., Xu, B., Warde-Farley, D., Ozair, S., Courville, S. & Bengio, Y. (2014). Generative adversarial nets. *Adv. in Neural Inf. Processing Systems*, 27, 2672–2680. https://doi.org/10.48550/arXiv.1406.2661

Ho, J., Jain, A., & Abbeel, P. (2020). Denoising diffusion probabilistic models. *Adv. in Neural Information Processing Systems*, 33, 6840–6851. https://doi.org/10.48550/arXiv.2006.11239

Hong, Y. Feng, J., Chen, H., Lan, J., Zhu, H., Wang, W., & Zhang, J. (2025). WildFake: A large-scale and hierarchical dataset for AI-generated images detection. *In Proc. AAAI Conf. Artificial Intelligence*, 39 (4), pp. 3500–3508. https://doi.org/10.1609/AAAI.V39I4.32363.

Ju, C., Bibaut, A., & van der Laan, M. (2018). The relative performance of ensemble methods with deep convolutional neural networks for image classification. *Jl. Applied Statistics*, 45(15), 2800–2818. https://doi.org/10.1080/02664763.

Kaggle: https://www.kaggle.com/datasets/xhlulu/140k-real-and-fake-faces?select=real_vs_fake

Kingma, D. P., & Welling, M. (2019). An Introduction to Variational Autoencoders. Foundations and *Trends in Machine Learning*, 12, 307-392. https://doi.org/10.1561/2200000056

Karras, T., Laine, S., & Aila, T. (2019). A style-based generator architecture for generative adversarial networks. *In Proc. IEEE/CVF Conf. Computer Vision and Pattern Recognition* (pp. 4401–4410). IEEE. https://doi.org/10.1109/TPAMI.2020.2970919

Lamichhane, D. (2024). Advanced detection of AI-generated images through vision transformers. *IEEE Access*, 13, 3644–3652. https://doi.org/10.1109/ACCESS.2024.3522759.

Li, C., Wang, X., Li, M., Miao, B., Sun, Zhang, Y., Ji, X., & Zhu, Y. (2025). Bridging the gap between ideal and real-world evaluation: Benchmarking AI-generated image detection in challenging

scenarios. *In Proc. IEEE/CVF Int. Conf. Computer Vision* (pp. 20379–20389). https://doi.org/10.1109/ ICCV51701.2025.01895. IEEE.

Marra, F., Gragnaniello, D., Verdoliva, L., & Poggi, G. (2019). Do GANs leave artificial fingerprints?. *In Proc. IEEE Conf. Multimedia Information Processing and Retrieval* (pp. 506–511). IEEE. https://doi.org/10.1109/MIPR.2019.00103

Martin-Rodriguez, F., Garcia-Mojon, R., & Fernandez-Barciela, M. (2023). Detection of AI-created images using pixel-wise feature extraction and convolutional neural networks. *Sensors*, 23(22), 9037. https://doi.org/10.3390/s23229037

Ojha, U., Li, Y., & Lee, Y. J. (2023). Towards universal fake image detectors that generalize across generative models. *In Proc. IEEE/CVF Conf. Computer Vision and Pattern Recognition (pp. 24480–24489).* IEEE.

Pellegrini, l., Cozzolino, d., Pandolfini, S., Maltoni, D., Ferrara, M., Verdoliva, L., Prati, M., & Ramilli, M. (2025). AI-GenBench: A new ongoing benchmark for AI-generated image detection. *In Proc. International Joint Conference on Neural Networks* (pp. 1–9). https://doi.org/10.1109/ IJCNN64981.2025. 11228377.

Raut, & Manish. (2026). Comparative analysis of neural network architectures: CNNs, RNNs, and Transformers in real-world applications. *Jl. Applied Technology and Innovation*, 10(1), 103-12. https://doi.org/10.1109/10.65136/jati.v10i1.320.

Radford, A., Jong Wook Kim, Chris Allacy, and Aditya Ramesh (2021) Learning transferable visual models from natural language supervision. In Proc. Int. Conf. Machine Learning (pp. 8748–8763). https://doi.org/10.48550/arXiv.2103.00020

Rombach, R., Blattmann, A., Lorenz, D., Esser, P., & Ommer, B. (2022). High resolution image synthesis with latent diffusion models. *In Proc. IEEE/CVF Conf. on Computer Vision and Pattern Recognition (pp. 10684–10695)*. IEEE. https://doi.org/10.1109/CVPR52688.2022.01042

Saharia, C. (2022). Photorealistic text-to-image diffusion models with deep language understanding. *Adv. in Neural Information Processing Systems*. 35, 36479–36494. https://doi.org/10.48550/arXiv.2205.11487

Tabassum, S., Pasha, S. Z., & Anitha, T. (2025). Image classification and explainable identification of AI-generated synthetic images. *Int. Jl. of Information Technology and Computer Engineering*, 13 (2), 77–85. https://doi.org/10.62647/IJITCE2025V13I2sPP77-85.

Tiwari, M. R., & Patil, S. S. (2026) Deep-fake image detection using machine learning techniques: A comprehensive review. *Int. Jl. on Advanced Electrical and Computer Engineering*. 15(1S), 137–145. https://doi.org/10.65521/IJAECE.V15I1S.1350.

Verdoliva, l. (2020). Media forensics and deepfakes: An overview. *IEEE Journal of Selected Topics in Signal Processing*. 14 (5), 910–932.

Wang, J., Duan, H., Wang, J., Jia, Z., Yang, W. Y., Zhu, X., Zhao, Y., Qian, J., Xing, Y., Zhai, G. & Min, X. (2025). DFBench: Benchmarking deepfake image detection capability of large multimodal models. *In Proc. 33rd ACM International Conference on Multimedia*. https://doi.org/10.1145/3746027.3758204.

Yu, N., Davis, L. S., & Fritz, M. (2019). Attributing fake images to GANs: Learning and analyzing GAN fingerprints. *In Proc. IEEE/CVF Int. Conf. Computer Vision* (pp. 7556–7566). IEEE.

Wang, S, Y., Wang, O., Zhang, R., Owens, A., & Efros, A. A. (2020). CNN-generated images are surprisingly easy to spot… for now. *In Proc. IEEE/CVF Conf. Computer Vision and Pattern Recognition* (pp. 8695–8704). IEEE. https://doi.org/10.48550/arXiv.1912.11035

Yu, H., Ye, Y., Rong, J., Xuan, Q., & Ma, C. (2025). RealHD: A high-quality dataset for robust detection of state-of-the-art AI-generated images. *In Proc. 33rd ACM Int. Conf. Multimedia* (pp. 11394–11403). https://doi.org/10.1145/3746027.3755207.