Systemic Analytics

www.sa-journal.org

Syst. Anal. Vol. 3, No. 2 (2025) 152–163.

Paper Type: Original Article

Distinguishing AI-Generated Images from Real Images: A Multi-Criteria Analysis Using SF-AHP

Ahmet Suha Hancioglu[1,*], Ibrahim Yilmaz[1]

[1] Department of Industrial Engineering, Faculty of Engineering and Natural Sciences, Ankara Yıldırım Beyazıt University, Ankara, Turkey; suhahancioglu@gmail.com; i.yilmaz@aybu.edu.tr.

Citation:

Received: 25 February 2025 *Revised: 21 April 2025* *Accepted: 17 May 2025*	Suha Hancioglu, A. & Yilmaz, I. (2025). Distinguishing AI-generated images from real images: A multi-criteria analysis using SF-AHP. *Systemic Analytics, 3*(2), 152-163.

Abstract

Recent advancements in the widespread use of the internet and computers have led to the storage of vast amounts of data, increasing computational power, and enhanced computer capabilities, fundamentally transforming human life. The large data sets generated through the internet provide an unlimited source of data for Artificial Intelligence (AI). These developments, coupled with current AI techniques, challenge the traditional concepts of photography and images, particularly due to the ability to produce realistic visuals. As synthetic images produced by AI become increasingly indistinguishable from real, human-made photographs, the distinction between the two is becoming more difficult over time. Considering the development of Large Language Models (LLM), which can produce highly fluent and coherent communication almost indistinguishable from human-generated text within a few years of their introduction, it is expected that synthetic images will rapidly increase in realism as well. This situation has become a critical issue in preventing disinformation and maintaining visual credibility. This study addresses this problem and aims to present a holistic approach for distinguishing between real and synthetic images. In this context, a comprehensive evaluation framework consisting of six main criteria and eighteen sub-criteria is presented, and the relative importance of these criteria is analyzed using the Spherical Fuzzy AHP (SF-AHP) method. These importance levels should be continuously updated and discussed as technology evolves. In this study, an innovative approach is used, incorporating expert opinions (such as those from advanced language models like ChatGPT or OpenAI-3), and based on SF-AHP analysis, physical consistency and sensor trace analysis emerged as the most critical determinants for distinguishing between synthetic and real images. The findings emphasize the necessity of a multi-criteria approach in AI image detection and provide insights for future validation methods.

Keywords: AI-generated images, Image authentication, Spherical fuzzy AHP (SF-AHP), Multi-criteria decision making.

1 | Introduction

The widespread use of computers and the internet has accelerated technological developments worldwide. As a natural result, technological progress has gained momentum, and computational power has rapidly increased. The global spread of the internet and computers has led to the development of data storage technologies, resulting in the production of 149 zettabytes of data by the end of 2024 [1].

Corresponding Author: suhahancioglu@gmail.com

https://doi.org/10.31181/sa32202550

This massive volume of data has enabled rapid advancements in AI and Machine Learning (ML) methods, Deep Learning (DL) technologies, and generative models (e.g., Generative Adversarial Networks, diffusion models) capable of producing realistic images that resemble human-created ones [2]. For instance, models like DALL-E, Midjourney, and Stable Diffusion can synthetically generate photograph-like scenes from short text descriptions [3]. These advancements have raised new questions about the reliability of digital images in a world where data spreads globally in seconds via the internet and leaves a permanent trace. Photographs, once considered "physically real," can now be manipulated or artificially produced. In fact, in 2023, a realistic AI-generated image of Pope Francis wearing an extravagant coat spread rapidly on the internet. However, it was false; it initially deceived millions, demonstrating the potential impact of visual disinformation [4]. Similarly, AI-generated images of political figures, celebrities, or significant events can be produced and distributed to influence public opinion, contributing to disinformation and creating an "I can no longer trust anything I see" perception among the conscious members of society. Unfortunately, this situation facilitates social engineering activities due to widespread internet usage and contributes to disinformation through shaping public perception [5].

The inability to detect AI-generated images leads to the spread of misinformation, defamation campaigns, fake news, public safety risks, economic damage, and even the potential to affect vital national events such as general elections [6], [7]. Therefore, verifying the authenticity of digital images has become crucial not only for individuals but also for journalism and digital forensics [8]. However, this task is becoming increasingly complex as the differences between artificial images and real photographs are gradually becoming less distinct. In early-generation artificial images, visible anatomical errors (e.g., abnormal numbers of fingers or physically impossible body parts) or inconsistent object details are significantly reduced or completely absent in newer generation models. These newer image generation tools have reached the point where they can flawlessly render human hands and fingers, eliminate physical inconsistencies, greatly reduce anomalies in details, and remove obvious visual cues that would previously expose them as fake. Current research shows that the average person is not significantly better than chance at distinguishing whether an image is synthetic or real [9].

On the other hand, the innovation that synthetic image creation can bring to the art field is undeniable [10], [11]. These revolutionary innovations in visual creation have led to the emergence of new, more liberal areas. While the positive uses of these new methods of image generation will undoubtedly improve human life, making the distinction between good and bad uses is crucial in a world where global data transmission occurs via the internet.

Under these conditions, a multifaceted and systematic approach is needed to distinguish artificial and real images. Relying on a single feature or technique may be inadequate in light of rapidly developing technologies. Approaches must be continuously evaluated and updated based on evolving technologies. In the literature, various methods for detecting artificial images have been proposed: detecting physical inconsistencies in images [12], analyzing digital noise and sensor traces [13], examining statistical properties such as frequency spectrum or pixel distribution [14], evaluating content context and logical consistency, verifying metadata and image source, and using advanced DL-based detectors [15]. This study aims to integrate all these approaches, which are considered necessary in the literature and by experts, into a holistic framework.

In this study, six main criteria and eighteen sub-criteria have been defined to detect AI-generated images. Each criterion addresses the differences between artificial and real visuals from a different perspective. Then, the relative importance of these criteria is quantitatively determined using a multi-criteria decision-making method, Spherical Fuzzy AHP (SF-AHP). The required pairwise comparisons for the SF-AHP analysis were innovatively obtained by including ChatGPT-o-3, an advanced AI model capable of synthesizing literature knowledge, as a participant in the expert panel. This allows for the provision of rich expert input, reflecting both academic literature findings and practical experience. The access of LLM models to raw information in literature and the innovative use of this knowledge in academia are presented. The criterion weights calculated using SF-AHP highlight which indicators are more critical under current technological developments.

In the continuation of this paper, similar studies and existing approaches for detecting artificial images are summarized in the Literature Review section. In Criteria Selection, each criterion included in the study is explained. In the Methodology section, the process of determining the criteria and applying the SF-AHP analysis is detailed. In the Results and Discussion section, the numerical results obtained from the SF-AHP analysis are presented, and the most important criteria are discussed, concluding artificial image detection. Finally, in the Conclusion section, a general evaluation of the study is provided, along with suggestions and future research directions.

2 | Literature Review

A recent survey on face manipulation and deepfake detection highlights the value of machine-learning multi-cue strategies for spotting forged videos. The review systematically covers four main manipulation categories—complete face synthesis, identity swap, attribute editing, and expression change—comparing generation techniques, public datasets, and current detection benchmarks, and concludes that fused, machine-learning-based cues are crucial for reliable forgery recognition [16]. Another up-to-date study on synthetic content across different media stresses the need for versatile countermeasures against advanced fakes; comparative experiments show that stylometric methods alone reach about 80 % accuracy, whereas hybrid, modality-fusion models achieve up to 92 %, and the authors discuss challenges such as robustness under real-world distortions, privacy risks, false-positive impacts and the priority of developing adaptive, ethically supervised detectors [17]. One of the first deep-fake detectors, the lightweight CNN model MesoNet, proved effective at capturing artifacts in manipulated face videos [18]. A frequency-domain approach similarly revealed hidden regularities, enabling detection of deepfake images that look authentic in the spatial domain [19]. Exploiting physiological inconsistencies, another study showed that irregular pupil shapes in GAN-generated faces can be segmented automatically and used to distinguish synthetic portraits from FFHQ and StyleGAN2 data with high accuracy [12]. Finally, SSDNeRF introduces a single-stage paradigm that jointly trains a NeRF auto-decoder and a latent diffusion model, delivering accurate unconditional 3-D scene generation and sparse-view reconstruction within one unified framework [20].

3 | Criteria Selection

C1 Physical Consistency and Realism examines whether the scene's lighting, perspective, and object interactions are consistent with the laws of physics. C1-1 Light and Shadow Consistency captures inconsistencies typical in synthetic scenes by comparing the angle of shadows with the direction of the light source and ensuring continuity in reflections and gloss. C1-2 Perspective and Geometric Consistency checks the distance–near ratio, the convergence of parallel lines, and depth distribution; any deviation, such as the scaling of windows in buildings or distortions in interior lines, reveals synthetic production. C1-3 Object Interactions and Physical Logic identifies impossibilities in object contact, such as hand-object or wheel-ground interactions, and exposes flaws like floating objects or poorly fitted clothing.

C2 Detail Quality and Visual Anomalies investigates micro-level errors. C2-1 Anatomical and Fine Detail Consistency detects excess or missing fingers, asymmetric pupils, and synthetic hair textures; GAN flaws that remain detectable in complex poses become visible under this category. C2-2 Texture and Pattern Anomalies identifies repeating motifs in the background and unfinished edge patterns by comparing them with statistical randomness; pattern repetitions can recognize synthetic production. C2-3 Focus and Depth Consistency analyzes bokeh distribution and depth-of-field effects, comparing them with lens physics to reveal "over-perfect" focus errors, such as all plans being equally sharp or blurring randomly distributed.

C3 Sensor Traces and Metadata Analysis searches for camera-related signatures within the file structure. C3-1 Sensor Noise and PRNU Analysis identifies the absence or synthetic addition of unique photo-response inconsistencies in real cameras, detecting artificial images at the fingerprint level. C3-2 EXIF and File Metadata captures abnormalities such as unknown camera models, missing timestamps, or unusual resolutions, pointing to algorithmic origins. C3-3 Compression Artifacts and Pixel Distribution assesses

medium-scale unusual pixel patterns and color quantization deviations, identifying multi-stage production traces.

C4 Statistical Features and Model Trace compares the frequency and pixel statistics of the image to natural photographic distributions. C4-1 frequency spectrum analysis defines deviations from log-log energy decay and peak points in mid-high frequencies as model-specific signatures. C4-2 Color Statistics and Pixel Distribution measures RGB channel correlations and histogram deviations to find over-generalization or unusual color balance. C4-3 Artificial Model Fingerprints traces spectral-spatial imprints specific to models like StyleGAN or diffusion models by matching with reference libraries to identify the source. C5 contextual and logical consistency evaluates whether the scene aligns with real-world knowledge. C5-1 reality and logic check searches for evident contradictions like seasonal mismatches, physically impossible positions, or historical anachronisms within the same frame. C5-2 In-Scene Context and Consistency examines story-object alignment, perspective-light cohesion, and period-costume matching to detect logical inconsistencies. C5-3 External Source and Reverse Search Verification follows the digital trace of the visual to check for similar sources in trusted platforms; community notes on social networks and reverse search results can quickly reveal synthetic origins.

C6 Expert/Tool-Supported Analysis Methods combine human experience and technological aids. C6-1 automatic AI detection tools enable DL-based detectors to classify the absence of sensor noise or spectral anomalies quickly; they offer initial filtering power by scanning thousands of images per minute. C6-2 human expert review and forensic analysis uncovers subtle clues missed by automatic tools through error-level analysis or pixel-level spectral filters; critical judgments in forensic cases depend on this. C6-3 digital watermarks and verification technologies verify the source by checking cryptographic signatures like content credentials or C2PA, detecting invisible platform stamps; in watermark-free images, suspicion of tampering arises. This multi-layered schema combines physical and technical findings with contextual and expert evidence, calculating weightings through SF-AHP to create the most reliable decision structure.

4 | Methodology

This study is designed with a holistic approach to identify the criteria that can be used for detecting AI-generated images and to determine the importance levels of these criteria quantitatively. In the first phase, based on a literature review and expert opinions in the field of visual media verification, six main criteria and three sub-criteria under each were defined. In selecting these criteria, a comprehensive approach was adopted to encompass all the methods and insights discussed in the previous section; different dimensions were considered, ranging from physical consistency and metadata analysis to content consistency and automatic detection tools. In the second phase, the SF-AHP, an advanced fuzzy logic-based variant of the Analytic Hierarchy Process (AHP), was used to calculate the relative importance of these identified criteria. This approach integrated uncertainties and indecisiveness in expert judgments, resulting in a more flexible and reliable decision model.

In complex decision-making scenarios, experts often face challenges in providing entirely specific assessments, especially in the absence of quantitative data. In such situations, intuitive judgments and expertise become even more crucial. Fuzzy logic, introduced as a solution to these challenges, provides a framework for modeling uncertainty through the concept of fuzzy sets, allowing decision-makers to obtain more objective outcomes. Initially introduced by Lotfi A. Zadeh in 1965, fuzzy logic extends classical logic by assigning a degree of membership for an element to a particular property on a continuous scale between 0 and 1 [21]. This approach makes it possible to represent ambiguous areas or partial membership, replacing the rigid dichotomies such as black-and-white or yes-and-no, thereby enabling the quantification of uncertainty or subjective judgments. Over time, various fuzzy set models have been proposed, one of the more recent being spherical fuzzy sets, introduced by Gündoğdu and Kahraman in 2019 [22]. These sets were designed to offer decision-makers a more expansive framework for expressing their preferences [23].

For a fuzzy set $\widetilde{A}$ defined on a universe of discourse X, the membership function $\mu_{\widetilde{A}}(x)$ quantifies the degree to which an element $x \in X$ belongs to $\widetilde{A}$, as shown in *Eq. (1)*.

$$\widetilde{A} = \{(x, \mu_{\widetilde{A}}(x)) | x \in X\}. \quad (1)$$

Spherical fuzzy sets were introduced to enhance the flexibility of Intuitionistic and Pythagorean fuzzy sets, offering a more adaptable approach to represent uncertainty. A spherical fuzzy set $\tilde{A}$ over a universe X is characterized by three degrees: membership $\mu_{\tilde{A}}(x)$, non-membership $v_{\tilde{A}}(x)$, and hesitation $\pi_{\tilde{A}}(x)$, as shown in *Eq. (2)*, and subject to the constraint in *Eq. (3)*.

$$\widetilde{A} = \{(x, \mu_{\widetilde{A}}(x), v_{\widetilde{A}}(x), \pi_{\widetilde{A}}(x)) | x \in X\}. \quad (2)$$

$$\mu_{\widetilde{A}}(x), v_{\widetilde{A}}(x), \pi_{\widetilde{A}}(x) \leq 1. \quad (3)$$

This condition ensures that the degrees remain within the unit sphere in a three-dimensional Cartesian coordinate system.

AHP-based criteria weighting using Spherical Fuzzy Numbers

The AHP helps to determine the relative importance of various criteria via pairwise comparisons. When paired with spherical fuzzy sets, expert evaluations are expressed using Spherical Fuzzy Numbers (SFNs).

The decision problem is structured hierarchically, with the main goal at the top level and criteria at the next level. Let $X = \{C_1, C_2, \ldots, C_n\}$ represent the set of n criteria to be evaluated. Decision-makers offer linguistic judgments for pairwise comparisons of the criteria, which are mapped into SFNs. This forms the pairwise comparison matrix $\widetilde{A} = [\tilde{a}_{ij}]_{nxn}$, where $\tilde{a}_{ij} = (\mu_{ij}, v_{ij}, \pi_{ij})$ denotes the spherical fuzzy preference of criterion C_i over C_j. If there are multiple experts (k decision-makers), their individual judgments $\tilde{a}_{if}^{(k)}$ are aggregated using the Spherical Fuzzy Weighted Averaging (SFWA) operator, where w_k represents the weight of the k-th expert, as shown in *Eq. (4)*.

$$\tilde{a}_{ij} = \left(\left(\sum_{k=1}^{K} w_k \mu_{ij}^{(k)}\right), \left(\sum_{k=1}^{K} w_k v_{ij}^{(k)}\right), \left(\sum_{k=1}^{K} w_k \pi_{ij}^{(k)}\right)\right). \quad (4)$$

For each criterion C_i, the synthetic extent value $\tilde{S}_i$ is calculated as the geometric mean of its fuzzy judgments, as shown in *Eq. (5)*.

$$\tilde{S}_i = \left(\prod_{j=1}^{n} \mu_{ij}\right)^{\frac{1}{n}}, \left(\prod_{j=1}^{n} v_{ij}\right)^{\frac{1}{n}}, \left(\prod_{j=1}^{n} \pi_{ij}\right)^{\frac{1}{n}}. \quad (5)$$

To obtain crisp weights, the spherical fuzzy weights $\widetilde{w}_i$ are defuzzified using a score function. A commonly used score function is shown in *Eq. (6)* [14].

$$S(\widetilde{w}_i) = \left(\mu_{w_i} - \pi_{w_i}\right)^2 - \left(v_{w_i} - \pi_{w_i}\right)^2. \quad (6)$$

Using the score function, the normalized weight w_i for each criterion is calculated, where $\sum_{i=1}^{n} w_i = 1$, as shown in *Eq. (7)*.

$$w_i = \frac{S(\tilde{s}_i)}{\sum_{j=1}^{n} S(\tilde{s}_j)}. \quad (7)$$

Finally, the consistency of the pairwise comparison matrix is assessed to ensure the reliability of the judgments. First, the spherical fuzzy matrix $\widetilde{A}$ is converted to a crisp matrix $A = [a_{ij}]$ using the score function $a_{ij} = S(\tilde{a}_{ij})$. The Consistency Ratio (CR) is calculated as shown in *Eq. (8)*, where the Consistency Index (CI) is calculated as presented in *Eq. (9)*.

$$CR = \frac{CI}{RI}. \quad (8)$$

$$CI = \frac{\lambda_{max} - n}{n - 1}. \quad (9)$$

In SF-AHP, linguistic evaluations are converted into SFNs using predefined scales, as presented in *Table 1*. This facilitates a more expressive representation of pairwise comparisons.

Table 1. Linguistic scale and corresponding SFNs.

Linguistic Term	Membership (m)	Non-membership (v)	Hesitation (π)	Spherical Importance (SI)
Very high	0,9	0,1	0	9
High	0,8	0,2	0,1	7
Medium high	0,7	0,3	0,2	5
Moderate	0,6	0,4	0,3	3
Medium low	0,5	0,4	0,4	1
Low	0,4	0,6	0,3	1/3
Very low	0,3	0,7	0,2	1/5
Critically low	0,2	0,8	0,1	1/7
Extremely low	0,1	0,9	0	1/9

Six main criteria have been identified, each providing critical evidence from different perspectives for detecting AI-generated fake images. These criteria are C1 (Physical consistency and realism), C2 (Detail quality and visual anomalies), C3 (Sensor Traces & Metadata Analysis), C4 (Statistical Features & Model Traces), C5 (Contextual and logical consistency), and C6 (Expert-/Tool-Supported Analytical Methods). The global fuzzy pairwise comparison matrix created to determine the relative importance of these criteria is presented in *Table 2*. The table shows, in each cell, the degree of superiority of criterion i over criterion j, represented by the values (μ, ν, π). During the comparisons, the main criteria were first compared in terms of their importance, followed by comparisons of the sub-criteria under each main criterion. This process resulted in separate pairwise comparison matrices for the two levels of the hierarchical model. The comparison scale used included linguistic evaluations commonly employed in fuzzy AHP studies (such as equal, slightly more important, markedly more important, much more important, and extremely more important) along with their global fuzzy counterparts. For example, when ChatGPT was asked, "Which is a more important indicator: physical consistency or sensor trace analysis? To what degree is one more important than the other?" a fuzzy triplet (μ, ν, π) value was derived from the response and incorporated into the matrix. The same process was carried out with the views of other members of the expert panel, resulting in a single fuzzy comparison matrix reflecting consensus. In this matrix, for each pair of *i* and *j* criteria, a (μ, ν, π) triplet is provided.

Table 2. Fuzzy pairwise comparison matrix for criteria importance.

	C1	C2	C3	C4	C5	C6
C1	0,5 0,4 0,4	0,7 0,3 0,2	0,6 0,3 0,2	0,7 0,3 0,2	0,8 0,2 0,1	0,8 0,2 0,1
C2	0,3 0,7 0,2	0,5 0,4 0,4	0,3 0,7 0,2	0,6 0,3 0,2	0,6 0,3 0,2	0,7 0,3 0,2
C3	0,4 0,6 0,3	0,7 0,3 0,2	0,5 0,4 0,4	0,6 0,3 0,2	0,7 0,3 0,2	0,7 0,3 0,2
C4	0,3 0,7 0,2	0,4 0,6 0,3	0,4 0,6 0,3	0,5 0,4 0,4	0,6 0,3 0,2	0,6 0,3 0,2
C5	0,2 0,8 0,1	0,4 0,6 0,3	0,3 0,7 0,2	0,4 0,6 0,3	0,5 0,4 0,4	0,5 0,4 0,4
C6	0,2 0,8 0,1	0,3 0,7 0,2	0,3 0,7 0,2	0,4 0,6 0,3	0,5 0,4 0,4	0,5 0,4 0,4

The weights obtained from the calculations are presented and discussed in the findings and discussion section. The criterion weights provide a quantitative indicator of which criteria play a more dominant role in distinguishing AI-generated from real images. Additionally, the consistency of the expert evaluations was examined, and it was found that the results obtained by SF-AHP, which incorporates uncertainties, align with the expectations in the literature.

Table 3 presents the evaluations between the sub-criteria of C1 (Physical consistency and real-world realism) — C1-1 Light and Shadow Coherence, C1-2 Perspective and Geometric Consistency, and C1-3 object

Interaction Logic. *Table 4* shows the sub-criteria of C2 (Micro-detail quality and visual anomalies), *Table 5* presents the sub-criteria of C3 (Sensor traces and meta-data examination), *Table 6* illustrates the sub-criteria of C4 (Statistical signatures and model fingerprints), *Table 7* shows the sub-criteria of C5 (Contextual and logical coherence), and *Table 8* presents the sub-criteria of C6 (Expert/tool-assisted diagnostic methods). In each cell, the triplet (μ, ν, π) represents the degree to which the sub-criterion in the row is more important than the sub-criterion in the column, expressed in global fuzzy terms; the main diagonal in all tables (0.5, 0.4, 0.4) represents the "equal importance" value.

Table 3. Spherical-fuzzy pairwise matrix for C1 sub-criteria.

	C1-1	**C1-2**	**C1-3**
C1-1	0,5 0,4 0,4	0,8 0,2 0,1	0,9 0,1 0
C1-2	0,2 0,8 0,1	0,5 0,4 0,4	0,6 0,4 0,3
C1-3	0,1 0,9 0	0,4 0,6 0,3	0,5 0,4 0,4

Table 4. Spherical-fuzzy pairwise matrix for C2 sub-criteria.

	C2-1	**C2-2**	**C2-3**
C2-1	0,5 0,4 0,4	0,8 0,2 0,1	0,9 0,1 0
C2-2	0,2 0,8 0,1	0,5 0,4 0,4	0,7 0,3 0,2
C2-3	0,1 0,9 0	0,3 0,7 0,2	0,5 0,4 0,4

Table 5. Spherical-fuzzy pairwise matrix for C3 sub-criteria.

	C3-1	**C3-2**	**C3-3**
C3-1	0,5 0,4 0,4	0,8 0,2 0,1	0,7 0,3 0,2
C3-2	0,2 0,8 0,1	0,5 0,4 0,4	0,6 0,4 0,3
C3-3	0,3 0,7 0,2	0,4 0,6 0,3	0,5 0,4 0,4

Table 6. Spherical-fuzzy pairwise matrix for C4 sub-criteria.

	C4-1	**C4-2**	**C4-3**
C4-1	0,5 0,4 0,4	0,7 0,3 0,2	0,8 0,2 0,1
C4-2	0,3 0,7 0,2	0,5 0,4 0,4	0,6 0,4 0,3
C4-3	0,2 0,8 0,1	0,4 0,6 0,3	0,5 0,4 0,4

Table 7. Spherical-fuzzy pairwise matrix for C5 sub-criteria.

	C5-1	**C5-2**	**C5-3**
C5-1	0,5 0,4 0,4	0,7 0,3 0,2	0,6 0,4 0,3
C5-2	0,3 0,7 0,2	0,5 0,4 0,4	0,6 0,4 0,3
C5-3	0,4 0,6 0,3	0,4 0,6 0,3	0,5 0,4 0,4

Table 8. Spherical-fuzzy pairwise matrix for C6 sub-criteria.

	C6-1	**C6-2**	**C6-3**
C6-1	0,5 0,4 0,4	0,4 0,6 0,3	0,7 0,3 0,2
C6-2	0,6 0,4 0,3	0,5 0,4 0,4	0,8 0,2 0,1
C6-3	0,3 0,7 0,2	0,2 0,8 0,1	0,5 0,4 0,4

The methodology developed applies not only to the current set of criteria but also to potential future updates to the criteria. As the nature of artificial images evolves over time, new clues and challenges will emerge. In

such cases, the established set of criteria can be revised, and the SF-AHP analysis can be repeated to detect how priorities shift. Thus, the proposed approach offers a dynamic and updatable decision support framework.

5 | Findings and Discussion

Table 9 reports the spherical-fuzzy synthesis values (μ, ν, π) and the normalized SF-AHP weights for the six main criteria; the CR of 0.081 confirms that expert judgements are acceptably coherent. *Tables 10–15* break these results down one level deeper: each table lists, respectively, the local weights of C1 through C6 sub-criteria, so the reader can see which physical-realism cue (*Table 10*), micro-detail anomaly (*Table 11*), sensor–metadata signal (*Table 12*), statistical fingerprint (*Table 13*), contextual check (*Table 14*) or expert/tool indicator (*Table 15*) exerts the most significant influence inside its own group. Finally, Table 16 aggregates those local weights with the parent weights from *Table 9* to yield the global ranking of every sub-criterion; this consolidated view highlights, for example, that light-and-shadow coherence (*Table 10*) and PRNU absence (*Table 12*) emerge as the two strongest individual detectors. In contrast, digital watermark verification (*Table 15*) sits at the lower end of the priority list. Together, *Tables 9–16* provide a complete, traceable weighting chain from top-level criteria to the most granular diagnostic cues.

Table 9. Main criteria weighting results based on the spherical fuzzy AHP method.

	μ	ν	π	Weights
C1	0,57702	0,416017	0,285419	0,174511
C2	0,785172	0,213983	0,175336	0,24935
C3	0,621437	0,370629	0,268539	0,190084
C4	0,504415	0,492876	0,295988	0,150078
C5	0,411423	0,551785	0,321354	0,117989
C6	0,411423	0,551785	0,321354	0,117989
				1

Table 10. Evaluation of sub-criteria under C1.

	μ	ν	π	Weights
C1-1	0,695421	0,28845	0,238397	0,457082
C1-2	0,392841	0,5769	0,31075	0,237963
C1-3	0,49118	0,482028	0,320517	0,304955
				1

Table 11. Evaluation of sub-criteria under C2.

	μ	ν	π	Weights
C2-1	0,695421	0,28845	0,238397	0,457082
C2-2	0,49118	0,482028	0,320517	0,304955
C2-3	0,392841	0,5769	0,31075	0,237963
				1

Table 12. Evaluation of sub-criteria under C3.

	μ	ν	π	Weights
C3-1	0,695421	0,28845	0,238397	0,457082
C3-2	0,49118	0,482028	0,320517	0,304955
C3-3	0,392841	0,5769	0,31075	0,237963
				1

Table 13. Evaluation of sub-criteria under C4.

	μ	ν	π	Weights
C4-1	0,836494	0,15874	0,151927	0,491561
C4-2	0,596516	0,416017	0,234679	0,337886
C4-3	0,326675	0,660385	0,267467	0,170553
				1

Table 14. Evaluation of sub-criteria under C5.

	μ	ν	π	Weights
C5-1	0,532832	0,457886	0,273936	0,313687
C5-2	0,350003	0,631636	0,281956	0,195062
C5-3	0,792737	0,2	0,178235	0,491252
				1

Table 15. Evaluation of sub-criteria under C6.

	μ	ν	π	Weights
C6-1	0,49118	0,482028	0,320517	0,304955
C6-2	0,695421	0,28845	0,238397	0,457082
C6-3	0,392841	0,5769	0,31075	0,237963
				1

Table 16. Comprehensive weighting results of sub-criteria using the SF-AHP method.

Sub-Criterion	Weight	Rank
Light and shadow coherence (C1.1)	0,07976575	3
Perspective and geometric consistency (C1.2)	0,04152712	13
Object interaction logic (C1.3)	0,05321783	10
Anatomical and fine detail consistency (C2.1)	0,113973273	1
Texture and pattern anomalies (C2.2)	0,076040278	4
Focus and depth consistency (C2.3)	0,059336011	6
Sensor noise and PRNU analysis (C3.1)	0,08688391	2
EXIF and file metadata consistency (C3.2)	0,0579669	7
Compression artifacts and pixel distribution (C3.3)	0,04523293	12
Frequency spectrum analysis (C4.1)	0,07377265	5
Color statistics and channel correlation (C4.2)	0,05070946	11
Model fingerprint identification (C4.3)	0,02559634	17
Reality and logic check (C5.1)	0,03701149	14
Intra-scene context consistency (C5.2)	0,02301511	18
External source and reverse image verification (C5.3)	0,05796218	8
Automated AI detector output (C6.1)	0,03598121	15
Human expert / forensic review (C6.2)	0,05393058	9
Digital watermark / content credential check (C6.3)	0,02807698	16

The SF-AHP analysis indicates that micro-level detail quality and visual anomaly detection (C2) wield the greatest influence, accounting for approximately 24.9 % of the total importance. Within this group, the sub-criterion concerning anatomical and fine-detail consistency (C2-1, 11.4 %) emerges as especially decisive, because residual distortions around eyes, hands, or hair strands are still produced by state-of-the-art generators and can be exposed through high-resolution inspection. Texture and pattern anomalies (C2-2, 7.6 %) further support this finding, as repeating or misaligned motifs remain characteristic artefacts of convolution-based synthesis pipelines.

The sensor trace and metadata cluster (C3) follows with about 19.0 %. Its foremost sub-criterion, sensor-noise and PRNU analysis (C3-1, 8.7 %), is weighted highly because synthetic imagery has yet to reproduce the stochastic fingerprints left by physical camera sensors; in practice, the absence of a coherent PRNU field or the presence of conflicting EXIF tags functions as a reliable provenance test. Together, C2 and C3 constitute roughly 44 % of the overall weight, implying that pixel-level irregularities and file-level provenance should be examined before contextual diagnostics are attempted.

Physical consistency and real-world realism (C1) stand in third position with 17.5 %. Although light-and-shadow coherence (C1-1, 8.0 %) and object-interaction logic (C1-3, 5.3 %) still reveal tampering, improvements in physically based rendering are gradually narrowing classical illumination gaps, which explains the observed decline in relative weight. Statistical signatures and model fingerprints (C4) contribute about 15.0 %; frequency-spectrum deviations (C4-1, 7.4 %) and colour-channel correlations continue to offer complementary evidence and are expected to gain importance if physical and sensor artefacts are further concealed.

Contextual and logical consistency (C5) together with expert- or tool-assisted assessment (C6) each receive roughly 11.8 %. Reverse-image search performance (C5-3, 5.8 %) underscores the current value of open-source intelligence, whereas the low score of content-credential checks (C6-3, 2.8 %) points to procedural under-utilisation rather than diminished conceptual relevance. Consequently, the hierarchy that emerges places greatest diagnostic power in "what the pixels reveal" (C2, C3), assigns a corroborative role to "how the scene behaves" (C1, C4), and reserves "where the image fits" (C5, C6) for final confirmation. As diffusion and GAN models progressively close physical and sensor gaps, statistical fingerprints and advanced contextual tests are expected to assume greater strategic weight. Periodic recalibration of the criterion set is therefore essential if verification resources are to remain aligned with the highest-yielding stages of the review workflow.

6 | Conclusion

The distinction between AI-generated images and real photographs is emerging as a crucial issue that needs to be addressed in the digitized world. This paper systematically examines the main criteria required for making this distinction, and the relative importance of these criteria was analyzed using the SF-AHP method. Six main criteria (physical consistency, detail anomalies, sensor traces, statistical features, contextual consistency, and expert/tool support) and their sub-criteria were identified through a systematic approach based on current literature findings and expert opinions, using the advanced LLM model ChatGPT o-3. SF-AHP analysis revealed that areas like physical consistency and sensor trace analysis play the most critical roles. At the same time, anatomical/detail errors and statistical/spectral cues are also significant supporting indicators. Contextual checks and expert support, although ranked lower, are considered additional elements for confirming suspicious cases.

These results confirm the necessity of a multifaceted approach to detecting artificial images and highlight the need for continuous updates to the criteria based on evolving technologies. Rather than relying on a single method, the combined use of different criteria will reduce error margins and increase reliability. For example, detecting both physical inconsistencies (such as shadow mismatch) and failing the sensor fingerprint analysis will almost certainly confirm an image as artificial. On the other hand, if an image passes all physical and technical tests but raises contextual suspicions, it is best to proceed cautiously and resort to external validation (such as source checks) or semi-automated systems for verification. The study also presents a methodological innovation: the integration of an AI model (ChatGPT-o-3) as an expert opinion in the SF-AHP process demonstrates how human and artificial expertise can be hybridized in decision-making mechanisms. This approach could be reliably used in future multi-criteria problems under expert supervision. The AI's fast access to up-to-date information and its ability to blend a broad perspective can offer valuable contributions to decision-making processes when directed appropriately. In particular, in solving AI-generated image detection, leveraging AI capabilities seems a logical strategy to stay ahead of AI developments.

Looking forward, it can be predicted that the struggle between artificial and real images will continue as an ongoing balancing/unbalancing race. Advancements in Generative AI will close many existing gaps; soon, artificial images may have perfect physical consistency, or they may come with fake EXIF data. In contrast, new techniques will be developed on the detection side, such as discovering entirely new signatures specific to artificial images. This study presents the criteria and priorities considered critical with today's knowledge base, but these priorities are not static. There will be a growing need for adaptive validation systems that are continuously updated.

Furthermore, policymakers and platform developers must also contribute to the solution. Transparent and verifiable infrastructure for content reliability should be encouraged. Indeed, the C2PA standard and similar content identification initiatives led by actors such as Adobe and Microsoft hold promise. If we transition to an ecosystem where every AI-generated image carries a cryptographic stamp indicating its source, the need for technical analysis may decrease somewhat. However, until then, the best tools available will remain the criteria discussed in this paper.

In conclusion, predicting the future has become a new dimension of digital literacy and forensic informatics. The multi-criteria framework and findings presented in this paper provide a comprehensive perspective on detecting AI-generated images. This set of criteria could serve as a guide for practitioners, journalists, regulators, and researchers, and it could also form a foundation for academic work. In the future, testing the weight of these criteria on real-world cases, expanding them, if necessary, with new criteria, and training detection algorithms to include this multifaceted approach will be crucial. The fight to distinguish between real and fake is a struggle with technological, ethical, and societal dimensions, and its success will depend on the collaboration of stakeholders.

Conflict of Interest

The authors declare no conflict of interest.

Data Availability

All data are included in the text.

Funding

This research received no specific grant from funding agencies in the public, commercial, or not-for-profit sectors.

References

[1] Statista. (2024). *Amount of data created, consumed, and stored 2010-2023, with forecasts to 2028*. https://www.statista.com/statistics/871513/worldwide-data-created/#statisticContainer

[2] Cao, H., Tan, C., Gao, Z., Xu, Y., Chen, G., Heng, P. A., & Li, S. Z. (2024). A survey on generative diffusion models. *IEEE transactions on knowledge and data engineering*, *36*(7), 2814–2830. https://doi.org/10.1109/TKDE.2024.3361474

[3] Borji, A. (2022). *Generated faces in the wild: Quantitative comparison of stable diffusion, midjourney and dall-e 2*. ArXiv Preprint ArXiv:2210.00586. https://doi.org/10.48550/arXiv.2210.00586

[4] Perrigo, B., & Johnson, V. (2023). *How to spot an AI-Generated image like the 'Balenciaga Pope*. https://time.com/6266606/how-to-spot-deepfake-pope/

[5] Hausken, L. (2024). Photorealism versus photography. AI-generated depiction in the age of visual disinformation. *Journal of aesthetics & culture*, *16*(1), 2340787. https://doi.org/10.1080/20004214.2024.2340787

[6] Bontcheva, K., Papadopoulous, S., Tsalakanidou, F., Gallotti, R., Dutkiewicz, L., Krack, N., … & Srba, I. (2024). *Generative AI and disinformation: Recent advances, challenges, and opportunities*. https://lirias.kuleuven.be/retrieve/758830

[7] Yoo, Y., Na, D., Nathanson, S., Cao, Y., & Watkins, L. (2024). Disinformation at scale: detecting ai-human composite images via convolution ensembles. *MILCOM 2024-2024 IEEE military communications conference (milcom)* (pp. 621–626). IEEE. https://doi.org/10.1109/MILCOM61039.2024.10773642

[8] Jagadish, T., & Jasmine, S. G. (2024). Detection of ai-generated image content in news and journalism. *2024 15th international conference on computing communication and networking technologies (ICCCNT)* (pp. 1–6). IEEE. https://doi.org/10.1109/ICCCNT61001.2024.10724589

[9] Moeßner, P., & Adel, H. (2024). *Human VS. AI: A novel benchmark and a comparative study on the detection of generated images and the impact of prompts*. ArXiv Preprint ArXiv:2412.09715. https://doi.org/10.48550/arXiv.2412.09715

[10] Cetinic, E., & She, J. (2022). Understanding and creating art with AI: Review and outlook. *ACM transactions on multimedia computing, communications, and applications*, *18*(2), 1–22. https://doi.org/10.1145/3475799

[11] Bellaiche, L., Shahi, R., Turpin, M. H., Ragnhildstveit, A., Sprockett, S., Barr, N., … & Seli, P. (2023). Humans versus AI: Whether and why we prefer human-created compared to AI-created artwork. *Cognitive research: principles and implications*, *8*(1), 1-22. https://doi.org/10.1186/s41235-023-00499-6

[12] Guo, H., Hu, S., Wang, X., Chang, M. C., & Lyu, S. (2022). Eyes tell all: Irregular pupil shapes reveal gan-generated faces. *ICASSP 2022-2022 IEEE international conference on acoustics, speech and signal processing (ICASSP)* (pp. 2904–2908). IEEE. https://doi.org/10.1109/ICASSP43922.2022.9746597

[13] Hua, M., Li, S., & Wang, J. (2025). A dual-stream model based on PRNU and quaternion RGB for detecting fake faces. *PloS one*, *20*(1), e0314041. https://doi.org/10.1371/journal.pone.0314041

[14] Xiao, S., Guo, Y., Peng, H., Liu, Z., Yang, L., & Wang, Y. (2025). *Generalizable AI-generated image detection based on fractal self-similarity in the spectrum*. ArXiv Preprint ArXiv:2503.08484. https://doi.org/10.48550/arXiv.2503.08484

[15] Kusuma, S. W., Natalia, F., Ko, C. S., & Sudirman, S. (2024). Detection of AI-generated anime images using deep learning. *ICIC express letters, part B: Applications*, *15*(3), 295–301. https://doi.org/10.24507/icicelb.15.03.295

[16] Tolosana, R., Vera-Rodriguez, R., Fierrez, J., Morales, A., & Ortega-Garcia, J. (2020). Deepfakes and beyond: A survey of face manipulation and fake detection. *Information fusion*, *64*, 131–148. https://doi.org/10.1016/j.inffus.2020.06.014

[17] Ghiurău, D., & Popescu, D. E. (2025). Distinguishing reality from AI: Approaches for detecting synthetic content. *Computers*, *14*(1), 1–33. https://doi.org/10.3390/computers14010001

[18] Afchar, D., Nozick, V., Yamagishi, J., & Echizen, I. (2018). Mesonet: A compact facial video forgery detection network. *2018 IEEE international workshop on information forensics and security (WIFS)* (pp. 1–7). IEEE. https://doi.org/10.1109/WIFS.2018.8630761

[19] Frank, J., Eisenhofer, T., Schönherr, L., Fischer, A., Kolossa, D., & Holz, T. (2020). Leveraging frequency analysis for deep fake image recognition. *International conference on machine learning* (pp. 3247–3258). PMLR. http://proceedings.mlr.press/v119/frank20a

[20] Chen, H., Gu, J., Chen, A., Tian, W., Tu, Z., Liu, L., & Su, H. (2023). *Single-stage diffusion NeRF: A unified approach to 3D generation and reconstruction.* https://doi.org/10.48550/arXiv.2304.06714

[21] Zadeh, L. A. (1965). Fuzzy sets. *Information and control*, *8*(3), 338–353. https://doi.org/10.1016/S0019-9958(65)90241-X

[22] Kutlu Gündoğdu, F., & Kahraman, C. (2019). Spherical fuzzy sets and spherical fuzzy TOPSIS method. *Journal of intelligent & fuzzy systems*, *36*(1), 337–352. https://doi.org/10.3233/JIFS-181401

[23] Kieu, P. T., Nguyen, V. T., Nguyen, V. T., & Ho, T. P. (2021). A spherical fuzzy analytic hierarchy process (SF-AHP) and combined compromise solution (CoCoSo) algorithm in distribution center location selection: A case study in agricultural supply chain. *Axioms*, *10*(2), 1–13. https://doi.org/10.3390/axioms10020053