Signal Processing: Image Communication 148 (2026) 117636

Contents lists available at ScienceDirect

Signal Processing: Image Communication

journal homepage: www.elsevier.com/locate/image

Towards compression-aware iris presentation attack detection ☆

Rocco Albano [a], Filippo Battaglia [b], Alessandro Gnutti [c], Emanuele Maiorana [a,*], Fabrizio Guerrini [c], Giuseppe Campobello [b], Pierangelo Migliorati [c], Patrizio Campisi [a]

[a] Department of Industrial, Electronic and Mechanical Engineering, Roma Tre University, Via V. Volterra 62, Rome, 00146, Italy
[b] Department of Engineering, University of Messina, Via Salita Sperone - c.da Papardo, Messina, 98166, Italy
[c] Department of Information Engineering, Università degli Studi di Brescia, Via Branze 38, Brescia, 25123, Italy

ARTICLE INFO

Keywords:
Biometrics
Iris recognition
Presentation attack detection
Image compression
Transformers

ABSTRACT

With the growing integration of biometric recognition systems into high-security and large-scale deployment scenarios, it is becoming increasingly important to ensure their robustness under realistic operational constraints. This implies designing solutions able to withstand potential adversarial threats that could affect their integrity and accountability, and also taking into account requirements of real-world operating systems such as limited availability of bandwidth and memory for data transmission and storage. Hence, the proposed study deals with presentation attack detection (PAD) for iris recognition, evaluating the effectiveness of Transformer-based frameworks at detecting spoofing attacks relying on fabricated or artificial biometric evidences. More specifically, we focus on the effects of image compression on the quality of iris images and on the resulting PAD performance, considering both traditional techniques such as JPEG as well as next-generation learning-based image codecs such as JPEG AI. We then examine the feasibility of mitigating compression-induced performance degradation by fine-tuning the adopted models on compressed images, achieving improvements in terms of half total error rate between 5% and 10% for images compressed at the worst JPEG and JPEG AI qualities. We also evaluate the generalizability of the developed solutions by testing them on learning-based codecs not considered during training, to check whether similar PAD-relevant artifacts are introduced by different compressions. Furthermore, we investigate the redundancy within the embeddings generated by the employed detectors, and demonstrated it is possible to significantly compress them while preserving the achievable PAD performance. Overall, the study provides a systematic analysis of iris PAD under compression constraints, offering insights into model adaptation, cross-codec robustness, and representation efficiency in scenarios where visual data coding plays a central role.

1. Introduction

Biometric recognition has increasingly established itself as one of the most reliable and versatile approaches to automated identity verification. By exploiting intrinsic physical or behavioral characteristics, such as fingerprints or speech, biometric systems provide security and usability that surpass those of traditional authentication paradigms based on knowledge (passwords, PIN codes) or tokens (ID cards, physical keys). The widespread adoption of biometric technologies across diverse fields, including border control, access regulation in critical facilities, digital financial services, and forensic applications, further underscores their strategic role in modern security and identification infrastructure [1].

Despite the clear advantages, the deployment of biometric systems does not eliminate potential vulnerabilities. In fact, each stage of a biometric pipeline can become a target for malicious interference. Components such as the sensing device, the feature extraction and matching modules, the storage infrastructure, and even the communication links between these elements, remain exposed to a broad spectrum of attacks designed to undermine system accuracy, integrity, or the privacy of enrolled users [1]. A notable example is the attack on biometric databases, often referred to as a template attack, in which adversaries attempt to intercept, replace, or reconstruct stored templates. These actions threaten the confidentiality of biometric data and may lead to identity fraud and privacy concerns. To counter them, the adoption of biometric template protection (BTP) mecha-

☆ This article is part of a Special issue entitled: 'IMAGE_Visual Data Coding' published in Signal Processing: Image Communication.

* Corresponding author.
E-mail address: emanuele.maiorana@uniroma3.it (E. Maiorana).

https://doi.org/10.1016/j.image.2026.117636
Received 22 December 2025; Received in revised form 1 April 2026; Accepted 10 June 2026
Available online 14 June 2026

nisms that ensure irreversibility, revocability, and unlinkability is essential [2].

A distinct and particularly insidious class of threats affects the acquisition stage: sensor-level attacks, also known as presentation or spoofing attacks. In this scenario, the adversary presents an artificial or manipulated biometric sample to the sensor in order to impersonate an enrolled individual. Unlike template attacks, spoofing does not require explicit knowledge of a recognition system's internal functioning. Biometric traits exposed in everyday contexts, such as face, fingerprint,[1] and even iris, can be covertly captured using high-resolution imaging devices [3]. The acquired data can then be reproduced as realistic forgeries that can deceive the recognition process and grant unauthorized access [3]. Because of the relative simplicity of executing such attacks and the potentially severe consequences of successful spoofing, robust mechanisms for presentation attack detection (PAD), also known as liveness detection, have become a crucial component of any operational biometric system.

PAD mechanisms aim to differentiate *bona fide* samples from fake or manipulated inputs, and their design is now considered as essential as improving raw recognition accuracy. Within this framework, the present work focuses on PAD in the context of iris biometrics, one of the most accurate and stable modalities for high-security and large-scale deployments. Although historically associated with applications such as automated border control or secure access to restricted environments, recent technological advances have encouraged the integration of iris recognition into everyday consumer devices. Actually, such an authentication approach is increasingly considered as a complementary or alternative solution in smartphones [4], incorporated into virtual and mixed reality headsets to ensure secure binding between the user and their digital avatar [5–7], and even evaluated as a potential mechanism for identity management in recently-introduced cryptocurrency ecosystems [8].

Building on this evolving scenario, the present study extends the authors' previous investigations in [9,10], where Transformer-based architectures have been used to discriminate between genuine and spoofed iris samples, by performing a detailed analysis on the effectiveness of iris PAD under compression constraints. This aspect is particularly relevant in practical biometric systems, where images are routinely compressed for storage and transmission, potentially altering the fine-grained texture information on which PAD algorithms rely, thereby affecting their reliability and raising important questions about the robustness of PAD solutions in real-world deployment scenarios. More in detail, the novel contributions of the current work include:

- a comprehensive evaluation on the effects of compression on iris image quality, considering both traditional techniques such as JPEG [11] and emerging learning-based codecs such as JPEG AI [12], and specifically analyzing to which extent iris PAD performance is affected by image compression;
- a strategy for mitigating the effects of compression-induced distortions by fine-tuning Transformer-based PAD models on datasets compressed at different rates, with the goal of maintaining high discrimination capability under varying compression conditions;
- an investigation of cross-compression generalizability, assessing whether PAD models trained on samples compressed with learning-based algorithms can successfully operate on images processed by alternative compression schemes not seen during training;
- an exploration of representational redundancy within the feature embeddings generated by the employed PAD models, with one of the main novelties of this work consisting in a targeted analysis aimed at understanding whether compact templates can be exploited for iris PAD without degrading the detectors' performance.

[1] https://www.bbc.com/news/technology-30623611

The rest of the paper is structured as follows: Section 2 presents a concise survey on iris PAD and reviews representative techniques proposed in the literature, as well as the Transformer-based architectures first proposed for iris PAD in our previous works [9,10]. Section 3 outlines the compression methods analyzed in this study, and summarizes prior work addressing the influence of compression on biometric recognition. Section 4 introduces a novel approach for iris PAD, where the used iris representations are derived from Transformer-based embeddings by retaining only the most relevant information, with the aim of designing a simple yet effective detector relying on compact templates. Section 5 then details the experimental setup and the obtained results, with discussions regarding the effects of compression on iris images and on iris PAD, the possibility of mitigating compression-induced performance degradation by means of fine-tuning, the effectiveness of the proposed simplified detector relying on compact templates, and the generalizability of the developed solutions on learning-based codecs not available during training. Finally, Section 6 summarizes the main contributions of this work and discusses its limitations, proposing possible directions for future research.

2. Iris PAD

As the adoption of iris recognition continues to grow, driven by its exceptional accuracy, long-term stability, and robustness against environmental fluctuations, systems relying on such a trait are also becoming increasingly attractive for malicious subjects trying to exploit it, or adversarial attempts aimed at breaking their security, particularly at the sensor interface where the biometric trait is first captured. Among these threats, presentation attacks (PAs) occupy a central role: by introducing counterfeit or altered biometric samples, adversaries can directly manipulate the input of the recognition pipeline and bypass the underlying security mechanisms. Numerous strategies for fabricating such spoofed iris samples have been documented in the literature [13], including printed iris images produced with high-resolution printers [14], textured cosmetic contact lenses engineered to simulate realistic iris patterns while worn by the attacker [15], digital display-based attacks using high-quality screens to project synthetic irises [16], prosthetic ocular models designed to closely resemble a living eye [17], and even samples obtained from cadavers during the early post-mortem interval [18].

Among these possibilities, attacks based on textured contact lenses are generally regarded as the most practical and alarming ones. Their attractiveness stems from several factors: they require minimal technical expertise, involve limited financial investment, and allow attackers to approach the sensor in a seemingly natural manner without arousing suspicion. For these reasons, textured lenses pose a realistic, high-impact threat in operational scenarios and must be prioritized in the design of robust iris PAD mechanisms [15].

Historically, iris PAD research focused on developing handcrafted descriptors tailored to capture subtle differences between bona fide and spoofed images [19]. These methods often relied on carefully designed texture operators, frequency-domain analysis, or features that model the physical properties of eye tissues. While such approaches provided meaningful insights and paved the way for standardized evaluations, the emergence of large-scale datasets and deep learning technologies has shifted the field toward data-driven solutions [20]. Contemporary methods increasingly exploit neural architectures that can discover intricate and discriminative patterns directly from raw data, yielding substantial performance improvements over traditional techniques [21, 22].

Within this paradigm shift, attention mechanisms have emerged as a powerful tool for enhancing the representational capacity of neural networks. By enabling models to dynamically weight salient regions or features, attention facilitates a more structured understanding of visual content and improves the ability to capture fine-grained artifacts

associated with presentation attacks [23]. In the context of iris PAD, several studies have integrated attention modules into convolutional neural network (CNN) architectures. For instance, [24] employs an attention-guided CNN to provide interpretable visual justifications for classification decisions. A pixel-level attention mechanism is proposed in [25] to highlight fine-grained spoofing cues, whereas [26] incorporates attention in the final CNN stages to refine the decision-making process. These recent contributions collectively demonstrate the potential of attention to improve the robustness and transparency of PAD systems.

Nevertheless, attention in these works is treated as an auxiliary component rather than the core computational principle. The possibility of using attention as the sole architectural foundation, following the pure Transformer paradigm introduced in [27], remained largely unexplored for PAD until the authors' preliminary studies in [9,10]. In these earlier works, we investigated the suitability of vision Transformers (ViTs) [28] and shifted-window (Swin) Transformers [29] for iris PAD, motivated by their remarkable success on a wide range of computer vision tasks. Such architectures rely exclusively on attention mechanisms to model spatial relationships, offering a fundamentally different representational framework compared to CNNs.

Several additional issues and challenges have also been recently considered to advance the field of iris PAD [30]. For instance, the capabilities of innovative sensors and their configurations have been analyzed in [31], where the usefulness of exploiting 3D information collected by stereo cameras has been evaluated. A noteworthy topic, especially for data-driven models, is cross-domain performance analysis, in which classifiers trained on data collected with a sensor are applied to samples collected by other devices [32,33]. Furthermore, the influence of many additional factors on iris PAD remains under investigation, with studies on image segmentation [34], demographic aspects of the subjects [35], and their gender [36].

Following this line of research, this work examines the impact of image compression on data-driven iris PAD and investigates strategies to account for compression requirements when developing effective iris PAD systems. As discussed in Section 3, compression is increasingly relevant to the design of large-scale biometric systems and communication pipelines, motivating research into its effects on both the treated biometric data and the associated performance of recognition systems. Here, we adopt the Transformer-based architectures first proposed in [9,10] and summarized in Section 2.1 for iris PAD. Furthermore, we deepen our investigation with respect to our previous contributions by verifying the feasibility of improving the achievable iris PAD performance through fine-tuning, by examining the generalizability of the adopted models to cross-compression scenarios, testing them on data coded with schemes not available during training, and by characterizing the structure and redundancy of the embeddings extracted from the adopted Transformer-based architectures to design novel, effective and compact iris PA detectors.

2.1. Adopted Transformer-based iris PAD approaches

To investigate iris PAD under compression constraints, we adopt two state-of-the-art attention-based architectures, i.e., ViT [28] and Swin Transformers [29]. Both architectures rely exclusively on self-attention mechanisms to model spatial dependencies, providing a fundamentally different representational paradigm compared to convolutional neural networks (CNNs) [27]. They have been applied to iris PAD in our previous studies [9,10], with the aim of exploiting the capability of attention mechanisms at modeling long-range spatial relationships, which are crucial for detecting subtle spoofing artifacts, and of supporting the interpretability of PAD systems. The most relevant characteristics of their architectures are summarized as follows.

2.1.1. Vision Transformer (ViT)

Inspired by the use of attention-based models in natural language processing (NLP), a ViT processes an input image by partitioning it into a grid of non-overlapping patches of fixed spatial resolution (16×16 pixels in our configuration). Each patch is flattened and linearly projected into a latent embedding space, generating a sequence of patch tokens. To preserve spatial ordering information lost during patch flattening, learnable positional embeddings are added to the token sequence. The resulting sequence is fed into a stack of Transformer encoder blocks, each consisting of:

- a multi-head self-attention (MHSA) module, which computes pairwise interactions between all tokens to achieve a global receptive field at every layer;
- a position-wise feed-forward network (FFN);
- residual connections and layer normalization.

In our implementation, 12 attention heads are employed, allowing the model to learn complementary spatial interaction patterns across different representation subspaces. The global self-attention mechanism allows ViT to capture well non-local texture inconsistencies and fine-grained structural irregularities [37] that may characterize presentation attacks, particularly textured contact lenses and print-based artifacts. For the PAD task, the final classification token is processed by a lightweight task-specific head composed of two fully connected layers with dropout regularization.

2.1.2. Swin Transformer

While ViT applies global self-attention over all image patches, the Swin Transformer introduces a more computationally-efficient architecture by restricting self-attention to local windows. Specifically, the input image is partitioned into non-overlapping windows of fixed size, and multi-head self-attention is computed independently within each window. To enable interaction across neighboring regions, the Swin architecture employs a shifted-window mechanism: in alternating Transformer blocks, the window partitioning is shifted by half the window size along both spatial dimensions, allowing tokens that were previously located in different windows to be jointly processed in subsequent layers, thus facilitating cross-window information exchange without significantly increasing the required computational cost.

Another distinctive characteristic of Swin is its hierarchical architecture: the network is organized into multiple stages, and a patch-merging operation is performed between stages. This operation reduces spatial resolution while increasing the channel dimensionality, producing multi-scale feature representations able to simultaneously capture both fine-grained local texture details and more global structural patterns. In the context of iris PAD, this combination of localized attention and hierarchical feature aggregation is particularly relevant: local window attention may focus on subtle texture irregularities introduced by presentation attacks, while the shifted-window mechanism and multi-stage structure support the modeling of broader spatial inconsistencies.

2.1.3. Relevance in the performed analysis

A notable benefit of using attention-based models in data-driven approaches lies in their natural support for explainability, a property that has become increasingly important as deep learning systems are entrusted with safety-critical decisions. Although data-driven approaches often outperform traditional handcrafted methods in recognition tasks, their limited interpretability raises concerns regarding reliability, trust, and fairness [38].

Self-attention mechanisms inherently produce attention weights that can be leveraged for model interpretability. In a Transformer layer, attention is computed through scaled dot-product operations between query and key projections of the input tokens, followed by a softmax normalization. The resulting attention matrix encodes pairwise interaction strengths between tokens, and indicates how much each

token contributes to the updated representation of the others. These normalized weights therefore provide a principled basis for identifying which image regions influence the final decision.

In our ViT-based iris PAD framework, the input image is divided into patches, which are embedded into token representations. Each attention layer produces multiple attention matrices (one per head), indicating how strongly each patch attends to every other one. To obtain spatial attention maps, we focus on the attention weights associated with the class token, which aggregates global information for classification. The attention vectors corresponding to this token are extracted and averaged across heads (and optionally across selected layers) to produce a single importance score per patch. These patch-level scores are then reshaped to the original patch grid and upsampled to the input resolution, yielding a heatmap that highlights the regions that contribute most to the bona fide/spoof decision. This procedure enables direct visualization of discriminative regions and facilitates analysis of how modifications to the original inputs, such as those introduced by compression, may shift or attenuate attention toward specific regions.

Swin Transformers also generate interpretable attention patterns, although their structure differs due to the local-window constraint. In each layer, attention is computed independently within fixed-size windows, resulting in block-diagonal attention matrices rather than fully global ones. To obtain image-level attention maps, attention weights from all windows are first aggregated within each layer, then mapped back to their corresponding spatial locations. Because successive layers adopt shifted window partitions, tokens that belong to different windows in one layer interact in the next, allowing attention information to propagate across the image. When visualized cumulatively across layers, the resulting maps may reveal fine-grained local cues, such as micro-textural irregularities introduced by presentation attacks, while also capturing broader spatial structures that emerge progressively through the hierarchical and shifting mechanisms.

Illustrative examples of attention maps generated from both architectures are presented in Section 5, where they are used to analyze the response of our detectors under different compression schemes and attack scenarios. Overall, these visualization procedures provide insights into the internal decision process of the considered Transformer-based detectors, allowing to assess whether distinct attention mechanisms exhibit different sensitivities in the considered iris PAD scenarios, as well as to analyze whether compression alters the spatial distribution of discriminative cues, thereby supporting a compression-aware interpretation of the model behavior.

3. Iris compression

Image compression serves the fundamental purpose of representing visual data in a compact form while retaining as much perceptual quality as possible. It plays a central role in virtually every modern digital imaging pipeline, from multimedia storage to real-time video streaming, mobile acquisition, and large-scale data archiving. As image volumes continue to grow and applications increasingly operate under stringent bandwidth and storage constraints, effective compression becomes essential to ensure both system scalability and user accessibility.

Practical implementations of biometric recognition systems are not an exception to this rule, with applications and effects of image compression on biometric traits being extensively studied [39], particularly for widely used modalities such as fingerprint [40], face [41], and iris [42]. These efforts typically aim to quantify how compression artifacts such as blur, blocking, quantization noise, and color distortions affect feature extraction and matching accuracy, with the aim of identifying which compression levels preserve the discriminative information necessary for recognition, enabling efficient transmission or storage of biometric samples without sacrificing system performance [43].

In this work, to analyze the interaction between compression and iris PAD performance, we consider both traditional transform-based coding and learning-based image coding. This dual choice allows us to investigate how different distortion models affect Transformer-based detectors.

JPEG [11] is the most widely adopted image compression standard and is therefore commonly used as a natural baseline for evaluating compression robustness. It operates by dividing the image into 8×8 blocks, applying a discrete cosine transform (DCT) to each block, then quantizing and entropy coding. The quantization stage is responsible for most of the information loss and typically introduces blocking artifacts, ringing effects, and attenuation of high-frequency components. Compression levels are controlled by the quality factor (QF), which yields progressively stronger distortions as QF decreases. Regarding iris biometrics, the effects of traditional compression algorithms on the achievable recognition performance have been addressed in several research papers. In particular, the authors in [44,45] evaluated the effects of different compression rates on the obtainable Iriscode templates, and assessed performance using fractional Hamming distances between the templates of original and decoded images. Similar studies considering traditional codecs have been published, for instance, in [46–48].

Unlike transform-based codecs, learning-based approaches rely on neural autoencoders trained end-to-end for rate–distortion optimization. Such an encoder maps an input image into a compact latent representation, which is quantized and entropy-coded, while the decoder reconstructs the image from the compressed latent code [49]. Because these models learn non-linear analysis and synthesis transforms, the resulting distortions significantly differ from those produced by block-based DCT coding. Rather than the structured blocking artifacts of conventional compression, learning-based compression may introduce smoother, spatially adaptive distortions. However, optimization is typically driven by pixel-level distortion metrics, which are not necessarily aligned with the preservation of biometric discriminative features. In the context of iris recognition, learning-based compression schemes have been considered in [50], where the authors analyzed the dependence of the achievable recognition performance on the applied rate–distortion trade-off. In this paper, we primarily focus on JPEG AI [12], a learning-based codec, to investigate whether neural compression better preserves spoofing-related cues. This algorithm is the international standard for learning-based compression and relies on an end-to-end-optimized neural architecture, typically comprising a convolutional analysis transform, latent-space quantization with learned entropy modeling, and a corresponding synthesis transform for reconstruction. The rate–distortion trade-off is controlled by a target compression rate (TR) parameter, which is used in the optimization process during training.

Since different learning-based codecs can be defined exploiting different architectures, besides testing the performance of the proposed Transformer-based architectures on JPEG-AI-compressed images, and investigating whether their effectiveness can be improved through fine-tuning, we also investigated the generalizability of the attainable performance by checking whether the artifacts introduced by JPEG AI are similar, and thus manageable by the detector, to those introduced by other learning-based codecs.

4. Reduced-Feature Classifier for iris PAD

The Transformer-based detectors described in Section 2.1 produce very large embeddings: the ViT architecture generates 768-dimensional vectors, while Swin Transformers output feature maps of size 49×768. While these representations carry rich discriminative information, their high dimensionality can pose challenges for computational cost, storage requirements, and interpretability, particularly in resource-constrained or distributed deployment scenarios.

Motivated by these observations, we propose a Reduced-Feature Classifier (RFC) for iris PAD that operates on a compact subset of features extracted from Transformer embeddings. The RFC framework comprises two stages: a supervised feature selection procedure that identifies a small set of highly discriminative features, and a lightweight classification strategy that combines threshold-based rules with normalized distance measures. The following subsections describe each stage in detail.

4.1. Feature selection via inter-class separability

The proposed feature selection adopts a supervised, filter-based approach that operates exclusively on the original (uncompressed) training images of each considered dataset. For each embedding dimension f, feature values $x_i^{(f)}$ are first partitioned by ground-truth class labels, and class-specific statistics are computed, namely the median $\tilde{\mu}$, standard deviation σ, minimum, and maximum. The use of the median, rather than the mean, as the central tendency measure is motivated by its robustness to outliers, which makes the selection process more stable. Features are then evaluated based on an inter-class separability criterion. Specifically, considering a general scenario in which iris PA detection is performed as a K-ary classification task, a feature f is considered discriminative for class C_k, $k \in [0, K-1]$, if it satisfies at least one of the following two conditions:

- **Lower Separation Condition (LSC):** the median of feature f computed over class C_k is strictly smaller than the minimum value of f observed across all other classes, i.e.,

$$\tilde{\mu}_k^{(f)} < \min_{j \neq k} \left(\min_{i \in C_j} x_i^{(f)} \right); \tag{1}$$

- **Upper Separation Condition (USC):** the median of feature f computed over class C_k is strictly larger than the maximum value of f observed across all other classes, i.e.,

$$\tilde{\mu}_k^{(f)} > \max_{j \neq k} \left(\max_{i \in C_j} x_i^{(f)} \right). \tag{2}$$

According to the definition of the median, a feature satisfying either LSC or USC guarantees that at least 50% of the training samples of class C_k can be unambiguously assigned to that class based on that single feature alone. This property provides a natural lower bound on the classification coverage achievable through the selected features.

In practice, multiple features may satisfy LSC or USC for the same class. To resolve such ambiguities, we select the feature exhibiting the smallest intra-class standard deviation $\sigma_k^{(f)}$ among those satisfying the corresponding condition. This criterion is motivated by the observation that features with lower variability within a class are more likely to yield stable, reliable separation boundaries. Moreover, such features are expected to remain robust under perturbations introduced by image compression, since low intra-class variance implies that the feature values are tightly concentrated around their median and are thus less susceptible to compression-induced shifts. Following this procedure, at most two features are retained per class: one satisfying LSC and one satisfying USC. The resulting feature sets are thus remarkably compact. Formally, for each class C_k, we denote by l_k the index of the feature satisfying LSC and by u_k the index of the feature satisfying USC. The corresponding class-specific statistics, $\tilde{\mu}_{l_k}$, σ_{l_k}, $\tilde{\mu}_{u_k}$, and σ_{u_k}, computed from the training data, are stored and used by the classification stage described next.

4.2. Classification strategy

Given the selected feature indices and their associated statistics, the RFC employs a two-stage classification strategy that combines deterministic threshold-based rules with a distance-based fallback mechanism. Algorithm 1 provides a concise summary of the complete RFC pipeline, encompassing both the feature selection and the classification stages.

Algorithm 1 Reduced-Feature Classifier (RFC) for Iris PAD

Feature Selection (Training Phase)
Input: Training embeddings $\mathbf{X}_{\text{train}}$, class labels $\{C_k\}_{k=0}^{K-1}$
Output: Feature indices $\{l_k, u_k\}$, statistics $\{\tilde{\mu}_{l_k}, \sigma_{l_k}, \tilde{\mu}_{u_k}, \sigma_{u_k}\}$ for each class

1. For each class C_k and each feature f, compute $\tilde{\mu}_k^{(f)}$, $\sigma_k^{(f)}$, $\min_k^{(f)}$, $\max_k^{(f)}$.
2. Identify features satisfying LSC (Eq. (1)) and USC (Eq. (2)).
3. Among LSC-satisfying features, select $l_k = \arg\min_f \sigma_k^{(f)}$.
4. Among USC-satisfying features, select $u_k = \arg\min_f \sigma_k^{(f)}$.
5. Store $\{l_k, u_k, \tilde{\mu}_{l_k}, \sigma_{l_k}, \tilde{\mu}_{u_k}, \sigma_{u_k}\}$ for each class.

Classification (Test Phase)
Input: Test embedding $\mathbf{x}$, stored indices and statistics
Output: Predicted class (bona fide or attack)

1. Evaluate DSC (Eq. (3)) for each class C_k.
2. If exactly one class satisfies DSC → assign sample to that class (direct classification).
3. Otherwise, compute d_k (Eq. (4)) for candidate classes and assign to $\arg\min_k d_k$.
4. Map to binary PAD decision: C_0 → bona fide, C_k $(k > 0)$ → attack.

4.2.1. Stage 1: Direct separation

Let $\mathbf{x} = \{x_i\}$ denote the full embedding vector associated with a test sample. In the first stage, the classifier evaluates whether the test sample can be directly assigned to a single class based on the selected features. Specifically, class C_k is considered a candidate if the following *direct separation condition* (DSC) is satisfied:

$$x_{l_k} < \tilde{\mu}_{l_k} \quad \text{or} \quad x_{u_k} > \tilde{\mu}_{u_k}. \tag{3}$$

If exactly one class satisfies the DSC, the sample is assigned directly to it. This rule exploits the separability guarantees provided by the LSC and USC conditions identified during feature selection: when DSC holds for a unique class, the sample lies in a region of the feature space that is exclusively associated with that class in the training data.

4.2.2. Stage 2: Distance-based resolution

If no class satisfies the DSC, or if multiple classes simultaneously satisfy it, the classifier resorts to a distance-based decision. For each candidate class C_k (or all classes, if no candidate was identified), a normalized Euclidean distance is computed using the two selected features:

$$d_k = \sqrt{\left(\frac{x_{l_k} - \tilde{\mu}_{l_k}}{\sigma_{l_k}} \right)^2 + \left(\frac{x_{u_k} - \tilde{\mu}_{u_k}}{\sigma_{u_k}} \right)^2}. \tag{4}$$

The sample is then assigned to the class with the smallest distance d_k. It is worth noting that the normalization by the class-specific standard deviations ensures that each feature contributes proportionally to its discriminative power, preventing features with larger absolute scales from dominating the distance computation.

This two-stage design ensures that the classifier is both efficient and robust: deterministic DSC rules are applied whenever possible, providing fast and confident decisions, while the distance-based fallback guarantees that a decision is always reached, even in ambiguous regions of the feature space.

It is worth emphasizing that, despite its simplicity, the RFC framework retains a direct connection to the Transformer-based representations from which the features are derived. The Transformer backbones, in fact, remain essential for generating the high-dimensional embeddings from which the discriminative features are selected. The RFC thus operates as a lightweight, interpretable decision layer on top of the learned representations, offering a favorable trade-off between computational efficiency and classification performance.

5. Experimental tests

An extensive set of experimental tests was first conducted to evaluate the effectiveness of a Transformer-based iris PAD and to assess the effects of image compression on iris image quality and on the associated achievable PAD performance. Then, the feasibility of improving results on compressed images, the applicability of the developed models to images processed with unseen codecs, and the possibility of leveraging simplified representations derived from the created attention-based embeddings were also investigated.

The results obtained in these experimental activities are reported in the following sections, after a brief description of the databases involved in the performed tests, the employed learning strategies, and the adopted performance metrics.

5.1. Iris databases

The iris images used in our experimental tests are drawn from two main sources: the Notre Dame Contact Lenses Dataset 2015 (NDCLD2015) [51], and the three publicly available subsets of the LivDet-Iris 2017 benchmark [52], namely the Notre Dame, Clarkson, and IIITD-WVU datasets (the Warsaw subset, originally part of the competition, is no longer distributed). These collections have been extensively used in the literature to assess the performance of several iris PAD strategies, as they provide a diverse set of acquisition conditions, sensor types, and presentation attack instruments.

The NDCLD2015 dataset contains 7300 iris images acquired in controlled indoor environments with uniform illumination and a stable imaging setup. It was specifically created to study the effect of contact lenses on iris recognition performance. The dataset comprises three balanced categories: bona fide irises without lenses, bona fide irises wearing soft, transparent lenses, and presentation attacks produced with textured contact lenses. Authentic samples correspond to subjects either not wearing any lenses or wearing standard, cosmetic-free soft lenses, whereas attack samples are generated by impostors using commercially available textured lenses from various brands and patterns. Images were captured during multiple acquisition sessions to introduce natural variability in pupil size, alignment, and eye appearance.

The LivDet-Iris 2017 database was designed to support rigorous evaluation under cross-PA and cross-dataset conditions. It consists of three independent subsets collected by different research groups using different sensors and environmental settings:

- the Notre Dame subset comprises 4800 images depicting both authentic irises and textured-lens attacks. The training partition includes 1200 samples (600 bona fide and 600 attacks), while the remaining 3600 images form the test set. Importantly, the test set includes both known attacks, produced using lens brands also present in the training data, and unknown attacks generated with different manufacturers and patterns not seen during training, thus allowing us to assess generalization capabilities across unseen PA categories;
- the IIITD-WVU subset was constructed explicitly to study cross-sensor and cross-environment generalization. The training set contains 6250 images, including bona fide samples, textured-lens attacks, printed irises, and lens-printouts, i.e., printed iris images photographed through contact lenses, all captured in controlled indoor settings. The test set consists of 4209 images collected from a disjoint subject pool, acquired with different sensors and under both controlled (indoor) and unconstrained (outdoor) conditions, thereby inducing variability in lighting, motion blur, and imaging distance;
- the Clarkson subset includes 4937 images in the training set and 3158 in the testing set. The training portion comprises bona fide images, irises wearing textured contact lenses, and printed iris attacks. Similarly to the Notre Dame subset, it mainly emphasizes cross-PA evaluation, including attacks in the test set samples generated using materials, printers, or printing resolutions that differ from those used for training, thus allowing for evaluating resilience to previously unseen print-based attacks. Acquisition conditions cover multiple illumination levels and slight variations in subject positioning.

Together, these datasets offer a broad spectrum of acquisition conditions, sensor technologies, and presentation attack methods. This variability enables a comprehensive assessment of PAD methods in realistic cross-PA and cross-dataset scenarios.

5.2. Training setup and evaluation metrics

The iris PAD classifiers employed in this study are built upon ViT and Swin Transformer backbones, both initialized using ImageNet-1K pre-trained weights to leverage strong general-purpose visual representations. The adopted implementations are available in the PyTorch Image Models (timm) library.[2] To enhance regularization and reduce overfitting in the downstream PAD task, we replaced each architecture's original final linear projection with a lightweight classification head comprising two fully connected layers separated by a dropout operation. This modification enables models to better adapt to dataset-specific cues while maintaining a controlled model capacity.

For each dataset, an independent model was trained, resulting in four ViT-based and four Swin-based classifiers. For NDCLD2015, we followed common practice in the literature and used 6000 images for training and the remaining samples for testing. For the LivDet-Iris 2017 subsets, we adhered to the official data partitions to preserve comparability with prior works and ensure a fair evaluation under cross-PA and cross-dataset conditions. All models were trained using cross-entropy as loss function and optimized with stochastic gradient descent (SGD) and momentum set at 0.9. A batch size of 128 was adopted to have a good trade-off between convergence stability and training speed across all networks. Early stopping was applied when no significant improvement in validation performance was observed for 10 consecutive epochs. To improve model robustness, a set of lightweight data-augmentation operations was applied during training. These included random horizontal flips, small in-plane rotations, and controlled variations of image brightness and contrast. No additional preprocessing or enhancement techniques, such as iris segmentation refinement, contrast equalization, or specular highlight suppression, were applied, so that the classifiers operate directly on the input images provided by each dataset. All tests were performed on an 4 × NVIDIA® Tesla V100 GPUs with 5120 CUDA cores and 32 GB GPU memory, on a personal computing platform with an Intel® Xeon® Gold 5218 CPU @ 2.30 GHz CPU using Ubuntu 18.04.6 LTS.

An aspect worth emphasizing concerns the label-space configuration used during training: empirically, we found that a three-class formulation led to improved discrimination for the NDCLD2015 (live, soft lens, textured lens) and Clarkson (live, textured lens, printout) datasets, likely because it allows the models to explicitly separate bona fide categories from different PA types. For the Notre Dame and IIITD-WVU subsets, however, the conventional binary classification scheme (bona fide *vs.* attack) yielded comparable or superior performance, and was therefore adopted. During evaluation, all systems were assessed under standard binary PAD metrics, with transparent-lens images categorized as bona fide to align with common PAD reporting protocols. Performance is quantified using the attack presentation classification error rate (APCER), the bona fide presentation classification error rate (BPCER), and their arithmetic mean, i.e., the average classification error rate (ACER) [13]. This latter metric has been employed during training to select the hyperparameter configurations giving the lowest ACER.

[2] https://github.com/huggingface/pytorch-image-model

Table 1
Iris PAD performance of the proposed methods, and comparison with literature algorithms. Bold values indicate the lowest ACERs for each iris database.

Database	Metric	D-NetPAD [53]	PBS [25]	A-PBS [25]	d-CBAM [26]	DFCANet [54]	ELF [33]	FAM [55]	AA-PAD [56]	ViT	Swin
NDCLD2015	BPCER	–	0.00%	0.06%	0.60%	21.35%	-	-	-	0.00%	0.00%
	APCER	–	1.09%	0.08%	0.80%	7.16%	-	-	-	0.00%	0.00%
	ACER	–	0.54%	0.07%	0.70%	13.13%	0.04%	-	-	**0.00%**	**0.00%**
Notre Dame	BPCER	3.32%	1.06%	0.00%	4.37%	–	-	0.00%	0.10%	0.61%	0.00%
	APCER	10.38%	8.89%	7.88%	11.67%	–	-	8.06%	12.20%	17.38%	6.42%
	ACER	6.81%	4.97%	3.94%	8.02%	–	3.75%	4.03%	6.15%	9.00%	**3.21%**
IIITD-WVU	BPCER	10.12%	8.26%	4.13%	–	6.75%	-	12.68%	0.10%	21.51%	11.41%
	APCER	36.41%	5.76%	8.86%	–	16.80%	-	1.00%	32.00%	24.07%	13.39%
	ACER	23.27%	7.01%	**6.50%**	–	11.78%	3.54%	6.84%	16.05%	22.79%	12.40%
Clarkson	BPCER	0.94%	0.00%	0.81%	0.87%	5.08%	-	0.81%	0.10%	0.35%	0.40%
	APCER	5.78%	8.97%	6.16%	5.79%	1.81%	-	6.10%	8.90%	9.10%	4.02%
	ACER	3.36%	4.48%	3.48%	3.33%	3.44%	-	3.45%	4.50%	4.72%	**2.21%**

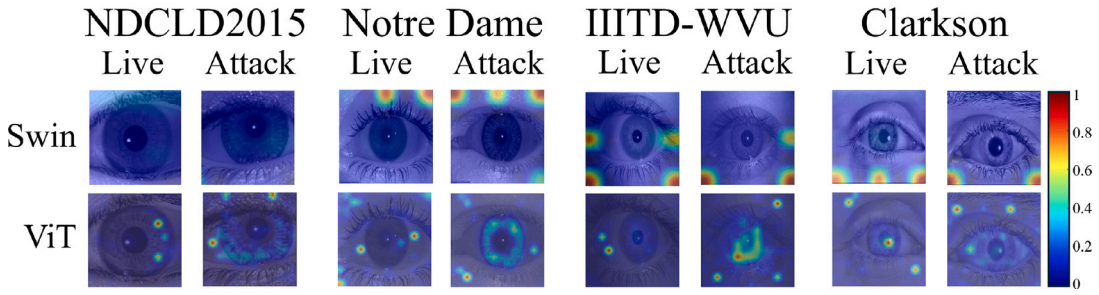


Fig. 1. Attention maps generated by ViT and Swin models for bona fide and spoofing test images from the databases considered, with the corresponding colormap used to display attention values superimposed to the original images.

5.3. Iris PAD on original images

Table 1 reports the results obtained when applying the adopted ViT- and Swin-based classifiers to the considered iris databases, as well as the performance obtained in state-of-the-art approaches relying on attention modules in convolutional architectures. More in detail, the approaches considered for comparison include both standard CNN-based approaches such as early–late fusion (ELF) [33] and adversarially-augmented PAD (AA-PAD) [56], as well as methods exploiting attention-based strategies as in D-NetPAD [53], pixel-wise binary supervision (PBS) and attention-based deep PBS (A-PBS) [25], dense convolutional block attention module (d-CBAM) [26], frequency-based attention module (FAM) [55], and dense feature calibration attention-assisted network (DFCANet) [54].

It can be seen that ViTs generally achieve good performance, whereas the Swin framework achieves the best results among the considered methods, except for the IIITD-WVU iris database, for which results align with state-of-the-art methods. Thus, both ViT and Swin models demonstrate strong performance on samples with visual characteristics encountered during training, as reflected by the results on NDCLD2015. At the same time, the Swin framework exhibits greater generalization, adapting better to the heterogeneous scenarios of the Notre Dame and Clarkson subsets. Its limitations appear when applied to the IIITD-WVU dataset, where substantial discrepancies between training and testing conditions, including different sensors, acquisition settings, and subject populations, pose a significantly more demanding challenge.

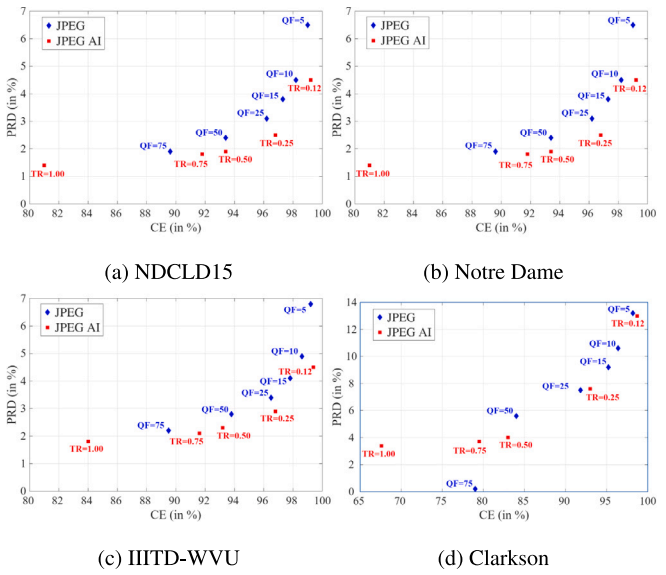
It is worth remarking that, as mentioned in Section 2.1, besides exploring the recognition performance achievable by the considered approaches, it would be also relevant to investigate their capabilities in providing interpretable results, describing in intelligible ways which aspects are mostly exploited to reach specific decisions.

To this aim, Fig. 1 illustrates the attention maps produced by the ViT and Swin models when processing both bona fide and spoofing samples taken from all the considered datasets. From the depicted maps, a consistent observation is that the two classes tend to activate different regions of interest, reflecting each model's attempt to isolate discriminative texture cues. It can be seen that ViTs generally provide more spatially coherent and fine-grained attention patterns, thanks to their global self-attention mechanism, which evaluates relationships across the entire image. In contrast, the Swin Transformer, which achieves computational efficiency and scalability through shifted-window-based attention mechanisms, yields more fragmented and localized activation maps. These smaller, window-based patches may contain useful cues but are harder to interpret visually and capture less of the global iris structure, making them less preferable than ViT for explainability. In particular, the attention patterns of ViTs associated with bona fide and attack samples appear remarkably stable across datasets. Such behavior indicates that, for explainability purposes, ViT models are more informative and appealing than Swin frameworks, although the latter may achieve better PAD performance, and both approaches warrant further investigation.

It has to be also observed that, despite the noticed consistency across datasets, each collection's unique visual properties, such as sensor-induced texture variations, imaging noise, or presentation attack differences, still have a systematic impact on performance. This result reinforces the importance of conducting dataset-specific evaluations in PAD research. It also highlights that, while Transformer-based models can learn generalizable cues, their robustness and interpretability remain tightly linked to the diversity and representativeness of the training data.

5.4. Quality of compressed iris images

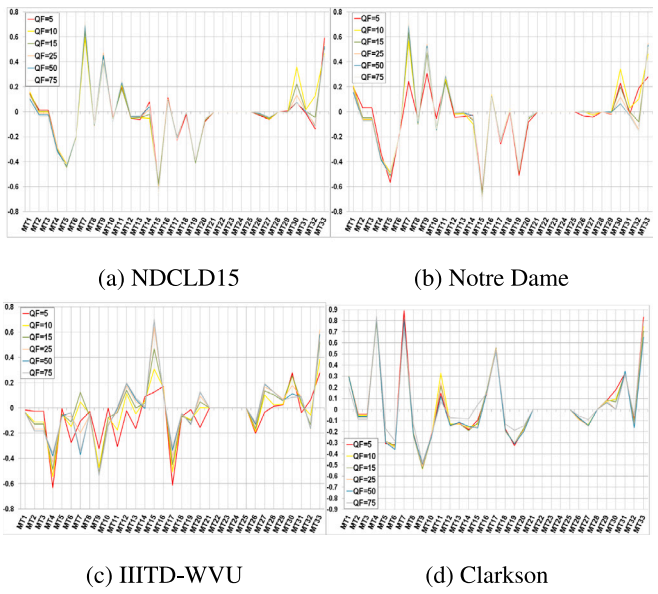
Several experimental tests were conducted to investigate the effects of compression on iris images and the consequences on iris PAD. More in detail, we applied both JPEG and JPEG AI codecs to the iris images in the considered databases. Quality factors (QFs) belonging to the set $\{75, 50, 25, 15, 10, 5\}$ were used for increasing JPEG compression, and target rates (TRs) equal to $\{1.00, 0.75, 0.50, 0.25, 0.12\}$ bit-per-pixel (bpp) were selected when running the JPEG AI reference software [57]. While JPEG QFs deterministically control the quantization table (Q-table)



(a) NDCLD15 (b) Notre Dame

(c) IIITD-WVU (d) Clarkson

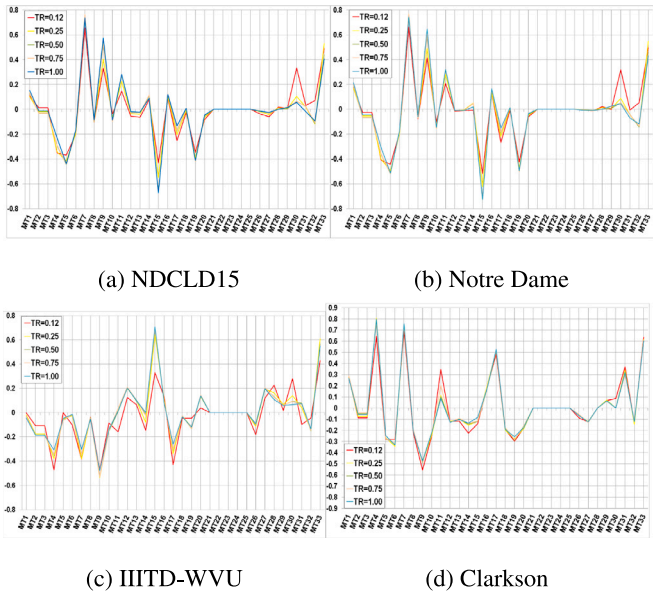
Fig. 2. Scatterplots showing CE *vs* PRD behaviors when applying JPEG and JPEG AI compression on the considered iris databases.



(a) NDCLD15 (b) Notre Dame

(c) IIITD-WVU (d) Clarkson

Fig. 3. Correlation between PRD and BIQT iris quality metrics, for JPEG compression.



(a) NDCLD15 (b) Notre Dame

(c) IIITD-WVU (d) Clarkson

Fig. 4. Correlation between PRD and BIQT iris quality metrics, for JPEG AI compression.

used in the standard, JPEG AI TRs specify a desired compression level, determined using different pre-trained fixed-rate encoding models, thus implementing a stochastic compression process.

Fig. 2 depicts the scatterplots associated with the average values of percentage root-mean-square difference (PRD) and compression efficiency (CE) on the considered iris datasets for the evaluated codecs. Here, CE is defined as $CE = 100 \cdot (1 - S_c/S_o)$, with S_o and S_c representing the sizes (in bytes) of the original and compressed images, respectively. The PRD measures the distortion between original and compressed images as a percentage [58]. From the reported results, it can be seen that very similar behaviors are observed for the NDCLD15, Notre Dame, and IIITD-WVU datasets, whereas notable differences are observed for the Clarkson datasets, because this latter dataset contains images natively JPEG-compressed with QF = 75. However, as expected, JPEG AI outperforms JPEG across all considered scenarios, enabling higher CE values at the same PRD level. Note that the default fixed-rate encoding configuration [57] was used for JPEG AI since it is computationally faster, yet it does not fully implement rate control. As a result, the actual rate may differ slightly from the nominal target for individual images, and thus the average rate across the dataset is used to compute the CE.

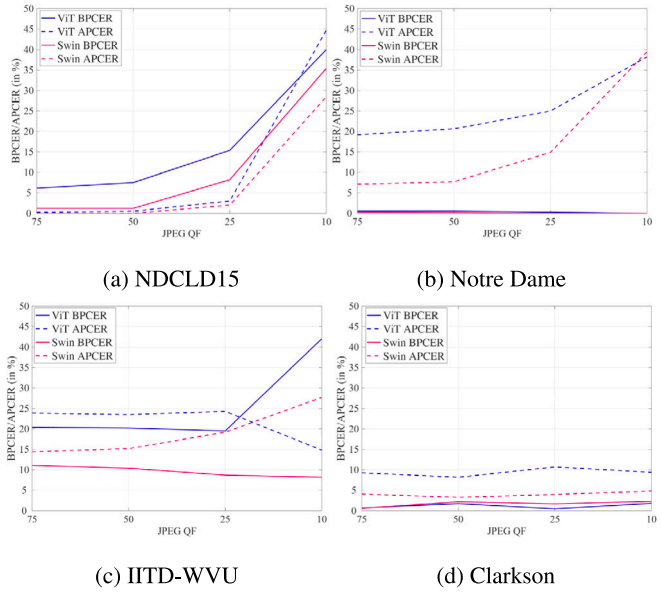
The performed analysis on the effects of compression on iris images shows that similar results in terms of PRD/CE values are generally achieved for images compressed with (JPEG, JPEG AI) settings given by (QF = 75, TR = 0.75), (QF = 50, TR = 0.50), (QF = 25, TR = 0.25), and (QF = 10, TR = 0.12) pairs. To provide a fair comparison of the effects of JPEG and JPEG AI compression on iris PAD, the results reported in the following sections will be limited to scenarios obtained with the aforementioned parameters for the employed codecs.

A deeper investigation on the effects of the distortions introduced by JPEG and JPEG AI on iris images was conducted by specifically considering aspects associated with the use of iris as a biometric trait in automatic recognition systems. To this aim, we have considered the metrics commonly employed to evaluate iris image quality, as defined by the ISO/IEC 29794-6:2015 standard [60,61]. More in detail, the parameters summarized in Table 2 were computed for the iris images in the considered databases using BIQTIris [59], an open-source reference library designed for determining statistics on iris image quality.

Then, the correlation between the PRD values of the compressed images and each of the tested quality metrics was evaluated across increasing JPEG and JPEG AI compression levels, and the results are reported in Figs. 3 and 4, respectively. As can be observed, the Notre Dame and NDCLD15 datasets exhibit similar trends. In particular, an absolute correlation coefficient greater than 50% is observed for three ISO/IEC 29794-6 metrics when considering both JPEG- and JPEG AI-compressed images, namely MT7 (ISO greyscale utilization), MT9 (ISO iris–pupil contrast), and MT15 (ISO sharpness). Also, for the IIITD-WVU dataset, MT9 and MT15 remain strongly correlated with the PRD distortion. For the Clarkson dataset, MT7 remains highly correlated with the PRD distortion, while MT9 also exhibits a good level of correlation. However, unlike the other datasets, under both JPEG and JPEG AI compression, high correlation values are observed for MT4 (contrast) and MT17 (normalized contrast). The different behavior observed for the Clarkson dataset can be explained by the fact that its images are originally stored in JPEG format, whereas the other datasets consist of uncompressed images.

Table 2
BIQTIris iris quality metrics [59].

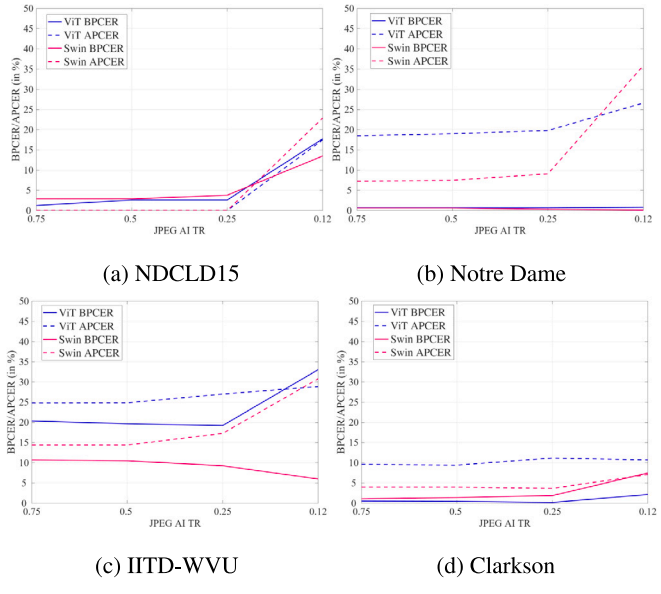
ID	Feature	ID	Feature	ID	Feature
MT1	iris diameter	MT12	iso margin adequacy	MT23	normalized iso iris-pupil concentricity
MT2	pupil diameter	MT13	iso overall quality	MT24	normalized iso iris-pupil contrast
MT3	pupil radius	MT14	iso pupil boundary circularity	MT25	normalized iso iris-pupil ratio
MT4	contrast	MT15	iso sharpness	MT26	normalized iso iris-sclera contrast
MT5	iris-pupil greyscale	MT16	iso usable iris area	MT27	normalized iso margin adequacy
MT6	iris- sclera greyscale	MT17	normalized contrast	MT28	normalized iso sharpness
MT7	iso greyscale utilization	MT18	normalized iris diameter	MT29	normalized iso usable iris area
MT8	iso iris-pupil concentricity	MT19	normalized iris-pupil grayscale	MT30	normalized sharpness
MT9	iso iris-pupil contrast	MT20	normalized iris-sclera grayscale	MT31	pupil circularity average deviation
MT10	iris-pupil ratio	MT21	normalized iso greyscale utilization	MT32	quality
MT11	iso iris-sclera contrast	MT22	normalized iso iris diameter	MT33	sharpness



(a) NDCLD15 (b) Notre Dame

(c) IITD-WVU (d) Clarkson

Fig. 5. Iris PAD performance for increasing JPEG compression levels.



(a) NDCLD15 (b) Notre Dame

(c) IITD-WVU (d) Clarkson

Fig. 6. Iris PAD performance for increasing JPEG AI compression levels.

5.5. Iris PAD on compressed images

We have then investigated the effects of image compression on the achievable PAD performance. Figs. 5 and 6 depict the behavior observed when applying detectors trained on original, uncompressed samples to iris images compressed with JPEG and JPEG AI, respectively, across different compression levels of the used encoders. The results obtained on the considered datasets indicate that compression mainly affects the APCER, i.e., the capability to detect spoofing attacks, especially for low-quality images. The loss of details in the processed images prevents the adopted detectors from recognizing the typical patterns of fake irises observed and learned during training.

Fig. 7 shows the attention maps for the same iris images considered in Fig. 1, yet compressed at low quality, i.e., with a QF = 10 for JPEG and TR = 0.12 bpp for JPEG AI. ViTs are adopted for such analysis given their greater explainability capabilities, as discussed in Section 5.3. While global attention is still preserved to some extent, compression artifacts reduce sharpness and contrast, as observed in Section 5.4, particularly for spoofing samples, yielding maps similar to those obtained for bona fide samples. Specific examples leading to misclassifications are depicted in Fig. 8, where the maps generated by ViT models for bona fide and spoofing samples from the NDCLD2015 database are reported, considering images both uncompressed and compressed at different JPEG and JPEG AI qualities. It can be seen that the distortions introduced by the strongest compressions significantly alter the structures of the produced attention maps, whereas medium-level compressions affect them only slightly. The notable modifications of the detectors' interpretations lead to failures, with errors returned for JPEG and JPEG AI compressions at minimum rates.

The relations between the distortions introduced by compression and the consequent modifications in the created attention maps were further analyzed through a quantitative evaluation of the correlation between these two aspects. In detail, different metrics were used to compute the (dis)similarity between maps of uncompressed and compressed images, and the Pearson correlation between these measurements and the PRD between the uncompressed and compressed images was then estimated. For images compressed with (JPEG; JPEG AI), correlation values at (−0.81; −0.83) were obtained when using structural similarity index (SSIM) as attention map similarity measure, (0.72; 0.77) for PRD between maps, (−0.68; −0.62) for cosine similarity, and (−0.72, −0.73) for Spearman rank correlation between average attention on 16 × 16 blocks. All the considered tests confirm a strong connection between the distortions induced in the processed images and the modifications in the attention maps generated by the employed detectors.

Fig. 9 depicts the scatter plots of the SSIM *vs* PRD measures for the considered databases, with JPEG and JPEG AI compression, indicating a strong correlation between compression-induced image distortions and changes in the visual structure of the generated attention maps. It is worth observing that strong correlations were also observed between the considered similarity metrics for attention maps and the iris quality metrics most closely related to image PRD, that is, ISO sharpness and

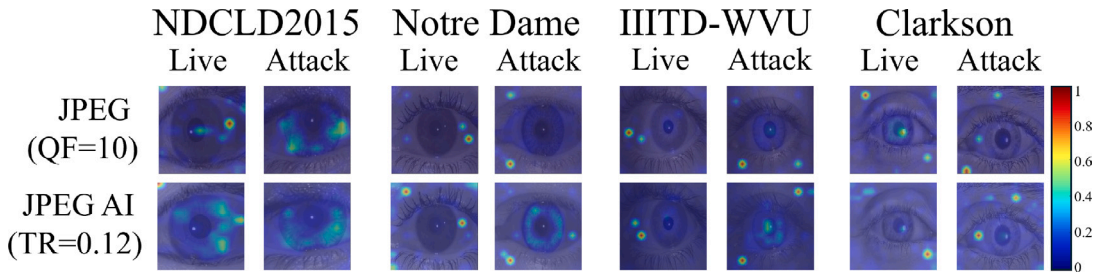


Fig. 7. ViT attention maps for JPEG and JPEG AI compressed images, for bona fide and spoof images from the databases considered.

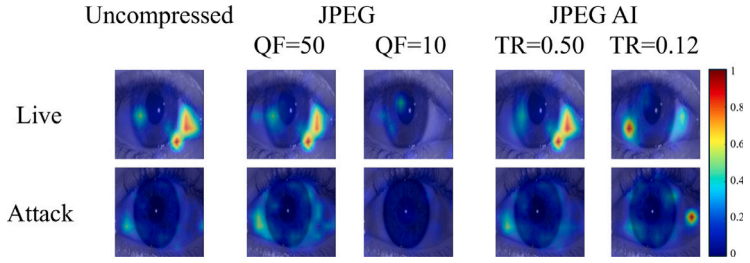
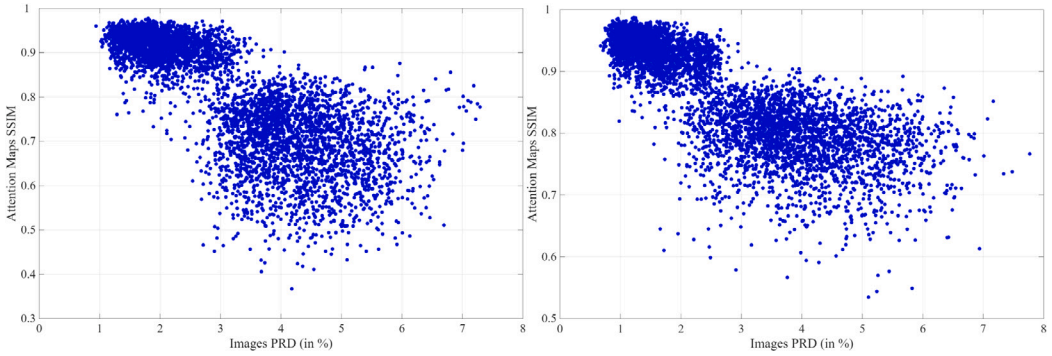


Fig. 8. Attention maps generated by ViT models for bona fide and spoofing samples from the NDCLD2015 database, for images uncompressed and compressed at different JPEG and JPEG AI levels. The depicted samples are correctly classified in their uncompressed and medium-quality compressed versions, and incorrectly classified in their low-quality compressed versions. In these latter cases, the introduced distortions alter the detectors' interpretation, leading to failures.



(a) JPEG (b) JPEG AI

Fig. 9. Scatterplots of SSIM between attention maps derived from original and compressed images *vs* PRD between original and compressed images. A significant correlation between compression-induced distortions and modifications on the generated attention maps is observed.

contrast as shown in Section 5.4, with SSIM having Pearson correlation at 0.78 and 0.76 with the two iris quality metrics, respectively.

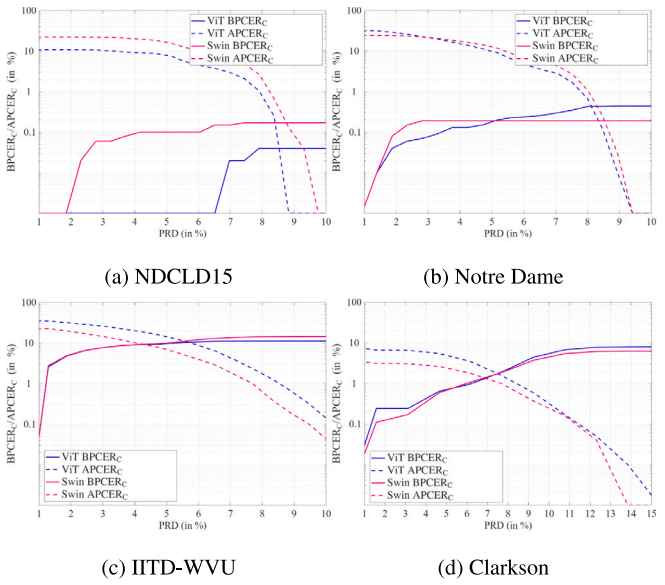
Furthermore, the trade-off between compression-induced image distortions and biometric recognition capabilities was also assessed by focusing on the relationship between the PRD and both the cumulative BPCER ($BPCER_C$) and the cumulative APCER ($APCER_C$). Specifically, here $BPCER_C(x)$ indicates the percentage of genuine samples with PRD $< x$ that are incorrectly rejected, thus expressing the system's ability to correctly recognize legitimate users under a given maximum level of distortion x. Conversely, $APCER_C(x)$ denotes the percentage of fake samples with PRD $> x$ that are incorrectly accepted as genuine.

The performance obtained when applying the ViT- and Swin-based detectors trained on uncompressed images on samples compressed with JPEG and JPEG AI are reported in logarithmic scale in Figs. 10 and 11. It is worth noting that the reported values of $BPCER_C$ and $APCER_C$ were computed over all the available compressed images, regardless the specific TR for JPEG AI, or QF for JPEG, used during the compression.

As already shown in Section 5.3, Swin models generally outperform ViTs, achieving in most cases lower error rates (especially $BPCER_C$) at the same PRD level for both JPEG- and JPEG-AI-compressed images. The plots in Figs. 10 and 11 allow determining, for each dataset and model, the PRD value corresponding to the equal error rate (EER), identified as the operating condition at which the $APCER_C$ equals the $BPCER_C$.

5.5.1. *Fine-tuning for improved iris PAD on compressed images*

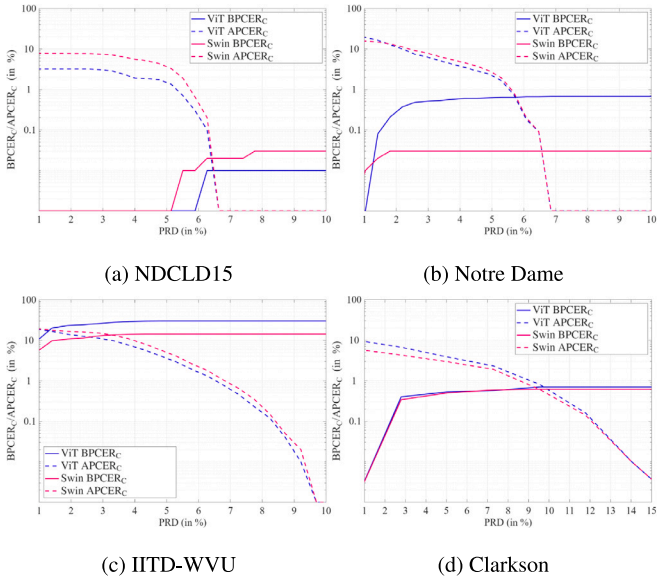
Having observed the considerable effects of compression on PAD performance, we investigated whether countermeasures can be designed to improve results on low-quality iris images. To this end, we evaluated whether fine-tuning models trained on uncompressed images, using samples compressed at increasing compression levels, may improve the attainable behavior. A learning rate (10^{-4}) approximately one order of magnitude lower than that used at the end of the initial training on uncompressed images was adopted, together with a cosine decay schedule, to enable stable adaptation while preserving the pretrained representations. Training was conducted by fine-tuning the whole networks (no backbone freezing) with stochastic gradient descent (SGD) with momentum set at 0.9 and batch size of 64 samples for a maximum of 1000 epochs and monitored on a validation subset, with early stopping applied when no significant improvement in validation performance was observed for 10 consecutive iterations.



(a) NDCLD15 (b) Notre Dame

(c) IITD-WVU (d) Clarkson

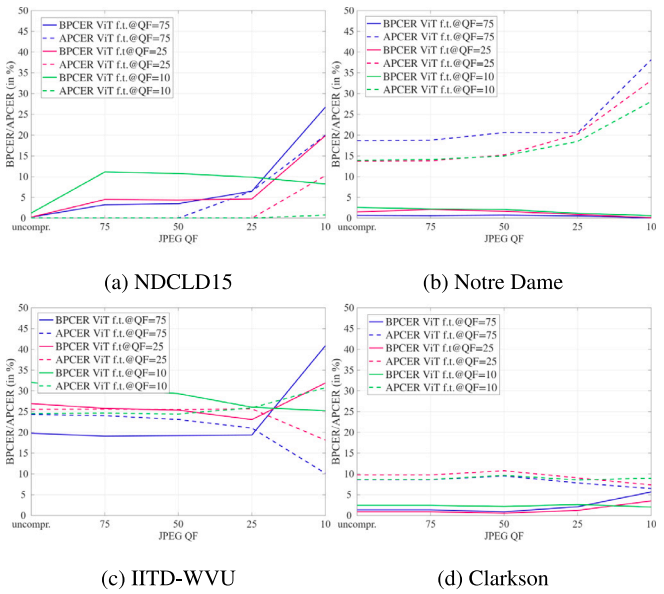
Fig. 10. BPCER$_C(x)$ and APCER$_C(x)$ curves for JPEG-compressed images.



(a) NDCLD15 (b) Notre Dame

(c) IITD-WVU (d) Clarkson

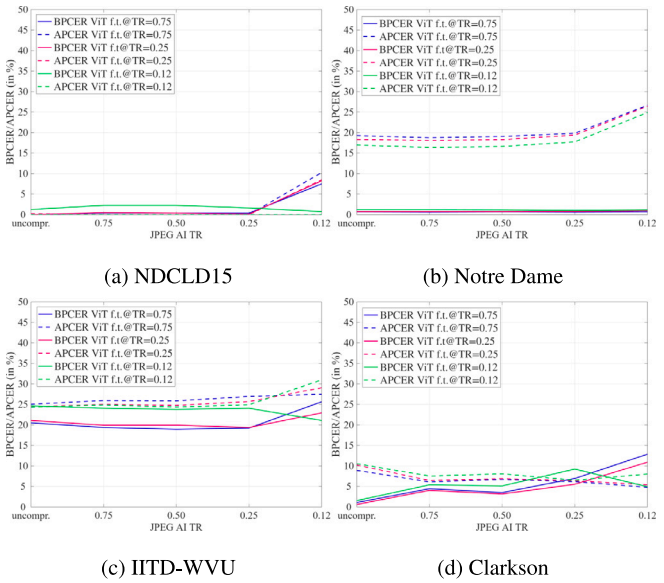
Fig. 11. BPCER$_C(x)$ and APCER$_C(x)$ curves for JPEG-AI-compressed images.



(a) NDCLD15 (b) Notre Dame

(c) IITD-WVU (d) Clarkson

Fig. 12. Iris PAD performance for ViT models fine-tuned with images JPEG-compressed at different QFs, at increasing JPEG compression levels.



(a) NDCLD15 (b) Notre Dame

(c) IITD-WVU (d) Clarkson

Fig. 13. Iris PAD performance for ViT models fine-tuned with images JPEG-AI-compressed at different TRs, at increasing JPEG AI compression levels.

The results obtained for ViT models are reported in Figs. 12 and 13, respectively for JPEG and JPEG AI compressions. The observed behaviors suggest that fine-tuning the original models, exploiting samples compressed at a given quality, consistently improves PAD results for that compression level and typically also yields beneficial effects for less severe compressions. For instance, across the considered databases, an average improvement of 10.8% in terms of the half total error rate (HTER), compared to models trained only on uncompressed images, is observed for images compressed at the lowest JPEG quality, and an improvement in HTER of 5.9% is achieved for the lowest-quality JPEG-AI-compressed images.

The performance on uncompressed images are generally only slightly affected (average modifications lower than 0.5%), thus overall achieving more stable behaviors across different compression levels, with respect to what can be accomplished without fine-tuning, as shown by comparing the obtained results with those in Figs. 5 and 6.

An analogous behavior is observed when considering Swin Transformers: for reasons of compactness and readability, these results are reported in Tables 3 and 4 for models trained on the original images as well as for those fine-tuned using the images with the largest considered compression factors for JPEG and JPEG AI, i.e., QF = 10 for JPEG and TR = 0.12 for JPEG AI. An average HTER improvement of 9.80% is here obtained for images compressed at the lowest JPEG quality, with an HTER improvement of 7.33% for the lowest-quality JPEG-AI-compressed images.

5.5.2. *Generalizability of iris PAD for learning-based compression*

To further assess the robustness of the proposed PAD solutions under learning-based compression, we investigate how well models fine-tuned on JPEG AI-compressed images generalize to iris samples encoded with a different learning-based image codec. In particular, we consider the ELIC (Efficient Learned Image Compression) codec [62], a state-of-the-art neural compression framework that differs from JPEG

Table 3
Iris PAD performance at different compression levels for Swin models trained on uncompressed images and fine-tuned with images compressed with JPEG @ QF = 10.

Metric	QF	NDCLD15		Notre Dame		IITD-WVU		Clarkson	
		Original	Fine-tuned	Original	Fine-tuned	Original	Fine-tuned	Original	Fine-tuned
BPCER	uncompressed	0.00%	0.00%	0.00%	0.00%	11.41%	28.04%	0.40%	1.91%
	75	1.22%	0.37%	0.31%	0.00%	11.17%	28.04%	0.65%	2.86%
	50	1.29%	0.37%	0.23%	0.00%	10.43%	26.06%	2.23%	4.95%
	25	8.22%	0.37%	0.00%	0.00%	8.76%	26.60%	1.78%	2.72%
	10	35.41%	0.12%	0.00%	0.00%	8.22%	24.19%	2.31%	1.18%
APCER	uncompressed	0.00%	0.00%	6.42%	9.12%	13.39%	7.22%	4.02%	2.91%
	75	0.00%	0.00%	7.12%	11.11%	14.43%	8.80%	4.12%	3.64%
	50	0.00%	0.00%	7.74%	11.33%	15.21%	8.80%	3.38%	3.20%
	25	2.05%	0.00%	14.95%	14.15%	19.25%	8.86%	4.38%	3.90%
	10	28.51%	0.00%	39.58%	26.94%	27.73%	11.85%	4.81%	4.10%

Table 4
Iris PAD performance at different compression levels for Swin models trained on uncompressed images and fine-tuned with images compressed with JPEG AI @ TR = 0.12.

Metric	TR	NDCLD15		Notre Dame		IITD-WVU		Clarkson	
		Original	Fine-tuned	Original	Fine-tuned	Original	Fine-tuned	Original	Fine-tuned
BPCER	uncompressed	0.00%	0.00%	0.00%	0.00%	11.41%	26.51%	0.40%	9.16%
	0.75	2.90%	0.62%	0.66%	0.70%	10.78%	30.74%	1.18%	10.90%
	0.50	2.90%	0.52%	0.66%	1.11%	10.53%	30.74%	1.46%	10.90%
	0.25	3.83%	0.52%	0.32%	1.94%	9.34%	31.41%	1.97%	9.17%
	0.12	13.51%	0.00%	0.12%	0.00%	6.07%	15.90%	7.56%	6.89%
APCER	uncompressed	0.00%	0.00%	6.42%	3.24%	13.39%	5.67%	4.02%	1.20%
	0.75	0.00%	0.00%	7.33%	4.23%	14.41%	8.23%	4.08%	1.91%
	0.50	0.00%	0.00%	7.33%	4.38%	14.41%	8.15%	4.08%	1.91%
	0.25	0.00%	0.00%	9.12%	4.83%	17.34%	8.04%	3.75%	2.31%
	0.12	22.91%	0.00%	35.91%	23.70%	30.83%	18.10%	7.12%	0.59%

Table 5
Iris PAD performance for ViT and Swin models trained on uncompressed images and fine-tuned with images compressed with JPEG AI @ TR = 0.12, when applied to images compressed with JPEG AI @ TR = 0.12 and images compressed with ELIC at a comparable rate.

Model	Metric	Test	NDCLD15		Notre Dame		IITD-WVU		Clarkson	
			Original	Fine-tuned	Original	Fine-tuned	Original	Fine-tuned	Original	Fine-tuned
ViT	BPCER	JPEG AI	17.75%	0.75%	0.83%	1.16%	33.04%	21.08%	2.15%	4.21%
		ELIC	0.00%	0.00%	0.66%	0.44%	31.48%	23.81%	9.22%	8.00%
	APCER	JPEG AI	17.50%	0.00%	26.61%	24.28%	28.85%	30.11%	10.70%	8.01%
		ELIC	36.25%	12.25%	34.70%	33.64%	30.76%	32.77%	7.94%	6.99%
Swin	BPCER	JPEG AI	13.51%	0.00%	0.12%	0.00%	6.07%	15.90%	7.56%	6.89%
		ELIC	0.00%	0.00%	0.00%	0.00%	21.79%	19.37%	0.88%	1.68%
	APCER	JPEG AI	22.91%	0.00%	35.91%	23.70%	30.83%	18.10%	7.12%	0.59%
		ELIC	74.00%	21.75%	47.23%	33.81%	27.93%	29.21%	11.35%	9.74%

AI in several fundamental aspects. While JPEG AI employs a standardized architecture featuring a single-level hyperprior, context modeling, and dedicated tools for both human- and machine-centric reconstruction, ELIC adopts a lighter encoder–decoder structure with a hierarchical latent representation and a hybrid entropy model that combines hyperpriors with partially autoregressive components. These design choices allow ELIC to achieve competitive rate–distortion performance with lower computational complexity, while also producing compression artifacts whose statistical characteristics differ from those of JPEG AI. As a result, ELIC provides an appropriate test case to evaluate whether PAD models fine-tuned on JPEG AI artifacts can generalize to unseen neural compression schemes.

In this analysis, all ViT- and Swin-based PAD models are first trained on uncompressed images and subsequently fine-tuned using JPEG AI at TR = 0.12 bpp. The resulting networks are then evaluated on uncompressed images, and on images compressed with both JPEG AI at TR = 0.12 bpp and ELIC at a rate comparable to TR = 0.12 bpp. The corresponding results are summarized in Table 5.

When tested on JPEG-AI–compressed images, both ViT and Swin architectures exhibit the benefits of fine-tuning already discussed in Section 5.5.1: for most datasets, BPCER values remain low or decrease after fine-tuning, while APCER is significantly reduced, confirming that codec-specific adaptation can effectively mitigate the distortions introduced by JPEG AI. For ELIC-compressed images, the performance attained when applying the detectors trained on uncompressed images is typically worse (in terms of ACER) than that obtained for JPEG AI images, which suggests that ELIC likely introduces more distortions than JPEG AI at comparable compression rates. It can be yet observed that adopting detectors fine-tuned on images severely compressed with JPEG AI partially relieves the issues associated with ELIC. This means that some of the introduced artifacts are common to those produced by JPEG AI, and thus manageable by a network that had experiences with samples characterized by a similar kind of compression noise. It can also be observed that Swin-based detectors, which generally outperform ViTs in matched-codec conditions, tend to be more sensitive to learning-based codec mismatch, especially in terms of APCER, whereas

Table 6
Classification accuracy on JPEG-compressed images at different compression levels, for detectors based on ViT and RFC. Values of average differences and paired t-test p-values (in columns for databases, in rows for compression levels) are also reported. The overall average difference across all the obtained results is −0.17%, with a *p*-value of 0.898.

QF	NDCLD15		Notre Dame		IITD-WVU		Clarkson		Avg. diff.	p-value
	ViT	RFC	ViT	RFC	ViT	RFC	ViT	RFC		
75	95.92%	95.50%	99.67%	99.44%	76.72%	82.18%	94.78%	94.65%	+1.17%	0.473
50	94.92%	94.92%	99.67%	99.50%	77.05%	82.18%	94.84%	94.49%	+1.15%	0.449
25	89.00%	89.92%	99.61%	99.56%	76.50%	81.14%	94.11%	93.63%	+1.25%	0.359
10	59.25%	61.67%	97.06%	97.22%	80.64%	79.23%	94.17%	75.93%	−4.26%	0.433
Avg. diff.	+0.73%		−0.07%		+3.45%		−4.77%			
p-value	0.330		0.461		0.124		0.363			

Table 7
Classification accuracy on JPEG-AI-compressed images at different compression levels, for detectors based on ViT and RFC. Values of average differences and paired t-test p-values (in columns for databases, in rows for compression levels) are also reported. The overall average difference across all the obtained results is 0.86%, with a *p*-value of 0.225.

TR	NDCLD15		Notre Dame		IITD-WVU		Clarkson		Avg. diff.	p-value
	ViT	RFC	ViT	RFC	ViT	RFC	ViT	RFC		
0.75	99.31%	99.03%	99.61%	99.39%	75.93%	81.28%	94.65%	94.58%	+1.19%	0.452
0.50	98.54%	98.47%	99.61%	99.56%	76.00%	81.61%	94.77%	94.55%	+1.31%	0.425
0.25	98.54%	97.78%	99.61%	99.50%	74.27%	78.88%	93.98%	94.01%	+0.94%	0.500
0.12	86.39%	86.88%	99.33%	99.22%	70.44%	74.86%	93.31%	88.56%	+0.01%	0.995
Avg. diff.	−0.15%		−0.12%		+5.00%		−1.25%			
p-value	0.592		0.410		0.040		0.362			

ViTs exhibit slightly more stable behavior across codecs at the cost of somewhat higher error rates in some scenarios.

The obtained results thus imply that cross-codec generalization represents an interesting aspect to be investigated, with the performed analysis representing the first attempt in the literature regarding this aspect. Codec-specific adaptation should be analyzed to further improve the performance achievable in applications involving cross-compression scenarios.

5.6. Experimental evaluation of the reduced-feature classifier

Having described the RFC methodology in Section 4, we here evaluate its effectiveness across all considered datasets and compression conditions. The analysis comprises classification accuracy (Section 5.6.1), a detailed breakdown of PAD metrics (Section 5.6.2), and an evaluation of the direct separation mechanism's efficiency (Section 5.6.3). Unless otherwise stated, the reported results are obtained using ViT-based embeddings. Regarding the size of the templates adopted in the proposed RFC detectors, as detailed in Section 4, at most two features are retained per class. This means that four features are selected from Transformer-based architectures trained on two classes (bona fide vs. attack) for the Notre Dame and IIITD-WVU datasets, while six features are extracted from detectors trained on three classes to better characterize the attacks for the NDCLD15 and Clarkson datasets. These numbers represent a reduction of approximately 128× to 192× relative to the original 768-dimensional ViT embeddings.

5.6.1. Classification accuracy

Tables 6 and 7 report the binary classification accuracy of the proposed RFC compared to the original ViT-based models under JPEG and JPEG AI compression, respectively. For each dataset, pairs of values are reported for different quality factors (QF) or target compression rates (TR) for both the original ViT model and the proposed RFC. The tables also include the average differences in accuracy between ViT- and RFC-based detectors (positive values indicate an average improvement of RFC over ViT), for each considered database and compression level, as well as the p-values obtained from paired-sample t-tests on the means of the accuracy differences. Typically, high p-values are obtained in the performed tests, testifying that it is not possible to reject the null hypothesis of having the same mean values for the performance obtained with the two distinct approaches, thus clearly demonstrating that the proposed RFC allows for preserving the achievable PAD performance while relying on extremely compact image representations.

More in detail, the RFC model performs on par with or even better than the full 768-dimensional ViT embeddings for the NDCLD15 and Notre Dame datasets, despite relying on only four or six features. In the IITD-WVU dataset, RFC shows a clear average improvement over ViT of approximately +5% for JPEG AI (only scenario with a statistically significant difference, given the low accuracies achieved for this database) and around +3.5% for JPEG. The RFC achieves lower accuracy than ViT only on the Clarkson dataset under high-distortion settings (QF = 10 and TR = 0.12). This behavior is likely attributable to the fact that the original images in the Clarkson dataset, on which the model was trained and the features were selected, are already JPEG-compressed.

We also applied the proposed selection process to Swin embeddings and compared RFC with the original Swin-based detectors for the considered datasets. Although full quantitative results are omitted for brevity, RFC achieved performance comparable to Swin across most compression settings, confirming its competitiveness despite relying on only 4 or 6 features rather than the original 49 × 768-dimensional embeddings. As a reference, for the NDCLD15 dataset, the average accuracy difference between RFC and the Swin-based classifier is −0.79% under JPEG compression and −1.41% under JPEG AI. This latter result is entirely driven by the highest-distortion setting, i.e., TR = 0.12, whereas for all higher TR values the difference was at most 0.2% in absolute value.

5.6.2. APCER, BPCER, and ACER analysis

To provide a more detailed characterization of the RFC behavior and enable direct comparison with the PAD metrics reported throughout this work, Tables 8 and 9 report the APCER, BPCER, and ACER values achieved by the RFC under JPEG and JPEG AI compression, respectively, alongside the corresponding values obtained by the ViT-based detector.

The results show that RFC performance degrades more rapidly than ViT under strong compression (QF = 10, TR = 0.12), especially on IITD-WVU and Clarkson, where both BPCER and ACER increase significantly. At moderate compression (QF≥25, TR≥0.25), the difference between the two methods reduces, with both detectors achieving consistently

Table 8
APCER, BPCER, and ACER for ViT and RFC on JPEG-compressed images at different quality factors.

Metric	QF	NDCLD15		Notre Dame		IITD-WVU		Clarkson	
		RFC	ViT	RFC	ViT	RFC	ViT	RFC	ViT
BPCER	75	7.75%	4.75%	0.78%	0.33%	39.17%	20.37%	1.21%	0.67%
	50	5.50%	2.75%	0.67%	0.33%	37.75%	20.23%	2.22%	1.75%
	25	2.50%	1.25%	0.33%	0.11%	36.04%	19.52%	0.67%	0.47%
	10	3.75%	0.75%	0.22%	0.00%	69.52%	42.02%	41.55%	1.82%
APCER	75	2.88%	3.75%	0.33%	0.33%	13.54%	23.87%	9.03%	9.26%
	50	4.88%	6.25%	0.33%	0.33%	13.83%	23.50%	8.43%	8.19%
	25	13.88%	15.88%	0.56%	0.67%	15.43%	24.29%	11.24%	10.70%
	10	55.63%	60.75%	5.33%	5.89%	11.01%	14.83%	8.55%	9.38%
ACER	75	5.31%	4.25%	0.56%	0.33%	26.27%	22.12%	5.12%	4.97%
	50	5.19%	4.50%	0.50%	0.33%	25.79%	21.86%	5.33%	4.97%
	25	8.19%	8.56%	0.44%	0.39%	25.73%	21.90%	5.96%	5.59%
	10	29.69%	30.75%	2.78%	2.94%	40.26%	28.43%	25.05%	5.60%

Table 9
APCER, BPCER, and ACER for ViT and RFC on JPEG-AI-compressed images at different target rates.

Metric	TR	NDCLD15		Notre Dame		IITD-WVU		Clarkson	
		RFC	ViT	RFC	ViT	RFC	ViT	RFC	ViT
BPCER	0.75	2.08%	0.83%	0.89%	0.44%	37.61%	20.37%	0.74%	0.54%
	0.50	1.88%	1.46%	0.56%	0.44%	37.32%	19.66%	0.94%	0.47%
	0.25	5.63%	3.12%	0.67%	0.44%	37.18%	19.23%	0.67%	0.20%
	0.12	10.21%	7.08%	0.89%	0.78%	61.40%	33.05%	15.09%	2.16%
APCER	0.75	0.42%	0.62%	0.33%	0.33%	14.94%	24.81%	9.57%	9.63%
	0.50	1.35%	1.46%	0.33%	0.33%	14.60%	24.86%	9.45%	9.45%
	0.25	0.52%	0.62%	0.33%	0.33%	17.91%	27.03%	10.71%	11.18%
	0.12	14.58%	16.88%	0.67%	0.56%	17.88%	28.86%	8.19%	10.71%
ACER	0.75	1.25%	0.73%	0.61%	0.39%	26.27%	22.59%	5.16%	5.08%
	0.50	1.61%	1.46%	0.44%	0.39%	25.96%	22.26%	5.20%	4.96%
	0.25	3.07%	1.88%	0.50%	0.39%	27.54%	23.13%	5.69%	5.69%
	0.12	12.40%	11.98%	0.78%	0.67%	39.64%	30.95%	11.64%	6.43%

Table 10
Percentage of test samples classified via direct separation (DSC%) and corresponding RFC accuracy, for each dataset under JPEG and JPEG AI compression.

Codec	Level	NDCLD15		Notre Dame		IITD-WVU		Clarkson	
		DSC%	Acc.	DSC%	Acc.	DSC%	Acc.	DSC%	Acc.
JPEG	QF = 75	69.25%	99.33%	73.17%	99.44%	70.11%	94.65%	69.39%	94.58%
	QF = 50	66.58%	98.92%	73.44%	99.50%	72.70%	94.49%	68.82%	94.55%
	QF = 25	50.92%	95.00%	71.61%	99.56%	67.92%	93.73%	68.16%	94.01%
	QF = 10	34.08%	59.17%	45.39%	97.22%	38.51%	75.93%	60.65%	88.56%
JPEG AI	TR = 0.75	73.13%	99.03%	74.17%	99.39%	22.832%	81.278%	69.39%	94.58%
	TR = 0.50	72.01%	98.47%	73.39%	99.56%	22.571%	81.611%	68.82%	94.55%
	TR = 0.25	70.97%	97.78%	76.28%	99.50%	25.731%	78.879%	68.16%	94.01%
	TR = 0.12	59.72%	86.88%	77.11%	99.22%	28.748%	74.863%	60.65%	88.56%

low error rates on NDCLD15 and Notredame. Under mild compression (QF = 75, TR = 0.75), RFC and ViT show similar performance across most datasets. These trends indicate that RFC is more sensitive to compression artifacts, but remains competitive when the compressed images preserve sufficient discriminative detail.

5.6.3. Analysis of direct separation classifications

A distinctive feature of the RFC is its ability to classify a subset of test samples through the direct separation condition (DSC), without resorting to the distance-based fallback. The proportion of samples classified via DSC, denoted DSC%, provides a meaningful indicator of classifier confidence: higher DSC% values indicate that a larger fraction of samples lie in well-separated regions of the feature space, where classification can be performed with high certainty based on a single feature comparison.

Table 10 reports the DSC% values obtained for each dataset under different compression levels for both JPEG and JPEG AI codecs. These values are computed on the test sets using the feature indices and thresholds determined during training. Overall, the DSC% values mirror the behavior observed in APCER, BPCER, and ACER: strong compression sharply reduces the proportion of samples classified through the RFC's high-certainty DSC mechanism, especially on IITD-WVU and Clarkson. Conversely, moderate to mild compression preserves a higher DSC% on NDCLD15 and Notre Dame, indicating that the separability structure learned during training remains largely stable for these datasets. JPEG AI consistently maintains slightly higher DSC% than JPEG, confirming that it preserves discriminative information more effectively under comparable compression levels.

5.6.4. Summary of RFC results

The experimental results presented in this subsection demonstrate that the proposed RFC achieves classification performance competitive with, and in some cases superior to, the full-dimensional Transformer-based detectors, despite operating on only 4 to 6 features extracted from the original 768-dimensional embeddings. The RFC exhibits particular strengths on the IITD-WVU dataset, where it consistently outperforms the ViT baseline, and on the NDCLD15 and Notre Dame datasets,

where performance is closely matched. The analysis of DSC% provides additional insight into the classifier's operating regime, confirming that the feature selection criterion based on inter-class separability and minimum intra-class variance identifies features that remain discriminative under moderate compression levels.

These results support the viability of the RFC as a lightweight, interpretable alternative to full Transformer-based classification for iris PAD, particularly in scenarios where computational resources, storage capacity, or bandwidth are limited, or where model interpretability is a priority.

6. Conclusions

This work has investigated presentation attack detection (PAD) for iris recognition under compression constraints. Pure-attention-based architectures, namely ViTs and Swin Transformers, have been used as detectors, with experimental results confirming their suitability for iris PAD under standard and cross-PA settings, with expected limitations under severe cross-dataset shifts. ViTs benefit from global self-attention, yielding more interpretable and coherent attention patterns, while Swin Transformers achieve even better accuracy, at the cost of a reduced interpretability, due to their hierarchical, window-based design. Properly addressing this trade-off between detection performance and interpretability is relevant for security-critical applications such as PAD, where guaranteeing high detection accuracy remains a priority, yet interpretability can support system auditing and analysis of potential failure cases, besides contributing to an increased user acceptability thanks to the possibility of providing hints about the models' behavior.

The analysis about the effects of compression on iris PAD has revealed that both traditional JPEG and learning-based JPEG AI codecs introduce artifacts that may affect the quality of iris images, and notably influence PAD accuracy, particularly at high compression levels. To deal with such issues, fine-tuning on compressed data has been adopted to improve PAD robustness against compression and achieve stable behaviors across different quality levels.

The generalizability of such approach to different learning-based compression schemes has been also investigated, noticing that relying on detectors fine-tuned on compressed images may help even when applying them to images compressed with codecs other than those considered during training, showing that different learning-based codecs may introduce similar artifacts. Yet, such scenario may represent a notable challenge when significantly-different distortions are introduced, suggesting that, in practical deployments, the performance of PAD systems may depend on the compression pipeline used in the acquisition or transmission process. Thus, models should ideally be validated or adapted for the specific compression scheme(s) expected in the operational environment, adopting for instance strategies involving training or fine-tuning on data compressed with a mixture of codecs to learn more compression-invariant features, motivating further research in cross-codec scenarios.

Furthermore, we have also explored the feasibility of significantly compressing the learned iris representations for PAD purposes, proposing a detection strategy relying on extremely-compact templates derived from the original embeddings, able to guarantee approximately the same detection performance while also showing robustness against compression schemes, a result particularly relevant for resource-constrained or distributed biometric systems.

CRediT authorship contribution statement

Rocco Albano: Software, Investigation. **Filippo Battaglia:** Writing – original draft, Software, Methodology, Investigation, Formal analysis. **Alessandro Gnutti:** Data curation, Conceptualization. **Emanuele Maiorana:** Writing – review & editing, Writing – original draft, Supervision, Methodology, Investigation, Formal analysis, Conceptualization. **Fabrizio Guerrini:** Software, Data curation, Conceptualization. **Giuseppe Campobello:** Writing – original draft, Supervision, Methodology, Funding acquisition. **Pierangelo Migliorati:** Supervision, Funding acquisition. **Patrizio Campisi:** Writing – review & editing, Funding acquisition.

Declaration of competing interest

The authors declare the following financial interests/personal relationships which may be considered as potential competing interests: Rocco Albano, Filippo Battaglia, Alessandro Gnutti, Emanuele Maiorana, Fabrizio Guerrini, Giuseppe Campobello, Pierangelo Migliorati, Patrizio Campisi reports financial support was provided by European Commission. Alessandro Gnutti, guest editor for the journal, is a co-author of this paper. Given his role as guest editor, he had no involvement in the peer review of this article and had no access to information regarding its peer review. Full responsibility for the editorial process for this article was delegated to another journal editor. If there are other authors, they declare that they have no known competing financial interests or personal relationships that could have appeared to influence the work reported in this paper. If there are other authors, they declare that they have no known competing financial interests or personal relationships that could have appeared to influence the work reported in this paper.

Acknowledgments

This work was partially supported by the European Union - Next Generation EU under the Italian National Recovery and Resilience Plan (NRRP), Mission 4, Component 2, Investment 1.3, CUP E13C22001870001, partnership on "Telecommunications of the Future" (PE00000001 - program "RESTART"), and conducted within the framework of the following cascade call projects: "EXperience and Privacy for Extended RealiTy" ("EXPERT" - CUP C89J24000280004), "Compression Of Biometric Signals for FUture Network Applications" ("COBS4FUN" - CUP C49J24000250004), "Formats for Representation, Analysis and Manipulation of Energy-friendly multimedia" ("FRAME" - CUP C89J24000250004).

Data availability

The authors do not have permission to share data.

References

[1] A.K. Jain, D. Deb, J.J. Engelsma, Biometrics: Trust, but verify, IEEE Trans. Biometrics Behav. Identity Sci. 4 (3) (2022) 303–323.
[2] K. Nandakumar, A.K. Jain, Biometric template protection: Bridging the performance gap between theory and practice, IEEE Signal Process. Mag. 32 (5) (2015) 88–100.
[3] A. Makrushin, A. Uhl, J. Dittmann, A survey on synthetic biometrics: Fingerprint, face, iris and vascular patterns, IEEE Access 11 (2023) 33887–33899.
[4] Irisguard, 2026, https://www.irisguard.com/iris-hardware/eyepay-phone/. (Accessed 30 March 2026).
[5] Apple optic id, 2026, https://support.apple.com/en-us/118483. (Accessed 30 March 2026).
[6] F. Boutros, et al., Iris and periocular biometrics for head mounted displays: Segmentation, recognition, and synthetic data generation, Image Vis. Comput. 104 (2020) 104007.
[7] G. Kang, J. Koo, Y. Kim, Security and privacy requirements for the metaverse: A metaverse applications perspective, IEEE Commun. Mag. 62 (1) (2024) 148–154.
[8] Orb, 2026, https://www.iso.org/standard/54066.html. (Accessed 30 March 2026).
[9] R. Albano, F. Battaglia, E. Maiorana, G. Campobello, P. Campisi, Effects of compression on attention-based iris presentation attack detection, in: 33rd European Signal Processing Conference, EUSIPCO, 2025.
[10] R. Albano, F. Battaglia, A. Gnutti, E. Maiorana, F. Guerrini, G. Campobello, P. Migliorati, P. Campisi, Transformers for iris presentation attack detection: Effectiveness and behavior under image compression, in: 24th International Conference of the Biometrics Special Interest Group, BIOSIG, 2025.

[11] G. Wallace, The JPEG still picture compression standard, IEEE Trans. Consum. Electron. 38 (1) (1992) xviii–xxxiv.

[12] E. Alshina, J. Ascenso, T. Ebrahimi, JPEG AI: The first international standard for image coding based on an end-to-end learning-based approach, IEEE MultiMedia 31 (4) (2024) 60–69.

[13] A. Czajka, K.W. Bowyer, Presentation attack detection for iris recognition: An assessment of the state-of-the-art, ACM Comput. Surv. 51 (4) (2018) 1–35.

[14] V. Ruiz-Albacete, et al., Direct attacks using fake images in iris verification, in: Biometrics and Identity Management, BioID, 2008.

[15] D. Yadav, N. Kohli, M. Vatsa, R. Singh, A. Noore, Detecting textured contact lens in uncontrolled environment using densepad, in: IEEE/CVF Conference on Computer Vision and Pattern Recognition Workshops, CVPRW, 2019.

[16] K.B. Raja, R. Raghavendra, C. Busch, Color adaptive quantized patterns for presentation attack detection in ocular biometric systems, in: 9th International Conference on Security of Information and Networks, SINCONF, 2016.

[17] J. Zuo, N.A. Schmid, X. Chen, On generation and analysis of synthetic iris images, IEEE Trans. Inf. Forensics Secur. 2 (1) (2007) 77–90.

[18] M. Trokielewicz, A. Czajka, P. Maciejewicz, Presentation attack detection for cadaver iris, in: IEEE 9th International Conference on Biometrics Theory, Applications and Systems, BTAS, 2018.

[19] R. Mahmood, I. Ahmed, Performance analysis of textured contact lens iris detection based on manual feature engineering, in: International Conference on Reliable Information and Communication Technology, IRICT, 2023.

[20] L. Pereira, et al., The rise of data-driven models in presentation attack detection, in: R. Jiang, C.-T. Li, D. Crookes, W. Meng, C. Rosenberger (Eds.), Deep Biometrics, Springer International Publishing, 2020.

[21] D. Yambay, et al., Review of iris presentation attack detection competitions, in: S. Marcel, J. Fierrez, N. Evans (Eds.), Handbook of Biometric Anti-Spoofing: Presentation Attack Detection and Vulnerability Assessment, Springer Nature, 2023.

[22] K. Nguyen, H. Proenca, F. Alonso-Fernandez, Deep learning for iris recognition: A survey, ACM Comput. Surv. 56 (9) (2024) 1–35.

[23] J. An, J. Inwhee, Attention map-guided visual explanations for deep neural networks, Appl. Sci. 12 (8) (2022).

[24] C. Chen, A. Ross, An explainable attention-guided iris presentation attack detector, in: IEEE Winter Conference on Applications of Computer Vision Workshops, WACVW, 2021.

[25] M. Fang, N. Damer, F. Boutros, F. Kirchbuchner, A. Kuijper, Iris presentation attack detection by attention-based and deep pixel-wise binary supervision network, in: IEEE International Joint Conference on Biometrics, IJCB, 2021.

[26] V.S. Swarup, D. Sadhya, V. Patel, K. De, Presentation attack detection in iris recognition through convolution block attention module, in: IEEE International Joint Conference on Biometrics, IJCB, 2022.

[27] A. Vaswani, N. Shazeer, N. Parmar, J. Uszkoreit, L. Jones, A.N. Gomez, L. Kaiser, I. Polosukhin, Attention is all you need, in: International Conference on Neural Information Processing Systems, 2017.

[28] A. Dosovitskiy, et al., An image is worth 16x16 words: Transformers for image recognition at scale, 2021, arXiv.

[29] Z. Liu, Y. Lin, Y. Cao, H. Hu, Y. Wei, Z. Zhang, S. Lin, B. Guo, Swin transformer: Hierarchical vision transformer using shifted windows, in: IEEE/CVF International Conference on Computer Vision, ICCV, 2021.

[30] A. Boyd, Z. Fang, A. Czajka, K.W. Bowyer, Iris presentation attack detection: Where are we now? Pattern Recognit. Lett. 138 (2020) 483–489.

[31] Z. Fang, A. Czajka, K.W. Bowyer, Robust iris presentation attack detection fusing 2D and 3D information, IEEE Trans. Inf. Forensics Secur. 16 (2021) 510–520.

[32] A. Kuehlkamp, A. Pinto, A. Rocha, K.W. Bowyer, A. Czajka, Ensemble of multi-view learning classifiers for cross-domain iris presentation attack detection, IEEE Trans. Inf. Forensics Secur. 14 (6) (2019) 1419–1431.

[33] A. Agarwal, A. Noore, M. Vatsa, R. Singh, Generalized contact lens iris presentation attack detection, IEEE Trans. Biom. Behav. Identity Sci. 4 (3) (2022) 373–385.

[34] M.R. Dronky, W. Khalifa, M. Roushdy, Impact of segmentation on iris liveness detection, in: 14th International Conference on Computer Engineering and Systems, ICCES, 2019.

[35] M. Fang, N. Damer, F. Kirchbuchner, A. Kuijper, Demographic bias in presentation attack detection of iris recognition systems, in: 28th European Signal Processing Conference, EUSIPCO, 2021.

[36] J. Tapia, C. Arellano, Gender classification from iris texture images using a new set of binary statistical image features, in: International Conference on Biometrics, ICB, 2019.

[37] K. Han, et al., A survey on vision transformer, IEEE Trans. Pattern Anal. Mach. Intell. 45 (1) (2022) 87–110.

[38] K. Santosh, C. Wall, Trustworthy and explainable AI for biometrics, in: AI, Ethical Issues and Explainability – Applied Biometrics, 2022.

[39] E. Caldeira, P.C. Neto, M. Huber, N. Damer, A.F. Sequeira, Model compression techniques in biometrics applications: A survey, Inf. Fusion 114 (2025) 102657.

[40] B. Stojanovic, A. Neskovic, Impact of PCA based fingerprint compression on matching performance, in: Telecommunications Forum, TELFOR, 2012.

[41] T. Schlett, S. Schachner, C. Rathgeb, J. Tapia, C. Busch, Effect of lossy compression algorithms on face image quality and recognition, in: IEEE International Conference on Acoustics, Speech and Signal Processing, ICASSP, 2023.

[42] I. Shaikh, S. Mukane, Iris image compression using JPEG & its effect on recognition performance, Int. J. Sci. Eng. Res. 6 (2) (2015).

[43] F. Battaglia, G. Gugliandolo, G. Campobello, N. Donato, EEG-over-BLE: A novel low-power architecture for multi-channel eeg monitoring systems, in: IEEE International Symposium on Measurements & Networking, M&N, 2022.

[44] R. Ives, B. Bonney, D. Etter, Effect of image compression on iris recognition, in: IEEE Instrumentationand Measurement Technology Conference, IMTC, 2005.

[45] R. Ives, D. Bishop, Y. Du, B. Craig, Iris recognition: The consequences of image compression, EURASIP J. Adv. Signal Process. 2010 (680845) (2010) 1–9.

[46] J. Daugman, C. Downing, Effect of severe image compression on iris recognition performance, IEEE Trans. Inf. Forensics Secur. 3 (1) (2008) 52–61.

[47] H. Hofbauer, C. Rathgeb, J. Wagner, A. Uhl, C. Busch, Investigation of better portable graphics compression for iris biometric recognition, in: International Conference of the Biometrics Special Interest Group, BIOSIG, 2015.

[48] G.A.S. Martinez, R.E.B. Moralde, N.B. Linsangan, R.M.L. Ang, A comparative analysis between the performance of the extracted features of JPEG and PNG on a raspberry pi iris recognition system, in: IEEE Region 10 International Conference, TENCON, 2023.

[49] A. Alkhateeb, A. Gnutti, F. Guerrini, R. Leonardi, J. Ascenso, F. Pereira, JPEG AI compressed domain face detection, in: IEEE 26th International Workshop on Multimedia Signal Processing, MMSP, 2024.

[50] E. Jalilian, H. Hofbauer, A. Uhl, Iris image compression using deep convolutional neural networks, Sensors 22 (7) (2022).

[51] J.S. Doyle, K.W. Bowyer, Robust detection of textured contact lenses in iris recognition using BSIF, IEEE Access 3 (2015) 1672–1683.

[52] D. Yambay, et al., Livdet iris 2017 - iris liveness detection competition 2017, in: IEEE International Joint Conference on Biometrics, IJCB, 2017.

[53] R. Sharma, A. Ross, D-NetPAD: An explainable and interpretable iris presentation attack detector, in: IEEE International Joint Conference on Biometrics, IJCB, 2020.

[54] G. Jaswal, A. Verma, S.D. Roy, R. Ramachandra, Learning joint local-global iris representations via spatial calibration for generalized presentation attack detection, IEEE Trans. Biom. Behav. Identity Sci. 6 (2) (2024) 195–208.

[55] Y. Li, Y. Lian, J. Wang, Y. Chen, C. Wang, S. Pu, Few-shot one-class domain adaptation based on frequency for iris presentation attack detection, in: IEEE International Conference on Acoustics, Speech and Signal Processing, ICASSP, 2022.

[56] D. Pal, R. Sony, A. Ross, A parametric approach to adversarial augmentation for cross-domain iris presentation attack detection, in: IEEE/CVF Winter Conference on Applications of Computer Vision, WACV, 2025.

[57] JPEG AI, reference software for Rec. ITU-T T.840.1 | ISO/IEC 6048-1, 2024–2025, 2026, https://gitlab.com/wg1/jpeg-ai/jpeg-ai-reference-software. (Accessed 30 March 2026).

[58] G. Campobello, G. Gugliandolo, N. Donato, A simple and efficient near-lossless compression algorithm for multichannel EEG systems, in: European Signal Processing Conference, 2021.

[59] BIQTIris open source library, 2026, https://www.iso.org/standard/54066.html. (Accessed 30 March 2026).

[60] ISO/IEC 29794-6:2015, 2026, https://www.iso.org/standard/54066.html. (Accessed 30 March 2026).

[61] N.G. Venkataswamy, Y. Liu, S. Singh, S. Dey, S. Schuckers, M.H. Imtiaz, Smartphone-based iris recognition through high-quality visible spectrum iris capture, 2024, arXiv.

[62] D. He, Z. Yang, W. Peng, R. Ma, H. Qin, Y. Wang, Elic: Efficient learned image compression with unevenly grouped space-channel contextual adaptive coding, in: IEEE/CVF Conference on Computer Vision and Pattern Recognition, CVPR, 2022.