frontiers | Frontiers in Imaging

TYPE Original Research
PUBLISHED 18 December 2024
DOI 10.3389/fimag.2024.1478783

OPEN ACCESS

EDITED BY
Christian Kraetzer,
Otto-von-Guericke University, Germany

REVIEWED BY
Luis Arturo Soriano,
Chapingo Autonomous University, Mexico
A. Ryad Soobhany,
Heriot-Watt University Dubai, United Arab Emirates

*CORRESPONDENCE
Juan E. Tapia
✉ juan.tapia-farias@h-da.de

RECEIVED 10 August 2024
ACCEPTED 26 November 2024
PUBLISHED 18 December 2024

CITATION
Valenzuela A, Tapia JE, Chang V and Busch C (2024) Presentation Attack Detection using iris periocular visual spectrum images. *Front. Imaging* 3:1478783. doi: 10.3389/fimag.2024.1478783

Presentation Attack Detection using iris periocular visual spectrum images

Andrés Valenzuela[1], Juan E. Tapia[2]*, Violeta Chang[1] and Christoph Busch[2]

[1]Departamento de Ingeniería Informática, Universidad de Santiago de Chile, Santiago, Chile, [2]da/sec-Biometrics and Internet Security Research Group, Hochschule Darmstadt, Darmstadt, Germany

In this work, we analyse the comparison between using the periocular area instead of the full face area for Presentation Attack Detection (PAD) in the visual spectrum (RGB). The analysis was carried out by evaluating the performance of five Convolutional Neural Networks (CNN) using both facial and periocular iris images for PAD with two different attack instruments. Additionally, we improved the CNN results by integrating the ArcFace loss function instead of the traditional categorical cross-entropy loss, highlighting that the ArcFace function enhances the performance of the models for both regions of interest, facial and iris periocular areas. We conducted Binary and Multiclass comparisons, followed by cross-database validation to assess the generalization capabilities of the trained models. Our study also addresses some of the current challenges in PAD research, such as the limited availability of high-quality face datasets in the desired spectrum (RGB), which impacts the quality of Presentation Attack Instruments (PAI) examples used in training and evaluation. Our goal was to address the challenge of detecting Iris periocular presentation attacks by leveraging the ArcFace function. The results demonstrate the effectiveness of our approach and provide valuable insights for improving PAD systems using periocular areas in the visual spectrum.

KEYWORDS

biometrics, Presentation Attack Detection, face, iris, periocular

1 Introduction

Biometric systems for personal authentication have gained significant attention in recent years due to the need for secure identification and access control. However, the vulnerability of these systems to detect presentation attacks, in which an attacker tries to bypass the system by presenting fake biometric traits, poses a significant threat to their security. Facial and periocular recognition systems (Minaee and Abdolrashidi, 2019; Hu et al., 2015; Tapia et al., 2022) have been widely used modalities due to their non-intrusive nature and ease of use. However, these systems are also susceptible to presentation attacks, such as printed photos, video replay, contact lenses, and masks. The development of effective PAD methods (Ramachandra and Busch, 2017) has thus become crucial to ensure the security and reliability of biometric authentication systems.

Nowadays, PAD is a very active research area. Several databases are constituted in the state-of-the-art using images extracted from videos (Zhang et al., 2012; Chingovska et al., 2012; Wen et al., 2015). One of the main challenges identified is that many databases present a low-quality, small image size and do not represent an operational scenario in an actual remote biometric system. Currently, the images are captured from smartphones

with high-quality and higher resolutions. This previous condition allows exploring other face areas to detect fake images, such as periocular iris images.

One of the challenges in developing PAD methods is the limited availability of high-quality face datasets in the desired spectrum (NIR or RGB) for research, which in turn affects the quality of PAI examples used for training and evaluation. In addition, deep learning models have been widely used for PAD, with several networks being employed, such as MobileNet (Sandler et al., 2018; Howard et al., 2019), DenseNet (Huang et al., 2017), and EfficientNet (Tan and Le, 2019).

The choice of network architecture and its hyperparameters can significantly impact the performance of the PAD system. Furthermore, the selection of the loss function in training deep learning models also plays an important role in the performance of the PAD system. CNN commonly uses Categorical Cross-Entropy (CCE) loss and has been shown to be effective in many deep-learning applications. Today, the ArcFace (Deng et al., 2019) loss has been shown to improve the performance of face recognition tasks. ArcFace is a margin-based penalty that enhances the discriminative power of the learned features and may have potential applications in the PAD domain.

In this article, we provide a comprehensive assessment for detecting presentation attacks using the face and periocular iris in the visual spectrum, focusing on the instruments employed for these attacks and the challenges posed by the limited availability of high-resolution bona fide and attack images. We also explored various CNNs proposed, including loss functions for PAD, and compared their performance. The metrics used for evaluation will follow the definitions of the ISO 30.107-3 standard (ISO/IEC JTC 1/SC 37 Biometrics, 2021).

The main contributions of this work can be summarized as follows:

- A comprehensive comparison between face and periocular iris images in the visual spectrum was reported.
- A benchmark of five different deep learning-based network architectures for PAD was performed.
- An assessment of the effectiveness of two different loss functions, CCE and ArcFace loss, in PAD models was reported.
- An analysis of the challenges and potential solutions for improving the performance of PAD methods is proposed.
- A new iris periocular dataset in the visual spectrum was presented and will be available for further research (upon acceptance).

By highlighting the strengths and limitations of existing methods and discussing potential solutions and future research directions, we aim to advance the field of biometric security further and ensure the reliability of PAD methods in real-world scenarios.

The remaining sections of this article are structured as follows: Section 2 summarizes the relevant studies on PAD. Section 3, elaborates on the description of the database used. Section 4, explains the metrics employed. The methodology is presented in Section 5, followed by the experiment and results in Section 6. Finally, Section 9 describes the conclusions.

2 Related work

Numerous PAD systems have been introduced in the literature for face and iris, as the utilization of these systems has grown in different applications in many biometric modalities, leading to a higher risk of attacks on these systems due to their sensitivity (Czajka and Bowyer, 2018; Tolosana et al., 2019; Tapia et al., 2021; Dhar et al., 2022). However, only a few of them are focussing on the iris periocular area using the visual spectrum for PAD.

2.1 Face PAD

Pasmino et al. (2023) addressed the need for improved PAD databases by introducing a new database called "F-PAD". Existing databases often suffer from low-quality, small-sized images that do not accurately represent real-world scenarios, such as remote biometric systems. In contrast, the F-PAD database is based on high-quality images sourced from the Face-HQ Dataset, offering a more comprehensive range of image quality and resolution. The database consists of 3,000 bona fide face images and 11,000 attack images. The bona fide images were carefully selected, focusing on portrait and selfie-like photos with evident facial biometric characteristics such as open eyes and a full mouth, while the attack images were created by dividing the bona fide images into three groups of Presentation Attack Instruments (PAIs) such as paper matte, glossy, and bond. Several devices, including screens from laptops, smartphones (e.g., iPhone-XI, LG, Huawei), and tablets (e.g., iPad, Microsoft Surface), were used to capture the attack images. Three deep learning models were implemented to evaluate PAD performance: MobileNet-V3 (small and large) and EfficientNet-B0. These models were trained and evaluated using the F-PAD database as well as four state-of-the-art datasets. All the evaluations were performed using full-face images.

Yu et al. (2020) proposed an algorithm based on a frame-level face anti-spoofing method with a Central Difference Convolution (CDC), which can capture intrinsic detailed patterns via aggregating both intensity and gradient information. Furthermore, over a specifically designed CDC search space, Neural Architecture Search (NAS) is utilized to discover a more robust network structure (CDCN++), which can be assembled with a Multiscale Attention Fusion Module (MAFM) for further boosting performance. The author evaluated their model on CASIA-MFSD (Zhang et al., 2012), MSU-MFSD (Wen et al., 2015), and Replay-Attack (Chingovska et al., 2012) datasets using full-face images in different resolutions.

Fang et al. (2023) proposed the Competition on Face Presentation Attack Detection Based on Privacy-aware Synthetic Training Data (SynFacePAD 2023) held at the 2023 International Joint Conference on Biometrics (IJCB 2023). The solutions were evaluated on four publicly available authentic face PAD benchmarks. The competition showcased various innovative approaches, resulting in improved performance compared to the baseline methods. The Solutions using transformer-based architecture as the base network generally exhibited higher

PA detectability compared to CNNs. All the evaluations were performed using full-face images.

Gonzalez-Soler et al. (2023) explored the utility of using different facial regions for PAD. In this context, a new metric, Face Region Utility, was proposed, which indicates the usefulness of a particular test region in spotting an attack attempt based on another training region. The left and the right eyes are explored separately in the visual spectrum. The full face was identified as the most helpful part compared with the face's left and right sides and other different areas. The central region of the faces could outperform the results achieved by the full face on a masked database.

2.2 Periocular iris PAD

Upon reviewing the state-of-the-art, it stands out that the vast majority of works report results using databases in the Near-InfraRed spectrum (NIR), while only a few (Pasmino et al., 2023) in the visual spectrum (RGB) use databases of intermediate or low quality. This fact highlights the need to explore and improve the capability of PAD models to operate in a high-quality visual spectrum. Additionally, only results using faces (Pasmino et al., 2023; Gonzalez-Soler et al., 2023) and monocular zones (Tapia et al., 2021, 2022) has been reported.

Dhar et al. (2022) introduced a multitask dual system called EyePAD++, which performs eye authentication and PAD using periocular images in the Near InfraRed spectrum based on ND-LivDet (2013-2017) databases (Yambay et al., 2017). This work proposes a whole system that employs a teacher-student framework with Multitask Learning Networks, where the teacher network is trained only for Eye Authentication (EA) and the student network is specialized in detecting Presentation Attacks (PA). The EyePAD++ system demonstrates effectiveness in combining both tasks (Face and iris), but it relies on datasets captured in the NIR spectrum, which are primarily in low resolution. This may present challenges when detecting fine-grained details that are critical for high-precision PAD systems. Although the approach works well in controlled environments, the dependence on low-resolution datasets limits the performance of the system in real-world scenarios where images with higher resolution may be required to ensure a high-quality Presentation Attack Detection.

Hoffman et al. (2019) proposed an iris plus ocular PAD using Multiple CNNs. This work extracted multiple patches from different eye positions. Three different solutions were proposed based on the region that is input to the CNN. The first solution, which they call the Iris CNN (I-CNN), looks primarily at the iris region. The second solution, called the Full image CNN (F-CNN), looks at the full ocular image, whereas the third solution, called Sampled Ocular CNN (S-CNN), looks at a subset of patches sampled from the ocular region. All the images are also in the NIR spectrum.

In our previous work (Tapia et al., 2021), we also explored the influence of iris periocular images for selfie-biometric using images in the visual spectrum captured in "the wild condition" based on a smartphone App and applying a Super-Resolution algorithm. The iris images were captured in a selfie mode with three different distances based on arm extension (Tapia et al., 2019). No PAD exploration was reported.

Motivated by challenges previously identified in PAD methods for both the NIR and VIS spectrum, we propose a novel approach focused on the visual spectrum (RGB) using high-resolution images and diverse presentation attack instruments (PAIs). Our method compares the performance of PAD systems when using the full face versus the iris periocular area. Through the use of different deep learning models, we demonstrate that training with these high-resolution bona fide and attack images significantly increases the discriminative power of the models. In alignment with recent advancements in PAD research (Yu et al., 2022), we conducted cross-validation to check the generalizability of our models, showing that the performance drops when tested on previously unseen databases. In this way, we highlight the well-known challenge of dataset variability in PAD, emphasizing the need for robust models that generalize well across different conditions. Our work builds upon the existing state of the art by exploring alternative regions of interest, integrating different loss functions, and using high-resolution RGB datasets that contribute to the ongoing development of more reliable PAD systems in the visual spectrum.

3 Databases

Two different datasets that accomplish our requirement of high-resolution were used for this work: F-PAD (Pasmino et al., 2023), and we also created from scratch a Private PAD dataset (P-PAD), which was developed with the support of a Biometric company only for research purposes and will be available for other researchers. In line with the problem identified by Pasmino et al. (2023) referring to the low availability of high-resolution datasets, we focused on using high-quality image datasets (F-PAD and P-PAD) to improve the generalizability of PAD models across different attack scenarios. Also, it has been highlighted in the state-of-the-art (Yu et al., 2022) that PAD methods obtain good results in intra-dataset scenarios, which means that test images were created under the exact conditions as the training set. However, these models tend to decrease their performance noticeably when inferring from other databases on cross-dataset scenarios or unknown attacks with attacks not previously considered. It is essential to highlight that most of the datasets in the literature present low-resolution images.

Therefore, high-resolution images of $1,280 \times 720$ pixels were needed and used for this study, and images for the print and screen classes with a resolution of $5,120 \times 3,840$. These images were then cropped to the periocular regions and resized to the input size of each CNN to be trained.

It is essential to consider that an image's quality and level of detail are closely related to its resolution. An image with a higher resolution will have more pixels, which translates to more outstanding sharpness and detail. In this context, the original high-resolution images were used to generate printed and screen images manually.

For the first scenario, the original images were printed on glossy paper and photographed with a high-end mobile device. Different collaborators captured the original images from different computer

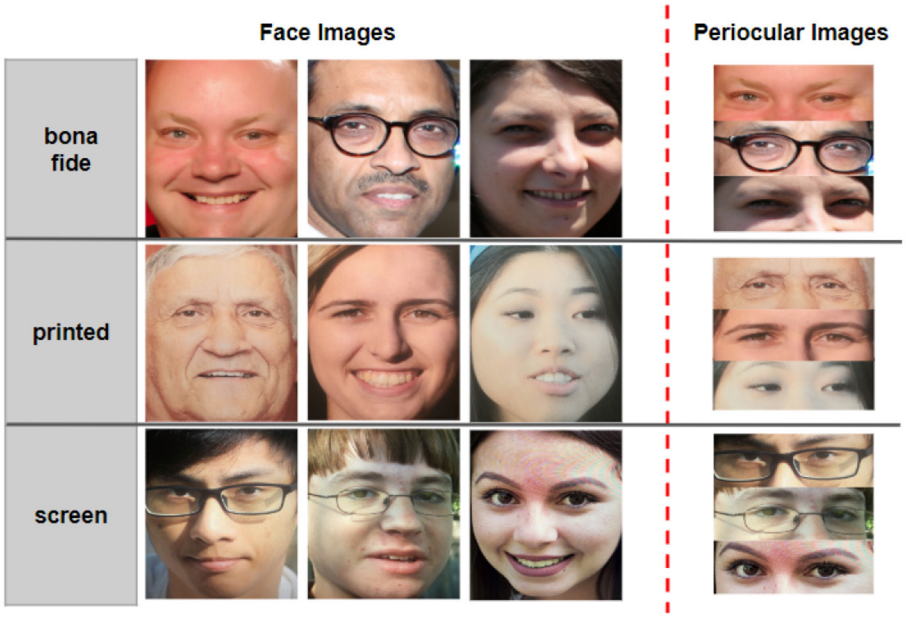

FIGURE 1
Examples of facial and periocular images with corresponding printed and screen classes.

TABLE 1 Database summary for bona fide and attack.

Classes	F-PAD	P-PAD	Total
Databases			
Bona fide	3,000	5,200	8,200
Printed	6,000	10,000	16,000
Screen	5,000	9,998	14,998
Total	14,000	25,198	39,198

TABLE 2 Summary of data sampling for F-PAD database.

Classes	Train 60%	Test 20%	Val 20%	Total
F-PAD binary splits				
Bona fide	1,783	606	604	2,993
Attack	6,585	2,189	2,194	10,968
Total	8,368	2,795	2,798	13,961
F-PAD multiclass splits				
Bona fide	1,783	606	604	2,993
Printed	3,590	1,193	1,196	5,979
Screen	2,995	996	998	4,989
Total	8,368	2,795	2,798	13,961

screens for the screen image scenario. The database provided High-Definition (HD) images of 1280 × 720 pixels for capturing selfie images and creating presentation attacks with high-end mobile devices by four collaborators tasked with manual work. As a result of using these high-end devices, the generated images have four times the resolution of the original ones, giving them a much higher level of detail and sharpness. The F-PAD and the P-PAD periocular test set datasets will be available only by request for research purposes. Figure 1 shows examples of images.

Table 1 shows the number of images per class for both mentioned databases.

Furthermore, the datasets were split into training, validation, and test sets while maintaining a balanced distribution of 60%, 20%, 20% for each class, as illustrated in Tables 2, 3. Alongside the aforementioned process, three text files were generated to retain a record of the corresponding image lists for each dataset split, ensuring consistency throughout the experiments and results. These files will also become publicly available to ensure the reproducibility of this work.

3.1 Preprocessing

All the selfie images were preprocessed, and the face was detected and cropped using the MTCNN face detector (Zhang et al., 2016). Each image's key points of the regions of interest were used to crop the periocular area. In the case of MTCNN, the detected key points are those of the eyes, nose, mouth corners, and face location.

Further, the MTCNN network only delivers key points of the mentioned face parts but not of the desired element (periocular area). To crop this area, the Euclidean distance between the pairs of points (eyes–nose) and (eyes–face upper limit) was calculated to obtain the midpoint of the periocular area and thus crop it.

Another challenge that arose while creating the databases was the presence of images with too large dimensions, which led to a significant increase in the MTCNN network's inference time. To improve that, the images were reduced by a factor of 5, and the face

TABLE 3 Summary of data sampling for P-PAD database.

Classes	Train 60%	Test 20%	Val 20%	Total
P-PAD binary splits				
Bona fide	3,598	1,218	1,200	6,016
Attack	11,360	3,768	3,787	18,915
Total	14,958	4,986	4,987	24,931
P-PAD multiclass splits				
Bona fide	3,598	1,218	1,200	6,016
Printed	5,986	1,997	1,996	9,979
Screen	5,374	1,771	1,791	8,936
Total	14,958	4,986	4,987	24,931

and landmark were inferred from these small images. Afterwards, the key points were interpolated or scaled to the original image. This way, all the necessary information for detection could be used without affecting the image resolution.

Despite the solutions applied, the challenge of undetected face images arose, where the MTCNN detector failed to detect regions of interest in some images. Thus, it is essential to mention that the number of examples was reduced from 14,000 to 13,961 for the F-PAD database and from 25,198 to 24,931 for the P-PAD database.

Figure 2 visually represents the face detection steps, while Figure 1 shows a compilation of the resulting images using the facial and periocular detection method previously described.

4 Metrics

The ISO/IEC 30.107-3[1] standard provides guidelines for evaluating the performance of PAD algorithms in biometric systems. The Attack Presentation Classification Error Rate (APCER) metric measures the proportion of attack presentations that are incorrectly classified as bona fide presentations for each type of Presentation Attack Instrument (PAI). This metric is calculated separately for each PAI, and the worst-case scenario is considered. Equation 1 outlines how to compute the APCER metric. In this equation, the value of N_{PAIS} represents the number of attack presentation images, where RES_i for the ith image is 1 if the algorithm classifies it as a spoofed image, and 0 if it is classified as a bona fide presentation.

$$APCER = 1 - \left(\frac{1}{N_{PAIS}}\right) \sum_{i=1}^{N_{PAIS}} Res_i \tag{1}$$

In addition, the Bona fide Presentation Classification Error Rate (BPCER) metric evaluates the proportion of bona fide (live) presentations that are incorrectly classified as attacks to the biometric capture device or the ratio between false rejections and total bona fide attempts. The BPCER metric is calculated using Equation 2, where N_{BF} represents the number of bona fide (live) presentation images, and RES_i takes the same values as those used in the APCER metric.

$$BPCER = \left(\frac{1}{N_{BF}}\right) \sum_{i=1}^{N_{PAIS}} Res_i \tag{2}$$

The experiments included a Detection Error Trade-off (DET) curve, which shows the relationship between the false acceptance rate (APCER) and the false rejection rate (BPCER). The Equal Error Rate (EER) value is the point where the APCER and BPCER are equal. The results included two operational points based on the ISO/IEC 30.107 standard: BPCER10, which is the BPCER when the APCER is fixed at 10%, and BPCER20, which is the BPCER when the APCER is fixed at 5%.

5 Methodology

This section outlines the methodology employed to train CNNs for face and periocular presentation attack classification and the application of data augmentation techniques. Additionally, it describes the utilization of the ArcFace loss function for improved performance in PAD classification tasks. The experiments conducted involve the following key steps.

5.1 Training of CNN architectures

Four state-of-the-art methods were explored including MobileNet V2 (Sandler et al., 2018), MobileNet V3 (Howard et al., 2019) (small and large versions), EfficientNetB0 (Tan and Le, 2019), and DenseNet121 (Huang et al., 2017). These architectures were trained on the face and periocular databases separately to explore the optimal set of hyperparameters through grid search tuning. The experiments consisted of four distinct trials, outlined as follows:

- Exp 01: Facial images were used, and the CCE loss function was applied.
- Exp 02: Periocular images were used, and the CCE loss function was applied.
- Exp 03: Facial images were used, and the ArcFace loss function was applied.
- Exp 04: Periocular images were used, and the ArcFace loss function was applied.

Table 4 presents the hyperparameters selected based on the training process for the CNN architectures.

5.2 Data augmentation techniques

Data augmentation techniques were employed during training to enhance the model's robustness and generalization capabilities. Specifically, data augmentation was applied only to the training set, while the validation and test sets retained the original images. This step is crucial when training these CNN's models, particularly for tasks like Presentation Attack Detection (PAD).

1 https://www.iso.org/standard/79520.html

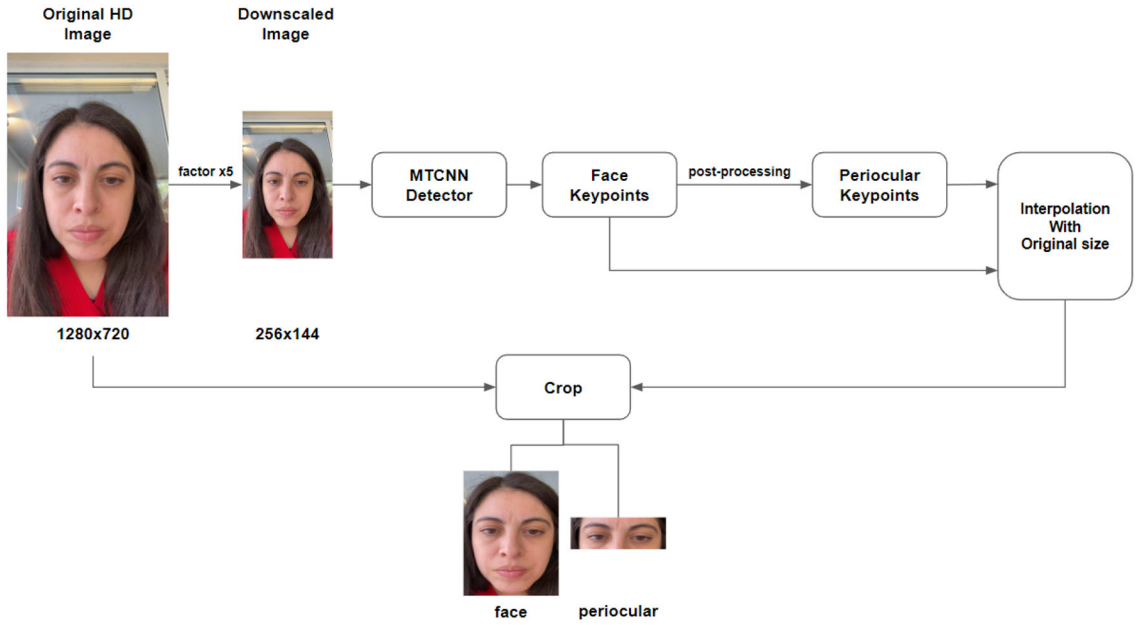

FIGURE 2
Face and periocular detection flow using MTCNN (Zhang et al., 2016).

TABLE 4 Summary of hyperparameters used for each model.

Test ID	Model arch	# params	Epochs	Batch size	Input size	Optimizer	LR
Hyperparameters							
01	MobileNet V2	2.260.546	200	16	224 × 224	RMSProp	10e-4
02	MobileNet V3 Small	940.274	200	16	224 × 224	Adam	10e-4
03	MobileNet V3 Large	2.999.232	200	16	224 × 224	RMSProp	10e-4
04	EfficientNet B0	4.052.133	200	16	224 × 224	SGD	10e-3
05	DenseNet121	7.039.554	200	16	224 × 224	SGD	10e-4

For both bona fide and presentation attack images, augmentation helped to improve the model's ability to generalize across severe conditions and attack types. The following table defines the different augmentators used, which include horizontal and vertical image flipping, Gaussian and median filter blurring, and brightness, contrast, and color adjustments. Additional transformations, such as perspective changes, were also applied. Table 5 describes the details of DA applied.

5.3 Categorical cross entropy loss (CCE)

A set of models was trained with the *CCE* loss function to assess performance improvement in PAD classification tasks. In the *CCE* function, the entropy of each class is calculated as the sum of the probability of the ground truth class multiplied by the logarithm of the predicted probability for that class. This dissimilarity shows the difference between the predicted class and the ground truth, as is depicted in Equations 3, 4:

$$-(y\log(p)+(1-y)\log(1-p)) \tag{3}$$

In the case of multiple classes, the function can be represented by the following equation:

$$-\sum_{c=1}^{M} y_{o,c}\log(p_{o,c}) \tag{4}$$

Where M corresponds to the number of classes, *log* is the natural logarithm, p is the probability of the predicted class, and y is the binary indicator (1 or 0) depending on whether the class was correctly classified or not.

5.4 ArcFace loss function

A set of models trained with the state-of-the-art ArcFace loss function was employed to assess performance improvement in PAD classification tasks. This loss function enhances the network's discrimination ability by learning relevant features in the embedding space. It is widely used in deep learning to improve the accuracy of data classification tasks. This loss function compares

TABLE 5 Summary parameters applied for DA.

Transformation	Description	Probability/range
Flipping	Randomly apply horizontal or vertical flipping to the image.	75% chance for each flip
Coarse dropout	Randomly removes parts of the image by dropping pixels.	25% chance
Brightness adjustments	Randomly adds or subtracts values from pixel intensity, changing brightnesfrom	Range: -30 to +30
Grayscale	Converts the image to grayscale by adjusting the color channels.	Alpha range: 0.0–1.0
Contrast adjustment	Modifies the contrast by increasing or decreasing it randomly.	Contrast range: 0.25–2.0
Noise addition	Adds Gaussian, Laplace, or Poisson noise to the image.	Noise intensity range: 0.01 * 255–0.1 * 255
Blurring	Applies different blur types: Gaussian, average, or motion blur.	Blur range: Sigma (0.01–2.5), Kernel size (1 to 5)
Gamma contrast	Adjusts the gamma contrast levels to enhance contrast.	Gamma range: 0.01–1.0

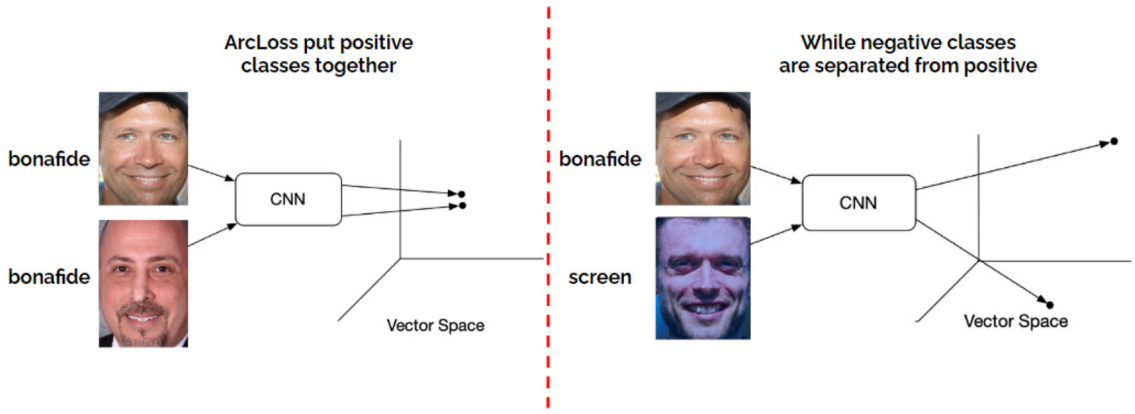

FIGURE 3
Visual representation of class separation in a vector space.

the input images with their respective projections and quantifies the amount of information lost during the projection process.

The models were trained using the designated databases, and their performance in PAD classification was evaluated following the ISO 30.107-3 standard.

The ArcFace function can be represented mathematically as:

$$L_{ArcFace}(x) = -\log \frac{e^{s(cos(\theta_{y_i}+m))}}{e^{s(cos(\theta_{y_i}+m))} + \sum_{j=1,j\neq y_i}^{N} e^{scos(\theta_j)}}$$

Where:

- x: is an input image representing a class.
- s: is a scaling factor used to control the magnitude of the loss function. A high value of s means that the loss will be larger and, therefore, the separation between classes will be more significant.
- m: an angular margin added to increase the separation between classes. This margin is used to force a clearer separation between classes, which in turn helps to reduce the probability of error in classification.
- θ_y: is the angle between the input images x and the desired output y. This angle represents the similarity between the images.
- θ_j: is the angle between the input image x and another image or class j other than the desired output. This angle represents the similarity between the input image and other classes.

The ArcFace function[2] combines these parameters to produce a measure of information loss when classifying images into different classes. The idea is to maximize the separation between the classes (positive and negative) and minimize the information loss to improve classification accuracy. In summary, the function ArcFace is a state-of-the-art loss function that is used to calculate and maximize the separation angle between two or more datasets (see Figure 3), in addition to applying a separation margin between their classes. In contrast, the classical CCE loss function only measures the discrepancy between the probability distributions of the classes.

6 Experiments and results

6.1 Experiment 01—Facial and categorical cross entropy

For this experiment, five different CNNs were trained to explore different optimiser and learning rate parameters. The parameter

2 https://github.com/yinguobing/arcface/blob/main/train.py

TABLE 6 Exp. 01. Results of models trained with face images and CCE.

Test ID	Model	EER (%)	APCER (%)	BPCER (%)	BPCER10 (%)	BPCER20 (%)
01	MobileNet V2	10.72	10.69	10.72	11.07	17.64
02	MobileNet V3 Small	26.47	26.43	26.47	46.88	61.24
03	**MobileNet V3 Large**	**8.82**	**8.79**	**8.82**	**7.09**	**15.74**
04	EfficientNetB0	23.52	23.50	23.52	48.44	65.57
05	DenseNet121	19.55	19.53	19.55	31.31	46.88

The best result is highlighted with bold text.

TABLE 7 Exp. 02. Results of models trained with periocular images and CCE.

Test ID	Model	EER (%)	APCER (%)	BPCER (%)	BPCER10 (%)	BPCER20 (%)
01	**MobileNet V2**	**2.42**	**2.39**	**2.42**	**0.86**	**1.21**
02	MobileNet V3 Small	18.85	18.85	18.85	33.56	45.67
03	MobileNet V3 Large	2.59	2.57	2.59	0.86	1.73
04	EfficientNetB0	26.12	26.11	26.12	42.90	60.38
05	DenseNet121	19.20	19.17	19.20	34.42	52.07

The best result is highlighted with bold text.

TABLE 8 Exp. 03. Results of models trained with facial images and ArcFace loss function.

Test ID	Model	EER (%)	APCER (%)	BPCER (%)	BPCER10 (%)	BPCER20 (%)
01	**MobileNet V2**	**11.41**	**11.41**	**11.41**	**11.59**	**21.45**
02	MobileNet V3 Small	17.54	17.54	17.64	50.51	90.86
03	MobileNet V3 Large	18.67	18.67	18.85	42.56	83.06
04	EfficientNetB0	14.87	14.83	14.87	16.78	24.74
05	DenseNet121	15.91	15.87	15.91	24.74	36.67

The best result is highlighted with bold text.

TABLE 9 Exp. 04. Results of models trained with periocular images and ArcFace function.

Test ID	Model	EER (%)	APCER (%)	BPCER (%)	BPCER10 (%)	BPCER20 (%)
01	**MobileNet V2**	**1.90**	**1.89**	**1.90**	**0.86**	**1.21**
02	MobileNet V3 Small	11.59	11.59	11.76	14.01	64.53
03	MobileNet V3 Large	1.21	1.17	1.21	1.21	1.21
04	EfficientNetB0	2.43	2.43	2.59	2.07	2.07
05	DenseNet121	8.47	8.43	8.47	8.13	12.80

The best result is highlighted with bold text.

numbers were also explored in order to look for a trade-off between a lower EER and a reduced number of parameters. Table 4 shows a summary of hyperparameters for each CNN trained and the input sizes, learning rate, epochs and optimisers used for each.

Table 6 lists the results obtained by each model for Experiment 1. In this experiment, a full-face image was used as input and using CCE. This table reported EER, APCER, BPCER, BPCER10 and BPCER20.

6.2 Experiment 02—Periocular and categorical cross entropy

In this experiment, the CNNs model training is similar to Experiment 01, but this time, it uses the periocular areas of the faces as an input to the network using CCE. The training parameters are reported in Tables 4, 7 lists the results obtained. This table reported EER, APCER, BPCER, BPCER10 and BPCER20.

6.3 Experiment 03—Facial and ArcFace

We propose replacing the CCE with the ArcFace function in the classification framework to improve the results. The ArcFace is a loss function that measures the information loss when data is projected into a lower-dimensional subspace. This experiment used a full-face image as input, and the training parameters are reported in Table 4.

Table 8 lists the results of each training, separated by model and using the ArcFace loss function. This table reported EER, APCER, BPCER, BPCER10 and BPCER20.

6.4 Experiment 04—Periocular and ArcFace

In experiment 04, the CNN models are trained using the periocular areas of the faces and the ArcFace loss function. The training parameters are reported in Table 4.

Table 9 lists the obtained results. This table reported EER, APCER, BPCER, BPCER10, and BPCER20.

6.5 Results

This section analyses and compares the results obtained in each experiment carried out. The MobileNetV3-Large and MobileNetV2 networks stand out above the others in terms of performance. Specifically, in Experiment 01 Section 6, the MobileNetV3-Large network achieves an EER of 8.82%, an APCER of 8.79%, and a BPCER of 8.82% when using the combination of Face and CCE loss function. Then, when switching the region of interest to the periocular area in Experiment 02 (Section 6.2), the MobileNetV2 network achieves better results, with an EER of 2.42%, an APCER of 2.39%, and a BPCER of 2.42% (see Tables 6, 7) which are better than the results reported by Pasmino et al. (2023). The improvement in performance between using periocular regions over face regions becomes more evident when we compare the improvements in the EER, APCER and BPCER metrics. All three metrics show a performance increase of 6.4% points in each of them. Thus, the models trained with periocular regions are observed to outperform those trained with full-face images.

Even so, when using the ArcFace loss function instead of CCE, the combination of the periocular region and ArcFace function proved to be competitive with the state of the art. In Experiment 04, the MobileNetV2 network achieved an EER of 1.90%, an APCER of 1.89%, and a BPCER of 1.90%, and the MobileNetV3-Large network achieved an EER of 1.21%, an APCER of 1.17%, and a BPCER of 1.21% (see Tables 8, 9). Both networks demonstrated a significant improvement when switching to the ArcFace loss function. Specifically, MobileNetV2 showed an average improvement of 0.5 percentage points across all metrics, while MobileNetV3-Large achieved a greater improvement with a 1.38 percentage points reduction in error rates across EER, APCER and BPCER (see Tables 6, 9).

These results indicate that the periocular region is more discriminative than the full-face images for Presentation Attack Detection (PAD), allowing the networks to distinguish between bona fide and attack samples more effectively. Moreover, when

TABLE 10 Results obtained from multiclass model evaluation for facial and periocular modalities.

Test set			
Facial		Periocular	
Threshold	0.21	Threshold	0.23
DEER	12.65 %	DEER	1.98 %
APCER(τ)	12.65 %	APCER(τ)	1.90 %
BPCER(τ)	12.70 %	BPCER(τ)	1.98 %
BPCER10	**15.18 %**	**BPCER10**	**1.15 %**
BPCER20	**26.56 %**	**BPCER20**	**1.32 %**

Bold indicates BPCER10, which is the BPCER when the APCER is fixed at 10%, and BPCER20, which is the BPCER when the APCER is fixed at 5%.

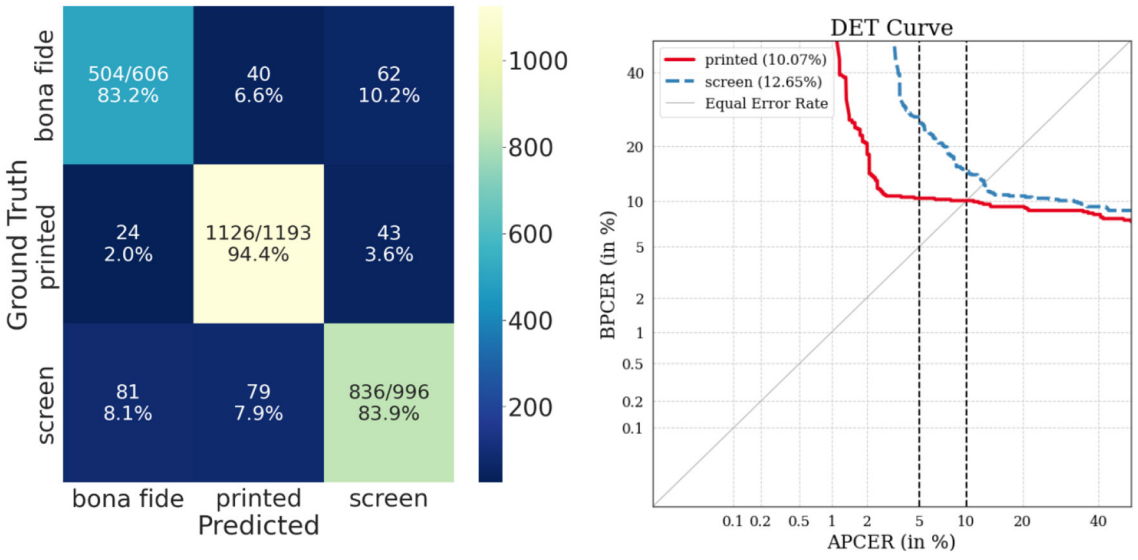

FIGURE 4
Confusion matrix **(left)** and DET curves **(right)** of multiclass evaluation *using facial modality*. This curve shows that the network's predictions have a high error rate, with the screen class performing the worst. In parentheses, EER is reported in percentages.

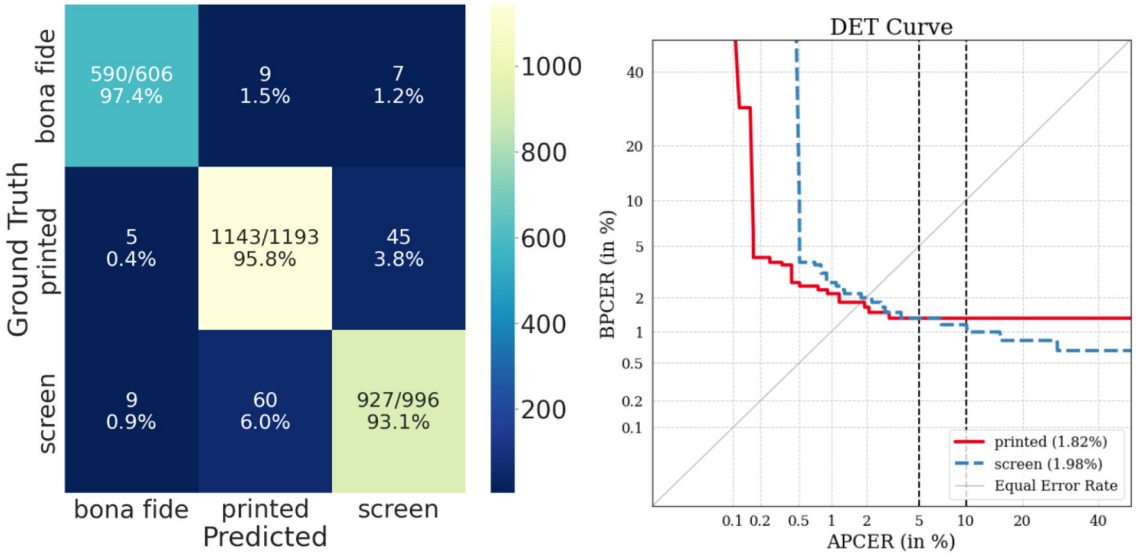

FIGURE 5
Confusion matrix **(left)** and DET curves **(right)** of multiclass evaluation *using periocular modality*. In this curve, both classes have a lower error rate than the facial modality, translating to a better classification of presentation attacks. In parentheses, EER is reported in percentages.

TABLE 11 Cross-validation results on the test dataset from the *P-PAD* database.

P-PAD test set			
Facial		Periocular	
Threshold	0.21	Threshold	0.23
DEER	24.71 %	DEER	10.58 %
APCER(τ)	24.67 %	APCER(τ)	3.10 %
BPCER(τ)	24.71 %	BPCER(τ)	17.40 %
BPCER10	**33.00 %**	**BPCER10**	**10.59 %**
BPCER20	**54.84 %**	**BPCER20**	**12.23 %**

Bold indicates BPCER10, which is the BPCER when the APCER is fixed at 10%, and BPCER20, which is the BPCER when the APCER is fixed at 5%.

complementing the training with a more advanced loss function such as ArcFace, both networks exhibited further improvements in all metrics. This highlights the importance of selecting the correct region of interest (periocular) and using a loss function that enhances feature discrimination, leading to superior performance in terms of EER, APCER, and BPCER.

This section analyses and compares the results obtained in each experiment carried out. The MobileNetV3-Large and MobileNetV2 networks stand out above the others in terms of performance. Specifically, the MobileNetV3-Large network achieves an EER of 8.82%, an APCER of 8.79%, and a BPCER of 8.82%. The MobileNetV2 network achieves an EER of 2.42%, an APCER of 2.39%, and a BPCER of 2.42% (see Tables 6, 7) which are better than the results reported by Pasmino et al. (2023).

Moreover, the models trained with periocular regions are observed to outperform those trained with full faces. On the other hand, when using the ArcFace function instead of CCE, the combination of the periocular region and ArcFace function proved to be competitive with the state of the art. The MobileNetV2 network achieved an EER of 1.90%, an APCER of 1.89%, and a BPCER of 1.90%. The MobileNetV3-Large network achieved an EER of 1.21%, an APCER of 1.17%, and a BPCER of 1.21% (see Tables 8, 9).

Although the results are satisfactory, it is important to consider that the complexity and limitations of evaluating RGB-PAD systems differ significantly from those associated with detecting attacks in NIR spectrum images.

7 Multi-class validation

To perform a more detailed analysis of the previous results, the class "attack" was divided into two subclasses: printed and screen. Then, the parameters of the best experiment (MobileNetV2, see Tables 8, 9) were replicated to train a multi-class classifier and compare the results between the facial and periocular modalities using the ArcFace loss function.

As stated above, Table 10 shows the metrics results obtained for each modality, while Figures 4, 5 show the confusion matrix and DET curves obtained from the multi-class evaluation for each modality. It can be observed from both Table 10 and both figures that the use of the periocular region again shows an improvement trend in the results.

The DET curve shown in Figure 4 indicates that the most challenging facial attack instrument to predict is the screen class with an EER of 12.65%. In contrast, the DET curve in Figure 5 indicates that switching to periocular modality substantially improves the results, achieving EER of 1.98% for the screen class. It is also worth noting that the confusion matrices reflect the performance improvement for the bona fide and screen classes when switching between facial and periocular modalities.

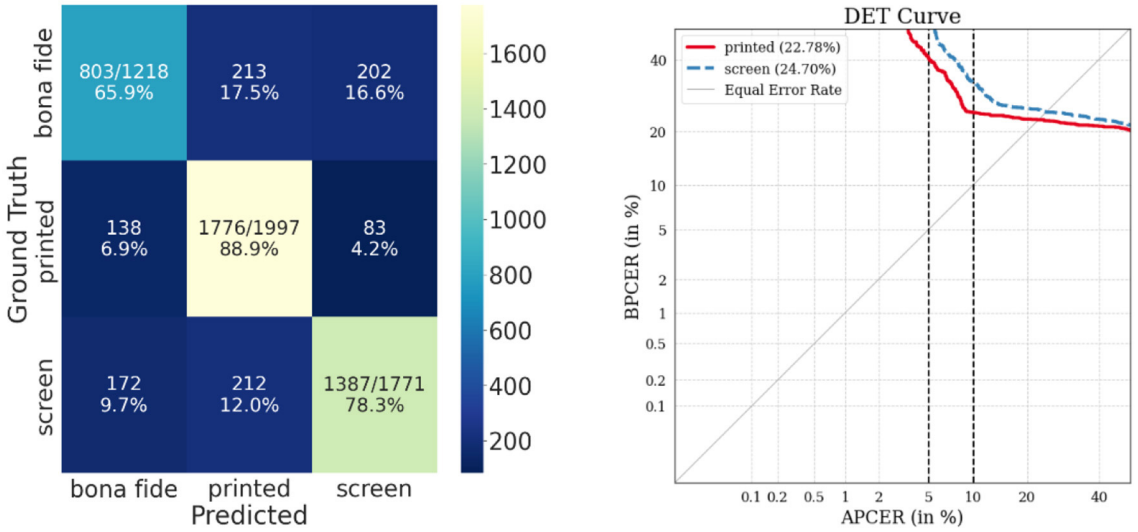

FIGURE 6
Confusion matrix **(left)** and DET curves **(right)** of multiclass cross dataset evaluation with *facial modality*. Both curves represent a high error rate, which is reflected in the number of errors made in the confusion matrix.

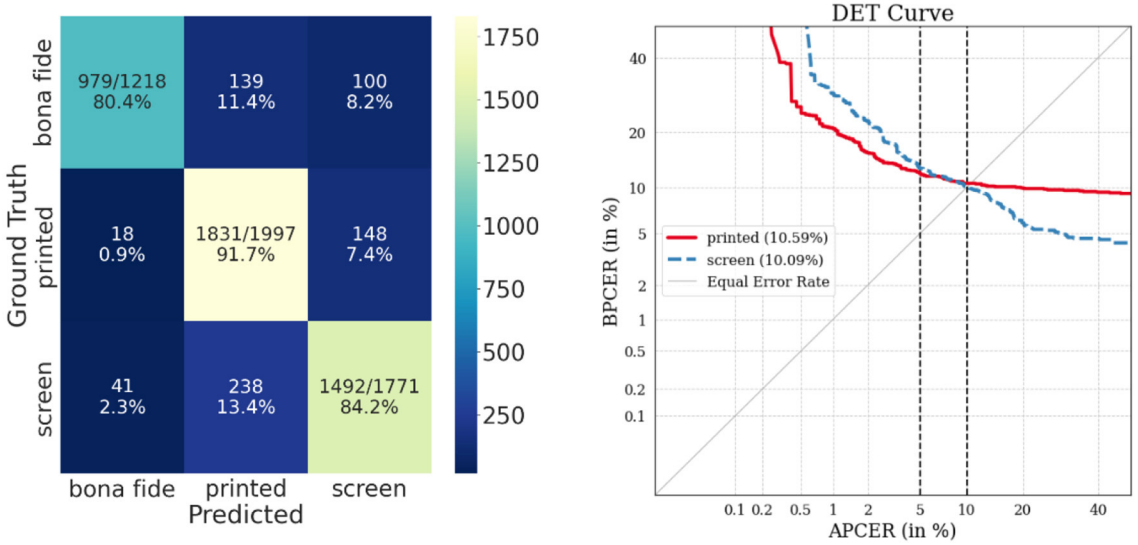

FIGURE 7
Confusion matrix **(left)** and DET curves **(right)** of multiclass cross dataset evaluation with *periocular modality*. Both curves represent a high error rate, but it is twice as low compared to the error rates obtained with the facial modality. In parentheses, EER is reported in percentages.

8 Cross dataset validation

To ensure comprehensive results, the best experiment's parameters (MobileNetV2, as shown in Tables 8, 9) were used throughout this section.

The training was performed on the *F-PAD* database, while the *P-PAD* database was utilized for cross-validation, as indicated in Tables 2, 3.

The results presented in Table 11 are aligned with the state of the art, exhibiting a decrease in performance for both facial and periocular modalities. The EER obtained were 24.71% for facial modality and 10.58% for periocular modality, both percentages being higher than those reported in Tables 8, 9.

Figures 6, 7 show the confusion matrix and DET curves obtained from the cross-dataset evaluation for each modality.

Then, final cross-validation is performed corresponding to training with the *P-PAD* database and evaluation with the *F-PAD* database (see Tables 2, 3). Table 12 shows the results of the final cross-validation evaluation.

It is worth noting that the results using the periocular area and the ArcFace function significantly improve due to the fact that the

P-PAD database has 10,000 more images than the *F-PAD* database. This is referenced in Tables 2, 3.

Table 12 presents the results of the cross-validation evaluation on the *F-PAD* database, showing a significant decrease in the performance of both methods. The EER of 38.95% and 8.88% were obtained for facial and periocular modalities, respectively.

Only facial results are higher compared to those obtained in Table 11, while periocular modality obtained a lower EER and BPCER but higher APCER.

Figures 8, 9 display the confusion matrix and DET curves obtained from the multiclass evaluation for each modality (facial/periocular), also showing graphically the model performance for each presentation attack class.

9 Conclusion

Previous studies on presentation attack detection (PAD) for the iris and face have primarily concentrated on specific presentation attack instruments (PAIs) using traditional state-of-the-art datasets. These datasets often reflect operational conditions characterized by low-resolution and controlled environments. To address this issue, we introduce the P-PAD dataset, as discussed in Section 3, which is captured in the visual spectrum (RGB). This dataset comprizes high-resolution images and a diverse range of PAIs, allowing for a more comprehensive evaluation of PAD systems under real-world conditions. This enhances the robustness of these systems against both known and unknown types of attacks. It's important to emphasize that the P-PAD database created in this study differs from existing state-of-the-art databases in terms of resolution, capture spectrum (utilizing visible light instead of infrared), and the inclusion of high-resolution presentation attacks. Most databases noted in the literature have lower resolution and/or a limited variety of presentation attacks.

In addition to the dataset, we explored the impact of two loss functions on the performance of PAD models and the use of the periocular area instead of the full-face image. Our results showed that the use of the ArcFace loss function, when combined with the periocular region, outperforms the BPCER score of the traditional Categorical Cross Entropy (CCE) loss function. This improvement highlights the importance of employing an advanced loss function to enhance the feature discrimination, particularly in regions of interest like the periocular area, which consistently demonstrated superior performance over full-face images (see Tables 6–9).

When the proposed models (see Table 4) for this study are trained using periocular images and the ArcFace function, better results were obtained compared to training with facial images and the same loss function. Additionally, through the conducted experimental search and the evaluation of the obtained results, it was identified that the MobileNetV2 network achieves the best performance obtaining an EER of 1.90%, an APCER of 1.89%, a BPCER of 1.90% and a $BPCER_{10}$ of 0.86% (see Table 9). These metrics highlight the network's effectiveness in accurately

TABLE 12 Cross-validation results on the test dataset from the *F-PAD* database.

F-PAD test set			
Facial		Periocular	
Threshold	0.13	Threshold	0.28
DEER	38.95 %	DEER	8.88 %
ACER(τ)	38.91 %	ACER(τ)	8.89 %
APCER(τ)	38.89 %	APCER(τ)	8.88 %
BPCER(τ)	38.94 %	BPCER(τ)	8.91 %
BPCER10	**41.41 %**	**BPCER10**	**8.58 %**
BPCER20	**41.58 %**	**BPCER20**	**9.40 %**

Bold indicates BPCER10, which is the BPCER when the APCER is fixed at 10%, and BPCER20, which is the BPCER when the APCER is fixed at 5%.

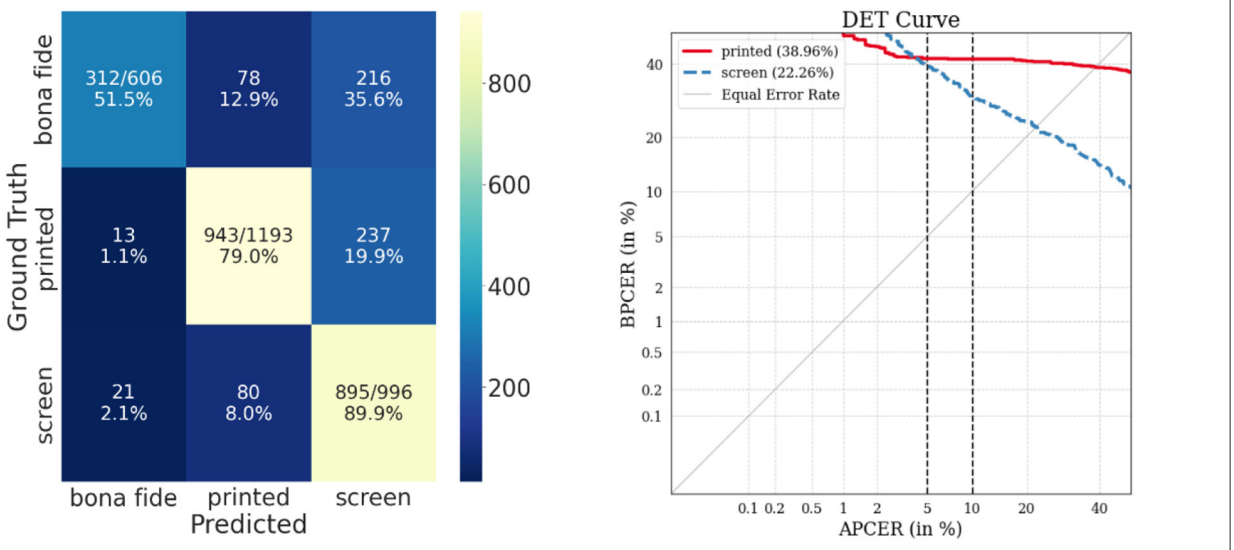

FIGURE 8
Confusion matrix **(left)** and DET curves **(right)** of multiclass cross dataset evaluation with facial modality. The curve for the "printed" class indicates that it is the attack instrument class with the higher error. In parentheses, EER is reported in percentages.

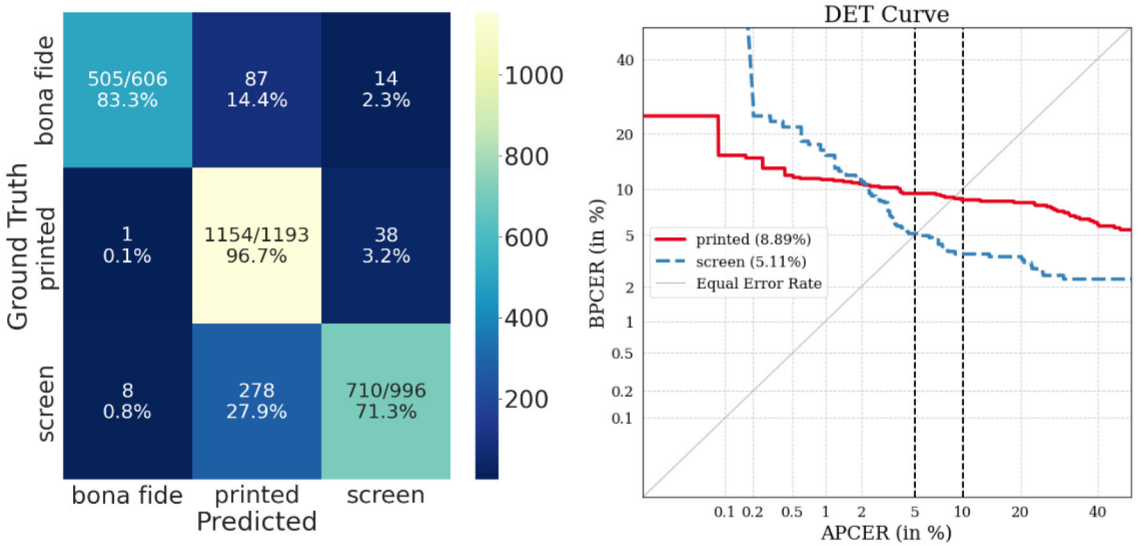

FIGURE 9
Confusion matrix **(left)** and DET curves **(right)** of multiclass cross dataset evaluation with periocular modality. Compared to the facial modality, both curves indicate a lower error rate. The "printed" class has a higher error percentage compared to the "screen" class. In parentheses, EER is reported in percentages.

classifying both bona fide and attack presentations, demonstrating its superior performance compared to other models evaluated in the study.

In conclusion, we placed significant emphasis on the cross-validation evaluations we conducted (see Tables 11, 12). These evaluations reveal the impact of different datasets on model performance, as previously discussed in the state-of-the-art (Yu et al., 2022). Our experiments demonstrate the effectiveness of focusing on the periocular region and show that the performance of Presentation Attack Detection (PAD) models can be further enhanced by using advanced loss functions like ArcFace. While Categorical Cross Entropy (CCE) and ArcFace are commonly employed, we propose exploring other loss functions in future work, such as Triplet Loss or Angular Margin Loss. These alternatives may improve feature separation and classification accuracy. Additionally, they could enhance the robustness of PAD systems by emphasizing the discriminative power of features, especially in complex scenarios involving unknown attack types.

Data availability statement

The raw data supporting the conclusions of this article will be made available by the authors, by request and for research purposes only. More info: https://github.com/Choapinus/ArcFace-tf.

Author contributions

AV: Investigation, Methodology, Software, Visualization, Writing – original draft. JT: Investigation, Methodology, Software, Visualization, Conceptualization, Formal analysis, Writing – review & editing. VC: Conceptualization, Supervision, Writing – review & editing. CB: Funding acquisition, Supervision, Writing – review & editing.

Funding

The author(s) declare financial support was received for the research, authorship, and/or publication of this article. This work was supported by the European Union's Horizon 2020 research and innovation program under grant agreement No. 883356 (iMARS) and No. 101121280 (EINSTEIN), the German Federal Ministry of Education and Research, and the Hessen State Ministry for Higher Education, Research and the Arts within their joint support of the National Research Center for Applied Cybersecurity ATHENE.

Conflict of interest

The authors declare that the research was conducted in the absence of any commercial or financial relationships that could be construed as a potential conflict of interest.

Publisher's note

References

Chingovska, I., Anjos, A., and Marcel, S. (2012a). "On the effectiveness of local binary patterns in face anti-spoofing," in *2012 BIOSIG - Proceedings of the International Conference of Biometrics Special Interest Group (BIOSIG)* (Bonn: Gesellschaft für Informatik e.V.), 1–7.

Czajka, A., and Bowyer, K. W. (2018). Presentation attack detection for iris recognition: an assessment of the state-of-the-art. *ACM Comput. Surv.* 51:3232849. doi: 10.1145/3232849

Deng, J., Guo, J., Xue, N., and Zafeiriou, S. (2019). "ArcFace: Additive angular margin loss for deep face recognition," in *Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition* (Long Beach, CA: IEEE), 4690–4699.

Dhar, P., Kumar, A., Kaplan, K., Gupta, K., Ranjan, R., and Chellappa, R. (2022). "Eyepad++: A distillation-based approach for joint eye authentication and presentation attack detection using periocular images," in *2022 IEEE/CVF Conference on Computer Vision and Pattern Recognition (CVPR)* (New Orleans, LA: IEEE), 20186–20195.

Fang, M., Huber, M., Fierrez, J., Ramachandra, R., Damer, N., Alkhaddour, A., et al. (2023). "Synfacepad 2023: Competition on face presentation attack detection based on privacy-aware synthetic training data," in *2023 IEEE International Joint Conference on Biometrics (IJCB)* (Ljubljana: IEEE), 1-11.

Gonzalez-Soler, L. J., Gomez-Barrero, M., and Busch, C. (2023). Toward generalizable facial presentation attack detection based on the analysis of facial regions. *IEEE Access* 11, 68512–68524. doi: 10.1109/ACCESS.2023.3292407

Hoffman, S., Sharma, R., and Ross, A. (2019). "Iris + ocular: generalized iris presentation attack detection using multiple convolutional neural networks," in *2019 International Conference on Biometrics (ICB)* (Crete: IEEE), 1–8.

Howard, A., Sandler, M., Chu, G., Chen, L.-C., Chen, B., Tan, M., et al. (2019). "Searching for mobilenetv3," in *Proceedings of the IEEE/CVF International Conference on Computer Vision* (Seoul: IEEE), 1314–1324.

Hu, G., Yang, Y., Yi, D., Kittler, J., Christmas, W., Li, S. Z., et al. (2015). "When face recognition meets with deep learning: an evaluation of convolutional neural networks for face recognition," in *Proceedings of the IEEE International Conference on Computer Vision Workshops* (Santiago: IEEE), 142–150.

Huang, G., Liu, Z., Van Der Maaten, L., and Weinberger, K. Q. (2017). "Densely connected convolutional networks," in *Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition* (Honolulu, HI: IEEE), 4700–4708.

ISO/IECJTC1/SC 37 Biometrics (2021). *ISO/IEC CD 30107-3, Information Technology – Biometric Presentation Attack Detection – Part 3: Testing and Reporting*. Geneva, CH: Standard, International Organization for Standardization.

Minaee, S., and Abdolrashidi, A. (2019). Deepiris: Iris recognition using a deep learning approach. *arXiv* [preprint] arXiv:1907.09380. doi: 10.48550/arXiv.1907.09380

Pasmino, D., Aravena, C., Tapia, J. E., and Busch, C. (2023). "Flickr-PAD: new face high-resolution presentation attack detection database," in *11th International Workshop on Biometrics and Forensics (IWBF)* (Barcelona: IEEE), 1–6.

Ramachandra, R., and Busch, C. (2017). Presentation attack detection methods for face recognition systems: a comprehensive survey. *ACM Comp. Surv.* 50, 1–37. doi: 10.1145/3038924

Sandler, M., Howard, A., Zhu, M., Zhmoginov, A., and Chen, L.-C. (2018). "Mobilenetv2: Inverted residuals and linear bottlenecks," in *Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition* (Silver Spring: IEEE), 4510–4520.

Tan, M., and Le, Q. (2019). "Efficientnet: Rethinking model scaling for convolutional neural networks," in *International Conference on Machine Learning* (New York: PMLR), 6105–6114.

Tapia, J., Arellano, C., and Viedma, I. (2019). *Sex-classification from Cellphones Periocular Iris Images*. Cham: Springer International Publishing, 227–242..

Tapia, J. E., Gonzalez, S., and Busch, C. (2021). Iris liveness detection using a cascade of dedicated deep learning networks. *IEEE Trans. Inform. Forens. Secur.*17, 42–52. doi: 10.1109/TIFS.2021.3132582

Tapia, J. E., Valenzuela, A., Lara, R., Gomez-Barrero, M., and Busch, C. (2022). Selfie periocular verification using an efficient super-resolution approach. *IEEE Access.* 10, 67573–67589. doi: 10.1109/ACCESS.2022.3184301

Tolosana, R., Gomez-Barrero, M., Busch, C., and Ortega-Garcia, J. (2019). Biometric presentation attack detection: beyond the visible spectrum. *IEEE Trans. Inform. Forens. Security* 15, 1261–1275. doi: 10.1109/TIFS.2019.2934867

Wen, D., Han, H., and Jain, A. K. (2015). Face spoof detection with image distortion analysis. *IEEE Trans. Inform. Forens. Secur.* 10, 746–761. doi: 10.1109/TIFS.2015.2400395

Yambay, D., Becker, B., Kohli, N., Yadav, D., Czajka, A., Bowyer, K. W., et al. (2017). "Livdet iris 2017 iris liveness detection competition 2017," in *2017 IEEE International Joint Conference on Biometrics (IJCB)* (Denver, CO: IEEE), 733–741.

Yu, Z., Qin, Y., Li, X., Zhao, C., Lei, Z., and Zhao, G. (2022). Deep learning for face anti-spoofing: a survey. *IEEE Trans. Pattern Anal. Mach. Intell.* 45, 5609–5631. doi: 10.1109/TPAMI.2022.3215850

Yu, Z., Zhao, C., Wang, Z., Qin, Y., Su, Z., Li, X., et al. (2020). "Searching central difference convolutional networks for face anti-spoofing," in *Conference on Computer Vision and Pattern Recognition*.

Zhang, K., Zhang, Z., Li, Z., and Qiao, Y. (2016). Joint face detection and alignment using multitask cascaded convolutional networks. *IEEE Signal Process. Lett.* 23, 1499–1503. doi: 10.1109/LSP.2016.2603342

Zhang, Z., Yan, J., Liu, S., Lei, Z., Yi, D., and Li, S. Z. (2012). "A face antispoofing database with diverse attacks," in *5th IAPR International Conference on Biometrics (ICB)*, 26–31.