Online-Journals.org

iJIM International Journal of Interactive Mobile Technologies

iJIM | eISSN: 1865-7923 | Vol. 18 No. 10 (2024) | OPEN ACCESS

https://doi.org/10.3991/ijim.v18i10.46981

PAPER

Dual Spectral Attention Model for Iris Presentation Attack Detection

Noura S. Al-Rajeh(✉), Amal A. Al-Shargabi

Department of Information Technology, College of Computer, Qassim University, Buraydah, Saudi Arabia

411200195@qu.edu.sa

ABSTRACT

The widespread use of iris recognition systems has led to a growing demand for enhanced security measures to counter potential iris presentation attacks, also known as anti-spoofing. To enhance the security and reliability of iris recognition systems, researchers have developed numerous methods for detecting presentation attacks. Most of these methods lack precision in detecting unknown attacks compared to known attacks. In addition, most literature on iris presentation attack detection (PAD) systems utilizes near-infrared (NIR) samples as inputs. These samples produce superior-quality and robust images with less reflection in the cornea of the eye. Despite this, due to the widespread use of smartphones and the necessity for unsupervised identity verification, visible-light samples play a crucial role in detecting presentation attacks. These samples can be easily captured using smartphone cameras. In this paper, a dual-spectral attention model has been developed to train a unified model for multiple real-world attack scenarios. Two different scenarios were tested. In the first scenario, the model was trained as a one-class anomaly detection (AD) approach, while in the second scenario, it was trained as a normal two-class detection approach. This model achieved the best result for the attack presentation classification error rate (APCER) of 4.87% in a one-class AD scenario when tested on the attack dataset, outperforming most studies on the same test dataset. These experimental results suggest that future research opportunities in areas such as working with visible light images, using an AD approach, and focusing on uncontrolled environment samples and synthetic iris images may improve iris detection accuracy.

KEYWORDS

presentation attacks, biometric, iris presentation attack detection (PAD), attack detection, spoofing, anomaly detection (AD), one-class classification

1 INTRODUCTION

Recently, biometrics have been used under supervision to verify identities, such as in airports or at borders, making it difficult to spoof another person's identity in a controlled environment [1]. The global biometric authentication market is expected

Al-Rajeh, N.S., Al-Shargabi, A.A. (2024). Dual Spectral Attention Model for Iris Presentation Attack Detection. *International Journal of Interactive Mobile Technologies (iJIM)*, 18(10), pp. 71–89. https://doi.org/10.3991/ijim.v18i10.46981

Article submitted 2023-11-24. Revision uploaded 2024-01-15. Final acceptance 2024-02-18.

to grow by approximately $100 billion by 2027, according to recent market research reports. Moreover, the market is expected to grow at a compound annual growth rate of 14.6% between 2019 and 2027 [2].

Iris recognition is increasingly being used in unsupervised environments, such as to identify authentication on smartphones [3] and in financial applications. According to the Center for Global Development, "iris scans are far more inclusive than fingerprints and are also more precise for authentication, with a lower trade-off curve between errors of acceptance and rejection" [4]. Such robust findings suggest that the significance of iris recognition and its applications will continue to grow daily. However, the potential for tampering and spoofing with these systems appears significant due to their limited supervision.

Spoofing iris recognition systems involves presenting fake or altered iris samples in front of the sensor. Attempts to spoof biometrics are referred to as presentation attacks, and the techniques used for spoof detection are known as presentation attack detection (PAD) methods [5].

Most of the studies in the literature on iris PAD systems use NIR images as inputs. NIR imaging is preferred due to its ability to reduce reflection in the cornea of the eye, leading to higher-quality and more robust input images. This, however, requires the use of highly sophisticated NIR sensors. In an era where smartphones are ubiquitous, detecting presentation attacks using visible-light images for iris recognition in unsupervised environments is advantageous. Visible-range images can be easily captured with smartphone cameras. Iris images captured under visible illumination can be utilized by various mobile apps, surveillance applications, and home security systems. Accurate detection of visible-light iris images is highly desired.

Additionally, it is remarkable that studies on iris PAD do not consider circumstances in which the testing samples are composed of different types of presentation attacks than those present in the training dataset. Existing detection models have failed to accurately detect unknown attacks (previously unseen), as demonstrated in [6], [7], [8], [9], [10], [11]. Thus, the approaches to iris recognition require significant effort to enhance the capability to detect unknown attack types while maintaining high accuracy on known attacks. One-class AD approaches for training and classification may be an excellent solution for real-world iris PAD if the issues associated with this type are resolved. Additionally, to address the weakness in visible-light iris PAD capabilities, this study will develop a dual spectral attention model for iris PAD using the 'train one for all' principle. This approach entails training the proposed model on both visible light and NIR spectra simultaneously to enhance the accuracy of attack detection in contrast to existing models.

2 RELATED WORK

Approaches in iris recognition need to enhance the capability to detect unknown attack types while maintaining high accuracy on known attacks, as this issue appears to be a significant gap in the literature [12]. The path to take is to rely less on data from presentation attack samples and more on data from real samples. Thus, the composition of a one-class anomaly-based approach in various biometric PAD scenarios demonstrates useful characteristics when working with real samples.

Inspired by the positive outcomes of studies on face and fingerprint PAD, we believe that one-class AD methods for training and classification could be a promising excellent solution for real-world iris PAD, provided the challenges associated with this approach are addressed.

2.1 Anomaly detection

One of the new approaches recently used is AD. AD is the process of distinguishing between different samples of data [13]. AD techniques have been used for years in various fields, including fraud detection [14], intrusion detection for cybersecurity [15], medical diagnosis [16], protecting web servers [17], video prediction [18], and recently for face-to-face PAD [19]. The performance of most biometric recognition methods deteriorates when models are tested on previously unseen datasets. To address some of these challenges, some authors have defined biometric PAD as an AD problem; genuine samples are considered normal, while manipulated data are treated as anomalies. The authors concluded that to detect unseen attacks, it is preferable not to make assumptions about the real or fake samples that will be used by an attacker [1].

At the same point, the study by [20] demonstrated that employing identical lens types in both training and testing samples can lead to a deceptive perception of the accuracy that will be achieved when encountering a new lens type. The experiment testing the model with a new lens type resulted in a decrease in classification accuracy to 75%. In contrast, the accuracy was about 95% when the test data only included the same training images.

The work of [21] made the model less reliant on presentation attack samples by requiring it to learn from real samples. Despite the small size of the dataset, the authors found that their model achieved better loss and accuracy, as well as lower error rates, in detecting attacks and real iris samples, despite the dataset's limited size. AD methods could include traditional machine learning approaches, such as one-class support vector machines (OC-SVM) [1], which often do not perform well in data-rich systems. Traditional approaches require numerous essential feature extraction processes before classification, which consume time and resources. Given the continuously expanding size and complexity of data across different fields, there is a growing need for effective and efficient methods to detect anomalies in vast datasets.

2.2 One-class classification

One-class classification is a method for detecting abnormal data samples compared to samples of the known class. One-class classification is another term for AD [22]. The key feature of OCC is its ability to distinguish one-class samples from other samples through one-class learning. One-class classification consists of two types: novelty detection and outlier detection [23]. In novelty detection, abnormalities are identified in the test dataset even though no abnormal data samples are present in the training dataset. It is a good solution for detecting abnormalities in heterogeneous or high-dimensional data [23]. On the other hand, the training dataset may include both normal and abnormal data samples for outlier detection, with the goal of establishing boundaries between them. Following that, the boundaries are applied to the test dataset, which may contain normal and abnormal samples.

The work of [1] proposed a PAD system based on single-class training. This method relies on decision models that utilize information from real data to detect a presentation attack. Three classification scenarios were evaluated, including one-class, unknown attack, and single-class scenarios. The dataset used is the GUC visible spectrum iris artifact (VSIA), as all samples are from the visible spectrum. They found that when the classifier is trained using only real iris samples, the ACER for

most feature extraction methods and attack types is lower than in the unknown attack scenario.

Several studies have compared the traditional two-class classification approach, which assumes prior knowledge of attack types, to one-class classification, which relies solely on real samples for learning. They discovered that the one-class classifier performs well in the presence of unknown attacks. This approach would be more realistic for detecting iris presentation attacks in the real world [1]. If the issues associated with this type are resolved, one-class training and classification may be an excellent solution for real-world iris PAD. Combining deep learning with one-class AD shows promise in detecting abnormal samples; however, further research in various domains is needed to validate this potential [23].

Inspired by these insights and the beneficial features of the one-class anomaly-based approach in biometric PAD and to tackle the limitations of visible-light iris PAD capabilities, this study aims to develop a dual spectral attention model for iris PAD based on the 'train one for all' principle. As opposed to existing models, the proposed model will be trained on both the visible light spectrum and NIR spectrum simultaneously to enhance the accuracy of attack detection.

3 METHODOLOGY

This work aims to detect iris presentation attacks. To accomplish this, two recent datasets were used. Some arrangement tasks are then performed on the training and testing samples. After developing a dual spectral attention model, the model is applied using two approaches (one-class AD and two-class) to detect iris presentation attacks. The evaluation process would be carried out using samples that differ from those used in the training phase. Various matrices were utilized in this process. The research framework summary is illustrated in Figure 1.

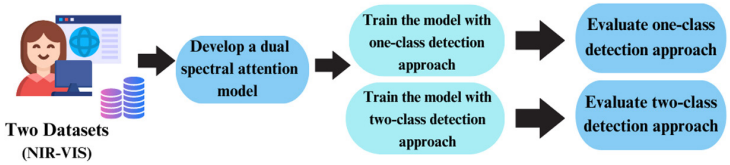

Fig. 1. Research framework

3.1 Datasets

This section introduces the datasets used in the experiments. Two large-scale datasets are used: the PolyU cross-spectral iris dataset [24, 25] for the training phase and the IIITD-WVU mobile iris spoofing dataset [26–28] for the testing phase.

- PolyU cross-spectral iris image dataset: The PolyU cross-spectral iris image dataset is a collection of multispectral real samples that consist of a combination of NIR and VIS images. It includes 12,540 iris images from 209 different subjects. Moreover, this dataset of iris samples was captured using a simultaneous bi-spectral imaging setup from both the right and left iris. Each image is 640×480 pixels in size. Each subject's data is saved in separate folders numbered 001 through 209. Samples from this dataset shown in Figure 2.

- IIITD-WVU dataset: The IIITD-WVU mobile iris spoofing dataset was a subset of the LivDet Iris 2017 competition. Sensors and acquisition conditions for training and testing sets differ because the evaluation was designed as a cross-dataset assessment. In the training dataset, there are 2,250 real iris samples with contact lenses, 1,000 textured samples with contact lenses, and 3,000 print attack samples. All these images were acquired using different sensors. The testing dataset contains iris images acquired with mobile sensors. The dataset contains 4,209 iris samples captured with patterned contact lenses, with or without textured contact lenses (real iris images). The dataset was captured in various settings and environmental conditions. Table 1 below shows details of the datasets used. Samples from this dataset shown in Figure 3.

Note: Only the test part of the IIITD-WVU mobile iris spoofing dataset will be used. This is because an AD approach will be employed during in training, where the model will exclusively train on real samples from the PolyU cross-spectral iris dataset.

Table 1. Details of the used datasets

Dataset	Spectra	Number of Images
PolyU Cross-spectral	NIR, VIS	12,540
IIITD-WVU	NIR	10,459

Fig. 2. PolyU Cross-spectral [24, 25]

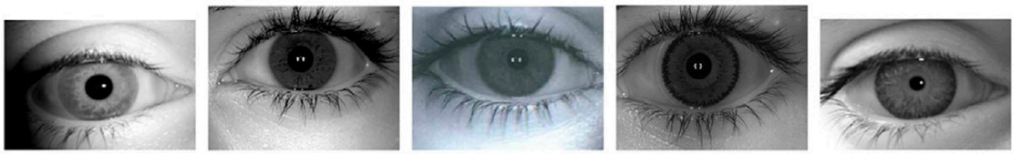

Fig. 3. IIITD-WVU mobile iris spoofing images [26–28]

3.2 Dual spectral attention model

While developing this model and inspired by the success of transformer scaling in natural language processing (NLP), we applied a vision transformer structure similar to the one used in the model of [29] directly to images without any preprocessing. The vision transformer outperformed the model on many image classification datasets while having a lower pretraining cost [29]. Figure 4 illustrates the design of the dual spectral attention model.

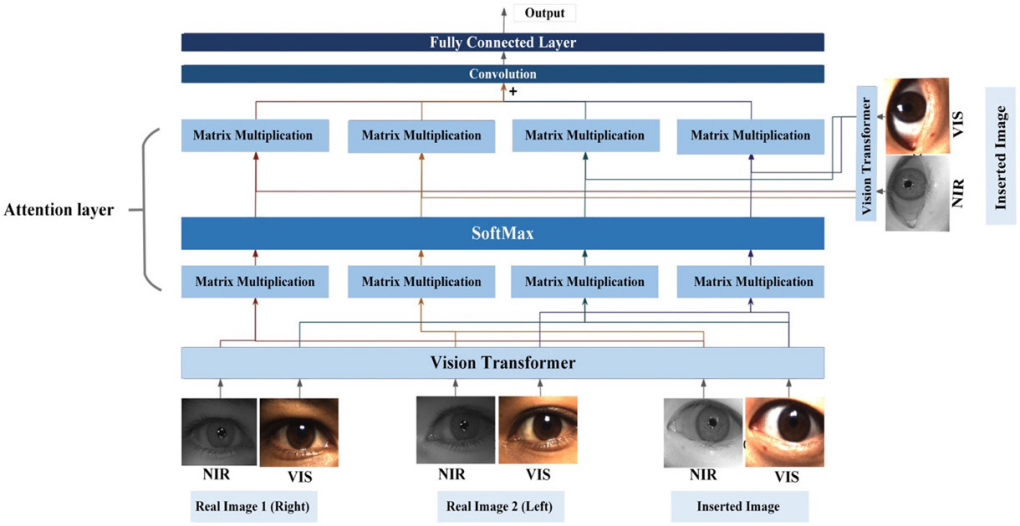

Fig. 4. Dual spectral attention model structure

In this model, the vision transformer (ViT) functions as a feature extractor, eliminating the need for a separate feature extraction step. The vision transformer divides the image into fixed-size patches, embeds each patch linearly, adds position embeddings, and then feeds the resulting vector sequence to a transformer encoder. This projection's output is known as patch embeddings. Figure 5 shows an overview of the transformer structure.

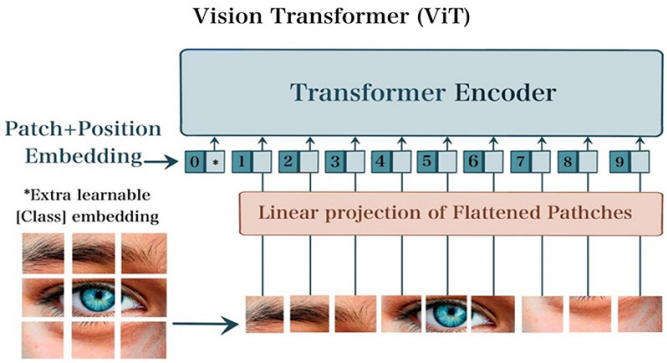

Fig. 5. overview of the vision transformer structure

Particularly the vision transformer's first layer linearly projects the flattened patches into a lower-dimensional space. Following the projection, the patch representations are enhanced with a learned position embedding. The ViT extracts features from all input images and then forwards them to the attention layer. After that, the relationship is mapped via cross-attention (CA) between the inserted image (on the right in Figure 4) and the real images 1 and 2 (NIR and VIS spectral for both). This is where our model comes into play. In the attention layer, the normalized modality-interacted maps are multiplied to form cross-attention features. For the second time, the original inserted image and cross-attention features from the previous step are combined and aggregated using a convolution layer to extract the features.

In this dual spectral attention model, we utilized the SoftMax function between two layers of matrix multiplications. The SoftMax function takes the input and produces probabilities that sum up to one. The SoftMax function generates a vector

containing probability distributions for possible outcomes. Then, the outcomes were compared with the features of the same inserted image for the second time using matrix multiplication. Finally, the results of the attention layer are fed into the convolution layer, which extracts features from the previous step before passing them on to a fully connected layer. The fully connected layer then generates the decision to determine if the sample is real or under attack. The convolution layer used to extract features from the last step is a composite function of three different operations: convolution, batch normalization (BN), and rectified linear unit (ReLU).

3.3 Performance measures

The performance of the methods is crucial for the practical application of various iris PAD solutions. The performance of biometric systems has been evaluated using a variety of measures. The ISO/IEC 1979 series of standards standardizes the assessment of biometric performance, which is conducted in collaboration with ISO/IEC [30].

We used three metrics to evaluate the performance of the iris PAD model: attack presentation classification error rate (APCER), bonafide presentation classification error rate (BPCER), and half total error rate (HTER). The relative importance of APCER and BPCER will vary depending on the use-case scenario. Low BPCER values are especially crucial for low-security applications such as phone unlocking, while low APCER values are more critical for high-security applications like encryption and financial services.

- Attack presentation classification error rate: The APCER measure represents the percentage of misclassified attacked samples. Therefore, the lower the APCER, the better the performance. The formula for the APCER is as follows:

$$APCER = FP/(TN + FP) \quad (1)$$

  Where FP is false-positive, and TN is true negative samples.
- Bonafide presentation classification error rate: The BPCER measure is the percentage of real images misclassified. Therefore, the lower the BPCER, the better the performance. The formula for the BPCER is as follows:

$$BPCER = FN/(TP + FN) \quad (2)$$

  Where FN is false-negative, and TP is true positive samples.
- Half total error rate: The HTER is a common metric for evaluating a detection system's performance. The HTER value is calculated as the average of the APCER and BPCER values. A low HTER value indicates that the detection system is working properly. The formula for the HTER is as follows:

$$HTER = (APCER + BPCER)/2 \quad (3)$$

3.4 Statistical analysis of the proposed model

This section identifies two different scenarios. The experiments have been designed to compare the performance of these two distinct iris PAD approaches.

Scenario 1: one-class anomaly detection approach. In this experimental setting, we explore the generalizability of the one-class AD approach for detecting unknown iris presentation attacks. A dual spectral attention model is applied to detect presentation attacks or misclassification of real irises (not belonging to the same person). Two cases are designed by considering the types of attacks: 1) mixed attacks, and 2) contact lens and printout attacks separately. In this scenario, we will use the PolyU cross-spectral iris image dataset, which contains real iris samples without attacks for training, and the attacks from the IIITD-WVU dataset for testing. Using real sample datasets for training the model and testing it with attack samples is an AD approach. The AD approach (one-class classification) was chosen due to its remarkable performance in addressing biometric presentation attacks [1], [19], [31], [32], [33], [34], [35]. The training dataset would be the PolyU cross-spectral iris image dataset (session 1). During the testing process, we will utilize one of the most challenging datasets provided in the Iris Liveness Detection Competition 2017 [28], IIITD-WVU, to assess the model. Table 2 below summarizes the number of iris samples used in training, validation, and testing in this scenario. Figure 6 shows the number of iris samples and datasets during training, validation, and testing in scenario 1.

Table 2. Total number of images in training, validation, and testing (Scenario 1)

No of Training Images	No of Validation Images	No of Testing Images
6270	1640	4209

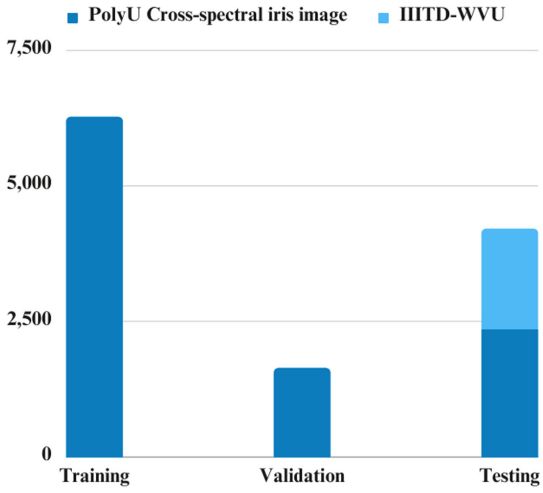

Fig. 6. Number of iris samples and datasets over the training, validation, and testing (Scenario 1)

Scenario 2: two-class detection approach. In this experimental setting, we investigate the impact of training our model with various types of attacks to differentiate between the one-class anomaly classification approach (scenario 1) and the standard two-class classification approach (this scenario). In this scenario, the training set comprises real iris images, printouts, and contact lens attacks, while the testing set includes the same types of attacks but with different samples from the training set. That means we will use both datasets (PolyU cross-spectral iris image and IIITD-WVU) in both the training and testing phases. Table 3 below summarizes the number of iris samples used in training, validation, and testing in this scenario. Figure 7 shows the number of iris samples and datasets during training, validation, and testing in scenario 2.

Table 3. Total number of images in training, validation, and testing (Scenario 2)

No of Training Images	No of Validation Images	No of Testing Images
10895	2730	8479

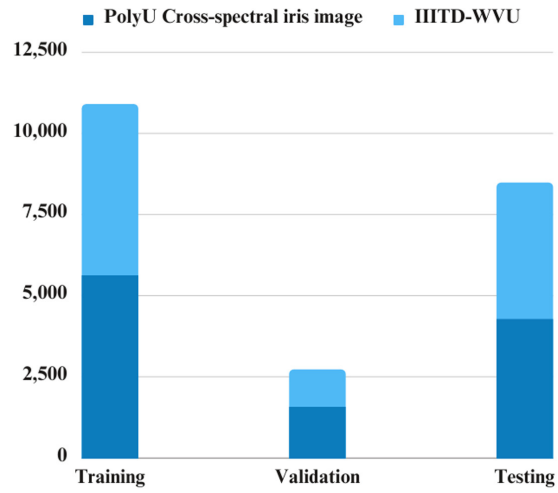

Fig. 7. Number of iris samples and datasets over the training, validation, and testing in (Scenario 2)

4 EXPERIMENTAL RESULTS

Initially, the results demonstrate the performance of detecting this model for real, attacked, and misclassified samples of subjects in scenario 1 (one-class anomaly approach), while the model is trained on real samples from the PolyU cross-spectral iris image dataset. Then, we explore the influence of training on attacks on the model performance in scenario 2 (two-class approach) using the IIITD-WVU with PolyU cross-spectral iris datasets. In addition, we separately examine the performance of detecting different types of attacks available in the IIITD-WVU datasets. The results are also analyzed with the normalized confusion matrix. Finally, we compare the achieved results of the model with the state-of-the-art iris PAD algorithms that use the same dataset for testing. Since the LivDet-Iris protocol does not specify parameters for speed efficiency analysis, the primary emphasis during the implementation of iris PAD models was on normal classification accuracy. The LivDet-Iris 2020 evaluation datasets and comparisons are not included due to the absence of officially provided training data and publicly available test data. We plot the confusion matrix to aid in understanding the model's performance in detecting iris presentation attacks. A confusion matrix is a collection of predicted and actual classification data utilized in a specific system. During analysis, a confusion matrix is created with true positive and negative rates (both true and false) [36]. For the experimental setups, the Adam optimizer with a learning rate (lr) of 1e-5 was used. In addition, the initial learning rate was 0.00001, and the model was trained for a maximum of 30 epochs. Table 4 summarizes the experimental setup details.

Table 4. Experimental setups

Learning Rate	0.00001
Optimizer	Adam
Epochs	30

4.1 Scenario 1: one-class anomaly detection results

The results of this approach in two cases are presented. The first case involves a mixed type of attack, while in the second case, three different attacks were conducted separately on the IIITD-WVU dataset. In this scenario, the model was trained on a PolyU cross-spectral iris image dataset that does not contain any attack types, and the IIITD-WVU dataset was used for the attack samples in the testing set.

Mixed type of attacks detection results. In this case, the model achieved good performance, with an APCER of 4.87%. On the other hand, the BPCER was 20.68%. This occurred because the trained model occasionally confused the real samples belonging to the correct persons and misclassified them due to the similarity in features between these two types. To illustrate the performance of the proposed model in this scenario, see Figure. 8, which shows the confusion matrix in this case.

Contact lens and printouts attacks results. In this experiment, we explore the PAD performance for each type of presentation attack in the test IIITD-WVU dataset. According to the results of the LivDet-Iris 2017 competition, printed iris images are easier to detect than textured contact lenses. This was demonstrated in our study, where printed images yielded better results than contact lenses in terms of detection accuracy. In most cases, contact lens attacks have higher APCER values than printout attacks. This can be seen in the confusion matrix in Figures 9 and 10. Contacts achieved a 2.39% APCER, while printouts achieved a 1.90% APCER. Hence, we conclude that, in most cases, contact lenses are more challenging to detect than printouts. Table 5 summarizes the performance of the model in scenario 1.

Table 5. Performance of the model in Scenario 1

Case	APCER	BPCER	HTER
Mixed types	4.87	20.68	12.78
Contact lens	2.39	12.85	7.62
Printouts	1.90	9.65	5.28

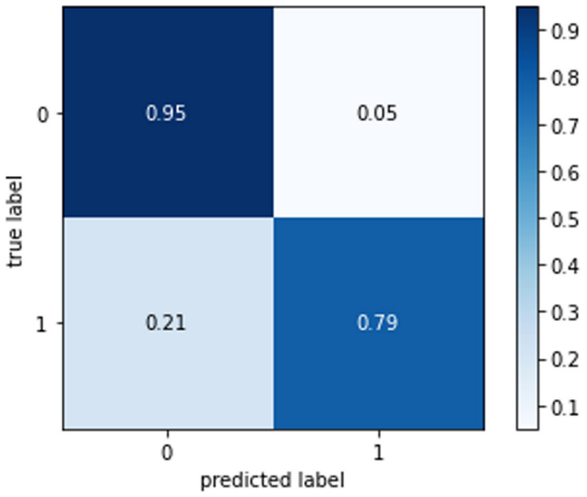

Fig. 8. Normalized confusion matrix of mixed type of attacks (Scenario 1)

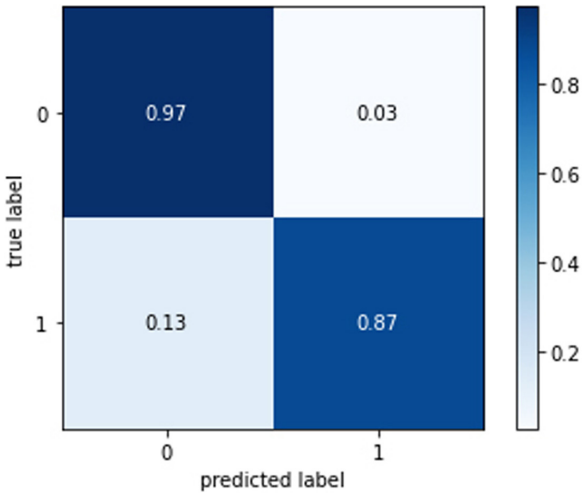

Fig. 9. Normalized confusion matrix of contacts attack (Scenario 1)

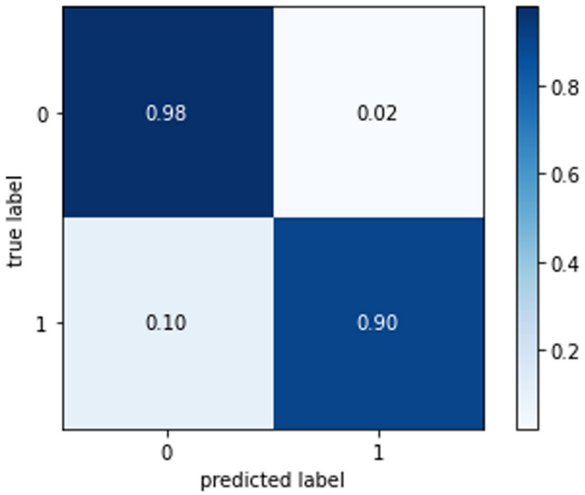

Fig. 10. Normalized confusion matrix of printouts attack (Scenario 1)

4.2 Scenario 2: two-class detection results

The results of the two-class detection experiments in two cases are presented in this section. In this scenario, the model was trained on real iris samples from the PolyU cross-spectral iris image dataset alongside samples of attacks from the IIITD-WVU dataset. Additionally, since the IIITD-WVU dataset comprises contact lens attack samples and printout attack samples, this experiment examines the outcomes of training the model using the three folders of the dataset: 1) contact lenses, 2) printouts, and 3) printouts combined with contact lenses.

Mixed types of attacks detection results. It is noticeable from the results of this experiment that training the model on all types of attacks increased the value of APCER from 4.87% in experiment one to 11.8%. At the same time, the BPCER value decreased from 20.68% to 19.6% in the experiment. Despite this increase in APCER, our proposed model still achieves the lowest HTER compared to other models in the same dataset. Furthermore, our model achieved the best results in detecting printout attacks and similar results in detecting contact lenses. This may be due to training our model with multi-spectrum iris samples (NIR and VIS) simultaneously, unlike

other studies where models were not trained on multi-spectrum samples together. See Figure 11 shows the confusion matrix in this case.

Contact lens and printouts attacks results. The value of APCER in this case was 9.2% while detecting the contact lens. This may be attributed to the small number of contact lens samples in the training phase. On the other hand, BPCER achieved 23.40%, indicating a relatively significant difference between the values of these two measures. This is one of the clear observations in this field, where the values between the two measures are very different. Furthermore, the printouts achieved an APCER 7.80% and BPCER of 20.65%, which are higher than the values observed in the experiment in scenario one. This leads us to infer a preference for a one-class AD approach. Please refer to Table 6 below for the detailed results of this experiment. Figures 12 and 13 show the confusion matrix for this case.

Table 6. Performance of the model in Scenario 2

Case	APCER	BPCER	HTER
Mixed types	11.8	19.6	15.7
Contact lens	9.2	23.40	16.3
Printouts	7.80	20.65	14.33

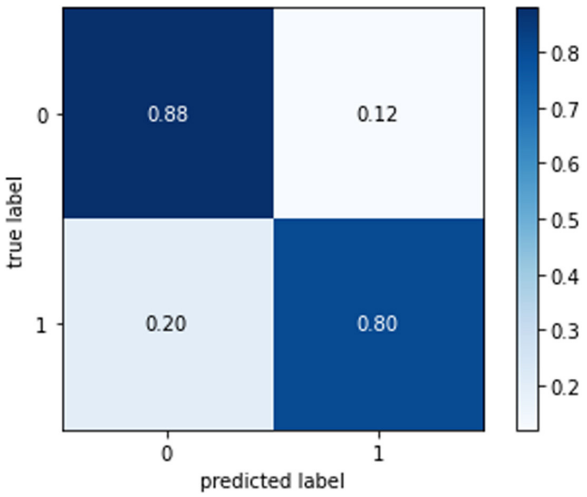

Fig. 11. Normalized confusion matrix of mixed type of attacks (Scenario 2)

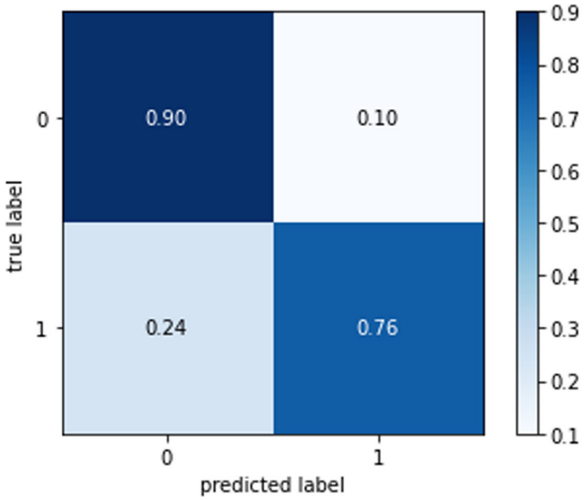

Fig. 12. Normalized confusion matrix of contacts attack (Scenario 2)

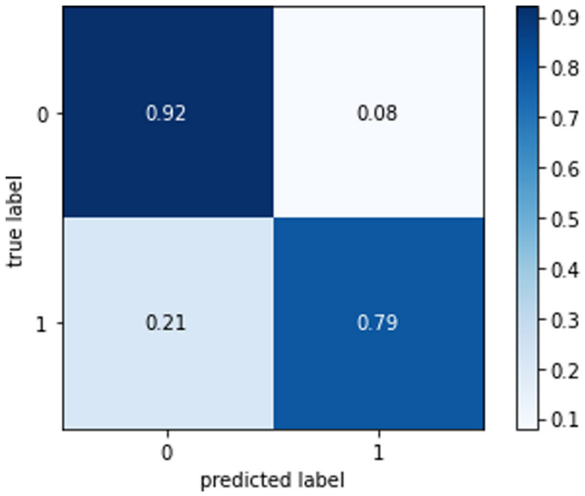

Fig. 13. Normalized confusion matrix of case printouts attack (Scenario 2)

4.3 Evaluation

The obtained results of the model are compared to state-of-the-art PAD models in the IIITD-WVU databases in the Table. Before analyzing the comparison outcomes, it is important to note that some studies tested their algorithms on various types of attacks that were not present during the training phase (unknown attacks). However, it should be considered that all the algorithms being compared were trained using both genuine and forged iris samples. While the dual spectral attention model in scenario 1 was fully trained on real iris samples, it was not trained on samples of attack types at all. Despite the differences in training approaches, this comparison of other results of the proposed model with the works on the same test dataset is the only one possible. This is because no other studies in iris PAD have employed a similar training approach on the same test dataset. Table 7 summarizes the APCER, BPCER, and HTER results from the IIITD-WVU test dataset.

Table 7. Performance of Dual spectral attention model, and existing state-of-the-art models on IIITD-WVU dataset in terms of APCER (%), BPCER (%) and HTER (%)

Method	Cross-Spectrum Test	Total of Training Images	APCER	BPCER	HTER
Winner [28]		16,900	29.40	**3.99**	16.70
SpoofNet [37]		16,900	**0.34**	36.89	18.62
D-NetPAD [38]		10,536	36.41	**10.12**	23.27
MLF [39]		7450	5.39	24.79	15.09
UNINA [28]		16,900	23.18	35.75	29.44
BWWVA [6]		16900	12.29	33.05	22.67
A-PBS [40]	✓	25,413	36.30	15.95	26.13
Dual spectral attention model (Scenario 1)	✓	6270	**4.87**	20.68	**12.78**
Dual spectral attention model (Scenario 2)	✓	10895	11.8	19.6	**15.7**

Note: Bold numbers indicate the lowest two values which mean better performance.

Before analyzing the comparison results, here is a brief description of the most recent methods. Cascade SpoofNets was designed based on the GoogleNet architecture to detect presentation attacks in the Livdet-Iris 2017 competition [28]. In the competition, they placed first on the dataset. The authors in [37] then proposed hyperparameter fin-tuning of the CASIA SpoofNets methods in 2020 to improve PAD performance. Furthermore, the D-NetPAD detector is built in the DenseNet convolutional neural networks (CNN) architecture [38]. The work [39] utilized MobileNet V3 Small, which comprises 2.5 million parameters, and employed depth-wise convolution and squeeze-and-excitation. In addition, the UNINA approach was based on iris segmentation, dense feature extraction, and SVM classification [28]. Finally, the work [6] leveraged multiple pre-trained BSIF filters to effectively train lightweight convolutional neural networks.

It can be observed in Table 7 that the dual spectral attention model architecture achieves significantly improved performance compared to UNINA [28] and a slightly lower HTER value than the A-PBS model [40]. For instance, the HTER was reduced from 15.09% by MLF [39] and 16.70% by the winner of the Livdet-Iris 2017 competition to 12.78%. Besides, the model achieves an APCER value of 4.87%, which is better than the winner of the competition [28] with 29.40% and similar to the best performance of 0.34%. In addition, the proposed dual spectral one-class attention model achieves similar or better results compared to state-of-the-art algorithms. For instance, it reduces the BPCER value from 36.89% by the state-of-the-art algorithm SpoofNet [37], 35.75% by the algorithm UNINA [28], 33.05% by the algorithm BWWVA [6], and 24.79% by the algorithm MLF [39] to 20.681%.

A possible reason for not improving the value of BPCER in experiment one is that some genuine samples were mistakenly classified as attacks because they did not belong to the same person. As a result, the model could misclassify these types of attacks (misclassified samples of subjects) and consider them genuine. Whereas other models being compared typically handle real samples without taking into account people's affiliations. Despite the fact that the IIITD-WVU test dataset has different sensor features and acquisition environments, the model does not perform any pre-processing on the iris image before training and testing. In contrast, the majority of the compared works utilize contrast enhancement and remove illumination influence [39].

The results indicate that our model has better generalizability than the other models compared.

4.4 Discussion

The main contribution is the development of a new dual-spectral attention model. This model focuses on training a unified model for multiple real-world attack scenarios and investigates two approaches (one-class anomaly and two-class) to detect iris presentation attacks. First, multi-spectral model training performs well in detecting both types of attacks and misclassified samples in both scenarios. Most of the studies compared focused on the NIR spectrum. In the Clarkson dataset used in [6], [28], [38], [39], and [37], some samples were printouts created from visible light images of the eye. These images were then processed to extract the red channel and convert it to a grayscale image, which was printed and presented to the iris camera. While it was the only study [40] that focused on multispectral imaging, our work achieved better results in detecting attacks on the same dataset compared to that study. Summing up all the results, we can see that training with multispectral data

significantly improves PAD performance compared to training only with data from a single spectrum. Considering that our multi-spectral model achieved similar or better results than the state-of-the-art models that primarily focus on one spectral band (NIR), it is advantageous to integrate various spectrums into network architectures intended for biometric recognition in general and iris recognition in particular.

Second, one-class AD in the iris PAD task has shown that it is not inferior to two-class methods. In one-class classification, the decision boundary is enhanced using only the real samples [1]. Because the model is built using only real data, it is immune to the negative impacts of attacks on diverse data on performance. Furthermore, because only real-access data is needed for training, the training set can be easily expanded [19]. In the first scenario, our model outperformed the two-class detection experiment in scenario two and other studies that employed a two-class classification method with limited training data, as our model was trained solely on real data. This helped us overcome the limitations of the training data because there are not many available datasets containing various attack types. Third, the attention mechanism is designed to automatically acquire essential discriminative features from input data relevant to PAD [40]. An attention mechanism was employed in this dual model to gather fine-grained pixel and patch-level cues and utilize regions that contribute the most to a correct PAD decision. The attention mechanism demonstrated superior performance compared to more traditional methods in directing the focus of the PAD models and enhancing the accuracy of detection. This helped us eliminate the need for any pre-processing or data enhancement steps. However, the results do not exhibit exact consistency across all attacks due to the limited availability of test datasets. In this context, we acknowledge that the most significant limitation of our work is the utilization of an attack dataset that comprises only one spectrum (NIR), whereas the actual sample dataset was multispectral (NIR-VIS). This is due to limitations in the attack datasets at the present time, stemming from general data protection regulation (GDPR) issues and complex license agreements. However, the dual model performs well in detecting NIR spectrum attacks, even those unseen in scenario one. This paves the way for incorporating multi-spectrum training to effectively address various image capture environments and lighting conditions that pose significant challenges in different biometric recognition systems. Our model not only detects presentation attacks but can also identify misclassified iris samples (from different subjects), making it suitable for unsupervised applications like mobile application locks, home security, and door access control.

Finally, although businesses and governments have widely embraced iris recognition technology, research on PAD is still in its early stages. So, like other researchers, we encountered some limitations during this study. The main limitation was the lack of available datasets containing visible light spectrum attack samples. We used an attack dataset that only included one spectrum (NIR), while the real samples dataset was multispectral (NIR-VIS). This limitation is attributed to the current constraints of the dataset, stemming from issues related to the GDPR, and intricate license agreements, and legal restrictions. In addition, data set administrators should streamline administrative processes to provide researchers with clear guidance when accessing a dataset. This is because certain datasets require the signature of a legal or institutional representative, which poses a significant hurdle for researchers. In our case, this process took longer than expected. Moreover, the increasing complexity of digital media and technologies reinforces data monopolies, enabling individuals with technological skills to control the information that is presented. Unfortunately, a similar situation is occurring in the domain of iris datasets, where some dataset administrators and educational institutions restrict access to researchers based only

in the United States. This barrier can hinder researchers from other countries from participating in the field, ultimately diminishing scientific benefits and outcomes.

5 CONCLUSION

To address the weaknesses of the existing models in detecting visible-light iris attacks, we need to consider the diversity of image capture environments and lighting conditions, as well as the challenges in accurately detecting unknown attacks compared to known ones. We developed a dual-spectral attention model that trained a unified model for multiple real-world attack scenarios. We investigated two approaches (one-class anomaly and two-class) to detect iris presentation attacks. An attention strategy was utilized as an effective approach to distinguishing real and fake features from different modalities. These two approaches achieved the best detection results compared to other models on the same test dataset. Furthermore, the one-class AD scenario achieved the best result of APCER (4.87%) and has proven that using one-class AD outperforms the two-class detection approach in detecting unknown attacks. The dual model excelled in detecting NIR spectrum attacks, which suggests the possibility of incorporating multi-spectrum training to better handle diverse image capture environments and lighting that pose significant challenges to biometric recognition systems. These experimental results suggest that future research opportunities in areas such as working with visible light images and focusing on uncontrolled environmental samples and synthetic iris images may improve iris detection accuracy. Additionally, one-class AD approaches may overcome the current and anticipated limitations of the training data.

6 ACKNOWLEDGMENT

The authors gratefully acknowledge Qassim University, represented by the Deanship of Scientific Research, for the financial support provided for this research under the grant number COC-2022-1-2-J-30393 during the academic year 1444 AH/2022 AD

7 REFERENCES

[1] A. F. Sequeira, S. Thavalengal, J. Ferryman, P. Corcoran, and J. S. Cardoso, "A realistic evaluation of iris presentation attack detection," in *2016 39th International Conference on Telecommunications and Signal Processing (TSP)*, 2016, pp. 660–664. https://doi.org/10.1109/TSP.2016.7760965

[2] J. A. Sava, "Biometric authentication and identification market revenue worldwide in 2019 and 2027," *Statista*, 2021. https://www.statista.com/statistics/1012215/worldwide-biometric-authentication-and-identification-market-value/

[3] S. Thavalengal, P. Bigioi, and P. Corcoran, "Evaluation of combined visible/NIR camera for iris authentication on smartphones," in *Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition Workshops (CVPRW)*, 2015, pp. 42–49. https://doi.org/10.1109/CVPRW.2015.7301318

[4] A. Gelb and J. Clark, "Performance lessons from India's universal identification program," *CGD Policy Paper*, vol. 20, 2013.

[5] C. Sousedik and C. Busch, "Presentation attack detection methods for fingerprint recognition systems: A survey," *IET Biometrics,* vol. 3, no. 4, pp. 219–233, 2014. https://doi.org/10.1049/iet-bmt.2013.0020

[6] A. Kuehlkamp, A. Pinto, A. Rocha, K. W. Bowyer, and A. Czajka, "Ensemble of multi-view learning classifiers for cross-domain iris presentation attack detection," *IEEE Transactions on Information Forensics and Security,* vol. 14, no. 6, pp. 1419–1431, 2018. https://doi.org/10.1109/TIFS.2018.2878542

[7] S. Hoffman, R. Sharma, and A. Ross, "Convolutional neural networks for iris presentation attack detection: Toward cross-dataset and cross-sensor generalization," in *Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition Workshops (CVPRW)*, 2018, pp. 1701–17018. https://doi.org/10.1109/CVPRW.2018.00213

[8] C. Chen and A. Ross, "Exploring the use of iriscodes for presentation attack detection," in *2018 IEEE 9th International Conference on Biometrics Theory, Applications and Systems (BTAS)*, 2018, pp. 1–9. https://doi.org/10.1109/BTAS.2018.8698581

[9] M. Trokielewicz, A. Czajka, and P. Maciejewicz, "Presentation attack detection for cadaver iris," in *2018 IEEE 9th International Conference on Biometrics Theory, Applications and Systems (BTAS)*, 2018, pp. 1–10. https://doi.org/10.1109/BTAS.2018.8698542

[10] S. Hoffman, R. Sharma, and A. Ross, "Iris+ ocular: Generalized iris presentation attack detection using multiple convolutional neural networks," in *2019 International Conference on Biometrics (ICB)*, 2019, pp. 1–8. https://doi.org/10.1109/ICB45273.2019.8987261

[11] D. Yadav, N. Kohli, M. Vatsa, R. Singh, and A. Noore, "Detecting textured contact lens in uncontrolled environment using densepad," in *2019 IEEE/CVF Conference on Computer Vision and Pattern Recognition Workshops (CVPRW)*, 2019, pp. 2336–2344. https://doi.org/10.1109/CVPRW.2019.00287

[12] A. Boyd, Z. Fang, A. Czajka, and K. W. Bowyer, "Iris presentation attack detection: Where are we now?" *Pattern Recognition Letters,* vol. 138, pp. 483–489, 2020. https://doi.org/10.1016/j.patrec.2020.08.018

[13] L. Ruff *et al.*, "Deep one-class classification," in *International Conference on Machine Learning*, 2018, pp. 4393–4402.

[14] C. Phua, V. Lee, K. Smith, and R. Gayler, "A comprehensive survey of data mining-based fraud detection research," *arXiv Preprint,* no. 1009.6119, 2010.

[15] P. Garcia-Teodoro, J. Diaz-Verdejo, G. Maciá-Fernández, and E. Vázquez, "Anomaly-based network intrusion detection: Techniques, systems and challenges," *Computers & Security,* vol. 28, nos. 1–2, pp. 18–28, 2009. https://doi.org/10.1016/j.cose.2008.08.003

[16] O. Salem, A. Guerassimov, A. Mehaoua, A. Marcus, and B. Furht, "Sensor fault and patient anomaly detection and classification in medical wireless sensor networks," in *2013 IEEE International Conference on Communications (ICC)*, 2013, pp. 4373–4378. https://doi.org/10.1109/ICC.2013.6655254

[17] K. L. Ingham and H. Inoue, "Comparing anomaly detection techniques for HTTP," in *International Workshop on Recent Advances in Intrusion Detection*, 2007, vol. 4637, pp. 42–62. https://doi.org/10.1007/978-3-540-74320-0_3

[18] W. Liu, W. Luo, D. Lian, and S. Gao, "Future frame prediction for anomaly detection–a new baseline," in *Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition*, 2018, pp. 6536–6545. https://doi.org/10.1109/CVPR.2018.00684

[19] S. Fatemifar, S. R. Arashloo, M. Awais, and J. Kittler, "Client-specific anomaly detection for face presentation attack detection," *Pattern Recognition,* vol. 112, p. 107696, 2021. https://doi.org/10.1016/j.patcog.2020.107696

[20] K. W. Bowyer and J. S. Doyle, "Cosmetic contact lenses and iris recognition spoofing," *Computer,* vol. 47, no. 5, pp. 96–98, 2014. https://doi.org/10.1109/MC.2014.118

[21] P. M. Ferreira, A. F. Sequeira, D. Pernes, A. Rebelo, and J. S. Cardoso, "Adversarial learning for a robust iris presentation attack detection method against unseen attack presentations," in *2019 International Conference of the Biometrics Special Interest Group (BIOSIG)*, 2019, pp. 1–7.

[22] M. M. Moya, M. W. Koch, and L. D. Hostetler, "One-class classifier networks for target recognition applications," *NASA STI/Recon Technical Report N,* vol. 93, p. 24043, 1993.

[23] N. Seliya, A. Abdollah Zadeh, and T. M. Khoshgoftaar, "A literature review on one-class classification and its potential applications in big data," *Journal of Big Data,* vol. 8, pp. 1–31, 2021. https://doi.org/10.1186/s40537-021-00514-x

[24] "The Hong Kong Polytechnic University cross-spectral iris images database," 2023. http://www4.comp.polyu.edu.hk/~csajaykr/polyuiris.htm. [Accessed: 24 Feb, 2023].

[25] P. R. Nalla and A. Kumar, "Toward more accurate iris recognition using cross-spectral matching," *IEEE Transactions on Image Processing,* vol. 26, no. 1, pp. 208–221, 2017. https://doi.org/10.1109/TIP.2016.2616281

[26] D. Yadav, N. Kohli, J. S. Doyle, R. Singh, M. Vatsa, and K. W. Bowyer, "Unraveling the effect of textured contact lenses on iris recognition," *IEEE Transactions on Information Forensics and Security,* vol. 9, no. 5, pp. 851–862, 2014. https://doi.org/10.1109/TIFS.2014.2313025

[27] P. Gupta, S. Behera, M. Vatsa, and R. Singh, "On iris spoofing using print attack," in *2014 22nd International Conference on Pattern Recognition*, 2014, pp. 1681–1686. https://doi.org/10.1109/ICPR.2014.296

[28] D. Yambay *et al.*, "LivDet iris 2017—Iris liveness detection competition 2017," in *2017 IEEE International Joint Conference on Biometrics (IJCB)*, 2017, pp. 733–741. https://doi.org/10.1109/BTAS.2017.8272763

[29] A. Dosovitskiy *et al.*, "An image is worth 16 × 16 words: Transformers for image recognition at scale," *arXiv Preprint,* no. 2010.11929, 2020.

[30] S. Khade, S. Ahirrao, S. Phansalkar, K. Kotecha, S. Gite, and S. D. Thepade, "Iris liveness detection for biometric authentication: A systematic literature review and future directions," *Inventions,* vol. 6, no. 4, p. 65, 2021. https://doi.org/10.3390/inventions6040065

[31] S. Fatemifar, S. R. Arashloo, M. Awais, and J. Kittler, "Spoofing attack detection by anomaly detection," in *ICASSP 2019–2019 IEEE International Conference on Acoustics, Speech and Signal Processing (ICASSP)*, 2019, pp. 8464–8468. https://doi.org/10.1109/ICASSP.2019.8682253

[32] S. R. Arashloo, J. Kittler, and W. Christmas, "An anomaly detection approach to face spoofing detection: A new formulation and evaluation protocol," *IEEE Access,* vol. 5, pp. 13868–13882, 2017. https://doi.org/10.1109/ACCESS.2017.2729161

[33] S. R. Arashloo and J. Kittler, "Client-specific anomaly detection for face presentation attack detection," *arXiv Preprint,* no. 1807.00848, 2018.

[34] S. Fatemifar, M. Awais, S. R. Arashloo, and J. Kittler, "Combining multiple one-class classifiers for anomaly based face spoofing attack detection," in *2019 International Conference on Biometrics (ICB)*, 2019, pp. 1–7. https://doi.org/10.1109/ICB45273.2019.8987326

[35] T. Ohki, V. Gupta, and M. Nishigaki, "Efficient spoofing attack detection against unknown sample using end-to-end anomaly detection," in *2019 Asia-Pacific Signal and Information Processing Association Annual Summit and Conference (APSIPA ASC)*, 2019, pp. 224–230. https://doi.org/10.1109/APSIPAASC47483.2019.9023183

[36] J. Bhattacharjee, S. Santra, and A. Deyasi, "Chapter 10 – Novel detection of cancerous cells through an image segmentation approach using principal component analysis," in *Recent Trends in Computational Intelligence Enabled Research*, 2021, pp. 171–195. https://doi.org/10.1016/B978-0-12-822844-9.00035-9

[37] G. Y. Kimura, D. R. Lucio, A. S. Britto Jr, and D. Menotti, "CNN hyperparameter tuning applied to iris liveness detection," *arXiv Preprint,* no. 2003.00833, 2020. https://doi.org/10.5220/0008983904280434

[38] R. Sharma and A. Ross, "D-NetPAD: An explainable and interpretable iris presentation attack detector," in *2020 IEEE International Joint Conference on Biometrics (IJCB)*, 2020, pp. 1–10. https://doi.org/10.1109/IJCB48548.2020.9304880

[39] M. Fang, N. Damer, F. Boutros, F. Kirchbuchner, and A. Kuijper, "Deep learning multi-layer fusion for an accurate iris presentation attack detection," in *2020 IEEE 23rd International Conference on Information Fusion (FUSION)*, 2020, pp. 1–8. https://doi.org/10.23919/FUSION45008.2020.9190424

[40] M. Fang, F. Boutros, and N. Damer, "Intra and cross-spectrum iris presentation attack detection in the NIR and visible domains using attention-based and pixel-wise supervised learning," *arXiv Preprint*, no. 2205.02573, 2022. https://doi.org/10.1007/978-981-19-5288-3_8

8 AUTHORS

Noura S. Al-Rajeh received her Master's degree in Cybersecurity from Qassim University. Her research interests include machine learning and cybersecurity. (E-mail: 411200195@qu.edu.sa).

Amal A. Al-Shargabi received Master's and Ph.D. degrees from Universiti Teknologi MARA (UiTM), Malaysia. She is an associate professor at the College of Computer, Qassim University. Her research interests include program comprehension, empirical software engineering, and machine learning.