Tech Science Press

Unconstrained Gender Recognition from Periocular Region Using Multiscale Deep Features

Raqinah Alrabiah, Muhammad Hussain* and Hatim A. AboAlSamh

Department of Computer Science, College of Computer and Information Sciences, King Saud University, Riyadh, 11451, Saudi Arabia
*Corresponding Author: Muhammad Hussain. Email: mhussain@ksu.edu.sa

Abstract: The gender recognition problem has attracted the attention of the computer vision community due to its importance in many applications (e.g., surveillance and human–computer interaction [HCI]). Images of varying levels of illumination, occlusion, and other factors are captured in uncontrolled environments. Iris and facial recognition technology cannot be used on these images because iris texture is unclear in these instances, and faces may be covered by a scarf, hijab, or mask due to the COVID-19 pandemic. The periocular region is a reliable source of information because it features rich discriminative biometric features. However, most existing gender classification approaches have been designed based on hand-engineered features or validated in controlled environments. Motivated by the superior performance of deep learning, we proposed a new method, PeriGender, inspired by the design principles of the ResNet and DenseNet models, that can classify gender using features from the periocular region. The proposed system utilizes a dense concept in a residual model. Through skip connections, it reuses features on different scales to strengthen discriminative features. Evaluations of the proposed system on challenging datasets indicated that it outperformed state-of-the-art methods. It achieved 87.37%, 94.90%, 94.14%, 99.14%, and 95.17% accuracy on the GROUPS, UFPR-Periocular, Ethnic-Ocular, IMP, and UBIPr datasets, respectively, in the open-world (OW) protocol. It further achieved 97.57% and 93.20% accuracy for adult periocular images from the GROUPS dataset in the closed-world (CW) and OW protocols, respectively. The results showed that the middle region between the eyes plays a crucial role in the recognition of masculine features, and feminine features can be identified through the eyebrow, upper eyelids, and corners of the eyes. Furthermore, using a whole region without cropping enhances PeriGender's learning capability, improving its understanding of both eyes' global structure without discontinuity.

Keywords: Gender recognition; periocular region; deep learning; convolutional neural network; unconstrained environment

1 Introduction

People can naturally recognize soft biometric traits, such as gender, age, and ethnicity, at first sight using facial characteristics. Unlike age, gender and ethnicity are permanent attributes and do not change over time. Gender classification (GC) would boost the performance of several fields of study and applications, such as human–computer interaction (HCI), surveillance systems, access control, demographic collection, forensic identification, content-based retrieval, and advertisements [1]. Different modalities (e.g., the face, iris, and periocular regions) have been employed to solve the GC problem. Some methods have relied on face images acquired through careful preparation and posing in a constrained environment; thus, these methods usually achieve good accuracy [2]. However, their performances decline significantly when photos feature real scenarios in which faces are presented under different conditions [3], such as the occlusions shown in Fig. 1. For instance, criminals cover all parts of their faces except for their eyes when committing crimes. In this situation, recognizing gender based only on the periocular region can narrow search results, thus streamlining criminal investigations. In addition, certain countries have gender-specific environments stemming from cultural/religious reasons, and access to a prohibited area cannot be controlled if someone has a partially occluded face. When confronted with such a range of situations, the periocular region can be more effective than the whole face in allowing artificial intelligence (AI) and machine learning systems to recognize gender.

Figure 1: Situations where the periocular region is effective as a biometric trait (source: web, [4,5])

GC using the iris is an invasive approach, and it is not socially acceptable or suitable for unconstrained scenarios [6]. The periocular region outperforms the iris in the GC task [7]. Due to facial and iris recognition limitations, an alternative modality is the periocular region, the facial area surrounding the eyes (Fig. 2), which can be used as a cue for soft biometrics. Brown and Perrett [8] reported that eyebrows and eyes together, eyebrows alone, and eyes alone are the top three facial areas that carry the most discriminative gender information.

GC using the periocular region has received extensive research interest [9–21]. Some studies have employed handcrafted features [9–13], and others have used deep learning to complete the GC task [14–21]. However, most methods were evaluated using images captured under constrained settings. A few approaches [9,10] considered wild image problems and validated them on challenging datasets. However, these methods employed custom descriptors, and the dimensions of the feature space were very high. The best-obtained accuracy reached 83.41% [10] on the GROUPS dataset, meaning that this field requires further exploration.
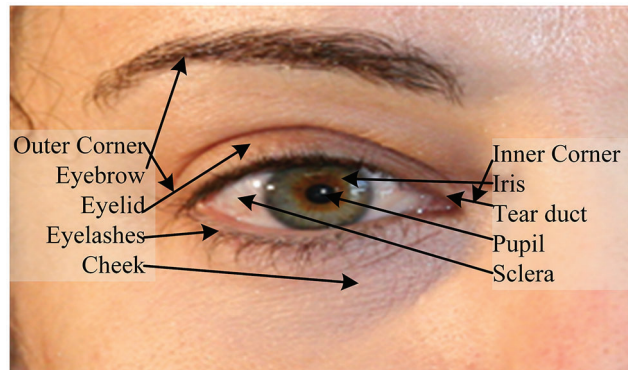
Figure 2: Structure of the periocular region; left eye image from the UBIPr dataset

Convolutional neural networks (CNNs) have shown promising solutions to computer vision problems. Previous studies [22,23] showed that CNN has a significant sensitivity to face-salient regions, which are the center of the mouth and the periocular region. Xu et al. [23] confirmed previous findings [24] that the most salient area of the human face is the periocular region. They trained the DeepGender network with whole face images rather than only periocular region images and found that the latter achieved better accuracy than images featuring whole faces with occlusions, illumination variations, and low resolution.

Motivated by these studies, we proposed a CNN model for GC that extracts and fuses features from the periocular region at different scales. The specific contributions of this research are as follows:

- We introduce a reliable gender recognition system based on a deep CNN model for periocular recognition. The system is invariant in different illumination, resolution, and occlusion conditions.
- We introduce a deep CNN model (PeriGender) that extracts and fuses features of different scales through skip connections between low-, medium-, and high-level layers. The fusion of features at different scales enhances the model's discriminative ability.
- We validate PeriGender's performance on five challenging datasets— the facial GROUPS dataset and the periocular UBIPr, UFPR, IMP, and Ethnic-Ocular datasets—using intra-and cross-database evaluation protocols. Experiments were performed using the open-world (OW) and closed-world (CW) protocols. The system was evaluated using standard performance metrics.
- We utilize explainable artificial intelligence (XAI) based on Gradient-weighted Class Activation Mapping (Grad-CAM) visualization [25], which reveals distinctive gender regions learned by deep learning techniques.

The remainder of this paper is structured as follows: Section 2 reviews the related literature. Section 3 describes the data preprocessing step and the proposed method. The datasets and evaluation methods used are presented in Section 4. Section 5 presents the experimental results, and Section 6 offers a discussion of the main findings. Section 7 concludes the paper.

2 Literature Review

The problem with GC using images of the periocular region taken in the wild has been addressed by some researchers. The following paragraphs provide an overview and analysis of these methods, which can be categorized as conventional and deep learning-based methods.

The method proposed by Merkow et al. [24] is a pioneering work that jumpstarted this field of research. Santana et al. [9] extracted histograms of oriented gradients (HOGs), local binary patterns (LBPs), local ternary patterns (LTPs), Weber local descriptors (WLDs), and local oriented statistics information booster (LOSIB) features. For classification, they used a support vector machine (SVM). The system was evaluated on the GROUPS dataset. The study also fused the periocular pattern with features of the face,

head, and shoulders. In [10], the authors used holistic and inner facial patches to extract multiscale features, such as WLDs, local salient patterns (LSPs), and LTPs. They were then combined using a score-level fusion (SLF) method. They proved that the multiscale features improved the GC system's accuracy. Woodard et al. [11] extracted local appearance features, such as LBPs, LTPs, LSPs, local phase quantizations (LPQs), local color histograms (LCHs), and histograms of Gabor ordinal measures (HOGOMs). Rattani et al. [12] studied the problem of uncontrolled mobile devices considering motion blur, occlusion, reflection, and illumination caused by mobility. LBPs, LTPs, LPQ, binarized statistical image features (BSIFs), and HOGs were extracted and passed to the SVM and multilayer perceptron (MLP). Kao et al. [13] addressed the occlusion problem by jointing multiple patterns and using HOGs, principal component analysis (PCA), and the flood fill algorithm to extract the features. For classification, they used a random forest (RF) and an SVM.

An extended study by Rattani et al. [14] continued the exploration into integrated biometric authentication and mobile health care systems using deep learning. They also studied human ability for comparative analysis. Tapia and Aravena [15] produced two modified LeNeT-5 CNNs, one for the left eye and the other for the right eye, and fused them. Manyala et al. [16] extracted the periocular region hierarchically and studied the effects of the absence of eyebrows on performance. They explored two types of CNNs: a pre-trained CNN as a feature extractor (VGG-Face) and a fine-tuned CNN as an end-to-end gender classifier. Tapia et al. [17] extracted features using super-resolution CNNs (SRCNNs) and classified them using RF. They enhanced the resolution of low-quality periocular selfies. However, a limitation of these previous methodologies [12,14] is their limited use of biometric ocular databases, such as VISOB [26]. This leads to difficulties in comparing the proposed method with other methods in the same domain. Until 2020, there was a lack of challenging wild biometric databases with gender metadata. Zanlorensi et al. [19] launched an unconstrained large-scale UFPR dataset that is useful for investigating our research problem. The dataset features images with different illumination, resolution, and occlusion conditions. The authors employed a multi-task approach based on MobileNetV2. Fernandez et al. [20] extracted features using AlexNet, ResNet-50, ResNet-101, DenseNet-201, VGG-Face, and MobileNetV2 and then classified them using an SVM. They reduced the dimensionality using PCA. In [21], they utilized lightweight CNN models, such as SqueezeNet and MobileNetV2. They validated their work on the Adience dataset. However, the highest accuracy for GC was 84.06% using a genetic algorithm [18]. They employed a facial patch-based approach for GC, extracted compass LBP features, and classified them using SVM and an artificial neural network (ANN).

The previous methods are summarized in Tab. 1. The overview indicates that most of the existing methods do not provide good gender recognition performance. Most current methods were designed to have two phases: first, the features are extracted through the descriptor, and second, they are passed to the classifier for decision-making. These methods are based on custom features that do not consider the nature of images, and they require much time and experience to design. Some techniques [11,12] were validated on small databases not applicable to wild images. As such, they do not yield good results, even when evaluated on a challenging dataset [9,10]. CNN models have been proposed to address the deficiencies of previous approaches and have achieved good accuracy [14,16]. They hierarchically extracted deep features learned directly from observations of the periocular region. Some approaches have used pre-trained CNN models due to limitations in the training data available. A custom CNN was shown to outperform custom techniques by almost 4%, confirming the superiority of deep learning-based approaches against hand-designed approaches [14]. In addition, the former outperformed human capabilities by 3%, meaning that it succeeded in emulating human abilities. Given the advances in deep learning, specifically in CNN architectures and their success in other pattern classification problems [27–29], these architectures should be explored to enhance gender recognition performance using only images of the periocular region. In the literature, there are two performance evaluation protocols: the OW and the CW. There are different subjects in the training and test sets in the OW protocol. In the CW protocol, different samples of the same subject may exist in both the training and test sets. Previous studies [9–10,14,16,21] have performed 5-fold cross validation (CV) for the OW protocol. The classification performance in these studies was lower than that of other studies implemented in the CW protocol [16,19]. Reference [11] did not indicate whether a validation protocol was used for an OW or a CW scenario.

Table 1: The state-of-the-art gender classification methods using periocular region modality. The letters L, H, F, and V in the factors column stand for low, high, frontal, and varied, respectively, while the letters l and r in the accuracy (ACC) column stand for the left and right ocular regions

Paper	Dataset	Images	Variability factors						Features/ Classifier	Protocol	ACC (%)
			Illumination	Expression	Resolution	Pose	Occlusion	Blur			
Hand-designed methods											
Santana et al. [9]	GROUPS	28,000	Y	N	V	Y	Y	N	HOG/SVM	OW	82.91
Santana et al. [10]	GROUPS	28,000	Y	N	V	Y	Y	N	WLD + LSP + LTP/SVM	OW	83.41
Woodard et al. [11]	Pinellas	1000	Y	N	N	Y	Y	Y	LBP/SVM	-	l 88.86 r 89.48
Rattani et al. [12]	VISOB	1600	Y	N	V	F	Y	Y	HOG/MLP	CW	90.20
Kao et al. [13]	AR	88	Y	N	Y	F	Y	Y	HOG, PCA, Flood Fill/RF, SVM	OW	85.61
	FERET	748								OW	85.47
Deep-learning-based methods											
Rattani et al. [14]	VISOB	2917	Y	N	V	F	Y	Y	ResNet/SVM, MLP	OW	90
Tapia and Aravena [15]	ND-GFI	19,000	Y	N	L	F	Y	N	LeNeT-5	OW	87.26
Manyala et al. [16]	In-house	576	Y	Y	V	Y	N	Y	VGG-Face/ SVM	OW	l 96.37 r 96.67
	MBGC	909								OW	93.13 86.34
										CW	98.20 96.60
	IMP	620								OW	86.11 84.77
Tapia et al. [17]	INACAP	450	Y	N	Y	F	N	N	SRCNNs/RF	OW	87.15
Bhattacharyya et al. [18]	Adience	1757	Y	Y	Y	F	Y	Y	LBP/SVM, ANN	OW	84.06
Zanlorensi et al. [19]	UFPR	33,660	Y	Y	Y	Y	Y	Y	MobileNetV2	CW	97.80
Fernandez et al. [20]	LFW	12,700	Y	Y	Y	Y	Y	Y	MobileNetV2 + PCA/SVM	OW	93.40
Fernandez et al. [21]	Adience	11,299	Y	Y	Y	F	Y	Y	MobileNetV2/ SVM	OW	78.90

3 Proposed Method

This study aimed to develop a robust system for gender recognition based on deep learning techniques using images of the periocular region taken under various illumination, resolution, and occlusion conditions. An overview of the proposed method is shown in Fig. 3. First, preprocessing procedures were implemented; the eye was localized, and the face was normalized. Then, the periocular region as the region of interest (ROI) was extracted. This ROI was captured in two ways: as the entire periocular region and as the region cropped into left-and right-side patterns. Next, the ROI was passed to the CNN model to classify the gender class.
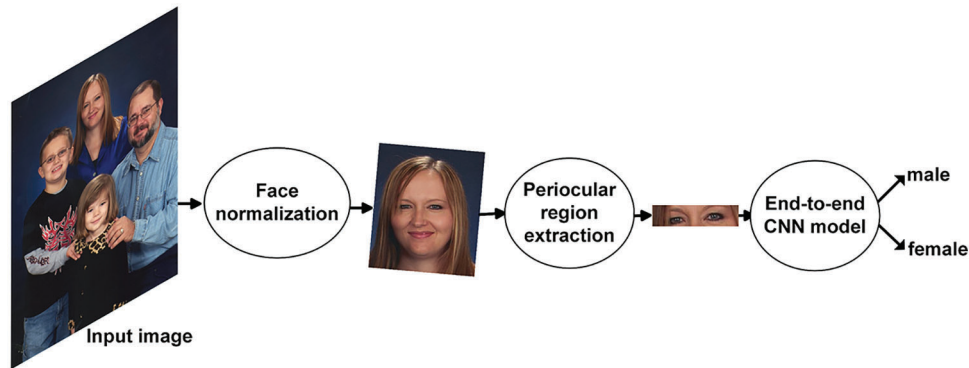

Figure 3: The framework of the periocular gender classifier; sample from the GROUPS database

3.1 Data Preprocessing

Prior to extracting the periocular region, the face must be normalized in terms of scale and rotation. The procedures for face normalization and periocular region extraction are explained in subsequent subsections.

3.1.1 Face Normalization

The face images were aligned using eye center coordinates because they are easily recognizable against other points, such as nose or mouth landmarks, which may be occluded by artifacts, such as scarfs, medical masks, or eyeglass frames. We rotated the face images horizontally based on the line connecting the eye centers, as demonstrated in [30,31]. An illustration of the face normalization procedure is shown in Fig. 4. This process was repeated for each subject in the image.
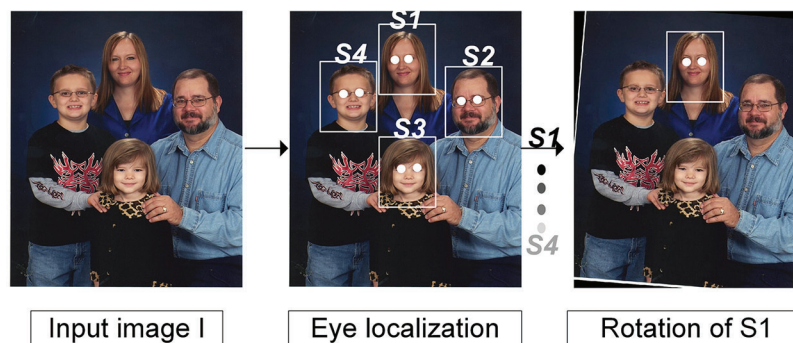

Figure 4: The procedure for face normalization. S_1, S_2, S_3, and S_4 are subjects in the image

3.1.2 Periocular Region Extraction

Two approaches were used to extract the periocular ROI from a face image. First, the ROI was extracted and cropped into separate left and right images. The cropping technique used was similar to the one presented

in [32]. This technique computes the inter-pupillary distance (d_{ip}) between the centers of the eyes when the visual axis is horizontally aligned (Fig. 5a). Point O between the eyes was chosen as the origin for the periocular region frame. Upper (b_u), lower (b_d), left (b_l), and right (b_r) bounds were computed according to the following equations:

$$b_u = O_y - (p_1 \times d_{ip}) \tag{1}$$

$$b_d = O_y + (p_2 \times d_{ip}) \tag{2}$$

$$b_l = O_x - (p_3 \times d_{ip}) \tag{3}$$

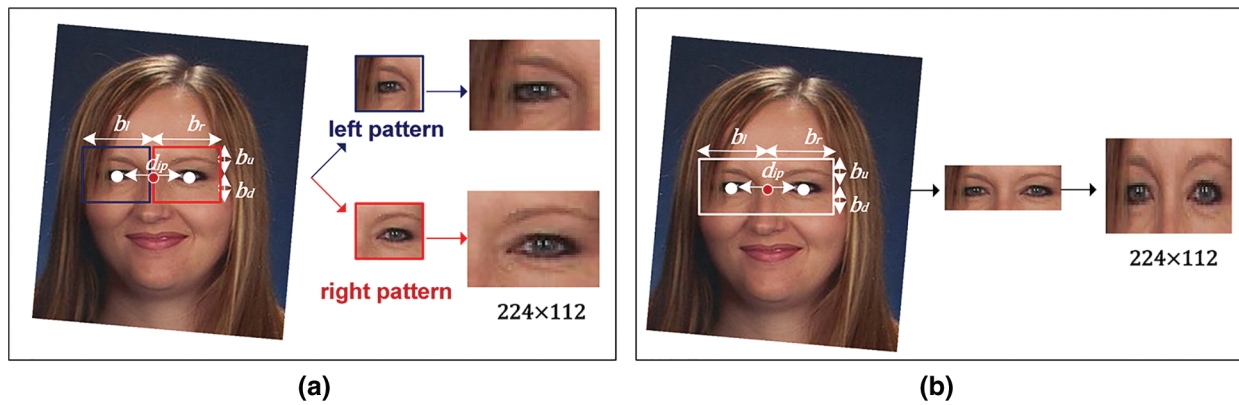$$b_r = O_x + (p_4 \times d_{ip}) \tag{4}$$

Figure 5: The procedure for extracting periocular regions (a) Extracted left and right regions. First, O, represented by a red circle, and d_{ip} (50 pixels for this sample) were calculated. Second, the left and right periocular regions were cropped. Third, ROIs were resized to 224×112 pixels (b) The extracted whole periocular region

The coefficients p_i, $i = 1, 2, 3, 4$; determine the size of the periocular region. Their values were determined empirically ($p_1 = 0.40$, $p_2 = 0.35$, $p_3 = 1$, and $p_4 = 0$) for the left region and ($p_1 = 0.40$, $p_2 = 0.35$, $p_3 = 0$, and $p_4 = 1$) for the right region. Since the extracted periocular regions differed in size from face to face, the image for each region was resized to 224×112 pixels to maintain the shape of the periocular area, including the eyes.

The second approach extracted the whole periocular region in the same way without cropping, as shown in Fig. 5b. The coefficients that determine the whole periocular region are $p_1 = 0.40$, $p_2 = 0.35$, $p_3 = 1$, and $p_4 = 1$.

3.2 CNN Model for Periocular Recognition-PeriGender

3.2.1 Motivation

Motivated by the stellar performance of deep learning [33,34] and CNNs [14], we designed a CNN model for gender recognition (PeriGender) using ResNet-18 [35] as a backbone model. As the features in the periocular region exist at different scales, analysis of the multiscale features is important for better discrimination. The receptive field of neurons increases as we go deeper into the architecture of a CNN model, which means that neurons in shallow layers learn small-scale features, and those in deeper layers learn large-scale features. Thus, in PeriGender, we fused multiscale features using skip connections. The skip connections also provide the added advantage of enhancing the network's learning speed and

improving the flow of information to avoid a vanishing gradient. The idea of skip connections was inspired by the philosophy of DenseNet [36], which reuses features to enhance classification performance. The multiscale features are fused using concatenation. The small-scale features encode texture patterns, whereas large-scale features encode the shape of the eye and periocular area; in this way, the multiscale features help distinguish males from females.

3.2.2 Model Architecture

The architecture of PeriGender consists of four residual groups $Res\_G^i(F, C), i = 1, 2, 3, 4$. Each Res_G^i is composed of two residual blocks $Res\_B^j(F, C), j = 1, 2,$ as shown in Fig. 6b. Each $Res\_B^j(F, C)$ consists of a convolutional layer (conv), batch normalization (BN), and a ReLU layer, as shown in Fig. 6a. We fused multiscale features after the first convolutional layer and each residual block, except for the last, by adding a skip connection $Skip^i(F)$. Each $Skip^i(F)$ consists of a 1×1 convolutional layer with three filters followed by an $F \times F$ max pooling layer (maxpool) with stride F, as shown in Fig. 6c. We unified the depth of features at different scales into three feature maps. The multiscale features were fused, adding a depth concatenation layer and resulting in $3 \times 7 \times 524$. The concatenation layer was added to extract a better feature representation for periocular GC. Then, the fused features were passed to the global average pooling and a fully connected layer with two neurons because there were two classes: male and female. The overall structure of the architecture is presented in Fig. 6d. Tab. 2 shows the architecture and detailed information about the model.
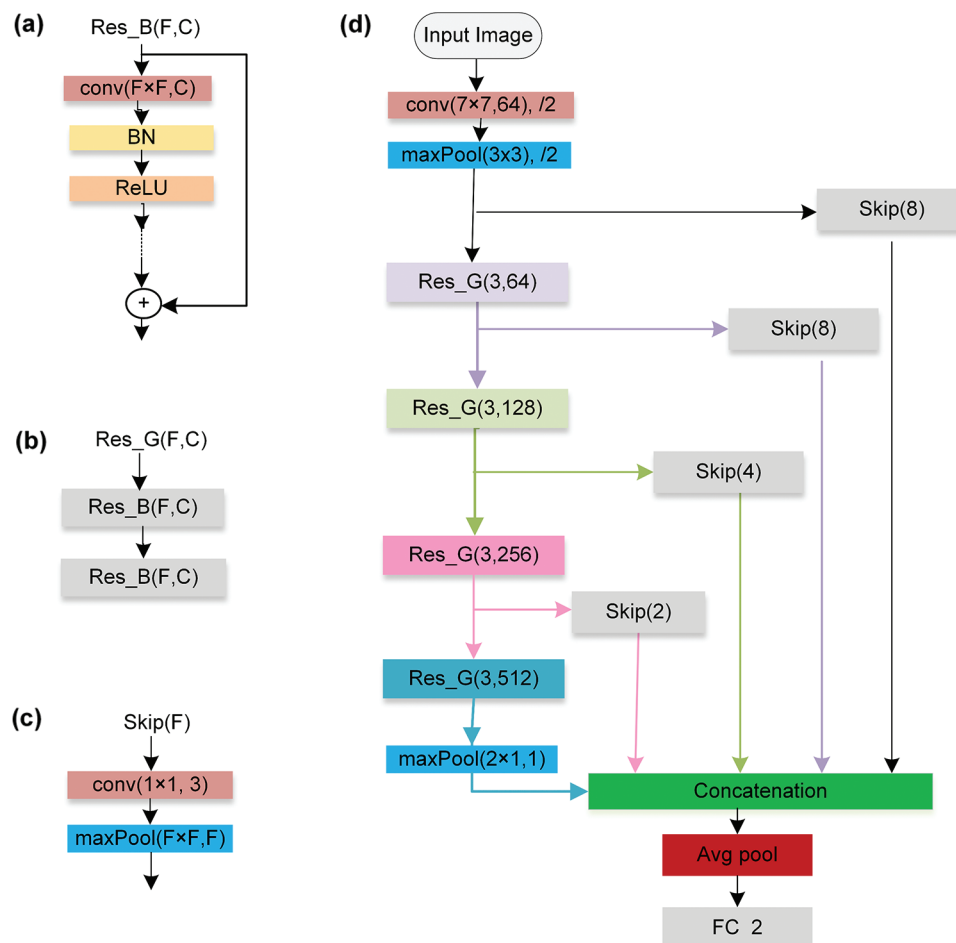

Figure 6: The architecture of PeriGender (a) The architecture of a residual block (b) The architecture of the residual group (c) The architecture of a skip module (d) The overall structure of the network architecture with four skip modules and four residual groups

Table 2: The proposed model architecture

Layer name	Output size	Architecture
Image Input	$112 \times 224 \times 3$	
Convolution	$56 \times 112 \times 64$	7×7, 64, stride 2
Pool	$28 \times 56 \times 64$	3×3 max, stride 2
Skip Module (1)	$28 \times 56 \times 3$	1×1 conv, 3
	$3 \times 7 \times 3$	8×8 max, stride 8
Residual group (1)	$28 \times 56 \times 64$	$\begin{bmatrix} 3 & \times & 3, & 64 \\ 3 & \times & 3, & 64 \end{bmatrix} \times 2$
Skip Module (2)	$28 \times 56 \times 3$	1×1 conv, 3
	$3 \times 7 \times 3$	8×8 max, stride 8
Residual group (2)	$14 \times 28 \times 128$	$\begin{bmatrix} 3 & \times & 3, & 128 \\ 3 & \times & 3, & 128 \end{bmatrix} \times 2$
Skip Module (3)	$14 \times 28 \times 3$	1×1 conv, 3
	$3 \times 7 \times 3$	4×4 max, stride 4
Residual group (3)	$7 \times 14 \times 256$	$\begin{bmatrix} 3 & \times & 3, & 256 \\ 3 & \times & 3, & 256 \end{bmatrix} \times 2$
Skip Module (4)	$7 \times 14 \times 3$	1×1 conv, 3
	$3 \times 7 \times 3$	2×2 max, stride 2
Residual group (4)	$4 \times 7 \times 512$	$\begin{bmatrix} 3 & \times & 3, & 512 \\ 3 & \times & 3, & 512 \end{bmatrix} \times 2$
Pool	$3 \times 7 \times 512$	2×1 max, stride 1
Concatenation	$3 \times 7 \times 524$	Depth concatenation
Classification layer	$1 \times 1 \times 524$	Global Average Pool
	$1 \times 1 \times 2$	Fully connected, SoftMax

To design PeriGender, we used the backbone model ResNet-18 [35] pre-trained on the ImageNet dataset. Training from scratch requires a huge amount of data, which was not available. Transfer learning was efficient in our case and required less data and time than training from scratch.

3.2.3 Architecture Selection and the Effect of Fusion

To determine which CNN model would be most suitable as the backbone model for the system, we concentrated on popular off-the-shelf CNN architectures that have demonstrated competitiveness against the ImageNet Large Scale Visual Recognition Challenge (ILSVRC) [37]. We carried out many experiments with different CNN architectures, including ResNet-18 [35], ResNet-50 [35], and DenseNet-201 [36]. We validated them on several scenarios that ranged from the simplest to the most challenging; the results are shown in Tab. 3. When we used deeper models with more layers and more learnable parameters, such as ResNet50, for the GC problem, the accuracy declined because the problem was a binary classification problem, not a complicated problem. ResNet-18 was chosen as the backbone model because it outperformed the other models. Additionally, DenseNet-201 achieved a GC score of 97.32% for the scenario without occlusions. We integrated important design principles to classify discriminative

gender features in the periocular region. As shown in Tab. 3, the proposed design that fused the multiscale features enhanced GC performance.

Table 3: Accuracy of gender classification using CNN models trained end-to-end on the GROUPS

CNN	GENDER CLASSIFICATION				
	D	E	F	G	H
ResNet-18 [35]	93.92	65.58	89.33	94.64	96.35
ResNet-50 [35]	93.33	64.86	89.08	93.65	93.43
DenseNet-201 [36]	93.92	63.04	89.67	93.85	97.32
PeriGender: Skip contain Avg pool	94.67	63.41	89.83	93.85	94.16
PeriGender: Skip contain Max pool	94.42	66.67	89.33	93.45	95.13
PeriGender: Skip contain (conv + Avg pool)	94.50	65.94	90.67	**96.03**	95.86
PeriGender: Skip contain (conv + Max pool)	**95.08**	**68.84**	**91.67**	95.83	**97.57**

The proposed model was implemented using MATLAB (2021b) on different datasets. The model's training was carried out using a computer with an Intel i9 core with a 3.6 GHz processor, 46 GB RAM, and an NVIDIA GeForce RTX 2080Ti (11 GB) GPU. The model was trained for 20 epochs with a batch size of 64. The optimizer used was SGDM, with an initial learning rate of 0.01 and a momentum of 0.9. The z-score was used to normalize the input layer.

4 Datasets and Evaluation Protocol

4.1 Datasets

We used benchmark databases that represent real-world settings to evaluate the effectiveness of the periocular gender recognition system. We evaluated the proposed method on the facial dataset GROUPS and periocular datasets UBIPr, UFPR, IMP, and Ethnic-Ocular.

4.1.1 Images of Groups (the GROUPS Dataset) [38]

Biometric classification problems were studied under challenging real-world conditions in terms of illumination, pose variations, occlusion, facial expressions, and distance from the camera. The GROUPS contains 28,231 labeled faces among 5,080 images. We used the evaluation protocol defined by Casas et al. [39]. The protocol is subject-disjoint for the GC problem. We also used scenarios employed in the literature on facial recognition systems to investigate the effects of various factors, such as occlusion, aging, eyebrows, and low resolution, as shown in Tab. 4. Scenario A followed Casas et al.'s protocol [39], which uses a predetermined 5-fold CV strategy. It covered a subset of faces whose inter-pupillary distance in the original images was larger than 20 pixels, discarding other low-resolution faces. Scenario A differed somewhat from Scenario B. The faces that were detected manually, as annotated in the provided metadata, were eliminated in Scenario B, keeping only auto-detected faces as in [9] for a fair comparison. Both scenarios ensured subject-disjoint in the training and test sets to enable the assessment of the performance of the proposed system in realistic conditions. In C and C*, we followed A and B, respectively, but split the periocular region into the left and right ocular areas.

Table 4: Specifications of the GROUPS dataset scenarios. The letters M and F in the last column stand for male and female samples, respectively

Code	Scenario Description	Images	M: F
A	Based on [39], 5-fold CV	14,760	1476:1476
	Interocular distance > 20, no cropping of eyes	2952/fold	
B	Based on [9], 5-fold CV	12,530	1253:1253
	Automatically detected faces, no cropping of eyes	2506/fold	
C	Same as A, with cropping eyes	29,520	7380:7380
C*	Same as B, with cropping eyes	25,060	6,262:6,262
D	Same as A, adults only (> 20 years)	12,004	5964:6040
E	Same as A, young and Children only (<= 20 years)	2756	1416:1340
F	Same as D, absence of eyebrows	12,004	5964:6040
G	Same as D, exclusion of periocular images (resolution < 63 × 26)	5043	2499:2544
H	Same as G, without glasses or sunglasses	4103	1897:2206
I	Same as G, only with glasses or sunglasses	940	602:338
J	Same as G, without dark sunglasses	4973	2463:2510
L	The whole dataset	28,231	13,672:14,559

Scenarios D and E focused on the effects of aging on gender features; D covered only wild periocular images featuring adults more than 20 years old, whereas Scenario E included periocular pictures of those 20 and younger. Scenario F investigated the effects of the eyebrows. The resolutions of some images were too low due to their small size and lack of clear details. In addition, images of extracted periocular regions after scaling could be dramatically degraded, blurry, and very noisy. Since they affected the recognition performance, we considered the periocular areas whose original size was 63×26 and up in Scenario G. This size range was selected because it would show the smallest clear resolution of the periocular region. The limitation of images of extremely low resolution mitigated challenges in the experiments. Scenarios H, I, and J dealt with the occlusion factor. H involved the same samples in G, except for images of adults wearing accessories, such as sunglasses or glasses; it included periocular region images partially occluded by hats or hair. Experiment I aimed to study the system's performance under full occlusion with glasses or sunglasses only. We derived experiment J from G using a subset containing an eye covered with medical eyeglasses, hair, or a hat to study the effects of both types of occlusions. The last scenario L contained the whole dataset, including images that were too difficult to recognize because they were either too blurry or too low a resolution.

4.1.2 UBI Periocular Recognition (UBIPr) [40]

UBIPr was the first periocular database developed for periocular research. The images were captured under different distances, illumination conditions, poses, and occlusion levels. Some images suffered from occlusion, either by glasses, hair, or pigmentation levels.

4.1.3 IIIT-D Multispectral Periocular (IMP) [41]

The IMP database features periocular images taken in three spectra: visible (VIS), near-infrared (NIR), and night vision (NV). In this study, only the subset containing VIS periocular images was considered. We followed the same gender labeling as that conducted by the research group in [16].

4.1.4 UFPR-Periocular (UFPR) [19]

UFPR is a large-scale color periocular dataset that was launched in 2020. It contains biometric metadata, such as gender and age.

4.1.5 Ethnic-Ocular [42]

It is an unconstrained dataset designed based on multi-ethnic groups. The images feature celebrities, politicians, and athletes with large variations in pose, illumination levels, appearance, and occlusions. The details of all periocular datasets used are summarized in Tab. 5. Please note that the letter w refers to the whole periocular region.

Table 5: Summary of periocular datasets

Dataset	Images	Subjects (M:F)	Sample	Eyes (M:F)	Region
UBIPr	10,252	342 (345:97)	15	5126 pairs (3671:1455)	l, r
IMP	620	62 (35:27)	5	310 pairs (175:135)	l, r, w
UFPR	33,660	1,122 (602:520)	15	16,830 pairs (9030:7800)	l, r, w
Ethnic-Ocular	170,808	1034	variable	85,404 pairs (53,758:31,646)	l, r

4.2 Evaluation Method

To evaluate the system's performance, 5-and 10-fold CV strategies were utilized for different image sets. The average performance was computed over folds using standard metrics [3], such as overall ACC and correct classification rates for males (MCR), females (FCR) [14], precision (Prec), F1-score (F1), and g-mean (GM).

5 Experimental Results

We performed experiments using the OW protocol, which ensured subject-disjoint, to develop our model so that there was no overlap between the training and test sets. The 5-and 10-fold CV techniques were used to perform the experiments, and the classic validation protocol specified for the dataset was also implemented. In addition, cross-database experiments were carried out to assess performance under realistic conditions.

5.1 Intra-Database Experiments

5.1.1 Classic Validation Strategy

We validated the model using a 90/10 split of the training/test datasets for the CW protocol on the GROUPS scenarios; the results are shown in Tab. 6. In the UFPR, we applied the CW and OW protocols specified in [19]. We followed the original subject-disjoint protocol from [42] for the Ethnic-Ocular dataset. Fig. 7a depicts the confusion matrices for Scenarios B and H from the GROUPS dataset in the CW protocol, while Fig. 7b shows the confusion matrices for the UFPR dataset in the CW and OW protocols.

Table 6: The results of PeriGender based on the 90/10 split strategy

Dataset	Protocol	Region	ACC (%)	MCR	FCR	Prec	F_1	GM
GROUPS A		w	92.14	92.82	91.46	91.58	92.19	92.14
B		w	92.17	91.85	92.49	92.44	92.15	92.17

(Continued)

Table 6 (continued)

| Dataset	Protocol	Region	ACC (%)	MCR	FCR	Prec	F_1	GM
C		l	90.38	88.48	92.28	91.97	90.19	90.36
		r	89.70	89.30	90.11	90.03	89.66	89.70
C*		l	90.26	89.94	90.58	90.51	90.22	90.26
		r	89.78	90.26	89.30	89.40	89.83	89.78
D	CW	w	95.08	95.13	95.03	94.97	95.05	95.08
E		w	68.84	70.42	67.16	69.44	69.93	68.77
F		w	91.67	90.60	92.72	92.47	91.53	91.65
G		w	95.83	95.20	96.46	96.36	95.77	95.83
H		w	97.57	98.42	96.83	96.39	97.40	97.62
I		w	86.17	85.00	88.24	92.73	88.70	86.60
J		w	95.77	95.93	95.62	95.55	95.74	95.78
L		w	89.37	87.86	90.80	89.96	88.90	89.31
UFPR	CW	l	96.93	97.41	96.38	96.89	97.15	96.90
		r	96.74	98.01	95.27	96.00	96.99	96.63
		w	**97.86**	98.87	96.69	97.19	98.02	97.78
	OW	l	93.83	94.71	92.94	93.20	93.95	93.82
		r	93.55	95.31	91.75	92.19	93.72	93.51
		w	**94.90**	95.13	94.58	96.12	95.62	94.85
Ethnic-Ocular	OW	l	**94.14**	98.14	86.59	93.26	95.64	92.18
		r	93.70	98.03	85.53	92.76	95.32	91.56
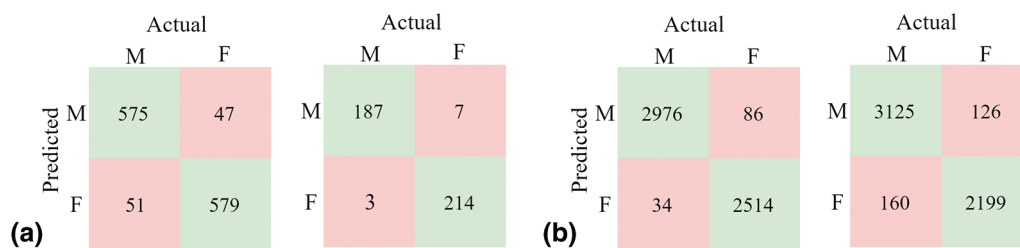
Figure 7: Confusion matrices of the system with (a) the CW protocol on the GROUPS-{B, H} and (b) CW and OW protocols on the UFPR: whole region

5.1.2 5-Fold CV and 10-Fold CV with OW Protocol

We also tested the model using 5-and 10-fold CV with the OW protocol. In the case of 5-fold CV, the data were divided into 5 folds, and 5 experiments were performed in such a manner that every time one fold (20%) was held out for testing, the remaining four folds (80%) were used as the training set. In this way, the performance of the model was tested on 5 different sets to ensure its generalization. The same procedure was applied in the case of 10-fold CV, where the data were divided into 10-folds, and 10 experiments were

performed. The results are shown in Tabs. 7 and 8; Fig. 8 shows the receiver operating characteristic (ROC) curves of applying the model to all GROUPS scenarios. For the UBIPr dataset, we performed two experiments: one on left-eye periocular samples and the other on right-eye samples. Additionally, experiments on the IMP dataset were performed on cropped eyes and the whole periocular region. The performance was better for the right-eye than left-eye images for both the last datasets.

Table 7: The results of PeriGender based on a 5-fold CV

Dataset	Region		ACC (%)				AVG	MCR	FCR	Prec	F$_1$	GM	
			fold1	*fold2*	*fold3*	*fold4*	*fold5*						
GROUPS	A	W	86.89	88.89	88.14	87.77	87.40	87.82	87.80	87.83	87.84	87.82	87.82
	B	W	86.63	88.31	87.19	87.37	87.35	87.37	87.36	87.37	87.39	87.37	87.36
	C	l	84.96	86.04	85.81	85.54	85.13	85.49	85.65	85.34	85.40	85.52	85.49
		r	85.20	84.99	84.89	85.81	84.69	85.11	85.65	84.58	84.75	85.19	85.11
	C*	l	84.56	85.36	85.63	84.00	84.60	84.83	85.08	84.58	84.66	84.87	84.83
		r	84.28	83.96	84.04	85.47	84.92	84.53	85.16	83.91	84.11	84.63	84.53
	D	W	93.04	92.42	92.96	92.17	92.13	92.54	92.86	92.24	92.19	92.52	92.55
	E	W	68.30	66.97	66.97	66.06	63.16	66.29	68.93	63.51	66.62	67.75	66.16
	F	W	86.13	86.38	86.38	86.76	85.54	86.24	87.11	85.38	85.47	86.28	86.24
	G	W	92.67	91.28	90.98	92.26	89.98	91.43	93.36	89.54	89.77	91.53	91.43
	H	W	93.54	92.20	92.94	92.32	93.05	92.81	93.31	92.38	91.33	92.31	92.84
	I	W	81.91	78.72	79.79	81.91	82.45	80.96	90.03	64.79	82.00	85.83	76.38
	J	W	93.17	91.46	92.66	92.66	90.04	92.00	92.33	91.67	91.58	91.95	92.00
	L	W	85.94	85.81	85.57	86.08	84.63	85.60	85.15	86.03	85.13	85.14	85.59
UBIPr		l	94.49	92.44	93.23	95.59	97.75	**94.70**	96.93	87.85	95.61	96.25	92.21
		r	95.17	93.31	96.66	95.10	95.59	**95.17**	97.84	89.07	95.35	96.55	93.33
IMP		W	95.38	95.38	100	100	98.33	97.82	99.43	95.87	96.75	98.05	97.61
		l	96.92	92.31	98.33	93.33	96.67	**95.51**	99.43	90.40	93.24	96.17	94.75
		r	95.38	95.38	100	100	95.00	97.15	98.86	95.20	96.26	97.49	96.98

Table 8: The results of PeriGender based on a 10-fold CV

Dataset	Region		ACC (%)										AVG	MCR	FCR	Prec	F$_1$	GM
			fold1	*fold2*	*fold3*	*fold4*	*fold5*	*fold6*	*fold7*	*fold8*	*fold9*	*fold10*						
GROUPS	A	w	86.92	86.65	86.92	88.96	87.13	87.06	89.50	86.72	89.70	87.20	87.68	88.24	87.11	87.26	87.75	87.67
	B	w	86.67	87.71	87.23	87.23	85.32	86.75	88.03	87.39	87.47	89.39	87.32	87.58	87.06	87.12	87.35	87.32
	C	l	85.37	84.69	85.09	86.45	86.31	86.11	84.55	85.91	84.82	86.11	85.54	86.15	84.93	85.11	85.63	85.54
		r	86.86	86.59	85.30	84.35	84.49	85.50	83.67	86.25	85.98	85.16	85.41	85.96	84.86	85.03	85.49	85.41
	C*	l	83.40	84.84	83.64	87.55	84.76	85.95	85.00	86.35	85.47	86.19	85.32	85.97	84.66	84.86	85.41	85.31
		r	85.47	85.32	86.59	84.84	84.36	85.32	85.71	85.71	84.84	87.15	85.53	85.76	85.30	85.37	85.56	85.53

Table 8 (continued)

| Dataset	Region	ACC (%)										AVG	MCR	FCR	Prec	F₁	GM
		fold1	fold2	fold3	fold4	fold5	fold6	fold7	fold8	fold9	fold10						
	D w	92.26	92.76	92.67	91.92	91.92	92.83	92.00	92.17	93.00	92.83	92.44	92.91	91.97	91.95	92.43	92.44
	E w	63.77	66.30	67.39	61.96	71.01	67.75	69.82	66.18	66.18	66.18	66.65	68.15	65.07	67.34	67.74	66.59
	F w	86.34	86.26	86.18	87.18	86.08	85.50	86.83	85.00	86.92	85.92	86.22	86.97	85.48	85.54	86.25	86.22
	G w	94.65	91.49	90.50	92.46	92.06	91.87	90.08	92.26	91.67	91.67	91.87	92.80	90.96	90.98	91.88	91.87
	H w	93.67	92.94	93.67	94.39	93.41	93.41	92.68	92.93	90.49	94.39	93.20	93.94	92.57	91.57	92.74	93.25
	I w	79.79	80.85	80.85	82.98	78.72	87.23	82.98	85.11	89.36	82.98	83.09	89.53	71.60	84.88	87.15	80.07
	J w	91.16	92.57	88.76	92.56	92.96	92.15	91.95	93.16	93.36	90.54	91.92	92.29	91.55	91.47	91.88	91.92
	L w	86.11	86.47	85.48	85.90	85.69	87.39	85.65	86.18	85.90	86.65	86.14	85.85	86.41	85.58	85.72	86.13
IMP	w	100	91.43	100	100	100	100	100	100	100	100	**99.14**	100	98	98.70	99.30	98.94
	l	100	91.43	100	100	93.33	96.67	93.33	90.00	93.33	90.00	94.81	93.83	96.00	96.38	94.94	94.83
	r	100	97.14	100	100	100	96.67	96.67	100	100	96.67	**98.71**	98.83	98.67	98.75	98.75	98.73
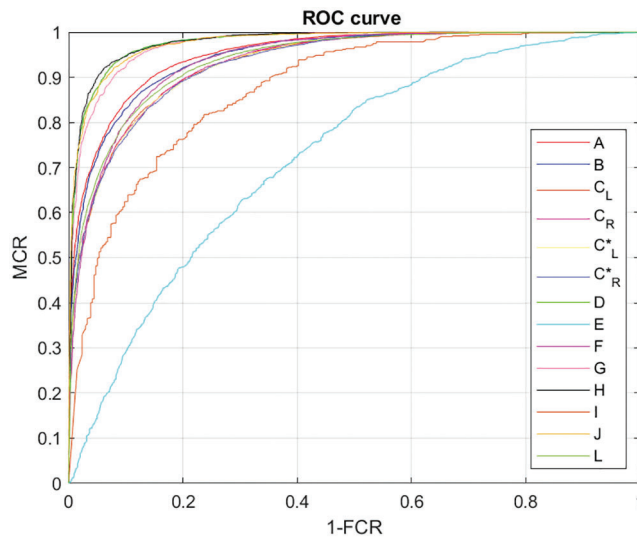
Figure 8: ROC curves of all GROUPS subsets using 5-fold CV

5.2 Cross-Database Experiments

To investigate PeriGender's generalizability, we performed cross-database experiments with the following details:

1) For the left and right periocular regions, the model was trained on scenario C (based on Casas' protocol) for the GROUPS database and tested on the UBIPr, UFPR, and Ethnic-Ocular databases as testing samples. Second, for the whole periocular region, the entire GROUPS database was used to train the model, and it was tested on the IMP and UFPR databases.

2) All images for each region from the UBIPr database were used to train the model, and they were then tested on the GROUPS-C, UFPR, and Ethnic-Ocular databases.

3) All images for each region from the UFPR database were used to train the model. The trained model was then tested on the GROUPS-C, UBIPr, and Ethnic-Ocular databases. Additionally, the evaluation was performed on the whole periocular region using the UFPR as a training set and the GROUPS-{A, H, L} and IMP datasets as test sets.

4) All images of the Ethnic-Ocular database for each region were used to train the model. The trained model was then tested on the GROUPS-C, UBIPr, and UFPR databases.

The IMP dataset has a small number of images, which is insufficient for training the neural network. Therefore, it was not considered for training. The results of the cross-database experiments are presented in Tab. 9. The results of the self-evaluation based on the 10-fold CV are also given for comparison. Generalization reduction is a common problem due to various factors. Despite this phenomenon, the generalization power of the model trained by the UFPR dataset was good for the UBIPr dataset, achieving an accuracy of 92.92%. Our model achieved high generalization scores on UBIPr as a testing dataset, reaching scores of 91.61%, and 91.05%, as trained by the UFPR and GROUPS datasets, respectively. It also achieved good results on the IMP dataset, reaching 92.26%. Failures occurred for several reasons:

Table 9: Generalization results (%) of PeriGender method for GC on five databases

Test	Cropped Periocular Regions							
	Left				Right			
Train	{C}	UBIPr	UFPR	Ethnic-Ocular	{C}	UBIPr	UFPR	Ethnic-Ocular
GROUPS-{C}	**85.54**	91.05	81.29	82.66	**85.41**	89.74	80.89	78.62
UBIPr	62.97	**94.70**	77.89	66.22	61.02	**95.17**	79.86	64.25
UFPR	68.24	91.61	**93.83**	78.96	71.34	92.92	**93.55**	69.53
Ethnic-Ocular	71.40	80.75	74.21	**94.14**	71.02	78.68	74.34	**93.70**

Test	Whole Periocular Region W				
Train	{A}	{H}	{L}	IMP	UFPR
GROUPS-{A}	**87.82**	-	-	89.35	83.20
GROUPS-{H}	-	**93.20**	-	91.29	84.89
GROUPS-{L}	-	-	**86.14**	87.42	82.89
UFPR	71.54	81.50	65.10	92.26	**94.90**

- The performance of the model was affected by different image settings and environmental conditions, such as variations in illumination and camera sensor differences, in the UFPR dataset against the GROUPS and the Ethnic-Ocular datasets.
- Some misclassified cases suffered from excessively narrow cropped eye regions, leading to the absence of the eyebrows and eyelids, as shown in Fig. 9a. For instance, we noticed that most images in the Ethnic-Ocular dataset were cropped tightly and did not contain eyebrows, as shown in Fig. 9b. The model did not learn to extract deep gender information from the eyebrow region, which is important for GC. This explains the decrease in accuracy when tested on the GROUPS or UFPR datasets.
- Blinking blocks the semantic eye features extracted by the model, as shown in Fig. 9c.
- The image resolution was severely degraded and blurred, preventing the model from recognizing region components, as in Fig. 9d.

- The periocular region was fully occluded by dark sunglasses, making it difficult to recognize gender information, as shown in Fig. 9e.
- There was difficulty distinguishing between masculine and feminine facial features among newborns, which mimics limitations in human ability, as shown in Fig. 9f.

For these reasons, generalization accuracy was usually lower than the intra-database evaluation results.
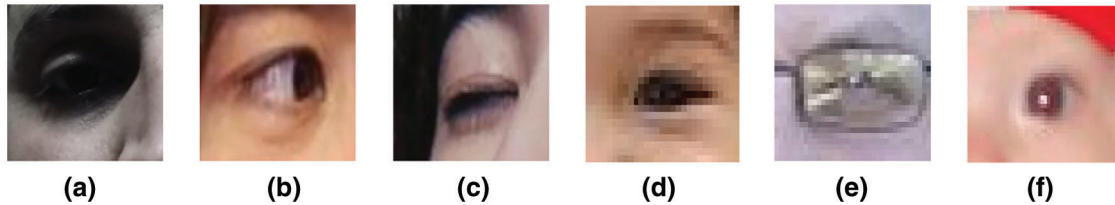

| (a) | (b) | (c) | (d) | (e) | (f) |

Figure 9: Misclassified male cases: (a) UBIPr *vs.* UFPR, (b) GROUPS *vs.* Ethnic-Ocular, (c) UBIPr *vs.* Ethnic-Ocular, (d) UBIPr *vs.* GROUPS, (e) UBIPr *vs.* GROUPS, and (f) UBIPr *vs.* GROUPS

6 Discussion

In this section, the experimental results are discussed and compared to those derived using state-of-the-art methods. PeriGender performed better when the whole periocular region was assessed rather than when it was split into left and right regions, as shown by the results reported in Tabs. 6, 7 and 8 on GROUPS-{A} and GROUPS-{C}, Therefore, we evaluated the method using the whole periocular region for other scenarios from the GROUPS database. Moreover, adjusting the size of the periocular area to 224×112 instead of 224×224 is a technique that has been commonly used to preserve aspect ratios for eye shape. The results shown in Tab. 10 provide further support for our hypothesis.

Table 10: The results on the GROUPS-{C} subsets of different sizes

Dataset	Region	Size	ACC(%)	MCR	FCR	Prec	F_1	GM
GROUPS-{C}	l	224×112	**85.17**	85.95	84.39	84.65	85.29	85.16
	r		**85.11**	85.65	84.58	84.75	85.19	85.11
	l	224×224	84.92	85.76	84.08	84.34	85.04	84.91
	r		84.87	85.62	84.12	84.35	84.98	84.87

The efficiency of the method was evaluated for images with occlusions. Set G contains occluded images featuring a variety of accessories (sunglasses or glasses). These images also include fully occluded regions with accessories or regions partially occluded by hair, hats, or masks. Despite all these occluded images, the model achieved an accuracy of 91.87%, less than the accuracy of Set H, which was 93.20%. Note that the contents of H is the same as G, excluding accessories. The proposed model achieved a 91.92% accuracy with set J, which contains the same occluded images as G but without dark sunglasses. This shows that the classification error rate decreased only slightly by 0.05%, indicating that the model was robust to occlusion. When we trained the proposed model with only images featuring glasses-induced occlusions in I, the model achieved 83.09% accuracy. The reason behind this poor performance is that the model was only trained on images wholly covered with dark glasses, preventing the learning of the periocular regions' structure. As a result, it failed to detect gender discriminative features, which led to unacceptable

results. In examining the effects of aging on gender recognition, we note that clear improvement was reached at 92.54% accuracy for the OW protocol and 95.08% for the CW protocol in Experiment D, which addressed images of adults (>20 years old) only. Whereas only youths and children (≤20 years old) were considered in Experiment E, the gender recognition accuracy dramatically declined to 66.65% for the OW protocol and 68.84% for the CW protocol. The statistics indicate that the younger a person (under the age of 20) is, the less visible the gender features around the eyes are.

In addition, we aimed to study the effects of the absence of eyebrows on gender recognition capabilities since the eyebrows are the main feature of the periocular region. For this reason, Experiment F was conducted. F includes ocular images of adults as in D but excludes images featuring eyebrows. As shown in Tab. 8, the results showed a significant reduction in recognition accuracy, which increased the classification error by roughly 6% (92.44% *vs.* 86.22%) for D and F, respectively. This indicator emphasizes that eyebrows are rich in gender discriminative information. This finding matches that observed in an earlier study [16].

To determine which features the proposed model focuses on during the GC task, we employed Grad-CAM visualization [25]. Fig. 10 shows the results; the model concentrated on prominent regions during the GC task. For males, the middle region between the eyes was significant, indicating the importance of using the whole periocular region instead of splitting it. The use of the entire periocular region allowed the model to learn the global structure of both eyes without discontinuity. Additionally, we noticed that the model focused more on the inner corners of the eyes. However, the regions around the eyebrows, upper eyelids, and corners of the eyes were distinctive for females. The model is robust to low-resolution images, as shown in Fig. 10a (right), and partial occlusions, such as hair and eyeglasses, as shown in Fig. 10b. In Fig. 10c (left), the misclassified child samples revealed that the ocular regions of children lack gender characteristics, which dramatically affected the model's performance in the experiments on subset {E}, which yielded a 66.65% accuracy. As seen in Fig. 10c (middle), the absence of eyebrows affected the learning and classification scores. This indicates the role of eyebrows in the classification of females because they contain gender-distinctive features. In Scenario I from the GROUPS dataset, the accuracy decreased due to periocular regions being fully covered by sunglasses, as shown in Fig. 10c (right). This hindered the work of the system, which was built to analyze the components of this hidden area. Therefore, this is not considered a defect of the system. For this reason, we developed different scenarios for the GROUPS dataset, such as H and J, to validate system performance using the appropriate dataset.
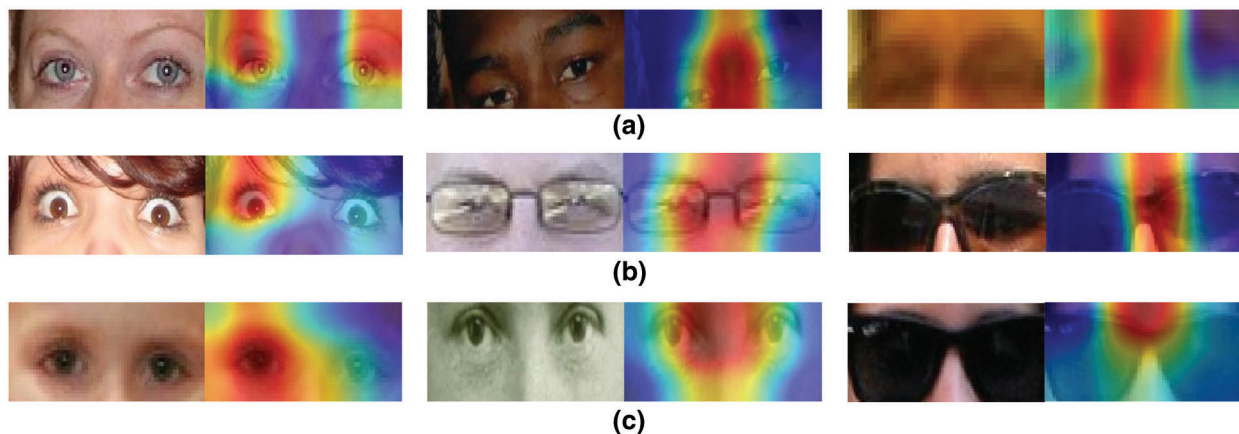

(a)

(b)

(c)

Figure 10: Grad-CAM visualization of most discriminative gender regions on the GROUPS dataset (a) Correctly classified samples: female, male, and male with low resolution (b) Correctly classified samples: female, male, and male with different occlusions (c) Misclassified samples: child male, female without eyebrows, and fully occluded female

To our knowledge, two studies, [9,10] have tested gender recognition based on handcrafted features on the GROUPS dataset using only the periocular region. The results of the proposed system, specifically for Scenario B, were compared because they followed the same protocol. PeriGender outperformed previous methods in terms of accuracy by 5%, reaching 87.37% *vs.* 82.91% accuracy for the OW protocol and 92.17% for the CW protocol. The results are presented for comparison in Tab. 11, which shows some of the state-of-the-art periocular methods reviewed, datasets, and accuracy results. Emulating reality, we performed experiments on the IMP and UFPR datasets based on the OW protocol to address GC problems without subject overlap between the training and test sets. This enhanced the generalization of the model, allowing it to learn discriminative representations for new subjects. However, the proposed method achieved higher accuracy for GC in the CW than in the OW scenarios for both whole and cropped images. This variance was possibly due to the model's learning of identity and gender characteristics. As can be seen from Tab. 11, PeriGender achieved higher classification accuracy when the whole periocular region was used than when the cropped region in the middle was used. Thus, using the whole region better facilitates the understanding of both eyes' global structure. PeriGender outperformed the existing method [16] by a large margin, exhibiting 9.4% and 13.94% ACC improvement for the left-eye and right-eye samples, respectively.

Table 11: Comparison of PeriGender with the state-of-the-art periocular methods

Paper	Dataset	Method	Protocol	ACC (%)		
				l	r	w
Santana et al. [9]	GROUPS	HOG	OW	-	-	82.91
Santana et al. [10]		WLD + LSP + LTP	OW	-	-	83.41
Proposed		PeriGender	OW	85.32	85.53	**87.37**
			CW	90.26	89.78	92.17
Manyala et al. [16]	IMP	VGG-Face	OW	86.11	84.77	-
Proposed		PeriGender	OW	**95.51**	**98.71**	99.14
Zanlorensi et al. [19]	UFPR	MobileNetV2	CW	-	-	97.80
Proposed		PeriGender	CW	96.93	96.74	**97.86**
			OW	93.83	93.55	94.90

For the CW protocol, PeriGender performed better than Zanlorensi et al.'s method [19] for the whole periocular region on the UFPR dataset. We validated the model on all UFPR images without omitting any existing images. The proposed model also worked well on the UBIPr dataset, achieving 94.70% and 95.17% accuracy. The proposed system outperformed custom techniques, such as HOGs, LBPs, and scale-invariant feature transform (SIFT), with logistic regression and MLP classifiers, as depicted by the ROC curves for the left and right regions [40]. To our knowledge, no prior studies have classified gender by evaluating their systems on the Ethnic-Ocular database, but we predict that it will play a role in future research due to the huge amount of and diversity in data.

7 Conclusion

In this study, we dealt with the problem of GC from images taken in the wild (i.e., in an uncontrolled environment). The images suffered from different illumination, resolution, and occlusion problems caused by

unconstrained conditions. Most existing methods were designed based on conventional techniques that do not learn the nature of data. They also require cumbersome manual tuning of parameters, and their performance depends on the designer's expertise in selecting hyperparameters. Most existing methods have been tested on controlled datasets and are thus not applicable in unconstrained environments. We built a robust classification system called PeriGender based on deep learning, which learns the hierarchy of features and fuses the features on multiple scales through skip connections. The architecture is based on residual learning to increase information exchange and overcome the degradation problem.

Particularly exciting is the finding that compared to other regions, the middle region between the eyes plays a crucial role in the recognition of masculine features. Another important finding is that feminine traits can be explicitly identified by the eyebrows, upper eyelids, and corners of the eyes. These findings suggest that developing a GC system that uses a whole periocular region without cropping enhances learning capabilities, fostering the understanding of both eyes' global structure. The system was validated using the wild datasets GROUPS, Ethnic-Ocular, UFPR, IMP, and UBIPr. The results showed that the proposed deep learning-based solution excels compared to other existing approaches. Therefore, the proposed method can address other classification problems, such as ethnicity or age, based on the periocular region. Especially during the Covid-19 pandemic, when people are required to wear masks everywhere, PeriGender can be used in a real environment for optical sorting at the entrances to buildings. The model can be combined with identification systems using other modalities to increase efficiency in future work.

Conflicts of Interest: The authors declare that they have no conflicts of interest to report regarding the present study.

References

[1] C. B. Ng, Y. H. Tay and B. M. Goi, "A review of facial gender recognition," *Pattern Analysis and Applications*, vol. 18, no. 4, pp. 739–755, 2015.

[2] S. S. Liew, M. K. Hani, S. A. Radzi and R. Bakhteri, "Gender classification: A convolutional neural network approach," *Turkish Journal of Electrical Engineering & Computer Sciences*, vol. 24, no. 3, pp. 1248–1264, 2016.

[3] M. Ngan and P. Grother, "Face recognition vendor test (FRVT) performance of automated gender classification algorithms," *US Department of Commerce, National Institute of Standards and Technology*, NISTIR 8271, 186, 2015. https://doi.org/10.6028/NIST.IR.8271.

[4] Beautiful Free Images & Pictures, Unsplash, 2021. [Online]. Available: https://unsplash.com/.

[5] Free Stock Photos, Royalty Free Stock Images, Pexels, 2021. [Online]. Available: https://www.pexels.com/.

[6] F. A. Fernandez and J. Bigun, "A survey on periocular biometrics research," *Pattern Recognition Letters*, vol. 82, no. 2, pp. 92–105, 2016.

[7] D. Bobeldyk and A. Ross, "Iris or periocular? exploring sex prediction from near infrared ocular images," in *Int. Conf. of the Biometrics Special Interest Group (BIOSIG)*, Darmstadt, Germany, 2016.

[8] E. Brown and D. I. Perrett, "What gives a face its gender," *Perception*, vol. 22, no. 7, pp. 829–840, 1993.

[9] M. C. Santana, J. L. Navarro and E. R. Balmaseda, "On using periocular biometric for gender classification in the wild," *Pattern Recognition Letters*, vol. 82, no. 2, pp. 181–189, 2016.

[10] M. C. Santana, J. L. Navarro and E. R. Balmaseda, "Multi-scale score level fusion of local descriptors for gender classification in the wild," *Multimedia Tools and Applications*, vol. 76, no. 4, pp. 4695–4711, 2017.

[11] D. L. Woodard, K. Sundararajan, N. Tobias and J. Lyle, "Periocular-based soft biometric classification," *Iris and Periocular Biometric Recognition*, vol. 5, no. 1, pp. 197, 2017.

[12] A. Rattani, N. Reddy and R. Derakhshani, "Gender prediction from mobile ocular images: A feasibility study," in *IEEE Int. Symp. on Technologies for Homeland Security (HST)*, Waltham, MA, USA, 2017.

[13] C. W. Kao, H. H. Chen, B. J. Hwang, Y. J. Huang and K. C. Fan, "Gender classification with jointing multiple models for occlusion images," *Journal of Information Science & Engineering*, vol. 35, no. 1, pp. 105–123, 2019.

[14] A. Rattani, N. Reddy and R. Derakhshani, "Convolutional neural networks for gender prediction from smartphone-based ocular images," *Iet Biometrics*, vol. 7, no. 5, pp. 423–430, 2018.

[15] J. Tapia and C. Aravena, "Gender classification from periocular NIR images using fusion of CNNs models," in *IEEE 4th Int. Conf. on Identity, Security, and Behavior Analysis (ISBA)*, Singapore, 2018.

[16] A. Manyala, H. Cholakkal, V. Anand, V. Kanhangad and D. Rajan, "CNN-Based gender classification in near-infrared periocular images," *Pattern Analysis and Applications*, vol. 22, no. 4, pp. 1493–1504, 2019.

[17] J. Tapia, C. Arellano and I. Viedma, "Sex-classification from cellphones periocular iris images," in *Selfie Biometrics*, Springer, Cham, Switzerland, pp. 227–242, 2019.

[18] A. Bhattacharyya, R. Saini, P. P. Roy, D. P. Dogra and S. Kar, "Recognizing gender from human facial regions using genetic algorithm," *Soft Computing*, vol. 23, no. 17, pp. 8085–8100, 2019.

[19] L. A. Zanlorensi, R. Laroca, D. R. Lucio, L. R. Santos, A. S. B. Jr et al., "UFPR-Periocular: A periocular dataset collected by mobile devices in unconstrained scenarios," arXiv:2011.12427, 2020.

[20] F. A. Fernandez, K. H. Diaz, S. Ramis, F. J. Perales and J. Bigun, "Soft-biometrics estimation in the era of facial masks," in *Int. Conf. of the Biometrics Special Interest Group (BIOSIG)*, Darmstadt, Germany, 2020.

[21] F. A. Fernandez, K. H. Diaz, S. Ramis, F. J. Perales and J. Bigun, "Facial masks and soft-biometrics: Leveraging face recognition CNNs for age and gender prediction on mobile ocular images," arXiv:2103.16760, 2021.

[22] G. Antipov, M. Baccouche, S. A. Berrani and J. L. Dugelay, "Effective training of convolutional neural networks for face-based gender and age prediction," *Pattern Recognition*, vol. 72, no. 1, pp. 15–26, 2017.

[23] F. J. Xu, E. Verma, P. Goel, A. Cherodian and M. Savvides, "DeepGender: Occlusion and low resolution robust facial gender classification via progressively trained convolutional neural networks with attention," in *Proc. of the IEEE Conf. on Computer Vision and Pattern Recognition Workshops*, Las Vegas, NV, USA, 2016.

[24] J. Merkow, B. Jou and M. Savvides, "An exploration of gender identification using only the periocular region," in *Fourth IEEE Int. Conf. on Biometrics: Theory, Applications and Systems (BTAS)*, Washington, DC, USA, 2010.

[25] R. R. Selvaraju, M. Cogswell, A. Das, R. Vedantam, D. Parikh et al., "Grad-CAM: Visual explanations from deep networks via gradient-based localization," in *Proc. of the IEEE Int. Conf. on Computer Vision*, Venice, Italy, 2017.

[26] A. Rattani, R. Derakhshani, S. K. Saripalle and V. Gottemukkula, "ICIP 2016 competition on mobile ocular biometric recognition," in *IEEE Int. Conf. on Image Processing (ICIP)*, Phoenix, AZ, USA, 2016.

[27] F. H. B. Zavan, O. R. P. Bellon, L. Silva and G. G. Medioni, "Benchmarking parts based face processing in-the-wild for gender recognition and head pose estimation," *Pattern Recognition Letters*, vol. 123, no. 1, pp. 104–110, 2019.

[28] A. Dhomne, R. Kumar and V. Bhan, "Gender recognition through face using deep learning," *Procedia Computer Science*, vol. 132, no. 1, pp. 2–10, 2018.

[29] M. C. Santana, J. L. Navarro and E. R. Balmaseda, "Descriptors and regions of interest fusion for in-and cross-database gender classification in the wild," *Image and Vision Computing*, vol. 57, no. 1, pp. 15–24, 2017.

[30] V. Struc, *The PhD face recognition toolbox*, MATLAB Central File Exchange, 2022. [Online]. Available: https://www.mathworks.com/matlabcentral/fileexchange/35106-the-phd-face-recognition-toolbox.

[31] V. Struc and N. Pavesic, "The complete gabor-fisher classifier for robust face recognition," *EURASIP Journal on Advances in Signal Processing*, vol. 2010, no. 1, pp. 1–26, 2010.

[32] V. Struc, J. Krizaj and S. Dobrisek, "MODEST face recognition," in *3rd Int. Workshop on Biometrics and Forensics (IWBF 2015)*, Gjovik, Norway, 2015.

[33] X. Zhang, X. Sun, W. Sun, T. Xu, P. Wang et al., "Deformation expression of soft tissue based on BP neural network," *Intelligent Automation and Soft Computing*, vol. 32, no. 2, pp. 1041–1053, 2022.

[34] W. Sun, L. Dai, X. Zhang, P. Chang and X. He, "RSOD: Real-time small object detection algorithm in UAV-based traffic monitoring," *Applied Intelligence*, vol. 51, pp. 1–16, 2021.

[35] K. He, X. Zhang, S. Ren and J. Sun, "Deep residual learning for image recognition," in *Proc. of the IEEE Conf. on Computer Vision and Pattern Recognition*, Las Vegas, NV, USA, 2016.

[36] G. Huang, Z. Liu, L. V. D. Maaten and K. Q. Weinberger, "Densely connected convolutional networks," in *Proc. of the IEEE Conf. on Computer Vision and Pattern Recognition*, Honolulu, HI, USA, 2017.

[37] O. Russakovsky, J. Deng, H. Su1, J. Krause, S. Satheesh *et al.,* "ImageNet large scale visual recognition challenge," *International Journal of Computer Vision*, vol. 115, no. 3, pp. 211–252, 2015.

[38] A. C. Gallagher and T. Chen, "Understanding images of groups of people," in *IEEE Conf. on Computer Vision and Pattern Recognition*, Miami, FL, USA, 2009.

[39] P. D. Casas, D. G. Jimenez, L. L. Yu and J. L. A. Castro, "Single-and cross-database benchmarks for gender classification under unconstrained settings," in *IEEE Int. Conf. on Computer Vision Workshops (ICCV Workshops)*, Barcelona, Spain, 2011.

[40] C. N. Padole and H. Proenca, "Periocular recognition: Analysis of performance degradation factors," in *5th IAPR Int. Conf. on Biometrics (ICB)*, New Delhi, India, 2012.

[41] A. Sharma, S. Verma, M. Vatsa and R. Singh, "On cross spectral periocular recognition," in *IEEE Int. Conf. on Image Processing (ICIP)*, pp. 5007–5011, Paris, France, 2014.

[42] L. C. O. Tiong, Y. Lee and A. B. J. Teoh, "Periocular recognition in the wild: Implementation of RGB-OCLBCP dual-stream CNN," *Applied Sciences*, vol. 9, no. 13, pp. 2709, 2019.