

Deep Ear Biometrics for Gender Classification



Ritwiz Singh, Keshav Kashyap, Rajesh Mukherjee, Asish Bera,
and Mamata Dalui Chakraborty

Abstract Human gender classification based on biometric features is a major concern for computer vision due to its vast variety of applications. The human ear is popular among researchers as a soft biometric trait, because it is less affected by age or changing circumstances and is non-intrusive. In this study, we have developed a deep convolutional neural network (CNN) model for automatic gender classification using the samples of ear images. The performance is evaluated using four cutting-edge pre-trained CNN models. In terms of trainable parameters, the proposed technique requires significantly less computational complexity. The proposed model has achieved 93% accuracy on the EarVN1.0 ear dataset.

Keywords Convolutional neural network (CNN) · Gender classification · Ear biometric · Soft biometrics

1 Introduction

Gender identification has become a major concern due to its vast variety of applications, including social communication and connection, commercial visual supervision, banking transactions, illness prognosis, demographic data collection, artificial

R. Singh · K. Kashyap · R. Mukherjee (✉)

Department of Computer Science and Engineering, Haldia Institute of Technology, Haldia, WB, India

e-mail: rajeshmukherjee.cse@gmail.com

A. Bera

Department of Computer Science and Information Systems, Birla Institute of Technology and Science Pilani, Pilani Campus, Rajasthan 333031, India

e-mail: asish.bera@pilani.bits-pilani.ac.in

M. D. Chakraborty

Department of Computer Science and Engineering, National Institute of Technology, Durgapur, West Bengal, India

e-mail: mamata.dalui@cse.nitdgp.ac.in

© The Author(s), under exclusive license to Springer Nature Singapore Pte Ltd. 2023

521

D. K. Sarkar et al. (eds.), *Proceedings of the 4th International Conference on Communication, Devices and Computing*, Lecture Notes in Electrical Engineering 1046, https://doi.org/10.1007/978-981-99-2710-4_42

intelligence (AI)-based user interface for customization, consumer analysis for business growth, and many more [17, 35]. Biometric traits have proven their suitability in gender classification as they are non-intrusive, remain invariant with time, and are less influenced by emotions and circumstances. Several approaches have been proposed using various biometrics traits and their fusion. Biometrics can be broadly categorized into the conventional (aka hard) biometrics (e.g., face, hand, etc.) [5, 18] and soft biometrics (e.g., ear, gait, etc.) [26, 34].

In classifying human gender, based on the iris samples, Bansal and Sharma [2] have employed wavelet transforms and statistical methods. A technique to identify the gender of complete body photographs using part-based gender recognition is proposed in [7]. To determine the dialect and gender from parlance data, supervised machine learning techniques have been employed [13]. Tapia et al. [30] used the local binary patterns (LBP) and Histograms of Gaussian (HOG) descriptors for gender identification from iris images. Gait-based gender categorization methods are suggested by Yu et al. [36] and Li et al. [22]. Shan [28] has used real-world facial pictures and the LBP description to categorize gender. A novel patch-based LBP with adjustable weights is presented by Chen and Jeng [9] for the classification of gender on the person's face.

Moreover, a human ear is an unique evidence among the existing biometric traits for gender identification. The anatomy of the human ear is regarded to be equally significant as compared to other biometric traits like the face, iris, and gait, for identifying an individual. Additionally, it is less affected by aging when compared to other biometric traits like the face [16] and gait [22]. Other biometric traits may change with age or changing circumstances, but the ear simply scales up. Numerous methods for ear identification have been proposed. Anwar et al. [1] use geometrical information from human ear images for identification. Various image descriptors have been studied including LPQ, BSIF, LBP, HOG, and surface descriptors in [27]. Gender classification from ear images has also been studied in [11, 12, 19, 34].

However, the effectiveness of ConvNet for gender identification is yet to be researched thoroughly. In this study, we suggest a basic yet successful CNN-based model that is computationally lightweight. For the experiment, we have considered the EarVN1.0 large-scale ear dataset and achieved significantly good accuracy.

2 Related Work

Existing works on gender identification can be divided into three types: appearance-based approaches [3, 4, 7, 14, 28, 33], biological feature-based approaches [6, 8, 29, 31, 32], and interpersonal information-based approaches [10, 23, 25]. External traits of an individual, such as the face, eye brows, hands, fingernails, attire, and footwear, are focused on an appearance-based approach. The biological feature-based techniques make use of biometric characteristics (e.g., fingerprint, ear, iris, etc.) and

biosignals to determine gender (ECG, EEG, etc.). The interpersonal information-based strategy uses email, blogs, handwriting, and other types of interpersonal information.

Nowadays, the research community has recently demonstrated a growing interest in biological feature-based techniques approaches since unique biological information of individuals does not change over long periods of time, is non-intrusive, is less influenced by emotions, and is appropriate for applications requiring real-time recognition. [20]. Because of various distinctive features, the ear has gained increased interest in biometric research for identification. Several works on gender categorization based on ear biometrics have been conducted. In [12], the authors considered measurements between the ear hole within the ear hole and seven ear features computed from the masked ear image for identification. Over 342 samples, they examined KNN, Bayes, and neural networks for classification accuracy. KNN performed the best, with a classification accuracy rate of 90.42%. Histogram intersection kernel is employed to classify profile face and ear images in [37]. To increase accuracy, they used Bayesian analysis to achieve score-level fusion. The trials were carried out using 2D images. Face and ear attained 95.43% and 91.78% perfection, respectively, while fusion achieved 97.68% precision. Yaman et al. [35] shows that much higher accuracy in gender identification (99%) can be achieved through fusion with facial profiles. Authors in [19] suggested a sparse representation-based method for identifying gender from ear images. They added Gabor filters for feature extraction to enhance accuracy. The authors in [21] have investigated the use of 3D ear images for gender identification. The authors have exploited curvature and indexed shapes of ears for the classification. In [34], authors with the aid of geometrical and appearance-based features identified age, gender, and other biometric traits. A methodology to identify gender and kinship using geometric aspects of ear pictures has been developed in [24]. The proposed model has an accuracy of 67.2% in gender classification.

In this study, a straightforward, but, effective CNN model is proposed and tested for gender classification using the EarVN1.0 [15] ear dataset, and performance is evaluated in contrast with four cutting-edge models.

3 Methodology

A novel CNN-based architecture is proposed for gender detection using ear images to increase classification accuracy and reduce model size. The proffered model's architecture is portrayed in Fig. 1, and Table 1 displays the output shape and parameters for each layer of the proposed model. Let, the input color image is $X \in I^{h \times w \times 3}$, where h , w , 3 represent height, width, and the number of channels. A sequel of convolutional extraction blocks $C(\sum_n \omega_n \cdot x_n + b)$ (where ω_n , x_n , and b are weight, input, and bias at n th layer) is employed for the extraction of deep feature map $FM = C(X) \in I^{h \times w \times c}$ from X . The resultant feature map is then split into two channels, one for global average pooling Pool_{avg} (GAP) and another for global max pooling Pool_{max}

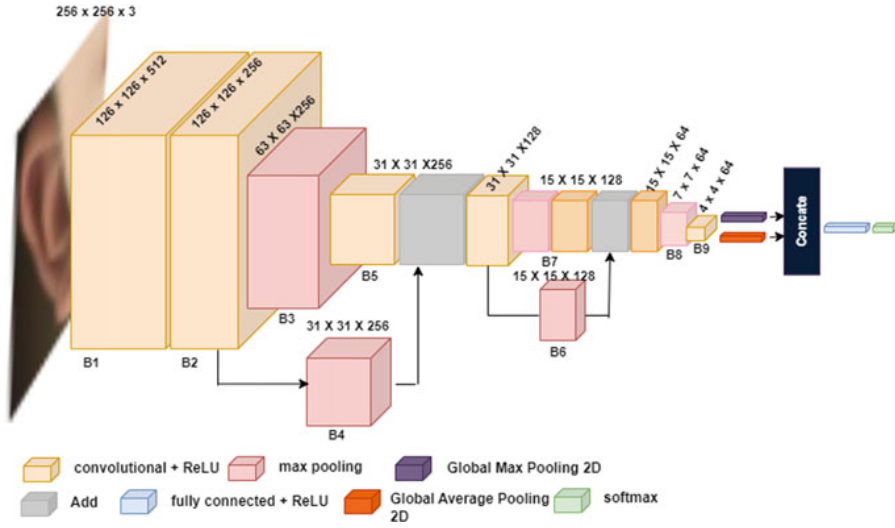


Fig. 1 Proposed deep model for ear biometrics

(GMP), in order to decrease the spatial aspects by identifying key features along both the pathways, $FM_a = \text{Pool}_{\text{avg}}(FM)$ and $FM_m = \text{Pool}_{\text{max}}(FM)$. It retains translation invariance while providing a summary of the feature maps. Then, a high-level feature map is created by fusing these two feature maps $FM_c = FM_a \oplus FM_m$, where \oplus represents linear concatenation function. And lastly, the Softmax estimates the likelihood (ρ) of gender classification from the input image, $\rho[0, 1] = \text{softmax}(FM_c)$. '0' is designated as the 'female' class in the binary classification issue, while '1' is designated as the 'male' class.

$$FM = C(X); \text{ where } FM \in I^{h \times w \times c}; X \in I^{h \times w \times 3}; \text{ and } C\left(\sum_n \omega_n \cdot X_n + b\right) \text{ is a CNN,} \quad (1)$$

$$FM_a = \text{Pool}_{\text{avg}}(FM) \text{ and } FM_m = \text{Pool}_{\text{max}}(FM), \quad (2)$$

$$FM_c = [FM_a \oplus FM_m]; \text{ and } \rho = \text{softmax}(FM_c). \quad (3)$$

Implementation: The proposed model comprises a sequel of convolution extraction blocks to extract the feature maps. Next, GAP and GMP are applied independently on the extracted feature map (Fig. 1). Finally, Softmax is applied to calculate output probabilities by concatenating the two acquired feature maps. To optimize the training job, binary cross-entropy (γ) is employed as a loss function.

Table 1 Layer outputs and parameters of proposed model

Blks	Layer (type)	Output shape	Parameter
	Input layer	(None, 256, 256, 3)	0
B1	Conv 2D 1	(None, 126, 126, 512)	38,912
B2	Conv 2D 2	(None, 126, 126, 256)	1,179,904
B3	Maxpooling 2D 1	(None, 63, 63, 256)	0
B4	Maxpooling 2D 2	(None, 31, 31, 256)	0
B5	Conv 2D 3	(None, 31, 31, 256)	590,080
	Add 1	(None, 31, 31, 256)	0
	Conv 2D 4	(None, 31, 31, 128)	295,040
B6	Maxpooling 2D 3	(None, 15, 15, 128)	0
B7	Maxpooling 2D 4	(None, 15, 15, 128)	0
	Conv 2D 5	(None, 15, 15, 128)	65,664
	Add 2	(None, 15, 15, 128)	0
	Conv 2D 6	(None, 15, 15, 64)	73,792
B8	Maxpooling 2D 5	(None, 7, 7, 64)	0
B9	Conv 2D 6	(None, 4, 4, 64)	36,928
GAP	GlobalAvgPooling 2D	(None, 64)	0
GMP	GlobalMaxPooling 2D	(None, 64)	0
Concat	Concatenate	(None, 128)	0
Dense	Dense	(None, 2)	258

$$\gamma = \frac{1}{n} \sum_{t=1}^k -[y_{t.} \cdot \log(P_t) + (1 - y_{t.}) \cdot \log(1 - P_t)], \tag{4}$$

where P_t and $(1 - P_t)$ denote the likelihood of female and male classes, respectively. y_t and \bar{y}_t indicate the actual and predicted class labels, and n represents the number of data samples. With a learning rate of 0.001, the Adaptive Moment Estimation (ADAM) optimizer is employed. The model is trained for 100 epochs with 32 mini-batch size.

The input shape is $256 \times 256 \times 3$, and also, the suggested model is trained over 100 epochs in 32-batch iterations. The data augmentation process is used to increase the data’s size and diversity by including ± 0.2 horizontal flips and ± 0.2 rotations. The output feature dimension (FM) is $4 \times 4 \times 64$. Pooling reduces the size of a feature map from $w \times h \times c$ to $1 \times 1 \times c$. Then, the feature map is further reduced to 64 ($FM_a = FM_m = 64$) by using the GAP and GMP layers in two distinct channels. Finally, feature maps from both channels (GAP and GMP) are concatenated ($FM_c = 128$). To avoid the overfitting problem, a dropout rate of 0.2 is used in the experiment as regularization.

4 Experiments and Result

In this section, we present the outcomes of the proposed model. In order to compare the performance of the proposed model, we have used four base models in which four CNNs, namely, VGG19, Xception, ResNet50, and MobileNet have been used as the backbone and tuned with the target dataset. For faster learning convergence, pre-trained ImageNet weights are employed in the considered base models. Features are extracted from the last layers of the backbone CNNs, and then, subsequent dense layer is applied for the classification. To assess the model's performance, we have used the EarVN1.0 [15] dataset. The dataset contains the large-scale ear images of 164 subjects and a total of 28,412 images.

For assessment, the Receiver Operator Characteristic (ROC) curve is used. AUC, which serves as a summary of the ROC curve, calculates the capacity to differentiate between the classes. In order to discriminate between the male and female categories, a higher AUC is indicative of greater performance. The accuracy (%) and AUC (%) metrics are employed. Model parameters of the CNN models are expressed in millions (M). Furthermore, both accuracy and recall are taken into account in order to properly assess the efficacy of the proposed framework. Higher recall demonstrates a higher likelihood that a specific category may be identified, whereas higher precision suggests the efficiency of a model used to predict a certain category.

The experiments are conducted on the Kaggle notebook platform equipped with an NVIDIA Tesla T4 GPU. We have tested and compared the performance of all four state-of-the-art pre-trained models and the proposed model on the EarVN1.0 [15] ear dataset with a 70:30 train–test ratio, and the results are displayed in Table 2. The accuracy indicates that our method outperforms four well-known state-of-the-art classification models, as stated above. Additionally, it should be highlighted that the computational complexity of the proposed model is considerably lower than the state-of-the-art models. We achieved 93% accuracy in the proposed model on the EarVN1.0 dataset, whereas only VGG-19 and MobileNet ensure the maximum 85% accuracy among the examined state-of-the-art models. The confusion matrix is used to assess a model's utility and envision its performance. In a confusion matrix, diagonal values correspond to predicted labels that are identical to actual labels, whereas non-diagonal values correspond to classifier erroneously classified labels. Figure 2 depicts the confusion matrix and training–testing accuracy during the training process for state-of-the-art models. Figure 3 depicts the same for the proposed model. The test results imply that our proposed deep architecture performs better than state-of-the-art computationally heavy models in the context of precision and recall metrics.

Feature Map Representation: Figure 4 displays the feature maps from various layers of the proposed model. The low-level features depict an ear shape with earlier levels having more distinct details. On the other hand, the network pipeline's deeper levels depict high-level characteristics with fewer details.

Table 2 Efficacy of the considered base models along with the proposed model

CNN model	Female		Male		Accuracy	Model parameters (M)
	Precision	Recall	Precision	Recall		
VGG-19	86	73	84	93	85	20.07
ResNet50	67	31	68	89	68	23.59
Xception	78	83	89	85	84	20.86
MobileNet	83	77	86	91	85	3.23
Proposed model	92	89	93	95	93	2.46

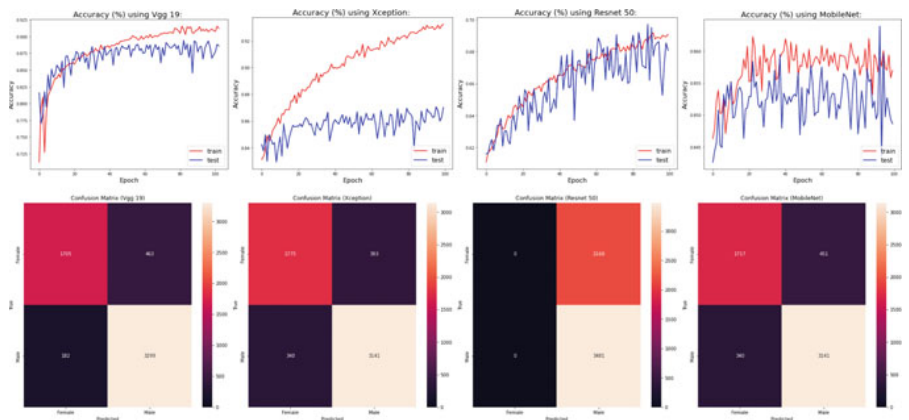


Fig. 2 Accuracy and confusion matrix of VGG-19, Xception, ResNet-50, and MobileNet (left to right)

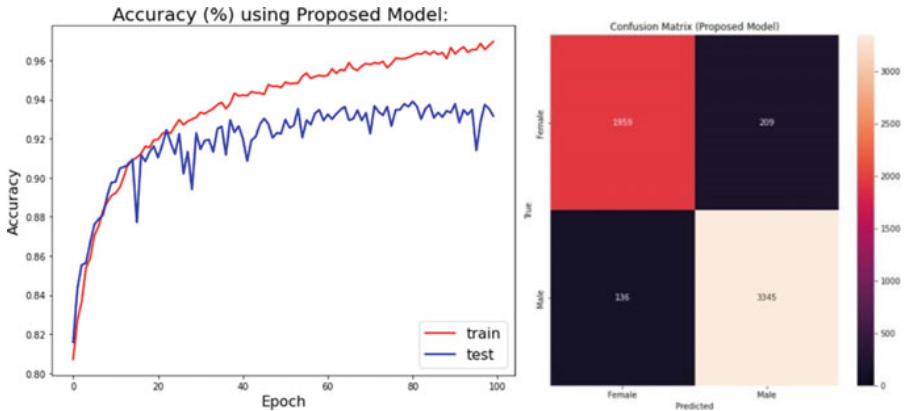


Fig. 3 Gender recognition performance of the proposed model

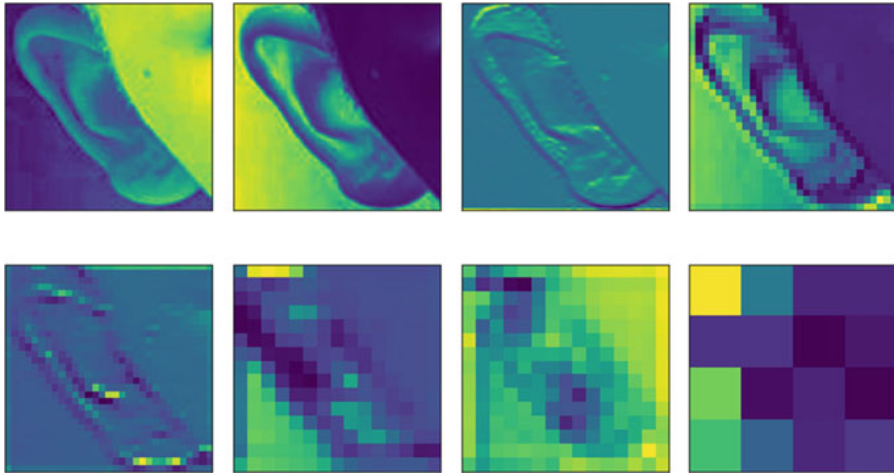


Fig. 4 Feature map representation of proposed model at different convolutional layers

5 Conclusion

In this paper, a simple and novel deep CNN method is devised for gender classification using ear biometric samples. The proposed model is tested on EarVN1.0 ear dataset. The proposed model outperforms four computationally demanding state-of-the-art model CNNs, namely VGG-19, Xception, ResNet-50, and MobileNet. Thus, the proposed approach makes the gender categorization more practical. We intend to develop a new deep model for soft biometric features and a fusion of them, to broaden the method's usefulness for real-world applications.

References

1. Anwar AS, Ghany KKA, Elmahdy H (2015) Human ear recognition using geometrical features extraction. *Proc Comput Sci* 65:529–537
2. Bansal A, Agarwal R, Sharma R (2012) SVM based gender classification using iris images. In: 2012 fourth international conference on computational intelligence and communication networks. IEEE, pp 425–429
3. Basha AF, Jahangeer GSB (2012) Face gender image classification using various wavelet transform and support vector machine with various kernels. *Int J Comput Sci Issues (IJCSI)* 9(6):150
4. Bekios-Calfa J, Buenaposada JM, Baumela L (2014) Robust gender recognition by exploiting facial attributes dependencies. *Pattern Recogn Lett* 36:228–234
5. Bera A, Bhattacharjee D (2017) Human identification using selected features from finger geometric profiles. *IEEE Trans Syst Man Cybernet: Syst* 50(3):747–761
6. Bera A, Bhattacharjee D, Nasipuri M (2014) Person recognition using alternative hand geometry. *Int J Biomet* 6(3):231–247
7. Cao L, Dikmen M, Fu Y, Huang TS (2008) Gender recognition from body. In: Proceedings of the 16th ACM international conference on Multimedia, pp 725–728

8. Carrier J, Land S, Buysse DJ, Kupfer DJ, Monk TH (2001) The effects of age and gender on sleep EEG power spectral density in the middle years of life (ages 20–60 years old). *Psychophysiology* 38(2):232–242
9. Chen WS, Jeng RH (2020) A new patch-based lbp with adaptive weights for gender classification of human face. *J Chin Inst Eng* 43(5):451–457
10. Corney M, De Vel O, Anderson A, Mohay G (2002) Gender-preferential text mining of e-mail discourse. In: 18th Annual computer security applications conference, 2002. Proceedings. IEEE, pp 282–289
11. Emersic Z, Struc V, Peer P (2017) Ear recognition: more than a survey. *Neurocomputing* 255:26–39
12. Gnanasivam P, Muttan S (2013) Gender classification using ear biometrics. In: Proceedings of the fourth international conference on signal and image processing 2012 (ICSIP 2012). Springer, pp 137–148
13. HaCohen-Kerner Y, Hagege R (2017) Language and gender classification of speech files using supervised machine learning methods. *Cybern Syst* 48(6–7):510–535
14. Hadid A, Pietikainen M (2009) Combining appearance and motion for face and gender recognition from videos. *Pattern Recogn* 42(11):2818–2827
15. Hoang VT (2019) Earvn1. 0: a new large-scale ear images dataset in the wild. *Data in brief* 27
16. Iannarelli A (1989) Forensic identification series: ear identification. Paramount, California 5
17. Kamboj A, Rani R, Nigam A (2022) A comprehensive survey and deep learning-based approach for human recognition using ear biometric. *Vis Comput* 38(7):2383–2416
18. Kar A, Pramanik S, Chakraborty A, Bhattacharjee D, Ho ES, Shum HP (2020) Lmzmpm: local modified zernike moment per-unit mass for robust human face recognition. *IEEE Trans Inf Forens Secur* 16:495–509
19. Khorsandi R, Abdel-Mottaleb M (2013) Gender classification using 2-d ear images and sparse representation. In: 2013 IEEE workshop on applications of computer vision (WACV). IEEE, pp 461–466
20. Kumar A, Wu C (2012) Automated human identification using ear imaging. *Pattern Recogn* 45(3):956–968
21. Lei J, Zhou J, Abdel-Mottaleb M (2013) Gender classification using automatically detected and aligned 3d ear range data. In: 2013 International conference on biometrics (ICB). IEEE, pp 1–7
22. Li X, Maybank SJ, Yan S, Tao D, Xu D (2008) Gait components and their application to gender recognition. In: IEEE transactions on systems, man, and cybernetics, part C (Applications and Reviews) 38(2):145–155
23. Liwicki M, Schlapbach A, Loretan P, Bunke H (2007) Automatic detection of gender and handedness from on-line handwriting. In: Proceedings of 13th conference of the graphonomics society. Citeseer, pp 179–183
24. Meng D, Nixon MS, Mahmoodi S (2019) Gender and kinship by model-based ear biometrics. In: 2019 International conference of the biometrics special interest group (BIOSIG). IEEE, pp 1–5
25. Mukherjee A, Liu B (2010) Improving gender classification of blog authors. In: Proceedings of the 2010 conference on empirical methods in natural language processing, pp 207–217
26. Mukherjee R, Bera A, Bhattacharjee D, Nasipuri M (2021) Human gender classification based on hand images using deep learning. *Tech Rep, EasyChair*
27. Pflug A, Paul PN, Busch C (2014) A comparative study on texture and surface descriptors for ear biometrics. In: 2014 International Carnahan conference on security technology (ICCST). IEEE, pp 1–6
28. Shan C (2012) Learning local binary patterns for gender classification on real-world face images. *Pattern Recogn Lett* 33(4):431–437
29. Shue YL, Iseli M (2008) The role of voice source measures on automatic gender classification. In: 2008 IEEE international conference on acoustics, speech and signal processing. IEEE, pp 4493–4496

30. Tapia J, Viedma I (2017) Gender classification from multispectral periocular images. In: 2017 IEEE international joint conference on biometrics (IJCB). IEEE, pp 805–812
31. Thomas V, Chawla NV, Bowyer KW, Flynn PJ (2007) Learning to predict gender from iris images. In: 2007 First IEEE international conference on biometrics: theory, applications, and systems. IEEE, pp 1–5
32. Tripathy RK, Acharya A, Choudhary SK (2012) Gender classification from ECG signal analysis using least square support vector machine. *Am J Sig Proc* 2(5):145–149
33. Ueki K, Komatsu H, Imaizumi S, Kaneko K, Sekine N, Katto J, Kobayashi T (2004) A method of gender classification by integrating facial, hairstyle, and clothing images. In: Proceedings of the 17th international conference on pattern recognition, 2004. ICPR 2004. vol 4. IEEE, pp 446–449
34. Yaman D, Eyiokur FI, Sezgin N, Ekenel HK (2018) Age and gender classification from ear images. In: 2018 International workshop on biometrics and forensics (IWBF). IEEE, pp 1–7
35. Yaman D, Irem Eyiokur F, Kemal Ekenel H (2019) Multimodal age and gender classification using ear and profile face images. In: Proceedings of the IEEE/CVF conference on computer vision and pattern recognition workshops, pp 0–0
36. Yu S, Tan T, Huang K, Jia K, Wu X (2009) A study on gait-based gender classification. *IEEE Trans Image Process* 18(8):1905–1910
37. Zhang G, Wang Y (2011) Hierarchical and discriminative bag of features for face profile and ear based gender classification. In: 2011 International joint conference on biometrics (IJCB). IEEE, pp 1–8