Dual-Domain Learning for Gender Classification from Ear Images using Vision Transformers and Frequency-Aware CNNs

1st Sanjita Phijam
Computer Science and Engineering
National Institute of Technology Silchar
Silchar, India
phijamsanjk@gmail.com

2nd Aryan Kumar Singh
Computer Science and Engineering
National Institute of Technology Silchar
Silchar, India
aryan21_ug@cse.nits.ac.in

3rd Deep Saikia
Computer Science and Engineering
National Institute of Technology Silchar
Silchar, India
deep21_ug@cse.nits.ac.in

4th Md Shohan Mia
Computer Science and Engineering
National Institute of Technology Silchar
Silchar, India
md.shohan21_ug@cse.nits.ac.in

5th Debbrota Paul Chowdhury
Computer Science and Engineering
National Institute of Technology Silchar
Silchar, India
dchowdhury@cse.nits.ac.in

***Abstract*—The Human Gender Classification using Ear Biometrics is a limited and lesser explored area of Human Gender Classification, but is gaining popularity for its wide variety of applications. The Human Ear is a popular soft-biometric trait to study, for it is a dependable alternative to other facial features. It remains relatively constant throughout human life, unless affected by any cosmetic procedures or injury. In this study, we implement a deep learning-based model for gender classification using ear images, using the AWE and EarVN1.0 dataset, using modern Convolutional Neural Network(CNN) architecture. This proposed model achieved an accuracy of 97.30% on EarVN1.0 dataset and 88.39% on AWE dataset.**

***Index Terms*—Convolutional Neural Network(CNN), Deep Learning, Vision Transformer(ViT), Gender classification, Biometrics, Human Ear.**

I. Introduction

In the area of research in computer vision and pattern recognition, age and gender estimation are gaining popularity for their wide spectrum applications like age-specific human and machine interaction, gender and age estimation in mobile applications, and electronic customer relationship management (ECRM). Age and gender classification serves a critical function in the sale and customization of a wide range of products. The vendor can tailor advertisements based on characteristics and demographics, consequently optimizing the advertisement cost. All online shopping websites, music, and video streaming websites function according to this strategy. Age and gender verification are also an indispensable prerequisite for some forensic validation cases. Consequently, it has emerged as one of the most active areas of computer vision.

Although traditional biometric modalities such as the face, iris, or fingerprint have been extensively utilized for soft-biometric classification, human ears are considered a dependable alternative. Recent research has shown the potential of ear biometrics in gender classification functions. In comparison with other facial features, that are often influenced by factors such as expressions or makeup, the ear retains a relatively constant shape throughout the life of a person and is less impacted by facial expressions, cosmetics, or aging. On classifying human gender and improved verifications using ear samples, [19] and [1] have shown that deep convolutional neural networks (CNNs) can effectively extract gender-discriminative features from ear images, this can achieve significant accuracy on datasets like EarVN1.0. Yaman [22] proposed a broad study using both geometric and presence-based features for age and gender classification from ear images. Their CNN-based approach has achieved up to 94% accuracy for gender prediction, the soft feature shows effectiveness of ear biometrics in classifications. Yaman [23] also suggested a multimodal, multitask deep learning framework in combining with ear and profile face images for age and gender classification, using advanced fusion techniques such as data-level, feature-level, and score-level fusion. Despite these precedents, gender classification using ear images remains an underexplored area as compared to facial analysis. Previous studies have shown that Convolutional Neural networks (CNNs) can achieve committed results in gender classification from ear images, with some models achieving more than 90% accuracy on datasets like EarVN1.0. Furthermore, the introduction of transformer-based models and the use of overlapping patches further enhances the effectiveness of the ear recognition technique, although its application for gender classification is still limited. In unconstrained environments, human recognition based on ear is comprehensively discussed by Authors [10], proposing a deep learning-based model using modified Faster-RCNN for detection and VGG-19 for recognition tasks.

In this study, we propose a deep learning-based model for gender classification using ear images, leveraging the structural uniqueness of ears and the feature-learning capabilities of modern CNN architectures. We aim to develop a model that surpasses the performance of established approaches in both accuracy and generalization. We utilized two benchmark datasets: AWE and EarVN1.0, both known for their diversity in pose, illumination, occlusion, and subject demographics.

This study is organized as follows:
In Section II, related works are discussed. The proposed methodology for gender classification using ear images is discussed in Section III. Section IV shows the experimental results, and finally conclusion is drawn in Section V to summarize the study.

II. Related Works

Age classification using human face was first introduced by Know and Lobo [14]. They classified age based on three classes: baby, young adult, and senior adult. Hayashi *et al.* [8] used a wrinkle-based texture that appeared in the face image to estimate the age and gender of a person. Later, many researches [3], [4], [6], [7], [15], [18], [21], [24] have performed research for classifying age and gender using the shape and texture of the frontal face image. One of the problems in age and gender classification using face images is the unavailability of the frontal face. In that case, profile face is used and classification accuracy degrades due to a lack of key features.

This problem was first addressed by Zhang and Wang [25], using the texture of profile face and ear to classify gender. Features were extracted using a hierarchical and discriminative bag of features, and are classified using support vector classification (SVC). They experimented on the UND-F dataset with 942 profile images of 302 subjects, including 562 male and 380 female images, respectively. Genders are classified with an accuracy of 95.43% and 91.78% using profile face and ear images respectively, and fusion of both modalities yielded an accuracy of 97.65%. In 2013, Gnanasivam and Muttan [5] proposed gender classification methodology, based on the distance between ear identification point and reference point, where ear identification points are ear features: outer lobe edge, outer and inner curves of the helix, outer and inner curves of the antihelix and two edges of the concha, whereas earhole is taken as a reference point. Bayes, K-Nearest Neighbour (KNN), and neural network are used for classification on an internal database having 342 ear samples of bothnmales and females. Among three classifiers, KNN gave a better result, with an accuracy of 90.42%. A dictionary-based feature extraction from ear image is proposed in [12] for gender classification. Local features are extracted using Gabor filters on the UND-J dataset with 2430 images corresponding to 404 subjects. Based on sparse classification, 235 males and 169 females (training: 100 males and 120 females, test: 135 males, 49 females) samples were classified with an accuracy of 89.49% with 128 features. Lei *et al.* [16] proposed an algorithm for gender identification using 3D ear shape. Ear feature is extracted using Histogram of Indexed Shapes (HIS) and for the classification task, Support Vector Machine(SVM) is used. Experiments were conducted on UND-F (942 images) and UND-J (1800 images) datasets, and an accuracy of 92.94 % and 91.92 % was obtained, respectively. Both the soft-biometric traits, age and gender, were classified in [22] using appearance-based methods. It was observed that a few works are there in gender classification using ear images, but [22] did the first work on age detection using ear images. They used a popular CNN model, namely, AlexNet, VGG-16, GoogLeNet, and SqueezeNet for the classification task. These models are fine-tuned using profile images of the Multi-PIE face dataset. A small number of ear images (270) was used for further fine-tuning of the models. They created a dataset of 338 subjects, 188 males and 150 females, with one profile image for each subject. For the age classification task, these subjects are categorized into five distinct classes: 18-28, 29-38, 39-48, 49-58, 59-68+ respectively. GoogLeNet gave the highest accuracy of 94% and 52% for gender and age classification, respectively. In their next work [23], they tried to improve the age and gender classification accuracy using a multi-modal deep neural network frameworks. Around 9% improved classification accuracy is achieved as compared to previous work [22] using the FERET dataset. Considering some of the recent works presented in TABLE I, Authors in [17] used a KNN and Euclidean Distance-based model having 134 images from 67 subjects, resulting in a 67.2% accuracy. Applying the Convolutional Neural Network(CNN) method to increase the accuracy of gender classification, that consisted of 28,412 images from 164 subjects. Authors in [19]trained and tested the EarVN1.0 dataset, in a 70:30 ratio, getting 93% accuracy. Author [20] in this study used a Mask RCNN and Grabcut Segmentation method for gender based ear detection, with features extracted from Gabon filters and Histogram Oriented gradient, then classified by gender using Deep learning and Machine Learning techniques. The proposed models achieved an accuracy of 81.48% and 87.52% for the EarVN1.0 dataset, respectively. Utilising the EarVN1.0 dataset and a hybrid architecture of CNN, ViT and MLP-Mixer, the authors in [13] proposed a method that is of low parameters, trainable quickly, and is state-of-the-art. The model got an accuracy of 96.66%. In [11] authors have taken 508 participants, of which 264 are males and 244 females, within the ages of 18-35 years. They proposed a model PyCaret, that uses a train-evaluate-test method to improve gender classification, and achieved the highest accuracy of 86.75% from Logistic regression classifier. Applying a customised Vision Transformer(ViT) configuration, namely ViT-T(Tiny), ViT-S(Small), ViT-B(Base) and ViT-L(Large), the ear models are trained with the UERC2023 dataset in [1] and tested using OPIB, AWE, WPUT and EarVN1.0 datasets for improved gender classifications in ear biometrics, getting accuracies of 91.96%, 94.89%, 92.78% and 78.38% respectively. Table I gives a quick overview of recent studies using ear images for gender classification. It shows the methods, datasets, and how well each approach performed, helping to highlight progress in this area.

Author	Detection Method	Dataset	Dataset Size	# Train Image	# Test Images	Accuracy (in %)
Yaman *et al.* (2019) [23]	multimodal deep neural network	FERET	1397	1117 (80%)	140 (10%)	99.11
Meng *et al.* (2019) [17]	KNN, Eucledian distance	Own	134	N/A	N/A	67.2
Singh *et al.* (2023) [19]	Convolutional Neural Network(CNN)	EarVN1.0	28412	19888	8524	93
Srinivasan (2024) [20]	Mask RCNN, Grabcut Segmentation	EarVN1.0	28,412			87.52
Kılıç (2024) [13]	CNN + MLP Mixer + ViT	EarVN1.0	28,412	N/A	N/A	96.66
Kaur *et al.* (2025) [11]	PyCaret	Own	508	368	153	86.75
Arun *et al.* (2025) [1]	Vision Transformer	UERC 2023, VGGFace-Ear	2,78,181	2,47,655		For Training
		OPIB			907	91.96
		AWE			1000	94.89
		WPUT			2071	92.78
		EarVN1.0			28,412	78.38

TABLE I: Summary of Recent Works For Gender Classification Using Ear Biometrics

III. Methodology

In this section, we detail the proposed model architecture and the training pipeline used for gender classification from ear images. The methodology includes dataset preprocessing, the design of a dual-branch architecture combining spatial and frequency-domain features, and the training procedures employed to optimize model performance.

A. Dataset Preprocessing

We preprocessed each ear image into two parallel inputs, one for the Vision Transformer branch and one for the frequency-domain CNN branch, as follows:

1) Spatial preprocessing (ViT branch):

- **Resize & Color:** Images are resized to 224×224 pixels and converted to RGB (three channels).
- **Normalization:** We normalize each channel to zero mean and unit variance using ImageNet statistics (mean = [0.485, 0.456, 0.406], std = [0.229, 0.224, 0.225]), matching the pretrained ViT-Tiny expectations.

2) Frequency-domain preprocessing (CNN branch):

- **Grayscale conversion:** Input images are converted to single-channel grayscale.
- **Resize:** Grayscale images are resized to 224×224 pixels.
- **Discrete Wavelet Transform:** We apply a single-level 2D Haar DWT to each resized image, yielding four sub-bands: LL, LH, HL, and HH. We discard the high-frequency HH band (which primarily captures noise) and stack LL, LH, and HL into a three-channel tensor.
- **Scaling:** Each wavelet sub-band is standardized (zero mean, unit variance) across the training set to aid convergence.

These two preprocessing pipelines produce aligned 224×224 tensors, one RGB for spatial feature extraction via ViT, and one pseudo-RGB (LL/LH/HL) for frequency analysis via the CNN branch.

B. Model Architecture

Our dual-domain network consists of two parallel branches, one using a Vision Transformer for spatial feature extraction and the other a lightweight CNN operating on wavelet sub-bands, where the outputs are fused for final gender classification.

1) Vision Transformer branch: We adopt the `vit_tiny_patch16_224` architecture from the `timm` library, pretrained on ImageNet. We remove its original classification head and take the 192-dimensional embedding produced by the `[CLS]` token after the final transformer block. Formally, given an RGB input $X_{\text{RGB}} \in \mathbb{R}^{3\times224\times224}$, we patchify and project it into token embeddings, then process through 12 transformer layers with multi-head self-attention and MLP blocks. The resulting vector $\mathbf{f}_{\text{ViT}} \in \mathbb{R}^{192}$ captures global spatial patterns critical for gender cues.

2) Frequency-domain CNN branch: To include fine-grained texture and edge information, we designed a three-layer CNN that operates on the LL, LH, and HL coefficients from a single-level 2D Haar DWT of the grayscale image. Each convolutional block consists of:

$$\text{Conv2D}(C_{\text{in}} \to C_{\text{out}},\, 3\times3) \;\to\; \text{ReLU} \;\to\; \text{MaxPool}(2\times2),$$

with channel progression $3 \to 32 \to 64 \to 128$. After the third block, we applied global average pooling to obtain a 128-dimensional vector $\mathbf{f}_{\text{CNN}} \in \mathbb{R}^{128}$. A final fully-connected layer projects this into the same 128-d space to stabilize fusion.

3) Feature Fusion and Classifier: We concatenate the two embeddings into a joint feature $\mathbf{f} = [\mathbf{f}_{\text{ViT}}; \mathbf{f}_{\text{CNN}}] \in \mathbb{R}^{320}$. This vector feeds into a fusion head comprising:

$$\begin{aligned}\text{Dense}(320\to256) \;\to\; \text{ReLU} \;\to\; \text{Dropout}(p=0.5)\\ \to\; \text{Dense}(256\to1) \;\to\; \sigma(\cdot)\end{aligned}$$

where σ is the sigmoid activation. The output $\hat{y} = \sigma(\mathbf{W}_2(\text{ReLU}(\mathbf{W}_1\,\mathbf{f}+b_1))+b_2)$ represents the predicted prob-

ability of the "female" class. We train end-to-end, allowing both branches to fine-tune their weights toward the gender-classification objective.

C. Training Procedure

We trained the dual-domain network end-to-end on the dataset training split using the following settings:

- **Loss Function:** Binary Cross-Entropy Loss (BCELoss) is applied to the sigmoid output $\hat{y}$ against the ground-truth gender label $y \in \{0, 1\}$:

$$L = -\left[y \log(\hat{y}) + (1 - y) \log(1 - \hat{y})\right]$$

- **Optimizer:** We use the Adam optimizer with an initial learning rate of 1×10^{-4}, $\beta_1 = 0.9$, $\beta_2 = 0.999$, and $\epsilon = 1 \times 10^{-8}$.
- **Learning Rate Schedule:** A `ReduceLROnPlateau` scheduler monitors the validation loss and reduces the learning rate by a factor of 0.5 if no improvement is observed for 3 consecutive epochs, with a minimum LR floor of 1×10^{-6}.
- **Batch Size & Epochs:** Training is performed for up to 30 epochs with a batch size of 32.
- **Early Stopping & Checkpointing:** We apply early stopping with a patience of 5 epochs on the validation loss; the model achieving the lowest validation loss is saved to `best_model.pt`
- **Data Augmentation:** To improve generalization, during training we apply random horizontal flipping, random brightness/contrast adjustment ($\pm 10\%$), and random rotation ($\pm 15°$) to the RGB inputs; no augmentation is applied to the wavelet-domain inputs.
- **Training Loop:**
  1) Forward propagate a batch through both branches to obtain $\hat{y}$.
  2) Compute BCELoss and backpropagate gradients to update all trainable parameters.
  3) Every epoch, evaluate on the validation split to compute loss and metrics.
  4) Adjust learning rate and apply early stopping criteria.

Fig 1 illustrates the overall flow of our dual-domain architecture, highlighting the parallel ViT and DWT-CNN branches and their late fusion for final gender prediction.

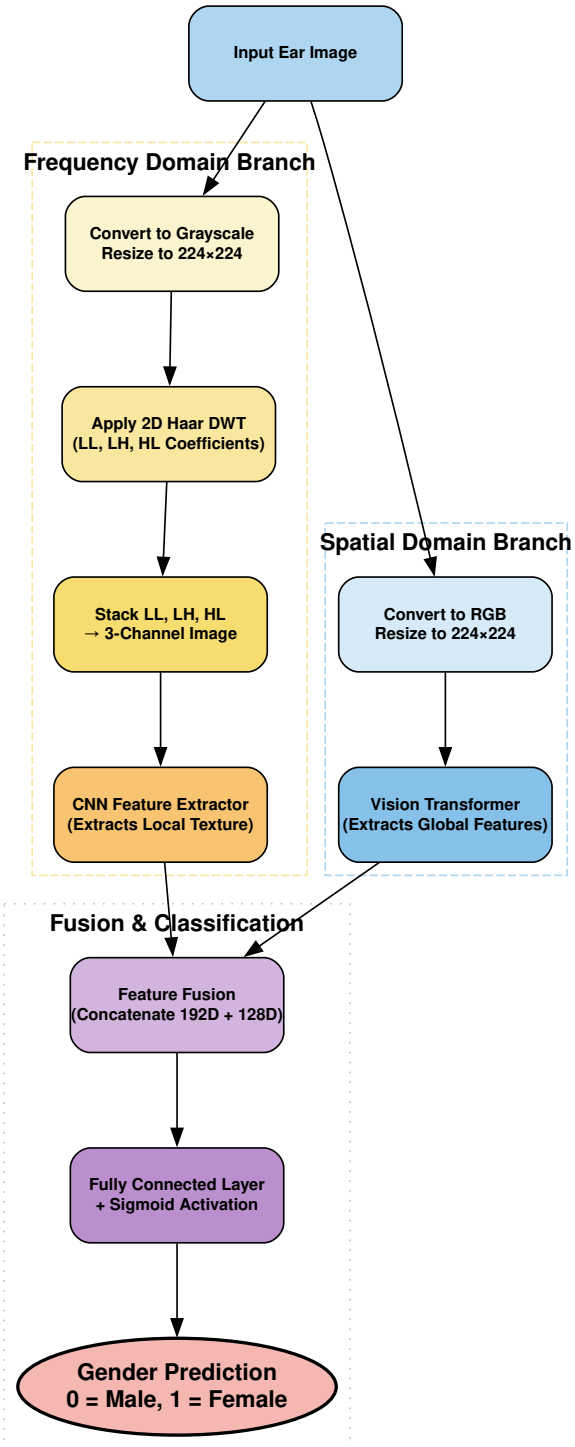


Fig. 1: Proposed dual-domain gender classification architecture using ViT and DWT-CNN fusion.

IV. Experimental Setup and Results

A. Dataset Description

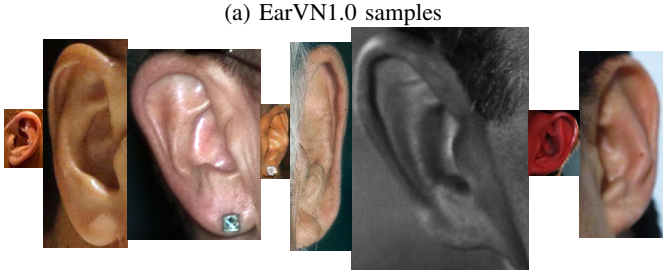
All experiments are conducted using two publicly available ear-image datasets: EarVN1.0 [9] and AWE [2].

1) EarVN1.0: EarVN1.0 is one of the largest ear-image collections to date, comprising 28,412 images from 164 Asian subjects (98 males, 66 females). Images exhibit wide variability in pose, illumination, occlusion, and image quality, and there are no duplicates or near-duplicates. Samples are shown in Fig. 2a. Data are organized under `IMAGES/001.<ID>` `...164.<ID>`, with subjects 001–098 labeled male (0) and

099–164 female (1). We perform a 80/20 split at the subject level (131 subjects, ≈22,729 images for training; 33 subjects, ≈5,683 images for validation).

2) AWE: The Annotated Web Ear (AWE) dataset contains 1,000 ear images collected from the web across 100 well-known individuals of diverse ethnicity. Of these, approximately 91% are male and 9% female. Images vary in background, pose, and lighting.

(a) EarVN1.0 samples



(b) AWE samples

Fig. 2: Example ear images from the EarVN1.0 and AWE datasets.

Fig 2 presents example ear images from the EarVN1.0 and AWE datasets, showcasing the visual diversity and varying conditions captured in each dataset.

B. Evaluation Metrics

We assessed model performance on the EarVN1.0 validation set using the following metrics:

- **Accuracy:** Overall fraction of correctly predicted labels.
- **Precision & Recall:**

$$\text{Precision} = \frac{\text{TP}}{\text{TP} + \text{FP}}, \quad \text{Recall} = \frac{\text{TP}}{\text{TP} + \text{FN}}$$

- **F1-Score:** Harmonic mean of precision and recall:

$$\text{F1} = 2 \cdot \frac{\text{Precision} \times \text{Recall}}{\text{Precision} + \text{Recall}}$$

- **ROC–AUC:** Area under the Receiver Operating Characteristic curve.
- **Confusion Matrix:** Counts of true positives (TP), true negatives (TN), false positives (FP), and false negatives (FN).

C. Results and Discussion

We evaluate the full Dual-domain model and conduct an ablation study against ViT-only and CNN-only variants on both EarVN1.0 and AWE.

1) Ablation Study: To quantify each branch's contribution, we compare:

- **ViT-only:** Vision Transformer branch alone.
- **CNN-only:** Wavelet-based CNN branch alone.
- **Dual-domain (ours):** Full fused model.

a) EarVN1.0 Ablation: Table II summarizes the three variants. Fig 3 shows a grouped bar chart of their validation accuracies.

Model	Acc.	Prec.	Rec.	F1	ROC–AUC
ViT-only	0.9342	0.9400	0.9330	0.9365	0.9539
CNN-only	0.6927	0.6924	0.6952	0.6915	0.7535
Dual-domain (ours)	0.9730	0.9725	0.9732	0.9729	0.9972

TABLE II: Ablation results on EarVN1.0.

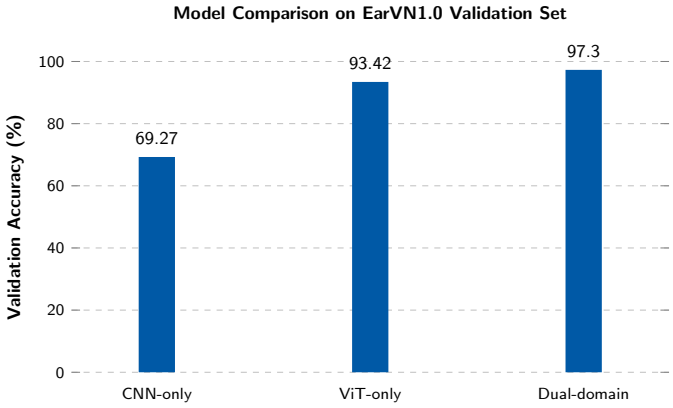


Fig. 3: Validation accuracy comparison (ViT-only, CNN-only, Dual-domain) on EarVN1.0.

b) AWE Ablation: Table III summarizes the three variants. Fig 4 shows a grouped bar chart of their validation accuracies.

Model	Acc.	Prec.	Rec.	F1	ROC–AUC
ViT-only	0.8615	0.4313	0.5010	0.4631	0.6950
CNN-only	0.8611	0.4306	0.5000	0.4627	0.4490
Dual-domain (ours)	0.8839	0.7661	0.5184	0.6182	0.7235

TABLE III: Ablation results on AWE.

Author	Detection Method	# Train Image	# Test Images	Accuracy (in %)
Singh *et al.* (2023) [19]	Convolutional Neural Network(CNN)	19888	8524	93
Srinivasan (2024) [20]	Mask RCNN, Grabcut Segmentation	–	–	87.52
Kılıç (2024) [13]	CNN + MLP Mixer + ViT	–	–	96.66
Arun *et al.* (2025) [1]	Vision Transformer	247655	28,412	78.38
The proposed	Dual Domain (CNN and ViT)	22729	5683	97.30

TABLE IV: Comparison of the Proposed Method with Others on EarVN1.0 Dataset

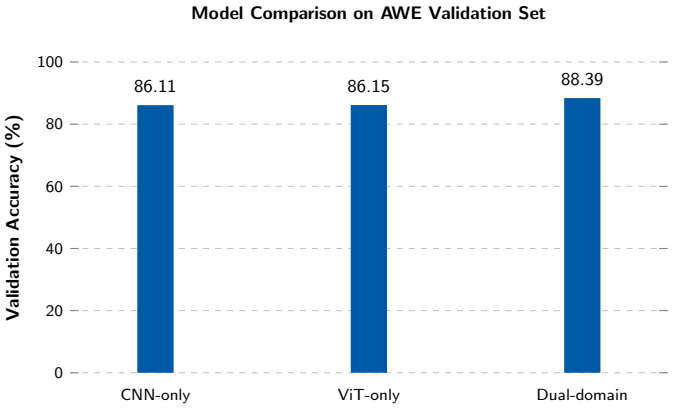


Fig. 4: Validation accuracy comparison (ViT-only, CNN-only, Dual-domain) on AWE.

2) Discussion: Across both datasets, the grouped bar charts clearly show that the Dual-domain model outperforms each single-branch variant, confirming that fusing spatial and frequency-domain features yields superior gender classification. Table IV provides a comparative analysis of the proposed method against recent approaches, highlighting its competitive accuracy.

V. Conclusion

In this work, we introduced a dual-domain architecture that brings together a Vision Transformer branch for capturing global spatial patterns and a lightweight, frequency-aware CNN branch for extracting fine-grained texture cues from ear images. By converting each ear into both an RGB patch sequence and a three-channel wavelet representation, our model learns complementary information that neither branch could fully exploit on its own.

Through extensive experiments on the large-scale EarVN1.0 and the more challenging AWE datasets, we showed that fusing spatial and frequency domains consistently boosts gender-classification accuracy, achieving over 97% on EarVN1.0 and nearly 89% on AWE datasets. Ablation studies further confirmed that each branch contributes uniquely: the ViT captures global shape and context, while the wavelet CNN highlights subtle edge and texture differences important for distinguishing male and female ear morphology. Looking forward, we plan to explore self-supervised pretraining for even richer feature representations, as well as multimodal extensions that jointly leverage ear and profile face. We believe this dual-domain paradigm can serve as a flexible foundation for a wide range of soft-biometric tasks beyond gender classification.

References

[1] D. Arun, K. Ozturk, K. Bowyer, and P. Flynn. Improved ear verification with vision transformers and overlapping patches. *arXiv preprint arXiv:2503.23275*, 2025. DOI: 10.48550/arXiv.2503.23275.

[2] Z. Emersic, V. Struc, and P. Peer. Ear recognition: More than a survey. *Neurocomputing*, 255:26–39, 2017. DOI:10.1016/j.neucom.2016.08.139.

[3] Y. Fu, Y. Xu, and T. S. Huang. Estimating human age by manifold analysis of face pictures and regression on aging features. In *International Conference on Multimedia and Expo*, pages 1383–1386, 2007. DOI:10.1109/ICME.2007.4284917.

[4] X. Geng, Z-H. Zhou, and K. Smith-Miles. Automatic age estimation based on facial aging patterns. *Transactions on pattern analysis and machine intelligence*, 29(12):2234–2240, 2007. DOI:10.1109/TPAMI.2007.70733.

[5] P. Gnanasivam and S. Muttan. Gender classification using ear biometrics. In *Proceedings of the Fourth International Conference on Signal and Image Processing*, pages 137–148, 2013. DOI: 10.1007/978-81-322-1000-9_13.

[6] G. Guo, Y. Fu, C. R. Dyer, and T. S. Huang. Image-based human age estimation by manifold learning and locally adjusted robust regression. *IEEE Transactions on Image Processing*, 17(7):1178–1188, 2008. DOI:10.1109/TIP.2008.924280.

[7] G. Guo and G. Mu. Human age estimation: What is the influence across race and gender? In *Computer Society Conference on Computer Vision and Pattern Recognition-Workshops*, pages 71–78, 2010. DOI:10.1109/CVPRW.2010.5543609.

[8] J. Hayashi, M. Yasumoto, H. Ito, and H. Koshimizu. Method for estimating and modeling age and gender using facial image processing. In *Seventh International Conference on Virtual Systems and Multimedia*, pages 439–448, 2001. DOI: 10.1109/VSMM.2001.969698.

[9] V. T. Hoang. Earvn1. 0: A new large-scale ear images dataset in the wild. *Data in brief*, 27:104630, 2019.

[10] A. Kamboj, R. Rani, and A. Nigam. A comprehensive survey and deep learning-based approach for human recognition using ear biometric. *The Visual Computer*, 38(7):2383–2416, 2022. DOI: 10.1007/s00371-021-02119-0.

[11] T. Kaur, K. Krishan, A. Sharma, A. Guleria, and V. Sharma. Sex classification accuracy through machine learning algorithms-morphometric variables of human ear and nose. *BMC Research Notes*, 18(1):169, 2025. DOI: 10.1186/s13104-025-07185-4.

[12] R. Khorsandi and M. Abdel-Mottaleb. Gender classification using 2-d ear images and sparse representation. In *Workshop on applications of computer vision*, pages 461–466, 2013. DOI: 10.1109/WACV.2013.6475055.

[13] Şafak Kılıç and Yahya Doğan. Deep learning based gender identification using ear images. *Traitement du Signal*, 40(4), 2023.

[14] Y. H. Kwon and N. da Vitoria Lobo. Age classification from facial images. *Computer vision and image understanding*, 74(1):1–21, 1999. DOI: 10.1006/cviu.1997.0549.

[15] A. Lanitis, C. Draganova, and C. Christodoulou. Comparing different classifiers for automatic age estimation. *IEEE Transactions on Systems, Man, and Cybernetics, Part B*, 34(1):621–628, 2004. DOI:10.1109/TSMCB.2003.817091.

[16] J. Lei, J. Zhou, and M. Abdel-Mottaleb. Gender classification using automatically detected and aligned 3d ear range data. In *International Conference on Biometrics*, pages 1–7, 2013. DOI: 10.1109/ICB.2013.6612995.

[17] D. Meng, M. Nixon, and S. Mahmoodi. Gender and kinship by model-based ear biometrics. In *2019 International Conference of the Biometrics Special Interest Group (BIOSIG)*, pages 1–5, 2019.

[18] N. Ramanathan and R. Chellappa. Face verification across age progression. *IEEE Transactions on Image Processing*, 15(11):3349–3361, 2006. DOI:10.1109/TIP.2006.881993.

[19] R. Singh, K. Kashyap, R. Mukherjee, A. Bera, and M. Chakraborty. Deep ear biometrics for gender classification. In *International Conference on Communication, Devices and Computing*, pages 521–530, 2023. DOI: 10.1007/978-981-99-2710-4_42.

[20] L. Srinivasan. Gender identification using 2D ear biometric. In *Science and Information Conference*, pages 439–449, 2024. DOI:.

[21] K. Ueki, T. Hayashida, and T. Kobayashi. Subspace-based age-group classification using facial images under various lighting conditions. In *7th International Conference on Automatic Face and Gesture Recognition*, pages 6–pp, 2006. DOI: 10.1109/FGR.2006.102.

[22] D. Yaman, F. I. Eyiokur, N. Sezgin, and H. K. Ekenel. Age and gender classification from ear images. In *International Workshop on Biometrics and Forensics*, pages 1–7, 2018. DOI:10.1109/IWBF.2018.8401568.

[23] D. Yaman, F. Irem Eyiokur, and H. Kemal Ekenel. Multimodal age and gender classification using ear and profile face images. In *Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition Workshops*, pages 2414–2421.

[24] S. Yan, M. Liu, and T. S. Huang. Extracting age information from local spatially flexible patches. In *International Conference on Acoustics, Speech and Signal Processing*, pages 737–740, 2008. DOI:10.1109/ICASSP.2008.4517715.

[25] G. Zhang and Y. Wang. Hierarchical and discriminative bag of features for face profile and ear based gender classification. In *International joint conference on biometrics*, pages 1–8, 2011. DOI:10.1109/IJCB.2011.6117590.