International Journal of Advanced Trends in Engineering and Management
(ISSN 2583-7052)

Vol. V, No. 09, pp. 1-7, September

Hybrid Attention-Based CNN Framework for Interpretable Age and Gender Classification from Facial Images

Alpaben Remeshbhai Patel [1*], Amisha Shingala[2]

[1]Research Scholar, Department of Computer Science, Gujarat Technological University, Ahmedabad, India.

[2]Department of Computer Applications, Charusat University, Changa, India.

Corresponding Author E-mail: alpa8187@gmail.com

ABSTRACT: Facial age and gender classification from unconstrained facial images is still very difficult since there are many variations in pose, illumination, occlusion, and other real-world image capturing conditions. There are current state-of-the-art solutions based on deep learning that aim to improve classification accuracy but offer limited interpretability and deployability. In this paper, a novel explainable hybrid attention-based CNN with SHAP for multi-task classification of facial age and gender attributes is proposed. The proposed method consists of a channel and spatial attention module which helps to improve facial feature extraction along with using SHAP for making the prediction more transparent. At last, a Flask-based web application is developed for implementing the proposed method for inference and explanation purposes. In experiments on the Adience dataset, the proposed system achieved 72.80% accuracy in age classification and 91.60% accuracy in gender classification.

Keywords: Age estimation, Gender classification, Hybrid CNN, Attention mechanism, SHAP interpretability

1. Introduction

Facial image classification for age and gender is vital for applications like surveillance and human-computer interfaces, yet challenging due to pose, lighting, expressions, and occlusions. While traditional models fall short, recent deep learning approaches utilizing CNN architectures (such as VGGNet, ResNet, MobileNet, and EfficientNet), attention mechanisms, and explainable AI (like SHAP) have significantly enhanced recognition performance and model transparency.

This research introduces a novel framework combining hybrid channel-spatial attention mechanisms, multi-task learning, SHAP-based explainability, and a Flask deployment.

The following sections are ordered as follows: Section 2 provides an extensive literature survey of deep learning approaches used for age and gender classification, together with current developments in explainable artificial intelligence, focusing on the SHAP approach. Section 3 outlines the methodology used as a foundation for our proposed model design, while Section 4 shows experimental results measured in accuracy, precision, recall, and F1-scores with confusion matrices alongside SHAP-based model interpretation visualizations. The study concluded in Section 5, which summarizes the conclusion, limitations, and suggests further research investigation.

2. Recent Works

While traditional manual methods for age and gender classification from facial images struggled with real-world variables like lighting and pose, existing literature often prioritizes only accuracy, deployment, or explainability as discussed in

Table 1. This study addresses that gap by integrating all three aspects.

Table 1: *Comparative analysis of existing methods and Research Gaps*

Research Paper	Model Used	Key Findings	Research Gap
Aruleba&Viriri [1]	EfficientNet CNN	High accuracy in age estimation using deep CNN architecture	No explainability (XAI) and no multi-task learning
Benkaddour [2]	CNN feature extractor	CNN improves age estimation and gender classification accuracy	No attention mechanism and no explainable AI
Duan et al. [3]	CNN–ELM hybrid model	Hybrid CNN improves classification accuracy for age & gender	No explainability and no real-time deployment
Hamza [4]	MobileNetV2 CNN	Lightweight model suitable for gender classification on face images	No attention module and no XAI integration
Kale & Altun [5]	CNN-based features	CNN effective for age, gender, ethnicity classification	No deep attention mechanism or interpretability module
Kocoń&Pawlukiewicz [6]	Deep CNN models	High performance in age/gender prediction using deep learning	No explainability and no deployment system
Kumar et al. [7]	Deep CNN model	Strong performance in face-based age and gender classification	No attention mechanism or SHAP-based interpretability
Levi & Hassner [8]	Basic CNN	First CNN-based approach for age and gender classification	No attention, no XAI, no modern architectures
Mansour & Abu-Naser [9]	ResNet CNN	ResNet improves classification accuracy for age/gender tasks	No explainability or deployment framework
Mosayyebi et al. [10]	EfficientNet + Transfer Learning	Improved gender classification in masked faces	No SHAP/XAI and no real-time deployment
P. Dey et al. [11]	CNN-based DL Architecture	ResNet-50 outperformedVGG16 for age/gender detection in facial images	No explainability and no real-time deployment
Patel &Shingala [12]	Comparative CNN models	Evaluates multiple CNN architectures for prediction accuracy	No unified model or explainability integration
Patel M.A.M. &Shingala [13]	CNN-based survey	Summarizes CNN techniques for age/gender classification	No experimental contribution or model improvement
Lundberg& Lee [14]	SHAPframework	Introduced SHAP to interpret model predictions which is mathematically proved	Assumption of feature independence and high computational complexity
Rwigema et al. [15]	Hybrid NN + Decision Fusion	Improved accuracy using ensemble neural networks	No SHAP explainability or deep attention models
Patel, A. R., & Shingala, A.[16]	Hybrid CNN	Improved accuracy using attention neural networks	Assumption of feature independence and high computational complexity
Tariq et al. [17]	VGG19 CNN	Real-time classification using VGG19 architecture	No attention mechanism, no SHAP explainability, no Flask deployment

From the literature, it is clear that previous studies have faced the following limitations: Lack of integration between accuracy and interpretability, Limited use of SHAP within deep CNN pipelines, Absence of real-time explainable deployment systems, Insufficient focus on multi-task learning

for joint prediction. To overcome the above research gap, the proposed study introduces a unified framework that combines an attention-based CNN, SHAP explainability and real-time Flask deployment to improve both performance and transparency.

3. Proposed Work Explanation

This study proposes an Explainable Age and Gender Classification Framework that integrates a hybrid CNN (attention-based) with Shapley Additive explanations (SHAP) and a Flask-based real-time deployment. The proposed architecture was designed to mitigate the limitations of classical CNN models through multi-task learning, attention-driven feature refinement, and explainability-aware validation.

The proposed architecture has six stages: dataset acquisition, preprocessing, feature extraction using a hybrid attention CNN, multi-task classification output, XAI SHAP explainability generation, and deployment using a Flask Web Application. This is a combined approach for age and gender classification with explainability. It starts with image acquisition and then goes through various image preprocessing steps like face detection, resizing, normalization, and augmentation. After image preprocessing, a hybrid attention-based CNN is used for image classification, where convolutional layers are used for image features and channel-spatial attention to highlight important features in the image. This architecture uses multitask learning to classify age and gender. To make the model more interpretable, SHAP has been used for interpretability purposes. Finally, a Flask-based web application is used for model deployment. Four Convolutional Layers have been used in the proposed Hybrid Attention-Based CNN model. The proposed architecture of CNN model is depicted in Figure 1.

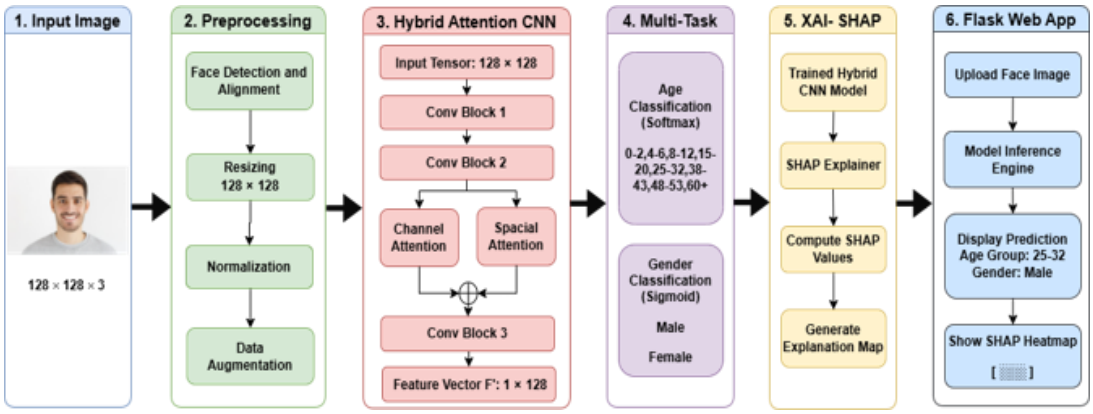


Figure 1: *Proposed Architecture of Hybrid Attention CNN*

3.1 Description of the Dataset

The Hybrid CNN Model uses the Adience database that consists of 16,574 face images. These images were labeled for both age and gender. Age is divided into eight categories, namely [0-2], [4-6], [8-12], [15-20], [25-32], [38-43], [48-53], and [60+]. The label for gender can be either "male" (8,341) or "female" (8,233). Some of the problems faced with the dataset include: imbalance between age groups; pose variation; expression variation; occlusion; lighting and noisy labels.

3.2 Data Preprocessing

In this work, 16,574 Adience dataset images preprocessed by detecting, cropping, and resizing facial regions to 128×128pixels, then normalizing pixel values to [0,1]. Eight age ranges and binary gender labels were encoded. To fix class imbalances and boost real-world generalization, we applied data augmentation (rotation, shifts, flipping) using ImageDataGenerator, expanding the dataset to 40,224 images—aligning with SDG 10 (Reduced Inequalities) by ensuring fair demographic

performance. In addition, data was partitioned into training (70%), validation (15%), and test (15%) datasets.

3.3 Hybrid CNN Architecture

The proposed hybrid attention-based CNN model architecture for age and gender recognition is demonstrated in Table 2. The framework includes four convolutional layers integrated with batch normalization and max-pooling operations to enable effective hierarchical feature extraction. To advance discriminative facial representation learning, both channel and spatial attention mechanisms are incorporated within the network. Furthermore, global average pooling and fully connected layers with dropout have been used before final classification layers. Overall, the architecture balances classification accuracy and interpretation efficiently with around 559,532 learnable parameters.

Table 2: *Detailed Architecture of the Proposed Hybrid Attention-Based CNN*

Layer	Layer Type	Output Size	Kernel Size / Operation	No. of Filters
1	Input Layer	128 × 128 × 3	RGB Image Input	–
2	Conv2D + ReLU	128 × 128 × 32	3 × 3	32
3	Batch Normalization	128 × 128 × 32	–	–
4	Max Pooling	64 × 64 × 32	2 × 2	–
5	Conv2D + ReLU	64 × 64 × 64	3 × 3	64
6	Batch Normalization	64 × 64 × 64	–	–
7	Max Pooling	32 × 32 × 64	2 × 2	–
8	Conv2D + ReLU	32 × 32 × 128	3 × 3	128
9	Batch Normalization	32 × 32 × 128	–	–
10	Max Pooling	16 × 16 × 128	2 × 2	–
11	Conv2D + ReLU	16 × 16 × 256	3 × 3	256
12	Batch Normalization	16 × 16 × 256	–	–
13	Channel Attention Module	16 × 16 × 256	Global Avg/Max Pool + MLP	–
14	Spatial Attention Module	16 × 16 × 256	7 × 7 Conv	–
15	Global Average Pooling	1 × 1 × 256	–	–
16	Dense + ReLU	512	Fully Connected	–
17	Dropout	512	Regularization	–
18	Age Output Layer	8	Softmax	–
19	Gender Output Layer	1	Sigmoid	–

3.4 Training Strategy

The model was optimized over a maximum of 65 epochs using the Adam optimizer, where the learning rate is set at 10 ^ (-4), while the batch size is between 32 and 64. To ensure generalization and prevent overfitting, certain regularizations were enforced in the model training process, such as the use of batch normalization, a dropout rate of 0.5, and early stopping according to validation loss. Additionally, the training process incorporates an adaptive learning rate scheduler and gradient clipping to enhance stability and convergence.

3.5 SHAP-Based Explainability

To render predictions transparent, Shapley Additive exPlanations (SHAP) were calculated using the DeepExplainer method. SHAP quantifies pixel- and feature-map contributions to yield visual explanation heatmaps for age and gender predictions.

3.6 Flask-Based Deployment Framework

The framework was deployed via Flask for real-time web inference, allowing users to upload images and view age/gender predictions alongside overlaid SHAP heatmaps. Supporting

multithreaded requests via an API, it ensures high-availability integration in line with UN SDG 9.

4. Results and Discussion

Results from experiments performed on the Hybrid Attention-Based CNN model are presented in this section. Table 3 offers an evaluation of the performance of the model using accuracy, precision, recall, F1-score, and confusion matrix metrics for age and gender classification tasks.

Table 3: *Performance Evaluation of the Proposed Model*

Task	Accuracy (%)	Precision (%)	Recall (%)	F1 Score (%)
Age Classification	72.80	71.50	70.90	71.20
Gender Classification	91.60	91.20	90.80	91.00

Training and validation graphs for age and gender classification using proposed model for 65 epochs are shown in Figure 2 below.

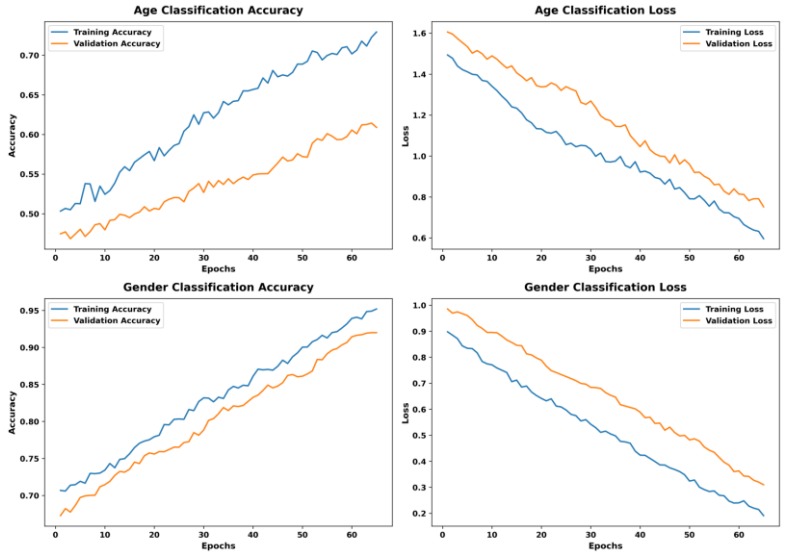


Figure 2: *Training and validation graphs for age and gender classification: (a) Accuracy of age classification task, (b) Loss of age classification task, (c) Accuracy of gender classification task, and (d) Loss of gender classification task*

The age classification confusion matrix in Figure 3(a) shows that the predictions were mostly accurate in the middle age range, but the classification error was slightly higher in the extreme age ranges. Figure 3(b) displays the confusion matrix for gender classification, where there is hardly any misclassification, with both male and female classes having relatively low error rates.

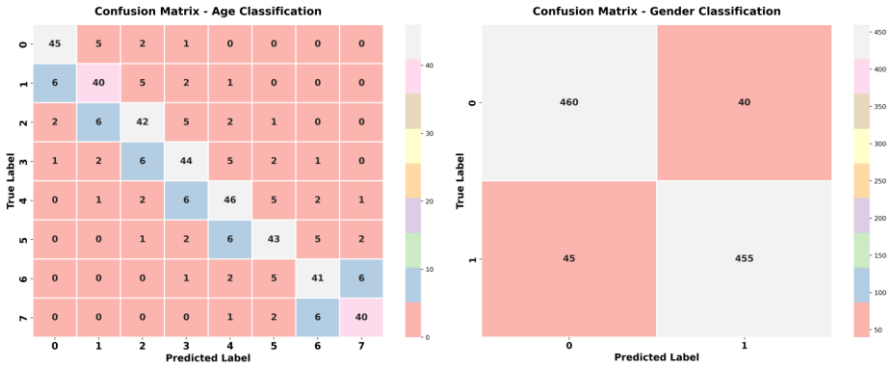


Figure 3: *Confusion Matrices for (a) age classification and (b) gender classification*

To add interpretability to the results, SHAP-based explanations were provided. The SHAP heatmaps presented in Figure 4 show the most impactful facial features for classification. The red color represents high positive impact, whereas blue represents low impact.

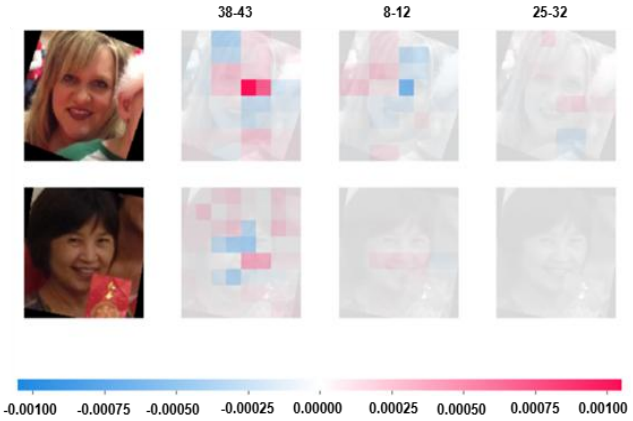


Figure 4: *SHAP-based explanation maps highlighting important facial regions contributing to age and gender predictions*

Real-time Deployment of Proposed System with Unknown Image is shown in Figure 5. In this figure, users can upload images and also can view the classification result in the same window of Flask web application.



Figure 5: *Output of the Flask-based web application showing real-time prediction and SHAP-based explanation for unseen input images*

5. Conclusion

This study presented an explainable attention-based CNN architecture was introduced for real-time age and gender classification using Flask, and tested using the Adience dataset. With accuracies of 72.80% and 91.60% for age and gender respectively, the developed architecture employs SHAP heat maps to identify important features in the face images and improve explainability, thus meeting UN SDG-9 and 10. Even though the system proves to be quite effective, it suffers from computational inefficiencies in generating SHAP values as well as poor performances on extreme ages; efforts will be made in future studies to optimize attention mechanisms for edge devices.

References

1. Idowu Aruleba; Serestina Viriri, Year: 2021, “Deep Learning for Age Estimation Using EfficientNet”, Proceedings of International Work-Conference on Artificial Neural Networks (IWANN 2021), Lecture Notes in Computer Science, Vol: 12861, Springer, pp. 392-403.
2. Mohammed Kamel Benkaddour, Year: 2021, “CNN Based Features Extraction for Age Estimation and Gender Classification”, Informatica, Vol: 45, No: 5, pp. 697-703.
3. Mingxing Duan; Kenli Li; Canqun Yang; Keqin Li, Year: 2018, “A Hybrid Deep Learning CNN-ELM for Age and Gender Classification”, Neurocomputing, Vol: 275, pp. 448-461.
4. Nisreen Ryadh Hamza, Year: 2025, “Gender Classification from Human Face Images Using Deep Learning Based on MobileNetV2 Architecture”, Journal of Al-Qadisiyah for Computer Science and Mathematics, Vol: 17, No. 1, pp. 145-156.
5. Ayşe Kale; Oğuz Altun, Year: 2021, “Age, Gender and Ethnicity Classification from Face Images with CNN-Based Features,” Proceedings of 2021 Innovations in Intelligent Systems and Applications Conference (ASYU), pp. 1-6.
6. M. Kocoń; S. Pawlukiewicz, Year: 2025, “Age Estimation and Gender Classification from Facial Images”, Applied Sciences, Vol: 15, No: 18, pp. 10212.

7. Rajiv Kumar; Kuldeep Singh; Dharmendra Prasad Mahato; Umesh Gupta, Year: 2024, “Face-based age and gender classification using deep learning model”, Procedia Computer Science, Vol: 235, pp. 2985-2995.
8. Gil Levi; Tal Hassner, Year: 2015, “Age and gender classification using convolutional neural networks”, In Proceedings of the IEEE conference on computer vision and pattern recognition workshops, pp. 34-42.
9. Aysha I. Mansour; S. S. Abu-Naser, Year: 2022, “Classification of age and gender using ResNet-deep learning”, International Journal of Academic Engineering Research (IJAER), Vol: 6, No: 8, pp. 20-29.
10. Faezeh Mosayyebi; Hadi Seyedarabi; Reza Afrouzian, Year: 2024, “Gender Recognition in Masked Facial Images Using EfficientNet and Transfer Learning Approach”, International Journal of Information Technology, Vol: 16, No: 4, pp. 2693-2703.
11. Puja Dey; Tanjim Mahmud; Mohammad Sanaullah Chowdhury; Mohammad Shahadat Hossain; Karl Andersson, Year: 2024, “Human age and gender prediction from facial images using deep learning methods”, Procedia Computer Science, Vol: 238, pp. 314-321.
12. Alpa Patel; Amisha Shingala, Year: 2024, “Comparative Study of Deep Learning Models for Age and Gender Prediction”, In International Conference on Data Analytics and Insights, pp. 417-425. Singapore: Springer Nature Singapore.
13. Ms Alpa Maulik Patel; Amisha Shingala, Year: 2023, “Age estimation and gender classification techniques using CNN: a survey”, EPRA International Journal of Multidisciplinary Research (IJMR), Vol: 9, No: 11, pp. 143-149.
14. Scott M. Lundberg; Su-In Lee, Year: 2017, “A Unified Approach to Interpreting Model Predictions”, Advances in Neural Information Processing Systems (NeurIPS), Vol: 30, pp. 4765-4774.
15. James Rwigema; Joseph Mfitumukiza; Kim Tae-Yong, Year: 2021, “A Hybrid Approach of Neural Networks for Age and Gender Classification Through Decision Fusion,” Biomedical Signal Processing and Control, Vol: 66, pp. 102459.
16. Alpaben Rameshbhai Patel; Amisha Shingala, Year: 2026, “Explainable Age and Gender Classification Using CNN with SHAP Interpretability on Human Facial Images,” Journal of Data Science and Intelligent Systems.
17. Muhammad Usman Tariq, A. R. Akram, S. O. Yaqoob, M. E. Rasheed, and M. S. Ali, Year: 2023, “Real-time age and gender classification using VGG19”, Adv. Mach. Learn. Artif. Intell, Vol: 4, pp. 56-65, 2023.