Journal of Electrical and Computer Engineering Innovations
(JECEI)
Journal homepage: http://www.jecei.sru.ac.ir

Research paper

Advanced Race Classification Using Transfer Learning and Attention: Real-Time Metrics, Error Analysis, and Visualization in a Lightweight Deep Learning Model

M. Rohani, H. Farsi, S. Mohamadzadeh *

Department of Electrical Engineering, Faculty of Electrical and Computer Engineering, University of Birjand, Birjand, Iran.

Article Info

Article History:
Received 29 September 2024
Reviewed 07 December 2024
Revised 06 January 2025
Accepted 10 January 2025

Keywords:
Race classification
Attention module
Efficient-Net network
Transfer learning
Real-time performance

*Corresponding Author's Email Address:
s.mohamadzadeh@birjand.ac.ir

Abstract

Background and Objectives: Recent advancements in race classification from facial images have been significantly propelled by deep learning techniques. Despite these advancements, many existing methodologies rely on intricate models that entail substantial computational costs and exhibit slow processing speeds. This study aims to introduce an efficient and robust approach for race classification by utilizing transfer learning alongside a modified Efficient-Net model that incorporates attention-based learning.

Methods: In this research, Efficient-Net is employed as the base model, applying transfer learning and attention mechanisms to enhance its efficacy in race classification tasks. The classifier component of Efficient-Net was strategically modified to minimize the parameter count, thereby enhancing processing speed without compromising classification accuracy. To address dataset imbalance, we implemented extensive data augmentation and random oversampling techniques. The modified model was rigorously trained and evaluated on a comprehensive dataset, with performance assessed through accuracy, precision, recall, and F1 score metrics.

Results: The modified Efficient-Net model exhibited remarkable classification accuracy while significantly reducing computational demands on the UTK-Face dataset. Specifically, the model achieved an accuracy of 88.19%, reflecting a 2% enhancement over the base model. Additionally, it demonstrated a 9-14% reduction in memory consumption and parameter count. Real-time evaluations revealed a processing speed 14% faster than the base model, alongside achieving the highest F1-score results, which underscores its effectiveness for practical applications. Furthermore, the proposed method enhanced test accuracy in classes with approximately 50% fewer training samples by about 5%.

Conclusion: This study presents a highly efficient race classification model grounded in a modified Efficient-Net architecture that utilizes transfer learning and attention-based learning to attain state-of-the-art performance. The proposed approach not only sustains high accuracy but also ensures rapid processing speeds, rendering it ideal for real-time applications. The findings indicate that this lightweight model can effectively rival more complex and computationally intensive recent methods, providing a valuable asset for practical race classification endeavors.

Introduction

The advances of recent years in the field of artificial intelligence (AI) and deep learning (DL) have significantly improved the accuracy of facial recognition, image classification, and object detection. Among these advancements, race classification (RC) from facial images remains a critical and inherently challenging task due to the subtle differences in facial features across various

racial groups and the extensive diversity among human faces [1], [2]. This capability holds immense potential for applications in security, human-computer interaction, and social and demographic analysis.

RC is not only an academic problem but also has significant real-world implications. For instance, in security and surveillance, accurate RC can enhance monitoring and identification [3], [4]. In personalized user experiences such as augmented reality (AR) and virtual reality (VR), understanding racial features can improve user interaction [5]. In healthcare, accurate RC can aid in providing tailored medical advice and interventions, as certain medical conditions are prevalent in different racial groups [6]. However, ethical considerations are crucial to address the potential biases and fairness issues in RC.

In addition to the major benefits such as model accuracy and efficiency, this study emphasizes the importance of lightweight models in resource-constrained environments. Lightweight architectures are especially useful for deployment on devices with limited computational resources, such as mobile phones and embedded platforms [7], [8]. This makes advanced RC features feasible and applicable across various domains, from consumer electronics to remote sensing technologies. The novelty of this research lies in its holistic approach, integrating real-time performance metrics, error analysis, and detailed visualization to provide a comprehensive understanding of model behavior and identify areas for improvement. Real-time performance metrics are crucial for applications demanding instant results, while error analysis can pinpoint specific challenges and potential biases in the classification process. Visualization techniques offer intuitive insights into how the model perceives diverse racial features, aiding in the refinement of the model.

This study demonstrates the powerful combination of convolutional neural network (CNN) architectures and transfer learning techniques when applied to complex classification tasks like race recognition [9]-[12]. The findings confirm the high accuracy and efficiency of the model, establishing a foundational benchmark for further research and development in AI and deep learning. This research lays the groundwork for future endeavors in developing robust and ethical racial classification systems applicable across diverse technological and societal contexts. The quest for high-performance, ethical, and practical racial classification models represents a critical and continuously evolving frontier in the field of artificial intelligence. By addressing the complexities of race classification, this work contributes to a more equitable and responsible application of AI technologies.

Related Work

In recent years, race recognition has become a prominent topic in facial recognition and image processing [13]-[15]. Numerous studies have been conducted to improve the accuracy and efficiency of various methods for this purpose. Al-Azani and El-Alfy (2019) examined race recognition methods in challenging conditions using Histogram of Oriented Gradients (HOG) features [16]. Their research demonstrated that HOG features could be used for race recognition under various lighting and background conditions [17]. However, a major drawback of this approach is its sensitivity to changes in scale and angle, which can result in decreased accuracy when dealing with noisy images or unexpected variations due to its reliance on low-level features. In a comparative study between machine learning and deep learning methods for age, gender, and race recognition, Hamdi and Moussaoui found that deep learning methods generally performed better than machine learning methods, especially in race recognition [18]. Krishnan et al. investigated the fairness of gender classification algorithms across different gender-race groups [19]. The results indicated that there were performance disparities among these algorithms across different groups, highlighting the need to address such disparities in the development of future algorithms. Ahmed et al. utilized deep networks for race estimation. Their study showed that deep networks could achieve high accuracy in race estimation, particularly when leveraging diverse data combinations [20]. Belcar et al. focused on race recognition using Convolutional Neural Networks (CNNs) and the middle part of the face [21]. Their results indicated that utilizing specific facial regions could enhance the accuracy of race recognition algorithms.

However, reliance on specific facial parts may lead to decreased overall model performance in scenarios where these parts are not fully visible or affected by external factors [22]. Patel et al. introduced a shift-invariant deep neural network for tri-fold classification [23]. This network demonstrated strong performance against spatial variations in input data, providing notable results. Lastly, Wirayuda et al. proposed a compact-fusion feature framework for race recognition, which improved the accuracy of recognition algorithms by using combined features [24]. However, the high complexity and need for combined processing of this framework could impact its real-time performance. Despite these advancements, there remains a need for improvements in race recognition accuracy in real-world conditions and in environments with limited computational resources. This research aims to address this gap by presenting an optimized model that reduces resource consumption and parameters while maintaining high accuracy. Our model, utilizing advanced optimization techniques and novel methods, is designed to offer robust performance under varying and challenging conditions while minimizing computational demands.

Proposed Method

In this section, proposed method is presented for the classification of races from facial images using transfer learning based on state-of-the-art deep learning architectures. At the center of the approach is the Efficient-Net model, which is well known for its trade-off between top results and low computation cost. Our method addresses many built-in difficulties of RC, such as subtle differences in facial features among racial groups and the need for a balanced, diverse dataset to train the model effectively. We first used Efficient-Net as the base model since it is already known for its success in image classification works [25]. The advantage of this architecture is that it adopts a compound-scaling approach, where all network dimensions are scaled up uniformly to bring about improved performance while consuming less computation cost. However, since RC is very specific in nature, using Efficient-Net directly would not be enough. We will make use of transfer learning to fine-tune the pre-trained Efficient-Net model to our targeted RC task. By this method, we retrain the upper layers to adapt the model with the unique characteristics of racial features. To enhance the preprocessing step, we employed the Multi-task Cascaded Convolutional Networks (MTCNN) for accurate face detection and alignment [26]. MTCNN is effective in extracting facial regions from images, ensuring that the input to the model focuses solely on the relevant facial features. This step is crucial for improving the overall accuracy of the RC process by eliminating background noise and variations in face alignment. Fig. 1 illustrates the progression of the proposed method.

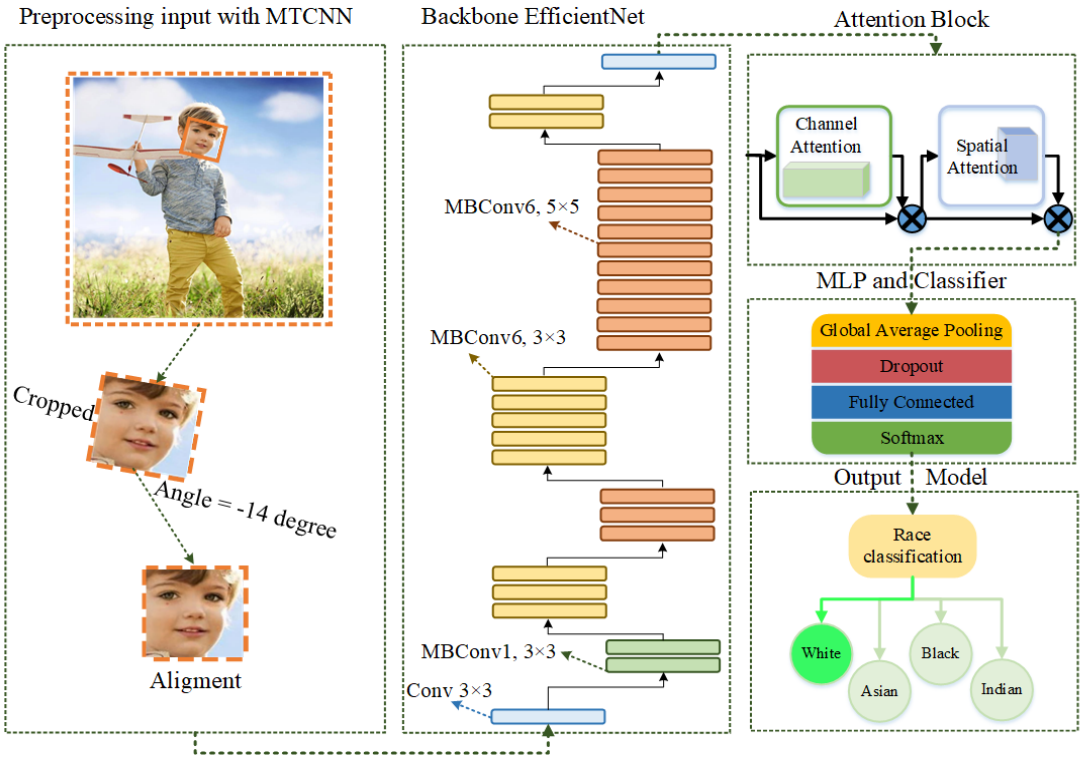


Fig. 1: An overview of proposed method.

A. Attention Mechanism

In the proposed methodology, the integration of the convolutional block attention module (CBAM) into the Efficient-Net architecture, specifically positioned after the convolutional layers, introduces a refined approach to enhancing feature extraction for race classification. This attention mechanism enables the network to dynamically focus on both salient channels and critical spatial regions within the feature maps, which is vital for effectively distinguishing subtle racial characteristics. In race classification, where minor facial feature variations across different ethnicities are key, conventional convolutional layers may fail to sufficiently capture these nuanced differences. CBAM addresses this limitation by applying dual attention mechanisms, improving the network's sensitivity to discriminative features [27]. CBAM module shows in Fig. 2.

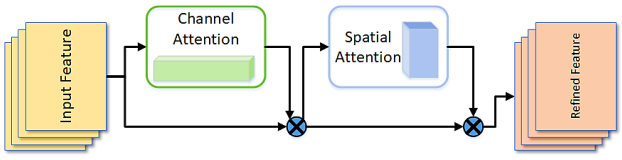


Fig. 2: Schematic representation of the CBAM architecture.

Formally, given an intermediate feature map FM, which belongs to a three-dimensional space denoted $H \times W \times C$, where H represents the height of the feature map, W represents its width, and C indicates the number of channels, CBAM first applies channel attention followed by spatial attention. In this context, the dimensions H, W, and C correspond to the spatial and depth characteristics of the feature map extracted from the convolutional layers, with the height and width representing the two-dimensional spatial resolution and the channel reflecting the depth or number of filters applied in the convolutional process as expressed in (1).

$$CA(FM) = \sigma(MLP(AvgPool(FM)) + MLP(MaxPool(FM))) \quad (1)$$

where MLP denotes a multi-layer perceptron, and σ represents the sigmoid activation function. This operation selectively enhances important channels by considering both average and max-pooled representations of the feature map. Subsequently, spatial attention is applied as expressed in (2).

$$SA(FM) = \sigma(Conv_{7\times7}(AvgPool(FM); MaxPool(FM)) \quad (2)$$

where $Conv_{7\times7}$ signifies a convolutional layer that processes the concatenation of average and max-pooled feature maps across the channel dimension, thereby refining spatial feature selection. The incorporation of CBAM in this manner allows for a more targeted feature representation, capturing both global dependencies and localized variations in facial structures pertinent to racial differentiation. By combining channel and spatial attention, the proposed approach offers a novel enhancement to Efficient-Net, enabling the model to better distinguish subtle racial traits, thereby improving classification accuracy in datasets with complex race-related features.

B. Data Balancing

Finally, the issue of data imbalance is addressed, which makes the classification task challenging with race datasets. An imbalanced dataset can further lead to biased models, performing poorly on the under-represented classes. In order to cope with that, we used a lot of data augmentation and resampling strategies. Data augmentation will create different training examples for rotations, scaling, flipping, and so on. Resampling can be implemented using techniques such as SMOTE (Synthetic Minority Over-sampling Technique) or oversampling for a certain group to ensure its representation in the dataset. Table 1 presents the pseudocode of the proposed method.

In the process of machine learning with imbalanced datasets, the SMOTE (Synthetic Minority Over-sampling Technique) is utilized as an advanced method for balancing the number of samples across classes.

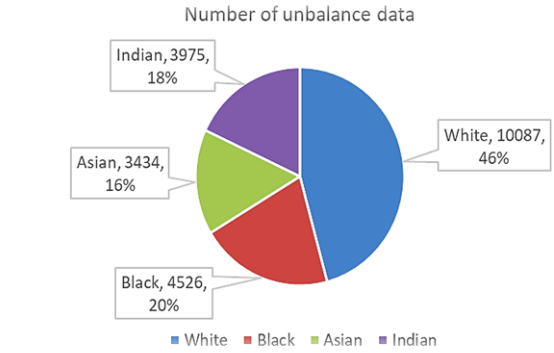
Table 1: Pseudocode of proposed method

- Start
1. # Load and preprocess the data
2. Data_list = Load data (dataset_path)
3. Extract face images by ***MTCNN*** **()**
4. Split data into ***training*** and ***testing*** sets
5. Balance ***training*** data with ***SOMTE*** method
6. Split balance data into ***training*** and ***validation*** sets
7. # Define the Efficient-Net model
8. Initialize ***Efficient-NetB3*** with ***Image-Net*** weights
9. Add ***Attention CBAM*** to ***base_model***
10. Add ***GlobalAveragePooling*** layer to ***base_model***
11. Add ***Dropout*** layer with a dropout rate of ***0.5***
12. Add ***Dense layer*** with ***softmax*** activation function
13. Compile model using ***'Adam'*** optimizer and ***'sparse_categorical_crossentropy'*** loss function
14. # Train the model
15. Trained_model = *Train model* (model, training_set_images)
16. # Metrics: ***Accuracy***, ***Precision***, ***Recall***, and ***F1 score***
17. # Evaluate the model
18. Validation_metrics = *EvaluateModel* (trained_model, validation_set_images)
19. # Test the model
20. Test_metrics = *TestModel* (trained_model, test_set_images)
- return ***Validation_metrics, Test_metrics***

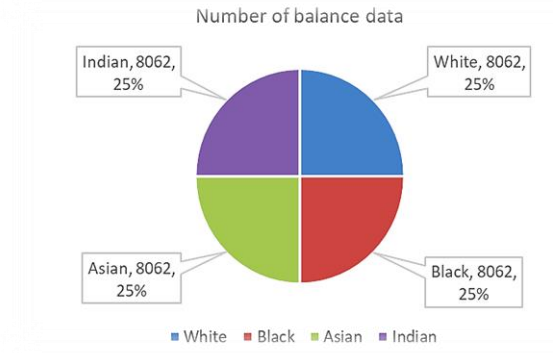
This technique generates new data by creating samples based on the existing minority class samples rather than merely duplicating them randomly. This approach not only enhances the diversity of the data but also mitigates the bias resulting from class imbalance. Consequently, the SMOTE technique has been employed in this study to address the issue at hand.

In Fig. 3, the total number of images is 22,013, whereas the total should amount to 22,022 based on the individual class sample counts. During the filtration process, nine samples were removed from the data frame due to being corrupted. For the training and processing of deep learning models, the dataset consisting of 22,013 images with four distinct racial labels (White, Black, Asian, and Indian) was utilized. To achieve balance in data distribution and enhance model quality, the dataset was initially divided into two main sections: training (train) and testing (test), with 80% of the data allocated for training and 20% for testing. Subsequently, the training data was further divided into two subsets for model validation: final training (train final) and validation. Accordingly, 90% of the training data was designated for final training and 10% for validation. This strategic division ensured that the racial distribution was preserved in each subset, allowing the model to be trained effectively on

balanced data. Finally, after applying the SMOTE technique to equalize the number of samples in each class, the new distribution resulted in 8,062 images for each racial category.



(a)



(b)

Fig. 3: Distribution of Images by racial labels before (a) and after (b) data filtration.

Results and Discussion

This section presents the main findings from our experiments, showcasing the performance of the proposed method across various evaluation metrics. The results are summarized in tables and figures to provide a clear and concise overview of the data. These findings will, in the following sections, serve as the basis for discussing their significance and relevance to existing research in the field.

A. Criterion

In evaluating the performance of machine learning models, four primary metrics are commonly used: accuracy, precision, recall, and the F1-score. These metrics provide a comprehensive assessment of a model's effectiveness [28].

Accuracy: This metric simply represents the ratio of correctly predicted instances to the total number of predictions. In other words, in (3) accuracy measures the overall correctness of a model by considering both true positives and true negatives.

$$Accuracy = \frac{TP + TN}{TP + TN + FP + FN} \tag{3}$$

where TP stands for true positives, TN for true negatives, FP for false positives, and FN for false negatives.

Precision: Precision indicates the proportion of positive predictions that are actually correct. This metric is particularly important when the cost of false positives is high in (4).

$$Precison = \frac{TP}{TP + FP} \tag{4}$$

Recall: Recall measures in (5) the percentage of actual positive instances that are correctly identified by the model. It is crucial when the cost of missing positive cases is high.

$$Recall = \frac{TP}{TP + FN} \tag{5}$$

F1-Score: The F1-score is the harmonic mean of precision and recall, providing a balanced measure that considers both metrics in (6). It is especially useful when dealing with imbalanced class distributions.

$$F1 - Score = 2 \times \frac{Precision \times Recall}{Precision + Recall} \tag{6}$$

RAM: Memory consumption (RAM) refers to the amount of memory required for storing data and model parameters during the training and evaluation of machine learning models. This metric can be calculated using (7).

$$RAM_{MB} = \frac{P \times Size\_of\_data\_type}{1024^2} \tag{7}$$

where P is the total number of parameters in the network and $Size\_of\_data\_type$ is the size of the data type (typically 4 bytes for float32). The number of parameters includes weights and biases across all layers, including convolutional and dense layers. Specifically, in a convolutional layer with K filters, C input channels, and kernel dimensions $H \times W$, the number of parameters is calculated in (8).

$$P_{conv} = K \times (H \times W \times C + 1) \tag{8}$$

B. Dataset

UTK-Face provides over 20,000 facial images with very wide coverage of racial backgrounds, age groups, and both genders. This will come in quite handy for research into racial classification, where the dataset is diverse and comes with a detailed label including race, age, and gender for each image [29]. The dataset allows for the making of a model capable of precise and fair RC from facial features, thus being important for its real-world application. Fig. 4 displays a diverse set of facial images representing different racial categories included in the dataset.

Fig. 4: Sample images from the UTK-Face dataset [29].

Table 2 provides a comparison between the base model and the proposed model in terms of parameters and memory usage. The base model comprises a total of 12,324,539 parameters, of which 12,237,236 are trainable, resulting in an approximate RAM consumption of 47.01 MB, indicating a higher resource requirement. In contrast, proposed model 1 demonstrates a reduced parameter count of 10,789,683, with 10,702,380 trainable parameters, leading to a lower memory consumption of 41.16 MB. Proposed model 2, although slightly more resource-intensive than Proposed model 1, still presents an improvement over the base model, featuring 11,379,605 total parameters and a RAM requirement of 43.41 MB. This reduction in both total and trainable parameters, along with decreased memory consumption, highlights the greater efficiency of the proposed models while ensuring a significant number of trainable parameters, which is essential for maintaining model performance. Additionally, the lower memory footprint of these models makes them particularly well-suited for deployment in resource-constrained environments, enabling wider implementation in scenarios where computational resources are limited.

Table 2: Model parameters and memory usage overview

Model	Parameters		
	Total	Trainable	RAM (MB)
Base model	12,324,539	12,237,236	47.01
Proposed model 1	10,789,683	10,702,380	41.16
Proposed model 2	11,379,605	11,292,302	43.41

Table 3 provides a comprehensive analysis of the performance metrics for the training and testing phases of the models evaluated on an NVIDIA Tesla T4 with 16GB GDDR6 RAM. As indicated in Fig. 3, the base model, trained on 22,013 images, achieved a training duration of 2249 seconds over 50 epochs. In contrast, both Proposed model 1 and Proposed model 2, which were trained on an expanded dataset of 32,248 images, required slightly longer training times of 2275 seconds and 2282 seconds, respectively.

In terms of testing efficiency, the base model demonstrates a processing time of 0.35 milliseconds per image, resulting in a frame rate of 2827 frames per second. Proposed model 1 shows improved performance with a reduced image processing time of 0.28 milliseconds and an increased frame rate of 3571 frames per second. Proposed model 2 also performs well, processing each image in 0.30 milliseconds and achieving a frame rate of 3352 frames per second. These results indicate that while the proposed models have a higher training time, they capitalize on a larger dataset to enhance their efficiency during testing. The substantial parameter reductions observed in the proposed models, as outlined in Table 2, coupled with their improved testing speeds, suggest that these models are not only more resource-efficient but also better suited for applications that demand high-speed image processing. This efficiency in real-time applications is paramount, highlighting the potential for deploying the proposed models in scenarios requiring rapid processing capabilities.

Table 3: Performance metrics for model training and testing on an INVIDIA Tesla T4 with 16GB GDDR6 RAM

Method	Training time (50 epoch) Second	Test (time per image) Millisecond	Test frame per second
Base model	**2249**	0.35	2827
Proposed model 1	2275	**0.28**	**3571**
Proposed model 2	2282	0.30	3352

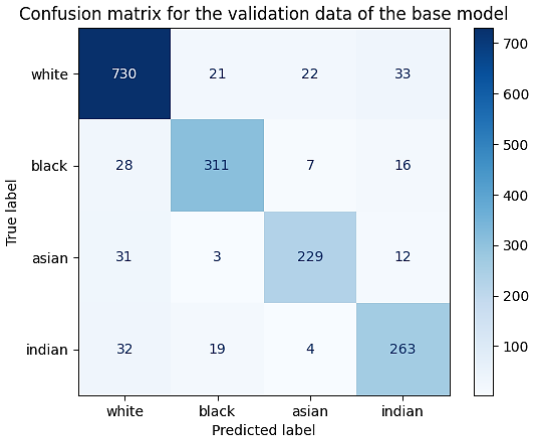
Table 4 provides a detailed comparison of model performance metrics across different racial categories, revealing significant insights into the effectiveness of the proposed methodologies. The base model without balance and attention modules shows robust accuracy, particularly in the White and Black categories, with notable recall for White (0.94). However, it struggles with the Indian category, exhibiting lower accuracy (0.85) and precision (0.68). Proposed Method 1, which employs balanced data, demonstrates improvements in precision for the White category (0.92) but experiences a decline in Indian category performance, achieving 0.74 accuracy. In contrast, Proposed Method 2, which integrates both balanced data and an attention module, achieves the highest overall metrics, particularly excelling in the Asian category (0.90 accuracy) and significantly improving Indian classification metrics to 0.86 accuracy and 0.76 precision. This comprehensive analysis underscores the importance of data balancing and advanced modeling techniques in enhancing classification accuracy, particularly for underrepresented groups, thereby demonstrating the proposed methods' superior capability in addressing the challenges faced by the base model.

Table 4: Detailed performance metrics for RC models

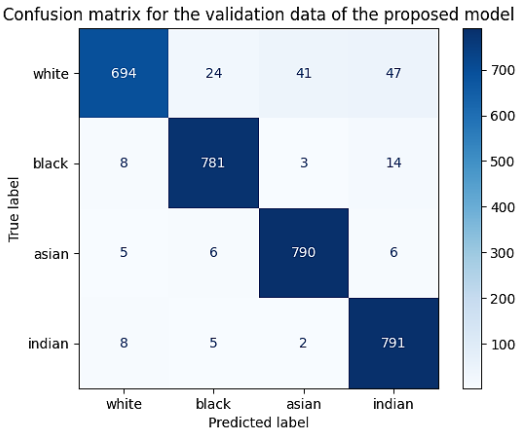
Method	Race	Criterion				
		Accuracy	Precision	Recall	F1-Score	Support-test
Base model without balance and attention module	White	0.86	0.84	**0.94**	0.89	2016
	Black		0.88	0.83	0.85	905
	Asian		0.88	0.84	0.86	687
	Indian		0.85	0.68	0.76	795
Proposed method_1 with balance data (Efficient-Net-BD)	White	0.87	**0.92**	0.88	0.90	2016
	Black		**0.88**	0.85	0.87	905
	Asian		0.88	0.87	0.87	687
	Indian		0.74	**0.85**	0.79	795
Proposed method_2 with balance data and attention module (Efficient-Net-BD-AM)	White	**0.88**	0.89	0.93	**0.91**	2016
	Black		0.87	**0.88**	**0.87**	905
	Asian		**0.90**	**0.89**	**0.90**	687
	Indian		**0.86**	0.76	**0.81**	795

Fig. 5 presents the confusion matrices for the validation dataset, featuring two distinct matrices that demonstrate the advantages of using a balanced dataset as well as the strengths of the proposed model in developing effective classification models. Fig. 5(a) illustrates the performance of the base model, while Fig. 5(b) pertains to proposed model 2. In Fig. 5(a), the unbalanced nature of the dataset is prominently displayed, highlighting its impact on classification accuracy across various racial categories. This matrix clearly shows the detrimental effect of imbalance, particularly in classes with fewer samples. In contrast, Fig. 5(b) showcases the application of a balanced dataset, emphasizing the strengths of the proposed method. For instance, the number of samples in the Indian class has increased from 318 to 806. This increase leads to a significant reduction in misclassifications, with errors decreasing from 55 to 15 during the validation phase. Moreover, similar trends can be observed across other classes, indicating a systematic improvement in classification performance. The results underscore the importance of data representation in training processes, suggesting that such enhancements can lead to a more robust and reliable model capable of better generalization to unseen data.

Fig. 6 presents the confusion matrices for the test dataset, consisting of two separate matrices. Since the test data were partitioned at the beginning of the process, these data were kept untouched for the testing phase in this research. Fig. 6(a) illustrates the performance of the base model, while Fig. 6(b) corresponds to proposed model 2. In Fig. 6(a), the performance of the base model on the test data is observed. In contrast, Fig. 6(b) shows the confusion matrix for the test data using proposed model 2. In these matrices, it can be seen that in matrix Fig. 6(a), the base method performed better in the white class compared to the proposed model. However, in this study, the classification of the white race has less challenge for the models, given that the database for the white class, as shown in Fig. 3, has greater diversity and accounts for 46% of the data. Therefore, improvements are needed in other classes. In Fig. 6(b), it is observed that the black class has 20 fewer error samples compared to the base model, the Asian class shows an improvement of 30 samples, and finally, the Indian class improved by 72 samples. Thus, the strength of proposed model 2 is demonstrated in other classes with fewer sample sizes. As a result, this model can be confidently utilized for RC applications.



(a)



(b)

Fig. 5: Confusion matrices for the validation dataset, illustrating the performance of the base model (a) and the proposed model 2 (b).

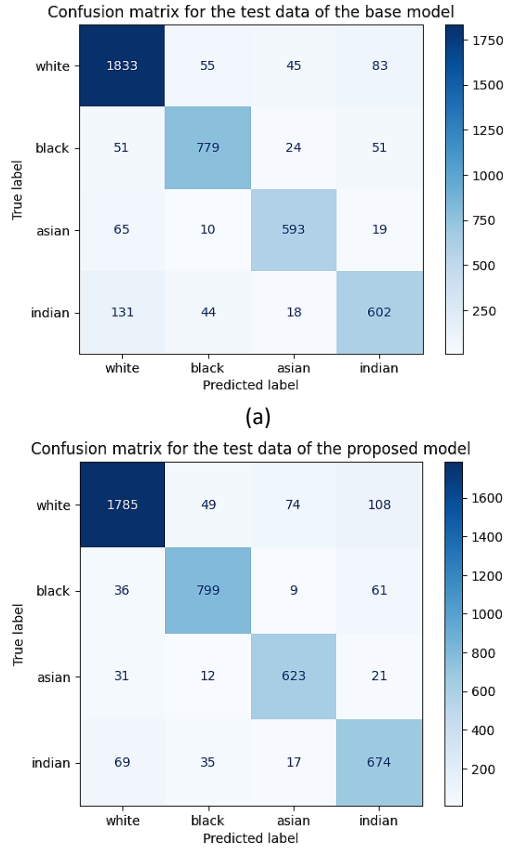


Fig. 6: Confusion matrices for the test samples, illustrating the performance of the base model (a) and the proposed model 2 (b).

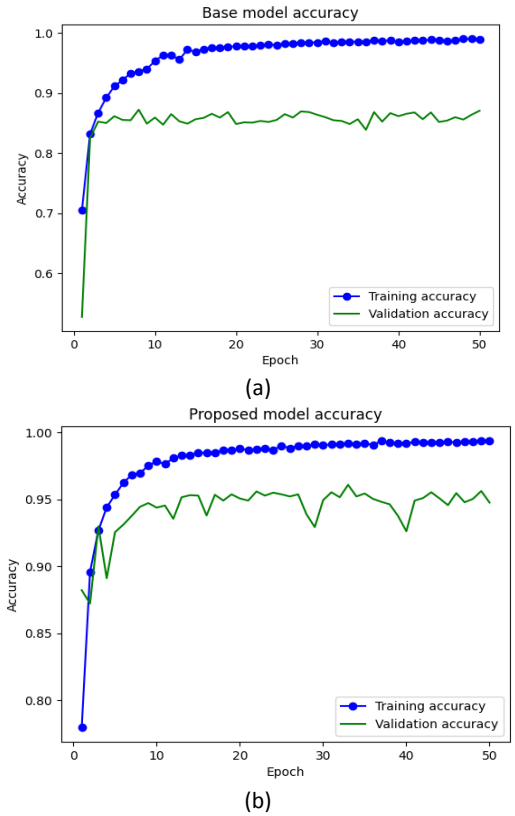


Fig. 7: Epoch-wise accuracy progression for RC model. (a): base model and (b) proposed model 2.

Fig. 7 illustrates the accuracy progression over training epochs for both the base model and proposed model 2. The x-axis represents the number of epochs, while the y-axis shows the model's accuracy across RC categories. In Fig. 7(a), which corresponds to the base model, the accuracy stabilizes at around 87% by the end of the training process. In contrast, Fig. 7(b) presents the performance of proposed model 2, where the accuracy reaches a significantly higher value of 95%. Additionally, it is evident that proposed model 2 not only achieves a better final accuracy but also starts with a higher accuracy at the beginning of the training process compared to the base model. This demonstrates the superior learning capability and faster adaptation of the proposed model in comparison to the base model.

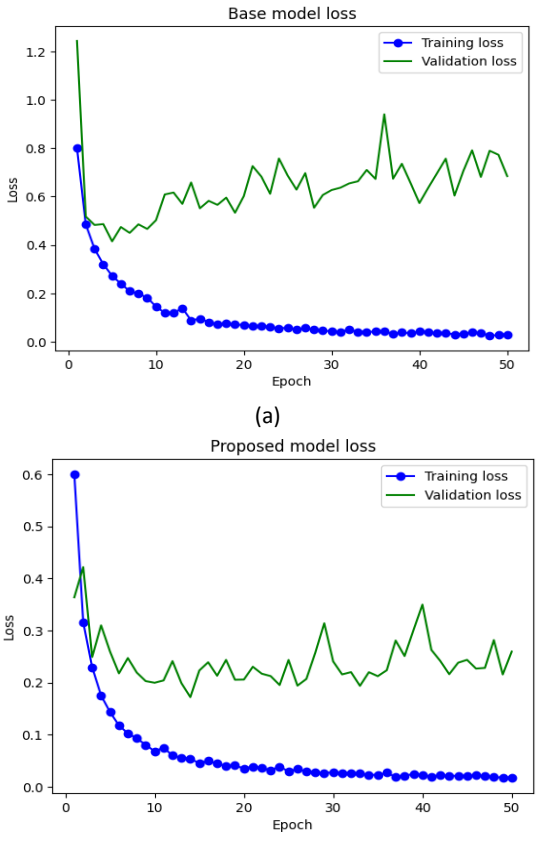


Fig. 8: Epoch-wise Loss trend for RC model. (a): base model and (b) proposed model 2.

Fig. 8 presents the loss progression over the training epochs for both the base model and proposed model 2, offering a detailed comparison of their performance. The x-axis represents the number of epochs, while the y-axis reflects the corresponding loss values during both the training and validation phases. In Fig. 8(a), which represents the base model, the initial loss is nearly double that of proposed model 2, as shown in Fig. 8(b). This disparity suggests that the base model faces greater difficulty in learning at the start of training. As training proceeds, the final validation loss for the base model

remains approximately three times higher than that of proposed model 2, underscoring a significant difference in convergence between the models. Moreover, the base model demonstrates increasing loss during later stages, indicating instability and a lack of convergence. In stark contrast, proposed model 2 exhibits a much more stable and lower loss curve, reflecting a more efficient learning process. Furthermore, the base model's loss curve shows greater variability, with frequent fluctuations throughout the training process, signaling potential issues with optimization. Conversely, proposed model 2 maintains a smoother and more consistent loss trajectory, indicating better control and robustness in learning. This clear improvement in stability and overall performance highlights the advantages of the proposed model in handling the classification task effectively. In Fig. 9 the plot illustrates the precision, recall, and F1-score for each class based on the model's performance. Each point represents the respective metric's score for the individual racial classes (White, Black, Asian, and Indian). Additionally, the dashed line indicates the overall accuracy of the model, providing a reference point to assess how well the model performs across different classes relative to its general performance. Fig. 9 illustrates the performance of the model using both macro and weighted averages. The macro-average reflects the model's performance equally across all classes, without considering the size of each class, providing insight into how the model performs on average for each class, regardless of the number of samples.

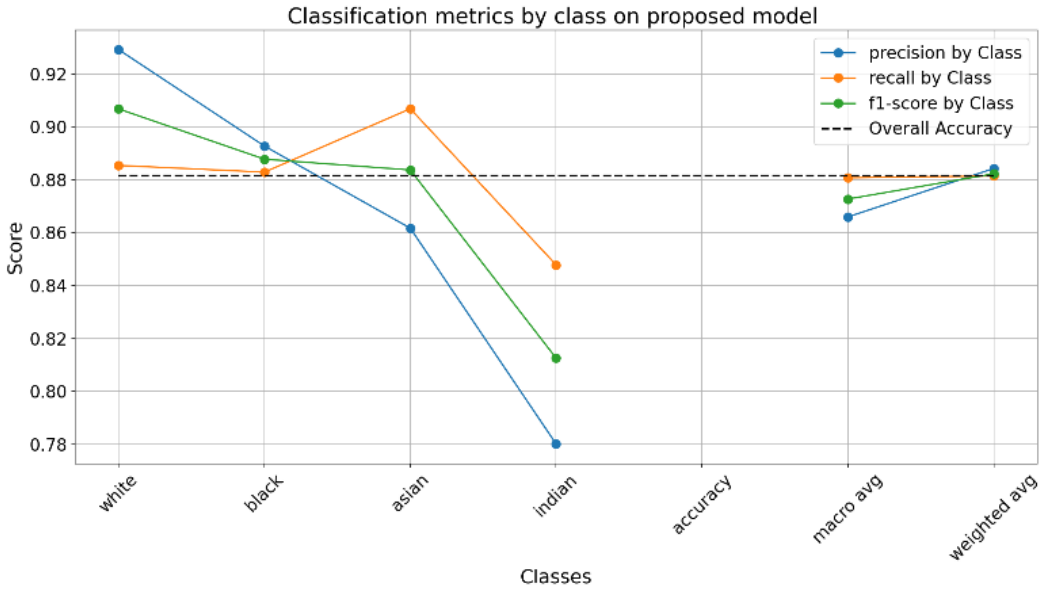


Fig. 9: Illustration of precision, recall, and F1-score for each class based on the proposed model 2 performance.

This allows for a balanced view of performance across minority and majority classes. In contrast, the weighted average evaluates the model's performance based on the sample size of each class, assigning greater importance to larger classes. This metric offers a more comprehensive understanding of the model's overall effectiveness, particularly in handling the class imbalance present in the dataset, by reflecting the influence of uneven class distributions on the final outcomes.

Table 5 reflects the continuous advancements in RC methodologies, particularly in the context of challenging and real-world conditions, as analyzed through the UTK-Face dataset. Al-Azani and El-Alfy (2019) laid the foundation by using Histogram of Oriented Gradients (HOG) features, achieving an accuracy of 69.68%. Despite the utility of HOG for handling variations in lighting and background, this approach showed limitations due to its sensitivity to scale and angle changes, leading to performance degradation in noisy environments.

Following this, Hamdi and Moussaoui (2020) demonstrated the superiority of deep learning methods over traditional machine learning approaches, achieving a higher accuracy of 78.88%. Similarly, Krishnan et al. (2020) highlighted performance disparities across different gender-race groups, reaching an accuracy of 79.49% but revealing the necessity of more equitable and robust models. Ahmed et al. (2022) advanced the field with deep networks, optimizing the use of diverse data combinations to achieve a notable accuracy of 77.50%. Meanwhile, Belcar et al. (2022) refined race recognition by concentrating on specific facial regions like the middle part of the face, reaching an accuracy of 80.34%. However, this approach's reliance on specific regions introduced vulnerabilities when those parts were occluded or affected. Deviyani (2022) marked a significant leap in accuracy, achieving 87.20% by employing StarGAN, and analyzing multiple datasets comprehensively. This method's versatility made it one of

the most thorough analyses in the domain. Patel et al. (2023) and Wirayuda et al. (2023) added further refinements with accuracies of 76.22% and 82.19%, respectively, employing shift-invariant architectures and compact-fusion frameworks, though these methods faced challenges in terms of complexity and real-time applicability.

Table 5: Numerical results on the UTK-Face dataset

Method	Accuracy (%)
Al-Azani and El-Alfy (2019) [16]	69.68
Hamdi and Moussaoui (2020) [18]	78.88
Krishnan et al. (2020) [19]	79.49
Ahmed et al. (2022) [20]	77.50
Belcar et al. (2022) [21]	80.34
Deviyani (2022) [30]	87.20
Patel et al. (2023) [23]	76.22
Wirayuda et al. (2023) [24]	82.19
Base model	86.46
Proposed model1	86.82
Proposed model2	**88.19**

In comparison, our base model shows substantial improvement, achieving 86.46% accuracy. The proposed model 1 slightly surpasses this with 86.82%, while proposed model 2 demonstrates the highest accuracy at 88.19%, outperforming all prior models. Notably, proposed model 2 leverages advanced optimization techniques and superior feature extraction, making it highly efficient under diverse and challenging conditions. Additionally, it maintains low computational costs, addressing the gap highlighted in previous studies regarding real-time performance and resource efficiency. This underlines the robustness and generalizability of the proposed model for race recognition tasks across different demographic groups.

Fig. 10 showcases a selection of correctly classified test samples, highlighting the model's capability to accurately predict various racial categories across a diverse set of facial images. The displayed images, along with their corresponding true and predicted labels, demonstrate the strong alignment between the model's predictions and actual labels. Notably, the proposed model exhibits robust performance across a wide range of racial categories and age groups, effectively handling the inherent diversity in both age and race within the dataset. This indicates the model's adaptability and precision in classifying facial images under varying demographic conditions, further validating its strength in real-world applications.

Fig. 11 presents a facial sample alongside its incorrect prediction in comparison to the actual label. The inaccuracies in test predictions can be attributed to three main factors: low image quality, which hampers the model's ability to distinguish subtle facial features; labeling errors, which result in incorrect associations between images and their racial labels—an example being a White male misclassified as Black. Additionally, the similarity among racial classes poses a challenge, as overlapping features can complicate differentiation. Another potential source of error arises from the data augmentation methods employed during the balancing process; if erroneous data exists in the dataset, it can propagate errors further. Addressing these issues through improved image quality, accurate labeling, and enhanced model sensitivity is crucial for bolstering prediction accuracy.

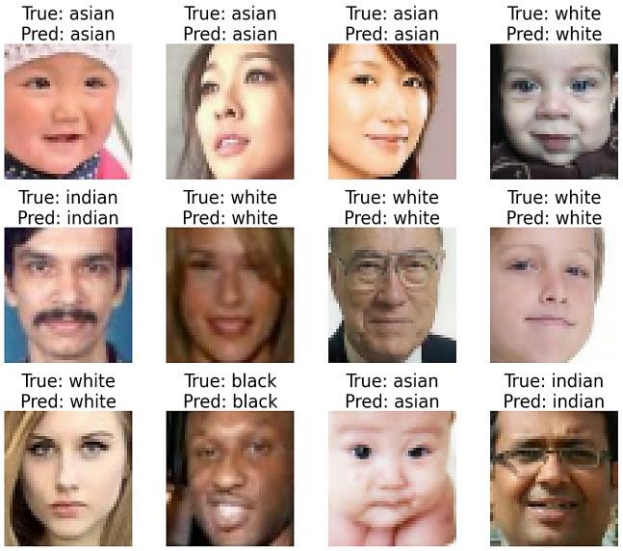


Fig. 10: Examples of true predicted test samples.

Fig. 11: Examples of incorrectly predicted test samples.

Conclusion

This study presents an innovative solution to the race classification (RC) problem by utilizing the Efficient-Net model in conjunction with transfer learning techniques to enhance performance. A key strength of this approach lies in the preprocessing of input images, achieved through the Multi-task Cascaded Convolutional Network (MTCNN) for accurate face detection and alignment. This preprocessing not only isolates facial features but also minimizes background noise, thereby establishing a robust foundation for effective classification. Addressing the issue of data imbalance was another crucial aspect of

our methodology. We implemented sophisticated techniques such as data augmentation and oversampling to generate a diverse set of training samples. Data augmentation involved applying transformations like rotation and scaling, which aided in enriching the training dataset. Oversampling was employed to mitigate class imbalance, particularly for racial categories with fewer training samples. This focus on enhancing dataset quality was effective in improving the model's generalization across different racial groups. Additionally, one of the key advantages of proposed model 2 is its ability to significantly reduce error rates compared to the base model in classes with limited data, positively impacting accuracy in these categories. The use of Efficient-Net, recognized for its optimal balance between accuracy and computational efficiency, has been specifically tailored for RC tasks. This adjustment enables the model to effectively capture subtle variations in facial features among different racial groups, thereby contributing to improved accuracy. Evaluation demonstrated that this model exhibits remarkable real-time performance. Assessment metrics, including accuracy, precision, recall, and F1 score, indicated overall high performance, although some variability was observed among racial categories. This variability highlights ongoing challenges related to classification accuracy, particularly in classes with limited training data. Factors such as low image quality, erroneous labeling, and similarities among specific racial features have been identified as contributors to classification errors. Future efforts should focus on enhancing data quality, correcting labeling inaccuracies, and refining the model to address these issues.

Author Contributions

M. Rohani designed the experiments and developed the overall methodology, wrote manuscript, conducted the data analysis, statistical evaluations and collected and preprocessed the dataset. H. Farsi interpreted the results and has drawn the general road map. S. Mohamadzadeh edited and revised the manuscript.

Acknowledgment

The authors wish to express their profound gratitude to the esteemed reviewers and editors of JECEI for their meticulous review, constructive feedback, and invaluable suggestions, which have significantly enhanced the quality of this article. Furthermore, the authors extend their sincere appreciation to the editorial board for their professional guidance and exemplary handling of the manuscript during the review process.

Conflict of Interest

The authors declare no potential conflict of interest regarding the publication of this work. In addition, the ethical issues including plagiarism, informed consent, misconduct, data fabrication and, or falsification, double publication and, or submission, and redundancy have been completely witnessed by the authors.

Abbreviations

AI	Artificial Intelligence
DL	Deep Learning
RC	Race Classification
AR	Augmented Reality
VR	Virtual Reality
CNN	Convolutional Neural Network
MTCNN	Multi-task Cascaded Convolutional Networks
SMOTE	Synthetic Minority Over-sampling Technique
HOG	Histogram of Oriented Gradients
CA	Channel attention
SA	Spatial attention
FM	Feature map
AvgPool	Average pooling
MaxPool	Maximum pooling
CBAM	Convolutional block attention module

References

[1] E. Ghasemi Bideskan, S. M. Razavi, S. Mohamadzadeh, M. Taghippour, "Facial expression recognition through optimal filter design using a metaheuristic kidney algorithm," J. Electr. Comput. Eng. Innovations (JECEI), 12(2): 425-438, 2024.

[2] M. Rohani, H. Farsi, S. Mohamadzadeh, "Deep multi-task convolutional neural networks for efficient classification of face attributes," Int. J. Eng., 36(11): 2102-2111, 2023.

[3] A. Nieves Delgado, "Race and statistics in facial recognition: Producing types, physical attributes, and genealogies," Social Stud. Sci., 53(6): 916-937, 2023.

[4] M. Rohani, H. Farsi, S. Mohamadzadeh, "Facial feature recognition with multi-task learning and attention-based enhancements," Iran. J. Energy Environ., 16(1): 136-144, 2025.

[5] D. M. Hilty, A. M. P. Schmid, R. E. Holbrook, J. P. Greer, "A review of telepresence, virtual reality, and augmented reality applied to clinical care," J. Technol. Behav. Sci., 5(1): 178-205, 2020.

[6] C. Lu, R. Ahmed, A. Lamri, S. S. Anand, "Use of race, ethnicity, and ancestry data in health research," PLOS Global Public Health, 2(9): 1060-1076, 2022.

[7] S. Minaee, A. Abdolrashidi, H. Su, M. Bennamoun, D. Zhang, "Biometrics recognition using deep learning: A survey," Artif. Intell. Rev., 56(8): 8647-8695, 2023.

[8] I. Adjabi, A. Ouahabi, A. Benzaoui, A. Taleb-Ahmed, "Past, present, and future of face recognition: A review," Electronics, 9(8): 1188-1202, 2020.

[9] K. Weiss, T. M. Khoshgoftaar, D. Wang, "A survey of transfer learning," J. Big Data, 3(1): 1-40, 2016.

[10] S. Zahiri, R. Iranpoor, N. Mehrshad, "Paying attention to the features extracted from the image to person re-identification," J. Electr. Comput. Eng. Innovations (JECEI), 13(2): 267-274, 2025.

[11] Z. Ghasemi-Naraghi, A. Nickabadi, R. Safabakhsh, "Multi-Task learning using uncertainty for realtime multi-person pose estimation," J. Electr. Compu. Eng. Innovations (JECEI), 12(1): 147-162, 2024.

[12] M. Rohani, H. Farsi, S. H. Zahiri, "Statistical analysis and comparison of the performance of meta-heuristic methods based on their powerfulness and effectiveness," J. Inf. Syst. Telecommun. (JIST), 10(37): 49-59, 2022.

[13] M. Wang, W. Deng, "Deep face recognition: A survey," Neurocomputing, 429(1): 215-244, 2021.

[14] S. Minaee, A. Abdolrashidi, H. Su, M. Bennamoun, D. Zhang, "Biometrics recognition using deep learning: A survey," Artif. Intell. Rev., 56(8): 8647-8695, 2023.

[15] M. J. A. Dujaili, "Survey on facial expressions recognition: databases, features and classification schemes," Multimedia Tools Appl., 83(3): 7457-7478, 2024.

[16] S. Al-Azani, E. S. El-Alfy, "Ethnicity recognition under difficult scenarios using HOG," J. Electr. Eng. Autom., 1(1): 1-10, 2019.

[17] M. Ruhani, H. Farsi, S. Mohamadzadeh, "Object tracking in video with correlation filter and using histogram of gradient feature," J. Soft Comput. Inf. Technol., 9(4): 43-55, 2020.

[18] S. Hamdi, A. Moussaoui, "Comparative study between machine and deep learning methods for age, gender, and ethnicity identification," in Proc. 2020 4th International Symposium on Informatics and its Applications (ISIA):1-6, 2020.

[19] A. Krishnan, A. Almadan, A. Rattani, "Understanding fairness of gender classification algorithms across gender-race groups," in Proc. 2020 19th IEEE International Conference on Machine Learning and Applications (ICMLA): 1028-1035, 2020.

[20] M. A. Ahmed, R. D. Choudhury, K. Kashyap, "Race estimation with deep networks," J. King Saud Univ. Comput. Inf. Sci., 34(7): 4579-4591, 2022.

[21] D. Belcar, P. Grd, I. Tomičić, "Automatic ethnicity classification from middle part of the face using convolutional neural networks," Informatics, 9(1): 18-32, 2022.

[22] S. Li, W. Deng, "Deep facial expression recognition: A survey," IEEE Trans. Affective Comput., 13(3): 1195-1215, 2020.

[23] S. Patel, V. Srivastava, A. Bajpai, "Three fold classification using shift invariant deep neural network," in Proc. 2023 9th International Conference on Advanced Computing and Communication Systems (ICACCS): 787-791, 2023.

[24] T. A. B. Wirayuda, R. Munir, A. I. Kistijantoro, "Compact-fusion feature framework for ethnicity classification," Informatics, 10(2): 51-84, 2023.

[25] M. Tan, Q. Le, "EfficientNet: Rethinking model scaling for convolutional neural networks," in Proc. International Conference on Machine Learning: 6105-6114, 2019.

[26] X. Li, Z. Yang, H. Wu, "Face detection based on receptive field enhanced multi-task cascaded convolutional neural networks," IEEE Access, 8: 174922-174930, 2020.

[27] S. Woo, J. Park, J. Y. Lee, I. S. Kweon, "CBAM: Convolutional block attention module," in Proc. European Conference on Computer Vision (ECCV): 3-19, 2018.

[28] R. Yacouby, D. Axman, "Probabilistic extension of precision, recall, and F1 score for more thorough evaluation of classification models," in Proc. First Workshop on Evaluation and Comparison of NLP Systems: 79-91, 2020.

[29] Z. Zhang, Y. Song, H. Qi, "Age progression/regression by conditional adversarial autoencoder," in Proc. IEEE Conference on Computer Vision and Pattern Recognition: 5810-5818, 2017.

[30] A. Deviyani, "Assessing dataset bias in computer vision," arXiv preprint arXiv: 2205.01811, 2022.

Biographies

Mehrdad Rohani received his M.Sc. degree in Telecommunications Engineering from Birjand University, Birjand, Iran, in 2018. He is currently pursuing a Ph.D. degree in Electrical Engineering with a focus on Telecommunications at Birjand University. His research interests encompass Machine Learning, Image Processing, Computer Vision, and Deep Learning Algorithms.

- Email: m.ruhani@birjand.ac.ir
- ORCID: 0000-0003-2930-019X
- Web of Science Researcher ID: LQJ-4143-2024
- Scopus Author ID: NA
- Homepage: NA

Hasan Farsi received his B.Sc. and M.Sc. degrees in Communication Engineering from Sharif University of Technology, in 1992 and 1994, and his Ph.D. in Communication Engineering from the University of Surrey, UK, in 2003. He is currently a Professor in the Department of Electrical and Computer Engineering at the University of Birjand. His research interests include deep learning, image processing, signal processing.

- Email: hfarsi@birjand.ac.ir
- ORCID: 0000-0001-6038-9757
- Web of Science Researcher ID: NA
- Scopus Author ID: NA
- Homepage: https://cv.birjand.ac.ir/hasanfarsi/fa

Sajad Mohamadzadeh received his B.Sc. degree in Communication Engineering from the University of Sistan and Baluchestan, Iran, in 2010, and his M.Sc. and Ph.D. degrees in Communication Engineering from the University of Birjand, Iran, in 2012 and 2016, respectively. He is currently an Associate Professor in the Department of Electrical and Computer Engineering at the University of Birjand. His research interests include image processing, deep neural networks and deep learning.

- Email: s.mohamadzadeh@birjand.ac.ir
- ORCID: 0000-0002-9096-8626
- Web of Science Researcher ID: NA
- Scopus Author ID: 57056477500
- Homepage: https://cv.birjand.ac.ir/mohamadzadeh/

How to cite this paper:
M. Rohani, H. Farsi, S. Mohamadzadeh, "Advanced race classification using transfer learning and attention: real-time metrics, error analysis, and visualization in a lightweight deep learning," J. Electr. Comput. Eng. Innovations, 13(2): 341-352, 2025.

DOI: 10.22061/jecei.2025.11318.793

URL: https://jecei.sru.ac.ir/article_2258.html