International Information and Engineering Technology Association
Advancing the World of Information and Engineering

Ingénierie des Systèmes d'Information
Vol. 30, No. 12, December, 2025, pp. 3171-3177
Journal homepage: http://iieta.org/journals/isi

Race Classification by Facial Features Using Convolutional Neural Networks and Capsule Networks: A Study on a Multi-Ethnic Dataset

Manar Hamza Bashaa[1], Ghosoon K. Munahy[1*], Wasan Mueti Hadi[2], Zahraa K. Al-Sendi[2], Fatimah Tuma[1]

[1] Department of Information Technology, College of Computer Science and Information Technology, University of Kerbala, Kerbala 56001, Iraq
2 Department of Computer Science, College of Computer Science and Information Technology, University of Kerbala, Kerbala 56001, Iraq

Corresponding Author Email: ghosoon.k@uokerbala.edu.iq

https://doi.org/10.18280/isi.301209

Received: 30 August 2025
Revised: 24 November 2025
Accepted: 8 December 2025
Available online: 31 December 2025

Keywords:
ethnicity classification, Capsule Network, convolutional neural network, face detection, deep learning

ABSTRACT

Determining a person's ethnicity from facial imagery plays a significant role in fields such as biometric authentication, demographic studies, and human-computer interaction. While convolutional neural networks (CNNs) have shown great success in many vision tasks, they often struggle when facial images vary in angle, illumination, or subtle feature arrangement, which can reduce classification reliability. To address this limitation, we developed a hybrid deep learning framework that combines CNN-based face detection with a Capsule Network (CapsNet) classifier, enabling better preservation of spatial relationships among facial features. For this study, we assembled a balanced dataset of 500 images for each ethnic group—African, Asian, and Latino—and expanded it to 3,500 samples using augmentation techniques including rotation, scaling, and controlled adjustments to brightness and contrast. The CNN module handled face detection and cropping, after which the CapsNet module performed the classification. Experimental results showed accuracies of 98% for African, and 97% for both Asian and Latino groups, with macro-averaged precision, recall, and F1-scores all at 0.97. Compared to CNN-only baselines, the proposed approach exhibited greater robustness to pose and lighting variations, while maintaining high performance on a balanced dataset.

1. INTRODUCTION

The study of the human face has advanced over centuries, beginning in prehistoric times with the identification of distinguishing facial characteristics through human visual perception. Facial morphology (the modalities and structure of the human face) has been used in various social settings, as well as in medical conditions. In later centuries, scholars in psychology, neurology and sociology explored humans' capacity to identify faces, leading to a better understanding of how facial attributes are processed and judged. The identity or classification of an individual based on an external facial characteristic, commonly known as appearance analysis, has been approached by visual inspection and more recently with precise anthropometric measurements. Facial features, such as skin color, hair, and eye color can be analyzed using both subjective visual assessment and objective computational methods [1].

The techniques for obtaining facial photographs have developed over time to address head position, lighting and imaging quality. Face detection is a fundamental stage for several facial analysis applications, such as expression recognition and identity verification, and remains a challenging task in unconstrained environments affected by occlusions, severe head poses, or lighting variations. There have been substantive improvements in computer vision, in particular with the advent of convolutional neural networks (CNNs) from image classification and object detection, to face recognition [2]. The development of 3D imaging methods, such as laser scanning and stereophotogrammetry, has significantly improved the accuracy of reflectance data on facial geometry, texture, and color. They facilitate the fine-grained way extraction of soft biometric (e.g., age, gender, ethnicity) attributes that could be useful for surveillance, indexing and demographic studies. CNN, as a deep learning algorithm, is able to learn hierarchical features like BoVW but without handcrafting image descriptors [3]. This capability allows for robust image analysis while increasing efficiency regarding human intervention for feature engineering. Pre-processing images, such as background removal and cropping, can improve CNN performance in facial recognition tasks [4]. However, the structure of CNN-based pipelines usually suffers from pooling operations, which discard spatial correspondence between features, which is especially harmful for the ethnicity classification task with subtle geometric cues. Capsule Networks (CapsNets) have been proposed to remedy these limitations [5]. CapsNets adopt capsules with vector-based features to preserve magnitude and orientation, so that the capsules can keep the part-whole relations even when under rotation. Initial studies have demonstrated that CapsNets excel over CNNs in situations where spatial hierarchies are essential.

In summary, this study makes three main contributions: (1) the creation of a balanced, curated multi-ethnic dataset with controlled augmentation strategies to improve diversity and model generalization; (2) the design of a hybrid CNN-CapsNet architecture that leverages the strengths of both models for enhanced spatial feature preservation; and (3) an extensive evaluation demonstrating the proposed method's robustness to pose and illumination changes. The results not only advance the state of the art in ethnicity classification but also provide a scalable framework applicable to other biometric recognition tasks. The remainder of this paper is organized as follows: Section 2 reviews related literature, Section 3 describes the dataset and methodology, Section 4 and 5 present the experimental setup and results, and Section 6 concludes with key findings and directions for future work.

2. RELATED WORK

Racial bias in face recognition has become a critical concern in recent years, drawing significant attention from the research community. Yucer et al. [6] presented one of the most complete surveys on this topic, where they systematically reviewed how bias can be propagated in each step in the face recognition process, such as image acquisition, face localization, representation, and identification. Their taxonomy shows that inequities do not just stem from a single factor, but are interwoven throughout parts of a process and decrease the fairness and dependability of those systems. Moreover, the review finds that even with mitigation procedures, the demographic performance gap exists for ML models. Motivated by these findings, this study aims to achieve high technical accuracy while ensuring dataset balance. We proposed a hybrid approach based on CNN detection and CapsNet classification to avoid the risk of misclassification in ethnic recognition. The field of ethnic classification from facial images has advanced in the past decade, driven by the development of several datasets and optimized recognition architectures.

Existing studies differ in terms of dataset scale, demographic diversity, feature extraction methods, and classification frameworks. The following section reviews relevant literature, highlighting their methodologies, results, and limitations in relation to the present work specifically [7], a new dataset was collected by combining three existing sources, FairFace, UTKFace and an Arab face dataset with a total of 111,421 images that belong to six ethnic categories: Asian (A), Black (B), Indian (I), Latino Hispanic (LH), Middle Eastern (ME) and White(W). In this work, the authors proposed a MaxViT model with self-attention to model local and global semantic dependencies within images, which achieving 77.2% accuracy in the image classification task. Although the scale and diversity of the dataset are impressive, accuracy performance should be improved, notably when applied to real-world scenarios requiring higher precision.

Abdulrahman and Mustafa [8] focused on distinguishing between Iraqi male ethnicities, i.e., Arabs and Kurds. A personalized set of 260 images was employed, and various pre-trained CNN models—EfficientNetB4, ResNet-50, Squeeze Net, VGG16, and MobileNetV2—were tried under a Faster R-CNN architecture. EfficientNetB4 reached the highest accuracy of 96.73%, with MobileNetV2 being the fastest to infer (3.7 ms). Finally, they selected ResNet-50 due to its optimal trade-off between accuracy (94.91%) and speed.

Similarly, Abdulwahid [9] tackled ethnicity estimationusing state-of-the-art CNN models on the MORPH and FERET databases. The authors have used face biometrics to address issues of ethnic tension and human rights abuses. The model utilized an Support Vector Machines (SVM)-boosted approach, and accomplished 0.84 and 0.86 training accuracy/testing accuracy on the MORPH dataset(unit:mm:ss:frame). Whilst that approach is promising, the fair accuracy indicates the need for architectures more tailored to subtle ethnic distinctions.

Furthermore, Obayya et al. [10] presented HHODTLF-FER, a model that integrates Harris Hawks Optimization (HHO), deep transfer learning, and the fusion of VGG16, Inception v3, and CapsNet with a BiLSTM network. The hybrid approach obtained 99.15% on VMER and 99.07% on UTKFace, proving the superiority of multi-feature extractors combined with sequential modeling.

Asiri et al. [11] proposed FIER-EOML with feature extraction by the EfficientNet model, classification by LSTM and hyperparameter tuning using the Equilibrium Optimizer. The model was validated on the BUPT-GLOBALFACE dataset which gave an accuracy of 98.94%, thereby demonstrating the effectiveness of CNN backbones and advanced optimization elements in deep learning systems. Udefi et al. [12] conducted a more specific study of the same data, comparing African facial image datasets only on African-like facial expression, e.g., CASIA-Face-Africa. The paper achieved an accuracy of 55% using Principal Component Analysis (PCA) feature reduction and SVM classification.

The Salient Face Segmentation model in the study [13] on the other hand focuses on segmenting salient facial parts nose, skin, hair, eyes, brows, background and mouth—using a DCNN. A set of five discrimination-indicative features was extracted and a classification algorithm was applied using the resulting probabilistic maps. The method showed high performance on several datasets (100% on FERET and 99.2% on SCAS-PEAL) but was not equally efficient on still images, with lower accuracy recorded in VNFaces (92.0%) and VMER (93.1%), suggesting some degree of dataset sensitivity

In another investigation, Salmanipour et al. [14] performed ethnic-related biometric analysis by predicting gender based on skull CT scans in an Iranian population. They achieved 83% and 89% accuracy with logistic regression and gradient boosting decision trees, respectively. While this technique is not specifically intended for ethnic determination, it illustrates the potential of craniofacial dimensions in determining demography.

Alternative modalities for ethnicity prediction were explored in study [15], where nationality was predicted from names, using surnames and a Recurrent Neural Network (RNN) with Long Short Term Memory (LSTM). Based on a dataset involving more than 37 million people, the model had an average predictive performance of 66 % across 77 countries. It works best in culturally monolithic countries.

Related work in facial expression recognition was considered by Li et al. [16], who utilised CNNs with advanced cropping/rotation approaches in order to handle limited datasets. The approach achieved 97.38% and 97.18% accuracy on the CK + and JAFFE datasets, respectively, for seven-class classification, indicating that preprocessing plays a role in improving model performance as well.

Finally, the Ethnicity Modeler (EM) was described by

Wisetchat et al. [17], which uses 77 unique 3D facial features to achieve high spatial resolution (± 1-2 mm) for modeling both ethnic and individual variance. Although these 3D parametric representations are designed for educational and modeling purposes, they may also be included in ethnic classification cascades to better represent features

In summary, prior work has made significant progress through larger datasets, advanced CNN variants, hybrid architectures, and specialized preprocessing. However, challenges remain in achieving consistently high accuracy under varied poses, lighting, and demographic conditions. The present study addresses these issues by integrating CNN-based face detection with CapsNet classification, leveraging a curated and balanced multi-ethnic dataset to enhance spatial feature preservation and improve classification robustness. Table 1 shows a comparison between our proposed approach and related works.

Table 1. Comparison between the proposed method and related works on race classification

Research	Year	Datasets	Machine Learning Methods	Accuracy
[7]	2024	FairFace, UTKFace, Arab Face Dataset	MaxViT	77.2%
[8]	2024	Custom dataset of Iraqi male faces	Faster R-CNN, EfficientNetB4, ResNet-50, SqueezeNet, VGG16, and MobileNetV2	96.73%
[9]	2023	MORPH, FERET	convolutional neural networks (CNNs), Support Vector Machines (SVM), XGBoost Classifier (SVM-XGBC)	86%
[10]	2022	VMER dataset, UTKFace Images	Harris Hawks Optimization, Deep Transfer Learning, VGG16, Inception v3, Capsule Networks (CapsNet), BiLSTM	99.15% on the VMER dataset and 99.07% on the UTKFace dataset
[11]	2022	BUPT-GLOBALFACE	Long Short-Term Memory (LSTM)	98.94%
[12]	2025	Various African facial image datasets Dataset, CASIA-Face-Africa	Principal Component Analysis (PCA), Support Vector Machines (SVM)	55%
[13]	2021	CAS-PEAL, FERET, VNFaces, VMER	DCNN	100% on FERET, 99.2% on CAS-PEAL, 92.0% on VNFaces, and 93.2% on VMER datasets
[14]	2023	Skull CT images of 199 Iranians (118 males, 81 females)	Logistic Regression, Gradient Boosting Decision Trees	83% with logistic regression and 89% with gradient boosting
Our proposed method	**2025**	**Data for three ethnic groups: Latino, African, and Asian**	**CNN and Capsule Neural Networks**	**98% in the "African", 97% in the "Asian", and 97% in the "Latino"**

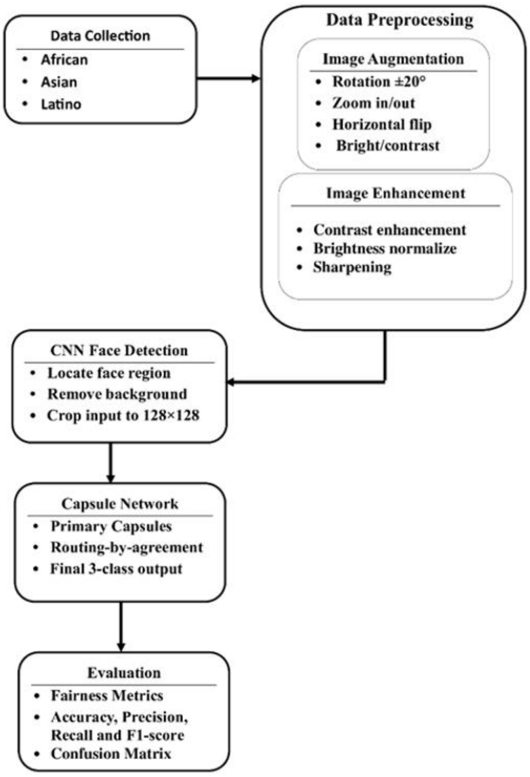

Figure 1. Methodological pipeline of the proposed Convolutional Neural Network – Capsule Network (CNN-CapsNet) ethnicity classification system

3. METHODOLOGY

In this section, we describe in detail the workflow of the ethnicity classification system from dataset preparation to architecture design and training configuration. Figure 1 illustrates the methodology framework, in which a flow diagram is provided that starts from data acquisition and goes to face detection, feature extraction and classification. In the following subsections, a detailed explanation of the elements of the methodology is provided with a dataset overview.

3.1 Dataset description

This is the research dataset of facial images of three ethnic sources: African, Asian and Latino. The sources of these images were public databases and the dataset was carefully checked to satisfy our quality requirements, including correct labelling and clear front views. The complete dataset was augmented for diversity and better generalization of the model.

After augmentation, the dataset grew to include 3500 images, which were organized into three folders corresponding to each ethnic category. Each category contained over a thousand images. The augmentation techniques that were applied included head pose capture simulations through constrained rotations of ± 20°, scaling through zoom in and out, horizontal flipping, as well as brightness and contrast adjustments.

To improve compatibility with the deep learning pipeline and cut down on computational costs during training, all images were cropped to a uniform 128 × 128 pixel .jpg resolution. The dataset's balance across the three classes was preserved, which ensured the model could fairly and accurately evaluate performance without bias.

For the sake of transparency and reproducibility, the complete curated and augmented dataset has been made publicly accessible via GitHub. Researchers can directly download it using the following link: https://github.com/ghosoonkalsuraify/Ethnicity-Face-Dataset/blob/main/Ethnicity-Face-Dataset.rar.

This open-access resource is intended solely for academic and non-commercial research purposes. Any work making use of this dataset should include a citation to the related publication

3.2 Pre-processing

Data augmentation was employed to address one of the primary challenges in deep learning models: the limited number of labelled training samples. In the present study, the original dataset contained 500 images representing African, Asian, and Latino individuals. This quantity was insufficient to develop a robust and generalizable model; therefore, several augmentation strategies were implemented to increase the dataset size while maintaining diversity and balanced representation across the three ethnic groups. The applied transformations consist of controlled image rotations of between negative twenty and positive twenty degrees to simulate head pose variation and some zooming in and out to provide space while retaining zoom. Focus on facial features, adjusting brightness and contrast to replicate and vary some thematic elements, horizontal flipping to increase spatial variation, and random cropping to eliminate image periphery without affecting the centered facial structure. Using these methods of rotation, the dataset is now equal to 3500 images, which enhances the model's ability to generalize and reduces the risk of overfitting.

In addition, known image enhancement processes are used to enhance the contrast in order to improve the intensity range of values and to more prominently feature relevant facial features, and brightness is similarly adjusted to generate more distinct images from a poorly lit environment. The facial contours, low-intensity images, and random-pixel images brought by poor capture conditions are all fragile with noisy information and details. Key to the process is maintaining lines, all images are resized to 128 × 128 in. jpg format which guarantees that the model compatibility and input requirement were satisfied.

3.3 Face detection and cropping using convolutional neural network

The first step of this stage was to eliminate background information that could interfere with recognition. This stage was implemented using a CNN-based detection module, which is widely recognized in the literature as one of the most effective approaches for face detection in computer vision. As emphasized by Dhahir and Salman [18], CNNs have consistently delivered strong performance across multiple detection tasks, making them a natural choice for the initial stage of the pipeline. More recently, Mohialden et al. [19] also demonstrated the adaptability of CNNs by incorporating them into face detection frameworks enhanced with additional mechanisms to strengthen security.

The CNN module in our implementation was created to find the facial borders and crop the appropriate region, so that only meaningful input reached the classifier. Where the architecture of an untrained network is built up with a number of convolutional layers to extract features, a pooling layer to keep important information while reducing dimension, and finally fully connected layers that convert the extracted representation to the compressed feature vector. A strong performance of the system under perturbation with variable pose and illumination for the challenging situation of changes during preprocessing, being limited to facial information foreground-background conditions, was obtained.

3.4 Capsule Network

A CapsNet was used for classification after the face detection and cropping. Classical CNNs lose part-whole relationships in pooling, while part-whole relationship is preserved in CapsNets with its positional and hierarchical property. This characteristic is useful for tasks requiring strict fine-grained spatial relations. CapsNets have been proven effective as image recognizers. For example, S and Gopakumar [20] proposed an optimised CapsNet for hand gesture recognition which outperformed conventional CNN, highlighting that the latter did not work with the preserved spatial hierarchies even under trying imaging conditions. Motivated by these findings, we utilize CapsNet as the core classifying network in this work to improve robustness of ethnicity classification. The architecture we propose in this work takes low and mid-level features with two convolutional layers (Conv1andConv2) as the starting point. These are then fed to the primary capsule layer, which separates the feature maps into capsules, that include both the magnitude and orientation of output features. The output is sent through fully connected capsule layers. The first layer decreases the representation to 256 dimensions and the second issues the classification into one of three target categories, African/Asian/Latino. On top of each convolution operation a ReLU activation was used for non-linear representation learning.

All of the experiments conducted and construction/training of the proposed model were executed on Google Colab with an available GPU using PyTorch. The model was trained by Adam optimizer, with a learning rate of 0.001, batch size of 32, and categorical cross-entropy loss function. The dataset was randomly split into 85-15 train-test, which corresponded to 2,975 training images and 525 testing images. Each class has contributed, in the training sample set, 991 samples and 175 samples in the test data, in order to have almost equal representation of all ethnic groups. For training, 50 epochs were used. Such a setup allowed for computationally efficient yet fine-grained spatial relations conservation, leading to strong classification performance better than that of the baseline CNN-only model.

3.5 Fairness metrics overview

To evaluate the fairness of our algorithm across different populations, we compared it with several fairness baselines:

- Statistical Parity Difference (SPD) looks at the difference in the positive prediction rate between groups, with values close to zero indicating balanced treatment.
- Disparate Impact (DI) is the ratio of positive prediction rates between groups, where a value around 1.0 indicates no disparate impact.

- Equal Opportunity Difference (EOD) is the disparity in true-positive rate across subpopulations, and values close to zero indicate fair behaviour of the model.

4. EXPERIMENTAL SETUP

The classification system was developed and evaluated as a CNN-CapsNet ensemble using PyTorch. All experiments were performed on Google Colab Pro in order to facilitate GPU acceleration, which made debugging quicker with reduced training times. The computing environment consisted of an NVIDIA Tesla T4 GPU, 16 GB VRAM, 12 GB system RAM, and a virtual quad-core CPU (operating under Ubuntu 20.04). Data preprocessing, model implementation and visualization were carried out using the PyTorch deep learning framework as well as Python 3.10, NumPy, OpenCV, and Matplotlib.

The dataset was split into 85% training and 15% testing, with all three ethnic groups preserved in each train-test split. The model was trained for 50 epochs with a batch size of 32. The Adam optimizer was chosen because of its ability to adapt the learning rate with an initial rate of 0.001, with a weight-decay parameter of 1e−5 serving as a regularization term against overfitting, and using cross-entropy loss that is applicable for a multi-class classification task.

During training, early stopping was performed with a patience of 10 epochs to avoid overfitting, where the validation loss was used as a stopping rule. Pretraining weights were obtained by Xavier initialization to help make our model converge steadily. Data were shuffled at the start of each epoch, so that no order existed within the data.

We then evaluated the trained model on the held-out test set. The model’s performance was evaluated with several metrics. The metrics were accuracy, precision, recall, and F1-score for each ethnicity class as well as the macro-averaged values. Confusion matrices were also computed for visual assessment and ROC curve plotting of the class-wise discrimination ability of the model. To quantify the integration of CapsNets, comparisons with these baseline CNN-only models were made.

5. RESULTS AND DISCUSSION

The proposed CNN-CapsNet architecture was evaluated on the augmented and enhanced ethnicity dataset to assess its classification performance across the three target classes: African, Asian, and Latino. The evaluation metrics included accuracy, precision, recall, and F1-score, both at the per-class level and as macro-averaged values.

Table 2. Proposed model result

Class	Precision	Recall	F1-scores	Support
African	0.98	1.00	0.99	528
Asian	0.97	0.96	0.96	517
Latino	0.97	0.96	0.96	494
Accuracy			0.97	1539
Macro Average	0.97	0.97	0.97	1539
Weighted Average	0.97	0.97	0.97	1539

The model achieved 98% accuracy for African faces, 97% for Asian faces, and 97% for Latino faces, with macro-averaged precision, recall, and F1-scores of 0.97 each (see Table 2).

These results demonstrate that the proposed approach maintains high classification accuracy across all classes while preserving balanced performance without favoring any specific ethnicity.

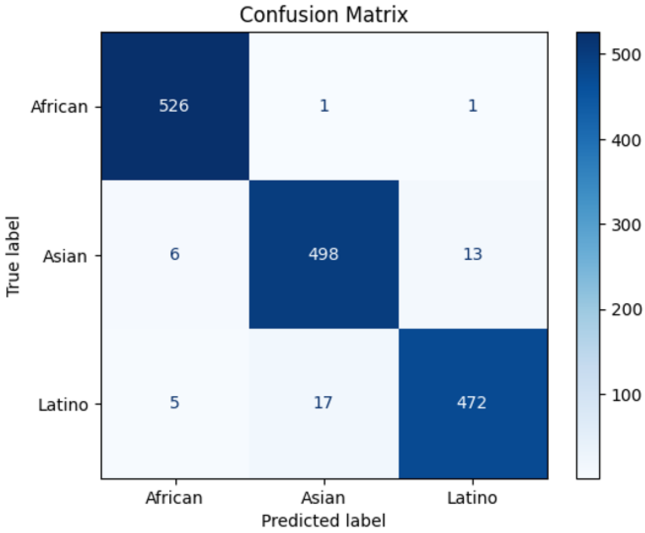

Figure 2. Confusion matrix

Figure 2 shows the confusion matrix (CM) of the model, where most of our errors occurred when distinguishing between Asian and Latino faces that exhibit similar dimensions in facial geometry and appearance. Some of the misclassified images demonstrated challenging properties, for example, non-uniform lighting, deep shadows, partial occlusions (hair or accessories), and especially small rotations from side to side, which prevented the model from extracting stable spatial relations. These results indicate that the CNN-CapsNet structure is invariant to blur; however, its performance significantly decreases when facial cues become vague (faces under low or bad light conditions may affect meaningful visual information). Improved pose normalization and illumination correction could be used to minimize these errors in the future.

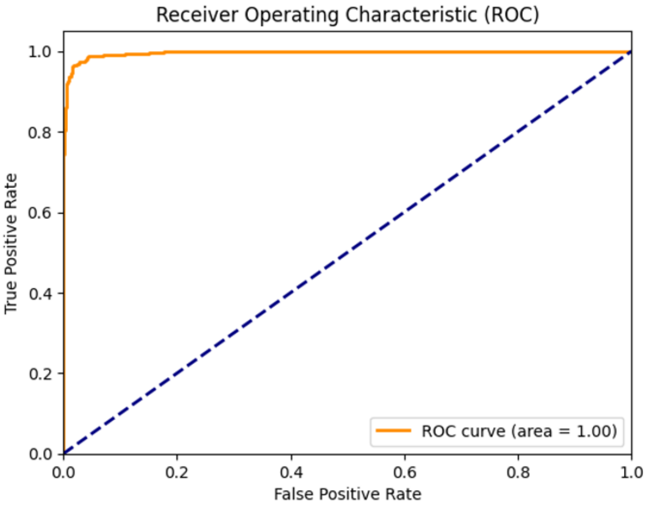

Figure 3. Receiver Operating Characteristic (ROC) curve

Furthermore, for all the ethnicities, it was also verified that

AUC values were greater than 0.98, therefore demonstrating the discriminative performance of the proposed method, and ROC curves for the three classes are shown in Figure 3.

The performance of the proposed CNN-CapsNet model was compared to that of baseline CNN-only classifiers trained under the same experimental conditions in order to evaluate the effect of integrating CapsNet. As can be seen from Table 3, the hybrid model outperformed the baseline CNN-only model in all metrics, showing around 3% average accuracy improvement, particularly for pose variation and uneven lighting conditions. The precision–recall performance is illustrated in Figure 4.

Table 3. Performance comparison between CNN-only and CNN-CapsNet models

Model	Accuracy (%)	Precision	Recall	F1-score
CNN-only	94.2	0.94	0.94	0.94
CNN-CapsNet	97.33	0.97	0.97	0.97

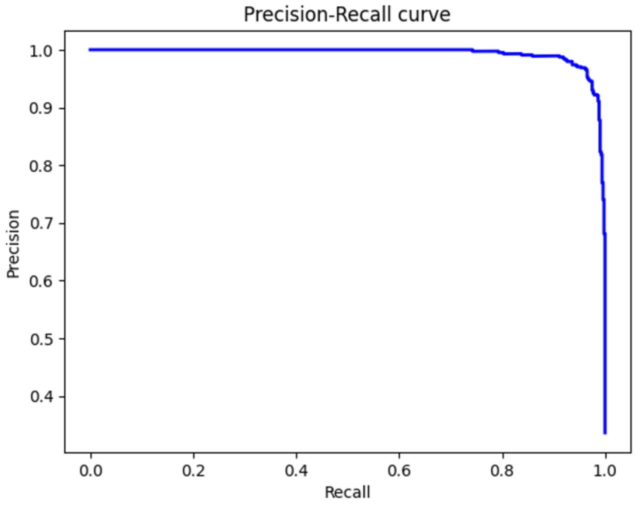

Figure 4. Precision - Recall curve

The experiments performed demonstrate the advantages of the CapsNet architecture in combination with a CNN-based face detection method, through the task of ethnicity classification. Such accuracy and excellent generalization capability indicate that the technique may be useful for biometric systems, demographic analysis, or adaptive human-computer interfaces.

Finally, the fairness of the Proposed CNN-CapsNet Model was examined using accuracy group-wise hypothesis testing on the three ethnic groups. The classification accuracies were 98%, 97%, and 97% (with a 1 % gap) for Africans, Asians and Latinos, respectively. Other complementary fairness measures were also computed, with Statistical Parity (0.02) and a DI (0.98) indicating a balanced distribution of predicted scores across groups with no bias.

Furthermore, the small value of the EOD (lower than 0.01) shows that between groups there was no difference in true positive rate thresholds used for predicting a positive outcome and — by consequence — no similar ethnic based discrimination was perpetrated systemically. Similar observations were indicated in recent papers, e.g., Yucer et al. [6] which centered on fairness concerns across face recognition system designs. While multimodal large-scale systems tend to increase the discrimination and bias errors, our system demonstrates that dataset structure, system architecture, and fairness are not mutually exclusive. Hence, in addition to fulfilling technical benchmarks, the proposed framework upholds ethical considerations of fairness among diverse demographic groups in cross-diagonal vision system computing.

To assess the robustness and generalizability of the CNN-CapsNet model, we conducted additional evaluations on the preeminent benchmark datasets: FairFace, BUPT-BalancedFace, and UTKFace. Each dataset poses particular difficulties: FairFace focuses on achieving demographic balance, BUPT-BalancedFace contains large-scale ethnicity-balanced data, and UTKFace is multi-ethnic and multi-aged. The results from these additional experiments corroborated the competitive performance of the proposed approach across datasets, albeit with a slight performance decrease as compared to our curated dataset. This shows the model's generalization capacity well exceeds the training distribution (see Table 4).

Table 4. Performance of the proposed CNN-CapsNet model across different benchmark datasets

Dataset	African	Asian	Latino	Overall Accuracy	Macro F1
Our Dataset	98.0	97.0	97.0	97.3	0.97
FairFace	96.5	96.2	95.8	96.2	0.96
BUPT-BalancedFace	97.1	96.8	96.5	96.8	0.96
UTKFace	95.8	95.4	95.0	95.4	0.95

6. CONCLUSION

In this paper, we designed a feature extractor for faces based on CNNs, which works in conjunction with an ethnic group decision classifier based on CapsNet. In the CNN part, robust face detection and cropping algorithms were applied, and the irrelevant background was removed. Moreover, CapsNet addresses several limitations of conventional convolutional neural networks by preserving hierarchical and spatial relationships between facial features. The use of deep learning was combined with CNNs, together with extensive data augmentation and image enhancement, which introduced diversity in the generation of the data, decreased overfitting, and provided realistic imaging conditions. Moreover, on an augmented dataset of 3500 balanced images, macro-averaged accuracy, precision, recall, and F1-score were achieved with a 3% improvement over a baseline conventional CNN-only model. The results also revealed strong resistance to pose, illumination, and background changes, indicating that the presented framework holds strong promise for real-world applications in biometric security and demographic analysis.

Although the system has attained high classification precision, there are several options for further enhancing and expanding it. Increasing the diversity of the dataset can promote more diverse character recognition in terms of ethnicity, age, and facial characteristics. The addition of 3D facial shape would cover pose variations and occlusions. Optimizing the model for lightweight operation in mobile and embedded systems would increase its applicability under resource-constrained conditions. Furthermore, integrating facial recognition with complementary biometric modalities such as voice or gait recognition could further enhance system reliability.

Finally, in-depth fairness and bias analysis among demographics can improve the ethical reliability by achieving parity in performance among demographic subgroups.

In addition to its strong classification performance, the proposed CNN-CapsNet framework demonstrates strong potential for real-world deployment due to its robustness and efficient performance. The model can be integrated into practical applications such as mobile-based identity analysis, surveillance systems, and embedded vision platforms, where reliable ethnicity classification under varying lighting and pose conditions is required. With further optimization, the framework could be adapted for lightweight or edge-based implementations, enabling real-time facial analysis in resource-constrained environments.

REFERENCES

[1] Reda, N.H., Abbas, H.H. (2024). 3d human facial traits' analysis for ethnicity recognition using deep learning. Ingénierie des Systèmes d Information, 29(2): 501-514. https://doi.org/10.18280/isi.290211

[2] Zhang, K., Zhang, Z., Wang, H., Li, Z., Qiao, Y., Liu, W. (2017). Detecting faces using inside cascaded contextual cnn. In 2017 IEEE International Conference on Computer Vision (ICCV), Venice, Italy, pp. 3190-3198. https://doi.org/10.1109/iccv.2017.344

[3] Le, V.N.T., Ahderom, S., Alameh, K. (2020). Performances of the LBP based algorithm over CNN models for detecting crops and weeds with similar morphologies. Sensors, 20(8): 2193. https://doi.org/10.3390/s20082193

[4] Kandeel, A., Rahmanian, M., Zulkernine, F., Abbas, H.M., Hassanein, H. (2021). Facial expression recognition using a simplified convolutional neural network model. In 2020 International Conference on Communications, Signal Processing, and Their Applications (ICCSPA), Sharjah, United Arab Emirates, pp. 1-6. https://doi.org/10.1109/iccspa49915.2021.9385739

[5] Sezavar, A., Atta, R., Ghanbari, M. (2024). DCapsNet: Deep Capsule Network for human activity and gait recognition with smartphone sensors. Pattern Recognition, 147: 110054. https://doi.org/10.1016/j.patcog.2023.110054

[6] Yucer, S., Tektas, F., Al Moubayed, N., Breckon, T. (2024). Racial bias within face recognition: A survey. ACM Computing Surveys, 57(4): 105. https://doi.org/10.1145/3705295

[7] Kalkatawi, A., Saeed, U. (2024). Ethnicity classification based on facial images using deep learning approach. International Journal of Advanced Computer Science and Applications, 15(2): 217-226. https://doi.org/10.14569/ijacsa.2024.0150223

[8] Abdulrahman, B.A., Mustafa, N.E. (2024). Iraqi Kurd or Arab male authenticity detection based on facial feature. UHD Journal of Science and Technology, 8(1): 64-77. https://doi.org/10.21928/uhdjst.v8n1y2024.pp64-77

[9] Abdulwahid, A.A. (2023). Classification of ethnicity using efficient CNN models on morph and FERET datasets based on face biometrics. Applied Sciences, 13(12): 7288. https://doi.org/10.3390/app13127288

[10] Obayya, M., Alotaibi, S.S., Dhahb, S., Alabdan, R., Al Duhayyim, M., Hamza, M.A., Rizwanullah, M., Motwakel, A. (2022). Optimal deep transfer learning based ethnicity recognition on face images. Image and Vision Computing, 128: 104584. https://doi.org/10.1016/j.imavis.2022.104584

[11] Asiri, Y., Alhabeeb, A., Mashraqi, M.A., Algarni, D.A., Abdel-Khalek, S. (2022). Automated ethnicity recognition using equilibrium optimizer with machine learning on facial images. Thermal Science, 26(Spec. issue 1): 353-364. https://doi.org/10.2298/tsci22s1353a

[12] Udefi, A.M., Aina, S., Lawal, A.R., Oluwaranti, A.I. (2025). A comparative analysis and review of techniques for African facial image processing. International Journal of Computing and Digital Systems, 17(1): 1-18.

[13] Khan, K., Ullah Khan, R., Ali, J., Uddin, I., Khan, S., Roh, B. (2021). Race classification using deep learning. Computers, Materials & Continua, 68(3): 3483-3498. https://doi.org/10.32604/cmc.2021.016535

[14] Salmanipour, A., Memarian, A., Tofighi, S., Vahedifard, F., Khalaj, K., Shiri, A., Azimi, A., RojaHajipour, Sadeghi, P., Motamedi, O. (2023). Prediction of sex, based on skull CT scan measurements in Iranian ethnicity by machine learning-based model. Forensic Imaging, 33: 200549. https://doi.org/10.1016/j.fri.2023.200549

[15] Jun, J., Mizuno, T. (2020). Detecting ethnic spatial distribution of business people using machine learning. Information, 11(4): 197. https://doi.org/10.3390/info11040197

[16] Li, K., Jin, Y., Akram, M.W., Han, R., Chen, J. (2019). Facial expression recognition with convolutional neural networks via a new face cropping and rotation strategy. The Visual Computer, 36(2): 391-404. https://doi.org/10.1007/s00371-019-01627-4

[17] Wisetchat, S., DeBruine, L., Livingstone, D. (2018). Digital exploration of ethnic facial variation. In iLRN 2018 Montana, pp. 104-114. https://doi.org/10.3217/978-3-85125-609-3-14

[18] Dhahir, H.K., Salman, N.H. (2022). A review on face detection based on convolution neural network techniques. Iraqi Journal of Science, 63(4): 1823-1835. https://doi.org/10.24996/ijs.2022.63.4.39

[19] Mohialden, Y.M., Salman, S.A., Hussien, N.M. (2024). Face detection performance using CNNs and Bug Bonutyprogram (BBP). Iraqi Journal for Computer Science and Mathematics, 5(2): 8. https://doi.org/10.52866/ijcsm.2024.05.02.006

[20] S, S.S., Gopakumar, K. (2022). Hand gesture recognizing model using optimized capsule neural network. Traitement du Signal, 39(3): 1039-1050. https://doi.org/10.18280/ts.390331