Contents lists available at Egyptian Knowledge Bank

Labyrinth: Fayoum Journal of Science and Interdisciplinary Studies

Journal homepage: https://lfjsis.journals.ekb.eg/

A new model for human ethnicity prediction from facial images

Mohammed M. Abd El-fatah[a,*] , Howida Youssry Abd El Naby [a] , Shereen A. Taie [a]

[a] *Computer Science Department, Faculty of Computers and Artificial Intelligence, Fayoum University, El Fayoum 63514, Egypt.*

ARTICLE INFO

Keywords:
Ethnicity Classification
Gender Classification
CNN

ABSTRACT

The humans' faces hold a lake of information that enables us to identify them. Classification of a human's ethnicity is important information in various areas, such as biometrics, security, and personal safety. Physical characteristics, such as skin color, hair type, and facial features are used by the human brain to divide people into different ethnic groups. This paper presents a model that classifies ethnicity and gender according to 5 ethnicity classes. Moreover, the proposed model classifies males and females for each ethnicity group. The proposed model uses deep learning to mimic the behavior of the human brain in distinguishing between different ethnic groups. The proposed model comprises two main phases, the first phase is the data preparation and preprocessing which is a crucial step to make the facial images meet the requirements of the second phase and emphasize that the proposed model is more robust and generalized, this phase includes three main steps; Data Augmentation by using images flipping and noise injection techniques, Face Detection, and Images Resizing, the second phase is features extraction and classification with convolutional neural network (CNN). The proposed model gives promising results with accuracy 99.78% and 100% for ethnicity and gender, respectively.

1. Introduction

In this era, facial analysis has become one of the most important and inevitable, because it contains so wealth data associated with humans. These data include the shape, and color of the eyes, the geometric shape of the facial bones, skin color, hair, and dimensions between different facial regions [1]. By obtaining this data, the identity of the person can be identified with the face like ethnicity, gender, and age. This technology is useful in various fields such as biometrics, personal safety, and security [2]. Humans can readily categorize ethnicity by the appearance of the face. Conversely, in the computer vision field identifying humans' ethnicity through the image of the face remains a research topic. Processing of face images is a challenge; this is due to the image formation process variability in terms of noise, illumination, falsification, geometry, and occlusion [3]. To overcome this challenge, this paper proposed a model that classifies 5-categories of ethnicities these listed as follows Asian American, Black, Indian, Latino-a, and White. Inside each ethnicity group, the proposed model distinguishes the gender as male or female.

The contribution of this paper is introducing a proposed model for ethnicity classification and gender classification. The proposed model classifies 10 classes with males and females; it achieves accuracies for ethnicity and gender 99.78% and 100% respectively, and overcomes the problem of overfitting in classification.

The remaining of this paper is organized as follows: Section 2 presents state-of-the-art studies related to ethnicity classification problem. The proposed model was illustrated in detail in section 3. Section 4 presents and discusses the achieved experimental results. Finally, concludes the paper in section 5.

* Corresponding author.
E-mail address: mm119@fayoum.edu.eg (M.M. Abd El-fatah); Tel.: +201009151205

DOI: 10.21608/ifjsis.2023.217835.1030
Received 18 may 2023; Received in revised form 17 july 2023; Accepted 17 August 2023
Available online 21 August 2023

2. Related work

Masood et al. [4] presented 2 experiments for distinguishing Mongolian, Caucasian, and Negroid ethnicities. Therefore, FERET database was used. Artificial Neural Network and Convolution Neural Network were used. In ANN, Viola Jones is used for detecting face region. Just face was detected, parts of the face such as mouth, right and left eye were marked. The distance and ratio between the marked parts were calculated. RGB images converted into YCbCr color to outperform illumination variation. Horizontal Sobel edge (Gx), vertical Sobel edge (Gy), and the intersection of Gx and Gy were applied to calculate Forehead area. For training, multi-layered perceptrons were implemented. The training task used 320 images, when 37 images were used for validation and the testing task was applied using 90 images. In CNN, the VGGNet pre-trained model which has 13 convolution layers followed by three fully connected layers. The VGGNet model was used to extract features and classify the ethnicity. For 3 classes of ethnicities, ANN and in CNN achieved accuracy was 82.4% and 98.6%, respectively.

Darabant et al. [5] collected frontal face images from 6 different databases. Afro-American, Asian, Caucasian, and Indian are the chosen classes. The face area was detected from the input images. Valuable information was determined from the area around the face such as the hair structure and texture. Then, according to the requirement of CNN, the images are resized. Four distinct convolutional neural network architectures, namely VGG19, Inception ResNet v2, Se-ResNet, and Mobilenet V3, were selected for training and comparison. Then, they trained the CNNs in around 10 days. In testing, images were collected from other 2 databases. 96.36% was the best obtained accuracy.

Anwar & Islam [6] distinguished between Asian, African-American, and Caucasian were 3 ethnicities. In their work, face images were collected from 10 different databases. Prior to training and testing, all face images are aligned using 68 dense fiducial points, which include the eyes, eyebrows, nose, and mouth. Fiducial point detection and alignment are performed for this purpose. After images preprocessing, CNN pre-trained VGG-Face model was used for features extraction. VGG-Face model consisted of 37 layers which are broken into 6 blocks. filter size of 3*3 was used with stride 1 and padding 1 in each convolutional layer. There were 2 convolutional layers in the first 2 blocks, followed by ReLU layers. After the second ReLU layer, a max pool layer was used to reduce the spatial size of the feature map to half. The next three blocks consist of three convolutional layers each, followed by ReLU layers. A max pool layer is added after the third ReLU layer. The last block is composed of three fully connected layers, followed by a soft-max layer. Then, SVM was performed the classification task. SVM achieved accuracy 98.28%, 99.66%, and 99.05% for Asian, African-American, and Caucasian, respectively.

Batsukh & Ts [1] proposed an approach for detecting different nationalities (Mongolian, Japanese, Chinese, and Korean) from the frontal face image. To detect the face, a cascade is used. The facial features are then extracted from the entire face region. edge detection was used to improve and enhance the contrast of the real input. human facial features were divided into base features and extra features. The base features consist of the shapes and location of the eyes, nose, mouth, and lips. On the other hand, the extra features include the forehead, mandible, eyebrows, eyelid, chin, and ear. The measurements of all these features based on their color, size, shape, and distance from their neighboring features were determined. After enhancement, PCA and eigenvector were used to select the gender. The geometric human face shapes were used for classifying. Using active appearance model (AAM) and ASM to recognize geometric shapes of face. The landmark and edge detection calculate the distance from the neighbors. The classification was performed via SVM. The best accuracy was 86.4%.

Chen et al. [7] used CNN to predict Chinese, Japanese, and Korean with their genders from face images. Images manually were collected from profile photos of university members from the three countries. data preprocessing was applied as follows. Haar Cascade Face Detection algorithm was used to crop the face. The noise was removed from the image background. The image size was normalized to 64 X 64 for CNN, but to 128 X 128 for the remaining methods. CNN is comprised of two convolutional layers that contain 64 and 128 filters respectively. Additionally, it includes a fully-connected layer with 1024 neurons, a dropout layer, and an output layer. L2 regularization was applied to each of these layers. Then, KNN, 2-Layer NN, SVM, and CNN were experimented for prediction, and the performance for each class with gender was 50.9%, 74.4%, 48.2%, 83.5%, respectively.

Momin & Raymond [3] conducted 3 experiments. Gabor features was used to extract features. Gabor filter was used with 5 distinct frequencies and 8 orientations. To create feature vectors, the mean value for each component across the 5 frequencies and 8 orientations was computed. Four classifiers K-means, Naive Bayesian, Multilayer Perceptron, and Support Vector Machines performed the classification task. The 1st experiment differentiated between Asian, Non-Asian with an accuracy 99.60%. The second one classified Asian, White, and Black with 90.21% accuracy. The last experiment achieved 87.67% in discrimination between Asian, Indian, White, and Black ethnicities

3. proposed model

This section illustrates the proposed model which consists of two phases. The first phase is data preparation and preprocessing which aims for converting the data into an appropriate form and satisfied the model pre-requests. The second is feature extraction and classification phase which receives the processed data as inputs, therefore the model collects the associated features to feed the top layer and learn to classify ethnicity and gender. The framework of the proposed model is shown in Fig.1 and it will be illustrated in detail in the following subsections.

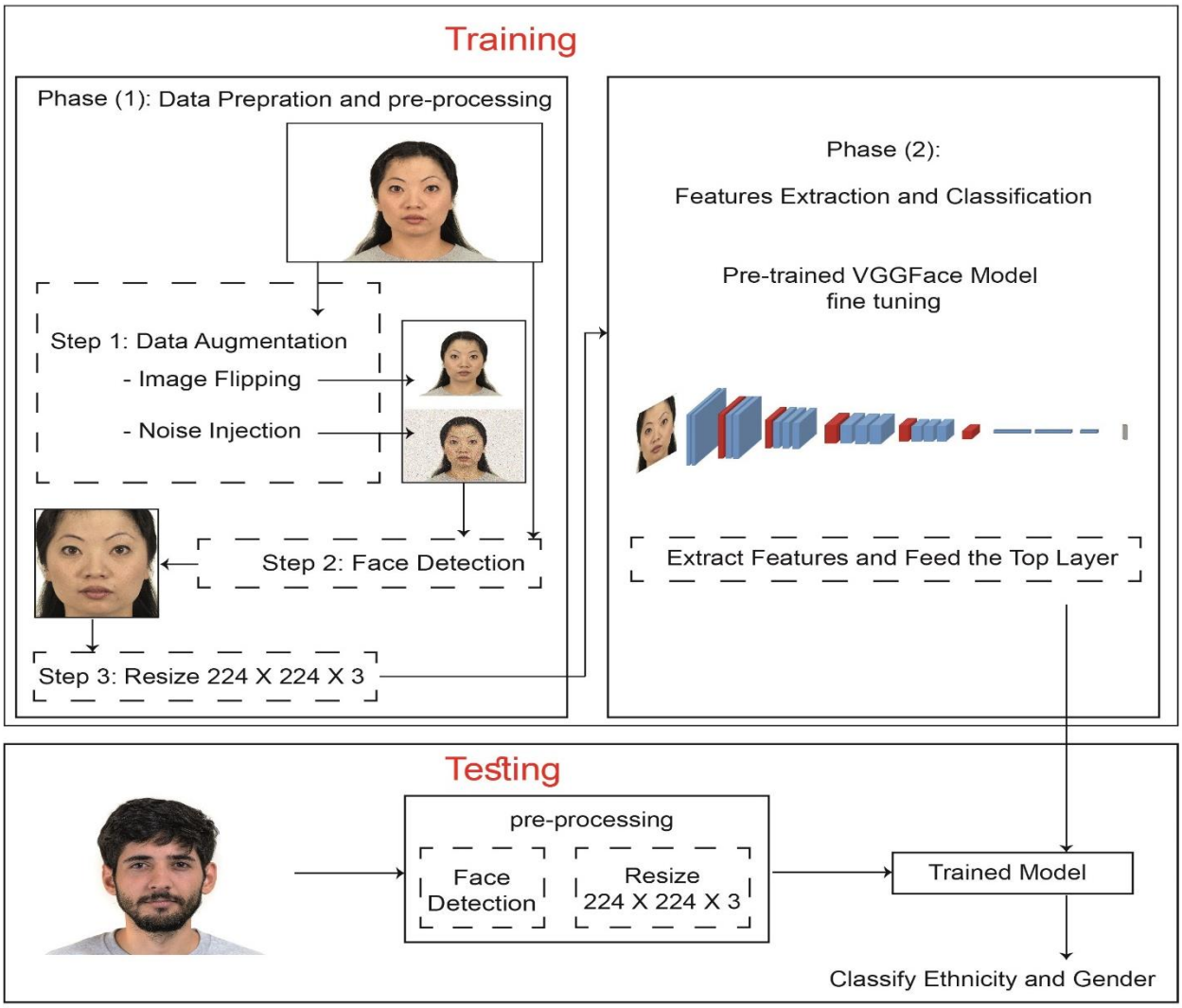

Fig. 1. Proposed Model Framework

3.1. Data Preparation and Preprocessing Phase

This phase is significant for data alignment. Data preparation and preprocessing were used to generate new samples for expansion of the little samples classes and obtain data that is appropriate to the pre-requisite of the CNN model. This phase is compromises of three main steps; Data Augmentation, Face Detection, and Resizing as follow:

Step 1: Data Augmentation

The purpose of this step is to increase the variety of dataset samples by creating modified versions of the existing data to avoid overfitting problem in classification. Data augmentation is a random transformation applied on Chicago Face Database Multiracial (CFD-MR) expansion and Chicago Face Database India (CFD-INDIA) which were presented in Ma et al. [8] and Lakshmi et al. [9], respectively to create new samples that are added to the range for each class.

The proposed model applied flipping, and noise injection transformations as follow:

-Image Flipping: Involves mirroring an image around its vertical, horizontal, or both axis (as shown in Fig.2 (A), flipping horizontally was applied in this model).

-Noise injection: Helps the model to learn to be more robust and generalize better to outperform the different inputs environments and different cameras resolution consequently classification is more accurate. In this work as shown in Fig.2 (B), salt and pepper noise was added; the added noise will appear as sprinkling white and black dots -as salt and pepper— on the original image.

The next step is extracting the Region of Interest (ROI), detect human frontal facial from the input images. It will be discussed in detail in the following step.

Step 2: Face detection

The Face detection step selects the human face from the images for getting the required features. These features conclude the ethnicity, and gender classifications. The proposed model uses Viola-Jones algorithm Viola & Jones [10] to detect the frontal face, as it performs this task with impressive speed and high performance.

Viola-Jones cascaded classifier consisting of 38 layers was trained for detecting frontal upright faces. A dataset consists of images containing faces and other images without, that used to train the detector. A collection of 4916 hand labeled faces adjusted to a resolution of 24 by 24 pixels which was used as face training set. Training set included 9544 non-face images which were manually checked and found to not contain any faces, that used to train the detector. 1, 10, 25, 25 and 50 were the number of features in the first five layers of the detector, respectively. The number of features in the rest layers increased. For all layers, the total number of features was 6061. Fig.2 (C) shows the frontal face that detected via Viola-Jones algorithm.

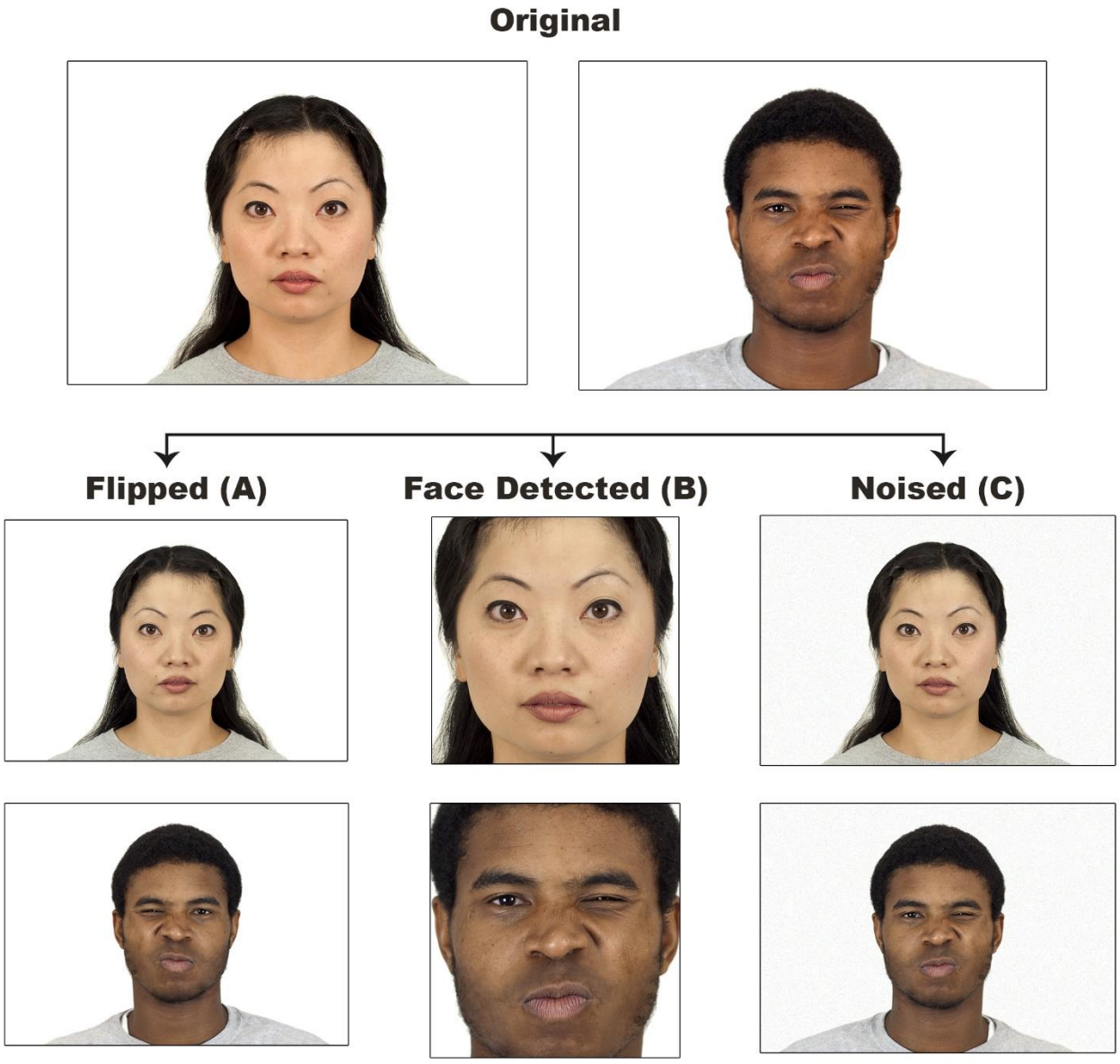

Fig.2. Data preparation and preprocessing

Step 3: Resizing

The final step of the preprocessing phase is the resizing for the face detected images. The images must be in size 224W X 224H X 3Channels to compatible with the pre-trained model – input images must be colored as the skin color is an important feature in ethnicity classification problem. The pre-trained VGGFace2 Cao et al. [11] model is used and feed the top layers with CFD dataset to classify gender and ethnicity.

3.2. Feature Extraction and Classification Phase

This phase determines the ethnicity, and gender via deep learning. The proposed model uses deep learning (DL) which mimics the human brain in learning process and knowledge gain based on Convolutional neural network (CNN). Convolutional Neural Network (CNN) has been used for features extraction, and classification problems, as it has worked well in classification tasks. The proposed model uses the pre-trained model VGGFace2 due to implementing a CNN model from scratch for robust functionality requires a huge dataset and therefore requires power resources and the computation needs a lot of time and cost.

This phase performs the classification using the pre-trained VGGFace2, which aims for face recognition and involves classifying images of faces into different identities. Powerful features can be extracted via VGGFace2 due to it was trained on 3.31 M images for 9,131 different humans' that have different identities Sayed et al. [12]. The VGGFace2 performs feature extraction and classification tasks. Finally, the last fully connected layer is replaced, it classifies ten classes using the softmax activation function. Hence, the fully connected layers are re-trained on ethnicity and gender classification feeding CFD preprocessed data.

4. Experimental Results and Discussion

This section illustrates the details of the experiments and discusses the obtained results. The proposed model has been implemented using python and Keras API on the Windows platform. The experiments were applied on 8 GB RAM, Intel Core i3-2120 CPU 3.30 GHz.

4.1. Dataset Description

In this paper, CFD was used which consists of human frontal faces with high-resolution, standardized images with a range of different ethnicity, ages, and facial expressions. The URL of CFD website [13].

The experiments conducted on 3,000 images for 5 different balanced classes of ethnicity (Asian American, Black, Indian, Latino-a, and White) with gender (male, and female) for each class were used to learn and test the proposed model. Each class contained 300 images, that split to 70% for training and 30% for validating and testing. Table 1 shows the dataset ingredients.

Table 1: Dataset ingredients

Ethnicity	Male	Female	Total in the same ethnicity
Asian American	300	300	600
Black	300	300	600
Indian	300	300	600
Latino-a	300	300	600
White	300	300	600
Total in the same gender	1500	1500	3000

4.2. Training configuration:

The training parameters of the model will be discussed in detail. The configuration setup was as follows:

* Batch size: 15,
* Learning rate was set to 10-5,
* epochs: 20.

As mentioned earlier in classification phase section, VGGFace2 pre-trained model was used and fine tuning technique was used to feed the fully connected layers with the preprocessed images of CFD. Fine tuning technique guide the model towards extract the features that distinguishes each class of ethnicity and gender, then re-trains the model to fit the ethnicity and gender classification problem.

The proposed model achieved performance with respect to training time it consumed 11.73 hours for training. Table 2 illustrates the comparing number of classes, datasets, training time, and results to the proposed model respect to other models. The comparison between the state-of-art models and the proposed model shows that, the proposed model distinguishes between the largest number of classes compared to other models.

The model in Darabant et al. [5] classified only ethnicity. It is only that recorded the time consumed in the learning process. this time was 10 days for the model learning. Compared to the time consumed to learn the proposed model where it was 11.73 hours, this shows the superiority of the proposed model in this factor.

According to the state-of-art models that differentiated between both ethnicity and gender, the obtained accuracy was less than 90%. As for the achieved accuracy from the proposed model is 99.78% and 100% for ethnicity and gender, respectively.

Table 2: Models comparison

Model	No. of classes	Dataset	Learning time	Classification accuracy	
				Ethnicity	Gender
[1]	4	Collected from every country	N/A	86.4%	
[7]	6	collected a set of facial images with labeled gender	N/A	for 3 ethnicities only: 89.2%	For 6-classes: 83.5%
[6]	3	CVL, CFD, FERET, MR2, UT Dallas face database, PICS, JAFFE, CAS-PEAL-R1, MSFDE, CUFC	N/A	Asian: 98.28% African-American: 99.66% Caucasian: 99.05%	N/A
[4]	3	FERET	N/A	98.6%	N/A
[5]	4	CFD, Minear-Park, JAFFE, MR2, IMFDB, KFDB	10 Days	96.36%	N/A
[3]	2 3 4	MORPH II database, Indian Face database, Asian Face database	N/A	99.6% 90.21% 87.67%	N/A
Proposed Model	10	CFD	11.73 Hours	99.78%	100%

The comparisons between state-of-the-art and the proposed model were showed in Fig. 3. Compared to the number of classes, the proposed model has superiority.

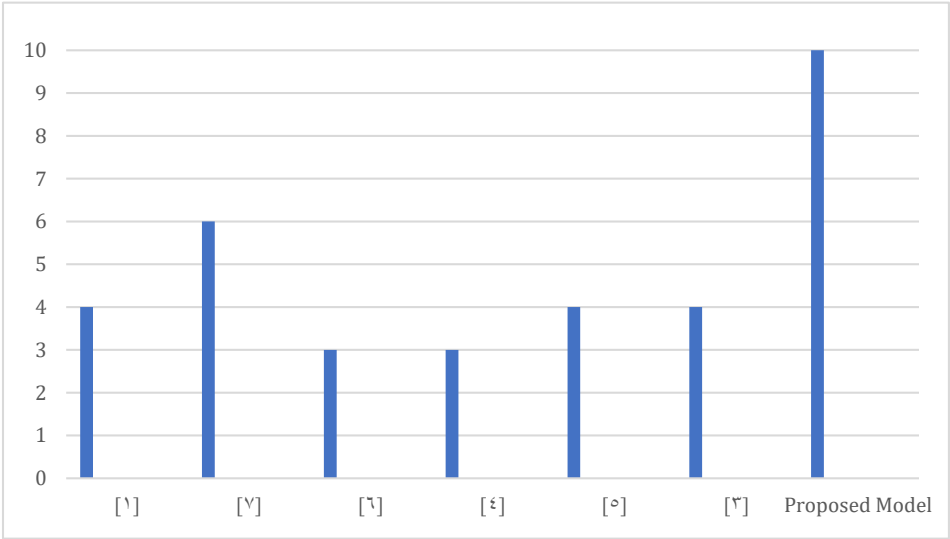

Fig. 3. Comparison in No. of classes between state-of-the-art and proposed model

Fig. 4. explores the classification accuracies for between state-of-the-art and the proposed model. Like the proposed model, Batsukh & Ts [1] and Chen et al. [7] classified both ethnicity and gender, while Momin & Raymond [3], Masood et al. [4], Darabant et al. [5], and Anwar & Islam [6] classified the ethnicity only. The proposed model is the superior compared to state-of-the-art respect to the ethnicity and gender classification accuracy.

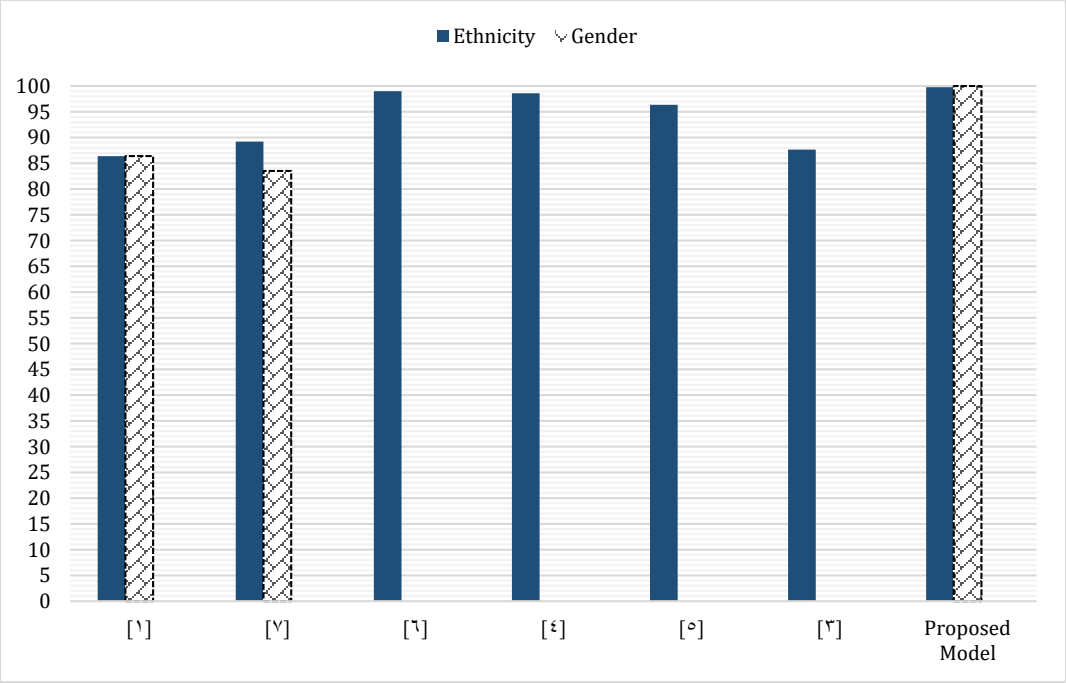

Fig.4. Comparison in accuracy between state-of-the-art and proposed model

4.3 Ethnicity Classification Evaluation

By obtaining predicted values and actual values respectively in a table form called a confusion matrix, the performance measurement of the model can be evaluated. The dimensions of the confusion matrix are N X N table which shows the number of true and false predictions, where N is the number of classes [14]. Fig. 5 presents the confusion matrix of the ethnicity and gender classification for the proposed model. There are various performance metrics which extremely useful for evaluating, namely, Precision, Recall, Accuracy, and F-Score, which were calculated according to equations (1) – (4).

$$\text{Precision} = \frac{TP}{TP+FP} \quad (1)$$

$$\text{Recall} = \frac{TP}{TP+FN} \quad (2)$$

$$\text{Accuracy} = \frac{TP+TN}{TP+TN+FP+FN} \quad (3)$$

$$\text{F-Score} = \frac{2 . \text{Precision} . \text{Recall}}{\text{Precision} + \text{Recall}} \quad (4)$$

where TP stands for the true positive predicted values, FP stands for the false positive predicted values, TN stands for the true negative predicted values, FN stands for the false negative predicted values.

The proposed model applied 5-fold cross-validation to find out if the proposed model is stabilized and generalized. The model has achieved similarity results about 99.78% accuracy for the ethnicity classification and 100% accuracy for gender classification. Table 3 outlined The performance of the proposed model.

Table 3: Proposed model performance

	Precision	Recall	F-Score
Asian American Female	1.00	1.00	1.00
Asian American Male	1.00	1.00	1.00
Black Female	1.00	1.00	1.00
Black Male	1.00	0.97	0.99
Indian Asian Female	1.00	1.00	1.00
Indian Asian Male	1.00	1.00	1.00
Latino/a Female	1.00	1.00	1.00
Latino/a Male	0.98	1.00	0.99
White Female	1.00	1.00	1.00
White Male	1.00	1.00	1.00
ACCURACY		99.78	

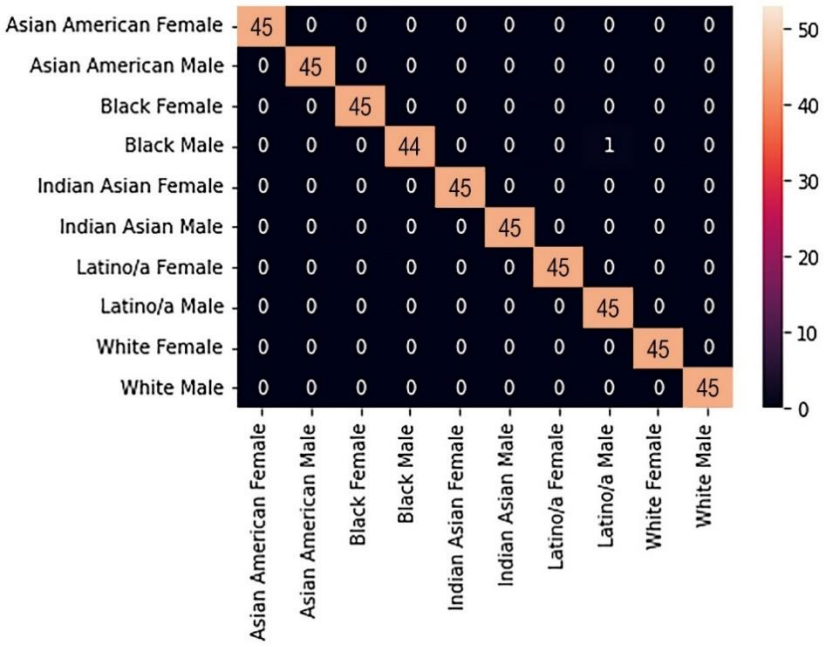

Fig. 5. Confusion matrix

5. Conclusion

This paper proposed a model for ethnicity and gender classification using CNN. The proposed model used CFD which contains frontal face images of humans of different ethnicities, and within each ethnicity, there are images of males and females. Flipping, and noise injection DA techniques were applied on CFD to outperform the overfitting problem. A human face was detected from dataset images. Then, resizing the image to 224 X 224 X 3 was performed to fit the pre-requirement of the VGGFace2 which is a deep learning pre-trained model. VGGFace2 model was based on CNN. Using a pre-trained model and the concept of fine-tuning in deep learning, the cost of implementing a new model from scratch is avoided. VGGFace2 pre-trained model extracts the discriminative features deservedly. Moreover, the model was retrained using these features. The proposed model improves the classification accuracy for 10 classes, 5 ethnicities and discrimination between males and females in each ethnicity group. The accuracy was achieved as follows ethnicity and gender classification 99.78% and 100%, respectively.

Acknowledgment

The authors would like to thank Fayoum University for supporting the publication of this work.

Author Contributions

All authors contributed to this work. Mohammed M. Abd El-fatah got and prepared the dataset and completed the experimental results and evaluations. Both Shereen A. Taie and Howida Youssry Abd El Naby followed the paper writing, analyzing the data, validation, and performance of the results. Mohammed M. Abd El-fatah completed the paper writing, Shereen A. Taie followed the revision and submission of the manuscript for publication.

Declaration of Competing Interest

The authors declare that they have no known competing financial interests or personal relationships that could have appeared to influence the work reported in this paper.

References

[1] B. Batsukh, G. Ts, Effective Computer Model For Recognizing Nationality From Frontal Image, International Journal of Advanced Studies in Computers, Science and Engineering 5, 3 (2016): 1.

[2] G. Muhammad, M. Hussain, F. Alenezy, G. Bebis, A. M Mirza, H. Aboalsamh, Race classification from face images using local descriptors, International Journal on Artificial Intelligence Tools 21, 05 (2012): 1250019.
[3] H. Momin, J. Raymond Tapamo, A comparative study of a face components based model of ethnic classification using gabor filters, Applied Mathematics & Information Sciences 10, 6 (2016): 2255-2265. http://dx.doi.org/10.18576/amis/100628
[4] S. Masood, S. Gupta, A. Wajid, S Gupta, M Ahmed, Prediction of human ethnicity from facial images using neural networks, In Data Engineering and Intelligent Computing: Proceedings of IC3T 2016, pp. 217-226. Springer Singapore, 2018.
[5] A. Sergiu Darabant, D. Borza, R. Danescu, Recognizing human races through machine learning—A multi-network, multi-features study, Mathematics 9, 2 (2021): 195. https://doi.org/10.3390/math9020195

[6] I. Anwar, N. Ul Islam, Learned features are better for ethnicity classification, Cybern. Inf. Technol 17, 3 (2017): 152-164.
[7] H. Chen, Y. Deng, S. Zhang, Where am i from?–east Asian ethnicity classification from facial recognition, Project study in Stanford University (2016).
[8] D. S. Ma, J. Kantner, B. Wittenbrink. "Chicago face database: Multiracial expansion." Behavior Research Methods 53 (2021): 1289-1300. https://doi.org/10.3758/s13428-020-01482-5
[9] A. Lakshmi, B. Wittenbrink, J. Correll, D. S. Ma, The India Face Set: International and cultural boundaries impact face impressions and perceptions of category membership, Frontiers in psychology 12 (2021): 627678.
[10] P. Viola, M. Jones, Rapid object detection using a boosted cascade of simple features, In Proceedings of the 2001 IEEE computer society conference on computer vision and pattern recognition. CVPR 2001, vol. 1, pp. I-I. Ieee, 2001.
[11] Q. Cao, L. Shen, W. Xie, O. M. Parkhi, A. Zisserman, Vggface2: A dataset for recognising faces across pose and age, In 2018 13th IEEE international conference on automatic face & gesture recognition (FG 2018), pp. 67-74. IEEE, 2018.
[12] H. M. Sayed, H. E. ElDeeb, S. A. Taie, Bimodal variational autoencoder for audiovisual speech recognition, Machine Learning 112, 4 (2023): 1201-1226. https://doi.org/10.1007/s10994-021-06112-5
[13] ChicagoFaces. https://www.chicagofaces.org/, (accessed 8 January 2023).
[14] Z. Karimi, Confusion Matrix, Encycl. Mach. Learn. Data Min., October (2021): 260-260.