Generating Embedding Features Using Deep Learning for Ethnics Recognition

Mohammed Alghaili
School of Electronic Engineering,
Xidian University, China
malghaili@stu.xidian.edu.cn

Zhiyong Li
College of Information Science and Engineering,
Hunan University, China
zhiyong.li@hnu.edu.cn

Ahmed AlBdairi
College of Information Science and Engineering,
Hunan University, China
ahmed_albdairi@hnu.edu.cn

Malasy Katiyalath
College of Economics and Trade,
Hunan University, China
malasykatiyalath@hnu.edu.cn

Abstract: *Although significant advances have been made recently in the field of ethnics recognition through face recognition, there is still a lack of studies of ethnics recognition through facial recognition. This study is concerned with ethnics recognition through facial representation using a few images used as samples for any selected group of ethnics using a deep neural network with a Variational Feature Learning (VFL) loss function that has been used to increase the performance accuracy during the evaluation process. The output of a deep neural network is an embedding of 128 bytes for each face image in each group of ethnics. After that, all embeddings of every face in each group of ethnics pass to a machine learning classification method like a Support Vector Machine (SVM). We achieved state-of-the-art ethnic recognition. The system achieved a classification accuracy of 97.3% on a collected group of image dataset collected from three different countries.*

Keywords: *Computer vision, machine learning, ethnics identification, deep learning.*

Received October 24, 2021; accept June 19, 2022
https://doi.org/10.34028/iajit/20/4/13

1. Introduction

To interact socially, we should be capable of analyzing human faces in various ways. The human face has a range of signs which form part of a person's identity. In the past, there is a massive literature available on social and cognitive psychology that recognized human abilities to identify a well-known face, as well as take out the information from the known and unknown faces, including the extracting information about the age, ra ce, and emotional gesture of the person [52]. Face recognition is an essential ability and plays a prominent role in forming a relationship in society [39]. Even though through this ability, humans are considered enormously adept, there are some kinds of other faces, which may belong to other groups, may not be very good in recognition as compared to those faces belonging to known groups.

The own groups' face recognition is fast and accurate of one's belongs to out-groups [6]. According to anthropometric studies [67], significant facial morphology variations exist among gender, age, and ethnicity groups. For instance, Sexual Dimorphism [7] research reveals that male face features are richer than female face features. Based on that knowledge, the male face has protrusive facial anthropometric measurements [67]. It further extends about the female face feature dimensions are smaller than the male one. Research [12] elaborates on ethnic differences and explains about white North American and Asian people face anthropometric measurements. So, besides age, gender and emotion, ethnic information is also an essential characteristic to represent the human face. Differences exist in the face characteristic among the various ethnic's groups. Different facial anthropometric measurements (facial features) were investigated by [10] from three ethnic minorities in China. In the late 60s, along with the development in computer technology and the research of intelligent computing methods, face recognition also entered the technological era to handle it with automatic recognition procedures, methods, and techniques. Face recognition was fully stretched out from criminal investigation to numerous daily life activities like sales, marketing, etc. In face recognition research, face feature extraction and face feature classification are very considerable, which directly affect face recognition accuracy [57]. Thus, human faces are necessary for the recognition and identification of every individual.

Recently, deep neural networks, especially Convolutional Neural Networks (CNNs), have become the most commonly used method for facial feature representation and have achieved good results in face recognition problems. Most existing works focused on face recognition have achieved a high level of success [30, 34, 44, 45, 46, 47, 49, 55, 56, 59, 64, 65, 66] and are based on facial representation. However, there is a lack of studies on ethnic's identification through facial recognition.

There are very limited studies on ethnic's identification. A lot of previous studies used only a simple classification for specific ethnics without using any facial representation. In this study, an embedding feature is used to get a good face representation for any set of ethnics as it has not been used before for ethnics identification.

Previous approaches on facial representation, in general, are based on two approaches of representation. The first approach is the discriminative classification model (face identification) trained on a dataset of known identities. An intermediate bottleneck layer is used as a representation for recognition. This approach generalizes a very large representation for each face, but some researchers have tried to reduce this dimensionality using PCA [46]. The second approach is used in FaceNet [41], where the output is directly trained to obtain 128-D embedding using a triplet loss function based on LMNN [56]. The triplets used in this study comprise two matching faces and a non-matching face. The triplet loss function aims to separate positive results from negative ones by a certain distance margin.

In the proposed approach, we use an unsupervised learning technique to obtain 128-D embeddings per face. The network is trained on several faces from different ethnicities. It then passes these 128-D embeddings to another machine learning classification algorithm such as SVM to find the most suitable representation for each group of ethnicities. The novalty of this work is using many ethnics types with small amount of number not even more than 100 images for each ethnics type and pass it to the SVM. Unlike other classifications researches [3, 11, 23, 28], we made the classification method in our study unlimited to any ethnics types depends on the samples of classification that our model will extract their features and pass them to SVM. Therefore, the number of classes will be detected after extracting the features using our model.

As an illustration, Figure 1 shows three pictures of multiple faces. Each picture belongs to a specific group. Three different groups were selected for the test; are Pakistani, Chinese, and Russian-the ethnics of the pictures in Figure 1 predicted with proposed system. Group A shows some people from Pakistan, group B shows some people from China, and group C shows others from Russia.

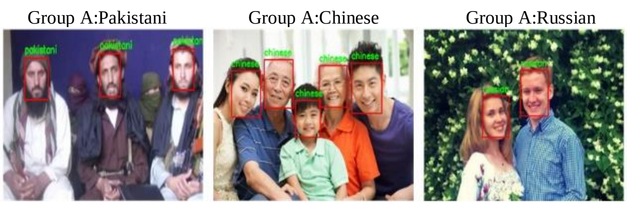


Figure 1. Three groups of different ethnics identified by the system.

The rest of the paper is organized as follows: The related work about face recognition is discussed in section 2. Section 3 describes our proposed method, including the feature extraction method through the deep neural network model. Sections 4 and 5 present some quantitative results and an evaluation, and section 6 shows the conclusion.

2. Related Work

Several computer vision researchers have conducted various studies to address the image-based facial representation such as gender [16, 18, 33], age [14, 16, 18, 36], and ethnicity [13, 33]. More specifically, some research work has tried to look into other races through traditional methods [15, 35]. In Furl *et al.* [15], the researcher proposed two algorithms for two different kinds of races, i.e., one for western and the other for Asian faces. The results conclude that the western fusion algorithm performs well and accurately recognizes Caucasian faces than East Asian faces. Similarly, the East Asian fusion algorithm performs accurately and classifies East Asian faces more than Caucasian faces. However, some other local feature descriptors have been used for face features extraction to consider multiple aspects of the human face, including face recognition. Commonly used descriptors for local features extractions are SIFT [61], LBP [22], HOG [60] and Gabor [25]. In spite, the face recognition research showed promising performance. However, there is still room for improvement regarding uncovering new methods and techniques that contribute well balanced and improved results in face technology. Therefore, the trend changes from traditional methods to Machine learning and deep learning approaches.

Many machine learning (ML) classification methods have been utilized for face recognition research [40, 63]. The most common include K-means classification [62], Support Vector Machine (SVM) [29], and BP neural network method [51]. Recently, the convolutional neural network has turned into the most prominent approach in the deep learning field. AlexNet [26] depends on the convolutional layered engineering, which comprises various numbers of neurons. Each layer of neurons consists of some inputs from the previous layer of neurons. Among the convolutional layers also consists of max-pooling layers and the activation functions (ReLUs). In this research, we also employed a deep learning approach to extract features for ethnic identification.

The proposed approach is similar to other recent works [30, 46, 50, 55]. It can learn its representation directly from the face. Each image in each group of ethnics gives a vector of 128 features when it passes to the network. Then, the generated vector obtained from the neural network is used as an input to a machine learning classification like SVM classifier to identify the ethnics of the face.

Although there are many studies in face recognition, ethnics' identification suffers from a lack of research. Furthermore, all previous studies in ethnics' identification used a simple classification. We will discuss in brief the most important relevant works.

One of the research studies about face recognition was proposed by Taigman *et al*. [50], called DeepFace. This approach is one of the earlier large-scale applications of 3D-model for face recognition. The face representation extracted using a nine-layer DeepFace model consists of two convolutional layers, three locally-connected layers, and two Fully-Connected (FC) layers with more than 120 million parameters that use numerous locally connected layers without sharing the weight. The network trained on 4.4M 2D facial images of 4,030 identities, and they achieve an accuracy of 97.35% on the benchmark Labeled Faces in the Wild (LFW) [20] dataset.

Another approach proposed a CNN used for face recognition and clustering called FaceNet [41] based on eleven convolutional and three FC layers. They trained a deep convolutional network on a dataset of 200M faces and 8M identities and triplet loss function to directly obtain the embedding instead of choosing the output of an intermediate bottleneck layer as in the previous works. The triplet sets used roughly aligned matching/non-matching face patches using an online triplet mining method, and they achieved a state-of-art face recognition with 128 bytes for each face.

Deep 3D face recognition results have been represented by Vijayan *et al*. [53]. The authors fine-tuned the VGGFace network [34] on 3D depth images. Then, they tested their approach on three public datasets. They also used an augmented dataset of 123,325 depth. They tested the model on Bosphorus [38], BU3DFE [45] and 3D-TEC (twins) [53] datasets. But the results do not achieve results as a state-of-art performance of the convolutional methods.

Gutta *et al*. [17] classified gender and ethnic through a hybrid classification architecture in face images. The authors show their results by evaluating their method using a set of 3006 face images corresponding to 1009 identities from FERET database. Their method contains an ensemble of REF networks and inductive decision trees. The results have 96% gender and 94% ethnic accuracy. In the research study, the total ethnic groups have been evaluated, namely Asian, Oriental, Caucasian, and African.

Shakhnarovich *et al*. [42] achieved an estimated rate of about 80% to distinguish the Asian/Non-Asian faces using some machine learning classifier where all Non-Asian faces in this study almost belong to Caucasians. The classifier output in this study is a linear combination of simple and weak single-feature discriminants.

Another approach in ethnicity estimation with facial images combines transformation of Gabor Wavelets with retina sampling to extract the most key facial features of the faces [19]. The authors in this study used Support Vector Machines (SVM) for ethnicity classification of three groups that have been used in this study these are Asian (Mongoloid), European (Caucasoid), and African (Negroid). The classification accuracy achieved in this study was 94% for ethnicity estimation under different illuminations.

Chen *et al*. [8] implemented the machine learning approach to classify and predict Chinese, Japanese, and Korean based on the facial images. They have used a series of computational machine learning-based algorithms, including the Convolutional Neural Network approach. Their results showed that CNN had the highest prediction accuracy of 89:2% in the 3-class classification.

Another attempt to distinguish three different ethnicities, Chinese, Korean or Japanese, was proposed by Yan [58]. Different variations of convolutional neural networks were applied to two new datasets constructed for this task. The author also used SVM and KNN classifiers for comparison.

A pre-trained VGG-face Convolutional Neural Network has been used to extract features from the face images for ethnic classification. There are three different major groups, Asian, Caucasian, and African-American, from ten other datasets that have been considered for this experiment. The extracted features were then classified through a linear kernel-based SVM classifier [5].

In [54], threw light on facial ethnicity classification. The authors consider two different combinations of ethnic groups and one sub-ethnic group for the research study. The d etails of groups evaluated through the proposed method include Black and white, Chinese and Non-Chinese people, while in sub-ethnic groups Han, Uyghurs are classified with Non-Chinese people. For features extraction, a deep convolutional neural network has been utilized and classifies each group individually. The effectiveness of the proposed method has been shown through the experimental results.

In another research study [9], two features have been taken from 3D faces models and combined. Furthermore, to get and uncover the local geometry of whole faces and their fine texture variations, Oriented Gradient Maps (OGMs) were formed. The extracted set of discriminative features was then utilized for ethnicity classification. FRGC v2.0 dataset has been evaluated to classify the Asians and non-Asians ethnic groups.

3. Method

We have used a deep neural network structure called NN4 as in FaceNet [41] and OpenFace [4] studies with some changes in the top of its structure as it is shown in Table 1 NN4 trained on CelebFaces Attributes Dataset (CelebA) [32] which contains several faces from different countries all over the world. Before we input the images to that network, we resized them to the size of 96×96×3. The first convolutional layer in the network has 64 kernels of the size 7×7×3 with stride two. The second convolutional layer has 64 kernels of the size 1×1×3 with stride two, and in the third convolutional layer, 192 kernels are used with the size 3×3×3 and stride 2.

Table 1. The structure of the neural network.

Type	Size in	Size out	Kernal	Feature Map
ZeroPadding	96, 96, 3	102, 102, 3		
Conv1	102, 102, 3	48, 48, 64	7×7×3, 2	64
Norm	48, 48, 64	48, 48, 64		
ZeroPadding	48, 48, 64	50, 50, 64		
Max pool	50, 50, 64	24, 24, 64	7×7×3, 2	
Conv2	24, 24, 64	24, 24, 64	1×1×3, 2	64
Norm	24, 24, 64	24, 24, 64		
ZeroPadding	24, 24, 64	26, 26, 64		
Conv3	26, 26, 64	24, 24, 192	3×3×3, 2	192
Norm	24, 24, 192	24, 24, 192		
ZeroPadding	24, 24, 192	26, 26, 192		
Max pool	26, 26, 192	12, 12, 192	3×3×3, 2	
inception 3a	12, 12, 192	12, 12, 256		
inception 3b	12, 12, 256	12, 12, 320		
inception 3c	12, 12, 320	6, 6, 640		
inception 4a	6, 6, 640	6, 6, 640		
inception 4e	6, 6, 640	3, 3, 1024		
inception 5a	3, 3, 1024	3, 3, 736		
inception 5b	3, 3, 736	3, 3, 736		
Average pool	3, 3, 736	1, 1, 736	3×3×3, 1	
fc1(μ)	1, 1, 736	128		
fc2(σ)	1, 1, 736	128		
fc3	128	1		

Seven blocks of inception architecture were used after all these layers labeled inception 3a, inception 3b, inception 3c, inception 4a, inception 4e, inception 5a and inception 5b [48]. Table 1 shows the network structure. At the top of the network, we added three fully connected layers. The output of the first and second fully connected layers are 128. The output of the second fully connected layer passed to the third fully connected layer. The output of the third fully connected layer is only 1 because we make the training as a classification but for only one class to act as unsupervised learning. This training aims to take the output of the first fully connected layer, which denotes the mean μ and has the output of 128-byte and not the last fully connected layer, which has the same number of classes in the training dataset, i.e., only one byte.

Before training, we loaded the weights of the FaceNet [41] network to the proposed network. Then, we trained the network using Stochastic Gradient Descent (SGD) with standard backprop [27, 37]. After training the network, the output of the first fully connected layer will be taken with the size of 128 bytes for each image passed to the network. This is called an embedding for each face in the image.

Another new dataset was collected to be used as a sample of some selected ethnicities. The images of this dataset will be passed to the model to extract their embeddings. These embeddings will be passed to SVM classifier for ethnics classification. It should be noticed that each group of ethnics can contain a limited number of faces as the network can extract the best feature representation for each face.

During training, we used a Kullback-Leibler (KL) loss function inspired by [2, 24]. The two fully connected layers used in the NN4 network were used to predict the mean μ and standard deviation σ of a Gaussian distribution. The last fully connected is used to predict the ethnicity while the two fully connected before that last one are used to produce a KL loss function that has been used in training. The LK loss function after that is combined with a categorical cross-entropy loss function. The following function explains the KL loss function, where n denotes the number of the vector size.

$$KL = -\frac{1}{2}\sum_{i=1}^{n}\left(1 + \log(\sigma_i) - \mu_i^2 - \sigma_i\right) \quad (1)$$

All samples in each group will pass the network model to generate an embedding for them. After that, we used SVM classifier to classify the embedding of the groups of ethnics and gives the results. Figure 2 shows the structure of the ethnics recognition system where it can classify n number of different ethnics depending on the embedding of the samples of the groups.

4. Evaluation Process

The deep neural network has been used to extract the features of all faces in each group in the dataset whereas each face can be represented by 128 bytes. After that, we used SVM classification to classify each group in the dataset and then generate the output to predict the ethnics of each face.

We evaluated the network on a large number of images collected from the street and the internet. We collected images for three groups they are Chinese, Pakistani, and Russian people. It is not important that the number of images in each group should be too large because the extracted features are too much enough to distinguish each group, i.e., a hundred images in each group used as samples in this study. Therefore, during training with SVM on the groups, we trained the features of very limited pictures for each group. The SVM classifier after that can be used to predict the ethnicity of any unseen image.

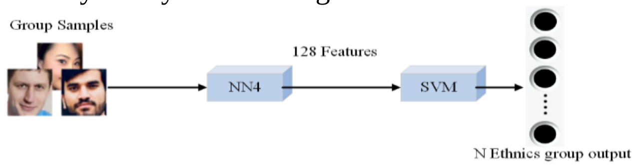


Figure 2. Feature of each group from NN4 passed to SVM.

5. Experiments

5.1. Training Dataset

CelebFaces Attributes Dataset (CelebA) [32] was used in this study as the training faces in this experiment because it contains many faces from many different ethnicities. It consists of 10,177 identities and 202,599 faces. Before the training, we used a face detector to extract the faces only from each image in CelebA dataset. After that, we resized each image to 96×96×3 as this size is the input size of the proposed networks.

5.2. Experimental Results

As there is no dataset available to be used to evaluate these kinds of studies, a new dataset was collected to use in the evaluation phase to identify the ethnics of people. This dataset is considered as the first dataset collected for ethnics' people [1]. The dataset was collected from three different ethnic groups they are Chinese, Pakistani, and Russian. Only a small part of each group was used as a sample for the ethnic's recognition system. Figure 3 shows a subset of the newly collected dataset.

The model was tested on a new dataset collected from the internet. The test phase shows good results on this dataset. We used a face detector to detect faces and pass each face to the model to extract its features and then pass the features to SVM classifier to predict its ethnicity. It should be noticed that the only drawback in all face studies is the lack of face detection in all previous face detection frameworks, especially for images with multiple faces and multiple poses. This drawback can negatively affect the results because some faces cannot be detected and therefore cannot be predicted in the proposed system. Figure 4 shows some samples of three tested groups in the collected dataset, i.e., Chinese, Pakistani, and Russian.

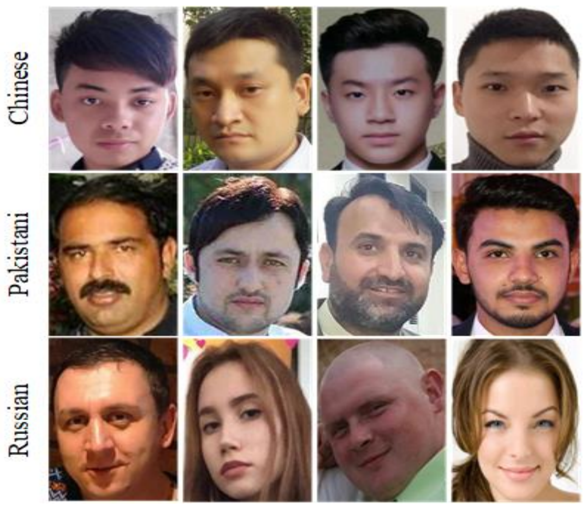


Figure 3. Three different subset images were collected from three different regions. A small part of each group was used as samples in NN4

The total number of images used for evaluation is 885. The accuracy performance achieved is 97.3% mPA. Table 2 shows the confusion matrix while Table 3 presents some parameters for measuring the performance and Table 4 presents a comparison of the achieved accuracy classification of the proposed system with previous studies.

Table 2. Confusion matrix.

	Chinese	Pakistani	Russian
Chinese	332	6	3
Pakistani	5	316	6
Russian	3	4	325

Table 3. Evaluation parameters for measuring the performance.

Recall	Precision	F1 Measure
0.97	0.97	0.97

Table 4. Comparison of accuracies with other previous studies.

Method	Accuracy	Type of Ethnics
Gender and Ethnic classification [17]	94%	Asian, Oriental, Caucasian, African
Features Extraction-Algebraic [10]	79%	Tibetan, Uighur, Zhuang
Features Extraction-Geometry [10]	90%	Tibetan, Uighur, Zhuang
PCA+SVM [21]	71.11%	British, Pakistan
PLS+SVM [21]	76.03%	British, Pakistan
Face Estimation-Linear [43]	Asian: 85.99% None-Asian: 92.21%	Asian, None-Asian
Face Estimation-polynomial [43]	Asian: 72.55% Non-Asian: 80.07%	Asian, None-Asian
Face Estimation-RBF [43]	Asian: 76.75% Non-Asian: 91.12%	Asian, None-Asian
Face Estimation-Sigmoid [31]	Asian: 72.55% Non-Asian: 90.27%	Asian, None-Asian
Influence of cross-ethnic social experience [32]	Asian: 73% Caucasian: 78% Gypsies: 86%	Asian, Caucasian and Gypsies
CNN approach [1]	96.9%	Pakistani, Russian and Chinese †
Embedding Method	97.3%	Pakistani, Russian and Chinese *

† We used this dataset evaluation.

* Type of ethnics can be any type. It's not limited to the three selected ethnics in this study because we are using some samples to be feed to the SVM after extracting their features from the model created from NN4.

6. Conclusions

We used a deep neural network trained on CelebA [32] to obtain 128-D embedding for facial representation to identify the ethnicity, then we used a machine learning classifier, e.g., SVM, to classify the extracted embeddings of any selected group of ethnics to identify the ethnicity. The number of samples in each group of ethnics can be small. It will not affect the classification accuracy due to the good network trained on a large number of images that contains many faces from different ethnicities and can generate a good feature representation for each face in each group. It should be noted that ethnicity does not mean nationality. There may be a person who has a nationality, but his origin belongs to another ethnicity. A new dataset was collected to test the model, and it will be openly available for research purposes. On the collected dataset, the system achieved a classification accuracy of 97.3%.

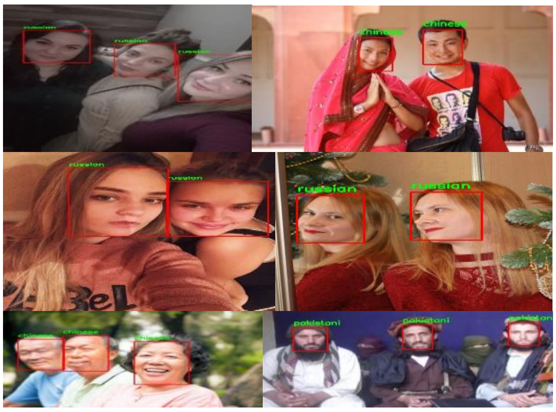


Figure 4. Output samples from the poposed system.

References

[1] AlBdairi A., Xiao Z., and Alghaili M., "Identifying Ethnics of People through Face Recognition: A deep CNN Approach," *Scientific Programming*, vol. 2020, 2020. https://doi.org/10.1155/2020/6385281

[2] Alfasly S., Hu Y., Liang T., Jin X, Zhao Q., and Liu B., "Variational Representation Learning for Vehicle Re-Identification," *in Proceedings of the IEEE International Conference on Image Processing*, Taipei, pp. 3118-3122, 2019. DOI:10.1109/ICIP.2019.8803366

[3] Alzahrani A. and Bhuiyan A., "On Satellite Imagery of Land Cover Classification for Agricultural Development," *The International Arab Journal of Information Technology*, vol. 20, no. 1, pp. 9-18, 2023. https://doi.org/10.34028/iajit/20/1/2

[4] Amos B., Ludwiczuk B., and Satyanarayanan M., "Openface: A general-purpose Face Recognition Library with Mobile Applications," *CMU School of Computer Science*, vol. 6, no. 2, pp. 1-20, 2016.

[5] Anwar I. and Islam N., "Learned Features are Better for Ethnicity Classification," *Cybernetics and Information Technologies*, vol. 17, no. 3, pp. 152-164, 2017. DOI: 10.1515/cait-2017-0036

[6] Blaine B. and Brenchley K., Understanding the Psychology of Diversity, *Sage Publications*, 2017.

[7] Bruce V., Burton A., Hanna E., Healey P., Mason O., Coombes A., Fright R., and Linney A., "Sex Discrimination: How Do We Tell the Difference Between Male and Female Faces?," *Perception*, vol. 22, no. 2, pp. 131-152, 1993. https://doi.org/10.1068/p220131

[8] Chen H., Deng Y., and Zhang S., "Where am I from? -East Asian Ethnicity Classification from Facial Recognition," *Project Study in Stanford University*, 2016.

[9] Ding H., Huang D., Wang Y., and Chen L., "Facial Ethnicity Classification based on Boosted Local Texture and Shape Descriptions," *in Proceedings of the 10th IEEE International Conference and Workshops on Automatic Face and Gesture Recognition (FG)*, Shanghai, pp. 1-6, 2013. DOI:10.1109/FG.2013.6553815

[10] Duan X., Wang C., Liu X., Li Z., Wu J., and Zhang H., "Ethnic Features Extraction and Recognition of Human Faces," *in Proceedings of the 2nd International Conference on Advanced Computer Control*, Shenyang, pp. 125-130, 2010. DOI: 10.1109/ICACC.2010.5487194

[11] Elbasi E. and Zreikat A., "Heart Disease Classification for Early Diagnosis based on Adaptive Hoeffding Tree Algorithm in IoMT," *The International Arab Journal of Information Technology*, vol. 20, no. 1, pp. 38-48, 2023. DOI:10.34028/iajit/20/1/5

[12] Farkas L., Katic M., Forrest C., and Alt K., "International Anthropometric Study of Facial Morphology in Various Ethnic Groups/Races," *Journal of Craniofacial Surgery*, vol. 16, no. 4, pp. 615-646, 2005. DOI:10.1097/01.scs.0000171847.58031.9e

[13] Fu S., He H., and Hou Z., "Learning Race from Face: A Survey," *IEEE Transactions on Pattern Analysis and Machine Intelligence*, vol. 36, no. 12, pp. 2483-2509, 2014. DOI: 10.1109/TPAMI.2014.2321570

[14] Fu Y., Guo G., and Huang T., "Age Synthesis and Estimation via Faces: A Survey," *IEEE Transactions on Pattern Analysis and Machine Intelligence*, vol. 32, no. 11, pp. 1955-1976, 2010. DOI: 10.1109/TPAMI.2010.36

[15] Furl N., Phillips P., and O'Toole A., "Another-race Effect for Face Recognition Algorithms. Face Recognition Algorithms and the Other-Race Effect: Computational Mechanisms for a Developmental Contact Hypothesis," *Cognitive Science*, vol. 26, no. 6, pp. 797-815, 2002. https://doi.org/10.1016/S0364-0213(02)00084-8

[16] Guo G., "Human Age Estimation and Sex Classification," *Video Analytics for Business Intelligence*, vol. 409, pp. 101-131, 2012. DOI:10.1007/978-3-642-28598-1_4

[17] Gutta S., Wechsler H., and Phillips P., "Gender and Ethnic Classification of Face Images," *in Proceedings of the 3rd IEEE International Conference on Automatic Face and Gesture Recognition*, Nara, pp. 194-199, 1998. DOI:10.1109/AFGR.1998.670948

[18] Han H., Otto C., Liu X., and Jain A., "Demographic Estimation from Face Images: Human vs. Machine Performance," *IEEE Transactions on Pattern Analysis and Machine Intelligence*, vol. 37, no. 6, pp. 1148-1161, 2014. DOI:10.1109/TPAMI.2014.2362759

[19] Hosoi S., Takikawa E., and Kawade M., "Ethnicity Estimation with Facial Images," *in Proceedings of the 6th IEEE International Conference on Automatic Face and Gesture Recognition*, Seoul,

pp. 195-200, 2004. DOI:10.1109/AFGR.2004.1301530

[20] Huang G., Mattar M., Berg T., and Learned-Miller E., "Labeled Faces in the Wild: A Database for Studying Face Recognition in Unconstrained Environments," pp. 1-11, 2008. http://vis-www.cs.umass.edu/lfw/lfw.pdf

[21] Jilani S., Ugail H., Bukar A., Logan A., and Munshi T., "A Machine Learning Approach for Ethnic Classification: The British Pakistani Face," *in Proceedings of the International Conference on Cyberworlds*, Chester, pp. 170-173, 2017. DOI: 10.1109/CW.2017.27

[22] Jin H., Wang X., Zhong Z., and Hua J., "Robust 3D Face Modeling and Reconstruction from Frontal and Side Images," *Computer Aided Geometric Design*, vol. 50, pp. 1-13, 2017. https://doi.org/10.1016/j.cagd.2016.11.001

[23] Jindal R. and Singla S., "Latent Fingerprint Recognition using Hybrid Ant Colony Optimization and Cuckoo Search," *The International Arab Journal of Information Technology*, vol. 20, no. 1, pp. 19-28, 2023. https://doi.org/10.34028/iajit/20/1/3

[24] Kingma D. and Welling M., "Auto-Encoding Variational Bayes," *arXiv Preprint arXiv*, pp. 1312-6114, 2013. https://doi.org/10.48550/arXiv.1312.6114

[25] Kisku D., Rattani A., Grosso E., and Tistarelli M., "Face Identification by SIFT-based Complete Graph Topology," *in Proceedings of the IEEE Workshop on Automatic Identification Advanced Technologies*, Alghero, pp. 63-68, 2007. DOI: 10.1109/AUTOID.2007.380594

[26] Krizhevsky A., Sutskever I., and Hinton G., "ImageNet Classification with Deep Convolutional Neural Networks," *Advances in Neural Information Processing Systems*, vol. 25, no. 2, pp. 1097-1105, 2012. https://scirp.org/reference/referencespapers.aspx?referenceid=3031254

[27] LeCun Y., Boser B., Denker J., Henderson D., Howard R., Hubbard W., and Jackel L.,"Backpropagation Applied To Handwritten Zip Code Recognition," *Neural Computation*, vol. 1, no. 4 pp. 541-551, 1989. DOI: 10.1162/neco.1989.1.4.541

[28] Lin Q., Man Z., Cao Y., and Wang H., "Automated Classification of Whole-Body SPECT Bone Scan Images with VGG-Based Deep Networks," *The International Arab Journal of Information Technology*, vol. 20, no. 1, pp. 1-8, 2023. https://doi.org/10.34028/iajit/20/1/1

[29] Liu C. and Wechsler H., "Independent Component Analysis of Gabor Features for Face Recognition," *IEEE Transactions on Neural Networks*, vol. 14, no. 4, pp. 919-928, 2003. DOI: 10.1109/TNN.2003.813829

[30] Liu W., Wen Y., Yu Z., Li M., Raj B., and Song L., "SphereFace: Deep Hypersphere Embedding for Face Recognition," *in Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition*, pp. 212-220, Honolulu, 2017. https://openaccess.thecvf.com/content_cvpr_2017/html/Liu_SphereFace_Deep_Hypersphere_CVPR_2017_paper.html

[31] Liu Z., Luo P., Wang X., and Tang X., "Large-Scale CelebFaces Attributes (CelebA) Dataset," *Retrieved August*, vol. 15, pp. 1-11, 2018. https://mmlab.ie.cuhk.edu.hk/projects/CelebA.html

[32] Megias A., Rzeszewska I., Aguado L., and Catena A., "Influence of Cross-ethnic Social Experience on Face Recognition Accuracy and the Visual Perceptual Strategies Involved," *International Journal of Intercultural Relations*, vol. 65, pp. 42-50, 2018. https://doi.org/10.1016/j.ijintrel.2018.04.008

[33] Ng C., Tay Y., and Goi B., "EtVision-based Human Gender Recognition: A Survey," *arXiv preprint arXiv*, pp. 1204-1611, 2012. https://doi.org/10.48550/arXiv.1204.1611

[34] Parkhi O., Vedaldi A., and Zisserman A., "Deep Face Recognition," *in bmvc*, pp. 1-12, 2015. https://www.robots.ox.ac.uk/~vgg/publications/2015/Parkhi15/parkhi15.pdf

[35] Phillips P., Jiang F., Narvekar A., Ayyad J., and O'Toole A., "An other-race Effect for Face Recognition Algorithms," *ACM Transactions on Applied Perception*, vol. 8, no. 2, pp. 1-14, 2011. https://doi.org/10.1145/1870076.1870082

[36] Ramanathan N., Chellappa R., and Biswas S., "Computational Methods for Modeling Facial Aging: A Survey," *Journal of Visual Languages and Computing*, vol. 20, no.3 pp. 131-144, 2009. https://doi.org/10.1016/j.jvlc.2009.01.011

[37] Rumelhart D., Hinton G., and Williams R., "Learning Representations by Back-Propagating Errors," *Cognitive Modeling*, vol. 5, pp. 1, 1988. https://www.scirp.org/(S(czeh2tfqyw2orz553k1w0r45))/reference/ReferencesPapers.aspx?ReferenceID=2239759

[38] Savran A., Alyüz N., Dibeklioglu H., Çeliktutan Q., Gökberk B., Sankur B., and Akarun L., "Bosphorus Database for 3D Face Analysis," *European Workshop on Biometrics and Identity Management*, vol. 5372, pp. 47-56, 2008. https://doi.org/10.1007/978-3-540-89991-4_6

[39] Sayers J., "The Bonds of Love-Psychoanalysis, Feminism and the Problem of Domination-Benjamin J," *Feminism and Psychology*, vol. 3, pp. 250-251, 1993. https://kar.kent.ac.uk/id/eprint/20663

[40] Schirrmeister R., Springenberg J., Fiederer L., and Glasstetter M.,"Deep Learning with Convolutional Neural Networks for EEG

Decoding and Visualization," *Human Brain Mapping*, vol. 38, no. 11, pp. 5391-5420, 2017. https://doi.org/10.1002/hbm.23730

[41] Schroff F., Kalenichenko D., and Philbin J., "FaceNet: A Unified Embedding for Face Recognition and Clustering," *in Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition*, Boston, pp. 815-823, 2015. DOI: 10.1109/CVPR.2015.7298682

[42] Shakhnarovich G., Viola P., and Moghaddam B., "A Unified Learning Framework for Real Time Face Detection and Classification," *in Proceedings of the 5th IEEE International Conference on Automatic Face Gesture Recognition*, Washington, pp. 16-23, 2002. DOI: 10.1109/AFGR.2002.1004124

[43] Shin M., Seo J., and Kwon D., "Face Image-Based Age and Gender Estimation with Consideration of Ethnic Difference," *in Proceedings of the 26th IEEE International Symposium on Robot and Human Interactive Communication (RO-MAN)*, Lisbon, pp. 567-572, 2017. DOI: 10.1109/ROMAN.2017.8172359

[44] Sun Y., Chen Y., Wang X., and Tang X., "Deep Learning Face Representation by Joint Identification-Verification," *Advances in Neural Information Processing Systems*, vol. 2, pp. 1988-1996, 2014. https://dl.acm.org/doi/10.5555/2969033.2969049#d21007992e1

[45] Sun Y., Wang X., and Tang X., "Deep Learning Face Representation from Predicting 10,000 Classes," *in Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition*, Columbus, pp. 1891-1898, 2014. DOI: 10.1109/CVPR.2014.244

[46] Sun Y., Wang X., and Tang X., "Deeply Learned Face Representations are Sparse, Selective, and Robust," *in Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition*, Boston, pp. 2892-2900, 2015. DOI: 10.1109/CVPR.2015.7298907

[47] Sun Y., Wang X., and Tang X., "Hybrid Deep Learning for Face Verification," *in Proceedings of the IEEE International Conference on Computer Vision*, Sydney, pp. 1489-1496, 2013. DOI: 10.1109/ICCV.2013.188

[48] Szegedy C., Liu W., Jia Y., Sermanet P., Reed S., and Anguelov D., "Going Deeper with Convolutions," *in Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition*, Boston, pp. 1-9, 2015. DOI: 10.1109/CVPR.2015.7298594

[49] Taigman Y., Yang M., Ranzato M. A., and Wolf L., "Web-Scale Training For Face Identification," *in Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition*, Boston, pp. 2746-2754, 2015. https://www.cv-foundation.org/openaccess/content_cvpr_2015/html/Taigman_Web-Scale_Training_for_2015_CVPR_paper.html

[50] Taigman Y., Yang M., Ranzato M., and Wolf L., "Deep-Face: Closing the Gap to Human-Level Performance in Face Verification," *in Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition*, Columbus, pp. 1701-1708, 2014. DOI: 10.1109/CVPR.2014.220

[51] Tan H., Yang B., and Ma Z., "Face Recognition Based on the Fusion of Global and Local HOG Features of Face Images," *IET Computer Vision*, vol. 8, no. 3, pp. 224-234, 2014. doi: 10.1049/iet-cvi.2012.0302

[52] Valentin D., Abdi H., O'Toole A., and Cottrell G., "Connectionist Models of Face Processing: A Survey," *Pattern Recognition*, vol. 27, no. 9, pp. 1209-1230, 1994. https://doi.org/10.1016/0031-3203(94)90006-X

[53] Vijayan V., Bowyer K., Flynn P., Huang D., Chen L., and Hansen M., "Twins 3D Face Recognition Challenge," *in Proceedings of the International Joint Conference on Biometrics*, Washington, pp. 1-7, 2011. DOI: 10.1109/IJCB.2011.6117491

[54] Wang W., He F., and Zhao Q., "Facial Ethnicity Classification with Deep Convolutional Neural Networks," *in Proceedings of the 11th Chinese Conference on Biometric Recognition*, *Chengdu,* pp. 176-185, 2016. https://doi.org/10.1007/978-3-319-46654-5_20

[55] Wang Y., Shen J., Petridis S., and Pantic M., "A Real-Time and Unsupervised Face Re-Identification System for Human-Robot Interaction," *Pattern Recognition Letters*, vol. 128, pp. 559-568, 2019. https://doi.org/10.1016/j.patrec.2018.04.009

[56] Weinberger K. and Saul L., "Distance Metric Learning for Large Margin Nearest Neighbor Classification," *Journal of Machine Learning Research*, vol. 10, pp. 207-244, 2009. https://jmlr.csail.mit.edu/papers/volume10/weinberger09a/weinberger09a.pdf

[57] Wu S. and Wang D., "Effect of Subject's Age and Gender on Face Recognition Results," *Journal of Visual Communication and Image Representation*, vol. 60, pp. 116-122, 2019. https://doi.org/10.1016/j.jvcir.2019.01.013

[58] Yan Y., "Is he Chinese, Korean or Japanese?," *Project study in Stanford University*, 2015. http://cs231n.stanford.edu/reports/2015/pdfs/yuyuan1_project.pdf

[59] Yi D., Lei Z., Liao S., and Li S., "Learning Face Representation from Scratch," *arXiv Preprint arXiv*, pp. 1411-7923, 2014. https://doi.org/10.48550/arXiv.1411.7923

[60] Zhang D., Meng D., and Han J., "Co-saliency Detection via a Self-Paced Multiple Instance-Learning Framework," *IEEE Transactions on*

Pattern Analysis and Machine Intelligence, vol. 39, no. 5, pp. 865-878, 2017. DOI: 10.1109/TPAMI.2016.2567393

[61] Zhang G., Huang X., Li S., Wang Y., and Wu X., "Boosting Local Binary Pattern (LBP)-Based Face Recognition," *in Proceedings of the* 5th *Chinese Conference on Biometric Recognition*, Guangzhou, pp. 179-186, 2004. DOI:10.1007/978-3-540-30548-4_21

[62] Zhang L., Gao Y., Xia Y., Da Q., and Li X., "A Fine-Grained Image Categorization System by Cellet-Encoded Spatial Pyramid Modeling," *IEEE Transactions on Industrial Electronics*, vol. 62, no. 1, pp. 564-571, 2015. DOI: 10.1109/TIE.2014.2327558

[63] Zhang L., Hong R., Gao Y., Ji R., Dai Q., and Li X., "Image Categorization by Learning a Propagated Graphlet Path," *IEEE Transactions on Neural Networks and Learning Systems*, vol. 27, no. 3, pp. 674-685, 2016. DOI: 10.1109/TNNLS.2015.2444417

[64] Zhu Z., Luo P., Wang X., and Tang X., "Deep Learning and Disentangling Face Representation by Multi-View Perceptron," *in Proceedings of the Advances in Neural Information Processing Systems*, 2014.

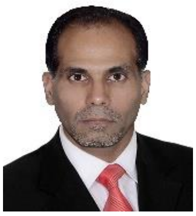
[65] Zhu Z., Luo P., Wang X., and Tang X., "Deep Learning Identity-Preserving Face Space," *in Proceedings of the IEEE International Conference on Computer Vision*, Sydney, pp. 113-120, 2013. DOI: 10.1109/ICCV.2013.21

[66] Zhu Z., Luo P., Wang X., and Tang X., "Recover Canonical-View Faces in the Wild with Deep Neural Networks," *arXiv Preprint arXiv*, pp. 1404-3543, 2014. https://doi.org/10.48550/arXiv.1404.3543

[67] Zhuang Z., Landsittel D., Benson S., Roberge R., and Shaffer R., "Facial Anthropometric Differences among Gender, Ethnicity, and Age Groups," *Annals of Occupational Hygiene*, vol. 54, no. 4, pp. 391-402, 2010. https://doi.org/10.1093/annhyg/meq007

Mohammed Alghaili received his PhD from the College of Computer Science and Electronic Engineering from Hunan University, Changsha. He received the B.Sc. from College of Computer Science, Thamar University, Thamar, Yemen. M.S degree from Department of Computer Science-the University of Peshawar-Pakistan. He worked as a Lecturer in Hajjah University, Hajjah, Yemen from 2004 to 2011. He published some papers in reputable journals. His current research interests include Machine Learning, Face Re-Identification, and Feature Extraction.

Zhiyong Li (Member, IEEE) received the Ph.D. degree in control theory and control engineering from Hunan University, Changsha, China, in 2004. He was a Postdoctoral Researcher with the Department of Computer Science, University of Dortmund, Germany, from 2006 to 2007. He is currently a Full Professor of computer science and technology with Hunan University. He has published more than 100 articles in international journals and conferences. His major research areas include Intelligent Optimization Algorithms, Intelligent Perception, and Machine Learning. He is also a member of the China Computer Federation (CCF) and Chinese Association for Artificial Intelligence (CAAI).



Ahmed AlBdair Computer Center, Babylon University, Babylon, Iraq. He received the BSc degree in Computer Science from College of Science, Babylon University, Babylon, Iraq, in 2004 and received the M.E. degree from School of Computer Science and Electronic Engineering, Hunan University, Hunan, Changsha, China, in 2012. Graduated Ph.D. in 2022 from Hunan University. His research work in Deep Learning, Face Recognition and Ethnic Recognition. Orcid number: https://orcid.org/0000-0002-3355-6596

Malasy Katiyalath Received the BSc degree in Finance-Banking from Rattana Business Administration College, Vientiane, Laos in 2008. Receaived Master degree in Finance, Zhongnan University of Economic and Law, Wuhan, China in 2013. And now Ph.D student in Hunan University, College of Economics and Trade, research in the Image Processing and Deep Learning.