Ethnicity Classification: A Machine Learning Approach

Angela U. Makolo, Oluwasegun V. Dada

Computer Science Department, Faculty of Science, University of Ibadan, Oyo State, Nigeria.
Tel: +2348038429948, +2348066657949
Email: aumakolo@gmail.com, dadaoluwasegunv@gmail.com

Abstract

Recently, researchers in the field of Machine Learning have paid a lot of interest to the human face. Soft biometric features taken from a facial image can be used to distinguish between racial classes. Other soft biometrics features include race, age, gender, and emotions. Research has employed different techniques (traditional and deep learning) in predicting the major racial classes (Asian, Hispanic, African, Caucasian) with outstanding cutting-edge performances. Recently, research has focused on identifying distinguishing characteristics in sub-racial (ethnic) groups. Racial profiling has been used in a variety of fields, including social media profiling, security surveillance, law enforcement, and targeted advertising. By seeing relevant studies in the field, we noted that the Black race (African/African American) is considered a single racial entity, models developed do not have practical application in the Nigerian domain, and most of the datasets available are racially imbalanced. As a result, the goal of this research is to create a unique dataset with accurate labels for Nigeria's three major ethnic groups, and then using deep learning techniques to classify these labels. There are three labels in the image dataset: Hausa, Igbo, and Yoruba. For feature extraction and classification, a pre-trained Convolutional Neural Network (CNN) was used. The model was evaluated on the test, and The Hausa ethnic group had the highest accuracy of 87.3%; lower accuracies were recorded from the Igbo and Yoruba subclass, which gave an accuracy of 56.0% and 56.0%, respectively. The result could be attributed and migration and inter-ethnic marriages which have dwindled the boundary between the ethnic groups.

Key words: *Racial Classification, Convolutional Neural Network, Computer Vision, Transfer Learning.*

1. Introduction

Human faces contain a wealth of soft biometric information, including height, weight, ethnicity, age, and gender. Height, weight, and age are examples of soft biometric traits that change over time. Gender and ethnicity, on the other hand, are permanent and stable characteristics. These characteristics can help with face recognition, re-identification, and racial classification.

Facial processing technology (FPT) has affected our lives in numerous ways [35]. It has found practical applications in self-driving cars, mobile devices, more innovative advertising, and finding missing persons. It has also been used in schools, banking halls, public squares, self-driving cars, concerts, apartment complexes, and airports to control access to restricted areas. Hundreds of millions of photos are uploaded to social media sites like Facebook, Instagram, and Flickr as well as cloud storage services like Dropbox and Google Drive due to the growing use of smart devices and high-speed internet. Organizing and retrieving relevant information that will be useful for decision-making from these uploaded photos is incredibly challenging and determines user experience on these platforms [34].

In computer vision, facial analytics is a challenging topic that has been extensively researched for more than two decades. The objective is to gather as much data as you can about a face, such as its location, pose, gender, ID, age, and emotion. This technology can be used for a variety of things, such as autonomous vehicles, active cell phone authentication, biometric payment, and person recognition and identification in surveillance footage [36]. Additionally, face recognition and verification from unrestricted images and videos, which includes tasks like face detection, facial landmark localization, etc., is gaining popularity.

In many different applications, the biometrics research community has frequently used soft traits: In situations where a face image is obscured or taken at awkward viewing angles, they can provide more useful information, improving the overall performance of face recognition. This could be used to gather important demographic data. Gender, race, and age data may be used to focus the search when comparing unknown faces to known faces in a database. Race-based classification of people has a big impact on social media platform monitoring, advertising, and profiling. Various visual complications, including complex variations in pose, facial expression, lighting conditions, racial and ethnic affiliation, age, gender, attire, hairdo, and other parameters, have been shown in numerous studies on

The term "race" refers to the division of people into relatively large and distinct populations or groups. These divisions are frequently made based on physical characteristics based on inherited phenotypic traits or geographic ancestry, but they are also frequently influenced and correlated with characteristics like culture, race, and socioeconomic status. People are categorized as belonging to an ethnic group based on shared traits that set them apart from other groups, such as a common set of customs, culture, nation, religion, language, history, society, ancestry, or values.

Numerous elements, such as the environment, genes, society, and others, have a big impact on how people look. It is challenging to distinguish between the genes of various racial groups due to different gene fragments. People of various races therefore share facial features. In order to test the idea that different variations might exist even within the same ethnic group, [8] examined three sub-ethnic groups (Zhuang, Tibetan, and Han). The need for more distinct and precise traits that can categorize intra-racial groups in nearby geographic areas has thus increased noticeably in recent years.

Due to their relative discriminatory nature and the availability of relevant preparation tools for training and testing, the primary race groups (Caucasians, African Americans, and Asians) have been the focus of many existing efforts in terms of racial categorization. In recent years, a lot of researchers have shifted their attention from popular racial groups like Blacks, Asians, and Caucasians (Whites) to sub-ethnic groups like Hans, Tibetans and Zhuang which are sub-ethnic groups that belong to the Asian race, but there has never been an attempt to categorize sub-ethnic groups of African descent. This research, which is motivated by this notion, attempts to examine a system that can identify race using deep learning for the various subethnic groups in Nigeria (Yoruba, Igbo, and Hausa)

2. Related Works

Computer vision researchers have already addressed race classification in numerous ways. But most of these approaches use landmark localization or consider a face image to be a one-dimensional feature vector. When comparing the two approaches, landmark localization has proven to perform better. In some critical scenarios, however, landmark localization is a daunting task. Landmark extraction methods, for example, fail when the facial expression is complicated, there are changes in face rotation or lighting conditions, and the image is taken from a distance. In this section we will look at past literatures and studies that employed CNN for image classification.

[1] proposed a CNN-based multi-task learning algorithm for simultaneously performing face detection, face alignment, pose estimation, gender recognition, smile detection, age estimation, and face recognition Parameters for concurrent learning were shared among the CNN's five sub-networks. They tested their method on various unconstrained datasets and discovered that it improved performance significantly. The accuracy and effectiveness of the machine learning model are improved through this process.

[2] used a Support Vector Machine classifier with a linear kernel and a convolutional neural network called VGGface to extract features from a large dataset of faces from over 2.6 million celebrity images. It was trained using facial images from ten different databases in three distinct classes: African-American, Asian, and Caucasian. For Asians, Caucasians, and African-Americans, in all these databases, the average classification accuracy is 98.28%, 99.05%, and 99.66%, respectively.

[3] used a pre-trained CNN for a variety of face-related tasks, including recognizing ethnicity. They increased the training set for their network using a data augmentation technique, which gave the FERET dataset an accuracy of 93.9%. on the three classes considered: White, Asian, and Other.

[4] Proposed a deep ConvNet classification model that has three convolutional and pooling layers, two fully connected layers at the end, and three layers total. In addition to two self-collected datasets, the model was applied to multiple datasets. Some datasets were used for testing purposes to assess how well images from different

trained datasets were classified. Separate classifications were made for white and black, Chinese and non-Chinese, and finally Han and Uyghurs.

[5] examined the issues with face images in relation to distance, darkness, and uncontrolled. They used NIR images from the LDWVU Database taken at 30, 60, 90, and 120 meters at night, as well as visible images at 1.5 meters. The LDHF database was also used. A Convolutional Neural Network with VGG architecture was utilized; Asians and Caucasians were categorized by their gender and ethnicity within these environments, and the results of 78.98% showed an improvement over earlier results.

[6] presented images of people from East Asian countries, both constrained and unconstrained, from a new dataset called WEAFD. WEAFD was also trained using a CNN model with three convolutional layers and several fully connected layers to classify age, gender, and ethnicity. They showed two networks, one with images of the entire face and the other with images of the face divided into regions. The first network produced better results for age, gender, and ethnicity than the second. However, compared to gender (88.02%, 84.70%), ethnicity (24.06%, 33.33%) and age (38.04%, 36.43%) performed poorly in both networks. They clarify that if training data for age and ethnicity were not sufficient, that might be the reason for the poor result. Another factor might be the accuracy of the labels, as they noted that gender labels are more susceptible to human error than age and ethnicity labels.

[7] used the FERET database, which contains 447 facial images, to predict three ethnicities: Caucasian, Mongolian, and Black race. The method, which is superior to a multiplayer perceptron (MLP) network, makes use of the geometric features and color information extracted from the convolutional layers. This is appropriate for categories that appear to be discriminatory, such as separating Black and White people from European and East Asian people, which was the ethnic group considered in this study. In general, classifying individuals with close geographic ties is more meaningful. [10].

[8] used a hybrid CNN classifier (VGG-16) that had been trained on ImageNet database, with an image ranking engine and trained the Support Vector Machine (SVM) using the hybrid features. An updated dataset of the faces of Bangladeshi, Chinese, and Indian people was used to assess the method. Improvement was seen in comparison to Faster R-CNN and [4] showing an accuracy of 95.2%.

[9] They explored the possibility of classifying three sub-ethnic Arab groups: those from the Levant, the Gulf Cooperation Council (GCC), and Egypt using a CNN model. In order to solve the issue, two different learning algorithms were considered. Unsupervised deep learning, a Convolutional Neural Network (CNN) pre-trained model, and supervised deep learning are the first types (deep clustering). When the labels in the Arab dataset were balanced, the best outcomes came from pre-training a CNN on the entire Arab dataset, then evaluating it on a different dataset. The results were 57% and 52%. A range of datasets were subjected to deep clustering techniques, producing ACCs between 32% and 59%.

It was clear from the earlier works that they had not considered the classification of African Sub-racial groups in Africa. Additionally, the application of pre-trained CNNs to this issue has not yet been investigated. In order to address this issue, this study considered training a deep convolutional neural network using a transfer learning strategy.

TABLE 1: THE LITERATURE SUMMARY OF ETHNICITY CLASSIFICATION STUDIES

Author(s)	Approaches (Feature + classifier)	Accuracy	Dataset	Details (Specific Information)
Ahmed *et al*., 2008	CNN + Transfer learning	93.9%	FRGC 2.0 DB (14714 Faces)	Asian, Caucasian, other
Narang and Bourlai, 2016	CNN architecture of VGG	95%	WVU and the LDHF database	Asian or Caucasian
Chen *et al*., 2016	KNN	57.5%	Self-collected with 1380 images of Chinese, Korean and Japanese subjects	Chinese, Japanese, and Korean
	SVM	62.1%		
	Two-layer Neural Net	64.7%		
	CNN	89.2%		
Anwar and Islam, 2017	SVM + pre-trained CNN	98.99%	FERET and CASPEAL	Asian, African Americans, Caucasian
Srinivas *et al*., 2017	CNN	24.06%: Full image 33.33%: face segmentation	WEAFD	Chinese, Japanese, Korean, Filipino, Indonesian, Malaysian, Vietnamese etc.
Masood *et al*, 2018	Pretrained VGGNet -10 model	98.6%	FERET	Mongolian, Caucasian, and Negroid
Heng *et al*, 2018	Pretrained CNN hybridized with image ranking engine.	95.2	Self-collected dataset	Bangladeshi, Indian and Chinese
Samir Brahim *et al*, 2020	DCNN VGG 16 and Gabor filter, classification using SVM	92.5%	CUHK and KFDB	Chinese vs Korean
Al-Humaidana *et al*, 2021	CNN and deep clustering	56.97%	Self-collected Arab image dataset	GCC, the Egyptian and the Levant

3. Methodology

In this section, we go over the data collection strategies and preprocessing techniques in detail and the Deep Learning Architecture used for training and evaluating the degree of correctness of our model.

3.1 Data Collection Strategies

Data was collected from various sources and later aggregated to form the custom dataset used for training, evaluation, and testing. The following strategies were employed:

- A sample dataset of 9000 images (3000 per class) was collected from various sources (Internet, Crowdsourcing, Photoshoot).
- To guarantee that images are distinct, the subjects for each ethnic group were carefully selected, i.e., demographic information was considered.
- Cleaning of data – to remove unrelated, duplicated images, images of inferior quality/resolution

3.2 Nature of Dataset

To increase variation, the dataset might contain more than one image for each subject. This dataset has images that depict a wide range of poses, expressions, lighting variations, and background clutter. The images vary significantly in age, ethnicity, and background. It also covers a wide range of accessories such as eyeglasses, sunglasses, hats etc. Unconstrained images obtained from profile pictures, mug shots, and selfies are considered.

Figure 1: Sample of Dataset

3.3 Preprocessing

Data preprocessing is necessary for cleaning the data collected from various sources before passing it through and machine-learning model. This process aids in enhancing the machine learning model's efficacy and accuracy. The stages are:

- Face detection and alignment
- Face recognition and extraction
- Image Enhancement
- Image cleaning
- Image resizing

On the facial images, the Multi-task Cascaded Convolutional Network (MTCNN) model, which recognizes faces and landmarks, was applied. The detected face is then cropped and resized to 224 x 224, a size that works with the pre-trained CNN of choice.

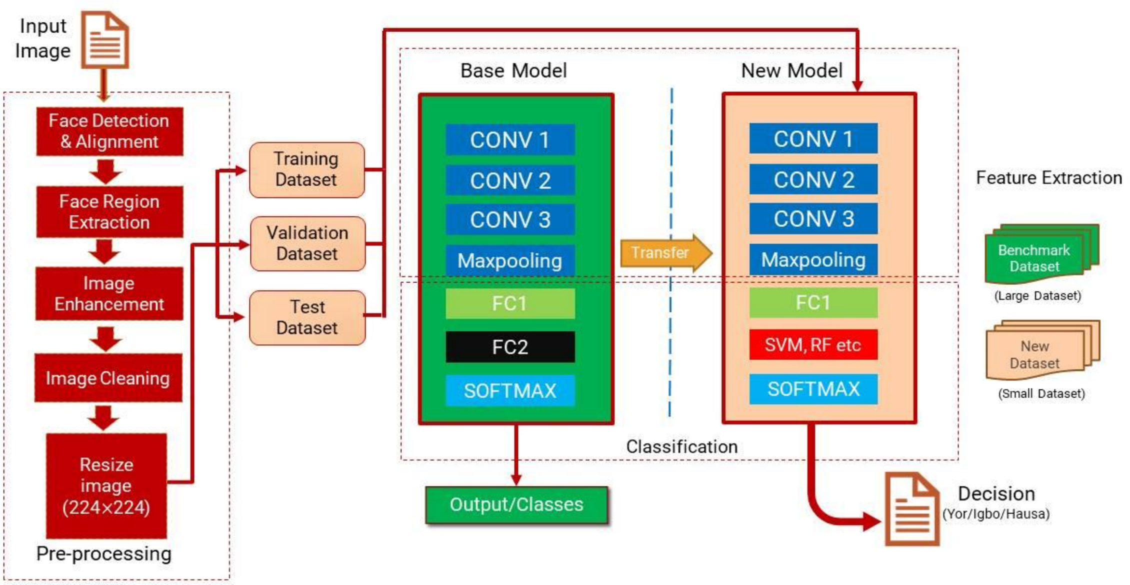

Figure 2: Image classification

Feature extraction represents the face with a more compact and discriminative set of features. CNN's convolutional layers serve as feature extractors, so they do away with the requirement for a separate feature extraction step. The fully connected (FC) layers, which serve as the classifier component and connect every node of the previous layer to every other node of the next layer, receive the features extracted by a series of these convolution layers.

The task of ethnicity identification can be considered as an N-way classification problem where N = 3 (Yoruba, Igbo, and Hausa) for the dataset

3.5 *Transfer Learning Approach*

In deep learning, transfer learning is a technique whereby learning from a similar problem is transferred to a new domain to reduce the training time and the amount of training set needed. Modern deep CNNs rarely start from scratch when training a convolutional network because of insufficient datasets [31]. Instead, they employ a network like VGG that has already been trained on a sizable dataset (like ImageNet), using it as an initialization or fixed feature extractor for new tasks.

In this experiment, we selected the CNNs that among contemporary deep network architectures are the most interesting and promising for our analysis. Specifically, ResNet-50, MobileNet v2, EfficientNet, VGG-16, and VGGFace. In the choice of CNN architecture, we considered the pre-trained model's speed, size, and scalability.

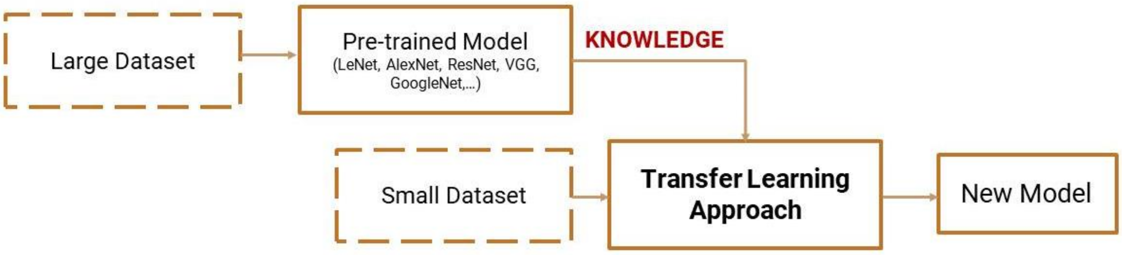

Figure 3: A transfer learning approach

The EfficientNetB3 pre-trained model was used, a convolutional neural network architecture and scaling method that uniformly scales all depth/width/resolution dimensions using a compound coefficient. The EfficientNet scaling method uniformly scales network depth, width, and resolution using a static set of scaling coefficients, unlike in traditional practice, where random values are chosen for scaling the network depth, width, and resolution.

Experimental setup

All images have been preprocessed to have a similar resolution (224×224) and are cropped to keep only the region of interest. The dataset was split into three: Training (70%), Validation (20%), and Test sets (10%)

We carried out several experiments on the faces of three different ethnic groups. using a number of convolution layers to extract ethnic salient features. A different mix of network hyper-parameters such as learning rate, epoch count, batch size, activation function, loss function, and optimizers was evaluated to achieve high accuracy.

4. Results and Analysis

4.1 Analysis

EfficientNet, a pre-trained network that was previously initialized with weights from ImageNet, was used to train the model. The classification outcome and score for all training and test images are then obtained using the trained CNN model. We conducted several experiments on the face images of three ethnic groups to extract ethnic salient features using a series of convolution layers. After this, the performance of ethnicity recognition models is evaluated using the following metrics: Accuracy, F1-score, Precision, and Recall.

$$Precision = \frac{TP}{TP + FP}$$

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

$$F1 = \frac{2 \times precision \times recall}{precision + recall}$$

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

Where:

- TP means True Positive
- FP means False Positive
- TN means True Negative
- FN means False Negative

To increase classification accuracy, we first tested a variety of network hyper-parameters, including batch size, number of epochs, learning rate, type of activation function, size of hidden layers, type of loss function, and optimizers. After conducting numerous tests to determine how hyper-parameters affect performance, The set of hyper-parameters with the best performance was chosen by

For the ethnicities Hausa, Igbo, and Yoruba, respectively, we achieved classification accuracies of 87.3%, 56%, and 56% using the selected base model. The overall accuracy and loss for the various ethnic groups are summarized in the table below.

Ethnic class	**Accuracy (%)**
Hausa	87.3
Igbo	56.0
Yoruba	56.0
Average	66.4%

The Hausa Ethnic class recorded the highest accuracy, followed closely by the Yoruba classes. We can conclude that the model is struggling to predict the Igbo class because of the close similarities between the two classes.

The training accuracy, validation accuracy, training loss, and validation loss are shown below

Figure 4: Training accuracy and validation accuracy

	precision	recall	f1-score	support
Hausa	0.77	0.87	0.82	600
Igbo	0.60	0.56	0.58	600
Yoruba	0.60	0.56	0.58	600
accuracy			0.66	1800
macro avg	0.66	0.66	0.66	1800
weighted avg	0.66	0.66	0.66	1800

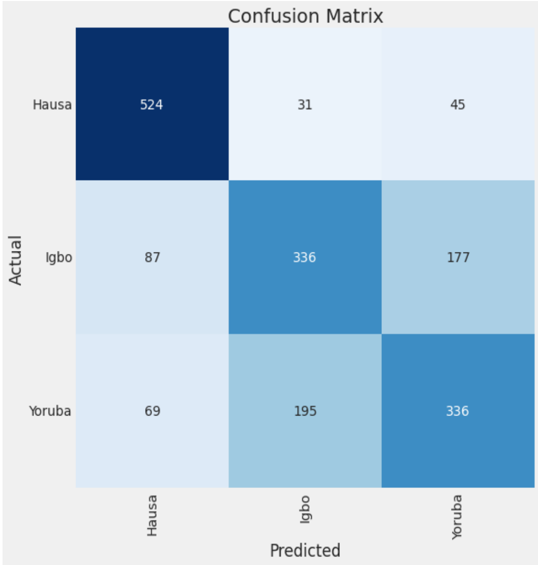

Figure 5: Confusion Matrix

From the confusion matrix, the Hausa class had the highest number of TP and TN hence the highest accuracy. A substantial part of the Igbo class was misclassified as Yoruba because of the similarities within the classes.

In this study, we investigated the possibility of categorizing the three main ethnic groups in Nigeria using a CNN model. First, we developed a unique Nigerian dataset with three labels selected in accordance with the distribution of countries into regions. And after image preprocessing, a Transfer learning technique was used for feature extraction and classification. The classification report showed close similarities between the Igbo and the Yoruba class based on the number of FP values recorded.

5. Conclusion

The classification of races can be difficult when taking similar appearances, to add to the complication, the increasing growth of the mixed-race population because of migration and intermarriage has further blurred the class boundary. Due to the strong similarities between the ethnic groups, especially the Igbo and Yoruba ethnic classes, the model struggled to differentiate between labels. An unbalanced dataset may lead to biased results. So, we used the equal contribution of each subclass in the main class. The quality of the images available for training significantly impacts the model's performance.

The following are open problem areas in which the extension and the improvement of this work can be realized: Extending the number of classes (ethnic groups within Nigeria), enhancing the robustness by investigating various image conditions, such as poses and changing lighting. A combination of the Handcrafted method could be investigated.

This research was funded by National Institutes of Health, grant number U41HG006941; H3Africa Bioinformatics Network. The funders or the H3Africa Initiative had no role in the design of the study; in the collection, analyses, or interpretation of data; in the writing of the manuscript; or in the decision to publish the results.

References

[1] R. Ranjan, S. Sankaranarayanan, A. Bansal (2018). Deep Learning for Understanding Faces: Machines May Be Just as Good, or Better, than Humans. IEEE Signal Processing Magazine, vol. 35, no. 1, pp. 66-83, Jan. 2018, doi: 10.1109/MSP.2017.2764116.

[2] I Answar and N. Ul Islam (2017). Learned features are better for ethnicity classification. Cybern. Inf. Technol., vol. 17, no. 3, pp. 152–164, Sep. 2017.

[3] Ahmed, A., Yu, K., Xu, W., Gong, Y., Xing, E (2008). Training Hierarchical Feedforward Visual Recognition Models Using Transfer Learning from Pseudo-Tasks. In: Forsyth, D., Torr, P., Zisserman, A. (eds.) Computer Vision – ECCV. pp. 69–82.

[4] W. Wang, F. He, and Q. Zhao (2016). Facial ethnicity classification with deep convolutional neural networks. in Biometric Recognition. Cham, Switzerland: Springer, 2016, pp. 176–185.

[5] N. Narang and T. Bourlai (2016). Gender and ethnicity classification using deep learning in heterogeneous face recognition. 2016 International Conference on Biometrics (ICB) pp. 1-8, doi: 10.1109/ICB.2016.7550082.

[6] N. Srinivas, H. Atwal, D. C. Rose, G. Mahalingam, K. Ricanek, & D. S. Bolme (2017). Age, Gender, and Fine-Grained Ethnicity Prediction Using Convolutional Neural Networks for the East Asian Face Dataset. 12th IEEE International Conference on Automatic Face & Gesture Recognition (FG, 2017) pp. 953-960, doi: 10.1109/FG.2017.118.

[7] Masood S., Gupta S., Wajid A., Gupta S., Ahmed M. (2018). Prediction of Human Ethnicity from Facial Images Using Neural Networks. In: Satapathy S., Bhateja V., Raju K., Janakiramaiah B. (eds) Data Engineering and Intelligent Computing. Advances in Intelligent Systems and Computing, vol 542. Springer, Singapore. https://doi.org/10.1007/978-981-10-3223-3_20

[8] Z. Heng, M. Dipu, and K.-H. Yap (2018). Hybrid supervised deep learning for ethnicity classification using face images," in Proc. IEEE Int. Symp. Circuits Syst. (ISCAS), May 2018, pp. 1_5.

[9] N. A. A. Humaidan. and M. Prince (2021). A Classification of Arab Ethnicity Based on Face Image Using Deep Learning Approach. IEEE Access, vol. 9, pp. 50755–50766, 2021, doi: 10.1109/ACCESS.2021.3069022.

[10] B. Wu, H.Z. Ai, and C. Huang (2004). Facial image retrieval based on demographic classification. Proc. IEEE Int. Conf. Pattern Recognition, vol. 3, pp. 914–917.

[11] Becerra-Riera F., Llanes N.M., Morales-González A., Méndez-Vázquez H., Tistarelli M. (2019) On Combining Face Local Appearance and Geometrical Features for Race Classification. In: Vera-Rodriguez R., Fierrez J., Morales A. (eds) Progress in Pattern Recognition, Image Analysis, Computer Vision, and Applications. CIARP 2018. Lecture Notes in Computer Science, vol 11401. Springer, Cham. https://doi.org/10.1007/978-3-030-13469-3_66

[12] Belhaouari, Samir Brahim, and Athmar N.M. Shamhan (2020). Fusion of Deep Learning and Handcrafted Features for Intra-Race Recognition. Advances in Natural and Applied Sciences, vol. 14, no. 3, Sept.-Dec. 2020, pp. 76+. Gale Academic OneFile.

[13] G. D. Guo and G. W. Mu (2010). A study of large-scale ethnicity estimation with gender and age variations. Proc. IEEE Comput. Soc. Sci. Comput. Vis. Pattern Recog. Workshops, pp. 79–86, 2010

[14] Inzamam Anwar and Naeem U Islam (2017). Learned features for better ethnicity. Cybernetics and Information Technologies, vol. 17, no. 3, pp. 152-164. DOI: 10.1515/cait-2017-0036

[15] Katia Huri, Eli David, & Nathan S. Netanyahu (2018). DeepEthnic: Multi-Label Ethnic Classification from Face Images. International Conference on Artificial Neural Networks (ICANN), Springer LNCS, Vol. 11141, pp. 604-612.DOI: 10.1007/978-3-030-01424-7_59

[16] Khalil Khan and Jehad Ali and Irfan Uddin and Sahib Khan and Byeong-hee Roh (2021). A Facial Feature Discovery Framework for Race Classification Using Deep Learning. arXiv:2104.02471

[17] Khan, Akbir & Mahmoud, Marwa. (2018). Considering Race, a Problem of Transfer Learning. arXiv:1812.04751

[18] Lu X., Chen H., Jain A.K. (2005) Multimodal Facial Gender and Ethnicity Identification. In: Zhang D., Jain A.K. (eds) Advances in Biometrics. ICB 2006. Lecture Notes in Computer Science, vol 3832. Springer, Berlin, Heidelberg. https://doi.org/10.1007/11608288_74

[19] Lyle, J.R., Miller, P.E., Pundlik, S.J., Woodard, D.L. (2010). Soft biometric classification using local appearance periocular region features. In: Fourth IEEE International Conference on Biometrics: Theory, Applications, and Systems (BTAS), pp. 1–7. https://doi.org/10.1016/j.patcog.2012.04.027

[20] M. Demirkus, K. Garg, and S. Guler (2010). Automated person categorization for video surveillance using soft biometrics," in Proc. SPIE, Biometric Technol. Human Identification VII, pp. 76670P.

[21] Mohammad, A.S., Al-Ani, J.A., (2018). Convolutional neural network for ethnicity classification using ocular region in mobile environment. In: 2018 10th Computer Science and Electronic Engineering Conference, CEEC 2018 - Proceedings. IEEE, pp. 293–298. https://doi.org/10.1109/CEEC.2018.8674194.

[22] Qiu X., Sun Z., Tan T. (2005) Global Texture Analysis of Iris Images for Ethnic Classification. In: Zhang D., Jain A.K. (eds) Advances in Biometrics. ICB 2006. Lecture Notes in Computer Science, vol 3832. Springer, Berlin, Heidelberg. https://doi.org/10.1007/11608288_55

[23] Roomi, S. Virasundarii, S. Selvamegala, S. Jeevanandham, D. Hariharasudhan (2011). Deriving Soft Biometric Feature from Facial Images. Third National Conf. on Computer Vision, Pattern Recognition, Image Processing, and Graphics, Massachusetts Ave., NW Washington, DC United States, pp. 54–57.

[24] Thanh Vo, Trang Nguyen and C. T. Le. (2018). Race Recognition Using Deep Convolutional Neural Networks. *Symmetry* 2018, *10*, 564. https://doi.org/10.3390/sym10110564

[25] Vo T, Nguyen T, Le CT. (2018). Race Recognition Using Deep Convolutional Neural Networks. *Symmetry*; 10(11):564. https://doi.org/10.3390/sym10110564

[26] Viola, P. and Jones, M. (2001). Rapid object detection using a boosted cascade of simple features. Proceedings of the IEEE Computer Society Conference on Computer Vision and Pattern Recognition (CVPR), Kauai, HI, USA, 8–14.

[27] Gao, Wen, Cao, Bo and Shan, Shiguang and Chen, Xilin and Zhou, Delong and Zhang, Xiaohua and Zhao, Debin (2008). The CAS-PEAL Large-Scale Chinese Face Database and Baseline Evaluations. IEEE Transactions on Systems, Man, and Cybernetics-Part A: Systems and Humans, vol. 38, no. 1, pp. 149–161, Jan. 2008, doi: 10.1109/TSMCA.2007.909557.

[28] Wang, C., Zhang, Q., Liu, W., Liu, Y., & Miao, L. (2019). Facial feature discovery for ethnicity recognition. WIREs Data Mining Knowl Discov., p. 9, e1278. https://doi.org/10.1002/widm.127

[29] Y. Xie, K. Luu and M. Savvides (2012). A robust approach to facial ethnicity classification on large-scale face databases. *2012 IEEE Fifth International Conference on Biometrics: Theory, Applications and Systems (BTAS)* pp. 143-149, doi: 10.1109/BTAS.2012.6374569.

[30] Farfade, S. S., Saberian, M., & Li, L. J. (2015). Multi-view face detection using Deep convolutional neural networks. ICMR 2015 - Proceedings of the 2015 ACM International Conference on Multimedia Retrieval, 643–650. https://doi.org/10.1145/2671188.2749408