iJRASET
International Journal For Research in Applied Science and Engineering Technology



INTERNATIONAL JOURNAL FOR RESEARCH

IN APPLIED SCIENCE & ENGINEERING TECHNOLOGY

Volume: 12 **Issue:** III **Month of publication:** March 2024

DOI: https://doi.org/10.22214/ijraset.2024.57989

www.ijraset.com

Call: 08813907089 | E-mail ID: ijraset@gmail.com

International Journal for Research in Applied Science & Engineering Technology (IJRASET)
ISSN: 2321-9653; IC Value: 45.98; SJ Impact Factor: 7.538
Volume 12 Issue III Mar 2024- Available at www.ijraset.com

Classification Based on Ethnicity and Race: A Comprehensive Review

Shruti Mishra[1], Kashish Srivastava[2], Dr. Pankaj Kumar[3]
Computer Science and Engineering, SRMCEM, Lucknow, India

Abstract: ***Ethnicity classification, a crucial facet in computer vision and artificial intelligence, has witnessed significant advancements in recent years. This review paper provides a comprehensive analysis of the state-of-the-art techniques employed in ethnicity classification based on facial characteristics using deep learning methodologies. The focus is on leveraging convolutional neural networks (CNNs) to discern and classify diverse human races.***
This paper encapsulates the journey of the project, offering a deep-dive into the methodologies, datasets, and models employed for ethnicity classification using deep learning. The insights garnered contribute to the ongoing discourse in the field, paving the way for enhanced accuracy and inclusivity in facial recognition technologies.
Keywords: Ethnicity and race classification, face biometrics, soft biometrics, Convolutional neural networks (CNNs), VGG-16, ResNet50, MobileNet v2, VGG-Face, transfer learning.

I. INTRODUCTION

In today's globalized world, embracing diversity and understanding the rich tapestry of human ethnicity is not only a matter of cultural significance but also holds profound implications across various domains. From improving facial recognition technology to enhancing the accuracy of demographic studies, the ability to classify individuals by their ethnicity is a multifaceted endeavor with far reaching applications.

Our project embarks on a transformative journey into the realm of ethnicity and human race classification, harnessing the formidable power of cutting edge deep learning techniques. Through the lens of deep learning, we seek to unravel the complexities of ethnicity and bring forth a robust classification system that can accurately identify an individual's ethnicity based on facial characteristics.

The scope of our project is both comprehensive and enlightening. It begins with the exploration and curation of a diverse and extensive dataset, carefully selected to represent a broad spectrum of ethnicities from around the world. At the heart of our endeavor lies the development of state of the art deep learning models, primarily Convolutional Neural Networks (CNNs). These models are designed to extract and analyze intricate facial features, discerning the subtle nuances that distinguish individuals across diverse ethnic backgrounds. With a focus on accuracy and inclusivity, we aim to create a classification system that transcends boundaries and prejudices..

In our quest for excellence, we draw upon the collective wisdom of established deep learning architectures, including VGG19, ResNet, Inception, and Xception. However, our approach is not one of complacency; instead, we embark on a rigorous journey of optimization. We conduct grid searches and hyper parameter tuning to ensure that our models achieve benchmark setting performance.

Beyond benchmarking, our project aspires to blaze new trails in the realm of ethnicity classification. We explore the possibilities of creating custom models tailored specifically to the challenges posed by this unique classification task. Our objective is clear: to develop a classification system that surpasses conventional standards and excels in capturing the intricate subtleties of ethnicity. Join us on this exciting journey as we leverage the capabilities of deep learning to advance the science of ethnicity and human race classification, fostering a world that values and appreciates the beauty of our differences.

II. DEEP LEARNING ARCHITECTURES AND EXISTING METHODS

A. CNN

A Convolutional Neural Network (CNN) is a type of deep learning algorithm that is particularly well suited for image recognition and processing tasks. It is made up of multiple layers, including convolutional layers, pooling layers, and fully connected layers.

The convolutional layers are the key component of a CNN, where filters are applied to the input image to extract features such as edges, textures, and shapes.

©IJRASET: All Rights are Reserved

The output of the convolutional layers is then passed through pooling layers, which are used to down sample the feature maps, reducing the spatial dimensions while retaining the most important information. The output of the pooling layers is then passed through one or more fully connected layers, which are used to make a prediction or classify the image.

CNNs are trained using a large dataset of labeled images, where the network learns to recognize patterns and features that are associated with specific objects or classes. Once trained, a CNN can be used to classify new images, or extract features for use in other applications such as object detection or image segmentation.

Convolutional neural networks are distinguished from other neural networks by their superior performance with image, speech, or audio signal inputs.

They have three main types of layers, which are:

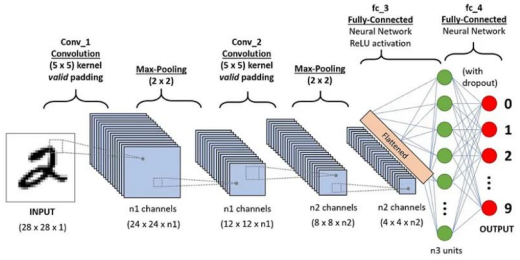
1) *Convolutional Layer:* This layer is the first layer that is used to extract the various features from the input images. In this layer, We use a filter or Kernel method to extract features from the input image.

$$\frac{W - F + 2P}{S} + 1$$

2) *Pooling Layer:* The primary aim of this layer is to decrease the size of the convolved feature map to reduce computational costs. This is performed by decreasing the connections between layers and independently operating on each feature map. We have Max pooling and average pooling.

$$\frac{W - F}{S} + 1$$

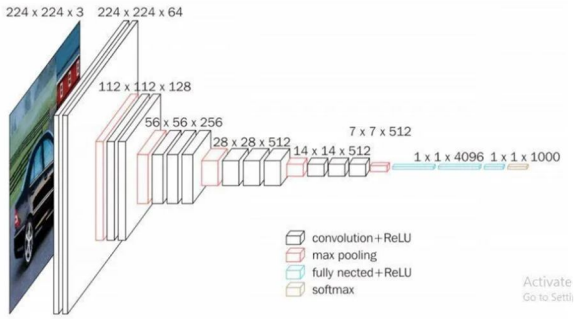
3) *Fully-connected (FC) Layer:* The Fully Connected (FC) layer consists of the weights and biases along with the neurons and is used to connect the neurons between two different layers. These layers are usually placed before the output layer and form the last few layers of a CNN Architecture.



B. VGG16

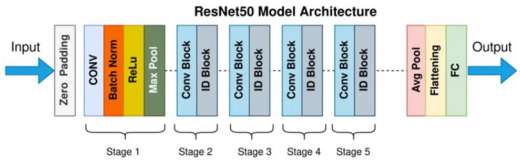
VGG16 is one of the most experimented CNN architectures for facial soft biometrics analysis. It achieved a significant success thanks to its shallow architecture (around 130K parameters, 13 convolutional layers and 3 fully connected layers), that allows to better generalize even in presence of small training sets. It achieves state-of-the-art age estimation accuracy and it is not a case, since there are not very large datasets for training deep networks for age estimation. Being one of the most popular CNNs, we include it in our performance analysis. As evident from the name, it consists of 16 levels and requires an input of 224×224224×224 pixels.

VGGFace is VGG16 trained from scratch for face recognition on almost 1,000,000 images. This CNN is probably the most adopted architecture for facial soft biometrics analysis. Indeed, the availability of weights pretrained on a very large number of face images and not for general image classification task (ImageNet) makes it very suited for transfer learning. VGGFace achieved an impressive accuracy in face recognition and age estimation and it has been successfully experimented also for gender recognition; for this reason, we believe it can be effective also for ethnicity recognition purposes.



C. ResNet50

ResNet50 is one of the residual networks proposed by the Microsoft research group, which won the ILSVRC and COCO 2015 competitions. The most important feature of such architecture is the introduction of the residual blocks, which allow ResNet to require less processing time for training and less extra parameters for increasing the depth of the network. Consequently, various versions of ResNet have been proposed with increasing number of layers (18, 34, 50, 101, 152). However, it has been demonstrated that ResNet50 is very effective for other facial soft biometrics analysis, namely age group classification and emotion recognition. For this reason, in this paper we use its version with 50 layers, which takes as input an image of 224×224224×224 pixels.



D. MobileNet v2

MobileNet v2 is one of the MobileNets architectures, very suited for mobile and embedded vision applications; think, as an example, to a cognitive robot or a smart camera that is able to perform face analysis in real time with a good accuracy even using normal CPUs. The software optimization which allows this network to significantly reduce the processing time is the transformation of the convolutional layers in depth wise and point wise operations, without paying a lot in terms of accuracy. In our opinion, this network architecture can be useful for realtime ethnicity recognition applications running on low cost devices. In this paper, we use the most popular v2 version, which consists of 17 layers and that, in its original version trained with ImageNet, requires an input of 224×224224×224 pixels.

Input	Operator	t	c	n	s
$224^2 \times 3$	conv2d	-	32	1	2
$112^2 \times 32$	bottleneck	1	16	1	1
$112^2 \times 16$	bottleneck	6	24	2	2
$56^2 \times 24$	bottleneck	6	32	3	2
$28^2 \times 32$	bottleneck	6	64	4	2
$14^2 \times 64$	bottleneck	6	96	3	1
$14^2 \times 96$	bottleneck	6	160	3	2
$7^2 \times 160$	bottleneck	6	320	1	1
$7^2 \times 320$	conv2d 1x1	-	1280	1	1
$7^2 \times 1280$	avgpool 7x7	-	-	1	-
$1 \times 1 \times 1280$	conv2d 1x1	-	k	-	

MobileNetV2 Overall Architecture

E. Transfer Learning

Transfer learning is a technique in machine learning where a model trained on one task is used as the starting point for a model on a second task. This can be useful when the second task is similar to the first task, or when there is limited data available for the second task. By using the learned features from the first task as a starting point, the model can learn more quickly and effectively on the second task. This can also help to prevent over fitting, as the model will have already learned general features that are likely to be useful in the second task.

III. LITERATURE REVIEW

Recent trends in race-based facial recognition have been proposed. Many plans focus on full-face images during design. This method uses mathematical techniques to extract individual features (Al-Humaidan and Prince, 2021). For example, Lu et al. have proposed a type of ethnic classification that examines facets at various scales (Lu, Jain et al., 2004). This method applies discriminant analysis (LDA) to facial images to improve classification, utilizing 2630 face models from 263 subjects. Although the method showed 96.3% accuracy, the data was divided into two groups (Asian and non-Asian).

Another model reported by Guo et al. relies on canonical correlation analysis (CCA) techniques to report race, gender, and age (Guo and Mu, 2014). Manesh et al. achieved one method using the golden ratio by applying a decision rule for the difference between different faces (Manesh, Ghahramani, and Tan, 2010). SVM technology then classifies the Gober results to be removed. Similar to Lu's method, this method splits the data into Asians and non-Asians.

Some methods identify race by skin or hair color (Demirkus, Garg and Güler, 2010). Two algorithms have been proposed for the classification of races: pixel density values and entertainment-based models (BIM) for facial color. The SVM classifier was used to classify 600 images into three ethnic groups (Asian, African, and Caucasian) and achieved an efficiency of 81.3%. Besserrilla et al. proposed a method that encodes information about different faces into appearance and geometric features, such as color and shape of the face (BecerraRiera et al., 2017; 2018). Both RF and SVM use a combination of the two techniques as classifiers. The model classified facial images in the FERET dataset into three groups: white, black, and Asian with 93.7% accuracy.

Wu et al. proposed a weekly review based on a backup method to extract demographic characteristics (such as race and gender) from faces (Wu, Ai, and Huang, 2004). This method consists of three modules: face area detection, face extraction, and population segmentation. The model was tested on a dataset with three groups: 1771 Africans, 2306 Caucasians, and 2411 Mongoloids. In the Mongoloid, Caucasian group, accuracy is 95.0%, 96.1%, 93.5%, and African. Muhammad et al. proposed an ethnic classification method based on two types of local descriptors: Weber Local Descriptor (WLD) and Local Binary Model (LBP) (Muhammad et al., 2012). The model also uses different variables such as Chi-square, Euclidean, and Cityblock. The results show that combining WLD with LBP and using city blocks as classifiers has better accuracy.

Some methods of solving race recognition problems focus on specific areas of the face. For example, Lyle et al. proposed a method to distinguish Asians from non-Asians by using LBP criteria to select the texture of the periocular region (Lyle et al., 2012). Sun et al. proposed classification methods based on hierarchical visual code books to classify iris images. Although this method provides good performance, its flaw is also obvious, as it is difficult to remove iris tissue due to the length of the camera for advice.

Xie et al. proposed a new ethnic classification system that combines relevant colored faces and seed categories as features to identify the ethnicity of data, including Asians, African Americans, and Caucasians (Xie, Luu, and Savvides, 2012). This method focuses on the upper part of the face, including the eye area. Hoos et al. proposed a new method combining retinal models and Gabor wavelet features for classification using support vector machine (SVM) techniques. This method manages to classify 93%, 94%, and 96% of Asians, Africans, and Europeans, respectively, but has difficulty distinguishing other races.

Rumi et al. proposed a face detection method using the Viola-Jones algorithm (Roomi et al., 2011). The authors focused on extracting features from different regions, including lips, skin color, and forehead regions. This method uses FERET, Yale's dataset of black, Mongolian, and Caucasian samples, and contains nine layers, the first six of which are convolutional layers. The last three layers are all layers. CFD and UTK data were used to evaluate the model, and the data was divided into four groups: Asian, Caucasian, American, and Indian. The model achieved a 97% accuracy rate.

Muhammad et al. proposed a new way to increase security in the mobile environment based on native biometric technology (Muhammad and Al-Ani, 2018). This study presents six different neural network models with different models for classifying five races: Middle Eastern, Asian, African, Caucasian, and Hispanic. Before classification, the eye region was extracted from the face image in the FERET dataset. Model-02 has the highest accuracy at 98.35%, while Model-5 has the lowest accuracy at 88.35%. Masood et al. utilized two strategies, namely artificial neural network and convolution neural network VGGNet, dividing FERET data into three races: Black, Caucasian, and Mongolian (Masood et al., 2018).

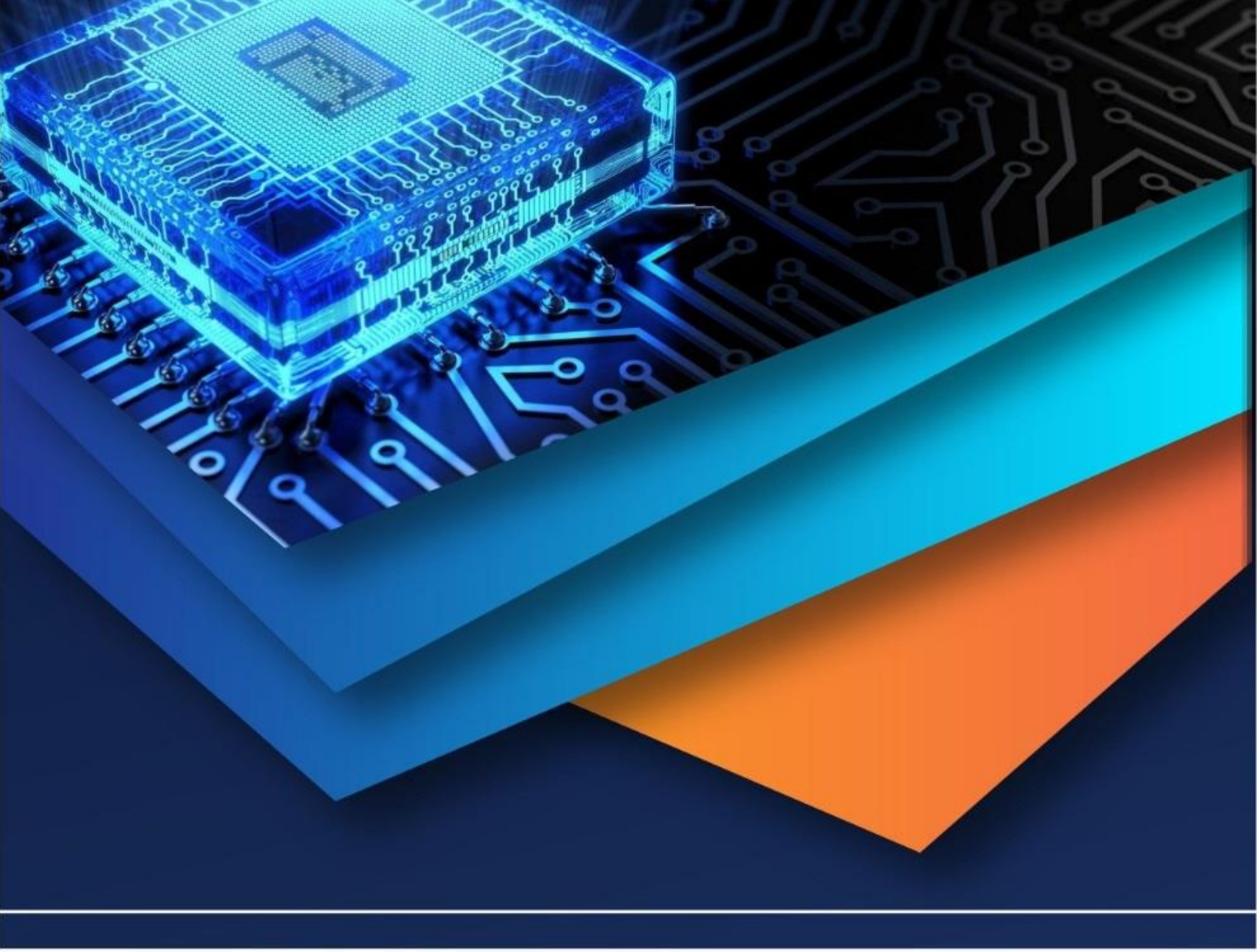
This method focuses on classifying color features and geometric features extracted from facial images. When ANN and CNN are used, the accuracy results are 82.4% and 98.6%, respectively.

IV. CONCLUSION

In conclusion, this review paper provides a comprehensive exploration of race and ethnicity classification methodologies, particularly focusing on advancements in deep learning techniques. The studies discussed underscore the complexities and challenges inherent in developing fair and accurate models for classifying individuals based on their racial and ethnic backgrounds. The surveyed literature highlights the importance of dataset diversity, model interpretability, and ethical considerations in mitigating biases and achieving more equitable outcomes. While significant strides have been made in leveraging Convolutional Neural Networks (CNNs) and other deep learning architectures, ongoing research is imperative to address the nuanced intricacies of race and ethnicity classification. As we navigate the evolving landscape of facial recognition technologies, it is crucial to remain cognizant of the ethical implications, biases, and societal impact of these systems. Future research directions may include refining model interpretability, addressing intersectional disparities, and establishing standardized practices to enhance the fairness and accuracy of race and ethnicity classification models. This synthesis of existing studies aims to contribute to a more nuanced understanding of the challenges and opportunities in this dynamic field, fostering continued advancements towards responsible and inclusive artificial intelligence applications.

REFERENCES

[1] Afifi, Mahmoud, and Abdelrahman Abdelhamed. 2019. “Afif4: Deep Gender Classification Based on Adaboost-Based Fusion of Isolated Facial Features and Foggy Faces.” Journal of Visual Communication and Image Representation 62:77–86.
[2] Al-Humaidan, Norah A., and Master Prince. 2021. “A Classification of Arab Ethnicity Based on Face Image Using Deep Learning Approach.” IEEE Access 9:50755–66.
[3] Baig, Talha Imtiaz, Talha Mahboob Alam, Tayaba Anjum, Sheraz Naseer, Abdul Wahab, Maria Imtiaz, and Muhammad Mehdi Raza. 2019.
[4] “Classification of Human Face: Asian and Non-Asian People.” Pp. 1–6 in 2019 International Conference on Innovative Computing (ICIC).
[5] Becerra-Riera, Fabiola, Nelson Méndez Llanes, Annette Morales-González, Heydi Méndez-Vázquez, and Massimo Tistarelli. 2018.
[6] “On Combining Face Local Appearance and Geometrical Features for Race Classification.” Pp. 567–74 in Iberoamerican Congress on Pattern Recognition. Springer Demirkus, Meltem, Kshitiz Garg, and Sadiye Guler. 2010.
[7] “Automated Person Categorization for Video Surveillance Using Soft Biometrics.” Pp. 236–47 in Biometric Technology for Human Identification VII. Vol. 7667. SPIE. Greco, Antonio, Gennaro Percannella, Mario Vento, and Vincenzo Vigilante. 2020.
[8] “Benchmarking Deep Network Architectures for Ethnicity Recognition Using a New Large Face Dataset.” Machine Vision and Applications 31(7):1–13.
[9] Guo, Guodong, and Guowang Mu. 2014. “A Framework for Joint Estimation of Age, Gender and Ethnicity on a Large Database.” Image and Vision Computing 32(10):761–70.
[10] Heng, Zhao, Manandhar Dipu, and Kim-Hui Yap. 2018. “Hybrid Supervised Deep Learning for Ethnicity Classification Using Face Images.” Pp. 1–5 in 2018 IEEE International Symposium on Circuits and Systems (ISCAS).
[11] Khan, Sajid Ali, Maqsood Ahmad, Muhammad Nazir, and Naveed Riaz. 2013. “A Comparative Analysis of Gender Classification Techniques.” International Journal of Bio-Science and Bio-Technology 5(4):223–43.
[12] Lu, Xiaoguang, Anil K. Jain, and others. 2004. “Ethnicity Identification from Face Images.” Pp. 114–23 in Proceedings of SPIE. Vol. 5404.
[13] Lyle, J. R., P. E. Miller, S. J. Pundlik, and D. L. Woodard. 2012. “Soft Biometric Classification Using Periocular Region Features.” Pattern Recognition 45(11):3877–3885. doi: 10.1016/j.patcog.2012.04.027.
[14] Manesh, F. Saei, Mohammad Ghahramani, and Yap-Peng Tan. 2010. “Facial Part Displacement Effect on Template-Based Gender and Ethnicity Classification.” Pp. 1644–49 in 2010 11th international conference on control automation robotics & vision.
[15] Marzoog, Zahraa Shahad, Ashraf Dhannon Hasan, and Hawraa Hassan Abbas. 2022. “Gender and Race Classification Using Geodesic Distance Measurement.” Indonesian Journal of Electrical Engineering and Computer Science 27(2):820–31.
[16] Masood, Sarfaraz, Shubham Gupta, Abdul Wajid, Suhani Gupta, and Musheer Ahmed. 2018. “Prediction of Human Ethnicity from Facial Images Using Neural Networks.” Pp. 217–26 in Data Engineering and Intelligent Computing. Springer.
[17] Mohammad, Ahmad Saeed, and Jabir Alshehabi Al-Ani. 2018. “Convolutional Neural Network for Ethnicity Classification Using Ocular Region in Mobile Environment.” Pp. 293–98 in 2018 10th Computer Science and Electronic Engineering (CEEC). IEEE. Muhammad, Ghulam, Muhammad Hussain, Fatmah Alenezy, George Bebis, Anwar



10.22214/IJRASET

45.98

IMPACT FACTOR: 7.129

IMPACT FACTOR: 7.429

INTERNATIONAL JOURNAL FOR RESEARCH

IN APPLIED SCIENCE & ENGINEERING TECHNOLOGY

Call : 08813907089 (24*7 Support on Whatsapp)