Face Sketch Recognition: Ethnic Groups Classification and Recognition Via a VGG16 Model Approach

Khalid OUNACHAD[1], Mohamed EL GHMARY[2]
USMS, Sultan Moulay Slimane University, Beni Mellal, Morocco[1]
ENSAM, Mohammed V University, Rabat, Morocco[2]

Abstract—**In the law enforcement investigation, the police use sketching techniques to identify suspects from an eyewitness's memory. Many automatic face sketch recognition systems that determine the perpetrator's appearance from the face image datasets have been proposed. The aim is to conduct the arrest of the right offender. We propose this work to carry out a search based on the ethnicity criterion to speed up this automatic identification and to help authorities execute fast responses by launching the retrieval process only in a part of the dataset of face images. The goal of this study is to enhance the accuracy of ethnic face sketch classification by using the convolutional neural network built on the VGG16 architecture. The FairFace dataset, which includes seven ethnic face images: White, Black, Indian, East Asian, Southeast Asian, Middle Eastern, and Latino|Hispanic, was employed in the study. We convert the face images dataset to face sketch images, and we optimize the VGG16 model for seven classification outputs. This work shows that the VGG16 deep learning model offers a reliable, automated approach for ethnic face sketch classification and recognition. The used model achieved an accuracy reaching above 94% and produced a low false negative rate, which is crucial for minimizing undetected cases.**

Keywords—Forensic face sketch; ethnic classification; deep learning; transfer learning; optimized VGG16 model

I. Introduction

Race is basically defined as a mixture of physical, behavioral, and cultural attributes, but ethnicity recognizes differences between people mostly based on language and shared culture [1]. In other words, race is inherent in our biology, and therefore inherited across generations, while ethnicity is based on native region or culture and acquired; it can be tied to the self-ascribed.

VGG16 (Visual Geometry Group-16 deep layers) is a deep convolutional neural network model introduced by the team at the University of Oxford [2]. It is most famous for its architectural simplicity and its performance, especially in image classification, but it needs many calculations, and it's a slow algorithm without a GPU. Among its applications, in addition to image classification, are feature extraction, object detection, and image segmentation. Generally, VGG16 is one of the most experimented CNN architectures for facial soft biometrics analysis. It achieved a significant success thanks to its deep architecture (around 138M parameters, 13 convolutional layers, and 3 fully connected layers), which allows it to better generate results even in the presence of small training sets.

Human Face Ethnic Groups Identification and Recognition (FEGIR) is an area of non-verbal language in day-to-day life communication. Ethnic facial features are some of the most important face features, but humans struggle and find it hard to try to differentiate between members of social outgroups, including different races [3]. Moreover, FEGIR systems have attracted numerous researchers and have been conducted and utilized in various fields since they attempt to overcome the problems and factors weakening these systems, including the problem of image classification, also due to their large-scale applications in face analysis, particularly face recognition [4]; sharpening the global focus on ethnicity and race in the time of COVID-19 disease [5],[6], career progression analysis [7], disentangling the stress process [8], and social crime [9].

Ethnic-based separation among humans is classified into six minimum categories: American Indian or Alaska Native, Asian, Black or African American, Hispanic or Latino, Native Hawaiian or Other Pacific Islander, and White [7]. However, based on their skin colors, we have five kinds of ethnicity:

red, white, black, brown, and yellow. It can also be tied up to some specific geographic regions, countries, or nationalities: symbolized relationships between the "races" and the countries. The U.S. Census Bureau also considers five race categories: White, Black or African American, American Indian or Alaska Native, Asian Native, and Hawaiian or Other Pacific Islander. The census considers some regions of origin for each race category. We use seven ethnic kinds of faces: White, Black, Indian, East Asian, Southeast Asian, Middle Eastern, and Latino, referring to the used dataset of face sketch images.

Ethnic group identification and recognition systems aim to automatically classify ethnicity in a dataset of photos or sketches (Fig. 1). It is based on a two-dimensional image of human subjects. The automatic task of ethnicity recognition is challenging work and explicitly difficult: Ethnic face classification and recognition can be done in many ways, but this study is concerned with the ethnic face sketch [10],[11] classification and recognition based on two-dimensional images of people's sketched faces.

Fig. 1. Examples of ethnic groups face sketch classification and recognition.

There are a large number of datasets available for human ethnicity classification research, some of which are private and some public. The FairFace database [10] is most commonly used in ethnicity classification, but until the redaction of this article, there was no face sketch dataset labeled with ethnic features used or cited in any related research. To resolve this situation and this problem, we created our ethnic face sketch dataset by converting face images of the FairFace dataset to face sketch images based on a method encompassing image inversion, Gaussian blur, and dodging fusion techniques.

VGGFace achieved an impressive accuracy in face recognition [12] and age estimation [13],[14] and has been successfully experimented with also for gender recognition [15],[16]. For this reason, we believe it can be effective for ethnicity recognition and classification purposes. In our proposed approach, we optimized the VGG16 model for seven classification outputs relative to our used face sketch dataset since it contains seven ethnic classes. The results of this work show that deep learning models, including VGG16, can offer a strong and automatic means of face sketch ethnicity classification and recognition. Beyond their intellectual value, these findings have practical ramifications for authorities in law enforcement investigation since they allow faster and more consistent ethnicity recognition and automatic face sketch recognition [17],[18].

The rest of the study is organized as follows: In Section II, related studies, an overview of the proposed VGG16 model is presented, emphasizing different modern searches when it's used. Many areas are mentioned, like computer vision, health disease classification, agriculture disease classification, and the civil engineering & construction sector. In Section III We present a brief explanation of the VGG16 models. In Section IV, we explore our approach in detail and describe the performance of our proposed algorithm. Section V presents the tests and experiments carried out during this study. Finally, we summarize this article with a conclusion.

II. Related Works

Ethnicity classification, a crucial facet in computer vision and artificial intelligence, has witnessed significant advancements in recent years. VGG16 is a VGGFace model 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, and we believe it can also be effective for ethnicity recognition and classification purposes. There has been a wealth of research on the VGG16 model, as well as on its application in areas such as computer vision. Specifically, it's used in various subfields such as image recognition
[19]-[21], health disease classification [22]-[24], agriculture disease classification or food security [25]-[26], and the civil engineering & construction sector [27], the last two areas were cited to indicate the performance of the use of VGG16 and considered as a related works to the use of the VGG16 model.:

In the image recognition subfield, for example, the study suggested by Zhu & Li [19] offers an improved version of VGG16 combined with wavelet analysis and marginal Fisher analysis (dimensionality reduction) to improve image recognition on different animal images. Their approach recognition accuracy reached 97%. Although in [20], R. Sewada and H. Goyal propose an adaptation of VGG-16 for the classification of multispectral satellite images. The used images are visible and near-infrared images. Their proposed modifications to the VGG-16 approach attain a higher classification accuracy rate of ±96.7%. While Genç & Yalman [21] propose an approach to recognize aircraft using satellite images based on CNN models. They detail a comparison study of VGG16 and VGG19 architectures. Based on their obtained results, VGG16 achieves an accuracy of approximately 82.67%, and VGG19 reaches ~89.29%.

However, in the medical domain, specifically in the health disease classification subfield, the authors of [22] propose a hybrid approach to detect heart disease by combining deep learning, machine learning, and explainable artificial intelligence. The explainable artificial intelligence techniques allow for the identification of the most decisive clinical variables in predicting cardiac risk. They use the pre-trained VGG16 model to extract 14 features from medical data converted to image format. These features are then merged with the original clinical data and classified using several machine learning algorithms: SVM, Random Forest, K-NN, etc. The VGG16 combined with the random forest model achieves the best performance, with an accuracy of approximately 92%. All the heart disease datasets used for their study are from Kaggle, comprising publicly available and synthetic data samples. In another health disease classification case [23], F. Tanjim and S. Hamdy describe and detail a method based on learning transfer with VGG16 to detect breast cancer from histopathological images. The authors use the pre-trained VGG16 model and adapt it to the BreaKHis medical dataset to distinguish between benign and malignant tumors. They improve the number of images using modern techniques such as data augmentation, dynamic tuning of the learning rate, and data regularization. They also propose a modified version of the pre-trained VGG16 model, which they named M-VGG16. The results show that the M-VGG16 model achieves an accuracy of approximately 94%. Although, in study [24], a retinal disease detection case, Jaimes et al. offer an approach for identifying four categories of retinal diseases from optical coherence tomography (OCT) images from the Kaggle database using a VGG16 transfer learning-based model. It relies on the pre-trained VGG16 model, adapted and refined for this specific task. The results achieve an accuracy of approximately 95%. The authors also use the Grad-CAM technique to visualize areas of the image that influence the prediction, thus improving the interpretability of the model.

Also, in the agriculture disease classification subfield related to the food security subdomain, in [25], the authors propose a CNN approach to improve the classification of tomato leaf diseases by adapting the classic VGG16 architecture to this agricultural domain and by optimizing its hyperparameters using Bayesian optimization. The innovative aspect of the used technique lies in the use of Bayesian optimization to automatically select the best hyperparameters for the model. The experiments are conducted on the Tomato Leaf Disease Detection (TLDD) dataset, a dataset of images of healthy and diseased tomato leaves, which is publicly available on the Kaggle platform. The results show that the proposed model achieves approximately 97% accuracy, surpassing the standard version of VGG16 (89.0%). The authors improve and modify the original VGG16 architecture by adjusting certain layers and adding regularization mechanisms to adapt it for plant disease detection. Although in a second agricultural classification disease case [26], Sahputra et al. compare two deep learning models, VGG16 and MobileNet, for classifying rice leaf diseases such as bacterial leaf blight, brown spot, leaf blast, and narrow brown spot. They use a set of images of infected and healthy leaves by using a dataset of 2,190 rice leaf images divided into five classes, apply preprocessing and data augmentation, and then evaluate the two models based on accuracy, speed, and complexity. The results show that VGG16 achieves superior accuracy, offering better classification performance. In contrast, MobileNet is lighter and faster. The VGG16 model achieved an accuracy of 98%, while the MobileNet model reached 95% accuracy.

In the civil engineering & construction sector, one example related to urban construction and infrastructure is the study carried out by Chen et al. [27]; it is a smart solution for structural health monitoring of buildings and dwellings. The authors propose a steel-reinforced concrete corrosion crack detection method based on improved VGG16. They propose a deep learning approach for detecting corrosion cracks in reinforced concrete using an improved VGG16 model. They adapt the VGG16 architecture to better capture fine crack details in images of concrete structures, while simultaneously reducing computational complexity. Chen et al. use the SDNET2018 and DeepCrack concrete crack datasets for training and testing experiments, and their proposed model achieved a precision of 94.4%.

As a contribution, this article proposes an ethnic face sketch classification and recognition method based on an improved VGG16 model. Our work is inspired in part by the recent and successful method that has shown that the use and integration of the VGG16 model in classification scenarios leads to obtaining a better score; it could be used to perform well. We believe it can also be effective for ethnicity recognition and classification purposes. As part of our research project, in which a face sketch recognition system was developed. We achieved an impressive accuracy in the face sketch recognition task [17], face sketch gender recognition task [11],[18],[28],[29], face sketch age estimation and recognition task [30], and face sketch emotions classification and recognition task [31]-[33]. What remains is the development of a phase of the overall project that allows for face sketch ethnicity recognition. For this reason, our project aspires to blaze new trails in the realm of face sketch ethnicity classification. We explore the possibilities of creating custom models tailored specifically to the challenges posed by this complex classification task. Our objective is to develop a classification system that surpasses conventional standards and excels in capturing the intricate subtleties of ethnicity. 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.

III. VGG16 Deep Convolutional Neural Network Model

VGG16 is a well-known pre-trained CNN model from ImageNet that consists of 16 layers with learned weights, mainly using 3×3 convolutions and 2×2 max pooling to extract and reduce image features. It's trained from scratch for face recognition on almost a million images from the ImageNet dataset. This pre-trained CNN model is one of the most experimented-with CNN architectures for facial soft biometrics analysis, and it is very effective at extracting useful features from images.

The diagram of Fig. 2 illustrates multiple layers of interconnected neurons (nodes), which allow the network to learn complex, hierarchical features from data. The different sections, indicated by varying colors and structures, include multiple and different layers: 2D convolution layers, 2D max pooling, flatten, and dense. We have already explained the main function and task of convolutional layers and max pooling in section 2.1, while the primary role of dense flatten is to convert the 3D output from the VGG16 base to 1D so that it can be used in fully connected layers, and the dense layers are the fully connected layers. As evident from the name, it consists of 16 levels and requires an input of 224×224 pixels.

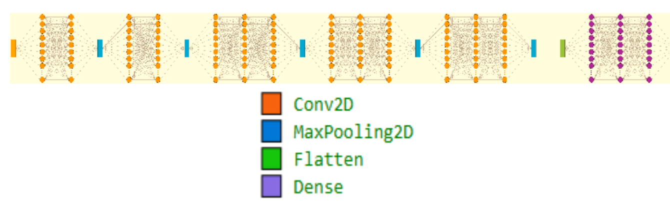

Fig. 2. VGG16 architecture scheme.

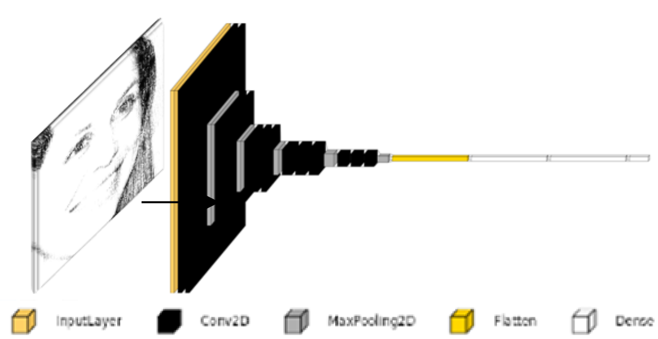

Fig. 3. VGG16 simple architecture 3D scheme.

Fig. 3, thus, clearly highlights the network's progression, in particular: gradual reduction of spatial size, increase in feature depth, and finally, classification via the fully connected layers. The architecture follows a very regular pattern: consistent filter sizes, double or triple convolution per block, then pooling, which simplifies optimization while allowing the extraction of increasingly abstract visual features. The VGG-16 architecture begins with an input layer receiving images of size 224×224×3, and during the progression, outputs a unique prediction, allowing classification among 1000 outputs. Fig. 3 clearly illustrates the model's ability to effectively reduce the dimensions of the input image before flattening it into a simple final vector.

The number of used parameters in each layer and for each block is detailed in Table I; in addition to the type of layer, it indicates the output shape obtained at the end of each layer. The number of parameters is zero for the input layer, all the MaxPooling2D layers used, and the flatten layer. The model uses 1 792 in the first 2D convolution layer and 4097000 in the last one. 102 764 544 parameters are used in the first dense layer; it’s the biggest number of parameters used in the model. The total parameters of the model are 138357544 (527.79 MB), and the trainable parameters are 138 57544 (527.79 MB), so the model doesn’t have a non-trainable parameter (0.00 B).

TABLE I. TYPE LAYERS, OUTPUT SHAPE AND PARAMETERS OF VGG16 MODEL

Layer (type)	Output Shape	Parameters
input_layer (InputLayer)	(None, 224, 224, 3)	0
block1_conv1 (Conv2D)	(None, 224, 224, 64)	1 792
block1_conv2 (Conv2D)	(None, 224, 224, 64)	36 928
block1_pool (MaxPooling2D)	(None, 112, 112, 64)	0
block2_conv1 (Conv2D)	(None, 112, 112, 128)	73 856
block2_conv2 (Conv2D)	(None, 112, 112, 128)	147 584
block2_pool (MaxPooling2D)	(None, 56, 56, 128)	0
block3_conv1 (Conv2D)	(None, 56, 56, 256)	295 168
block3_conv2 (Conv2D)	(None, 56, 56, 256)	590 080
block3_conv3 (Conv2D)	(None, 56, 56, 256)	590 080
block3_pool (MaxPooling2D)	(None, 28, 28, 256)	0
block4_conv1 (Conv2D)	(None, 28, 28, 512)	1 180 160
block4_conv2 (Conv2D)	(None, 28, 28, 512)	2 359 808
block4_conv3 (Conv2D)	(None, 28, 28, 512)	2 359 808
block4_pool (MaxPooling2D)	(None, 14, 14, 512)	0
block5_conv1 (Conv2D)	(None, 14, 14, 512)	2 359 808
block5_conv2 (Conv2D)	(None, 14, 14, 512)	2 359 808
block5_conv3 (Conv2D)	(None, 14, 14, 512)	2 359 808
block5_pool (MaxPooling2D)	(None, 7, 7, 512)	0
flatten (Flatten)	(None, 25088)	0
fc1 (Dense)	(None, 4096)	102 764 544
fc2 (Dense)	(None, 4096)	16 781 312
predictions (Dense)	(None, 1000)	4 097 000

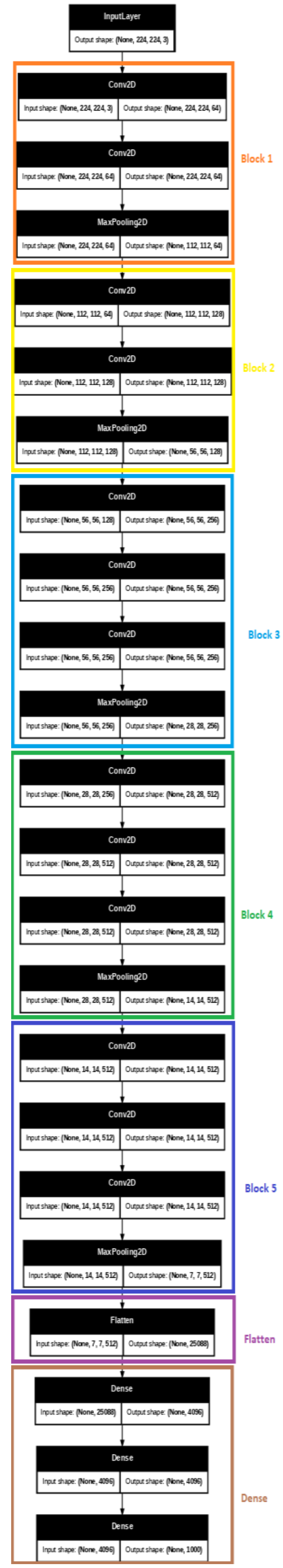

Fig. 4. VGG16 deep architecture scheme.

In Fig. 4, the VGG16 deep and detailed architecture is shown; it's composed of several successive blocks of convolutional layers followed by max-pooling layers. The first block applies two convolutions with 64 filters of size 3×3, maintaining the input resolution at 224×224, before reducing it to 112×112 via max-pooling. The second block repeats the same structure but with 128 filters, reducing the size to 56×56 after pooling. The third block consists of three convolutions with 256 filters each, followed by pooling that reduces the resolution to 28×28. The fourth block applies three convolutions of 512 filters before pooling, which reduces the size to 14×14. The fifth and final convolutional block applies three more layers of 512 filters, then a final max-pooling, resulting in a feature map of dimensions 7×7×512. The output of the convolutions is flattened into a vector of 25088 values, which feeds two fully connected layers of 4096 nodes each. The final dense layer has 1000 nodes and SoftMax activation, corresponding to 1000 possible output classes. The total of these output classes is chosen relative to the ImageNet dataset.

IV. Proposed Approach

The proposed framework for classifying ethnic groups from face sketches consists of several successive steps. First, images from a specialized dataset of face sketches are fed into the input system. This raw data then undergoes preprocessing, including automatic face detection, pixel intensity normalization, and image resizing to the format required by the VGG16 architecture (224×224×3).

The proposed system takes as input an image $I \in R^{H \times W \times 3}$ from the face sketches dataset. A facial detection operation $D(I)$ extracts the face region I_f, and the D operation encompasses exactly the HOG and SVM methods. The resulting image I_f is then normalized as in Eq. (1):

$$I_n = \frac{I_f}{255} \tag{1}$$

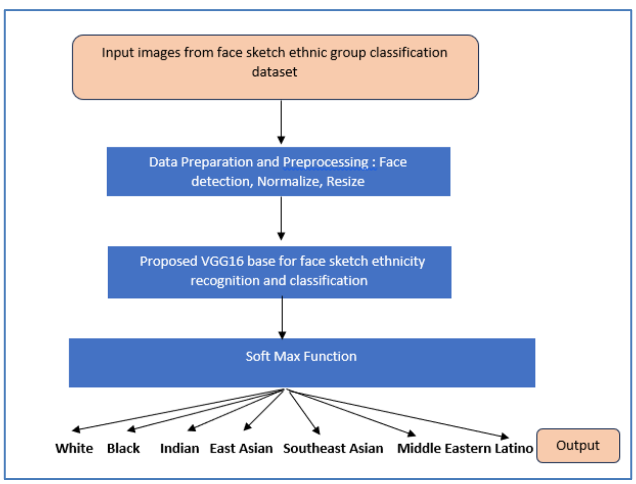

Fig. 5. Our proposed framework based on the VGG16 Model for face sketch ethnic group classification and recognition: An overview.

The normalization step converts pixel values to a 0-1 scale by dividing all pixel values by 255. Then, the resulting image I_n is resized to Ir through Eq. (2). The resize function resizes the image to 224x224 to conform to the VGG16 model shape.

$$I_r = Resize(I_n, 224 \times 224 \times 3) \tag{2}$$

In the feature extraction phase, a tailored VGG16 model is used as the basis network. This acts as a deep feature extractor, capable of efficiently capturing the facial structures present in the sketches. The pre-trained convolutional layers can be retained or fine-tuned to optimize performance on this type of non-photographic data.

The preprocessed image Ir is then propagated through a modified VGG16 network, consisting of 3×3 convolutional blocks followed by ReLU activation functions and max-pooling layers. Each block b performs a nonlinear transformation of the image described by Eq. (3):

$$F_l = \sigma(W_l * F_{l-1} + b_l) \tag{3}$$

where W_l is the convolutional filter of layer l, * is the convolution operation, and σ is the ReLU function.

After the last convolutional block, a flatten operation produces a feature vector $z \in R^d$. Finally, the representations generated by VGG16 are passed to a dense layer equipped with a SoftMax activation function, ensuring a probability distribution across all defined categories. The dense classifier then applies the following equation as in Eq. (4):

$$y = W_z + b \tag{4}$$

where $y \in R^K$ and K is the number of classes.

The final Softmax layer calculates a probability distribution as in Eq. (5):

$$P(y_i) = \frac{e^{y_i}}{\sum_{j=1}^{k} e^{y_j}} \tag{5}$$

For C ethnic groups, the predicted probability for class c is $P(y = c \mid F_l)$, calculated as mentioned in Eq. (5). The loss function L in Eq. (6) is categorical cross-entropy:

$$L = -\sum_{c=1}^{C} y_c \log \left(P(y = c|F_l) \right) \tag{6}$$

The predicted class corresponds to the result of the equation in Eq. (7):

$$\hat{c} = \mathrm{arg}max_i \;\; P(y_i) \tag{7}$$

The class with the highest probability is selected as the model output. An overview of our proposed framework based on the VGG16 model for face sketch ethnic group classification and recognition is shown in Fig. 5. This framework thus enables automatic and structured classification of face sketches according to the seven predefined categories in the used dataset.

The details of the proposed VGG16 base algorithm for face sketch ethnic group classification and recognition can be summarized as follows: Fig. 6. The pseudocode is also shown in Fig. 7.

Step1: A face sketch is input to the trained VGG16-based model (face sketch preprocessing)

Step 2: Deep feature extraction by convolutional layers (feature learning)

Step3: The classifier predicts the ethnic group label(classification)

Step4: The class with the highest softmax probability is selected (ethnic group recognition)

Fig. 6. Main stages of the proposed system.

Input: Face sketch dataset D = {(Is, C)}

Is=Face sketch image, C= Ethnic group class

Pre-trained VGG16 model

Output: Predicted ethnic group label ĉ

1- Load pre-trained VGG16 model

2- Remove fully connected layers

3-Freeze early convolutional layers

4- Fine-tune deeper layers on sketch dataset

5- For each sketch image Is in D do

Resize Is to 224×224

Normalize Is

Extract features with VGG16 Model

f = VGG16(Is)

end for

6- Train classifier Softmax using features f

7- For a test sketch Itest:

Extract feature ftest

Predict ethnic group ĉ

end for

8- Return ĉ

Fig. 7. Pseudo-code of the proposed algorithm: VGG16-based face sketch ethnic groups classification and recognition.

Each face sketch image is preprocessed by resizing to 224×224×3 and normalizing pixel values before being passed through the modified VGG16 network to extract high-dimensional deep feature vectors. Partial layer freezing allows for the preservation of generic descriptors. The extracted features are subsequently used to train the SoftMax classifier, which predicts the ethnic group label for test face sketches.

V. Experiments and Results

A. Dataset and Ethnicity Groups Taxonomy

To demonstrate the effectiveness of the proposed method, we proceeded to acquire a dataset of ethnic groups' face sketch pictures. We used the FairFace [34] face images dataset that we converted to face sketch images. It is used as a source of real face images to generate sketches using inversion, Gaussian blur, fusion, "dodging," and sketch-synthesis methods. In the context of ethnic face sketch classification and recognition, FairFace can serve as the best source of real face images from which sketches can be generated.

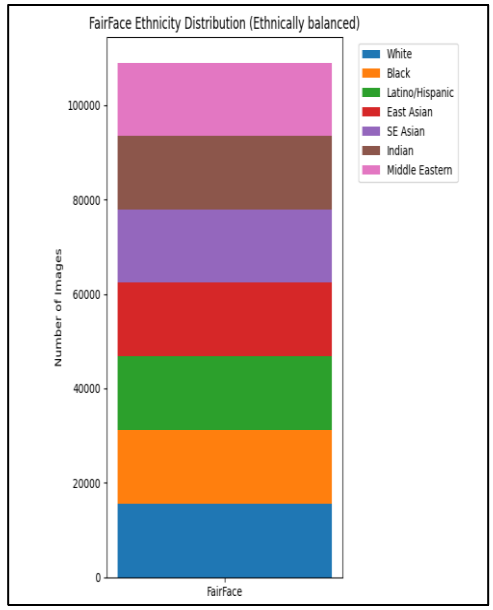

Fig. 8. FairFace ethnicity balanced distribution.

FairFace is a face image dataset that is ethnically balanced. It is a large-scale facial image dataset and contains 108501 images of faces in the wild, annotated with seven ethnicity categories [White, Black, Indian, East Asian, Southeast Asian, Middle Eastern, and Latino/Hispanic], gender labels (male and female), and age group ranges. Images were collected from the YFCC-100M Flickr dataset and labelled with race groups, gender, and age groups. FairFace [35],[36] is ethnically balanced, as shown in Fig. 8, ensuring that all groups are fairly represented, which makes it particularly suitable for training models for ethnicity group classification and recognition. Its images exhibit natural variations in pose, lighting, and background, providing realistic and perfect conditions for model training. Fig. 9 details the FairFace ethnic groups' class distribution, Table II. gives a summary dataset description, and Fig. 10 includes random samples from the same dataset, where each line in Fig. 10 represents and corresponds to an ethnic group. These ethnic groups are illustrated from the white ethnicity to the Latino| Hispanic ethnicity, suitable to order in Fig. 9.

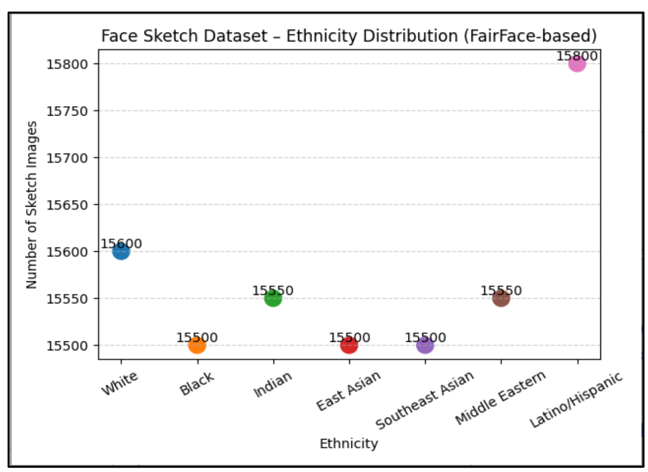

Fig. 9. FairFace ethnicity distribution.

TABLE II. FAIRFACE DATASET DESCRIPTION (ETHNICITY)

Number of images	Characteristics	Annotations	Ethnicity Classes	Ethnic groups Description	Number of Images (~)	% of Total
~108000	-Varied, mostly high-quality images - Contains in-the-wild images, (faces with natural conditions) -variations in pose, lighting, & background.	-Race/ Ethnicity: 7 categories -Gender: M|F -Age groups: 0–2, 3–9, 10–19, 20–29, 30–39, 40–49, 50–59, 60–69, 70+	-White -Black -Indian -East Asian -Southeast Asian -Middle Eastern -Latino / Hispanic	-Caucasian heritage -African descent -Indian subcontinent -East Asian heritage -Southeast Asian heritage -Middle Eastern heritage -Latin American heritage	15600 15500 15550 15500 15500 15550 15800	~14.4% ~14.3% ~14.3% ~14.3% ~14.3% ~14.3% ~14.6%

Fig. 10. Random samples from the FairFace dataset.

We divide the dataset into seven data subsets suitable for each kind of basic ethnicity: White, Black, Indian, East Asian, Southeast Asian, Middle Eastern, and Latino|Hispanic. Table II shows the percentage of each ethnic group; for each ethnicity class, the number of images is approximately 15500, with 14.4% of the average total. The reduction in the number of images is intended to simplify calculations, and the choice of images for the subclasses is arbitrary. Our chosen number of images in each ethnicity class is approximately seven thousand (~7000).

The generated face sketch ethnicity dataset from FairFace is used for training, validation, and testing our proposed approach. As shown in Table III., sixty percent (60%) of the images (29400) are used for the training phase, nineteen percent (20%) of the images (9800) are used for the validation task, while the remaining twenty percent (20%) of the images (9800) are reserved for the testing phase to assess the generalization performance of the proposed approach on unseen face sketch images. The dataset was divided proportionally to ensure that each ethnicity class was represented in all seven subsets. To achieve a complexity that is feasible on our machine, in our experimented and generated face sketch ethnicity dataset, the number of images in each ethnicity class is approximately seven thousand (~7000), 49000 images in total. Images were determined by data availability and the requirements of the VGG16 model training and testing process.

TABLE III. OUR DATASET SPLIT DISTRIBUTION ACROSS TRAINING, VALIDATION, AND TESTING PHASES

Phase task	Number of images	% of Total
Training	~29400	~60%
Validation	~9800	~20%
Testing	~9800	~20%

B. Tools

The proposed VGG16 model was developed, trained, and tested in Python using TensorFlow [39] to give access to pre-trained networks. Pandas was used for label file extraction and oversampling, CV2 for data preprocessing and facial recognition, and Scikit-learn for splitting data into training, validation, and testing sets. All computations were run in Google Colab using GPUs. The pre-trained VGG16 was downloaded via Keras through TensorFlow and subjected to transfer learning.

C. Image Pre-Processing

This stage prepares the data for the training process. Several steps are performed, including face detection, sketch image generation, normalization, and image resizing. We used the methods and operations detailed and mentioned in section IV to normalize and to resize to 224x224 all FairFace dataset images. We also convert all face images of the FairFace dataset to face sketch images by using a combination of inversion, Gaussian blur, and fusion image techniques. Fig. 11 shows a random sample of ethnic groups' face sketches generated and converted from FairFace by using these techniques. They include all ethnic classes:

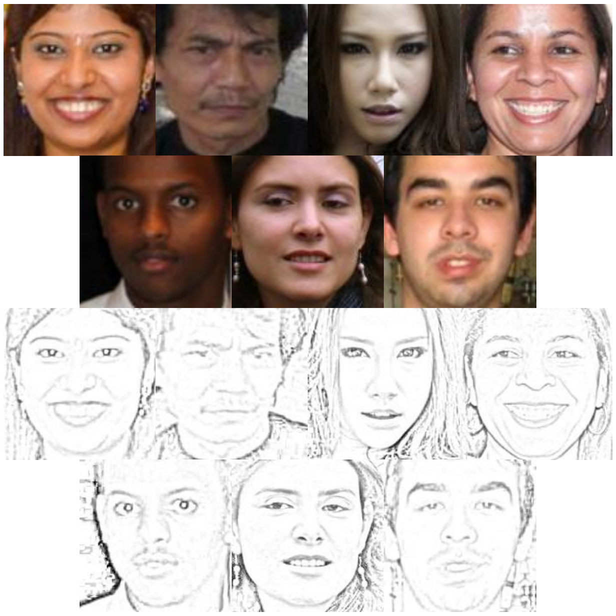

Fig. 11. Random samples of ethnic groups face sketches generated from the FairFace dataset images.

In our proposed approach, the first step in the preprocessing process is the face detection task. We use for this stage the famous Viola and Jones algorithm [40]-[42]. Fig. 12 illustrates some obtained results; the used method, based on the Viola-Jones technique, enables the faces to be detected perfectly.

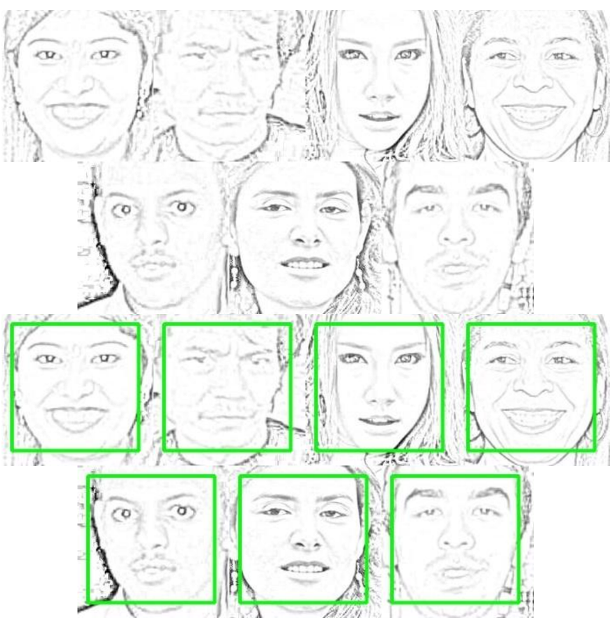

Fig. 12. Random Samples face sketch detection.

Fig. 13 summarize the experimental workflow for the data preparation and preprocessing task. The output image from the preprocessing phase will be the input image of the ethnic classifier based on the VGG16 model; it is an image containing only the detected face sketch area, which has been resized to a 224×224-pixel image and normalized.

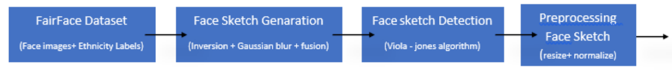

Fig. 13. Detailed experimental workflow for data preparation and preprocessing task.

D. Performance Metrics

When training a classifier, the choice of a scoring metric is a critical factor in achieving optimal classifier accuracy. The selection of the appropriate evaluation scale is of paramount importance to discriminate and ensure superior performance [37],[38]. Meanwhile, the model evaluation stage was carried out to assess the accuracy, consistency, and performance of the model in ethnic group face sketch classification and recognition. Model performance measurements used a confusion matrix and metrics such as accuracy (Eq. 8), precision (Eq. 9), recall (Eq. 10) and F1-score (Eq. 11). To compare model predictions with actual data, below are some equations for the proposed VGG16 model evaluation:

$$Accuracy = \frac{\mathrm{TP+TN}}{TP+TN+FP+FN} \quad (8)$$

$$Precision = \frac{\mathrm{TP}}{TP+FP} \quad (9)$$

$$Recall\ (sensitivity) = \frac{\mathrm{TP}}{TP+FN} \quad (10)$$

$$\mathrm{F1-Score} = 2 \times \frac{\mathrm{Precision} \times \mathrm{Recall}}{\mathrm{Precision+Recall}} \quad (11)$$

$$\mathrm{F}_{\beta} - \mathrm{Score} = (1 + \beta^2) \times \frac{\mathrm{Precision} \times \mathrm{Recall}}{\beta^2 \times \mathrm{Precision+Recall}} \quad (12)$$

The F_β-score in Eq. (12) is a generalized version of the classic F1-score, which allows more weight to be given to precision or recall, depending on the specific application. The score results will be the same as the classic F1-score, while ß is equal to one. While ß is greater than one, the focus is on recall; if less precision is needed and ß is less than one, more emphasis is placed on precision.

A True Positive (TP) occurs when the model predicts a positive output and the actual outcome is also positive, indicating a correct prediction for the positive class. Similarly, a True Negative (TN) occurs when the model predicts a negative output and the actual output is also negative, indicating a correct prediction for the negative class. Conversely, a False Positive (FP) occurs when the model incorrectly predicts a positive output when the actual output is negative. And finally, a false negative occurs when the model incorrectly predicts a negative output when the actual output is positive.

E. Results and Discussion

The generated dataset for this study consists of face sketches labeled by ethnicity. Seven classes were identified: White, Black, Indian, East Asian, Southeast Asian, Middle Eastern, and Latino|Hispanic. A total of 49000 face sketch images were randomly collected and generated from FairFace and divided into three subsets: training data, validation data, and test data. This proportional distribution ensured that each class was represented in all three subsets, thus optimizing the training, validation, and testing of the VGG16 model. The number of images used for each ethnic group class is balanced as detailed in Table IV below:

TABLE IV. FACE SKETCH IMAGE DATA DISTRIBUTION

Image Ethnicity Class	Training Data	Validation Data	Testing Data
White	~4200	~1400	~1'00
Black	~4200	~1400	~1400
Indian	~4200	~1400	~1400
East Asian	~4200	~1400	~1400
Southeast Asian	~4200	~1400	~1400
Middle Eastern	~4200	~1400	~1400
Latino|Hispanic	~4200	~1400	~1400
Total (7000)	~29400	~9800	~9800

Tests were conducted on 49000 data face sketch images distributed across seven classes of ethnic group face sketches: each ethnicity class contained approximately 7000 images. To evaluate the learning performance of the VGG16 model, we first present the performance of optimized VGG16-based ethnic groups classification on the face sketches generated from the FairFace dataset, and we present the accuracy and loss plots. The obtained results are shown in Fig. 14 and Fig. 15. These plots illustrate the model's ability to learn from the training and validation data, as well as the stability of the training process. The accuracy, precision, recall, $F_{\beta=1}$-score $F_{\beta=2}$- score and $F_{\beta=0.5}$-score for the face sketch ethnic group classification and recognition using our adopted approach are shown in Table V.

TABLE V. PERFORMANCE OF OPTIMIZED VGG16-BASED ETHNIC GROUPS CLASSIFICATION ON THE FACE SKETCHES GENERATED FROM THE FAIRFACE DATASET

Metric / Model	Accuracy	Precision	Recall	$F_{\beta=1}$-score	$F_{\beta=2}$ score	$F_{\beta=0.5}$ - score
Adopted VGG16 (Proposed)	94.00	93.80	94.60	94.70	34.64	93.62

The optimized VGG16 model accuracy graph (Fig. 14) shows a consistent increase in accuracy since the beginning of training, reaching above ~95% at the 100th epoch. The training and validation data accuracy patterns are similar, with no significant differences. This indicates that the model has good generalization capabilities and does not experience overfitting. The consistency between training and validation accuracy also indicates that the model is able to learn effectively from the data provided.

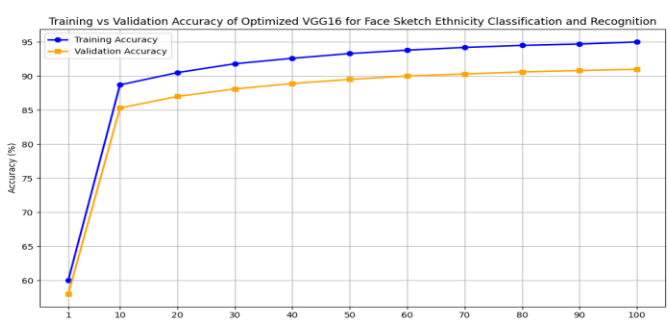

Fig. 14. Optimized VGG16 accuracy graph.

Meanwhile, the loss plot of the optimized VGG16 model(Fig. 15) shows a sharp decrease during the early epochs for both the training and validation data. Subsequently, the loss tends to stabilize at a low value (close to ~0), despite slight fluctuations in the validation loss at certain epochs. However, these fluctuations remain small and within acceptable limits. This behavior confirms that the training is proceeding correctly and that the model does not exhibit overfitting or underfitting problems.

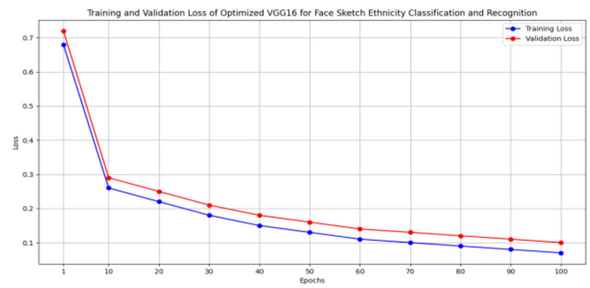

Fig. 15. Optimized VGG16 loss graph.

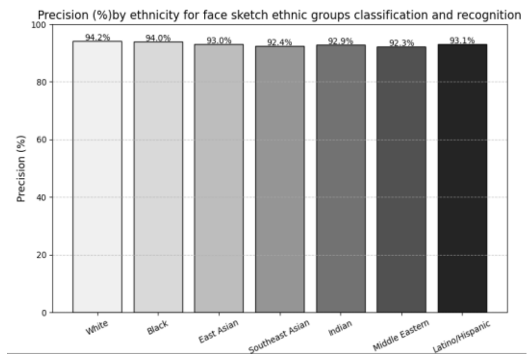

Fig. 16. Class-wise precision (%) for face sketch ethnicity classification and recognition (Optimized VGG16).

Once the model is trained, it is tested on the test data. The results of this test are presented in the visualization in Fig. 16. The figure illustrates the class-wise precision performance for face-sketch-based ethnicity classification and recognition across seven basic ethnic groups. The results show clearly high precision values for all ethnicity classes, ranging from 92.3% for Middle Eastern to 94.2% for the White ethnic group, indicating strong discriminatory capability of the optimized VGG16 proposed model. The highest precision is achieved for the White (94.2%) and Black (94.00%) ethnic groups, while slightly lower performances are obtained for Middle Eastern (92.3%) and Southeast Asian (92.4%) ethnic groups. Overall, the balanced distribution of precision scores across ethnicities demonstrates the robustness and fairness of the optimized VGG16 proposed model for face sketch ethnicity classification and recognition. We deduce that class-wise precision (%) for face sketches, ethnicity classification, and recognition using an optimized VGG16 model shows consistently high performance across all seven basic ethnic groups.

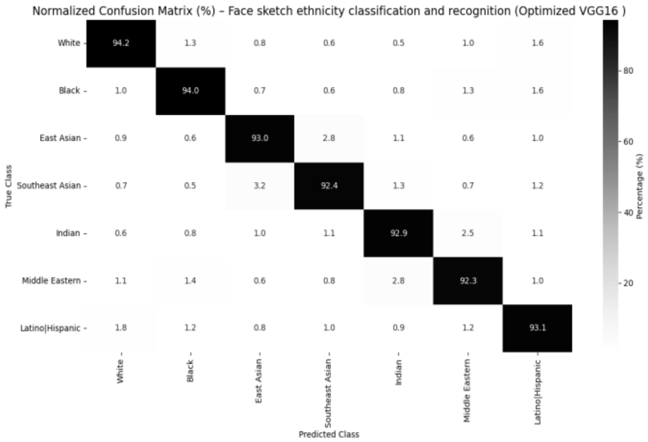

Fig. 17. Optimized VGG16 confusion matrix (%) in face sketch-based ethnicity classification and recognition.

Fig. 17 presents the normalized confusion matrix (%) for face sketch ethnic group classification and recognition using the optimized VGG16 model. The matrix brings to the fore strong diagonal dominance, indicating high correct classification rates across all seven basic ethnic groups. Precision values on the main diagonal exceed 92% for all classes, demonstrating the effectiveness of the proposed approach: White (94.2%), Black (94.0%), East Asian (93.0%), Southeast Asian (92.4%), Indian (92.9%), Middle Eastern (92.3%), and Latino|Hispanic (93.1%). Minor misclassifications are observed between visually similar ethnic groups, such as East Asian and Southeast Asian, as well as Middle Eastern and Indian.

Fig. 18 also illustrates the inter-class confusion matrix for face sketch–based ethnicity classification and recognition using the optimized VGG16 model. The matrix emphasizes misclassification patterns between ethnic groups (distribution of misclassification errors). Overall, the confusion rates remain low, indicating strong class separability.

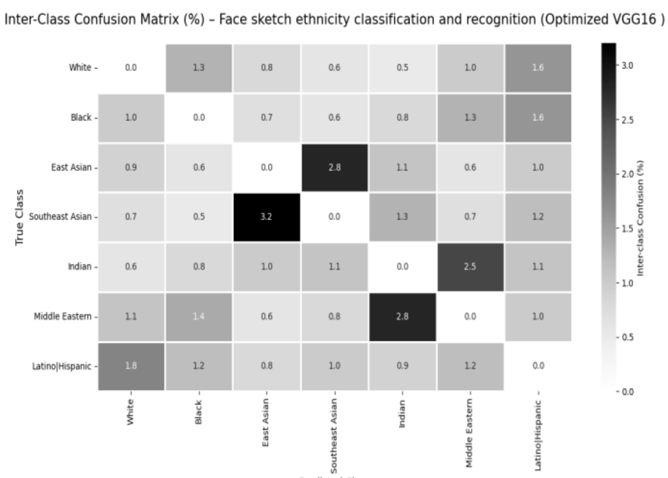

Fig. 18. Inter-class confusion matrix (%) for face sketch ethnic group classification and recognition using optimized VGG16.

The most noticeable and slight confusions are observed between visually related ethnic groups, such as between East Asian and Southeast Asian ethnic classes, which can be attributed to shared facial sketch features and characteristics. Similarly, moderate confusion is observed between Indian and Middle Eastern ethnicity classes, reflecting visual similarities in facial characteristics commonly found in sketch representations. Especially, these confusions are symmetrical and very low in magnitude, proving that the model does not exhibit systematic bias toward any particular ethnic group.

For example, VGG16 incorrectly classified some cases shown in Fig. 19. The first case is classified as Indian with 54% precision, so in fact the real class ethnic group is Southeast Asian. The prediction of the second case is East Asian with 57% precision; however, the real ethnicity is Southeast Asian. For the third case, in which the real class is East Asian, its prediction is 60% Southeast Asian. The last case is classified as Indian with 55% precision, so in reality, it is a Middle Eastern face sketch. This indicates that face sketches exhibit similar spot patterns or discoloration, making them difficult for the model to distinguish.

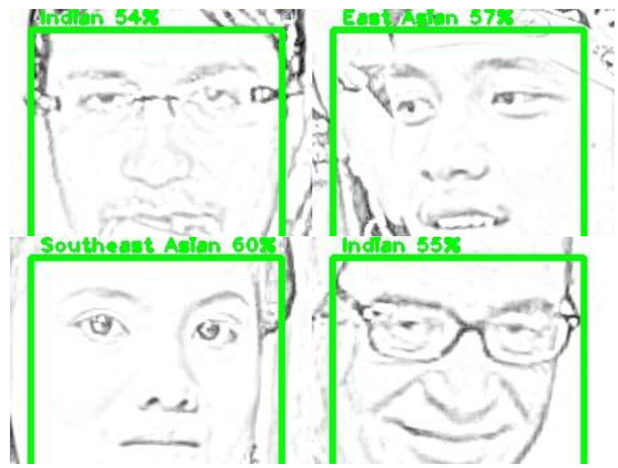

Fig. 19. Sample fails classified cases.

However, the model correctly identified the other ethnicity classes shown in Fig. 20. The figure illustrates some cases correctly classified as all seven basic ethnic groups, demonstrating the model's strong ability to identify ethnic groups with more distinctive characteristics. The first case in Fig. 20. is classified as Black with 91.7%; the second case is predicted Latino|Hispanic with 87.4%; the third one is classified as 82.6% Middle Eastern; the fourth case is predicted to be Southeast Asian with 81.9%; the fifth case is classified as 87.9% East Asian; the sixth case is predicted to be 95% White; and the last case is 93.16% Indian.

It is reported that for some face sketch images, the enhanced approach doesn't even allow for face detection in the image, and this problem is because the image doesn't include both eyes or the image isn't clear enough, as in the examples in Fig. 21:

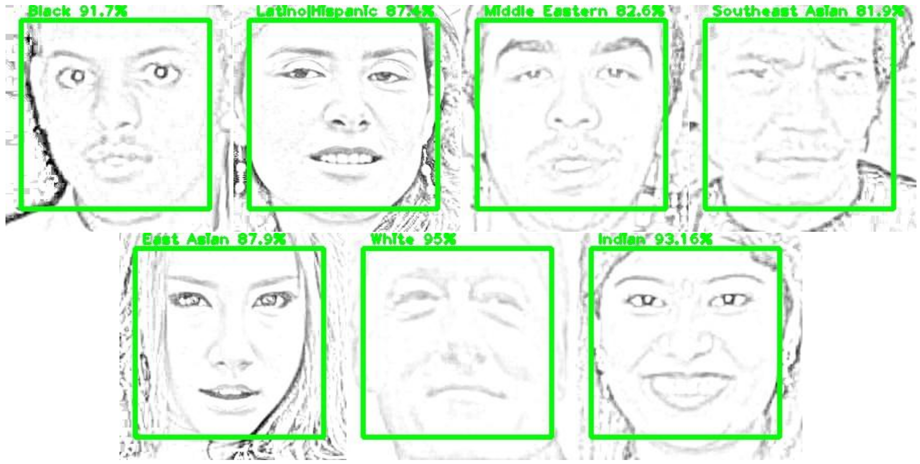

Fig. 20. Samples of the correct results of ethnic groups' face sketches classification and recognition.

Fig. 21. Samples of face sketches with failed face detection and their corresponding real photos from the FairFace dataset.

VI. Conclusion

This study proposes a new deep learning method for face sketch ethnic group classification and recognition. The method is based on the optimized VGG16 model. We used seven ethnic group sub-datasets for all ethnicities: White, Black, Indian, East Asian, Southeast Asian, Middle Eastern, and Latino. We tested our method on generated face sketches from the FairFace dataset using a combination of inversion, Gaussian blur, and fusion, and the results are very satisfactory. The accuracy is reaching above 94% and produced a low false negative rate. Our work is inspired by the recent successful methods that showed that a relatively mothered output VGG16 feature could be used to give good performance in a classification and recognition-based framework. To address the challenge of the lack of data and comparative models in the state of the art of the topic under study, the future work will include a decrease in the number of features used and the use of other competitive deep learning models.

References

[1] K.Suyemoto , M.Curley and S.Mukkamala, "What Do We Mean by “Ethnicity” and “Race”? A Consensual Qualitative Research Investigation of Colloquial Understandings" Genealogy journal, vol.4, no. 3, 10.3390,81, August 2020.

[2] S.Karen and A.Zisserman,“Very Deep Convolutional Networks for Large-Scale Image Recognition.” CoRR abs/1409.1556, 2014.

[3] S. Chen, B. Mulgrew, and P. M. Grant. "A clustering technique for digital communications channel equalization using radial basis function networks", IEEE Trans. on Neural Networks, Vol. 4, pp. 570-578, July 1993.

[4] H. Keat Wong, I. Stephen and D. Keeble "The Own-Race Bias for Face Recognition in a Multiracial Society", Front. Psychol, Sec. Cognition Vol. 11, March 2020.

[5] N.Bhala, G.Curry, A.Martineau, C.Agyemang and R.Bhopal, "Sharpening the global focus on ethnicity and race in the time of COVID-19”, The Lancet Journal, Vol. 395, no. 10238, pp. 1673-1676, May 2020.

[6] D..Pan, S.Sze, J. Minhas, M. Bangash, N. Pareek and P.Divall, "The impact of ethnicity on clinical outcomes in COVID-19: A systematic review", The Lancet Journal, eClinicalMedicine, Vol. 23, no. 100404, June 2020.

[7] K.Bhopal, " Gender, ethnicity and career progression in UK higher education: a case study analysis", Research Papers in Education, Vol. 35, no. 6, pp. 706-721, May 2020.

[8] L. Brown, U. Mitchell and J. Ailshire, "Disentangling the Stress Process: Race/Ethnic Differences in the Exposure and Appraisal of Chronic Stressors Among Older Adults", The Journals of Gerontology: Series B, Vol. 75, no. 3, pp. 650–660, March 2020.

[9] C. Rojas-Gaona, J. Hong and A. Peguero, "The significance of race/ethnicity in adolescent violence: A decade of review 2005–2015", J. Crim. Justice, Vol.46, pp.137–147, September 2016.

[10] K. Ounachad, M. Oualla, A. Souhar and A. Sadiq, "Face Sketch Recognition-An Overview”, NISS '20: Proceedings of the 3rd International Conference on Networking, Information Systems & Security, Article no. 38, pp. 1-8, March 2020.

[11] K. Ounachad, M. Oualla, A. Souhar and A. Sadiq, "Structured learning and prediction in face sketch gender classification and recognition", International Journal of Computational Vision and Robotics, Vol. 10, no. 6, January 2020.

[12] A. Gite, S. Gethe, S. Gupta and S. Pagare, "Implementation of VGG16 Based learning for Facial Recognition", International Journal of Progressive Research in Engineering Management and Science (IJPREMS), Vol. 5, no. 4, April 2025.

[13] V. Nushi, R. Santos, H. Brkić, A. Oliveira, A. Francisco and C. Pereira, "Pre-trained VGG16 model for forensic dental age estimation", Egyptian Journal of Forensic Sciences, Vol. 15, no. 73, October 2025.

[14] M. Kocoń and S. Pawlukiewicz, "Age Estimation and Gender Classification from Facial Images", Applied Sciences Journal, Special Issue Applications of Deep Learning and Artificial Intelligence Methods: 3rd Edition, Vol. 15, no. 18, September 2025.

[15] S. Ozechi, "African Gender Classification Using Clothing Identification Via Deep Learning", arXiv:2503.00058, Vol. 1, March 2025.

[16] S. Makinist and G. Aydin, "Gender Classification Using Face Vectors: A Deep Learning Approach Without Classical Models", Information, Vol. 16, no. 7, June 2025.

[17] K. Ounachad, A. Souhar and A. Sadiq, "Fuzzy Hamming Distance and Perfect Face Ratios Based Face Sketch Recognition", IEEE 5th International Congress on Information Science and Technology, (CiSt), Marrakech, Morocco, pp. 317-322, October 2018.

[18] K. Ounachad, M. Oualla and A. Sadiq, "Face Sketch Recognition: Gender Classification Using Eyebrow Features and Bayes Classifier", Innovations in Smart Cities Applications, Vol. 4, pp. 809-819, February 2021.

[19] X. Zhu & Y. Li, "Application analysis of computer vision and image recognition based on improved VGG16 network", Discov Appl Sci, Vol. 7, no. 867, August 2025.

[20] R. Sewada and H. Goyal, "A Novel VGG-16 Adaptation for Multi-band Satellite Image Classification: Optimized Preprocessing and Class-specific Augmentation", Journal of Computational and Cognitive Engineering, pp. 1-12, May 2025.

[21] M. Genç & Y. Yalman "Aircraft Recognition Based on CNN Using Satellite Images", Journal of Innovative Science and Engineering. Vol. 9, no. 1, pp. 1-14, 2025.

[22] E. Addisu, T. Yirga, H. Yirga and A. Yehuala,"Transfer learning-based hybrid VGG16-machine learning approach for heart disease detection with explainable artificial intelligence", Front. Artif. Intell, Vol. 8, February 2025.

[23] F. Tanjim and S. Hamdy, "Application of VGG16 Transfer Learning for Breast Cancer Detection", Information, Vol. 16, no. 227, pp.1-20, March 2025.

[24] W. Jaimes, W. Arenas, H. Navarro and M. Altuve, "Detection of retinal diseases from OCT images using a VGG16 and transfer learning", Discover Applied Sciences, Vol. 7, February 2025.

[25] T. Arkan, A. Sugiharto, and H. Wibawa, "Modified VGG16 Architecture with Bayesian Optimization for Tomato Leaf Disease Classification", Jurnal Masyarakat Informatika, Vol.16, no. 2, pp. 162-173, June 2025.

[26] I. Sahputra, A. Ulva, B. Putri and C. Eviyanti, "Comparative Study of VGG16 and MobileNet Architectures for Rice Leaf Disease Classification Using CNN", Journal of Informatics and Telecommunication Engineering, Vol. 9, no. 1, pp. 303-314, July 2025.

[27] L. Chen, Z. Wang and H. Liu, "Steel-Reinforced Concrete Corrosion Crack Detection Method Based on Improved VGG16", Coatings, Vol. 15, no.641, pp. 1-14, May 2025.

[28] K. Ounachad, M. Oualla, A. Sadiq and A.souhar, "Face Sketch Recognition: Gender Classification and Recognition", International Journal of Psychosocial Rehabilitation, Vol. 24, no. 03, pp. 1073-1085 February 2020.

[29] K. Ounachad, M. Oualla, A. Sadiq and A.souhar, "Golden Ratio and Its Application to Bayes Classifier Based Face Sketch Gender Classification and Recognition", International Journal of Emerging Trends in Engineering Research, Vol. 8, no. 7, pp. 3538-3545, July 2020.

[30] K. Ounachad, M. Oualla, A. Sadiq and A.souhar, "Human Face (Sketch/Photo) Age Group Estimation and Classification Using Perfect Face Ratios and Levenshtein Distance", International Journal of Emerging Trends in Engineering Research, Vol. 8, no. 7, pp.3191-3201, July 2020.

[31] K. Ounachad, M. Oualla, A. Sadiq and A.souhar, "Facial Emotion Recognition Using Average Face Ratios and Fuzzy Hamming Distance ", Journal of Automation Mobile Robotics & Intelligent Systems, Vol.14, no. 4, pp. 37-44, March 2020.

[32] K. Ounachad, M. Oualla and A. Sadiq, "Geometric Feature Based Facial Emotion Recognition", International Journal of Advanced Trends in Computer Science and Engineering, Vol. 9, no. 3, pp. 3417-3425, May – June 2020.

[33] K. Ounachad, M. Oualla, A.souhar and A. Sadiq, "Structured Learning and Prediction in Facial Emotion Classification and Recognition", International Journal of Engineering and Advanced Technology, Vol. 9, no. 4, pp. 152-160, April 2020.

[34] K.Karkkainen, & J.Joo, " FairFace: Face attribute dataset for balanced race, gender, and age for bias measurement and mitigation". In Proceedings of the IEEE/CVF Winter Conference on Applications of Computer Vision (WACV), pp. 1548–1558.January 2021.

[35] K. Karkkainen, & J. Joo, "FairFace: Face Attribute Dataset for Balanced Race, Gender, and Age", arXiv preprint arXiv:1908.04913, 2019.

[36] T.K. Bahiru, N. T. Sinshaw, T. H. Moges and D. K. Singh, "Auditing and Mitigating Bias in Gender Classification Algorithms: A Data-Centric Approach", arXiv preprint arXiv:2510.17873, 2025.

[37] J. Xu, C. Liu, X. Tan et al., "General information metrics for improving AI model training efficiency", Artif Intell Rev, Vol. 58, no.289, July 2025.

[38] C. Lehmann & Y. Paromau, "Quantifying Uncertainty and Variability in Machine Learning: Confidence Intervals for Quantiles in Performance Metric Distributions ", arXiv:2501.16931 [cs.LG], January 2025.

[39] Z. Ba Alawi, "A Comparative Survey of PyTorch vs TensorFlow for Deep Learning: Usability, Performance, and Deployment Trade-offs", arXiv:2508.04035 [cs.LG], August 2025.

[40] S. Abd, H. Mangi, A. Kadhum, R. kazim, B.Ali and F. AbdAlkarim, "Efficient face detection using viola-jones and neural networks: a comparative study", AJIEA, vol. 1, no. 1, p. 13, June 2025.

[41] S.A. Wijaya, T.S. Famuji, M.A. Mu'min, Y. Safitri, N. Tristanti, A. Dahmani, Z. Driss, A.Sharkawy and R. Sabur, "Trends and Impact of the Viola-Jones Algorithm: A Bibliometric Analysis of Face Detection Research (2001-2024)", Scientific Journal of Engineering Research, Vol. 1, no. 1, pp. 33-42, February 2025.

[42] S. Husain, & E. Abolkasim, "Estimating the Number of People in Digital Still Images Based on Viola-Jones Face Detection Algorithms", African Journal of Advanced Pure and Applied Sciences, Vol. 3, no. 2, pp. 146–154. June 2024.