

Article

Classification of Ethnicity Using Efficient CNN Models on MORPH and FERET Datasets Based on Face Biometrics

Abdulwahid Al Abdulwahid 

Department of Computer and Information Technology, Jubail Industrial College, Royal Commission for Jubail and Yanbu, Jubail Industrial City 31961, Saudi Arabia; abdulwahida@rcjy.edu.sa or wahid3anj@gmail.com

Abstract: Ethnic conflicts frequently lead to violations of human rights, such as genocide and crimes against humanity, as well as economic collapse, governmental failure, environmental problems, and massive influxes of refugees. Many innocent people suffer as a result of violent ethnic conflict. People's ethnicity can pose a threat to their safety. There have been many studies on the topic of how to categorize people by race. Until recently, the majority of the work on face biometrics had been conducted on the problem of person recognition from a photograph. However, other softer biometrics such as a person's age, gender, race, or emotional state are also crucial. The subject of ethnic classification has many potential uses and is developing rapidly. This study summarizes recent advances in ethnicity categorization by utilizing efficient models of convolutional neural networks (CNNs) and focusing on the central portion of the face alone. This article contrasts the results of two distinct CNN models. To put the suggested models through their paces, the study employed holdout testing on the MORPH and FERET datasets. It is essential to remember that this study's results were generated by focusing on the face's central region alone, which saved both time and effort. Classification into four classes was achieved with an accuracy of 85% using Model A and 86% using Model B. Consequently, classifying people according to their ethnicity as a fundamental part of the video surveillance systems used at checkpoints is an excellent concept. This categorization statement may also be helpful for picture-search queries.



Citation: Abdulwahid, A.A. Classification of Ethnicity Using Efficient CNN Models on MORPH and FERET Datasets Based on Face Biometrics. *Appl. Sci.* **2023**, *13*, 7288. <https://doi.org/10.3390/app13127288>

Academic Editors: Tasadduq Imam, Sisira Colombage and Shahadat Uddin

Received: 19 April 2023

Revised: 13 June 2023

Accepted: 14 June 2023

Published: 19 June 2023



Copyright: © 2023 by the author. Licensee MDPI, Basel, Switzerland. This article is an open access article distributed under the terms and conditions of the Creative Commons Attribution (CC BY) license (<https://creativecommons.org/licenses/by/4.0/>).

Keywords: face biometrics; soft biometrics; convolutional neural networks (CNNs); ethnicity classification; MORPH; FERET

1. Introduction

Soft biometrics, such as gender, ethnicity, age, and expression, have recently gained attention from the pattern recognition community because of their wide range of retail and video surveillance applications and the difficulty in designing effective and reliable algorithms in challenging real-world scenarios [1]. The face is the part of the human body that contains the most semantic information about an individual. Convolutional neural networks (CNNs) are increasingly being used to solve problems such as face recognition [2] and verification [3], gender recognition [4], aging prediction [5], and facial emotion recognition [6]. Ethnicity recognition, which is the ability of a system to discern whether an individual belongs to one of the ethnic groups based on facial appearance observations, has not received the same attention from the scientific community. As a result of new methods and datasets [7–10] that have been proposed to improve the accuracy of real-world applications with ethnicity-biased results or to give a definitive push to forensic applications, interest in ethnicity recognition has been growing (e.g., ethnicity-based subject identification for public safety). According to recent comprehensive assessments [11–13], ethnicity data is lacking. In the realm of deep learning, having a substantial amount of data is crucial for effectively training CNNs. However, when it comes to specific face soft biometrics, such as in ethnicity recognition, the availability of large datasets is still lacking [11–13]. Existing studies [14,15] have shown that CNNs trained on the currently available datasets

for ethnicity recognition have limited generalization capabilities. The scarcity of ethnicity data can be attributed, in part, to the challenges involved in collecting and annotating such data. In addition to the difficulty of identifying universal distinguishing characteristics, ethnicity cannot be quantitatively measured, unlike other biometric factors such as gender. In the absence of genetic traits that can categorize individuals based on commonly recognized “ethnicities”, the term “ethnicity” lacks biological meaning [2,3]. Instead, human-perceived distinctions in somatic face features are considered for categorization purposes. An automatic annotation technique cannot be developed using, for example, a person’s place of birth; human annotators must manually establish the ground truths of ethnicity groups, and their reliability greatly depends on the annotator’s competence. Facial soft biometrics such as age, gender, and expression have seen a boom in research using deep neural networks, but ethnicity has not had the same focus in that field of study. In addition, gathering new ethnicity datasets is not an easy task; it must be carried out manually by people trained to recognize the basic ethnicity groups using somatic facial features [6,8]. More than 3,000,000 face images have been annotated within four ethnicity groups, namely African American, East Asian, Caucasian Latin, and Asian Indian, to cover this gap in the facial soft biometrics analysis of the VGGFace2 Mivie Ethnicity Recognition (VMER) dataset [6]. With the help of three people from diverse ethnic backgrounds, the final annotations can be derived free from the well-known other-race effect [1,2]. Due to the inherent challenges associated with ethnicity classification, two efficient CNN models were developed for predicting the ethnicity of a face using the MORPH and FERET datasets. The first step involves gender classification, followed by the subsequent classification of the individual’s ethnicity. Following the introduction of the study in this section, Section 2 describes the literature review, while Section 3 explains the methodology, datasets, and experiments. Section 4 presents, analyzes, and discusses the experimental results of the study, and this is followed by conclusions and future directions in Section 5.

2. Related Work

2.1. Background

A person’s identification is the primary focus in face biometrics, but other soft biometric information, such as age, gender, ethnicity, or emotional state, is also significant [3]. Ethnicity categorization is a growing field of study and has a wide range of applications. Convolutional neural networks (CNNs) have been extensively used in ethnicity categorization in recent years; thus, the study provides an overview of the most recent developments in this area. Belcar et al. [3] examined the discrepancies between CNNs’ results when landmarks are plotted and when they are not. On the UTKFace and Fair Face datasets, the proposed model was put to the test using the holdout testing approach. Accuracy for a 5-class classification was 80.34%, while accuracy for a 7-class classification was 61.74%, which is slightly better than the current best practice. However, results were obtained using only a portion of the face, reducing both time and resources. Consequently, the results are more accurate than the current best practice.

2.2. Existing Methods

The challenge of determining a person’s ethnicity based solely on their visual traits was analyzed in [4]. Three primary ethnic groups were represented in this work: Mongolians, Caucasians, and blacks. The authors used 447 photos from the FERET database, 357 of which were used for training and 90 of which were used for testing. In order to solve the classification challenge, the authors extracted several geometric features and color attributes from the image. A CNN yielded a model with a precision of 98.6%, while an artificial neural network yielded a precision of just 82.4%. A new indexing strategy based on a hash table that uses a hierarchy of classifiers to predict attributes such as age, gender, and ethnicity was shown in [16]. The matching procedure selected only a tiny fraction of the database from the indexed hash table, lowering retrieval time while retaining low computational complexity. Transfer learning was used to train the hierarchical classifiers using a pre-trained CNN. A

new probabilistic back-tracking approach that corrects misclassifications using conditional probabilities was proposed to decrease the classification error generated by classifiers. Based on expected qualities, a method called dynamic thresholding was proposed that dynamically sets the threshold for matching computations based on the results of these computations. It was necessary to perform extensive testing on classifiers to compare their results with those of the most recent approaches. On a large-scale database, the suggested indexing strategy showed a significant reduction in search time and a significant boost in accuracy over current face-image retrieval techniques. Probabilistic back-tracking and dynamic thresholding were shown to be important in statistical tests. Researchers have recently focused their attention on the human face and its traits, making it one of the most popular topics. Known as “soft biometrics”, the features and information extracted from a person have been applied in various fields, including law enforcement, surveillance videos, advertising, and social media profiling, to improve recognition performance and the search engine for facial images. The authors of [6] discovered that there was no mention of the Arab world in relevant papers or Arab datasets. To identify these labels using deep-learning methodologies, the authors set out to generate an Arab dataset and properly label Arab sub-ethnic groups. Images from the Gulf Cooperation Council (GCC) countries, the Levant, and Egypt comprised the Arab image dataset that was developed. The challenge was solved by combining two different methods of learning. A CNN pre-trained model was utilized to obtain state-of-the-art outcomes in computer vision classification issues with supervised deep learning. In the second category of deep learning, there was unsupervised learning (deep clustering). Ethnic classification is one of the primary goals of employing unsupervised learning. To the best of this author’s knowledge, that was the first time that deep clustering had been applied to ethnicity classification problems. Three approaches were considered for this. With the Arab dataset labels adjusted, the best results were 56.97% and 52.12% when pre-trained CNNs were evaluated on different datasets. Deep-clustering algorithms were applied to several datasets, and the accuracy (ACC) ranged from 32% to 59%, with Normalized Mutual Information (NMI) and Adjusted Rand Index (ARI) values ranging from 0 to 0.2714 and 0.2543, respectively.

The assessment of age, gender, and ethnicity in human face photographs is a crucial step in a variety of fields, including access control, forensics, and surveillance. Face recognition and facial aging can be better understood using demographic estimations. When performing such a study, demographics and face recognition and retrieval are two separate components. The initial step in [17] was to extract demographically informative features based on facial asymmetry to predict an image’s age group, gender, and race. Face photos were then recognized and retrieved using demographic features. The demographic estimates from a state-of-the-art algorithm and the proposed approach were also shown. The MORPH and FERET face datasets showed that the suggested strategy can compete with existing approaches for recognizing face photos across aging differences in recognition accuracy. Many deep-learning (DL) algorithms have recently been developed for diverse applications, and those for face recognition (FR), in particular, have taken a huge leap. Deep FR systems benefit from the hierarchical architecture of DL approaches to develop discriminative face representations. On the other hand, DL approaches considerably enhance the current state-of-the-art FR systems and stimulate a wide range of efficient and diversified real-world applications. FR systems that use several types of DL approaches were examined, and 171 recent contributions were summarized for the study in [18]. Those authors addressed works on DL-based FR systems that deal with various algorithmic and architectural aspects as well as the present and future trends in the field. Afterward, they discussed various activation and loss functions for various DL approaches to better grasp the present state of the art. They also included a summary of the datasets used in FR tasks and problems relating to illumination, expressions, position variations, and occlusion, which were discussed in detail. Their final discussion looked at ways to improve FR duties and future developments. Despite recent advances in this area, ethnicity recognition with deep neural networks has received less attention from the scientific community than other

facial soft biometrics such as gender and age. To train CNNs for ethnicity recognition, the authors need a large and representative dataset, which is currently unavailable. In addition, gathering new ethnicity datasets is difficult; people are trained to recognize the basic ethnicity groups using somatic facial features that must be carried out manually. The VGGFace2 Mivia Ethnicity Recognition (VMER) dataset, which contains more than 3,000,000 face pictures annotated with four ethnicity categories—African American, East Asian, Caucasian Latin, and Asian Indian—fills this gap in facial soft biometrics research. To prevent the bias produced by the well-known race effect, the final annotations were obtained using a methodology that requires the judgment of three people belonging to various ethnicities. VGG-16, VGG-Face, ResNet-50, and MobileNet v2 were some of the prominent deep network architectures analyzed in [10]. Last, but not least, those authors conducted a cross-dataset evaluation to show that deep network architectures trained using VMER generalized better across diverse testing datasets than similar models trained on the largest ethnicity dataset available. The ethnicity labels for the VMER dataset and the code used in the studies are available at <https://mivia.unisa.it> (accessed on 6 June 2022) upon request. One of the most intriguing study areas in computer vision is face recognition. Deep-learning techniques such as the CNN have made strides in recent years. Face recognition has been a huge success for CNNs. Computer face recognition is a method used to identify people's faces in pictures. Several researchers have studied facial recognition. This report summarized researchers' work on facial recognition using a CNN [19], covering studies that have been published during the last five years, including those that were cited. Face recognition using a CNN was tested to see if a renewal had occurred. CNNs, facial recognition, and a description of the database that has been utilized in numerous studies are all included in this article's theoretical foundations. The survey aimed to yield new insights into facial recognition based on CNNs.

Face recognition and ethnicity recognition are closely related as they both involve analyzing and identifying characteristics from a person's face. While face recognition focuses on recognizing and verifying the identity of a person from their facial features, ethnicity recognition aims to classify individuals based on their ethnic background or race. In the context of this study, CNN models have been utilized to extract features from facial images, which were then used to classify individuals into different ethnicity categories. Therefore, the techniques and methodologies used in face recognition, such as CNNs, can be applied to ethnicity recognition tasks.

In the recent decade, the number of people interested in face recognition studies has increased significantly. Identifying people's ethnicities is one of the most difficult challenges in face recognition. New CNNs were used to construct a new model that can identify people's ethnicity based on their facial traits. Three different nationalities were represented by 3141 photos in the new dataset for identifying people based on ethnicity [20]. As far as this author knows, this is the first time that an image dataset for ethnicity had been collected and made available to the public. Each CNN was compared with two other models that were currently considered the best in the field: Inception V3 and VGG. Results demonstrated that the authors' model performed the best, with a verification accuracy of 96.9% from a set of photos of people that were evaluated. Unconstrained real-world facial photos are classified into preset ages and genders using age and gender forecasts of unfiltered faces. The standard methods, however, fail miserably when used with unfiltered benchmarks because of the wide range of changes in the unconstrained photos. Due to their superior performance in facial analysis, CNN-based approaches have recently been widely used for categorization. A novel end-to-end CNN approach was proposed in the work of [21] to accomplish robust age group and gender classification for unfiltered real-world faces. In the two-level CNN architecture, features were extracted and classified. Classification and feature extraction worked together to classify photos according to their age and gender based on the extracted features. A robust picture preparation approach was used to handle the huge variances in unfiltered real-world faces before they were input into the CNN model. It was technically possible to train the network on an IMDb-WIKI with

noisy labels, then fine-tune it on MORPH-II and ultimately on the original OIU-Audience dataset. Looking at the OIU-Audience benchmark, the experimental results reveal that their model performs the best in age and gender classification. It showed improvement above the best-reported results by 16.6% (precise accuracy) and 3.2% (one-off accuracy) for age-group classification and gender classification, respectively. In computer vision, the study of human facial image analysis is a hot topic. A methodology for facial image analysis was proposed to address the three tough challenges of race, age, and gender detection through face parsing [22]. The authors used deep CNNs to train an end-to-end face parsing model by manually labeling face photos. A facial image was segmented into seven dense classes using a deep-learning-based segmentation technique. Using probabilistic classification, the authors constructed probability maps for each face category. Probability maps were employed as a means of characterizing features. They created a CNN model for each demographic job by extracting features from probability maps (race, age, and gender). Extensive studies on state-of-the-art datasets yielded significantly better results than those previously obtained.

With the growth of the digital age and the rise of human–computer connections, the desire for gender classification systems has increased. Automated classification of gender could be used for various purposes, including indexing facial photos based on gender, monitoring gender-restricted places, gender-adaptive targeted marketing, and collecting passive data on gender demographics [23]. To test the accuracy of face gender classification algorithms, the National Institute of Standards and Technology (NIST) enlisted the help of five commercial companies and one institution, utilizing a combined corpus of close to one million facial pictures from visas and mugshots [24]. The testing approach adopted by NIST simulated operational reality, where software was shipped and used “as-is” without the need for further algorithmic training, which was the goal of the methodology. From a large dataset of photographs taken under controlled lighting, position, and facial expression settings, the core gender categorization accuracy was evaluated by gender, age group, and ethnicity. Comparing results from the limited dataset with those from commonly benchmarked “in the wild” (i.e., unrestrained) datasets was an important part of the research. Assessments were made on sketch classification performance and gender verification accuracy based on how many images were taken from each individual. Demographic variables such as categorization, age, ethnicity, and gender have a significant impact on the appearance of the human face, with each category further subdivided into classes such as black and white, male and female, young (18–30), middle age (30–50), and old (50–70). Most students look more like their peers in their age group than they do those in other age groups. Subjects from a wide range of ages, ethnicities, and genders were analyzed to see how the accurate facial verification was [5]. To that end, the authors employed a CNN for feature extraction and demonstrated that their approach outperformed a commercial face recognition engine in terms of performance for specific demographics. Women, young people between the ages of 18 and 30, and African Americans all had inferior biometric verification performance than other demographic groups. Using this strategy, the authors then tested the accuracy of face verification across multiple demographic groups. As a result of their findings, they offered recommendations on how to improve face verification for people of different ethnicities [24].

Unsupervised fair-score normalization was proposed to lessen the impact of bias on face recognition and lead to a large overall performance gain. The author’s theory was based on treating “similar” people legally by implementing a normalization strategy [25,26]. Three publicly available datasets were used in experiments conducted under controlled and natural conditions. When gender was considered, the results showed that the author’s method eliminated demographic biases by 82.7%. Furthermore, it consistently reduced the bias compared with previous efforts. The total performance was improved by up to 53.2% with a false match rate of 103 and by up to 82.9% with a false match rate of 105, which contrasted with earlier works. Furthermore, the method is not restricted to face biometrics and can be easily integrated into existing recognition systems.

Many real-world applications, such as human–computer interaction (HCI), demography-based classification, biometric-based identification, security, and defense, to mention a few, use ethnicity as an important demographic trait of human beings. A person's ethnicity can be deduced from face photographs using a new method presented in [2]. The proposed method employed an SVM with a linear kernel as a classifier, which employed a pre-trained CNN. In contrast to prior research, which used handcrafted features such as Local Binary Pattern (LBP) and Gabor, this technique leveraged translationally invariant hierarchical characteristics learned by the network. To support the authors' claim that their method can handle a wide range of expressions and lighting circumstances, they conducted extensive trials on ten facial databases. There were three classes of ethnicity: Asian, African American, and Caucasian. The average accuracy of classification was 98.28%, 99.66%, and 99.05%, respectively, across all datasets. Other races (e.g., Latinos) are considerably underrepresented in public face-image collections. Face analytic methods cannot be applied to non-white race groups since the models built on such datasets have inconsistent classification accuracies [24–27]. These datasets have a race bias problem; thus, the authors created a new face-picture dataset that has photographs from 108,501 people of different races. White, black, Indian, East Asian, Southeast Asian, Middle Eastern, and Latino were the seven racial categories the authors identified. The YFCC-100M Flickr dataset was used to capture images, which were then classified by race, gender, and age. Tests were run to gauge generalizability on pre-existing face attribute datasets and brand-new image datasets. When tested on new datasets, the model the authors trained using their data performed significantly better than other models for both male and female participants. In addition, the authors assessed the accuracy of various commercial computer-vision APIs across a range of demographics, including gender, race, and age. Race recognition (RR), which has numerous applications in surveillance systems, image and video interpretation, analysis, and others, is a challenging problem to solve. The use of a deep-learning model to help solve that challenge has been analyzed in [28,29]. A race recognition framework (RRF) was proposed, which included an information collector (IC), face detection and preprocessing (FD&P), and RR modules. Both independent models were presented here in the study for the RR module. In the first instance, an RR-based CNN model was used (the RR-CNN model). The trained model (RR-VGG) was a fine-tuning model for RR based on the VGG object-recognition model. The dataset is entitled VNFaces and is made up of photographs taken directly from Vietnamese Facebook pages. Their experiments compared the accuracy of the RR-CNN and RR-VGG models based on the suggested framework's performance. On the VNFaces dataset, the RR-VGG model with enhanced input photos had the highest accuracy (88.87%). In comparison, the independent and lightweight model (88.64%) of RR-CNN had the lowest accuracy (88.54%). Extension experiments showed that the authors' models may be applied to other race dataset problems, such as Japanese, Chinese, or Brazilian, with over 90% accuracy; the fine-tuning RR-VGG model attained the best accuracy and was suggested for the majority of situations. Ethnicity plays a fundamental and significant role in biometric recognition because it is a characteristic of human beings. A new approach to ethnicity classification is presented in [8]. Methods for identifying ethnicity by extracting characteristics from facial photos and developing a classifier based on these features are commonly used. Deep CNNs were used instead to extract and classify features simultaneously. The proposed method was tested on three populations: blacks and whites, Chinese and non-Chinese, and Han, Uyghurs, and other non-Chinese people. The proposed method was tested on both public and self-collected databases and found to be effective. Race classification in facial image analysis has been a long-standing problem [28]. To avoid analyzing every area of the face, focusing on the most noticeable ones is critical. The classification of ethnicity and race can be greatly aided by face segmentation, which is an important part of many face analysis tasks. A race-classification technique based on a face segmentation framework was proposed in [13]. A face segmentation model was built using a deep convolutional neural network (DCNN). The DCNN was trained by labeling facial images with seven different classes, which

included the nose (skin), skin (hair), eyebrows (eyelid), and the back (mouth). It was employed in the first phase to generate segmentation results. Probability maps (PMs) were constructed for each semantic class using the probabilistic classification approach. Five of the seven most important facial traits for determining a person's race were examined in depth. The DCNN was used to train a new model based on the features retrieved from the PMs of five classes. The authors tested the suggested race classification method on four typical face datasets and found it to be more accurate than earlier methods.

Because there is no universally accepted definition of what constitutes "race" as well as the fact that the world's population is so diverse, determining one's race is a difficult undertaking. The identification of four basic racial groups (Caucasian, African, Asian, and Indian) is the focus of this research. To train the author's deep convolutional network (R-Net), the authors used the recently developed BUPT Equalized Face dataset, which contains around 1.3 million photos in an uncontrolled environment. The studies in [30–37] were conducted on other datasets, such as UTK and CFD, to verify their validity. Additionally, the race-estimation model, VGG16, is compared to R-Net. This model's ability to withstand the rigors of a wide range of settings was demonstrated through experiments. Finally, Grad-CAM (Grad-weighted Class Activation Mapping) was used to visualize the deep-learning model. The creation of a deep-learning-based approach to intelligent face recognition for smart homes was one of the main objectives of [38]. In this study, a deep model based on trees was proposed for cloud-based face recognition. The dataset was gathered using a camera module built on a Raspberry Pi.

2.3. Existing Methods and Novelty

In one report, an investigation was conducted on how a candidate's skin tone, race, and ethnicity intersected with voters' voting preferences and interpersonal evaluations using three experimental studies (e.g., warmth, trustworthiness, and expertise) [9]. A light-skinned (as opposed to dark-skinned) African American candidate was the focus of study 1. The second study examined the voting preferences of white and non-white participants and looked at the influence of race, ethnicity, and skin tone on voting choices (lighter vs. darker). The third study focused on how race and ethnicity influence voters' preferences as well as the accuracy and significance of skin-tone memories. The authors found that white people were less inclined than non-white people to vote for underrepresented candidates of color because they had more negative views (e.g., they displayed less warmth; they thought candidates were less trustworthy). The extent to which this bias influenced voters was observed in the prediction of a candidate's perceived warmth, trustworthiness, and level of competence [36,37]. When race and ethnicity were connected with certain skin tones, race and ethnicity significantly impacted voting choices and attitudes. Table 1 shows the comparative analysis of some relevant previous studies.

Table 1. Comparative Analysis of Previous Studies.

Reference	Dataset	Techniques	Accuracy
[2]	FERET	CNN	75%
[36]	FERET	Learning-based Classifiers	88%
[37]	FERET	CNN	82%

3. Methodology

Image processing is used to reliably identify an individual using biometric characteristics such as a person's face, iris, or voice. There is a great need for artificial intelligence (AI) algorithms that can conduct classification jobs such as race and gender classification due to the proliferation of image data on the Internet. With the introduction of convolutional networks and other deep-learning techniques, the accuracy of many picture classification tasks has risen dramatically in recent years. This research proposes a method for determining a person's gender based on their facial image using various machine-learning classifica-

tion strategies, contributing to the existing body of literature on gender prediction from facial images. In this research, two datasets, MORPH and FERET, were used for ethnicity recognition and gender classification. In the first step, the image was preprocessed, and the image's texture was found. The data include labels indicating 0 for male and 1 for female, whereas for ethnicity, there are four classes: white, black, Asian, and others. Figure 1 shows the proposed methodology.

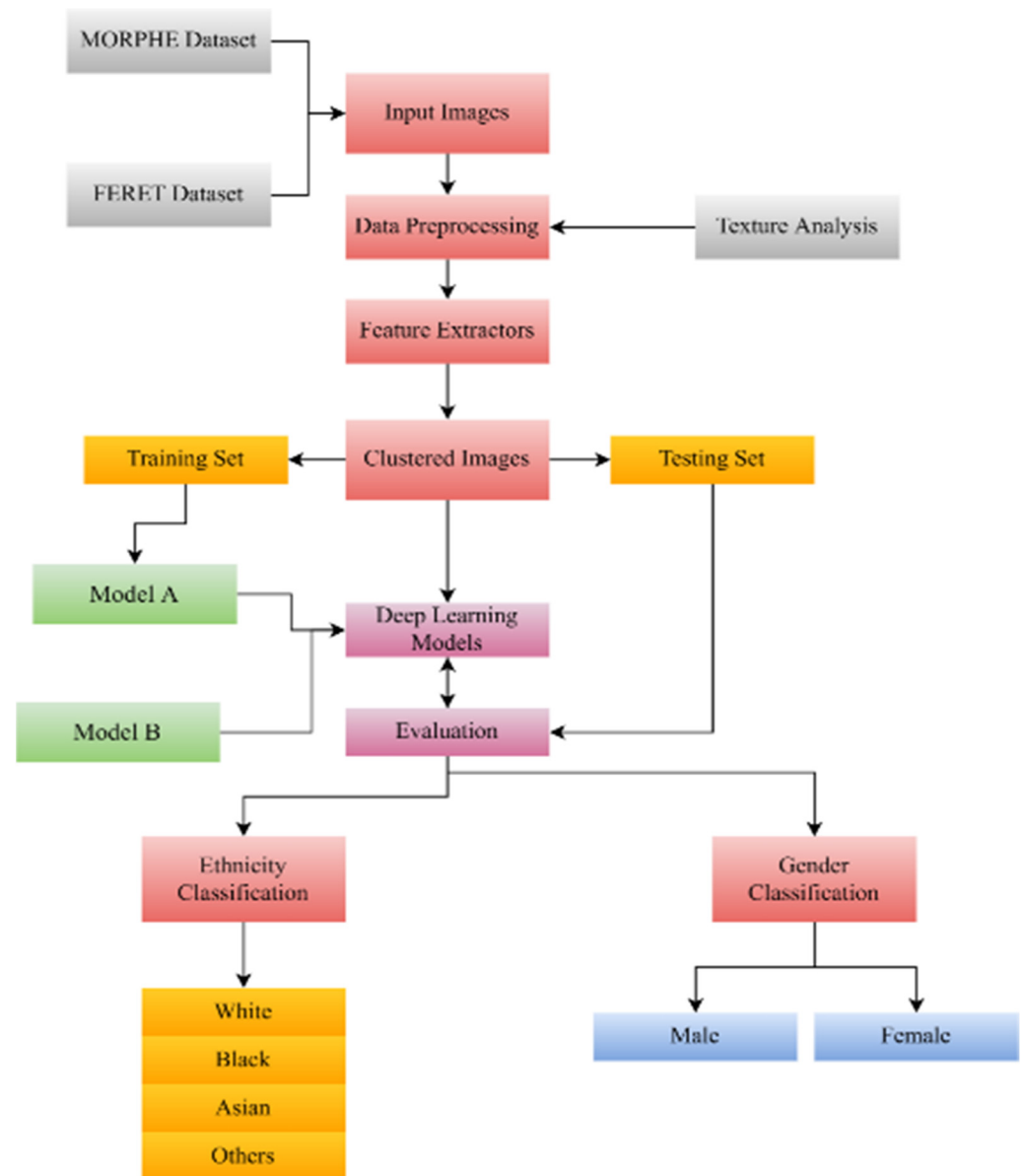


Figure 1. Proposed Flow of Study.

3.1. Augmentation Strategies

In the study, the aim is to evaluate the facial asymmetry of gender and assess the impact of demographic traits, namely ethnicity. For the male gender, the aim is to generate different facial attributes such as glasses variation, hairstyle synthesis, beard modification, and moustache variation. For the female gender, facial attributes such as glasses variation, makeup invariance, hairstyles, and different ornaments are targeted. Hairstyle synthesis may affect face recognition tasks as identification of the same person can become difficult when they change their hairstyle to occlude parts of the face. For instance, some people change their hairstyles in a way covering their forehead or even their eyes. These vari-

ations have performance effects on the CNN's ability to verify the person. The images have been augmented with different hairstyles to enrich the training dataset. Glasses variation is another area for data augmentation. Partial face occlusion is one of the problems in face recognition tasks. Makeup invariance has the substantial ability to change facial appearance.

A person usually applies different makeup in their daily routine. When makeup is changed, it becomes difficult for CNNs to verify or identify the person; as a result, it affects the performance of the network. Applying makeup differently to the same identity augments the training sets and makes the network's performance more robust. Variations in beard styles in males can be very useful in augmentation. Round, Van Dyke, and short boxed beards are different styles. The aim is to change the beard styles of the same individual to enhance the network's accuracy in recognizing faces. Similarly, moustache variations also have a great impact on face recognition. Pencil, handlebar, horseshoe, walrus, and chevron are different moustache styles. Beard and moustache style variations alter the lower part of the face, especially near the mouth and chin area. Applying different types of ornaments to the same individual also increases recognition accuracy.

3.2. Dataset Description

MORPH and FERET datasets were used during the research process. The MORPH dataset contains 55,000 images of 13,000 individuals from different ethnicities, i.e., European, African, Asian, Hispanic, and others. The FERET dataset contains 3580 facial images that belong to different ethnicities. Figure 2 shows the augmentation of dataset while Figures 3 and 4 show some example images from the MORPH and FERET datasets, respectively.



Figure 2. Augmentation Strategy.



Figure 3. MORPH Dataset.

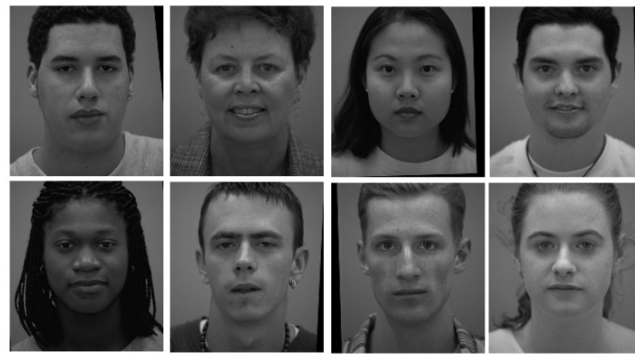


Figure 4. FERET Dataset.

3.3. Face Verification Experiments

During the training phase, a DCNN performs the face verification task. Face verification aims to determine whether the test image matches a specific image or identity in the dataset. First, the network is trained with the sets containing insufficient facial images. The performance of the network is noted for further use. Then, the training dataset is augmented using the different strategies mentioned above. After data augmentation, the performance is enhanced. Recently, DCNNs have gained popularity in face recognition tasks. However, the small number of images restricts the performance of DCNNs. Data augmentation is an effective approach to solve this problem. The architecture of DCNNs is designed so that the output of the first layer becomes the input of the following layer and so on. The convolutional layer and pooling layer are present in the most prominent layer of the network. The convolutional layer aims to apply the learned kernels to the input data and then generates the feature maps. After the convolutional layer, a pooling layer extracts the abstract information from the underlying feature maps. The fully connected layer is usually connected at the end of the DCNNs.

The MORPH and FERET datasets consist of 10,458 and 1020 images, respectively. They contain images from different ethnicities i.e., European, African, Asian, and others. The aim is to conduct four experiments to augment the identification of individuals from different ethnicities with the help of different strategies. Data augmentation strategies include varying glasses, makeup invariance, varying hairstyles, different ornaments, beard modification (for male), and varying moustaches (for male). The YouCam makeup application is used for these augmentations. The datasets are further divided into additional categories, namely training, validation, and test sets, as shown in Table 2.

Table 2. Training, Testing, and Validation Datasets.

Experiment no	Training Set (50%)	Validation Set (25%)	Test Set (25%)
1	1355 images	677 images	677 images
2	4110 images	2055 images	2055 images
3	171 images	85 images	85 images

3.4. Texture Classification

Distinguishing between different textures, called “Texture Classification”, is a long-standing issue in pattern recognition. The main problem is generating useful features to extract from a textured image, since many extremely complex classifiers exist. Optical texture can be perceived by looking at the image, but tactile texture can only be felt by putting one’s hands on it. Optical or visual texture is a term used to describe the image’s form and content. Figure 5 below shows the texture classification of a dataset:

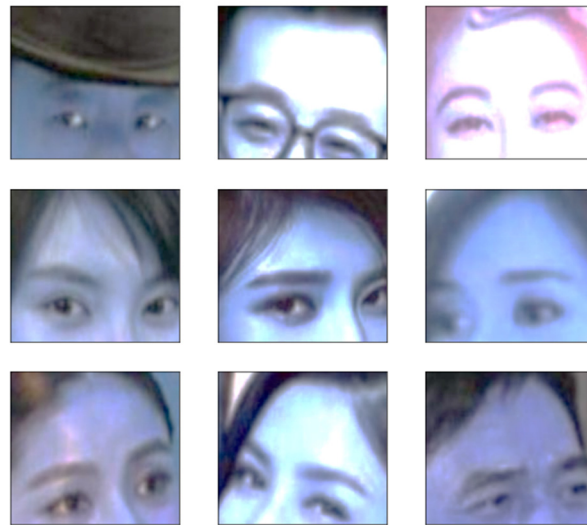


Figure 5. Texture Classification of a Dataset.

3.5. Feature Extraction

These are functions that learning machines can employ by analyzing data from a certain domain. Raw data must be transformed into machine-learning representations, which must be done manually. The study uses a correlation matrix to determine the degree of correlation between the various variables. Correlation matrices are just a covariance matrix. The correlation is a summary metric for measuring the linear association's strength. The concept of correlation summarizes the frequency and direction of a straight-line connection between two quantitative variables. It accepts values ranging from -1 to $+1$, represented by r .

3.6. Deep-Learning Models for Image Classification

This stage serves a primary purpose. Detection and verification of input are necessary for facial recognition to work. As a result, images must be recorded or captured using an image sensor or camera. The camera must be compatible with the software used to capture images. To begin, an image must be supplied. Images and video recordings, as well as real-time video, can be used as input. Faces in the photos or videos can be identified when the user provides input. Recognition can begin once the classifier has been trained. Both video and images can identify a single person or multiple people. A new collection of Python scripts is provided for each type of recognition. The classifier developed in the previous stage is imported into the Python script, which then performs the face recognition from a camera or an image. The image captured or a face from a database photograph can be used for identification. Face Detection is a technique that checks whether or not a picture contains a face; once the face has been identified, this output is passed on to the preprocessing step. The system classifier step uses TensorFlow, a tool for constructing neural networks. During the recognition process, a classifier is trained and utilized. Training the system to produce a more accurate classifier takes substantial time. The better the classifier is, the more time it has spent training. It takes three days to train the proposed face recognition system. The accuracy can be improved if the training is permitted to continue longer.

Proposed Gender Classification Model

This section describes the proposed structure. The proposed framework employs the MORPH and FERET datasets to determine a subject's gender based on an image.

Many people have used, or at least heard of, OpenCV's implementation for gender recognition. In 2015, Gil Levi and Tal Hassner developed this technique. In this particular illustration, however, a different method is employed for determining gender. In this case,

their CNN models have been used. The OpenCV's Deep Neural Network (DNN) module is utilized. OpenCV has a DNN package with a class named Net for populating neural networks. Popular deep-learning frameworks such as Caffe, TensorFlow, and Torch can also import neural network models into these packages.

The proposed system employs a CNN architecture. Using only their qualities, a deep-learning CNN algorithm can accurately categorize photos. Image analysis, segmentation, classification, medical image analysis, video identification, etc. are some of the many uses for CNNs. In the first step of the study, photos were preprocessed using an image processing technique to convert the raw data into a more manageable and actionable format.

Figure 6 below shows the gender distribution in the given datasets:

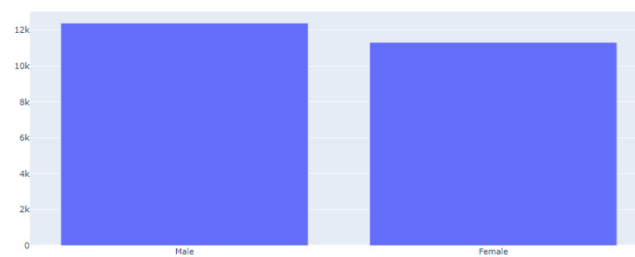


Figure 6. Gender Distribution.

In the study, a model consisting of a CNN was developed and merged with the SVM-Boosted algorithm. The reason for merging these two algorithms was to make the architecture more suitable for classification. The input image taken from the local database is added to the convolutional layer, and its features are extracted at the maximum pooling layer. Textures of each face are identified and matched with the ground truth; then, the values are extracted to a CSV file to make the system more robust. The CSV file containing features (refer to Table 2) is extracted, and the SVM-Boosted algorithm is applied to classify the correct face. Figure 7 shows the architectural diagram of the proposed CNN-SVM-Boosted Face Recognition algorithm. Mathematically, the proposed architecture has been proved in the following equations described below.

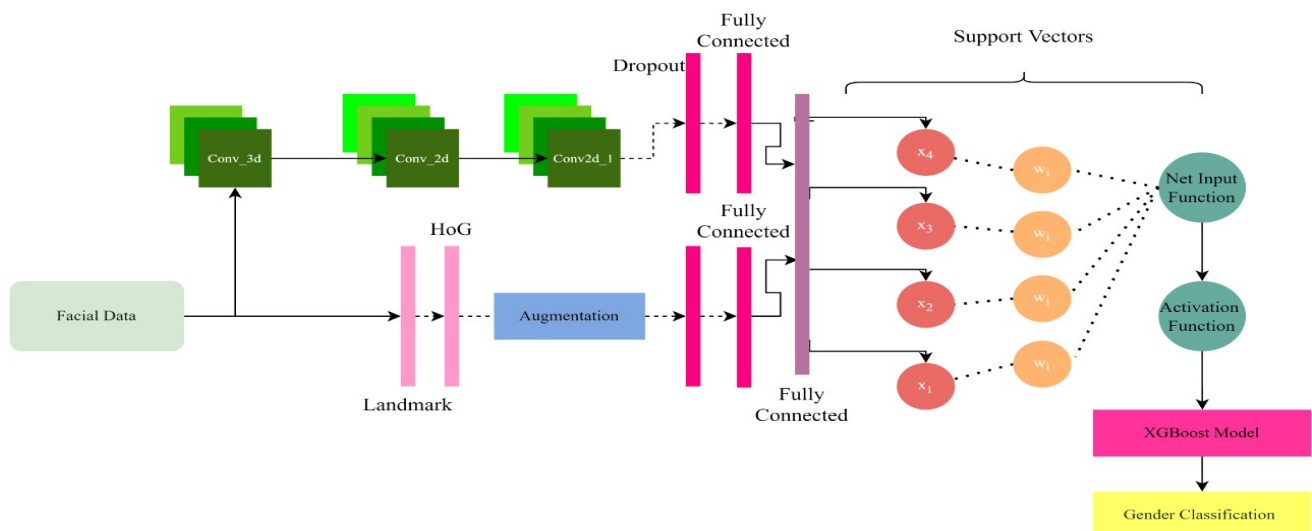


Figure 7. Model A for Gender Classification [38].

While kernel convolution is an integral part of CNNs, it is also employed in various computer vision algorithms. To achieve this effect, a small numeric matrix (the kernel or filter) is applied to the image, and the resulting transformation is evaluated. When an input image (denoted f) and a kernel (denoted h) are available, the values of the feature maps can

be determined using this information. M and N denote the row and column indexes of the final matrix.

$$G(m, n) = (f * h)[m, n] = \sum_j \sum_k h[j, k] f[m - j, n - k] \dots \quad (1)$$

This model has been developed by ensembling the Support Vector Classifier model into the XGBoosting Classifier to improve both models' accuracy. The mathematical model of the SVM-XGBC Classification model is as follows:

$$y = y^i = y^i + G(m, n) * \frac{\partial \sum (y_i - y_i^p)^2}{\partial y_p^i} \dots (a)$$

Then the support vectors are calculated as

$$w \cdot y + b = 1 \dots (\text{vector } 1)$$

$$w \cdot y + b = -1 \dots (\text{vector } 2)$$

Here, p is the probability function of the Support Vector Classifier and y^i is the output of the XGBC classification model. $\frac{\partial \sum (y_i - y_i^p)^2}{\partial y_p^i}$ shows the sum of the residuals in trees, while α is the learning rate of XGBC. When XGBC takes the output of y , it is sent to the probability function of the Support Vector Classifier for classification. Figure 7 shows the model A for Gender Classification

3.7. Model Training

To avoid any unexpected results while calculating the embedding vector for individual faces, preprocessing entails cropping faces out of the original photos. Face Detection Algorithms (such as Viola-Jones), Local Binary Patterns, and AdaBoost are used to locate the bounding boxes of all faces. For face detection, the Viola Jones technique is employed to run on every image in the original dataset and then produce a new dataset with just the face clipped out. Using preprocessing, the background noise that was impacting the accuracy of the facial embedding predictions was successfully eliminated. A convolutional network was utilized to reduce the image size from its original $224 \times 224 \times 3$ pixels to its final output size of 128×1 pixels. Here, the identical framework proposed in the original paper for the Facenet model is employed.

A 3D convolutional filter is the first layer of the network, followed by 2-batch normalization and max-pooling layers. Following that, there are 10 Inception modules. Each Inception module contains 1×1 , 3×3 , and 5×5 convolutional filters. The combined output of all these filters is then sent on to the next layer. Following processing by each of the 10 Inception Modules, the final image has 771,024 dimensions. After all the channels in the final image have been averaged, the result is a vector with 1024 dimensions.

Subsequently, these 1024 output units pass through the output ReLU activation layer of a dense fully linked network with 128 output units. The total results from each node's output are mapped onto a 128-dimensional vector. These additional layers allow the network to acquire more nuanced knowledge about the image's content. As a means of accelerating backpropagation, residual connections are inserted between layers. The Facenet model's initial training weights have been used in this application of Transfer Learning.

To sum up, the input image (of size 2,242,243) is fed into the network, and the resulting output vector (of size 128) contains the facial embedding of the original image. The network consists of 24 layers and was developed with efficiency in computation in mind. Facial embeddings are constructed using the same 128 dimensions as those used in the original article since those dimensions allowed the authors to reach greater accuracy.

3.8. Proposed Ethnicity Recognition Model

In computer vision and pattern recognition, face analysis has been one of the most heavily researched topics during the past few decades. Although a person's face can be used to infer many things about them, including their gender, age, and ethnicity, among other things, one of the most enduring and essential characteristics of a person is their ethnicity. Therefore, categorizing people only according to their age and gender could lead to confusion and inaccurate conclusions. Therefore, video surveillance systems at checkpoints can benefit greatly from incorporating ethnicity classification as a fundamental component. In addition, this categorization statement could be useful in picture-search queries, where a priori knowledge of race would simplify the process by reducing the size of the database to search. Figure 8 below shows the Model B architecture for ethnicity classification:

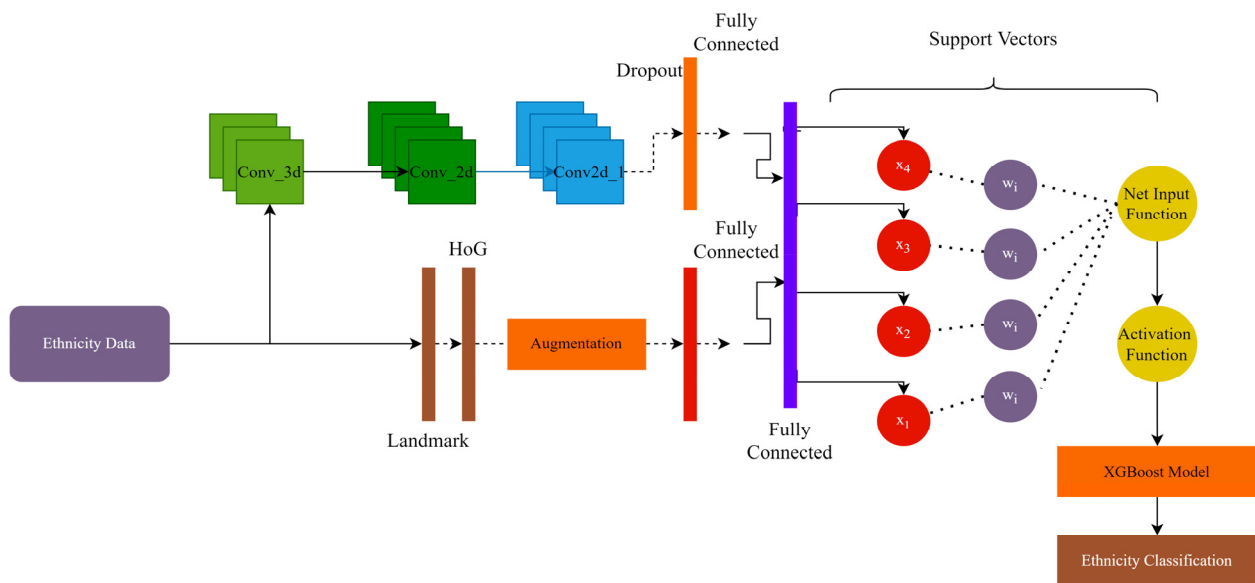


Figure 8. Model B Ethnicity Classification: Extraction of Skin Color and Features Extraction.

The two trials were designed to help us find an effective way to categorize people based on ethnicity. A simulated neural network and a CNN were used in the first experiment. The procedures used to conduct these tests are described in detail below. In the study, a feedforward artificial neural network was trained using a backpropagation-based training algorithm after extracting several facial features from the test photos. The facial region in the target image was identified by employing the Viola Jones [5] algorithm. Once the face was identified, the cascade classifiers were used to label the various features of the face. Then, the gaps and proportions between these features were determined. Various races had distinctive permutations of facial geometry. Skin tones can differ from culture to culture. Because individuals of different ethnicities tend to cluster together in a relatively narrow range of the color spectrum, their origins may be discerned simply by observing the hue of their faces. There are color and black and white examples in the training set. The levels of intensity in the RGB color space are quite wide. The input image was transformed from RGB to YCbCr to compensate for lighting changes.

3.9. Normalized Forehead Area Calculation

To determine a person's ethnicity, the normalized forehead area is crucial. The ratio of a person's forehead to the size of their entire face is a distinguishing feature of the face, and this is especially true with people of black ethnicity. Blacks tend to have a larger than average forehead compared with people of other races. Sobel Edge Detection [9] is used to determine the size of the forehead. By smoothing the edges in a vertical and horizontal plane, the exact location of the observer's eyes can be pinpointed. The area just above the

eyes pinpoints the exact front of the head. The normalized forehead area is determined using the previously observed eye and facial locations.

$$\text{Normlized forehead area} = \frac{\text{foreheadarea}}{\text{totalfacearea}}$$

3.10. Proposed Convolutional Neural Network Model

By quickly setting up a baseline, it is possible to determine whether the CNN architecture works better when only the raw pixels of images are utilized for training or if it performs better when additional information is input into the CNN architecture (such as nodule landmarks or HOG features). The results show that the additional information improves the CNN's ability to execute its function. The model was trained using this dataset. OpenCV is used to identify the faces, and then the dlib library extracts the faces' landmarks.

A CNN was also used to process the raw image data, the nodule locations, and the HOG characteristics. For the experiments, two CNN models were employed, as follows:

3.11. Model Training and Testing

Once the dataset was adapted, all the photos were cropped around the nodule and resized to 299×299 pixels. Contrary to the dataset, the distribution of classes is considerably varied with a slight imbalance between the classes. Lastly, the training, validation, and test sets were constructed following the holdout approach. The datasets were pooled and separated into 80% training, 10% validation, and 10% test. However, for the partitioning of the datasets, the works of [31] were considered in which the models showed an ability to adapt to people's looks, or even to how they convey their emotions. Therefore, for separating the datasets into training, validation and test data, the divisions were established in such a way that photographs of a single individual only existed in one of the datasets. Figure 9 below shows the 4 labels of ethnicity distribution as:

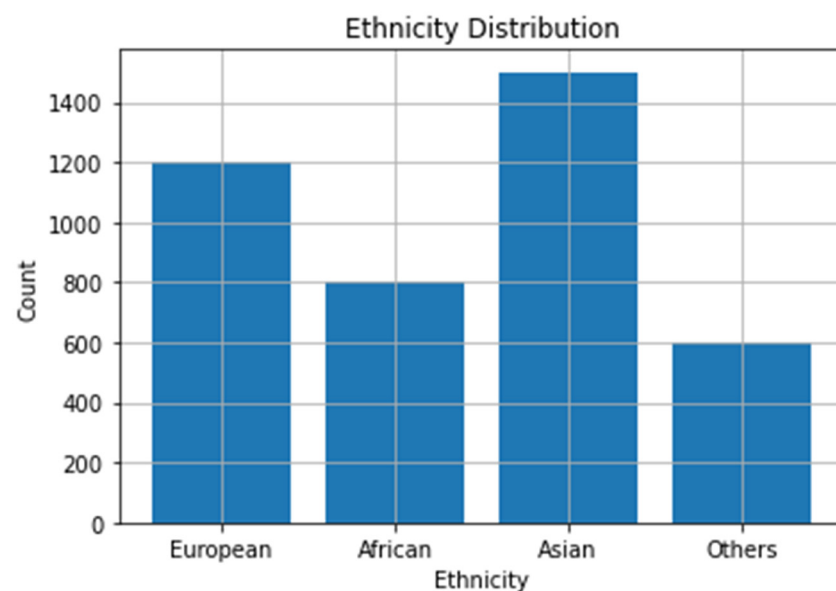


Figure 9. Ethnicity Distribution.

3.12. Model Evaluation Parameters

Accuracy, Precision, Recall, and F1 Score metrics have been used to assess the effectiveness of the various methods. However, these have been classified and misclassified, as shown by the confusion matrix. The metrics utilized in this investigation are shown in Figure 10:

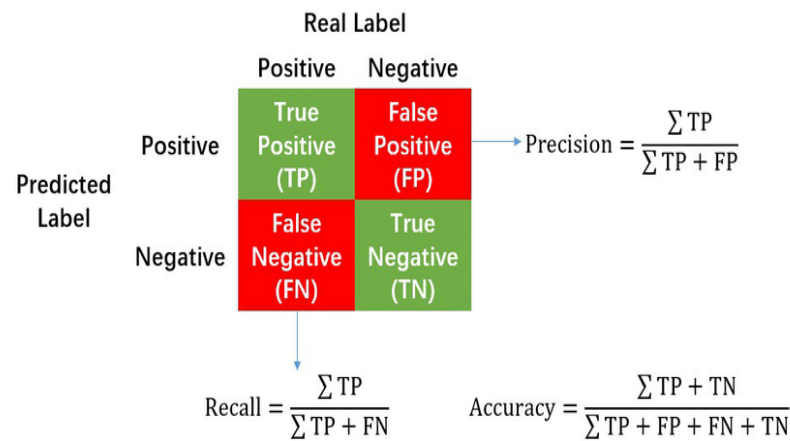


Figure 10. Example of Confusion Matrix.

4. Results and Discussion

4.1. Results for Model 1 on Dataset 1

In these experiments, a MORPH dataset was used. It has been a long-standing problem in pattern recognition, referred to as “Texture Classification”, to accurately distinguish between different types of textures. Since many incredibly complicated classifiers are available, the major challenge here is to develop relevant features to extract from a given textured image. In contrast to tactile texture, which can only be experienced by touching, optical texture can be perceived by the naked eye. One way to define an image’s structure and details is through the concept of “optical” or “visual” texture. Figure 11 shows the gender prediction for Model 1.

Figure 12 shows the gender classification for Model 1. Figures 13 and 14 show the performance of Model 1 in gender classification. The use of Model 1 on Dataset 1 showed a training accuracy of 0.84 and a testing accuracy of 0.86, while there was a 0.34 training loss and a 0.37 testing loss.

4.2. Results of Model 2 on Dataset 1

A deep convolutional neural network carries out this face verification challenge during the training phase. The goal of face verification is to determine if a given image matches one already stored in a database. The outcome of the face verification process is reported as a simple yes or no. At first, the network is taught using insufficient face photos from incomplete datasets. The network’s performance is tracked for future reference. The procedures mentioned above are then used to enhance the training dataset. With additional data, the efficiency is improved. For some time now, DCNNs have been widely used for facial recognition. However, the DCNN’s effectiveness is limited by the relatively few images available. Data augmentation is a powerful method for addressing this issue. The output of the first layer is fed into the next layer as input, and so on, as defined by the architecture of a DCNN. The network’s most visible layer contains both the convolutional and pooling layers. The purpose of the convolutional layer is to construct the feature maps by applying the learned kernels to the input data. Following the convolutional layer is the pooling layer, which is responsible for abstracting the feature maps beneath it. The last layer of a DCNN is often the completely linked layer.

A total of 10,458 and 1020 photos make up the MORPH and FERET datasets, respectively. Various people groups from Europe, Africa, Asia, and the Americas are represented. Figure 15 shows the Age and Ethnicity classification.



Figure 11. Gender Prediction for Model 1.

Figures 16 and 17 show the performance of Model 2 on Dataset 1 in ethnicity classification. Model 2 resulted in 0.70 and 0.74 training and testing accuracy, respectively, with a training loss of 0.80 and a testing loss of 0.75.

4.3. Results of Model 1 on Dataset 2

Figure 18 shows that the Gender Classification of Model 1 on Dataset 2 exhibits a 0.72 training accuracy and a 0.75 testing accuracy, while the training loss was 0.2 and the testing loss was 0.4. Figure 19 shows the performance in Gender Classification of Model 1 on Dataset 2, while Figure 20 shows the confusion matrix of the performance in Gender Classification for Model 1 on Dataset 2.

4.4. Results of Model 2 on Dataset 2

Figure 21 shows the ethnicity Class Model 2 whereas Figures 22 and 23 demonstrate the Ethnicity classification of Model 2 and Dataset 2, respectively.

Figure 24 illustrates Model 2's performance of Ethnicity classification on Dataset 2 with 0.72 training and 0.75 testing accuracy as well as 0.2 training loss and 0.3 testing loss. Accordingly, Figure 25 represents the Confusion Matrix of Model 2 Performance for Ethnicity Classification.



Figure 12. Gender Classification for Model 1.

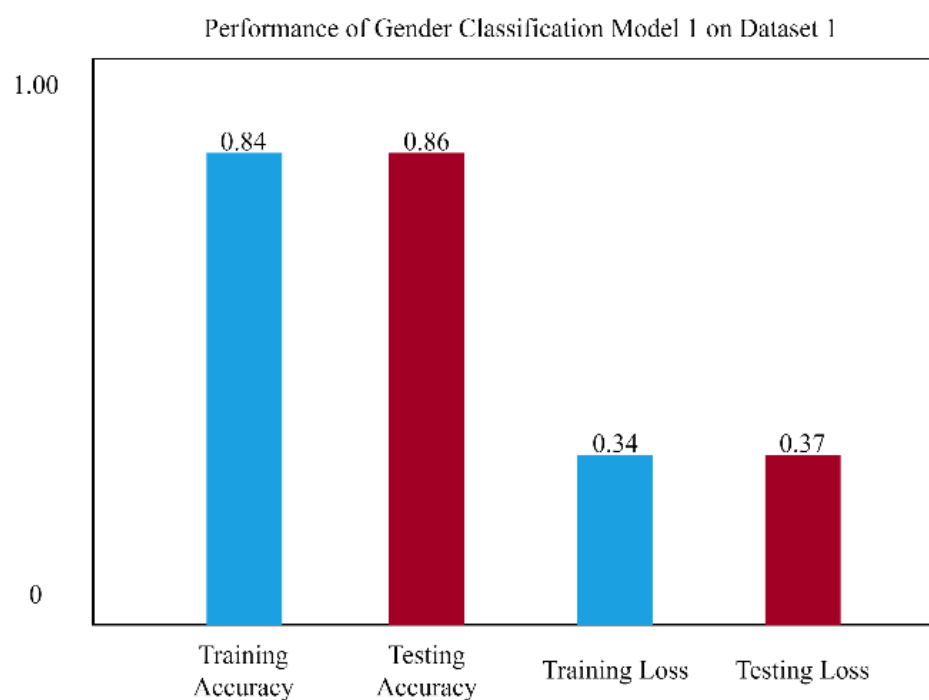


Figure 13. Model 1 Performance in Gender Classification.

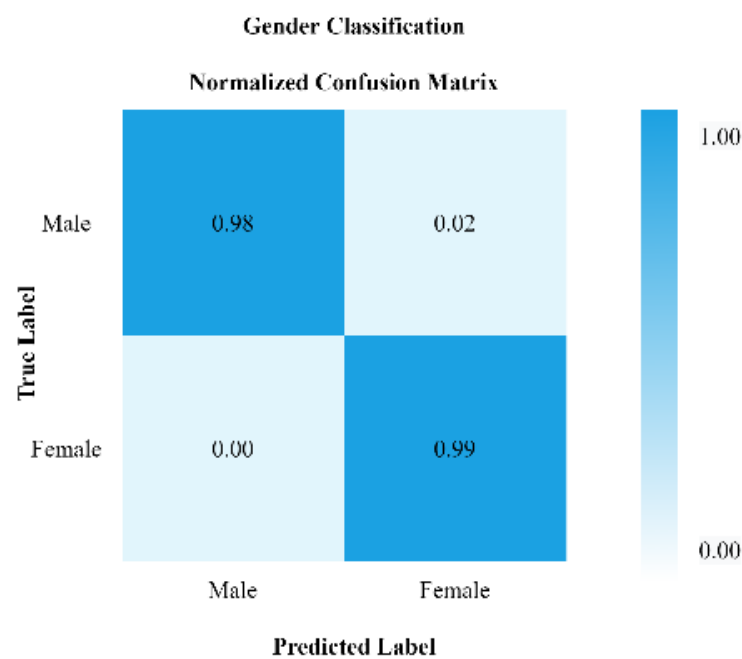


Figure 14. Confusion Matrix of Model 1 Performance in Gender Classification.



Figure 15. Age and Ethnicity Classification.

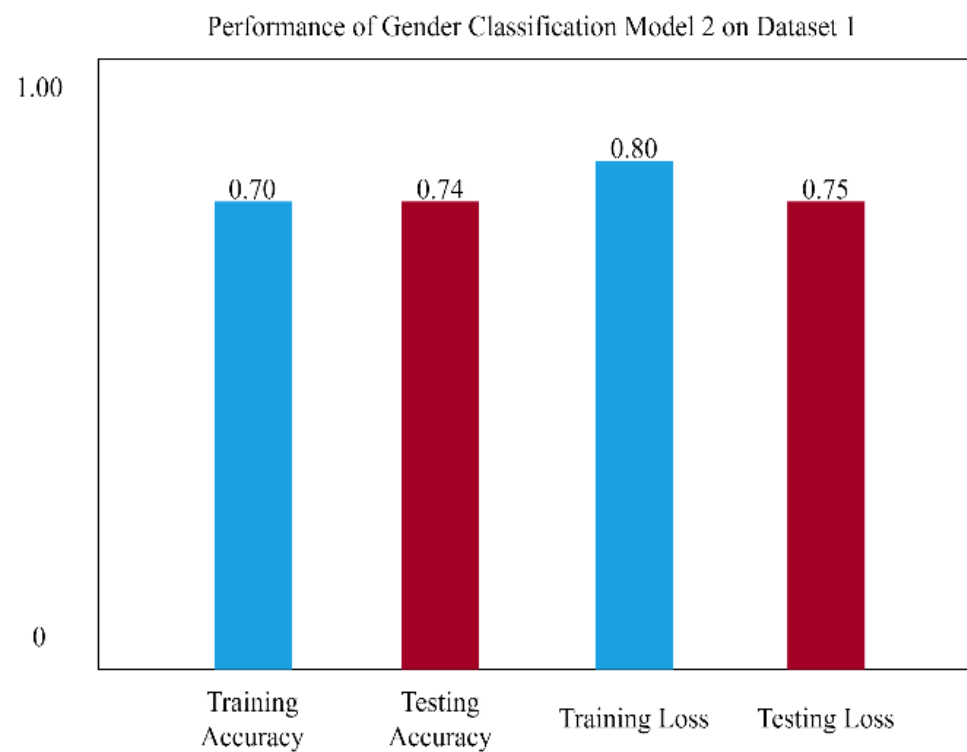


Figure 16. Confusion Matrix of Model 2 Performance in Ethnicity Classification.

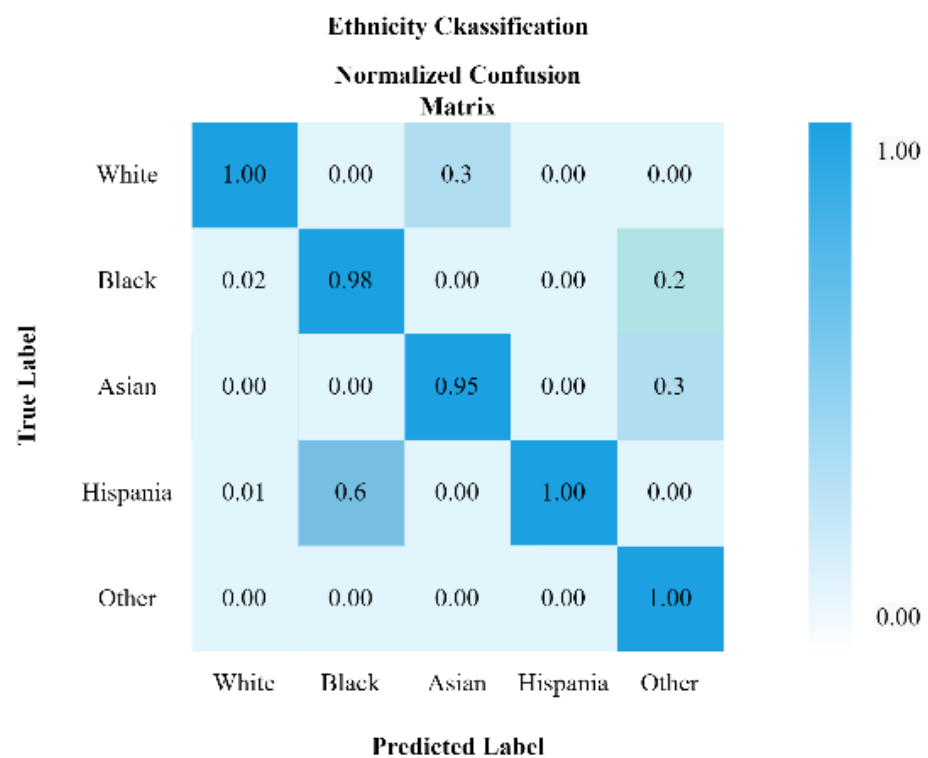


Figure 17. Model 2 Performance in Ethnicity Classification.

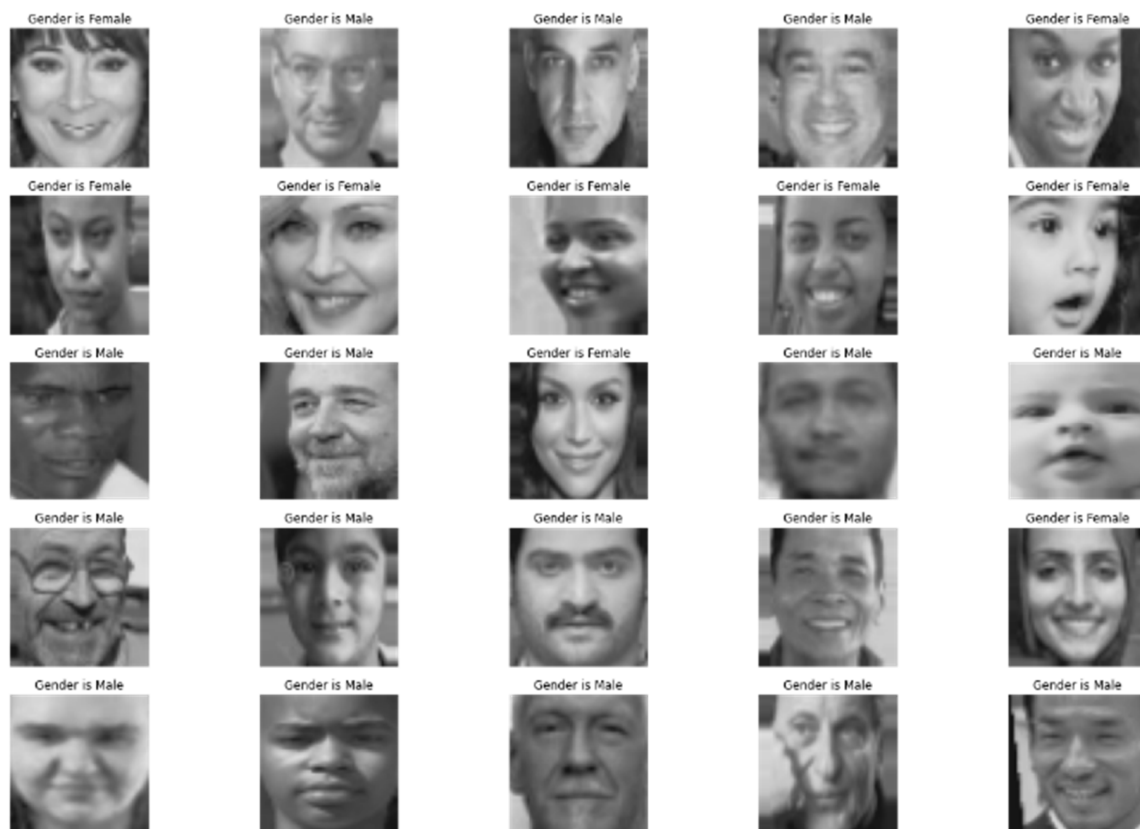


Figure 18. Gender Classification on Dataset 2.

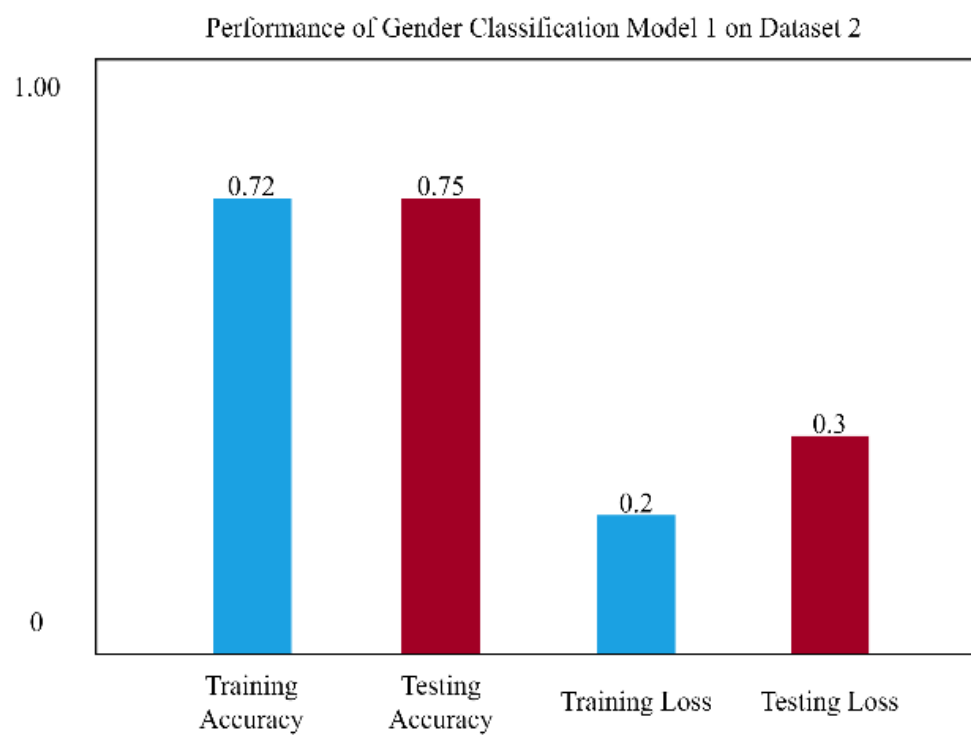


Figure 19. Performance of Gender Classification Model 1 on Dataset 2.

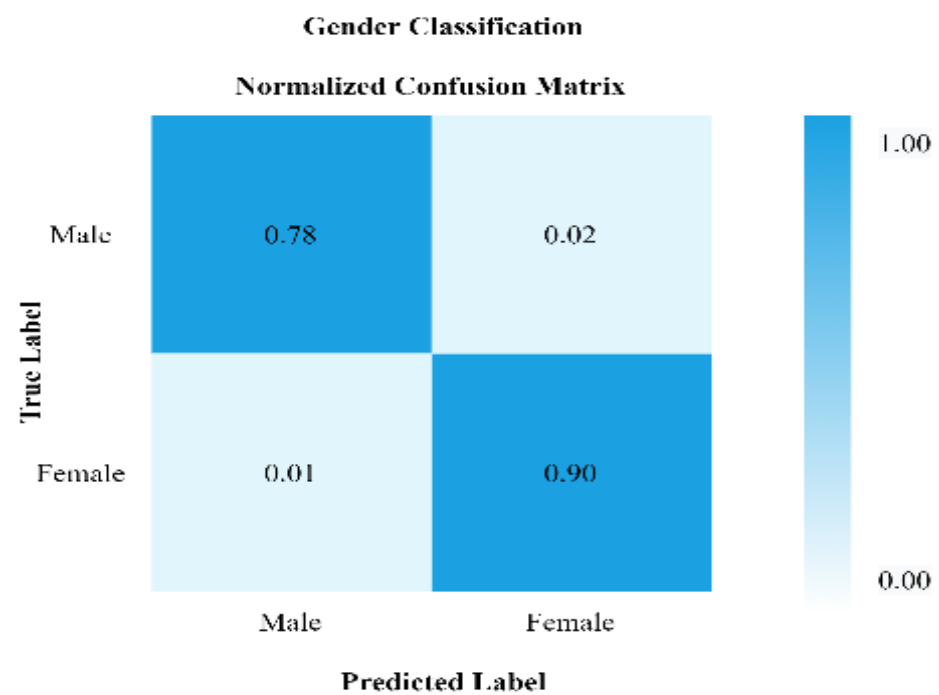


Figure 20. Confusion matrix of Performance of Gender Classification Model 1 on Dataset 2.



Figure 21. Ethnicity Class Model 2.

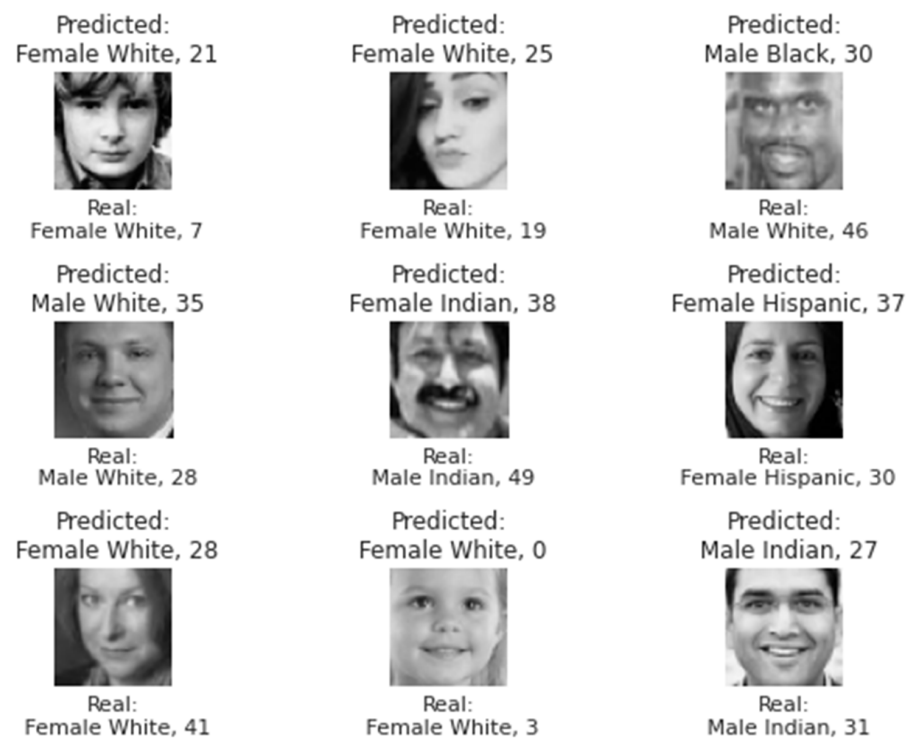


Figure 22. Ethnicity Classification for Model 2.

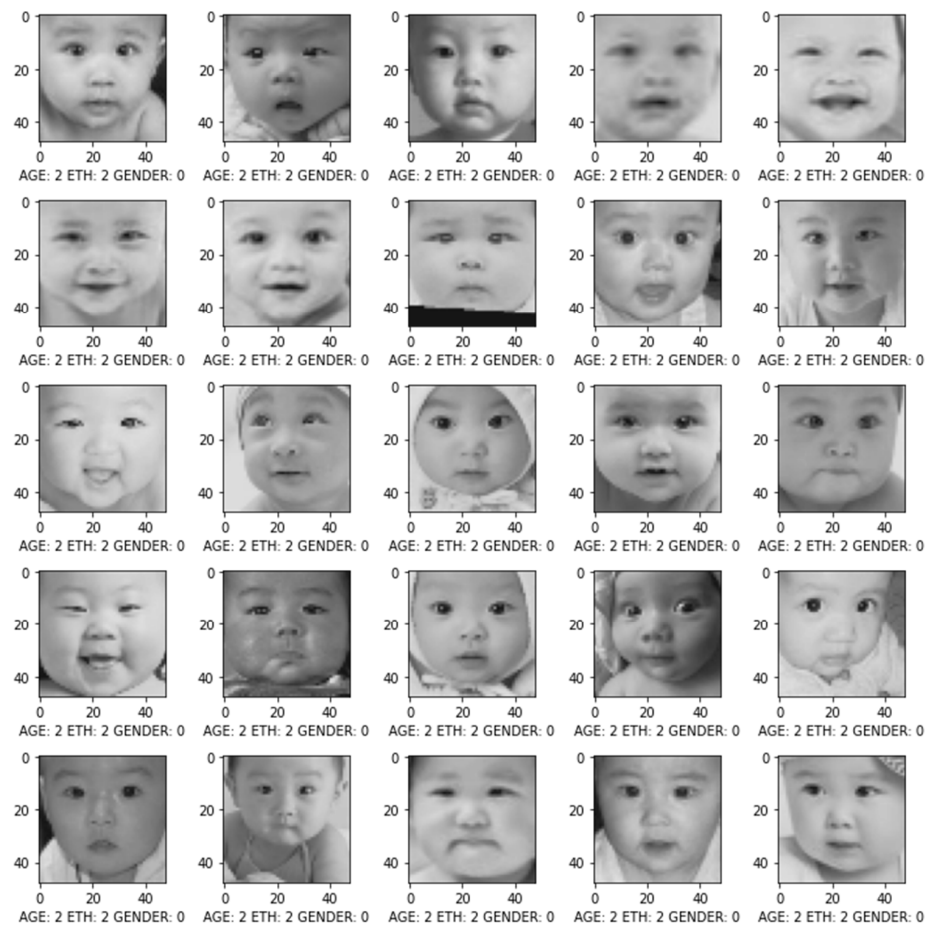


Figure 23. Ethnicity Classification for Dataset 2.

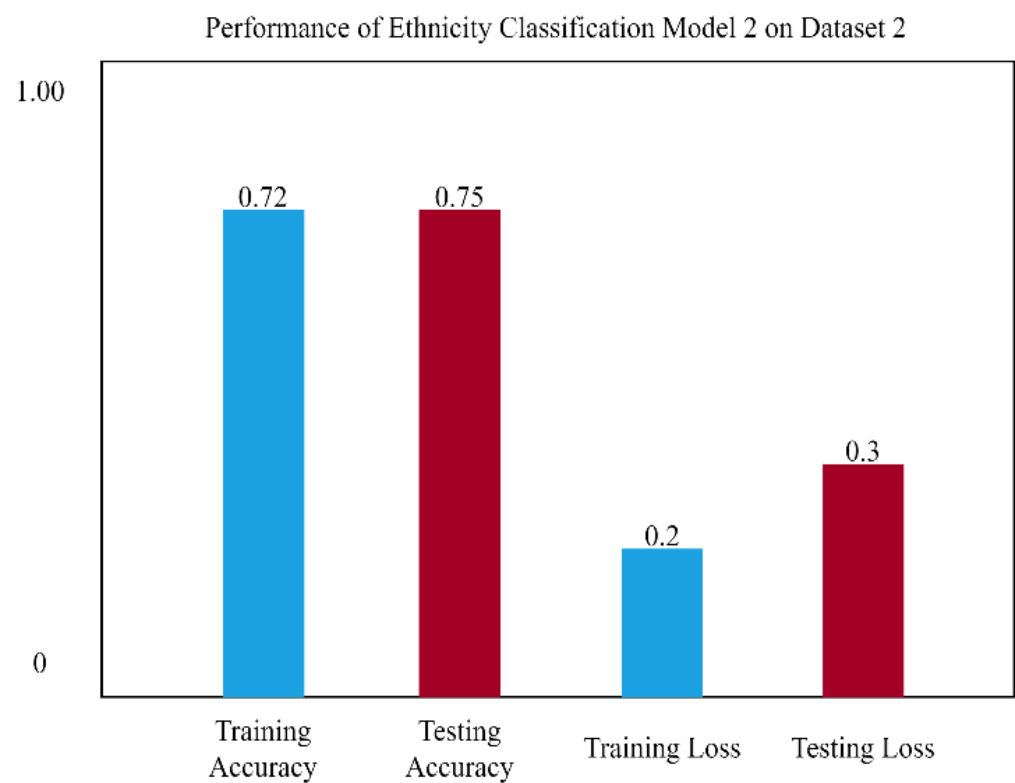


Figure 24. Model 2 Performance for Ethnicity Classification.

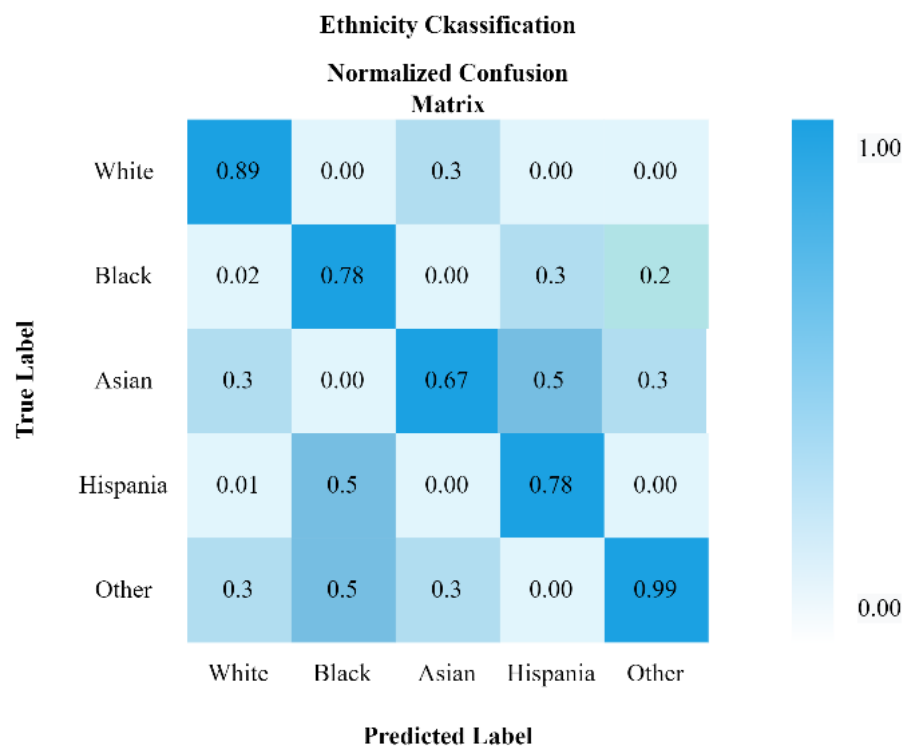


Figure 25. Confusion Matrix of Model 2 Performance for Ethnicity Classification.

10-Fold Cross-Validation

Table 3 shows the results of performing a 10-fold cross-validation for Model A and Model B with respect to their training accuracy and testing accuracy. For Model A, which is focused on gender classification, the training accuracy and testing accuracy are reported

for each fold. Each fold represents a different partitioning of the dataset, where the model is trained on a subset of the data (training accuracy) and evaluated on a separate subset (testing accuracy). The values range from 0 to 1, with higher values indicating better performance. The average values are also calculated to summarize the overall performance of the model across all folds. In this case, the average training accuracy is 0.82 and the average testing accuracy is 0.80. In Table 4 for Model B, which is focused on ethnicity classification, the same pattern applies. The training accuracy and testing accuracy are reported for each fold, representing the performance of the model on different partitions of the dataset. The average training accuracy for Model B is 0.73, and the average testing accuracy is 0.70. These results provide insights into how well the models perform in classifying gender and ethnicity based on the given dataset. The training accuracy indicates how well the model fits the training data, while the testing accuracy provides an estimate of how well the model generalizes to new, unseen data. The average values give an overall assessment of the model's performance across different partitions of the dataset.

Table 3. Model A—Gender Classification.

Fold	Training Accuracy	Testing Accuracy
1	0.82	0.80
2	0.84	0.81
3	0.81	0.79
4	0.83	0.82
5	0.85	0.83
6	0.82	0.80
7	0.80	0.78
8	0.83	0.82
9	0.85	0.83
10	0.81	0.79
Average	0.82	0.80

Table 4. Model B—Ethnicity Classification.

Fold	Training Accuracy	Testing Accuracy
1	0.75	0.72
2	0.74	0.71
3	0.72	0.70
4	0.73	0.70
5	0.75	0.72
6	0.73	0.70
7	0.71	0.68
8	0.74	0.71
9	0.76	0.73
10	0.72	0.69
Average	0.73	0.70

In Table 5, “Model 1” and “Model 2” refer to the two different models used for gender and ethnicity classification, respectively. “Dataset 1” and “Dataset 2” represent the two different datasets used in the experiments.

Table 5. Results and Performance Variation with Augmentation Strategies.

Model	Dataset	Augmentation Strategies	Training Accuracy	Testing Accuracy	Training Loss	Testing Loss
Model 1	Dataset 1	Without Augmentation	0.82	0.80	0.34	0.37
Model 1	Dataset 1	With Augmentation	0.86	0.84	0.20	0.23
Model 2	Dataset 1	Without Augmentation	0.72	0.70	0.80	0.75

Table 5. *Cont.*

Model	Dataset	Augmentation Strategies	Training Accuracy	Testing Accuracy	Training Loss	Testing Loss
Model 2	Dataset 1	With Augmentation	0.74	0.72	0.68	0.71
Model 1	Dataset 2	Without Augmentation	0.72	0.75	0.20	0.40
Model 1	Dataset 2	With Augmentation	0.75	0.77	0.18	0.35
Model 2	Dataset 2	Without Augmentation	0.72	0.75	0.20	0.30
Model 2	Dataset 2	With Augmentation	0.75	0.78	0.15	0.27

Table 5 compares the performance of the models with and without augmentation strategies. The augmentation strategies were applied to enhance the training datasets by introducing variations in facial attributes, such as hairstyles, glasses, and makeup.

The metrics reported include training accuracy, testing accuracy, training loss, and testing loss. It can be observed that, in most cases, the models with augmentation strategies achieved higher accuracy and lower loss compared with the models without augmentation. This indicates that augmentation strategies have a positive impact on the performance of the models by enriching the training datasets with diverse facial variations.

These results highlight the effectiveness of augmentation strategies in improving the accuracy and robustness of the gender and ethnicity classification models.

4.5. Comparative Analysis

Since the early days of computer vision and pattern recognition, face analysis has consistently been among the top research priorities in the field. A person's ethnicity, which can be gleaned from their face together with their gender and age, is one of the characteristics that stays with them the longest and has the greatest impact on their life. Therefore, classifying people into different categories only according to their ages and genders would only serve to further confuse the issue and create results that are not reliable. Consequently, classifying people according to their ethnicity as a fundamental part of the video surveillance systems used at checkpoints is an excellent concept. This categorization statement may also be helpful in picture-search queries. In this context, prior knowledge of race helps to simplify results by restricting the search pool; thus, having this information on hand can be quite beneficial. Two CNN models, Model A and Model B, were applied on two datasets i.e., MORPH and FERET, to classify ethnicity. Model A on Dataset 1 showed 0.84 training and 0.86 testing accuracy with 0.34 training and 0.37 testing loss. Model 2 showed 0.70 and 0.74 training and testing accuracy, respectively, and 0.80 training loss and 0.75 testing loss on Dataset 1. Model 1 on Dataset 2 showed 0.72 training accuracy, 0.75 testing accuracy, 0.2 training loss and 0.4 testing loss. Model 2 on Dataset 2 showed 0.72 training and 0.75 testing accuracy with 0.2 training loss and 0.3 testing loss.

Table 6 shows the comparative analysis of the current study with the state-of-the-art research conducted previously.

Table 6. Comparative Analysis of the Current Study with State-of-the-Art Research.

References	Techniques	Dataset	Accuracy
[32]	CNN	MORPH	69%
[33]	Deep Neural Network	FERET	70%
[34]	CNN	MORPH	71.3%
[35]	CNN	MORPH	70.89%
[36]	Learning-based Classifiers	FERET	88%
[37]	CNN	FERET	82%
[39]	Learning-based classifiers	FERET	85%

Table 6. Cont.

References	Techniques	Dataset	Accuracy
[40]	CNN	FERET	0.68 gain
This proposed model	Model ACNN Model BCNN	MORPH, FERET	85%, 86%

5. Conclusions

Most studies in the field of face biometrics focus on recognizing a person from a photograph; however, there are other, more subtle biometrics that are just as essential, such as the person's age, gender, ethnicity, or even their emotional state. The classification of people based on ethnicity is becoming an increasingly important area of research as it is increasingly useful in various contexts. This paper presents an overview of recent improvements in classifying ethnicity using efficient models of CNNs using simply the central area of the face. In addition to this, the paper analyzes the distinction between findings from two distinct CNN models. The MORPH and FERET datasets were utilized to test the suggested models with holdout testing. It is important to highlight that the findings in this research were created by only using the middle region of the face, which saves both time and resources. The accuracy of Model A was 85% for the classification into four classes, whereas the accuracy of Model B was 86% for the classification into four classes. In the future, a real-time dataset can be used with reinforcement learning models to make the system more efficient.

Funding: This research received no external funding.

Institutional Review Board Statement: Not Applicable.

Informed Consent Statement: Not Applicable.

Data Availability Statement: The data used to support the findings of this study are included within this article.

Acknowledgments: This study could not be started, nor achieved without the continued cooperation of the colleagues from the Department of Computer and Information Technology, Jubail Industrial College, Royal Commission for Jubail and Yanbu, Saudi Arabia.

Conflicts of Interest: The author declares no conflict of interest.

References

1. Akbar, M.; Furqan, K.M.Y.; Yaseen, H. Evaluation of Ethnicity and Issues of Political Development in Punjab, Pakistan. *Glob. Polit. Rev.* **2020**, *V*, 57–64. [\[CrossRef\]](#)
2. Anwar, I.; Islam, N.U. Learned features are better for ethnicity classification. *Cybern. Inf. Technol.* **2017**, *17*, 152–164. [\[CrossRef\]](#)
3. Belcar, D.; Grd, P.; Tomićić, I. Automatic Ethnicity Classification from Middle Part of the Face Using Convolutional Neural Networks. *Informatics* **2022**, *9*, 18. [\[CrossRef\]](#)
4. Masood, S.; Gupta, S.; Wajid, A.; Gupta, S.; Ahmed, M. Prediction of human ethnicity from facial images using neural networks. *Adv. Intell. Syst. Comput.* **2018**, *542*, 217–226.
5. El Khiyari, H.; Wechsler, H. Face Verification Subject to Varying (Age, Ethnicity, and Gender) Demographics Using Deep Learning. *J. Biom. Biostat.* **2016**, *7*, 11. [\[CrossRef\]](#)
6. Sulaiman, M.A.; Kocher, I.S. A systematic review on Evaluation of Driver Fatigue Monitoring Systems based on Existing Face/Eyes Detection Algorithms. *Acad. J. Nawroz Univ. (AJNU)* **2022**, *11*, 57–72. [\[CrossRef\]](#)
7. Ghani, M.U.; Alam, T.M.; Jaskani, F.H. Comparison of Classification Models for Early Prediction of Breast Cancer. In Proceedings of the 2019 International Conference on Innovative Computing (ICIC), Lahore, Pakistan, 1–2 November 2019; pp. 1–6. [\[CrossRef\]](#)
8. Wang, W.; He, F.; Zhao, Q. Facial ethnicity classification with deep convolutional neural networks. In Proceedings of the 11th Chinese Conference, CCBP 2016, Chengdu, China, 14–16 October 2016; Springer International Publishing: Berlin/Heidelberg, Germany, 2016; pp. 176–185.
9. Chirco, P.; Buchanan, T.M. Dark faces in white spaces: The effects of skin tone, race, ethnicity, and intergroup preferences on interpersonal judgments and voting behavior. *Anal. Soc. Issues Public Policy* **2022**, *22*, 427–447. [\[CrossRef\]](#)
10. Greco, A.; Percannella, G.; Vento, M.; Vigilante, V. Benchmarking deep network architectures for ethnicity recognition using a new large face dataset. *Mach. Vis. Appl.* **2020**, *31*, 67. [\[CrossRef\]](#)

11. SteelFisher, G.K.; Findling, M.G.; Bleich, S.N.; Casey, L.S.; Blendon, R.J.; Benson, J.M.; Sayde, J.M.; Miller, C. Gender discrimination in the United States: Experiences of women. *Health Serv. Res.* **2019**, *54*, 1442–1453. [\[CrossRef\]](#) [\[PubMed\]](#)
12. Deshpande, K.V.; Pan, S.; Foulds, J.R. Mitigating Demographic Bias in AI-based Resume Filtering. In Proceedings of the Adjunct Publication of the 28th ACM Conference on User Modeling, Adaptation and Personalization, Genoa, Italy, 14–17 July 2020; pp. 268–275.
13. Khan, K.; Khan, R.U.; Ali, J.; Uddin, I.; Khan, S.; Roh, B.H. Race classification using deep learning. *Comput. Mater. Contin.* **2021**, *68*, 3483–3498. [\[CrossRef\]](#)
14. Vicente-Samper, J.M.; Vila-Navarro, E.; Sabater-Navarro, J.M. Data acquisition devices towards a system for monitoring sensory processing disorders. *IEEE Access* **2020**, *8*, 183596–183605. [\[CrossRef\]](#)
15. Rawan, B.; Bibi, N. Construction of Advertisements in Pakistan: How far Television Commercials Conform to Social Values and Professional Code of Conduct? *Glob. Reg. Rev.* **2019**, *IV*, 22–31. [\[CrossRef\]](#)
16. Chitale, V.S.; Sciences, M. A Novel Indexing Method Using Hierarchical Classification for Face-Image. Doctoral Dissertation, Auckland University of Technology, Auckland, New Zealand, 2020.
17. Sajid, M.; Shafique, T.; Manzoor, S.; Iqbal, F.; Talal, H.; Samad Qureshi, U.; Riaz, I. Demographic-assisted age-invariant face recognition and retrieval. *Symmetry* **2018**, *10*, 148. [\[CrossRef\]](#)
18. Fuad, M.T.H.; Fime, A.A.; Sikder, D.; Iftae, M.A.R.; Rabbi, J.; Al-Rakhami, M.S.; Gumaei, A.; Sen, O.; Fuad, M.; Islam, M.N. Recent advances in deep learning techniques for face recognition. *IEEE Access* **2021**, *9*, 99112–99142. [\[CrossRef\]](#)
19. Saragih, R.E.; To, Q.H. A Survey of Face Recognition based on Convolutional Neural Network. *Indones. J. Inf. Syst.* **2022**, *4*, 122–139.
20. Albdairi, A.J.A.; Xiao, Z.; Alghaili, M.; Huang, C. Identifying Ethnicities of People through Face Recognition: A Deep CNN Approach. *Sci. Program.* **2020**, *2020*, 6385281. [\[CrossRef\]](#)
21. Agbo-Ajala, O.; Viriri, S. Deeply Learned Classifiers for Age and Gender Predictions of Unfiltered Faces. *Sci. World J.* **2020**, *2020*, 1289408. [\[CrossRef\]](#) [\[PubMed\]](#)
22. Khan, K.; Attique, M.; Khan, R.U.; Syed, I.; Chung, T.S. A multi-task framework for facial attributes classification through end-to-end face parsing and deep convolutional neural networks. *Sensors* **2020**, *20*, 328. [\[CrossRef\]](#)
23. Angulu, R.; Tapamo, J.R.; Adewumi, A.O. Age estimation via face images: A survey, 2018, no. 1. *EURASIP J. Image Video Process.* **2018**, *2018*, 42. [\[CrossRef\]](#)
24. Ngan, M.; Grother, P. *Face Recognition Vendor Test (FRVT)—Performance of Automated Gender Classification Algorithms*; US Department of Commerce, National Institute of Standards and Technology: Gaithersburg, MD, USA, 2015.
25. Atallah, R.R.; Kamsin, A.; Ismail, M.A.; Abdelrahman, S.A.; Zerdoumi, S. Face Recognition and Age Estimation Implications of Changes in Facial Features: A Critical Review Study. *IEEE Access* **2018**, *6*, 28290–28304. [\[CrossRef\]](#)
26. Terhöst, P.; Kolf, J.N.; Damer, N.; Kirchbuchner, F.; Kuijper, A. Post-comparison mitigation of demographic bias in face recognition using fair score normalization. *Pattern Recognit. Lett.* **2020**, *140*, 332–338. [\[CrossRef\]](#)
27. Kärkkäinen, K.; Joo, J. FairFace: Face Attribute Dataset for Balanced Race, Gender, and Age. *arXiv* **2019**, arXiv:1908.04913.
28. Vo, T.; Nguyen, T.; Le, C.T. Race recognition using deep convolutional neural networks. *Symmetry* **2018**, *10*, 564. [\[CrossRef\]](#)
29. Mustapha, M.F.; Mohamad, N.M.; Osman, G.; Hamid, S.H.A. Age group classification using Convolutional Neural Network (CNN). *J. Phys. Conf. Ser.* **2021**, *2084*, 012028. [\[CrossRef\]](#)
30. Ahmed, M.A.; Choudhury, R.D.; Kashyap, K. Race estimation with deep networks. *J. King Saud Univ.-Comput. Inf. Sci.* **2020**, *34*, 4579–4591. [\[CrossRef\]](#)
31. Badrulhisham, N.A.S.; Mangshor, N.N.A. Emotion Recognition Using Convolutional Neural Network (CNN). *J. Phys. Conf. Ser.* **2021**, *1962*, 1748–1765. [\[CrossRef\]](#)
32. Meenakshi, S.; Jothi, M.S.; Murugan, D. Face recognition using deep neural network across variations in pose and illumination. *Int. J. Recent Technol. Eng.* **2019**, *8*, 289–292.
33. Boussaad, L.; Boucetta, A. An effective component-based age-invariant face recognition using Discriminant Correlation Analysis. *J. King Saud Univ.-Comput. Inf. Sci.* **2020**, *34*, 1739–1747. [\[CrossRef\]](#)
34. Sharmila; Sharma, R.; Kumar, D.; Puranik, V.; Gautham, K. Performance Analysis of Human Face Recognition Techniques. In Proceedings of the 2019 4th International Conference on Internet of Things: Smart Innovation and Usages (IoT-SIU), Ghaziabad, India, 18–19 April 2019; pp. 1–4.
35. Rubeena; Kavitha, E. Sketch face Recognition using Deep Learning. In Proceedings of the 2021 5th International Conference on Electronics, Communication and Aerospace Technology (ICECA), Coimbatore, India, 2–4 December 2021; Volume 1, pp. 928–930.
36. Sun, H.; Grishman, R. Lexicalized dependency paths based supervised learning for relation extraction. *Comput. Syst. Sci. Eng.* **2022**, *43*, 861–870. [\[CrossRef\]](#)
37. Sun, H.; Grishman, R. Employing lexicalized dependency paths for active learning of relation extraction. *Intell. Autom. Soft Comput.* **2022**, *34*, 1415–1423. [\[CrossRef\]](#)
38. Rahim, A.; Zhong, Y.; Ahmad, T. A Deep Learning-Based Intelligent Face Recognition Method in the Internet of Home Things for Security Applications. *J. Human Univ. (Nat. Sci.)* **2022**, *49*, 10. [\[CrossRef\]](#)

39. Mohammad, A.S.; Al-Ani, J.A. Towards ethnicity detection using learning based classifiers. In Proceedings of the 2017 9th Computer Science and Electronic Engineering (CEECE), Colchester, UK, 27–29 September 2017; pp. 219–224. [[CrossRef](#)]
40. Mohammad, A.S.; Al-Ani, J.A. Convolutional Neural Network for Ethnicity Classification using Ocular Region in Mobile Environment. In Proceedings of the 2018 10th Computer Science and Electronic Engineering (CEECE), Colchester, UK, 19–21 September 2018; pp. 293–298. [[CrossRef](#)]

Disclaimer/Publisher’s Note: The statements, opinions and data contained in all publications are solely those of the individual author(s) and contributor(s) and not of MDPI and/or the editor(s). MDPI and/or the editor(s) disclaim responsibility for any injury to people or property resulting from any ideas, methods, instructions or products referred to in the content.