Facial Features Based Human Age, Gender And Ethnicity Identification System

A PROJECT REPORT

Submitted by

SHIVAGANESAN K	**(180701114)**
SHRUTHI E	**(180701116)**
YASHWIN VASANTH S	**(180701154)**

in partial fulfillment for the award of the degree

of

BACHELOR OF ENGINEERING

IN

ELECTRONICS AND COMMUNICATION ENGINEERING

SRI VENKATESWARA COLLEGE OF ENGINEERING
(An Autonomous Institution; Affiliated to Anna University, Chennai-600025)

ANNA UNIVERSITY :: CHENNAI 600 025

JUNE 2022

SRI VENKATESWARA COLLEGE OF ENGINEERING
(An Autonomous Institution; Affiliated to Anna University, Chennai-600025)

ANNA UNIVERSITY :: CHENNAI 600 025

BONAFIDE CERTIFICATE

Certified that this project report **"FACIAL FEATURES BASED HUMAN AGE,GENDER AND ETHNICITY IDENTIFICATION SYSTEM"** is the bonafide work of **"SHIVAGANESAN (180701114), SHRUTHI E (180701116), YASHWIN VASANTH S (180701154)**" **who carried out the project work under my supervision.**

SIGNATURE	**SIGNATURE**
Dr. S. MUTHUKUMAR, M.E., Ph.D.	Dr. S. VELUCHAMY, M.E., Ph.D.
HEAD OF THE DEPARTMENT	**SUPERVISOR**
	Assistant Professor,
Department of Electronics and	Department of Electronics and
Communication Engineering.	Communication Engineering.

Submitted for the project viva-voce examination held on ___________________
.

INTERNAL EXAMINER **EXTERNAL EXAMINER**

ABSTRACT

Since the rise of social platforms and social media, automatic age, ethnicity, and gender classification have been relevant to a growing number of applications. Nevertheless, the performance of existing methods on real-world images is still significantly lacking, especially when compared to the tremendous leaps in performance recently reported for the related task of face recognition. As a result, we show that learning representations using deep-convolution neural networks (CNN) can result in a considerable improvement in performance on these tasks. To this end, we propose a simple convolutional net design that can be employed even when there is a small amount of learning data. For image-based gender, age, and ethnicity estimates, deep neural networks with pre-trained weights are used. VGG is used to investigate transfer learning. To increase prediction accuracy, examine the effects of modifications in various design schemes and training settings on pre-trained models. Finally, a hierarchy of deep CNNs is explored, which first classifies participants by gender and then predicts age and ethnicity using separate models.

Keywords: Deep Learning, Convolution Neural Network, Deep Transfer Learning, VGG-16, RESNET-50

ACKNOWLEDGEMENT

We thank our Principal **Dr. S. Ganesh Vaidyanathan,** Sri Venkateswara College of Engineering for being the source of inspiration throughout our study in this college.

We express our sincere thanks to **Dr. S. Muthukumar,** Head of the Department, Electronics and Communication Engineering for his encouragement accorded to carry out this project.

With profound respect, we express our deep sense of gratitude and sincere thanks to our guide, **Dr. S. VELUCHAMY,** Assistant Professor for their valuable guidance and suggestions throughout this project.

We are also thankful to our project coordinators **Mrs. K.S. Subashini, Assistant Professor and Mr.S.P.Sivagnana Subramanian, Assistant Professor and and Dr.S.Veluchamy, Assistant Professor** for their continual support and assistance.

We thank our family and friends for their support and encouragement throughout the course of our graduate studies.

SHIVAGANESAN K (180701114)

SHRUTHI E (180701116)

YASHWIN VASANTH S (180701154)

TABLE OF CONTENTS

LIST OF FIGURES

LIST OF TABLES

LIST OF ABBREVIATIONS

CNN	Convolutional Neural Network
FRS	Face Recognition System
GPU	Graphics Processing Unit
HCI	Human-computer interaction
HOG	Histogram of Oriented Gradients
RGB	Red Green Blue
ReLU	Rectification Layer
VGG	Visual Geometry Group
LBP	Local Binary Patterns
IP	Image Processing
DL	Deep Learning
ML	Machine Learning
AI	Artificial Intelligence
FER	Facial Expression Recognition

CHAPTER 1

INTRODUCTION

1.1 GENERAL

The face is crucial for the identity of persons as a result of it contains a lot of info concerning personal characteristics. Thus, the face image is vital for many bioscience systems. The face image provides countless helpful info, as well as the person. Ås identity, gender, ethnicity, age, emotional expression, etc. He distinguishing proof normal for confronting pictures has been very much investigated in certifiable applications [1], including travel permits and driver licenses. Regardless of the broad investigation of individual distinguishing proof from confronting pictures, there is just a restricted measure of research [2] on the best way to precisely gauge and utilize the statistical data contained in face images such as age, gender, and race. For some viable applications, depending on people to supply statistic data from confronting pictures isn't achievable. Henceforth, there has been a developing enthusiasm for programmed extraction of statistic data from confronting pictures. As of late, a few applications that endeavor statistic properties have risen. “These applications embrace access management [3], re- identification in police work videos [4], the integrity of face pictures in social media [5], intelligent advertising, and human-computer interaction, law enforcement of the system. In daily life, human beings perform gender and face recognition as well as estimate the age of their peers naturally and expeditiously [6]. Many studies from totally different backgrounds (face and head measure, psychological science, clinical studies, etc.) have tried to know however the method works. Specifically, various anthropometric examinations [7] have uncovered that huge facial morphology contrasts exist among the sexual orientation, the ethnicity, and the age gatherings. For instance, when considering the Sexual Dimorphism (Male/Female contrasts) [8], analysts have discovered that male faces, for the most part, have more conspicuous highlights than female appearances. Male faces more often than not have more protuberant noses, eyebrows, more conspicuous buttons and jaws. The temple is all the more in reverse slanting, and the separation between top-lip and nose-base is longer. [7] Have likewise shown that all the concerned anthropometric estimations of females are littler. In the investigation of the ethnic contrasts [9], specialists have discovered that contrasted with the North America Whites, Asians more often than not have more extensive faces and noses, far separated eyes, and display the best distinction in the anatomical orbital locales (around the eyes and the eyebrows). In the clinical examination announced in [10], Alphonse et al. have uncovered that Caucasians have altogether brought down foetal Fronto-Maxillary Facial Angle (FMFA) estimations than Asians In [7], sixteen anthropometric estimations have been perceived as essentially extraordinary amongst Asian and Caucasian countenances. When considering the face maturing [11], analysts have reasoned that the craniofacial development is the fundamental change in child and young faces, which brings about the re-measuring and redistribution of facial highlights. Amid this period, for the most part, the bigger is the age, the greater is the measure of the face. At the point when the craniofacial development stops at 18-20 years of age, the face form and surface changes turn into the prevailing changes. Youthful grown-ups tend to have a triangle formed a face with a little measure of wrinkles. Interestingly, old grown-ups are generally connected with a U-molded face with noteworthy wrinkles on the face. Other than the presence of these Soft-Biometric Traits [12] in the face, sexual orientation, ethnicity, and age are additionally connected in portraying the facial shape [7]. For instance, as indicated by the anthropometric examinations referred to over, the state of the nose is impacted by all the three delicate biometric characteristics. In human recognition, female faces look smoother and more youthful than male countenances, and the Asian faces normally look more youthful than Non- Asian appearances [13].

We use the Deep Learning technology to reduce erroneous predictions and boost accuracy. The number of hidden layers in a neural network is commonly referred to as "deep." Deep neural networks can have up to 150 hidden layers, whereas traditional neural networks only have 2-3. Large sets of labeled data and neural network topologies that learn features directly from the data without the requirement for manual feature extraction are used to train deep learning models. Convolution neural networks (CNN) are one of the most common types of deep neural networks, and we plan to use them as a basis architecture at first. The purpose of this research is to train and assess a convolution neural network for gender categorization and age estimate on images using a transfer learning strategy based on the Vgg network for Face Recognition System. All of the implementation parts for showing the output in image format as well as showing the output based on the system's accuracy have been accomplished.

1.2 FACE RECOGNITION SYSTEM

Vital tasks, admire face recognition, face following, countenance recognition, head cause estimation, will like precise facial point localization. While face detection is generally regarded as the starting point for all face analysis tasks, confront arrangement can be viewed as a vital and fundamental middle person venture for some, resulting face investigations that range from biometric acknowledgment to mental state understanding. Solid assignments may vary in the number and kind of the required facial focuses, and the way these focuses are utilized

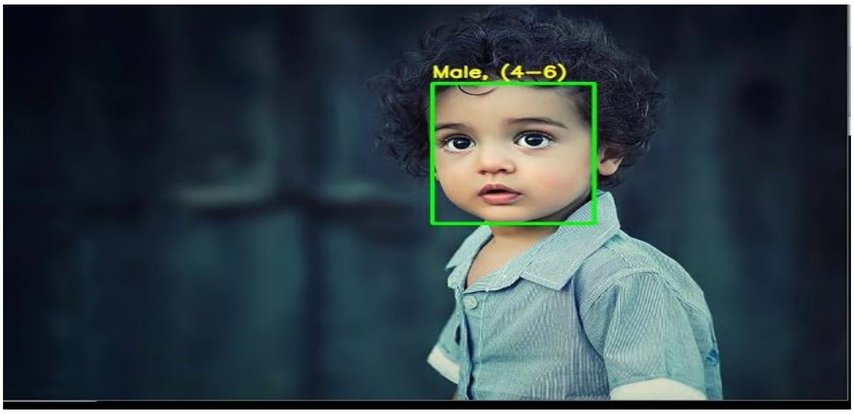

Fig 1.1 Face Recognized Image

In particular, the following factors have a significant influence on facial appearance and the states of local facial features:

- **Pose:** The looks of native facial features disagree greatly between totally different camera object poses (e.g., frontal, profile, top down), and a few facial elements comparable to the one aspect of the face contour, will even be utterly occluded in a very profile face.

- **Occlusion:** For face pictures captured in at liberty conditions, occlusion oft happens and brings nice challenges to face alignment. For example, the eyes could also be occluded by hair, sunglasses, or ametropia glasses with black frames.
- **Expression:** Some neighborhood facial highlights, for example, eyes and mouth are delicate to the difference in different articulations. For instance, giggling may make the eyes close totally, and generally disfigure the state of the mouth.
- **Illumination:** Lighting (differing in spectra, source dispersion, and power), may fundamentally change the presence of the entire face, and make the nitty gritty surfaces of some facial parts missing.

1.2.1 KINDS OF APPROACH USED FOR FACE RECOGNITION:

Generative ways: These methods build generative models for each the face form and look. They generally formulate face alignment as associate optimization downside to search out the form associated look parameters that generate an look model instance giving the simplest work four half dozen to the check face. Note that the facial look are often diagrammatic either by the full (warped) face or by the native image patches focused at the facial point.

Discriminative methods: These methods directly infer the target location from the facial look. this can be usually done by learning freelance native detector or regressor for every facial purpose and using a world form model to regularize their predictions, or by directly learning a vector regression operate to infer the complete face form, throughout that the form constraint is implicitly encoded.

1.3 GENDER CLASSIFICATION SYSTEM

After the extracting face region, Gender classification is to be performed. Facial attributes are derived by either appearance based method or geometric based methods. In appearance based method, the entire image is measured relatively with the local features which are equivalent to different parts of the face. In geometric based methods, the geometric features such as distance between eyes face, length and width are considered for gender estimation. After detecting the face, facial attributes have to be identified and extracted from the detected facial image. The classification of gender is performed by two main categories, namely Appearance based feature classification and Geometric based feature classification.

Appearance based feature classification: The name of the category 'Appearance based feature' itself entails that it furnishes the details of the features of the face appearance. Male skin is usually harder than the female skin. Female eyes and lips are not lighter than male eyes and lips. The luminance diversity around male and female eyes and lips are contrast. Female eyes and lips produce better luminance difference than male eyes and lips. The nose luminance diversity is also diverse from male to female. It is recognized and conventional that the intense and diverse levels of male and female are assorted from one human to another.

Geometric based feature classification: Geometric based feature classification method predominantly deals with the shape and size of the face and facial scraps. The geometric arrangement of male and female are different for each and every one. The shape and size of the facial scraps like eyes, nose, lips and eyebrows are dissimilar for male and female. Normally male have longer and bigger face than female. The female picture has round arcs. The head is curved, cheeks and the neck form is rounded quite than flatten. After the extraction of facial features, the procedure arrives to the end stages. It predicts that the male and female facial features contain apparent assets. These apparent assets provide the relative ratio and based on these ratios, the threshold value of male and female is determined. These values are then transferred to calculate the further steps of automatic human facial age estimation.

1.4 AGE ESTIMATION SYSTEM

Age estimation is used to estimate the age of an input facial image. The visualizes the stages of automatic facial age estimation method. The output of the extracted face region of the face detection method is treated as input for this method. After fetching the face region, certain features are extracted to estimate age. The age based feature extraction process is further categorized into various groups as Anthropometric Model, Active Appearance Model, Aging Pattern Subspace, Age Manifold, and Appearance based Feature Model.

- **Anthropometric model:** Anthropometry model is used to measure the dimension and proportions of human face. It utilizes the concept of cranio-facial theory and generated fiduciary points. These measurements are used to estimate the age of a person Extracted Face Region Feature Extraction Age Estimation 10 from Infancy to Adulthood. The constraint of this model is that it provides better performance for the images of younger people and also for smaller databases. This geometrical model depends on the value of the distance and ratio of the two dimensional image which are very sensitive.
- **Active Appearance Model (AAM):** AAM is a statistical representation of Age Estimation by generating both shape and appearance of face by utilizing several attribute points. The prime advantage of this model is that it deals well for images of all age group and does not have any constraints like anthropometric model. Even though, it produces high performance for age estimation, it is mainly depends on the time factor for fitting several points in a face.
- **Aging Pattern Subspace:** Aging Pattern Subspace method applied on individual of different ages and it is sorted according to the growth. If the training dataset contains all age group samples, then it is considered as a complete data set or else it is said to be an incomplete data set. In case of an incomplete dataset, then the missing age sample was reconstructed by using an algorithm. (For ex., if a sample’s Face is missing for the particular age, then the previous age facial images are utilized to form an image of that particular age.) The test face is compared with dataset and determines the age of a person according to samples

in the data set. The main difficulty of this method is to collect all photos of different age of a person.

- **Age manifold Method:** The Manifold embedding technique is used to acquire a low dimensional aging trend for many facial images of the same age. Age manifold Method leads the possible way to learn the common pattern of aging for more than one person at different ages instead of providing the specific aging pattern for each person's to represent the age. An Individual can have several face images in the same age or in a given age range. Therefore, this model is more flexible than AGES model, and it is easier to accumulate a larger number of samples (facial images) and generate a larger database than the other methods. The only requirement of this model is that the sample size should be large enough to engage so that the pattern can be taught with a statistical sufficiency to produce several facial images.
- **Appearance feature model:** Appearance feature model concentrates on the extraction of age related features such as skin texture showing wrinkles and shape analysis for determination of age. The result of this method is regarded as a portrayal when the face is not affected by any of the biological and natural disturbances of an individual. For instance, the facial scar due to accidents or any surgeries and the natural occurrences as the birth defect on the face will be difficult to analyses in this model. These are the various methods used to estimate age of a person in the form of accurate age or range of age. Mostly all the methods produce the output only as range of age because the system can't predict accurate age due to biological and natural factors. Age estimation plays a major role in the real world application which motivates to do the research work for the determination of age from given facial images.

1.5 RACE ESTIMATION SYSTEM

Race alludes to characterizations of people into moderately vast and unmistakable populaces or gatherings frequently in view of elements, for example, appearance in light of heritable phenotypical qualities or geographic heritage, yet additionally regularly affected by and associated with attributes, for example, culture, race and financial status. As a natural term, race signifies hereditarily disparate human populaces that can be set apart by normal phenotypic attributes. The statistic highlights, for example, race and sex, are engaged with human face personality acknowledgment. People are better at perceiving countenances of their own race than appearances of other race. In that, the face has just been identified by some earlier advance. In the first place, pre-preparing is connected to standardize the info confront (e.g., as for estimate, area, introduction, and lighting). At that point, include extraction is performed to speak to the face by a minimized, more discriminative arrangement of highlights. At long last, a choice instrument is utilized (e.g., classifier) to distinguish the race of the info confront.

1.6 OBJECTIVE OF OUR WORK

Accurate age forecasting is one of the main significant concerns in human communication. Age identification is an important measurement of human-computer interaction. The aim of this research is to estimate age and gender and race assortment based on facial features of human being and to increase the accuracy of those detections. The main objectives of the proposed work includes the following

- To extract face region from an input image and also upsurge the performance of face detection effectively.
- To categorize the gender from an obtained facial portion based on facial features
- To estimate the age as well estimate the race of the person from the derived facial features.
- To test and train the proposed system with use of UTK data to predict the result based on the deep learning concept of the system.

1.7 SCOPE OF OUR PROJECT

The goal of this research is to improve the efficiency of our proposed estimation system from the constructed facial images. Using facial features, create an autonomous age approximation, gender, and ethnicity prediction system. We use the Deep Learning technology and transfer learning to reduce erroneous predictions and boost accuracy. This work is estimated to identify the age, gender and race attribute of the system.

1.8 DESIGN CHALLENGES

- It is perceivable that our system may wrongly classify due to extremely challenging viewing conditions of the internet images.
- Gender estimations might occasionally be wildly off for infant age group.
- Gender and age estimation might occasionally be wronlgy classified due to extra accessories such as masks, specs, etc.
- Identifying race, which is a major physical feature in humans, is a challenging task owing much to the lack of a concrete definition of race and the diversity of population across the globe

1.9 APPLICATIONS OF THE SYSTEM

- **Human-computer interaction systems:** More sophisticated human computer interaction framework can be built if they are capable to discover a human's attribute such as gender. A simple situation would be a robot interacting with a human, it would require the information about gender to interact with the human correctly. (e.g. as Mr. or Ms.)
- **Surveillance systems:** In smart surveillance systems, it can aid in restricting areas to particular gender, in a train coach or hostel. Content-based indexing and searching: With the significant use of electronic devices such as cameras, a large number of photos and videos are being produced. Indexing or annotating records such as the variety of humans in the picture or video, their age and gender will become easier with computerized structures using computer vision.

- **Biometrics:** In biometric scheme, especially in face recognition, if the gender of the person is known first, then the time for seeking the face database can be reduced. The one more added advantage is, it can be used for improving face recognition by means of separate face recognizers for each gender to achieve high recognition rates.
- **Targeted advertising:** A digital billboard device is used in existing commercials on flat-panel displays. Targeted advertising is used to display commercials applicable to the person looking at the billboard totally based on attributes such as gender.

CHAPTER 2

LITERATURE REVIEW

A. J. O'Toole, et al., "Demographic effects on estimates of automatic face recognition performance". The intended applications of automatic face recognition systems include venues that vary widely in demographic diversity. Formal evaluations of algorithms do not commonly consider the effects of population diversity on performance. We document the effects of racial and gender demographics on estimates of the accuracy of algorithms that match identity in pairs of face images. In particular, we focus on the effects of the "background" population distribution of non-matched identities against which identity matches are compared. The algorithm we tested was created by fusing three of the top performers from a recent US Government competition. First, we demonstrate the variability of algorithm performance estimates when the population of non-matched identities was demographically "yoked" by race and/or gender (i.e., "yoking" constrains non-matched pairs to be of the same race or gender). We also report differences in the match threshold required to obtain a false alarm rate of .001 when demographic controls on the non-matched identity pairs varied. In a second experiment, we explored the effect on algorithm performance of progressively increasing population diversity. We found systematic, but non-general, effects when the balance between majority and minority populations of non-matched identities shifted. Third, we show that identity match accuracy differs substantially when the non-match identity population varied by race. Finally, we demonstrate the impact on performance when the non-match distribution consists of faces chosen to resemble a target face. The results from all experiments indicate the importance of the demographic composition and modelling of the background population in predicting the accuracy of face recognition algorithms.

Cavazos et al., "Accuracy comparison across face recognition algorithms: ".Previous generations of face recognition algorithms differ in accuracy for images of different races (race bias). Here, we present the possible underlying factors (data-driven and scenario modelling) and methodological considerations for assessing race bias in algorithms. We discuss data-driven factors (e.g., image quality, image population statistics, and algorithm architecture), and scenario modelling factors that consider the role of the "user" of the algorithm (e.g., threshold decisions and demographic constraints). To illustrate how these issues apply, we present data from four face recognition algorithms (a previous-generation algorithm and three deep convolutional neural networks, DCNNs) for East Asian and Caucasian faces. First, dataset difficulty affected both overall recognition accuracy and race bias, such that race bias increased with item difficulty. Second, for all four algorithms, the degree of bias varied depending on the identification decision threshold. To achieve equal false accept rates (FARs), East Asian faces required higher identification thresholds than Caucasian faces, for all algorithms. Third, demographic constraints on the formulation of the distributions used in the test, impacted estimates of algorithm accuracy. We conclude that race bias needs to be measured for individual applications and we provide a checklist for measuring this bias in face recognition algorithms.

Hosoi et al. [2] have integrated the Gabor wavelet features and retina sampling for their work. These features were then used with the Support Vector Machines (SVM) classifier. This approach has used three categories: Asian, African and European. And the accuracy achieved for each category is: 96%, 94% and 93% respectively. However their approach seemed to have issues when considering other ethnicities

Lagree, S et al., "Predicting ethnicity and gender from iris texture". Previous researchers have reported success in predicting ethnicity and in predicting gender from features of the iris texture. This paper is the first to consider both problems using similar experimental approaches. Contributions of this work include greater accuracy than previous work on predicting ethnicity from iris texture, empirical evidence that suggests that gender prediction is harder than ethnicity prediction, and empirical evidence that ethnicity prediction is more difficult for females than for males.

Lu et al. has proposed ethnicity classification algorithm in which image of the faces were examined at multiple scales .The Linear Discriminant Analysis (LDA) scheme is used for input face images to improve the classification result. The accuracy of the performance of this approach is 96.3% on the database of 2,630 sample images of 263 subjects. However, the dataset considered in this work consisted only of two classes i.e. Asian and non-Asian.

P. Viola and M. Jones has provided the efficient and rapid method of detecting face in input image. This is a novel approach which uses Adaboost classifier. This has high detection rate with very less computation time on the dataset consisting of images under varying condition like illumination, pose, color, camera variation; etc. This algorithm is used in our approach to detect face in the image which will be later processed further

S. Md. Mansoor has used Viola Jones algorithm for face detection problem. After the detection of face, various features namely skin color; lip color and normalized forehead area were extracted from the image. This classification problem has used the Yale, FERET dataset of Mongolian, Caucasian and Negroid images. The overall accuracy achieved in this work with these features was 81%.

Vangara, K et al., "Characterizing the variability in face recognition accuracy relative to race" Many recent news headlines have labelled face recognition technology as "biased" or "racist". We report on a methodical investigation into differences in face recognition accuracy between African-American and Caucasian image cohorts of the MORPH dataset. We find that, for all four matchers considered, the impostor and the genuine distributions are statistically significantly different between cohorts. For a fixed decision threshold, the African-American image cohort has a higher false match rate and a lower false non-match rate. ROC curves compare verification rates at the same false match rate, but the different cohorts achieve the same false match rate at different thresholds. This means that ROC comparisons are not relevant to operational scenarios that use a fixed decision threshold. We show that, for the ResNet matcher, the two cohorts have approximately equal separation of impostor and genuine distributions. Using ICAO compliance as a standard of image quality, we find that the initial image cohorts have unequal rates of good quality images. The ICAO-compliant subsets of the original image cohorts show improved accuracy, with the main effect being to reducing the low-similarity tail of the genuine distributions.

Terhörst, P et al., "Comparison-level mitigation of ethnic bias in face recognition". Current face recognition systems achieve high performance on several benchmark tests. Despite this progress, recent works showed that these systems are strongly biased against demographic sub-groups. Previous works introduced approaches that aim at learning less biased representations. However, applying these approaches in real applications requires a complete replacement of the templates in the database. This replacement procedure further requires that a face image of each enrolled individual is stored as well. In this work, we propose the first bias-mitigating solution that works on the comparison-level of a biometric system. We propose a fairness- driven neural network classifier for the comparison of two biometric templates to replace the systems similarity function. This fair classifier is trained with a novel penalization term in the loss function to introduce the criteria of group and individual fairness to the decision process. This penalization term forces the score distributions of different ethnicities to be similar, leading to a reduction of the intra-ethnic performance differences. Experiments were conducted on two publicly available datasets and evaluated the performance of four different ethnicities. The results showed that for both fairness criteria, our proposed approach is able to significantly reduce the ethnic bias, while it preserves a high recognition ability. Our model, build on individual fairness, achieves bias reduction rate between 15.35% and 52.67%. In contrast to previous work, our solution is easy to integrate into existing systems by simply replacing the systems similarity functions with our fair template comparison approach.

CHAPTER 3

PROPOSED WORK

The major goal of our proposed study is to use the deep Learning idea of the system to identify gender, age, and ethnicity or race. Dataset collection, pre-processing, feature extraction, classification, and data evaluation are the primary modules here. The remaining work described our System's information.

3.1 DATASETS COLLECTION

Datasets are a type of data collection. The data for this study was gathered as part of our project's UTK data. To illustrate the output, the data is separated into multiple folders. The system's folders are Gender classification, Age classification, and Race classification. In the Gender classification, we take the system's male and female images. Then, in the Age classification, we take data from people aged 1-100. Finally, in the race classification section, we use data from the system's five country ethnicity data sets.

3.2 DATA PREPROCESSING

Data cleaning, smoothing, grouping or Filtering the image. Data can require pre-processing techniques to ensure accurate, efficient, or meaningful analysis. Data cleaning refers to methods for finding, removing, and replacing bad or missing data. A real-world data generally contains noises, missing values, and maybe in an unusable format which cannot be directly used for Deep learning models. Data pre-processing is required tasks for cleaning the data and making it suitable for a machine learning model which also increases the accuracy and efficiency of a Deep learning model.

3.3 FEATURE EXTRACTION

Feature extraction starts from an initial set of measured data and builds derived values (features) intended to be informative and non-redundant, facilitating the subsequent learning and generalization steps, and in some cases leading to better human interpretations. Feature extraction is related to dimensionality reduction. When the input data to an algorithm is too large to be processed and it is suspected to be redundant (e.g. the same measurement in both feet and meters, or the repetitiveness of images presented as pixels), then it can be transformed into a reduced set of features is also named a feature vector. Determining a subset of the initial features is called feature selection. The selected features are expected to contain the relevant information from the input data, so that the desired task can be performed by using this reduced representation instead of the complete initial data. Feature extraction involves reducing the number of resources required to describe a large set of data. When performing analysis of complex data one of the major problems stems from the number of variables involved. Analysis with a large number of variables generally requires a large amount of memory and computation power, also it may cause a classification algorithm to over fit to training samples and generalize poorly to new samples. Feature extraction is a general term for methods of constructing combinations of the variables to get around these problems while still describing the data with sufficient accuracy.

3.4 GENDER CLASSIFICATION

A person can easily tell whether someone is male or female by looking at their face, but the PC has a difficult time doing so. To perform the identification, machines require a substantial amount of knowledge. The system uses certain recognisable possibilities between male and feminine to classify the gender of a facial image supplied gender.

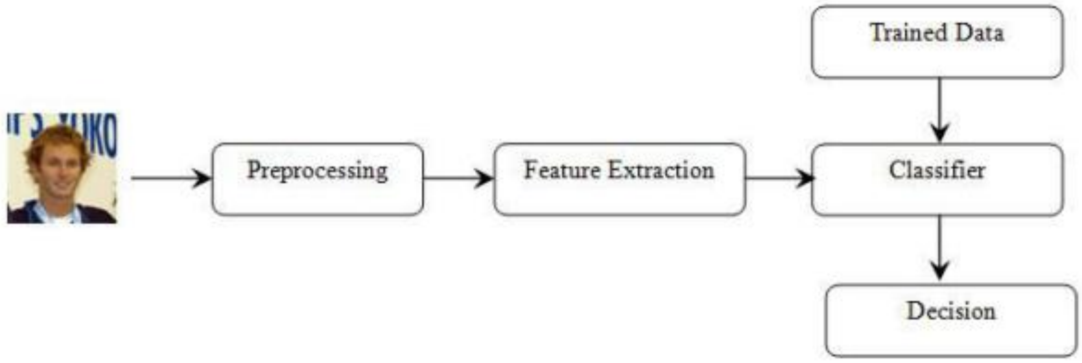

Fig 3.1 Block Diagram of Gender Classification

We wish to classify the gender in the gender classification model using the system's deep learning model. The data is first created in the folder. Next, we'll use pre-processing techniques to clean up the data and remove any undesirable noise from the image. After that, we'll extract certain features from the image. After the trained model is produced, these three stages are generated in the trained model, and the outcome is displayed in the system's testing section. A The data from the trained model is fed into the classifier, which is used to classify the data based on whether the system is male or female. Finally, the decision section is utilised to display the outcome based on the Gender Classification section of the questionnaire.

3.5 AGE ESTIMATION SYSTEM

Facelytics is a face recognition solution that is able to detect peoples' morphological criteria such as age and gender, by analyzing the video feed in real time. It relies on any type of camera and can be used directly within your platform or through a cloud-based solution. In social interaction, gender and age play an important role. Natural differences vary in gender, as are the terms used to identify people by their age. Programmed age estimation in human facial pictures is critical for some reasons. Individuals' behavioral examples, ways of life, and inclinations change alongside their age. Likewise favored methods of human-PC cooperation are distinctive for individuals of various age gatherings. In this way a decent comprehension of age can prompt more smooth and effective correspondence amongst individuals and machines, making programmed age estimation from facial acknowledgment a region with a considerable measure of potential for various viable applications identified with law authorization, multi-signal recognizable proof, age-particular access control and age-particular HCI. People build up the capacity to evaluate age ahead of schedule throughout everyday life and can be genuinely exact in their estimation. Human age estimation is however exceptionally touchy to factors like race, sexual orientation, facial connections, for example, glasses, whiskers, facial piercings, and other outwardly discerning components like engaging quality. Subsequently it would be exceptionally helpful to have a computerized Age Estimation System that takes a human facial picture as its information and 6 doles out a yield mark to the picture, where this name is the correct age (in years) or the age gathering (year run) of the individual face as appeared beneath of the system. Fig 3.2 shows the block diagram of age estimation system

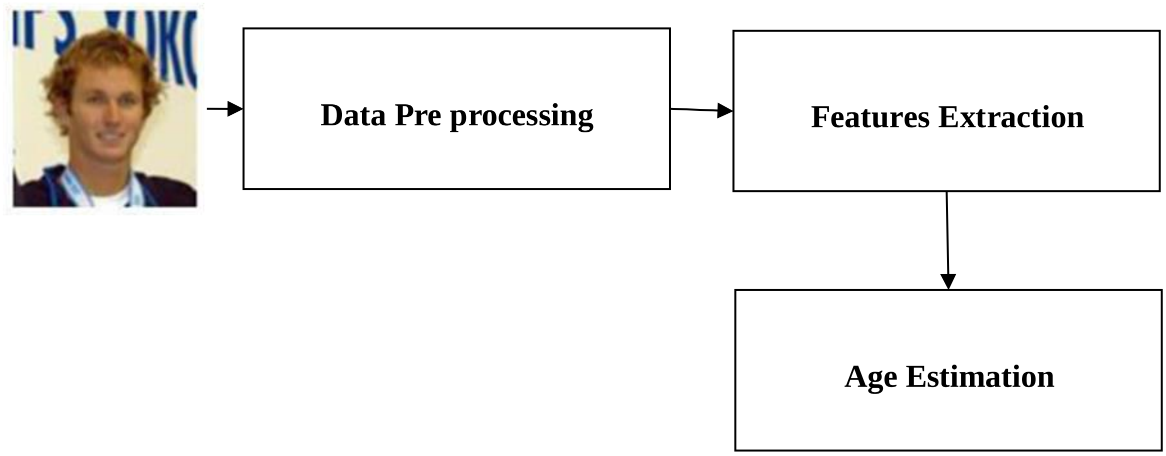

Fig 3.2 Block diagram of Age Estimation system

Our system's age estimation is depicted in a block diagram. As a result, the data is gathered on the kaggle website to determine the outcome. After the dataset process is complete, use Pre-Processing Techniques to pre-process the data. The next step is to extract some features from that system's image. Finally, we employ the system's deep learning idea for training and testing. Finally, depending on the image of the system's age prediction pattern, to display the output.

3.6 ETHIICITY ESTIMATION SYSTEM

Nationality, regional culture, ancestry, and language are all cultural components that make up ethnicity. Brown, white, or black skin (from diverse parts of the world) is an example of race, whereas German or Spanish origin (independent of race) or Han Chinese ancestry are examples of ethnicity. We have used the existing database which has been used for age and gender identification. We have categorised the race into five categories which includes White, Black, Asian, Indian and others (Latin, European). Since our dataset has majority of infant images ,output of the ethnicity identification isn’t accurate.

3.7 ARCHITECTURE OF OUR PROPOSED METHOD

Here, the architecture of our proposed method to show the result of the system.

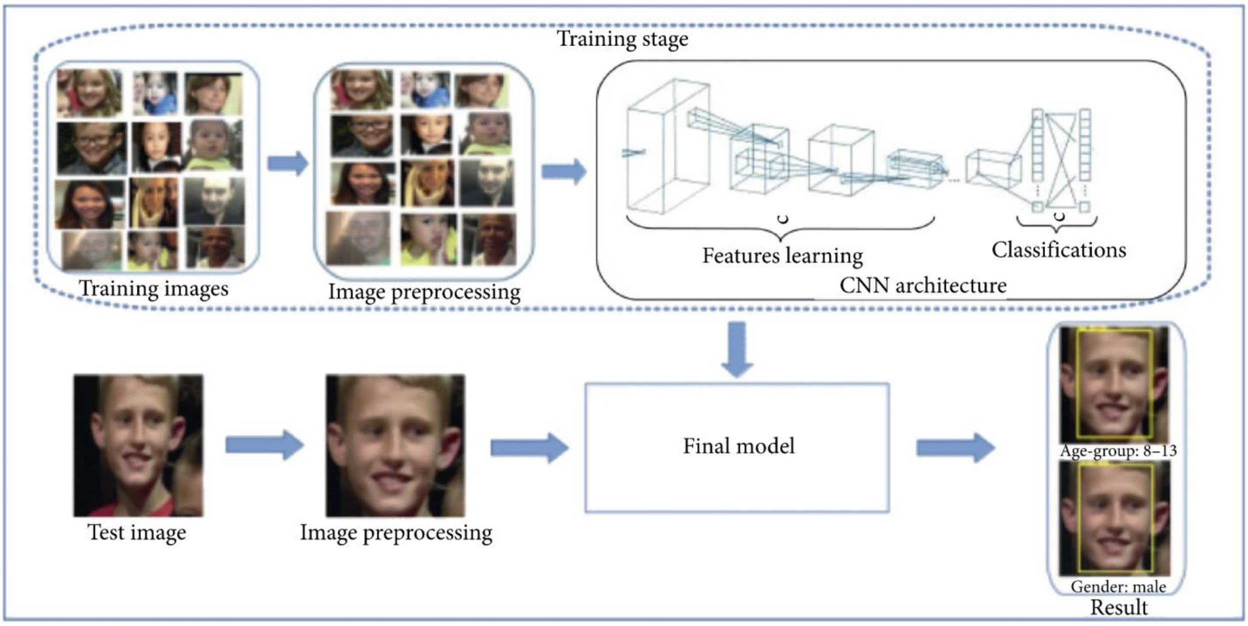

Fig 3.3 Architecture of proposed method

The architecture depicts the model's planned system. There are two parts to our model that are described. One portion of the system is for training, while the other is for testing. Both methods were combined to demonstrate the system's output.

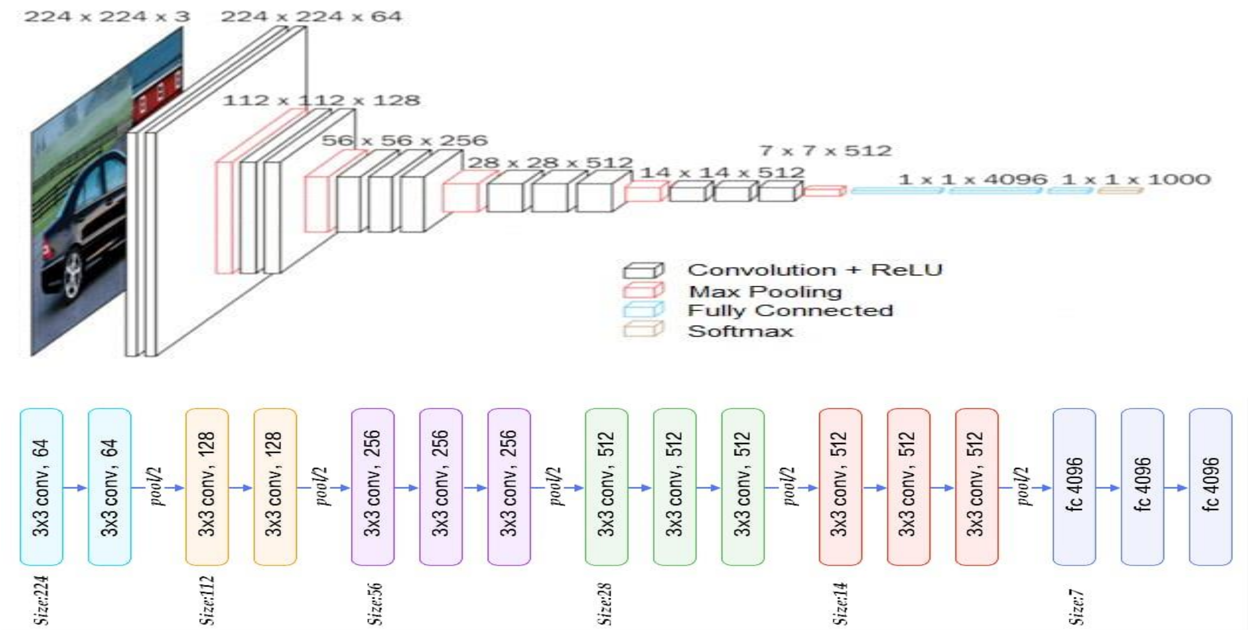

Fig 3.4 Architecture of VGG Network

VGG Face is a dataset of 2.6 million face images of 2,622 people that is used development face recognition technology. VGG16 is a convolution neural net (CNN) architecture which is trained from scratch for facial recognition using VGG Face Dataset. Since facial recognition neural networks (VGG) have already been trained to distinguish human features, the features that they extract may be more useful for determining age and gender from a photo than the features extracted by a more general neural network.

$$n_{out} = \left\lfloor \frac{n_{in} + 2p - k}{s} \right\rfloor + 1$$

n_{in}: number of input features
n_{out}: number of output features
k: convolution kernel size
p: convolution padding size
s: convolution stride size

Equation 3.1 Determines the output features

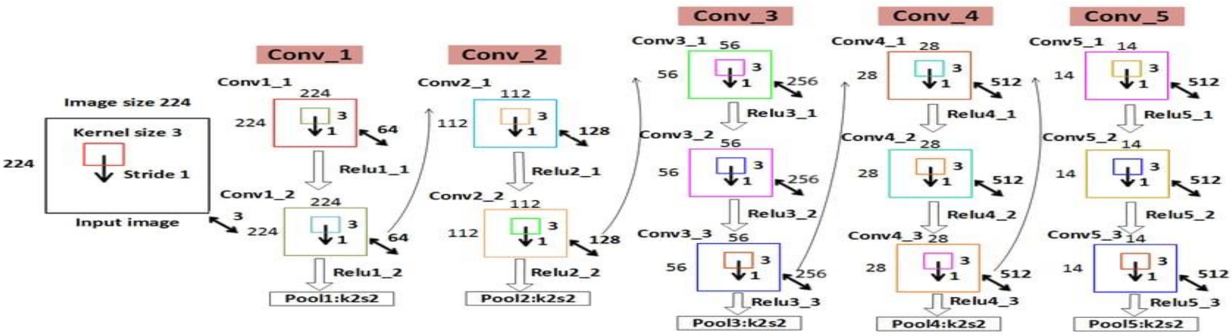

Fig 3.5 Layers of the Network

Here we use, Convolution layer of 3x3 kernel ,stride 1x1 and padding(1),starting with 224x224 input features. Max pool layer of 2x2 pool size and stride 2x2 and has no zero padding of the system.

3.8 VGG-16

VGG16 proved to be a significant milestone in the quest of mankind to make computers "see" the world. A lot of effort has been put into improving this ability under the discipline of Computer Vision (CV) for a number of decades. VGG16 is one of the significant innovations that paved the way for several innovations that followed in this field.

It is a Convolutional Neural Network (CNN) model proposed by Karen Simonyan and Andrew Zisserman at the University of Oxford. The idea of the model was proposed in 2013, but the actual model was submitted during the ILSVRC ImageNet Challenge in 2014. The ImageNet Large Scale Visual Recognition Challenge (ILSVRC) was an annual competition that evaluated algorithms for image classification (and object detection) at a large scale. They did well in the challenge but couldn't win.

Using the ImageNet database, Alex Krizhevsky proposed the AlexNet – a CNN-based network in 2012. He used deep learning for the first time to get the top-5 error rate in ImageNet Challenge to 16.4%. It was the first network to get the error rate below 25%. After the success of AlexNet, the non-CNN methods for computer vision were completely abandoned. In 2014, GoogLeNet by Google proposed the use of inception modules to reduce the problem of exploding or vanishing gradients. They could, in turn, make the networks deeper. It could use 22 layers and was eventually the winner of the 2014 ImageNet Challenge. Later in 2015, Kaiming He et al. at Microsoft proposed ResNet to make the networks truly deep – as deep as 152 layers. It was a significant jump from 22 to 152 layers. They broke the barrier of vanishing and exploding gradients by the use of skip connections. ResNet brought down the top-5 error rate to 3.57% – thanks to the 152 layers in the network.

Fig 3.6 ImageNet Results: VGG Vs Others

3.9 VGG CONFIGURATION

The idea behind using 3 x 3 filters uniformly is something that makes the VGG stand out. Two consecutive 3 x 3 filters provide for an effective receptive field of 5 x 5. Similarly, three 3 x 3 filters make up for a receptive field of 7 x 7. This way, a combination of multiple 3 x 3 filters can stand in for a receptive area of a larger size.

But then, what is the benefit of using three 3 x 3 layers instead of a single 7 x 7 layer? Isn’t it increasing the no. of layers, and in turn, the complexity unnecessarily? No. In addition to the three convolution layers, there are also three non-linear activation layers instead of a single one you would have in 7 x 7. This makes the decision functions more discriminative. It would impart the ability to the network to converge faster.

Secondly, it also reduces the number of weight parameters in the model significantly. Assuming that the input and output of a three-layer 3 x 3 convolutional stack have C channels, the total number of weight parameters will be 3 * 32 C2 = 27 C2. If we compare this to a 7 x 7 convolutional layer, it would require 72 C2 = 49 C2, which is almost twice the 3 x 3 layers. Additionally, this can be seen as a regularization on the 7 x 7 convolutional filters forcing them to have a decomposition through the 3 x 3 filters, with, of course, the non-linearity added in-between by means of ReLU activations. This would reduce the tendency of the network to over-fit during the training exercise.

Another question is – can we go lower than 3 x 3 receptive size filters if it provides so many benefits? The answer is “No.” 3 x 3 is considered to be the smallest size to capture the notion of left to right, top to down, etc. So lowering the filter size further could impact the ability of the model to understand the spatial features of the image.

The consistent use of 3 x 3 convolutions across the network made the network very simple, elegant, and easy to work with.

The authors proposed various configurations of the network based on the depth of the network. They experimented with several such configurations, and the following ones were submitted during the ImageNet Challenge.

A stack of multiple (usually 1, 2, or 3) convolution layers of filter size 3 x 3, stride one, and padding 1, followed by a max-pooling layer of size 2 x 2, is the basic building block for all of these configurations. Different configurations of this stack were repeated in the network configurations to achieve different depths. The number associated with each of the configurations is the number of layers with weight parameters in them.

The convolution stacks are followed by three fully connected layers, two with size 4,096 and the last one with size 1,000. The last one is the output layer with Softmax activation. The size of 1,000 refers to the total number of possible classes in ImageNet.

VGG16 refers to the configuration "D" in the table listed below. The configuration "C" also has 16 weight layers. However, it uses a 1 x 1 filter as the last convolution layer in stacks 3, 4, and 5. This layer was used to increase the non-linearity of the decision functions without affecting the receptive field of the layer.

Karen Simonyan and Andrew Zisserman proposed the idea of the VGG network in 2013 and submitted the actual model based on the idea in the 2014 ImageNet Challenge. They called it VGG after the department of Visual Geometry Group in the University of Oxford that they belonged to.

ConvNet Configuration					
A	A-LRN	B	C	D	E
11 weight layers	11 weight layers	13 weight layers	16 weight layers	16 weight layers	19 weight layers
input (224 × 224 RGB image)					
conv3-64	conv3-64 **LRN**	conv3-64 **conv3-64**	conv3-64 conv3-64	conv3-64 conv3-64	conv3-64 conv3-64
maxpool					
conv3-128	conv3-128	conv3-128 **conv3-128**	conv3-128 conv3-128	conv3-128 conv3-128	conv3-128 conv3-128
maxpool					
conv3-256 conv3-256	conv3-256 conv3-256	conv3-256 conv3-256	conv3-256 conv3-256 **conv1-256**	conv3-256 conv3-256 **conv3-256**	conv3-256 conv3-256 conv3-256 **conv3-256**
maxpool					
conv3-512 conv3-512	conv3-512 conv3-512	conv3-512 conv3-512	conv3-512 conv3-512 **conv1-512**	conv3-512 conv3-512 **conv3-512**	conv3-512 conv3-512 conv3-512 **conv3-512**
maxpool					
conv3-512 conv3-512	conv3-512 conv3-512	conv3-512 conv3-512	conv3-512 conv3-512 **conv1-512**	conv3-512 conv3-512 **conv3-512**	conv3-512 conv3-512 conv3-512 **conv3-512**
maxpool					
FC-4096					
FC-4096					
FC-1000					
soft-max					

Fig 3.7 VGG Configurations

3.10 VGG ARCHITECTURE

The input to any of the network configurations is considered to be a fixed size 224 x 224 image with three channels – R, G, and B. The only pre-processing done is normalizing the RGB values for every pixel. This is achieved by subtracting the mean value from every pixel.

Image is passed through the first stack of 2 convolution layers of the very small receptive size of 3 x 3, followed by ReLU activations. Each of these two layers contains 64 filters. The convolution stride is fixed at 1 pixel, and the padding is 1 pixel. This configuration preserves the spatial resolution, and the size of the output activation map is the same as the input image dimensions. The activation maps are then passed through spatial max pooling over a 2 x 2-pixel window, with a stride of 2 pixels. This halves the size of the activations. Thus the size of the activations at the end of the first stack is 112 x 112 x 64.

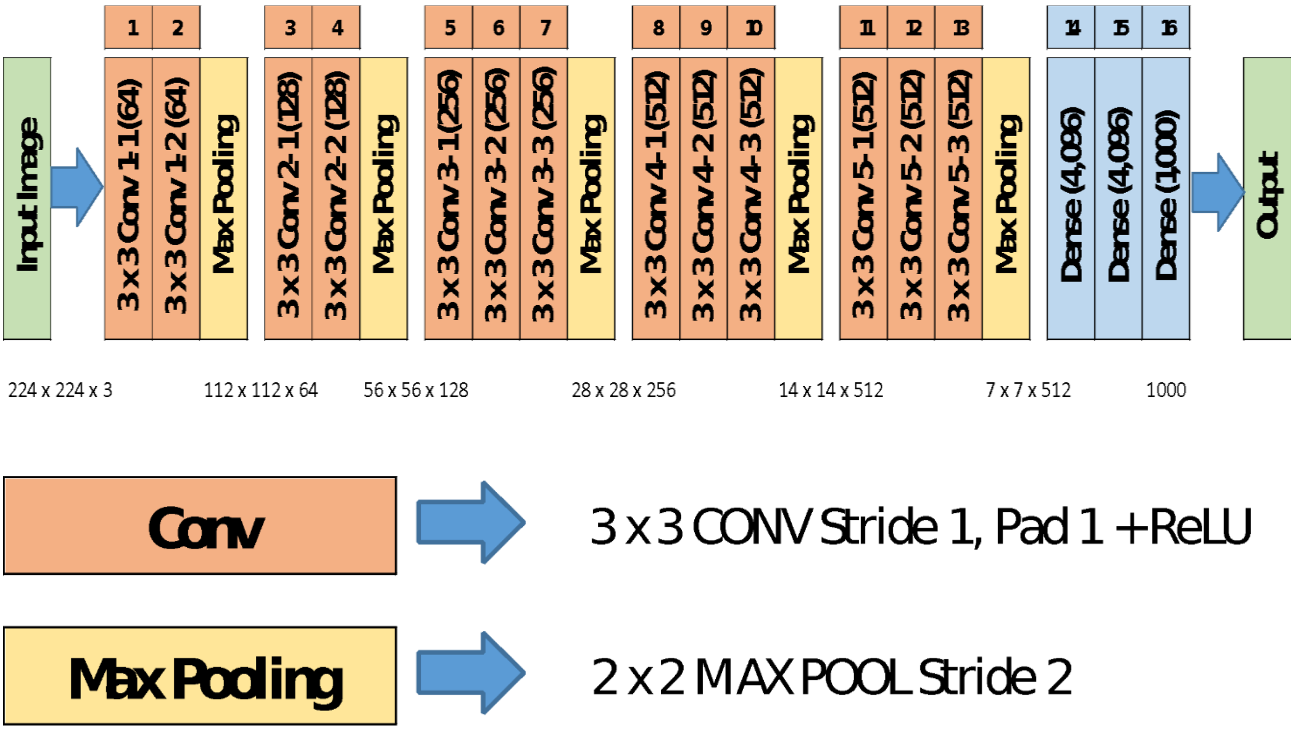

Fig 3.8 VGG16 Architecture

The activations then flow through a similar second stack, but with 128 filters as against 64 in the first one. Consequently, the size after the second stack becomes 56 x 56 x 128. This is followed by the third stack with three convolutional layers and a max pool layer. The no. of filters applied here are 256, making the output size of the stack 28 x 28 x 256. This is followed by two stacks of three convolutional layers, with each containing 512 filters. The output at the end of both these stacks will be 7 x 7 x 512.

The stacks of convolutional layers are followed by three fully connected layers with a flattening layer in-between. The first two have 4,096 neurons each, and the last fully connected layer serves as the output layer and has 1,000 neurons corresponding to the 1,000 possible classes for the ImageNet dataset. The output layer is followed by the Softmax activation layer used for categorical classification.

Though VGG16 didn’t win the ImageNet 2014 Challenge, the ideas it provided paved the way for subsequent innovations in the field of Computer Vision. The idea of stacked convolution layers of smaller receptive fields was a breakthrough innovation, in my opinion. It provided a number of advantages over the erstwhile state-of-the-art networks. Most of the modern CNN networks still use this block of stacked convolutions.

CHAPTER 4

CLASSIFICATION TECHNIQUES

4.1 STRUCTURE FOR PERFORMING IMAGE CLASSIFICATION

- **Image Pre-processing**: The aim of this process is to improve the image data (features) by suppressing unwanted distortions and enhancement of some important image features so that the computer vision models can benefit from this improved data to work on. Steps for image pre-processing includes Reading image, Resizing image, and Data Augmentation (Grey scaling of image, Reflection, Gaussian Blurring, Histogram, Equalization, Rotation, and Translation).
- **Detection of an object**: Detection refers to the localization of an object which means the segmentation of the image and identifying the position of the object of interest.
- **Feature extraction and training**: This is a crucial step wherein statistical or deep learning methods are used to identify the most interesting patterns of the image, features that might be unique to a particular class and that will, later on, help the model to differentiate between different classes. This process where the model learns the features from the dataset is called model training.
- **Classification of the object**: This step categorizes detected objects into predefined classes by using a suitable classification technique that compares the image patterns with the target patterns.

4.2 SUPERVISED CLASSIFICATION

Supervised classification is based on the idea that a user can select sample pixels in an image that are representative of specific classes and then direct the image processing software to use these training sites as references for the classification of all other pixels in the image. Training sites (also known as testing sets or input classes) are selected based on the knowledge of the user. The user also sets the bounds for how similar other pixels must be to group them together. These bounds are often set based on the spectral characteristics of the training area. The user also designates the number of classes that the image is classified into. Once a statistical characterization has been achieved for each information class, the image is then classified by examining the reflectance for each pixel and making a decision about which of the signatures it resembles most. Supervised classification uses classification algorithms and regression techniques to develop predictive models. The algorithms include linear regression, logistic regression, neural networks, decision tree, support vector machine, random forest, naive Bayes, and k-nearest neighbour.

4.3 UNSUPERVISED CLASSIFICATION

Unsupervised classification is where the outcomes (groupings of pixels with common characteristics) are based on the software analysis of an image without the user providing sample classes. The computer uses techniques to determine which pixels are related and groups them into classes. The user can specify which algorithm the software will use and the desired number of output classes but otherwise does not aid in the classification process. However, the user must have knowledge of the area being classified when the groupings of pixels with common characteristics produced by the computer have to be related to actual features on the ground. Some of the most common algorithms used in unsupervised learning include cluster analysis, anomaly detection, neural networks, and approaches for learning latent variable models.

4.4 CONVOLUTIONAL NEURAL NETWORK

Convolutional Neural Network (CNN, or ConvNet) are a special kind of multi-layer neural networks, designed to recognize visual patterns directly from pixel images with minimal pre-processing. It is a special architecture of artificial neural networks. Convolutional neural network uses some of its features of visual cortex and have therefore achieved state of the art results in computer vision tasks. Convolutional neural networks are comprised of two very simple elements, namely convolutional layers and pooling layers. Although simple, there are near-infinite ways to arrange these layers for a given computer vision problem. The elements of a convolutional neural network, such as convolutional and pooling layers, are relatively straightforward to understand. The challenging part of using convolutional neural networks in practice is how to design model architectures that best use these simple elements. The reason why convolutional neural network is hugely popular is because of their architecture, the best thing is there is no need of feature extraction. The system learns to do feature extraction and the core concept is, it uses convolution of image and filters to generate invariant features which are passed on to the next layer. The features in next layer are convoluted with different filters to generate more invariant and abstract features and the process continues till it gets final feature/output which is invariant to occlusions. The most commonly used architectures of convolutional neural network are LeNet, AlexNet, ZFNet, GoogLeNet, VGGNet, and ResNet.

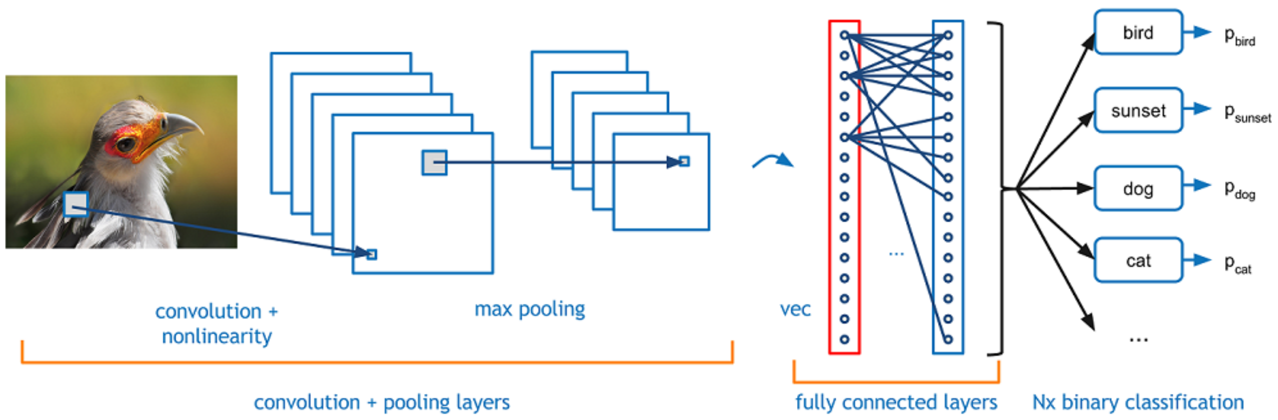

Fig 4.1 Classification using CNN

The objective of image classification is to identify and portray, as a unique gray level (or color), the features occurring in an image in terms of the object these features actually represent on the ground. Image classification is perhaps the most important part of digital image analysis. Classification between objects is a complex task and therefore image classification has been an important task within the field of computer vision. Image classification refers to the labelling of images into one of a number of predefined classes. There are potentially n number of classes in which a given image can be classified. Manually checking and classifying images could be a tedious task especially when they are massive in number and therefore it will be very useful if we could automate this entire process using computer vision.

4.5 ARTIFICIAL NEURAL NETWORK

Inspired by the properties of biological neural networks, Artificial Neural Networks are statistical learning algorithms and are used for a variety of tasks, from relatively simple classification tasks to computer vision and speech recognition. Artificial neural networks are implemented as a system of interconnected processing elements, called nodes, which are functionally analogous to biological neurons. The connections between different nodes have numerical values, called weights, and by altering these values in a systematic way, the network is eventually able to approximate the desired function. The hidden layers can be thought of as individual feature detectors, recognizing more and more complex patterns in the data as it is propagated throughout the network. For example, if the network is given a task to recognize a face, the first hidden layer might act as a line detector, the second hidden takes these lines as input and puts them together to form a nose, the third hidden layer takes the nose and matches it with an eye and so on, until finally the whole face is constructed. This hierarchy enables the network to eventually recognize very complex objects. The different types of artificial neural network are convolutional neural network, feedforward neural network, probabilistic neural network, time delay neural network, deep stacking network, radial basis function network, and recurrent neural network.

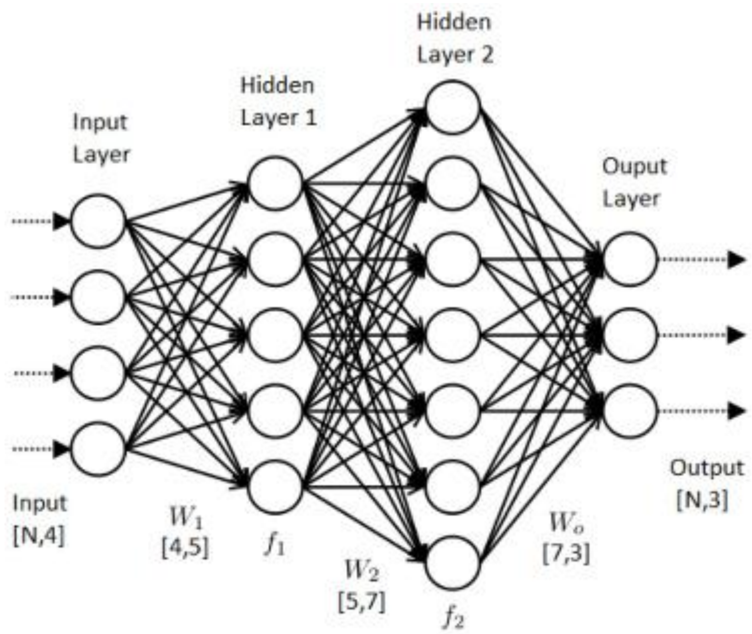

Fig 4.2 Classification using ANN

4.6 SUPPORT VECTOR MACHINE

Support vector machines (SVM) are powerful yet flexible supervised machine learning algorithms which are used both for classification and regression. Support vector machines have their unique way of implementation as compared to other machine learning algorithms. They are extremely popular because of their ability to handle multiple continuous and categorical variables. Support Vector Machine model is basically a representation of different classes in a hyperplane in multidimensional space. The hyperplane will be generated in an iterative manner by support vector machine so that the error can be minimized. The goal is to divide the datasets into classes to find a maximum marginal hyperplane. It builds a hyper-plane or a set of hyper-planes in a high dimensional space and good separation between the two classes is achieved by the hyperplane that has the largest distance to the nearest training data point of any class. The real power of this algorithm depends on the kernel function being used. The most commonly used kernels are linear kernel, gaussian kernel, and polynomial kernel.

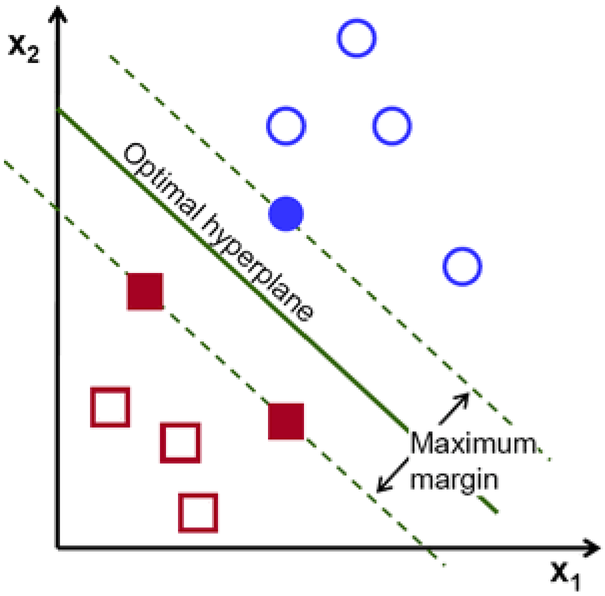

Fig 4.3 Classification using Support Vector Machine

4.7 K-Nearest Neighbor

K-Nearest Neighbor is a non-parametric method used for classification and regression. In both cases, the input consists of the *k* closest training examples in the feature space. It is by far the simplest algorithm. It is a non-parametric, lazy learning algorithm, where the function is only approximated locally and all computation is deferred until function evaluation. This algorithm simply relies on the distance between feature vectors and classifies unknown data points by finding the most common class among the k-closest examples. In order to apply the k-nearest Neighbor classification, we need to define a distance metric or similarity function, where the common choices include the Euclidean distance and Manhattan distance. The output is a class membership. An object is classified by a plurality vote of its neighbors, with the object being assigned to the class most common among its k nearest neighbors (k is a positive integer, typically small). If k = 1, then the object is simply assigned to the class of that single nearest neighbor. Condensed nearest neighbor (CNN, the Hart algorithm) is an algorithm designed to reduce the data set for K-Nearest Neighbor classification.

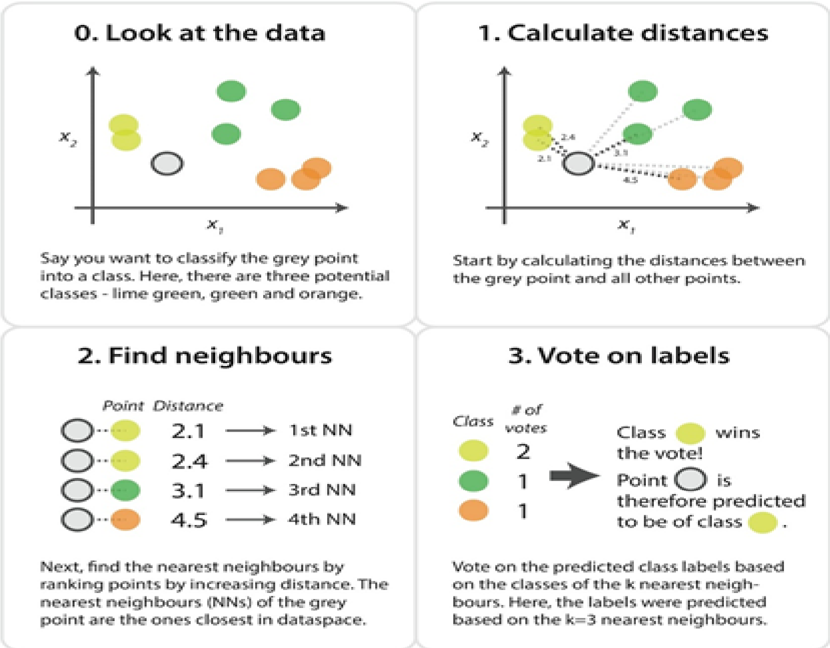

Fig 4.4 Classification using K-Nearest Neighbor

4.8 Naive Bayes Algorithm

Naive Bayes classifiers are a collection of classification algorithms based on Bayes' Theorem. It is not a single algorithm but a family of algorithms where all of them share a common principle, i.e. every pair of features being classified is independent of each other. Naive Bayes is a simple technique for constructing classifiers: models that assign class labels to problem instances, represented as vectors of feature values, where the class labels are drawn from some finite set. All naive bayes classifiers assume that the value of a particular feature is independent of the value of any other feature, given the class variable. Naive Bayes algorithm is a fast, highly scalable algorithm, which can be used for binary and multi-class classification. It depends on doing a bunch of counts. It is a popular choice for text classification, spam email classification, etc. It can be easily trained on small dataset. It has limitation as it considers all the features to be unrelated, so it cannot learn the relationship between features. Naive Bayes can learn individual features importance but can't determine the relationship among features. Different types of naïve bayes algorithms are gaussian naïve bayes, multinomial naïve bayes, and bernoulli naïve bayes.

4.9 Random Forest Algorithm

Random forest is a supervised learning algorithm which is used for both classification as well as regression. As we know that a forest is made up of trees and more trees means more robust forest, similarly, random forest algorithm creates decision trees on data samples and then gets the prediction from each of them and finally selects the best solution by means of voting. It is an ensemble method which is better than a single decision tree because it reduces the over-fitting by averaging the result. The random forest is a classification algorithm consisting of many decision trees. It uses bagging and feature randomness when building each individual tree to try to create an uncorrelated forest of trees whose prediction by committee is more accurate than that of any individual tree.

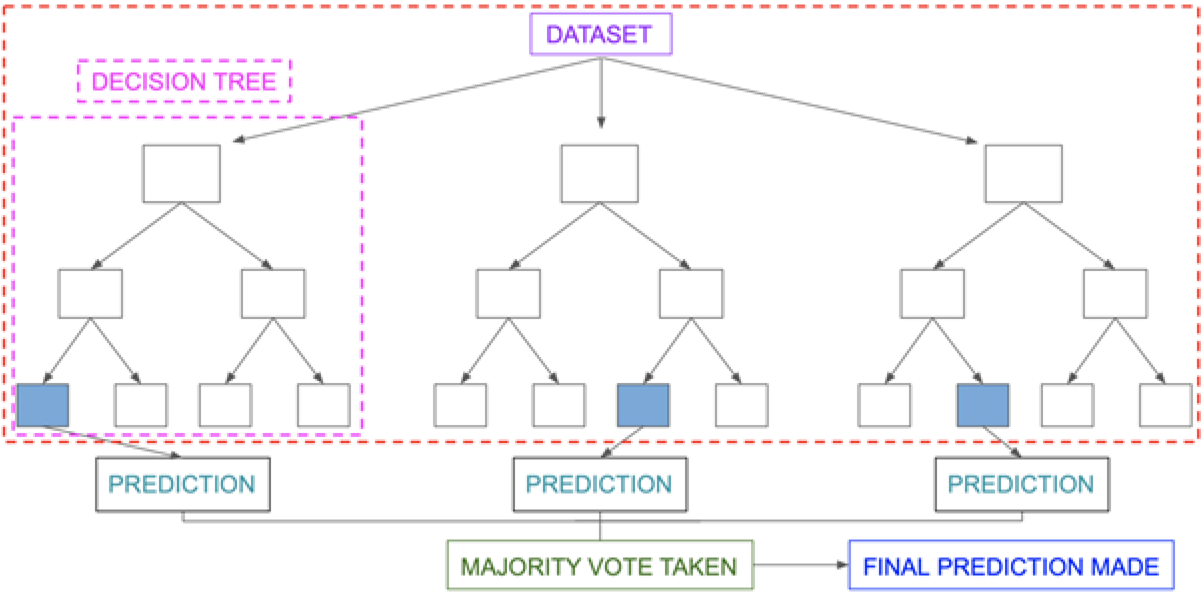

Fig 4.5 Classification using Random Forest Algorithm

4.10 IMPLEMENTATION ASPECTS

➢ **DETECT FACES:**

- When given a frame/video, the CNN algorithm first detects for faces in each frame of the system.

➢ **CLASSIFY INTO MALE/FEMALE:**

- Once it finds faces in the frame, the features of the faces are extracted, and the gender is determined using second layer of CNN. Based on the different parts of facial feature extraction, the gender classification via face can be divided into local feature extraction and global feature extraction methods. The local feature extraction method extracts features from certain facial points like the mouth, nose and eyes, whereas the global features extraction method extracts features from the whole face instead of extracting features from facial points.

➢ **CLASSIFY INTO ONE OF THE AGE RANGES:**

- In the third layer of CNN, the age of the faces is determined and falls under either of the age ranges for example [(0-2), (4-6), (8-12), (15-20), (25-32), (38-43), (48-53), (60-100)].

- **PUT THE RESULTS ON THE FRAME AND DISPLAY IT:**

- The result is displayed on the frame containing the age range and gender. The resulting frame consists of a square box around the faces with the estimated gender and age.

- **ETHINICITY ESTIMATION**

- We have used the existing database which has been used for age and gender identification. We have categorised the race into five categories which includes White, Black, Asian, Indian and others (Latin, European). Since our dataset has majority of infant images ,output of the ethnicity identification isn’t accurate.

CHAPTER 5

IMPLEMENTATION TOOLS

5.1 REQUIREMENT ANALYSIS AND SPECIFICATIONS

The requirement engineering process consists of feasibility study, requirements elicitation and analysis, requirements specification, requirements validation and requirements management. Requirements elicitation and analysis is an iterative process that can be represented as a spiral of activities, namely requirements discovery, requirements classification and organisation, requirement negotiation and requirements documentation.

5.2 SOFTWARE REQUIREMENT

- Operating System - Windows 11
- Software Programming Package - MATLAB R2022a

5.3 HARDWARE REQUIREMENTS

- Processor Type - Pentium -IV
- Speed - 2.4 GHZ
- Ram - 8 GB
- Hard disk - 552 GB SSD

5.4 SYSTEM USED

Name of the Company: MI

Version: 21H2

5.5 DESCRIPTION OF THE TOOLS USED

The project is mainly implemented in MATLAB using several packages. Deep learning modules, which are constructed with certain tools are used in our project to classify age, gender and race.

5.5.1 MATLAB

MATLAB (Matrix Laboratory) is a matrix-oriented language for technical computing. It is not only used for computation, but also for visualization and programming in an easy-to-use environment. It is an interpreted language (not compiled) that was conceived to provide easy access to matrix and linear algebra software that was written in FORTRAN. One of the main features of MATLAB is that it is oriented toward numerical computing, instead of symbolic computing (as e.g., Maple software, Mathematical).The software comes in the form of a core program and addition a libraries or toolboxes. A toolbox is a collection of MATLAB functions (called M-functions or M-files) that extend the capability of the core environment to solve specific topic problems. MATLAB is optimized to be relatively fast when performing array operations, so it is important to take this into account to write suitable instructions, for example, to avoid unnecessary 'for' loops that process individual array elements.

5.5.2 DEEP LEARNING TOOLBOX

Deep Learning Toolbox provides a framework for designing and implementing deep neural networks with algorithms, pre-trained models, and apps. You can use convolutional neural networks (Con-Nets, CNNs) and long short-term memory (LSTM) networks to perform classification and regression on image, time-series, and text data. You can build network architectures such as generative adversarial networks (GANs) and Siamese networks using automatic differentiation, custom training loops, and shared weights.

With the Deep Network Designer app, you can design, analyze, and train networks graphically. The Experiment Manager app helps you manage multiple deep learning experiments, keep track of training parameters, analyze results, and compare code from different experiments. You can visualize layer activations and graphically monitor training progress.

Create new deep networks for image classification and regression tasks by defining the network architecture and training the network from scratch. You can also use transfer learning to take advantage of the knowledge provided by a pretrained network to learn new patterns in new data. Fine-tuning a pretrained image classification network with transfer learning is typically much faster and easier than training from scratch. Using pretrained deep networks enables you to quickly learn new tasks without defining and training a new network, having millions of images, or having a powerful GPU.

After defining the network architecture, you must define training parameters using the training option function. You can then train the network using train network. Use the trained network to predict class labels or numeric responses.

You can train a convolutional neural network on a CPU, a GPU, multiple CPUs or GPUs, or in parallel on a cluster or in the cloud. Training on a GPU or in parallel requires Parallel Computing Toolbox™. Using a GPU requires a supported GPU device (for information on supported devices, see GPU Support by Release (Parallel Computing Toolbox)). Specify the execution environment using the training Options function.

5.5.3 DEEP NETWORK DESIGNER

The Deep Network Designer app lets you build, visualize, edit, and train deep learning networks. Using this app, you can:

- Build, import, edit, and combine networks.
- Load pretrained networks and edit them for transfer learning.
- View and edit layer properties and add new layers and connections.
- Analyze the network to ensure that the network architecture is defined correctly, and detect problems before training.
- Import and visualize datastores and image data for training and validation.
- Apply augmentations to image classification training data and visualize the distribution of the class labels.
- Train networks and monitor training with plots of accuracy, loss, and validation metrics.
- Export trained networks to the workspace or to Simulink®.
- Generate MATLAB® code for building and training networks and create experiments for hyperparameter tuning using Experiment Manager.

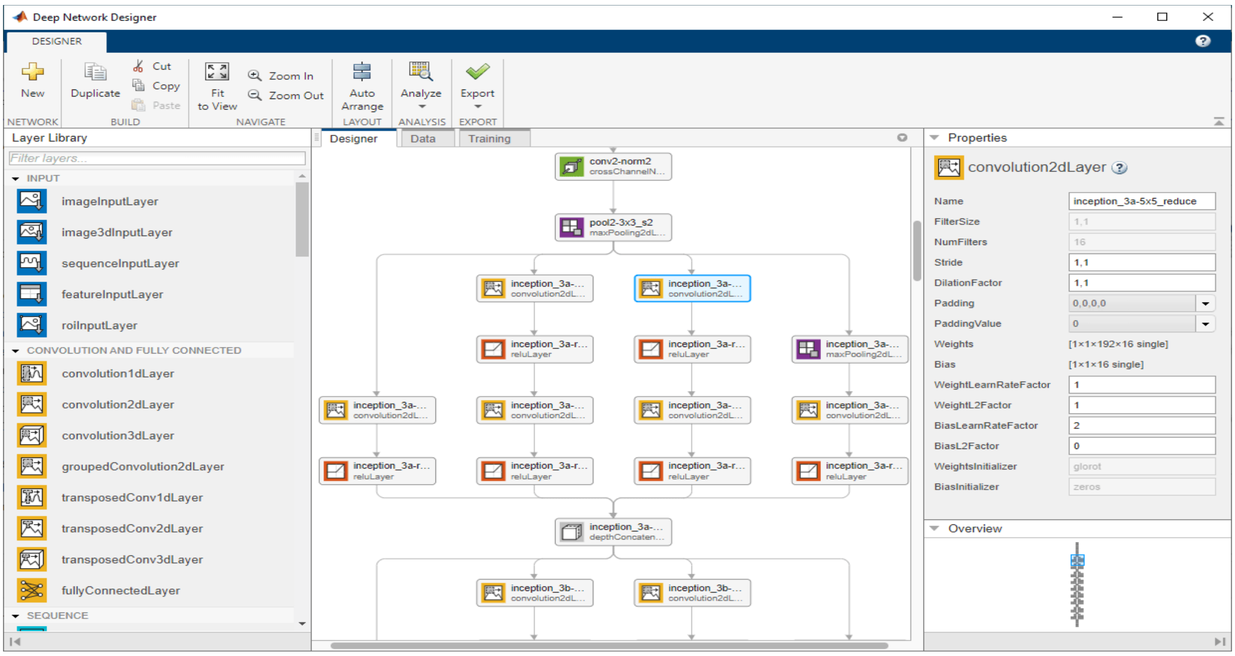

Fig 5.1 Deep Network Designer

5.5.3.1TRANSFER LEARNING WITH DEEP NETWORK DESIGNER

Transfer learning is the process of taking a pretrained deep learning network and fine-tuning it to learn a new task. Using transfer learning is usually faster and easier than training a network from scratch. You can quickly transfer learned features to a new task using a smaller amount of data.

Use Deep Network Designer to perform transfer learning for image classification by following these steps:

1. Open the Deep Network Designer app and choose a pretrained network.
2. Import the new data set.
3. Replace the final layers with new layers adapted to the new data set.
4. Set learning rates so that learning is faster in the new layers than in the transferred layers.
5. Train the network using Deep Network Designer, or export the network for training at the command line.

CHAPTER 6

RESULTS AND DISCUSSIONS

In the results and discussion part of the system, we show the final output model of the system. The output are shown in the three part of the system. They are Gender Classification, Age Classification and Ethnicity Classification model of the system. This section clarifies the performance assessment of the facial demographic estimation from free facial pictures. This demographic estimation experimented with the assistance of MATLAB (version R2022a), the experiments area unit performed on associate degree Intel(R) Core machine with a speed a pair of 2.30 GHz, and 8.0 GB RAM exploitation Windows seven 64-bit software system.

6.1 AGE CLASSIFICATION SYSTEM

6.1.1 AGE-TRAINING

Fig.6.1 shows the training process of the VGG16network for the age prediction system. For this training process, the Adam optimizer is chosen and the maximum number of epochs is fixed at 120. For training purposes, 80% of database images are used and the remaining 20% of images are used for the validation process. Then finally, the training process was completed with a maximum accuracy of 81%.

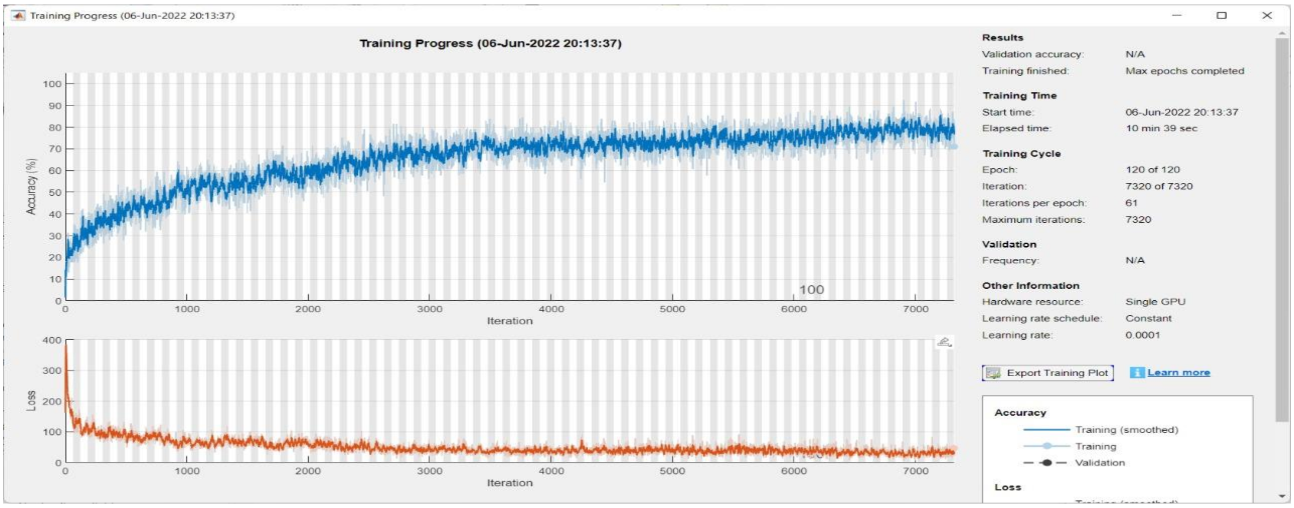

Fig 6.1 Age Accuracy

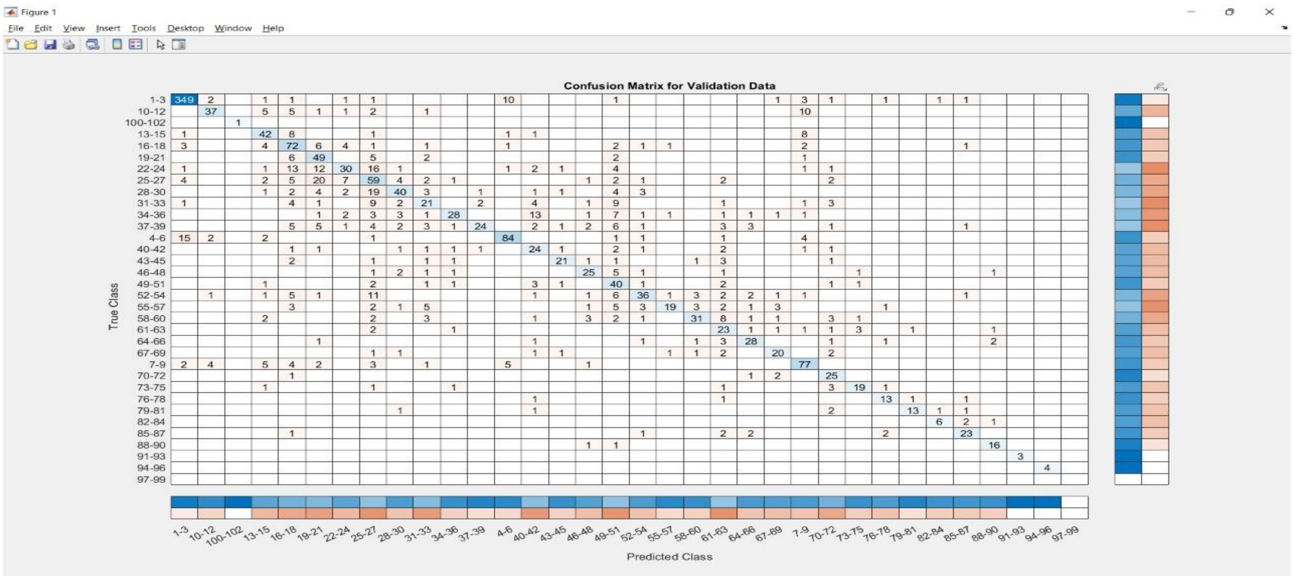

Fig 6.2 Confusion Matrix - Age

Fig 6.2 shows the performance evaluation of the age prediction system in the form of confusion matrix and the training process had a validation error of 33.13%.

6.1.2 AGE-TESTING

Open the GUI created. Click on load database**,** where the network which determines the age will be loaded.Then choose the image from folder images or through webcam (live image). Image chosen will be appeared on the GUI. Click on the AGE button which will show you the age(Fig 6.3).

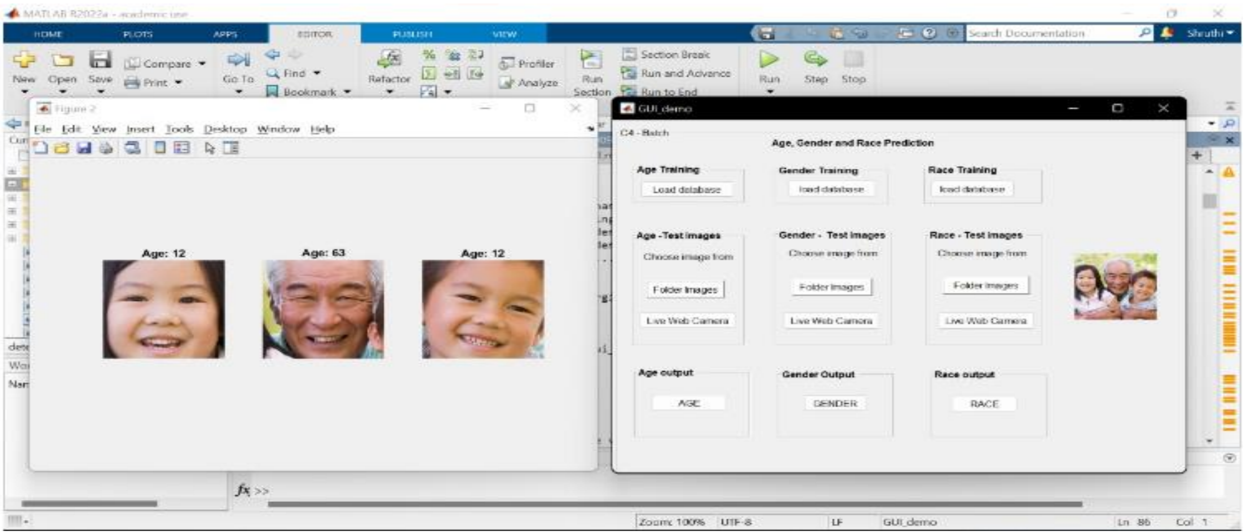

Fig 6.3 Age Classification Output

6.2 GENDER CLASSIFICATION SYSTEM

6.2.1 GENDER-TRAINING

Fig.6.4 shows the training process of the VGG16 network for the Gender prediction system. For this training process, the Adam optimizer is chosen and the maximum number of epochs is fixed at 120. For training purposes, 80% of database images are used and the remaining 20% of images are used for the validation process. Then finally, the training process was completed with a maximum accuracy of 91%.

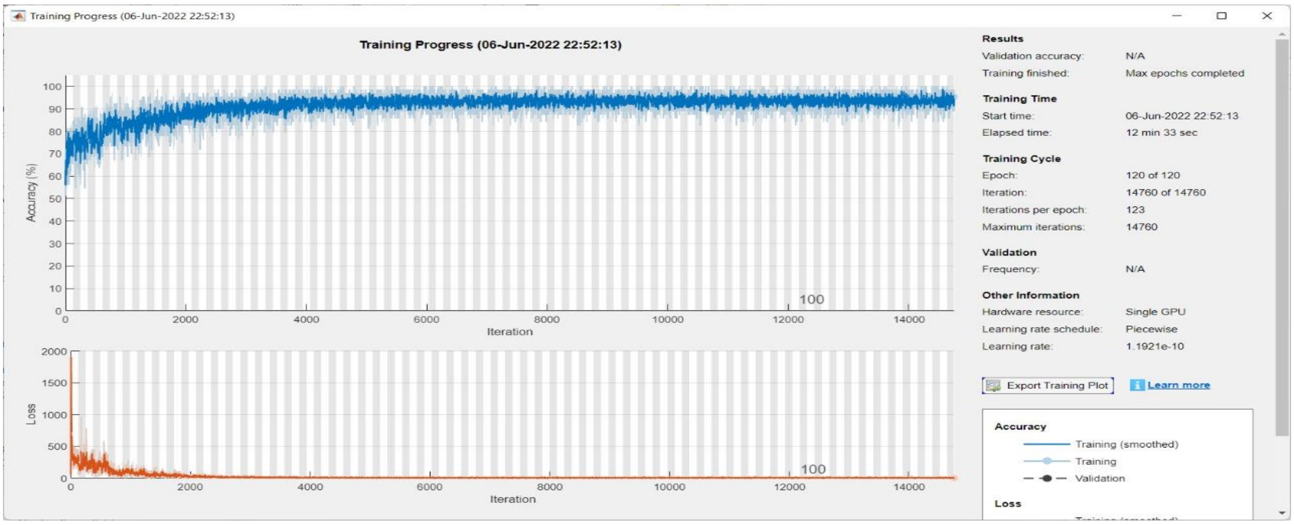

Fig 6.4 Gender Accuracy

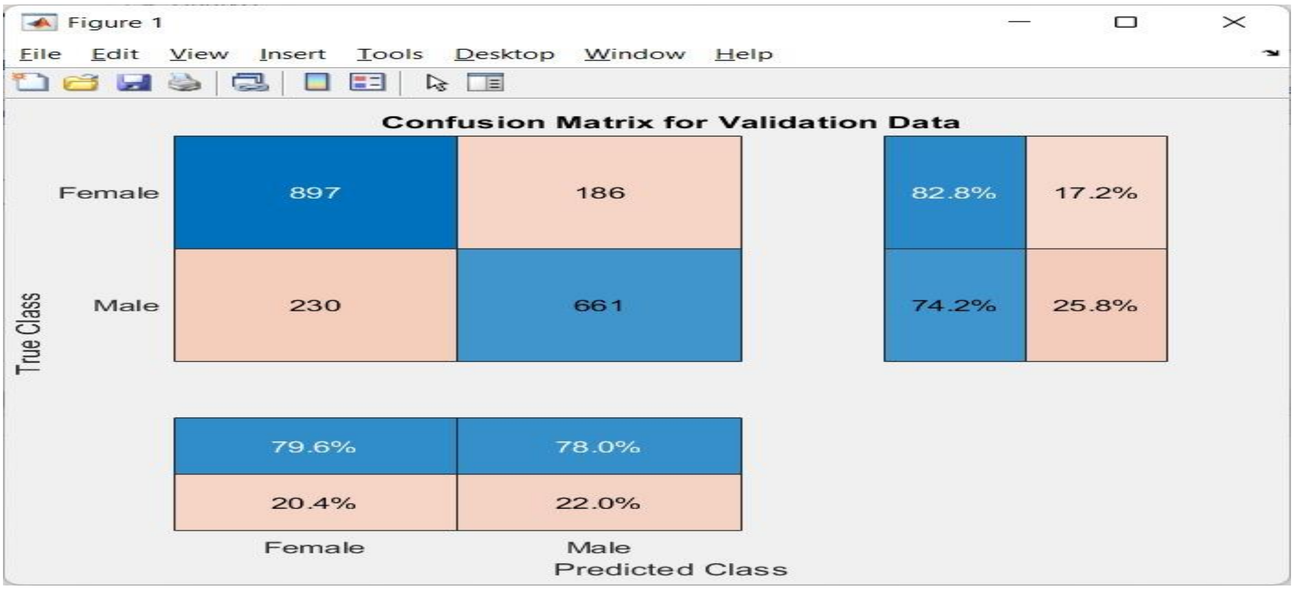

Fig 6.5 Confusion Matrix - Gender

Fig 6.2 shows the performance evaluation of the age prediction system in the form of confusion matrix and the training process had a validation error of 3.16%.

6.2.2 GENDER-TESTING

Open the GUI created. Click on load database**,** where the network which determines the gender will be loaded. Then choose the image from folder images or through webcam (live image). Image chosen will be appeared on the GUI. Click on the GENDER button which will show you the gender.

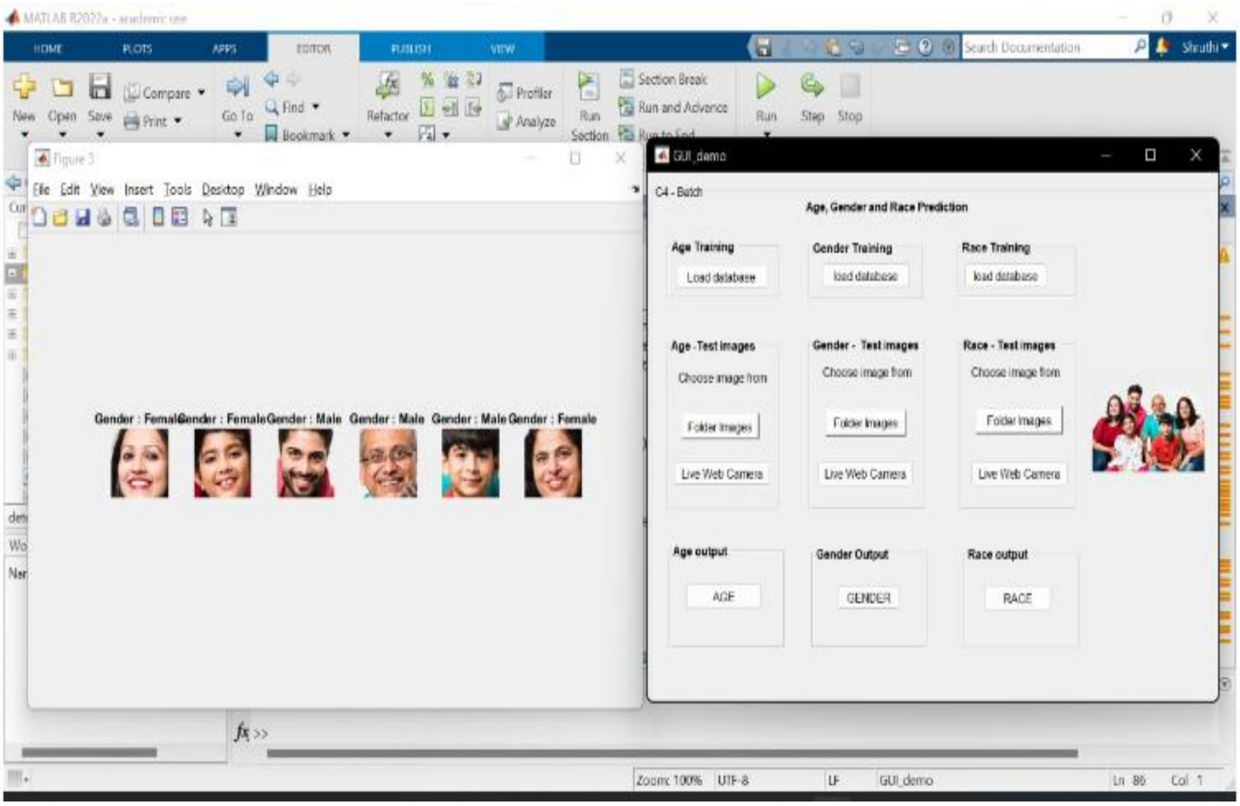

Fig 6.6 Gender Classification Output

6.3 ETHNICITY CLASSIFICATION SYSTEM

6.3.1 ETHNICITY-TRAINING

Fig.6.7 shows the training process of the VGG 16network for the race prediction system. For this training process, the Adam optimizer is chosen and the maximum number of epochs is fixed at 120. For training purposes, 80% of database images are used and the remaining 20% of images are used for the validation process. Then finally, the training process was completed with a maximum accuracy of 81%.

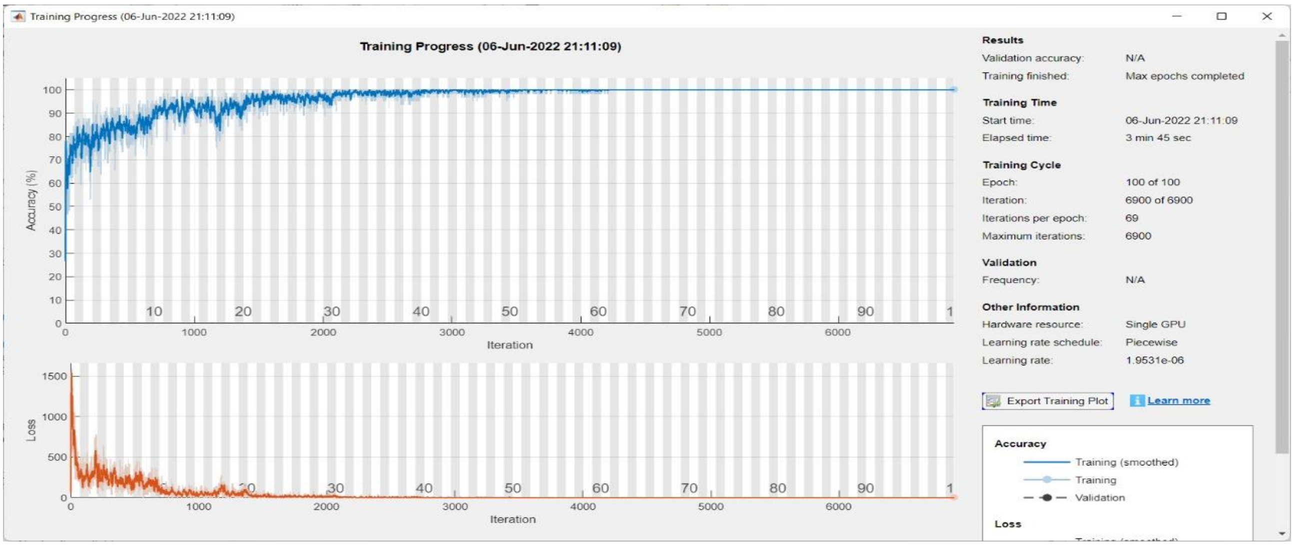

Fig 6.7 Race Accuracy

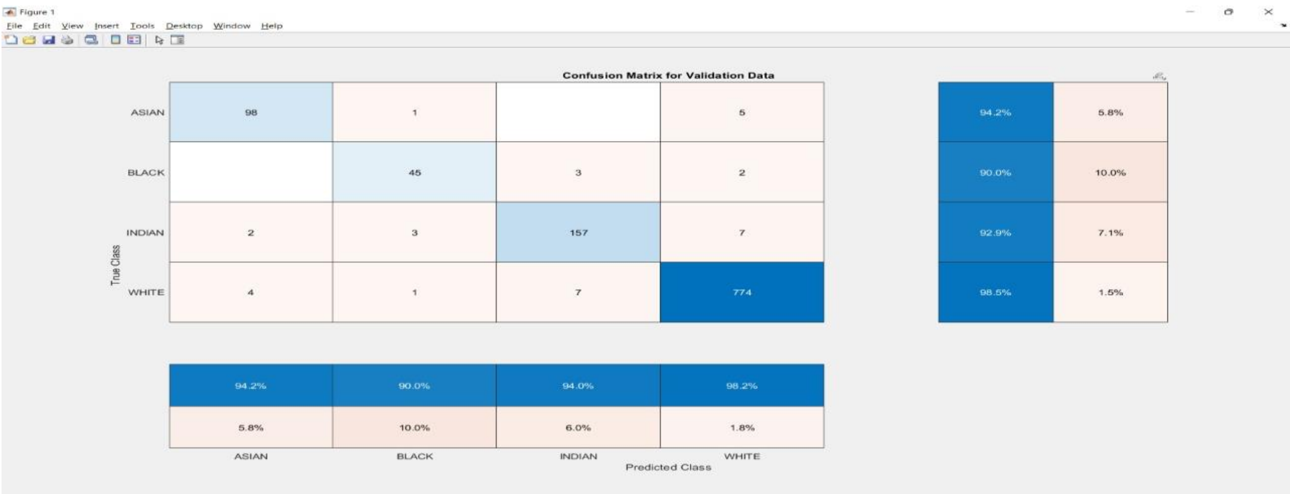

Fig 6.8 Confusion Matrix - Race

Fig 6.2 shows the performance evaluation of the age prediction system in the form of confusion matrix and the training process had a validation error of 2.98%.

6.3.2 ETHNICITY-TESTING

Open the GUI created. Click on load database, where the network which determines the race will be loaded. Then choose the image from folder images or through webcam (live image). Image chosen will be appeared on the GUI. Click on the RACE button which will show you the race.

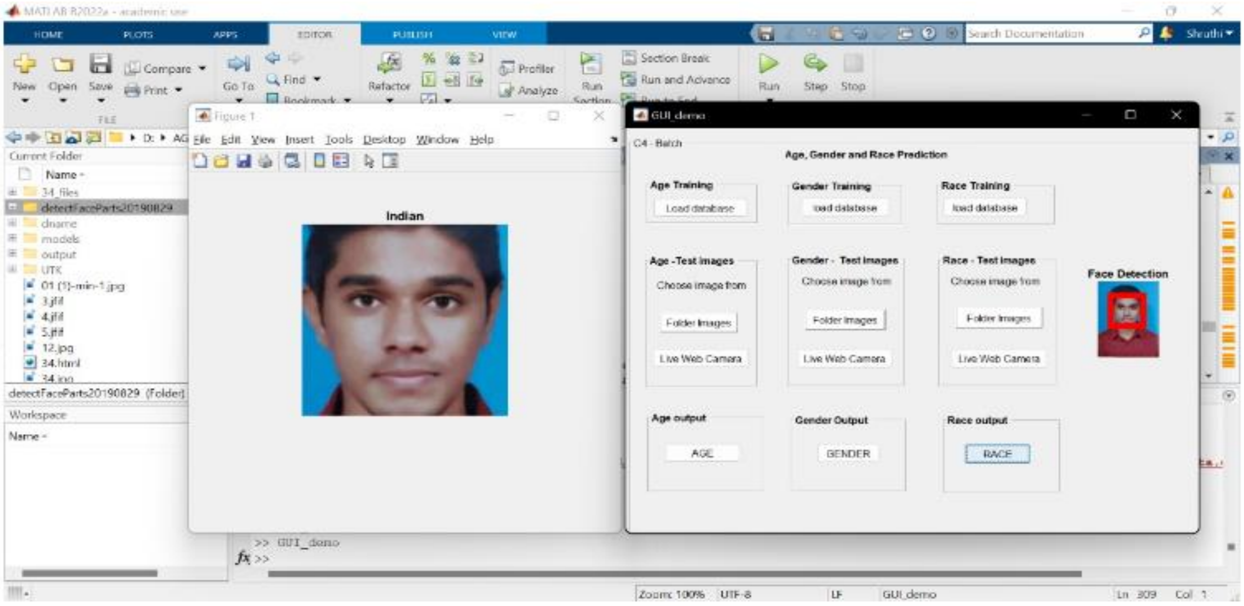

Fig 6.9 Race Classification Output

6.4 COMPARTIVE ANALYSIS

The three types of algorithm comparison table columns is shown in the final Concluded phase of our project.

SNO	ALGORITHM NAME	ACCURACY	PREDICTION	STABILITY	PERFORMANCE	RESULT
1.	VGG 16 Network	Good	Faster	Stable	High	Good
2.	Res Net 50	Better	Low	Unstable	Low	Low
3.	Alex Net	Low	Low	Unstable	Moderate	Better

Table 6.1 Comparison of Algorithm

CHAPTER 7

CONCLUSION AND FUTURE WORK

In this chapter, the results of the newly suggested algorithm VGG Network are computed using databases. Gender, Age, and Race performance accuracy is demonstrated by the system. This study proposes a generic framework for estimating automatic identification (age, gender, and race) using a database of face photos. The proposed approach utilizes a superior system performance in order to increase the estimation accuracy of demographic variables. It demonstrates that the suggested facial features-based Gender, Age, and Ethnicity estimation outperforms the majority of state-of-the-art techniques in UTK databases. In addition, the proposed method performed admirably in the difficult databases. Furthermore, the constructed real-time application exhibits the capability and efficiency of our method for extracting facial data from a standard computer.

In future different optimization and regularisation functions can be incorporated to the proposed system to increase the accuracy and also decrease the false prediction rate.

REFERENCES

[1] S. Z. Li and A. K. Jain (eds.) (2021), "Handbook of face recognition", 2nd edition. Springer Verlag, London, 2021.

[2] N. Kumar, A. Berg, P. Belhumeur, and S. Nayar. (2021), "Describable visual attributes for face verification and image search." IEEE Trans. PAMI, 33(10):1962–1977, 2021.

[3] Jain, A. K., Dass, S. C., & Nandakumar, K. (2020). "Can soft biometric traits assist user recognition?" (Vol. 5404). doi: 10.1117/12.542890

[4] Park, U., & Jain, A. K. (2020,). "Face matching and retrieval using soft biometrics." IEEE Transactions on Information Forensics and Security, 5(3), 406-415.

[5] Correa, T., Hinsley, A. W., & de Ziga, H. G. (2020). "Who interacts on the web? The intersection of user Ås personality and social media use." Computers in Human Behavior, 26 (2), 247 - 253.

[6] Kumar, N., Belhumeur, P., & Nayar, S. (2008). “Facetracer: A search engine for large collections of images with faces." In D. Forsyth, P. Torr, & A. Zisserman (Eds.), Computer vision – eccv 2020: 10th European conference on computer vision, Marseille, France, October 12-18, 2020, proceedings, part iv (pp. 340–353). Berlin, Heidelberg: Springer Berlin Heidelberg. doi: 10.1007/978-3-540-88693-825

[7] Zhuang, Z., Landsittel, D., Benson, S., Roberge, R., Shaffer, R. (2019), "Facial anthropometric differences among gender, ethnicity, and age groups." Annals of Occupational Hygiene (2019)

[8] Bruce, V., Burton, A.M., Hanna, E., Healey, P., Mason, O., Coombes, A. (2020) "Sex discrimination: How do we tell the difference between male and female faces?" Perception 22(2), 131–152(2020)

[9] Farkas, L.G., Katic, M.J., Forrest, C.R. (2018) "International anthropometric study of facial morphology in various ethnic groups/races." Journal of Craniofacial Surgery 16(4), 615–646 (2018) 76

[10] Jennifer, A., Jennifer, C., Jill, C., Philip, S., Andrew, M. (2021) "The effect of ethnicity on 2D and 3D frontomaxillary facial angle measurement in the first trimester." In: Obstetrics and Gynecology International (2021)

[11] Ramanathan, N., Chellappa, R., Biswas, S. (2019) "Computational methods for modeling facial aging: A survey." Journal of Visual Languages & Computing 20 (3), 131 – 144 (2019)

[12] Dantcheva, A., Elia, P., Ross, A. (2020) "What else does your biometric data reveal? a survey on soft biometrics." IEEE Transactions on Information Forensics and Security 11(3), 441–467 (2020). DOI 10.1109/TIFS.2015.2480381

[13] Shirakabe, Y., Suzuki, Y., Lam, S. (2020) "A new paradigm for the aging Asian face." Aesthetic Plastic Surgery 27(5), 397–402 (2020)

[14] Vignali, G., Hill, H., Vatikiotis-Bateson, E. (2021) "Linking the structure and perception of 3D faces: Gender, ethnicity, and expressive posture" In: AVSP 2003 - International Conference ,France, September 4-7, 2003, pp.193–198

[15] Gao, W., Ai, H. (2019) "Face gender classification on consumer images in a multiethnic environment." In: Advances in Biometrics, Lecture Notes in Computer Science, vol. 5558, pp. 169–178 (2019)

[16] O'Toole, A. J., Phillips, P. J., An, X., & Dunlop, J. (2020). Demographic effects on estimates of automatic face recognition performance. Image and Vision Computing, 30(3), 169-176.

[17] Cavazos, J. G., Phillips, P. J., Castillo, C. D., & O'Toole, A. J. (2020). Accuracy comparison across face recognition algorithms: Where are we on measuring race bias?. IEEE transactions on biometrics, behavior, and identity science, 3(1), 101-111.

[18] Lagree, S., & Bowyer, K. W. (2020, November). Predicting ethnicity and gender from iris texture. In 2020 IEEE international conference on technologies for homeland security (Hst) (pp. 440-445). IEEE.

[19] Vangara, K., King, M. C., Albiero, V., & Bowyer, K. (2019). Characterizing the variability in face recognition accuracy relative to race. In Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition Workshops (pp. 0-0).

[20] Terhörst, P., Tran, M. L., Damer, N., Kirchbuchner, F., & Kuijper, A. (2020, April). Comparison-level mitigation of ethnic bias in face recognition. In 2020 8th international workshop on biometrics and forensics (iwbf) (pp. 1-6). IEEE