A Deep Learning Approach Towards Soft Biometrics Attributes Prediction Using CNN

by

Maharab Kibria
17101319
Ilmi Tabassum
17101130
Fardin Ahmed
17101450
Nahian Habib
17101372

A Thesis Submitted to The Department of Computer Science and Engineering
in partial fulfillment of the requirements for the degree of
B.Sc. in Computer Science

Department of Computer Science and Engineering
BRAC University
September 2021

Declaration

It is hereby declared that

1. The thesis submitted is my/our own original work while completing degree at Brac University.
2. The thesis does not contain material previously published or written by a third party, except where this is appropriately cited through full and accurate referencing.
3. The thesis does not contain material which has been accepted, or submitted, for any other degree or diploma at a university or other institution.
4. We have acknowledged all main sources of help.

Student's Full Name & Signature:

M. Kibria

Maharab Kibria
17101319

Ilmi Tabassum

Ilmi Tabassum
17101130

Fardin Ahmed

Fardin Ahmed
17101450

Nahian Habib

Nahian Habib
17101372

Approval

The thesis titled ”A Deep Learning Approach Towards Soft Biometrics Attributes prediction using CNN” submitted by

1. Maharab Kibria(17101319)
2. Ilmi Tabassum(17101130)
3. Fardin Ahmed(17101450)
4. Nahian Habib(17101372)

Of Summer, 2021 has been accepted as satisfactory in partial fulfillment of the requirement for the degree of B.Sc. in Computer Science on September, 2021.

Examining Committee:

Supervisor: (Member)

Dr. Amitabha Chakrabarty, PhD
Associate Professor
Department of Computer Science and Engineering
BRAC University

Program Coordinator: (Member)

Dr. Md. Golam Rabiul Alam, PhD
Associate Professor
Department of Computer Science and Engineering
BRAC University

Head of Department: (Chair)

Dr. Sadia Hamid Kazi, PhD
Chairperson and Associate Professor
Department of Computer Science and Engineering
BRAC University

Ethics Statement

We, hereby declare that this thesis is based on the results we obtained from our work. Due acknowledgement has been made in the text to all other material used. This thesis, neither in whole nor in part, has been previously submitted by anyone to any other university or institute for the award of any degree.

Abstract

Any physical, behavioural or adhered human characteristics that we can observe from a person is known as Soft Biometric.The most common physical soft biometric attributes are height, age, ethnicity, facial hairs, gender, hair color etc. In this era of machine and deep learning, retrieving a person based on these semantic descriptions has become a major research interest. Face recognition and bounding boxes are now common implementations in IoT and surveillance systems because of the efficiency of training models. But the research on soft biometric attributes training models still lacks an amount. To overcome this, we have trained different CNN models for the best outcoming prediction result with a UTKface dataset. The dataset includes height, age and gender and 48x48 text pixels face images. The models include CNN, Multi-Headed CNN, DenseNet-169, Multi Label CNN and ResNet-50. After training all the models we have found that the DenseNet-169 model can achieve the most accuracy for all the soft biometric classes in our dataset. The accuracy we have achieved with our model is 96.16% for age, 97.74% for ethnicity and 99.2% for age on our UTKface dataset keeping a training loss below 0.1 for the three soft-biometric traits. All the models have been trained into the same environment and it is being uploaded with the source code to the given link below: https://github.com/Kibria10/machine-learning-works

Keywords: Deep Learning; Machine Learning;Prediction; soft biometrics; UTK-face dataset; CNN;DenseNet-169; ResNet; Keras; PyTorch; Multi Label CNN;

Dedication

We would like to dedicate this thesis to our loving parents and all amazing faculties we encountered and learnt from in the course of pursuing our Bachelor's degree.

Acknowledgement

Firstly, and foremost, we would like to thank our Almighty for enabling us to conduct our research, give our best efforts and complete it. Secondly, we would like to thank our supervisor Dr. Amitabha Chakrabarty sir for his feedback, support, guidance and contribution in conducting the research. We are grateful to him for his excellent supervision to successfully conduct our research.

In addition, we like to extend our gratitude to our faculty members, family and friends who guided us with kindness and gave inspiration and with their suggestions. Last but not the least, we thank BRAC University for providing us the opportunity of conducting this research and for giving us the chance to complete our Bachelor's degree with it.

Table of Contents

Declaration **i**

Approval **ii**

Ethics Statement **iii**

Abstract **iv**

Dedication **v**

Acknowledgment **vi**

Table of Contents **vii**

List of Figures **ix**

1 Introduction **1**
- 1.1 Introduction 1
- 1.2 Problem Statement 1
- 1.3 Aims and Objectives 2
- 1.4 Classification of Biometrics 2
- 1.5 Thesis Outline 2

2 Literature Review **4**
- 2.1 Related Works 4

3 Dataset Collection and Analysis **7**
- 3.1 Data Collection 7
- 3.2 Data Pre-Processing : 8
- 3.3 Data Analysis: 9

4 Model Implementations and Result Analysis **12**
- 4.1 Convolution Neural Network (CNN) Models Selection 12
- 4.2 CNN Model Implementation: 14
- 4.3 Multi Headed CNN Model Implementation: 19
- 4.4 DenseNet-169 Model Implementaion: 22
- 4.5 Multi Label CNN Model Implementation: 24
- 4.6 Resnet-50 Model Implementation 26
- 4.7 Result Analysis: 28

5 Conclusion **30**

Bibliography **32**

List of Figures

3.1 Our UTKFace dataset in CSV format 7
3.2 Gender Distribution 8
3.3 Ethnicity Distribution 8
3.4 Age Distribution 9
3.5 3D Channel Matrix 9
3.6 From reshaped 1D vector array to pixel image via 3D matrix conversion 10
3.7 Image presentation from Pixel Array 10
3.8 Ethnicity and Age vs Count 11
3.9 Age and Gender Density Graph 11

4.1 Convolutional neural network (CNN) architecture and the training processes with forward and backward Propagation layers 13
4.2 Sequential Model (CNN) Summary 15
4.3 Ethnicity Prediction for Sequential Model (Loss vs Epoch) 16
4.4 Gender Prediction (Loss vs Epoch) 17
4.5 Mean absolute error around 5.5 years (MEA vs Epoch) 18
4.6 Confusion Matrix of Gender Prediction 18
4.7 Confusion Matrix of Ethnicity Prediction 19
4.8 Three Headed CNN Model Summary 20
4.9 Graph of Epoch vs Loss for Age prediction 21
4.10 Graph of epoch vs loss for Gender and Ethnicity prediction 21
4.11 Age Mean Absolute Error (MEA vs Epoch) 22
4.12 Framework of DenseNet-169 [14] 23
4.13 Loss and Accuracy graph for DenseNet-169 Model (Loss vs Epoch) . 24
4.14 The model summary of Multi Label CNN 25
4.15 The training cell of multi label CNN 26
4.16 Block diagram of residual block in ResNet 26
4.17 Training result for prediction with ResNet-50 (Loss vs Iteration) . . . 27
4.18 Training result for prediction with ResNet-50 (Loss vs Iteration) . . 28

Chapter 1

Introduction

1.1 Introduction

Soft-Biometrics is the study of identifying a person based on physical or behavioral traits such as their appearance, iris pattern recognition, gait, voice identification, gender, ethnicity, height and so on. With advancement of technology, personal data security and privacy has become a serious problem for people all around the world. Biometrics system is used to secure data and verify identification and prediction . In soft biometric, data is collected from a subject to extract feature set and compares the feature set to templates in a database in order to verify a person's identity results in biometric identification, verification and prediction [8] . Human identification can be done by uniquely identifying a person on the basis of their physical attributes such as age, gender, height, face recognition. The results of soft biometric attributes are very useful in many other application domains in day to day lives which includes medical healthcare, forensic science, security, human identification. When approached with high variability situations, soft biometric information can be used with biometric identification algorithms to increase overall recognition and prediction. One notable example is visual surveillance, where face photos are typically collected in low-quality settings with considerable unpredictability and automatic face recognition algorithms do not function effectively [6]. In this case, soft biometric information classification is extremely useful for human identification and recognition.

1.2 Problem Statement

Face recognition is the most popular identification technique to identify a person but unfortunately, we can not always identify the age, ethnicity and gender of a person accurately. Also an individual soft biometric attribute is not unique enough to be used for identifying individual in a large spectrum of face dataset. By using multiple soft biometric traits we can classify the people into smaller dataset and later search an individual matching the domain of specified conditions. This method of identifying an individual will work similar to image queries as the soft biometric traits are part of human semantic description. Also, by predicting the attributes accurately the domains will results into more accurate in terms of human identification.

1.3 Aims and Objectives

Facial age estimation, gender classification and ethnicity prediction are used in various areas such as identification and prediction in surveillance videos, access control, security, law enforcement, integrity of facial images in social media, image query, human-computer interaction, intelligent advertising and many more. Age, gender, and ethnicity, for instance, can be automatically determined from a 2D facial image acquired in the visible spectrum. Reducing soft biometric attributes reduces the search space for biometric identification, gender or age-specific prediction and enhancing the biometric system's recognition accuracy, and so on [8] . Our main aim is to identify the optimum prediction model for the dataset and to pave the way for other researchers to progress with the results we have achieved by training the deep learning models. In our report, we are going to describe how accurately we can predict an individual's physical soft biometric traits, what related works have already been done, the methodologies we have used, the results we have achieved throughout the report and finally setting a new accuracy record for the implemented dataset.

1.4 Classification of Biometrics

Biometrics are divided into three categories which include: hard, soft and hidden. Classic or traditional biometrics such as faces, fingerprints or signatures represents hard biometric. The attributes which are related to faces, such as iris pattern, skin color, hair color or includes physical features such as height, weight, ethnicity or age, to wearing accessories, such as spectacles, hats, clothes represents soft biometrics [20] . Finally, attributes which are difficult to collect especially which are based on medical data, such as bio signals such as MRI images, X-Ray images or DNA are known as hidden biometrics. Soft biometric represents physical and behavioral attributes, such as gender, height, weight, age, ethnicity which are exceptional features and traits of a human [12]. They are necessary for human subjects identification, verification, prediction and description. It can be combined with other biometric attributes such as hair color, wearing glasses, walking pattern, and many more to enhance the accuracy of a biometric recognition system.

1.5 Thesis Outline

This report aims on constructing the optimum prediction model for human physical soft biometric attributes for age, gender and ethnicity. The selected dataset has the soft biometrics attributes labeled with human face data. The authors have converted the dataset into CSV format for downsampling purpose. The overall report focuses on implementing existing CNN models and custom CNN models on the downsampled dataset to reach the most accuracy from the other existing CNN models.

Firstly, In the introduction part (Chapter 1) states the importance and the application of soft biometric attributes and addresses particular problem statement, the aims and objectives of the authors and the classification of biometrics to briefly convey the report's theme.

In the Literature Review section (Chapter 2), we have gone through other researchers paperwork and found their approaches. In literature review, we focused on the methodologies and the models they have used to perform their research. There were some approaches which could not meet our requirements. Moreover, most of the papers have implemented only one or two models to get the results which is not enough for a comparative results analysis.

The Dataset Collection and Analysis (Chapter 3), states the purpose of collecting the specified dataset. In the data pre-processing section, we have discussed the modifications that was brought on the dataset and projects the soft biometrics classes (age, gender and ethnicity) condition of the dataset. In dataset analysis part, we have explained the downsamplification theory of the dataset and compared the data frequency over different attributes of soft biometrics.

In the Model Implementation and Result Analysis (Chapter 4) phase, we included the selection of implemented models and the reasons behind the selected model. This chapter describes five CNN models with proper explanations of each models and the conditions that was implemented while training the models. It also projects the training loss and accuracy graph of each of the models. Finally, the chapter comprehends a comparative result analysis among the implemented models and concludes with the newly found prediction accuracy.

Chapter 2

Literature Review

2.1 Related Works

As many works exist on both age and gender classification from facial images, this still remains an interesting subject in facial image analysis related to other facial attributes classification, prediction and analysis provides various important information. While a large number of attributes are classified and analysed on those works, we mainly focus on areas such as age, gender and ethnicity prediction which provides important factors and can not be easily modified compared to other soft biometric attributes such as hair color, clothes,gait or the use of glasses. Therefore, these attributes can be used for facial recognition, classification and prediction of soft biometrics attributes.

Age estimation, gender and ethnicity classification falls under soft biometrics that provides information of an individual's identity information. Determining a person's age by its biometric features is known as age estimation. Using images there are many ways to find out the age and gender of a person. An approach utilizing geometric features and Principal Component Analysis (PCA) was introduced in a paper by the author Khaung in 2011 [4]. In this paper mentioned some techniques which mainly emphasized on traditional face recognition and assumption to improve the parameters to get better accuracy results. Firstly, they have worked on the shape alignment for face models of their datasets. Secondly they used a "Feature Extraction Pattern" using two geometric features and the ratios were evaluated to calculate the distance between the elements like eyes, noses, and mouths. Finally, in the HSI paper they have classified the gender and age based on geometric features using the Principal Component Analysis (PCA) method for Enhancing the efficiency for facial recognition. Thus, [5] have achieved an overall prediction rate for all the 1300 experimental images is 92.5% in age prediction.

In another paper the writer has mentioned Residual Networks of Residual Networks (RoR) architecture [9] was used there as the main methods for age and gender detection. In this paper, they have advised a new CNN method for age and gender estimation using residual networks of residual networks known as RoR, which gives better optimization and classification ability for age group and gender comapred to any other CNN models. In this paper first the RoR model is pre-trained on ImageNet, and then it goes on the IMDB-WIKI-101 data set for learning the features of facial images. Lastly, they have used this model for further fine-tuning on Audience data set. Finally, the RoR-152+IMDB-WIKI-101 results on the Audience benchmark. Their exact accuracy varies from 92.43% but our model after training gives better results and we have trained our model for three parameters which includes age, gender and ethnicity.

In another paper, they find out the advantages and limitations of the automated classification of soft biometric traits in Near Infrared (NIR) long-range to differentiate and predict night-time or low resolution face images. The impact of these traits in terms of gender and ethnicity is mentioned in this paper in detail and how to improve traditional face recognition and enhance prediction results [7]. The main database is collected from NIR bands at low light and at variable distances and has used CNN to perform the classification for ethnicity and gender.

In this paper we have learned about the two-stage system,where the CNN predicts age and gender from all photos and extracts facial features which are required for face identification and prediction.In this paper they have modified the MobileNet, trained to perform face recognition [15], in order to get information about age and gender traits. In the second stage of their approach, the extracted face data are grouped into different stages. The birth year to get the age and gender of the person in every cluster are estimated using predictions from each photo. In this work, the first stage of the model based on MobileNet is used to recognize age and gender. In the second stage they have performed hierarchical agglomerative clustering to group faces. Finally, the results were refined with these clusters to get more accurate results. In this paper the models are pre-trained using ImageNet models. They applied a multi-output neural network to predict traits simultaneously and the final loss function is used on basis of the combination of age and gender losses [17]. In their approach, they have shown that the model has better efficiency than other non-pre-trained model.

We were inspired by the work of Kumar et.al. [3], where a total of 10 regions human images were used to classify 73 attributes. The main motive was to make the face recognition process more efficient. Edge magnitude, pixel intensities and orientation were used as low features to represent the facial components. Moreover, using a computationally expensive procedure, trained and evaluated with different classifiers to select the best combination for a given attribute. In this work, we propose to use keras sequential models which have shown to be powerful for training the attributes such as age, gender and ethnicity.

We have found some related works related to soft biometrics which focus on age and gender classification from images.For their work, they have used a multi-agent system [10] as well as a face recognition algorithm to classify the age and gender from the image. Moreover, to make the algorithm work with a high success rate it has to do preprocessing and training. To classify the age they have used fisherfaces which are implemented on an open CV framework. To classify different age groups, they have used children who are below 18 years, youths aged between 19 to 30 years, adults as 31 to 70 years and the elderly above 70 years. The success rate is between

80 to 93 percent depending on which face detection algorithm they are using.

We also got some vast knowledge from A. Fariza [13] where they have used Deep Residual Network to estimate the age. To estimate the age, they have used UTK-face dataset to perform the performance of the residual network. In the dataset, they have used 20,000 face images of people to estimate the age of the individuals. The researchers have mentioned that CNN is being used to identify age estimation as well. Moreover, they have referred to other researchers who have used CNN to estimate the age. In the dataset the age of the people were between 1 to 100 years old. The data length was 21,301 and testing data was 2,362. They have used 3 methods to train their dataset. From the 3 methods they have got 99.15% accuracy from Resnet-50 and ResNeXt-50 (32x4d).

In the papers of Sooyeon Lim [16], the researcher tries to predict the traits using CNN face recognising algorithm. In this method, the images are given as input. After that, the faces are detected. After detecting the face, it is pre-processed. The images are then trained using CNN. Finally age and gender is estimated. The researcher has used the IMDB learning dataset. There are 73,746 high quality images in the dataset. The face detection rate was 94.8%. Using this dataset, 98.4% gender estimation has been achieved. The accuracy of age estimation was 85%.

Classification and analysis studies have utilized face images which were captured in controlled conditions. Many variations may appear in the image of a person's face and these variations may affect the ability to estimate the age or recognize the gender. They are due to many factors which can be divided into: human factors such as race, facial expression etc and capture process factors such as pose, illumination, quality etc. Age estimation, ethnicity and gender classification is done by treating each face independently and is focused on appearance features.

Chapter 3

Dataset Collection and Analysis

3.1 Data Collection

Collecting a wide range of dataset is one of the major study into deep learning, and most essentially, gathering information from a tenable source with clear agreement is significant. We All know deep learning algorithms highly depend on data collection so data preparation is the most important procedure in terms of deep learning algorithms. Data preparation involves defining the appropriate method for gathering data. So, proper data collection and data preparation is a must before training our models. While searching for a dataset we considered the following attributes: Age data, Ethnicity data and Gender data and Facial Images. The reason behind searching such dataset is because these are the most major physical soft-biometric attributes.

UTKFace created this collection of data based on the age, gender and ethnicity. The dataset contains a total of 23705 person's data that includes their age, gender, ethnicity and the face image. The dataset was labeled with age, gender and ethnicity values with facial images in .jpg format. We have converted the images into RGB pixel values and then converted the respective classes into numerical values to build a CSV dataset file of the UTKface dataset(Figure 3.1).

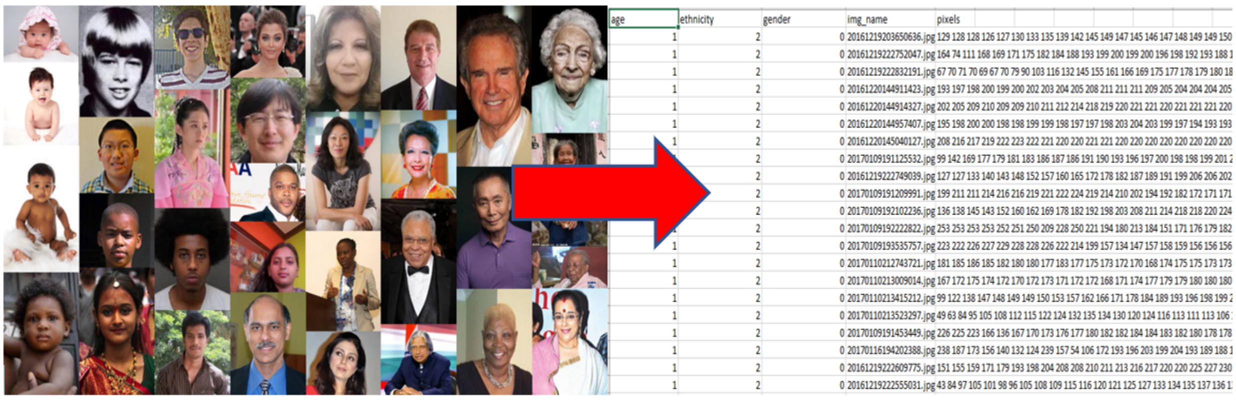

Figure 3.1: Our UTKFace dataset in CSV format

3.2 Data Pre-Processing :

In this stage we searched for any missing values or any other errors that might create problem while training the dataset. Then we applied normalization to fit the images. After applying the normalization method, we distributed the dataset.

According to dataset the parameters such as Age, Gender and Ethnicity are label and separated among all these 23705 data there are 12391 data from male (labeled as 0) and 11314 data from female (labeled as 1). The distribution clearly shown in the barchart showing clear separation between the male and female distribution given below (Fig 3.2).

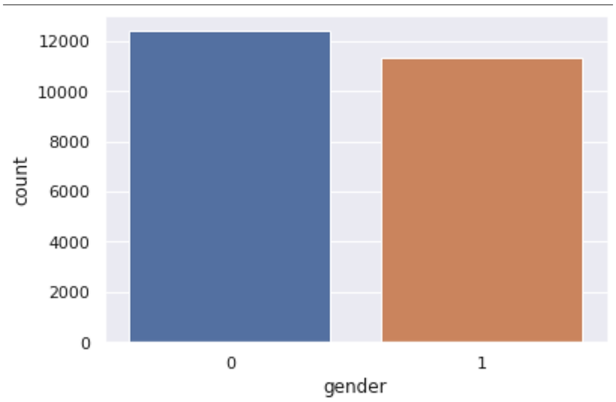

Figure 3.2: Gender Distribution

According to ethnicity class there are 10078 data of White (labeled as 0), 4526 data of Black (labeled as 1), 3434 data of Asian (labeled as 2), 3975 data of Indian (labeled as 3) and 1692 data of Hispanic (labeled as 4).The ethnicity is divided in the (Fig 3.3) shows the Ethnicity Distribution.

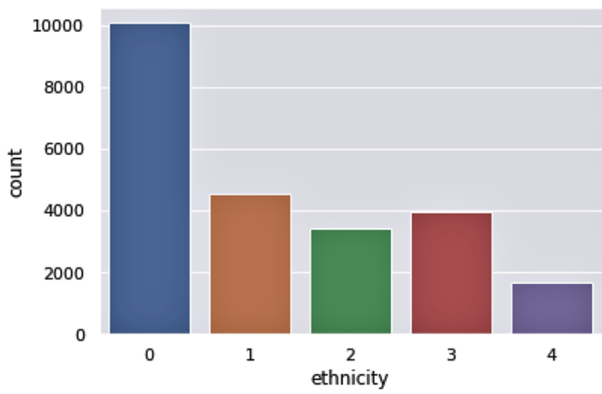

Figure 3.3: Ethnicity Distribution

The graph below shows age distribution over a range from 1 year to 100 years. According to age class of the dataset, there is a majority of 5893 people between the range 24 to 30 years. This number is higher than the mean average data rate for other ages (Fig: 3.4) . It can create overfitting if we try to train the age estimation model without labeling it properly.

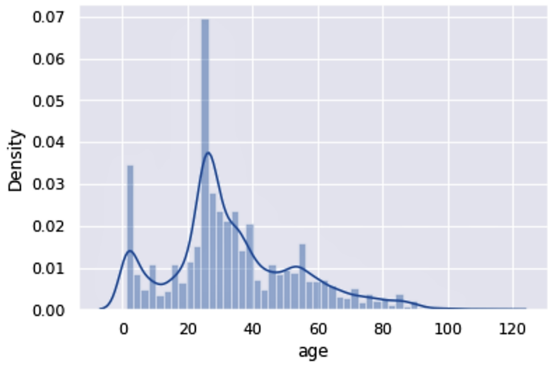

Figure 3.4: Age Distribution

3.3 Data Analysis:

To analyze our data, we imported NumPy, Matplotlib, Pandas and Plotly for visual data representations and segmentation. After distributing the data according to age, gender and ethnicity we converted the pixel array from 1D to 3D Channel. The 3D pixelated matrix below in (fig:3.5) contains a row, a column and a channel.

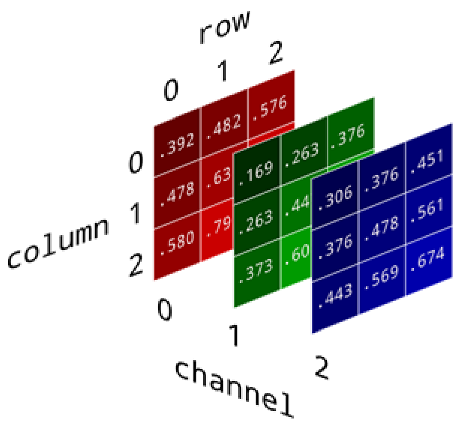

Figure 3.5: 3D Channel Matrix

The 3D pixelated matrix contains a row, a column and a channel. Here in the figure (Fig 3.6) the row and column contain the color density and the 3 channels contain Red, Green and Blue color subsections accordingly. With the integration of the values in these pixels a full visual image is reconstructed.

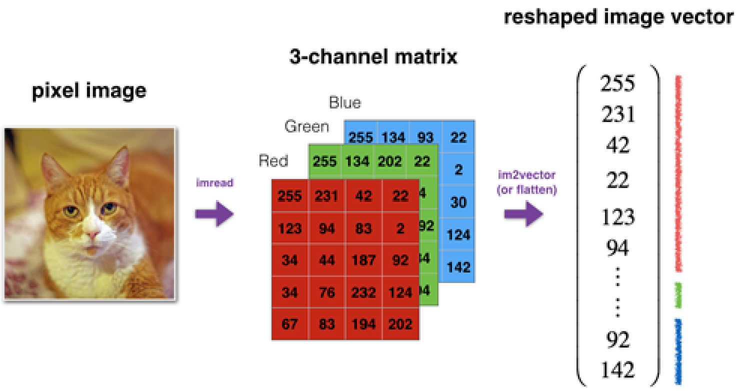

Figure 3.6: From reshaped 1D vector array to pixel image via 3D matrix conversion

Finally, after plotting this pixel arrays of our dataset on X-label and Y-label (matrix alignment), we get the images with their Age, Gender and Ethnicity (Fig 3.7). We have total 48 pixels for one image . We divided the 48 pixels into three 4*4 matrix for Red,Green and Blue Scale. After aligning the pixels according to these scale we finally get a fully formed image.

Figure 3.7: Image presentation from Pixel Array

Ethnicity and Age Density: In this set (figure 3.8), we can see that the maximum amount of face images are set into one race than the other four ethnic groups.This might create a problem while training the models as it will overfit data for that particular labeled data.

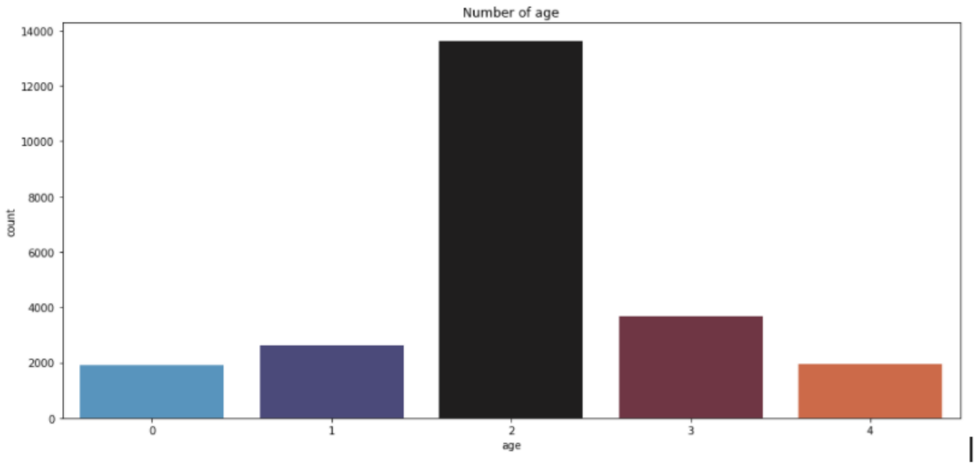

Figure 3.8: Ethnicity and Age vs Count

Age and Gender Density:

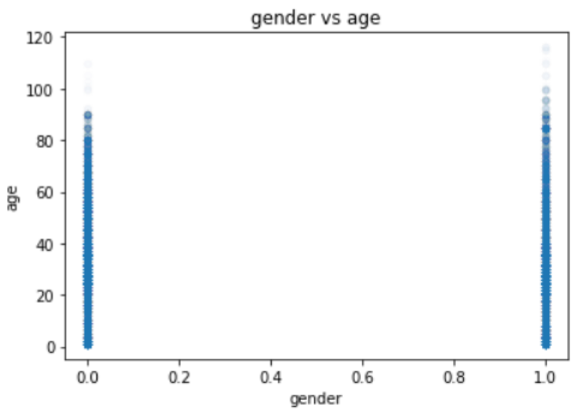

Figure 3.9: Age and Gender Density Graph

The figure 3.9, is a visual representation graph of Gender VS Age. We see there are only two specific genders but no mixed gender of any age.So,from this analysis we get that we have a boundary on detecting mixed ethnicity and mixed gender. We have defined 0 as male and 1 as female. Moreover, we have collected data of different ages. We have also found more aged people in the female category than male category.

Chapter 4

Model Implementations and Result Analysis

4.1 Convolution Neural Network (CNN) Models Selection

Convolutional layer has filters which do convolution operations and to do this operation it requires volume size of input image, filter size, stride, zero padding as parameters. With all the information as parameters, it calculates and gives an output which goes to the next layer of CNN. It is a subsection of Deep Learning which is mainly used for image classification and prediction. Many other deep learning algorithms are used nowadays for image classification and prediction but CNN gives better performance due to its architecture and other advanced features. CNN is mainly divided into three layers.The very first layer is called the input layer, the second layer is the hidden or covered layer and the final layer is the output layer. Most of the calculation is done in the covered layer. In CNN neurons of one layer are connected to some other particular neuron of its immediate next layer but in other deep learning algorithms, one neuron is attached with all the other neurons of the next layer which become very costly to calculate. Convolutional Neural Network (CNN) has few layers which work together to classify images and give better accuracy. It has a convolutional layer, pooling layer, ReLU layer and fully connected layer. We have implemented CNN in our dataset with Keras sequential model (chapter 4.2) and multi headed CNN (chapter 4.3) has been used in our work as it uses a linear stack of layers for image classification and predictions. We also implemented DenseNet-169 (chapter 4.4) to classify the soft biometric attributes. In the later chapter (chapter 4.5) we have implemented Multi Label CNN because in a multi label classification model it can predict mutually non-exclusive classes with different labels. Finally, we ended our training models with ResNet-50 (chapter 4.6)

Keras.layer.Conv2D: It is known as 2D Convolution Layer, this layer creates a convolution kernel that is wind with layers input which helps produce a tensor of outputs. In image processing (figure: 4.1) Conv2D is a convolution matrix or mask which can be used for edge detection, sharpening, blurring, embossing and more by doing a convolution between a kernel and an image.

Figure 4.1: Convolutional neural network (CNN) architecture and the training processes with forward and backward Propagation layers

A CNN is composed of several building blocks which includes convolution layers, max pooling layers, and fully connected layers. A model kernels and weights is calculated with a loss function through forward propagation on a training dataset, and learnable parameters are updated according to the loss value through backpropagation with gradient descent optimization algorithm and ReLU (rectified linear unit).

ReLU Layer: ReLU is an activation function in convolution neural networks and works as a function. If any negative value is given to the function, then the function converts the value into zero. But if any positive value is given, then it remains the same. This ReLU layer works in parallel with the convolution and pooling layer. We have used this function in our model to prevent all the negative values in the horizontal axis from being mapped to 0 in the vertical axis. Basically it helps us to get rid of negative values such as if the age by mistake is given negative the ReLU function will eventually reject those values and make them 0 as it allows non-negative values only. In our model all the activation layers are set to Relu.

Pooling Layer: Pooling layer is used to reduce the size of images which helps to classify images of our dataset easily. In this pooling layer, it takes filter size and stride as parameters. depending on the filter size and stride the output changes. Whatever image size is given in the convolution layer, it becomes half of that after the pooling layer. Three categories of pooling exist where the first one is known as the Max Pooling which estimates the highest pixel value among the batches, Second one is the Min pooling which estimates the lowest pixel value among the batches. Final one is the Average Pooling which averages all the pixel values in the batch.

MaxPooling2D(2,2): has been used as we are using Max Pooling in a 2D space. MaxPooling2D is used to get the maximum value within the 2D space and shows it as output. Here (2,2) means the pooling size is 2.

Batch Normalization layer: We used batch normalization as a technique for training a dataset that standardizes the inputs to a layer for each sub-batch. Batch Normalization layer has the effect of stabilizing the training process and reducing the number of training epochs that are required to train our datasets.Batch Normalization layer gives us the privileges to train independently every layer of our CNN. It normalizes the previous layer output. Batch normalization layer makes the learning more efficient as it is implemented as a regularization so that we can avoid overfitting the model by reshaping it. Batch Normalization is integrated into our sequential model to regulate the data input and data output. We are using Batch normalization here so many times.

Image Flattening : We converted the Image into an array before processing the information to our model. We are working with a dataset that contains a big amount of images. It means to flatten means all the layers are stacked and merged together into a single layer. Any hidden layers and layer effects are usually discarded in the process. Flattening and fully-connected layers are at the last stage of CNN. This is how flattening helps in reducing the memory and decreasing the time to train the model as well.

Dense layer : is the regular deeply connected neural network layer. It is the most common and frequently used layer in our training process. Dense layer does the below operation on the input and returns the output. For the multi class classification activation function is softmax we have used a dense layer in our model to get better results. Dense layer adds to the fully connected layer to the neural network CNN. Dense Layer is a Fully Connected Layer which describes how the neurons are connected to the next layer of neurons of an intermediate layer. [19] In our model we have used a dense layer of 5, 64 to integrate the image flattening. Dense adds the fully connected layer to the neural network.

4.2 CNN Model Implementation:

A sequential model is considered where response categories are reached successively step by step. The sequential models are derived from assumptions about an underlying step-wise response mechanism. The Sequential model is made up of a series of layers that are stacked in a linear fashion. The model must recognize the input shape that it will accept [1]. This is why the first layer of a Sequential model and the first subsequent layers can perform automatic shape inference requires input shape information.For the model epochs are assigned as 25 and the batch-size is set as 64. The callback functions are set for the sequential models. As sequential models work only with single data input and single data output, we built specific models for gender, age and ethnicity separately and trained all three attributes on the sequential model(Figure 4.2).

Our aim is to try and minimize the loss function, by updating the values of the parameters and making our predictions as accurate as possible. Optimizers are the algorithms or methods which are used to change the different attributes such as age, gender and ethnicity according to the training rate in order to reduce the losses. We have trained our dataset using three different Optimizers which includes Stochastic gradient descent, RMSprop(Root Mean Square Propagation), and Adam Optimizer. For better results we have increased the parameters of our model for training to get more accuracy from our training dataset.

To see the importance of functions in neural network algorithms, in the sequential

```
Model: "sequential_3"
_________________________________________________________________
Layer (type)                 Output Shape              Param #
=================================================================
conv2d_13 (Conv2D)           (None, 46, 46, 64)        640
_________________________________________________________________
max_pooling2d_10 (MaxPooling (None, 23, 23, 64)        0
_________________________________________________________________
batch_normalization_10 (Batc (None, 23, 23, 64)        256
_________________________________________________________________
conv2d_14 (Conv2D)           (None, 23, 23, 128)       73856
_________________________________________________________________
conv2d_15 (Conv2D)           (None, 21, 21, 128)       147584
_________________________________________________________________
max_pooling2d_11 (MaxPooling (None, 10, 10, 128)       0
_________________________________________________________________
dropout_9 (Dropout)          (None, 10, 10, 128)       0
_________________________________________________________________
batch_normalization_11 (Batc (None, 10, 10, 128)       512
_________________________________________________________________
conv2d_16 (Conv2D)           (None, 10, 10, 256)       295168
_________________________________________________________________
conv2d_17 (Conv2D)           (None, 8, 8, 256)         590080
_________________________________________________________________
max_pooling2d_12 (MaxPooling (None, 4, 4, 256)         0
_________________________________________________________________
dropout_10 (Dropout)         (None, 4, 4, 256)         0
_________________________________________________________________
batch_normalization_12 (Batc (None, 4, 4, 256)         1024
_________________________________________________________________
conv2d_18 (Conv2D)           (None, 4, 4, 512)         1180160
_________________________________________________________________
max_pooling2d_13 (MaxPooling (None, 2, 2, 512)         0
_________________________________________________________________
dropout_11 (Dropout)         (None, 2, 2, 512)         0
_________________________________________________________________
batch_normalization_13 (Batc (None, 2, 2, 512)         2048
_________________________________________________________________
flatten_3 (Flatten)          (None, 2048)              0
_________________________________________________________________
dense_6 (Dense)              (None, 128)               262272
_________________________________________________________________
dropout_12 (Dropout)         (None, 128)               0
_________________________________________________________________
dense_7 (Dense)              (None, 1)                 129
=================================================================
Total params: 2,553,729
Trainable params: 2,551,809
Non-trainable params: 1,920
```

Figure 4.2: Sequential Model (CNN) Summary

model for training of age and ethnicity attributes of the dataset we have used softmax and sigmoid function is being implemented on both of the models on the gender attribute of the dataset. Softmax allows multi classifications in our linear regression models of age and ethnicity. Sigmoid allows us to binary classify our gender model. In our callback function, reduced learning is implemented. When a statistic no longer improves, lower the learning rate. When learning becomes stagnant, models often benefit from slowing the learning rate by a factor of 2-10. This callback checks quantity and reduces the learning rate if no improvement is noticed after a 'patience' number of epochs. The number of epochs is the measure of patience. The learning rate is reduced. Patience is the number of epochs with no improvement after which learning rate will be reduced; this is set to patience=2 whereas the verbose can be either 0: quiet or 1: update messages. The verbose is set to 1 for both the sequential models.

Graphs of sequential model: The graphs given below are of the sequential CNN model: The figures below show the graph for Percentage against epoch. The graph shows results such as loss(represented by blue line), accuracy(represented by orange line), val-loss(represented by green line), val-accuracy(represented by red line), Learning rate(represented by purple line).

Model for Ethnicity Prediction : In figure: 4.3 (loss vs epoch) below the val-loss decreases rapidly with the increase in the number of ecoph. After 20 out of 25 epochs we get a loss: 0.3031, accuracy: 0.8907, val-loss: 0.5961 and val-accuracy: 0.8127. In the model below we can see that a sharp decrease in the val-loss from 35% to 7.5% and accuracy and val-accuracy is increasing as we do further training.

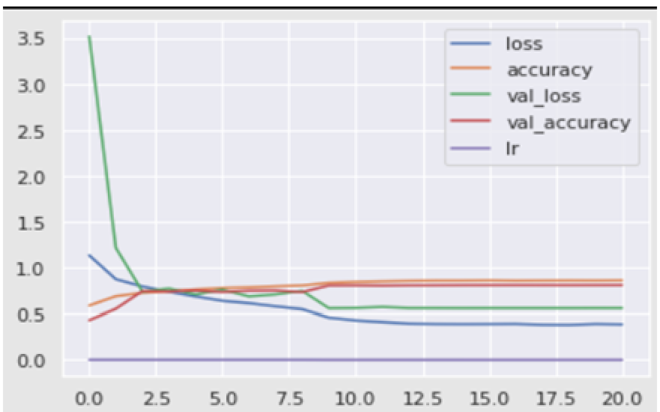

Figure 4.3: Ethnicity Prediction for Sequential Model (Loss vs Epoch)

Model for Gender Prediction : In the figure: 4.4 (Loss vs Epoch) of the Gender Prediction given below , after 21 out of 25 epochs we get loss: 0.1467, accuracy:

0.9391 , val-loss: 0.2369 and val-accuracy: 0.9095 shows a rapid drop in loss and val-loss as it decreases from 1400 to 200 after completing 5 epochs. The val-accuracy and accuracy is increasing with the number of epoch as the model is being trained further which leads to better accuracy. The red line shows an increase in accuracy from 500 to 800 which shows an increase in accuracy. From the graph below we can see a loss: 23.69% and the accuracy: 90.95% off the gender prediction.

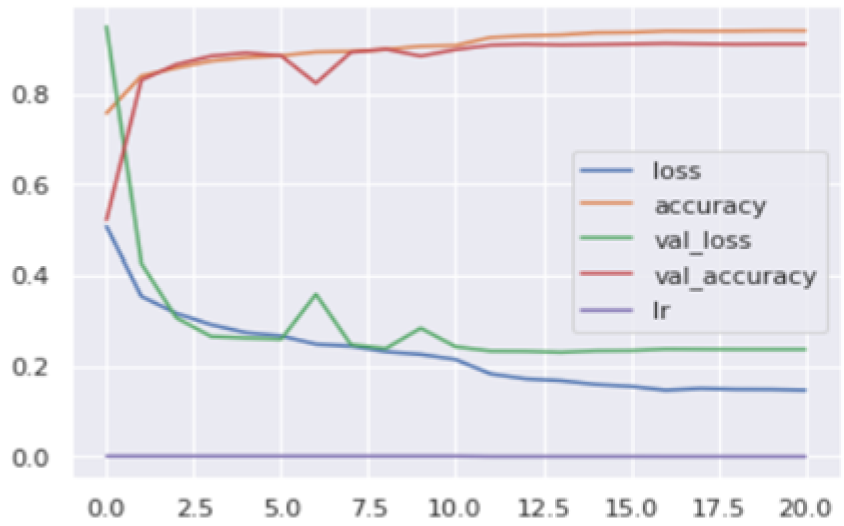

Figure 4.4: Gender Prediction (Loss vs Epoch)

In the figure 4.5 below we have calculated the absolute mean error which is a way to measure the accuracy of a Gender prediction model [18]. Effectively, MAE (Mean Absolute Error) describes the typical magnitude of the mean absolute error around 5.5years. The graph shows the mean age of 33 years which shows 17% of all the data we have collected.

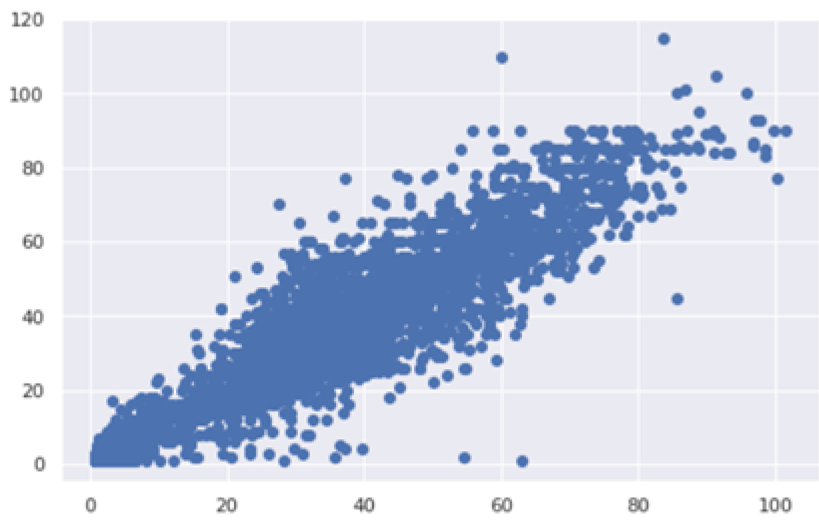

Figure 4.5: Mean absolute error around 5.5 years (MEA vs Epoch)

From the Figure: 4.6 [11] below shows a confusion matrix is a tabular summary showing displaying the number of accurate and erroneous predictions produced by our gender prediction model's classifier. It is used to calculate performance measures like accuracy and precision to assess the performance of our gender prediction model. In the below matrix there is a part showing false results as variation of children via gender using only images is different which results in some inaccuracy.

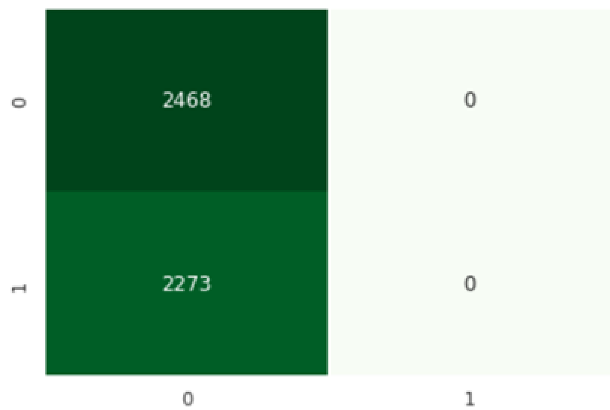

Figure 4.6: Confusion Matrix of Gender Prediction

From the Figure: 4.7 below shows a confusion matrix which is in a tabular summary showing the number of correct and incorrect predictions made by a classifier for our age prediction model. It can be used to evaluate the performance of our age prediction model through the calculation of performance metrics like accuracy and precision. In the below matrix, there is data showing data in the graph fill diagonal with 0 for better visualisation.

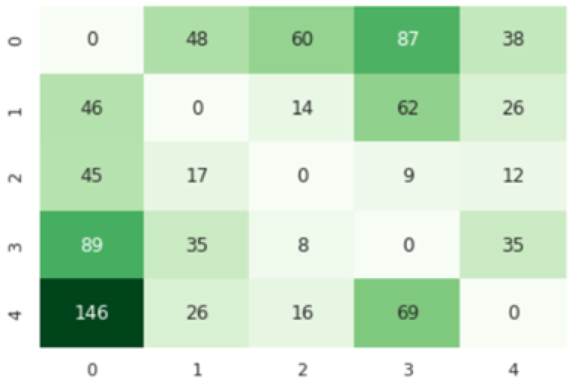

Figure 4.7: Confusion Matrix of Ethnicity Prediction

4.3 Multi Headed CNN Model Implementation:

We have gained good loss and accuracy results from the previous CNN implementation. So, we have decided to implement another CNN model with a different structure. In this model, we are also taking in account the limitations of our dataset, as the ethnicity data has a major imbalance in ethnicity class 0 and also has a slight imbalance in ethnicity class 4. That's why we have stratified the ethnicity class data. Three heads CNN is a Multi Headed CNN where we dense and do multiple classification from a single flow Polling and flattening. The reason we are implementing this model is to see if we can get better accuracy by applying dense layers for both age and ethnicity. As in the previous CNN model we could not confirm the accuracy by applying any extra dense layer. In our model, we used 3 classifiers: age, gender and ethnicity which can be stated as our three heads in our CNN as we are classifying each later after processing. After pooling and flattening the dense layer, we sum up the fully connected layer to the neural network in three different classifiers. In the figure 4.9, we can see that our data is going through the input layer for one time and then it passes Conv2D , Dropout2D , Max Pooling , Spatial Dropout2D, Batch normalization for several times and then it goes for flattening. After flattening the flattened data our dataset is going to be dense in 3 individual Classifiers which are age, ethnicity and gender and gives us 3 individual output based on these classifiers. Here (Figure 4.8) is the diagram model of multi-headed CNN approach.

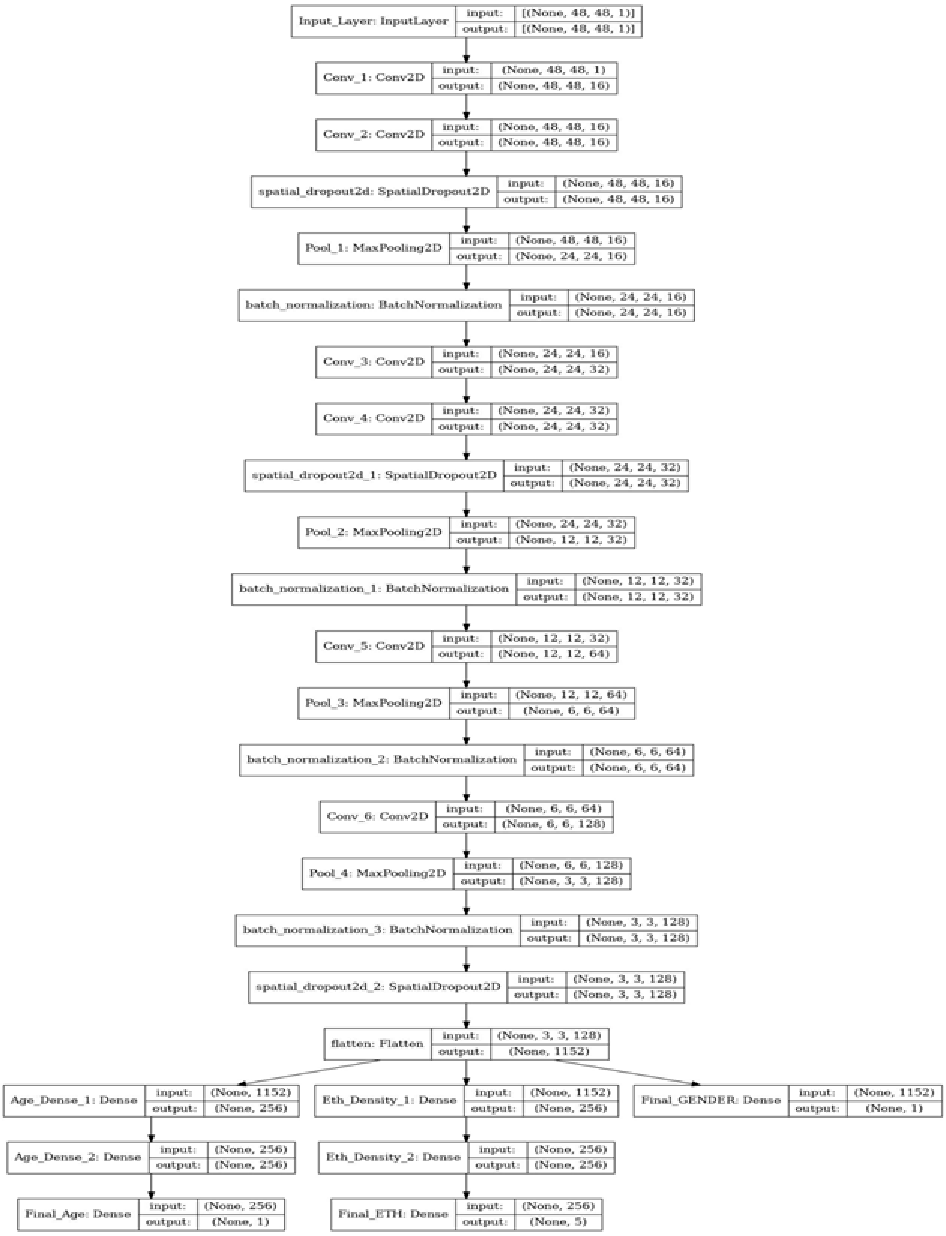

Figure 4.8: Three Headed CNN Model Summary

The model has been trained for a total of 50 epochs to reach its accuracy. Each and every class layer was later implemented so that we get better result and below we have described each graph and the result we have achieved after training the model several times.From the Figure: 4.9 below shows loss of 72.0958%, Final-Age-loss of 70.9618% and Final-Age-mean-absolute-error: 6.1624%. Mean Absolute Error of age from the graph is 6.1.In the diagram below the blue line represents Training of the dataset which we have continued till epoch 17 out of 50 epochs we have trained. After training the dataset several times the training loss in our model has fallen from 350 to 50. The validation loss has decreased from 270 to 100 after 4 epochs.

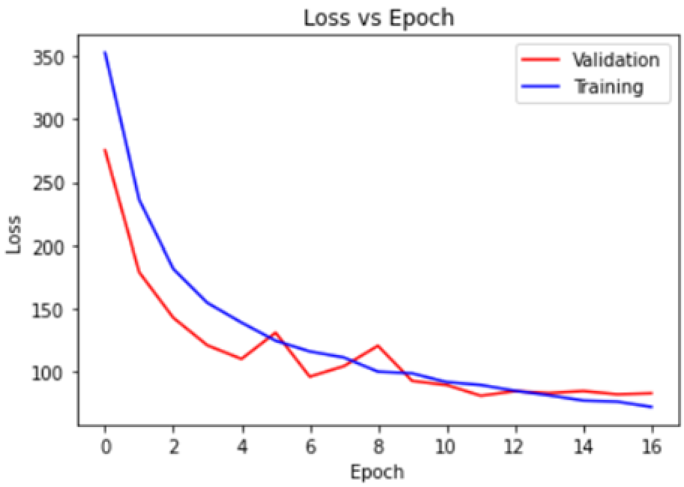

Figure 4.9: Graph of Epoch vs Loss for Age prediction

From the Figure: 4.10 below shows loss and epoch graph where the blue line represents the gender and the orange line represents the ethnicity. The value loss for ethnicity model becomes constant from 0.6 to 0.7 as the loss is becoming constant and after training the model using Multi headed CNN. The gender loss increases from 0.75 to 0.85 but as the epoch is increasing the rate of loss becoming constant.

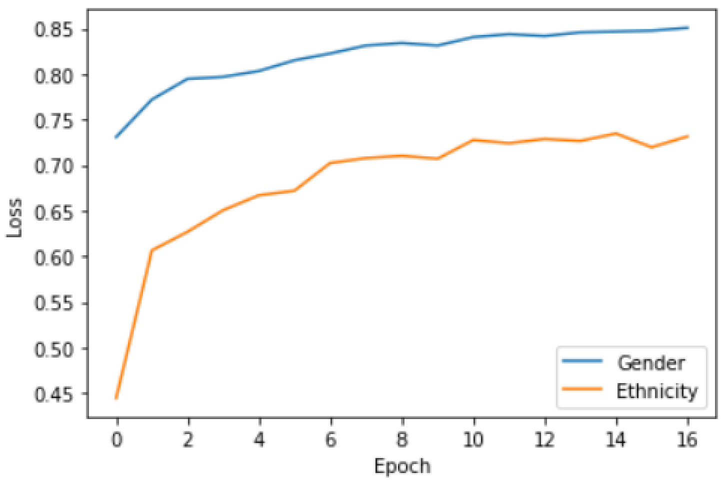

Figure 4.10: Graph of epoch vs loss for Gender and Ethnicity prediction

From the Figure: 4.11 shows val-Final-Age-mean-absolute-error which was 12.2331% changes to val-Final-Age-mean-absolute-error: 6.4658% The Mean Absolute Error falls from 14% to 6% after 17 of 50 as shown in the graph above.

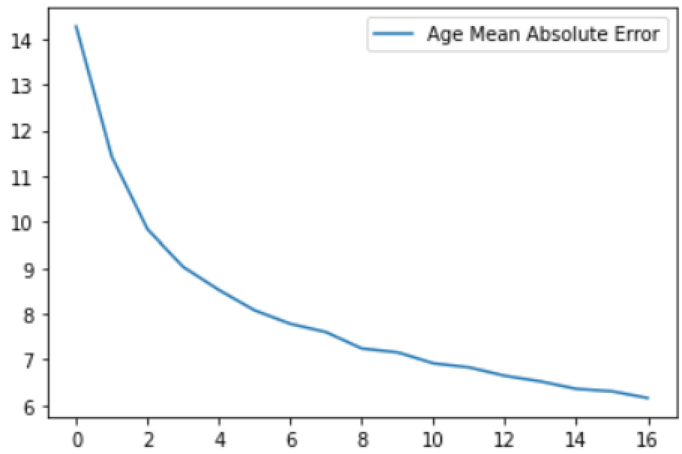

Figure 4.11: Age Mean Absolute Error (MEA vs Epoch)

4.4 DenseNet-169 Model Implementaion:

Densenet-169: DenseNet-169 model is one of the DenseNet group of models designed to perform image classification. The densenet-169 uses a large model which results in increased size and accuracy of the model. We use DenseNet as it can handle vanishing gradient problems compared to the other models. In multi-output classification the network branches creating multiple sets of fully-connected heads at the end of the network at least twice or more times. Our network follows multi-output classification such as ethnicity, age and gender to predict a set of class labels for each head, making it possible to learn disjoint label combinations. Thus unlike the previous sequential models we do not need to build class models for the soft-biometric attributes. The model is trained with the architecture of DenseNet-169 for three branches including age, gender and ethnicity estimation which is used as the base of the pre-trained model. The dataset trained under this model gives accurate results after setting a specific epoch which is 80. In this model we have converted the age into categorical values to categorise the age range and solve it as a classification problem instead of a regression problem. Throughout this model we have used 'adam' optimizer where both the learning rate and decay rate is already initialised. If age is trained as a regression model, the age values fluctuate too much during inference so we have to solve the age range using regression. While training for age and ethnicity, softmax is used for multi-classification in the logistic regression model. For the gender class, sigmoid function is used for binary classification in the logistic regression model. The binary classification of gender refers to female and male. During our training we make a gender branch for gender prediction, age branch for age estimation and ethnicity branch for ethnicity prediction on top of the base model. The loss value that will be minimized by the model will therefore be the weighted total of all individual losses, weighted by the loss-weights coefficients, to determine the weight loss contributions of distinct model outputs. The (Initial learning rate) init-lr is set to 1e-4. While training we split the data frame into 80-20% training and validation dataframes. The training data shape is (17449, 48, 48) (17449,) (17449,) (17449, 5) whereas Validation data shape is (4362, 48, 48) (4362,) (4362,) (4362, 5) respectively. Initially we set an epoch of 80 but after reaching 58 we get the result and the learning rate is over. This is because we follow the procedure decay based on the learning rate divided by the number of epochs, to slowly decrease the learning rate over the epochs.

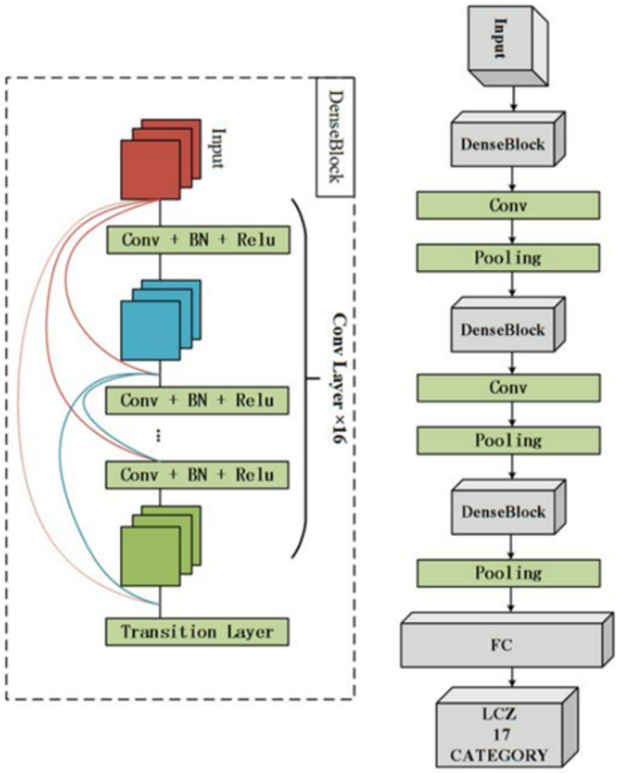

Figure 4.12: Framework of DenseNet-169 [14]

Loss and Accuracy Plots: The figure 4.13 shows the result of Loss and Accuracy plots which are plotted following the correlation between the soft biometric attributes of gender, age and ethnicity. The orange line represents the validation of loss or accuracy for all the attributes. Similarly, the blue line represents training loss or accuracy for all the attributes mentioned above. The result we have achieved after training the model is divided into three classes for age, gender and ethnicity. Gender training Loss of 0.1495, gender-output-binary-accuracy: 0.9920, ethnicity training loss: 0.4719, ethnicity-output-categorical-accuracy:0.9774,val-loss:1.4380,val-gender-output-loss:0.2735,val-ethnicity-output-loss:0.6661, val-gender-output-binary accuracy: 0.9296.

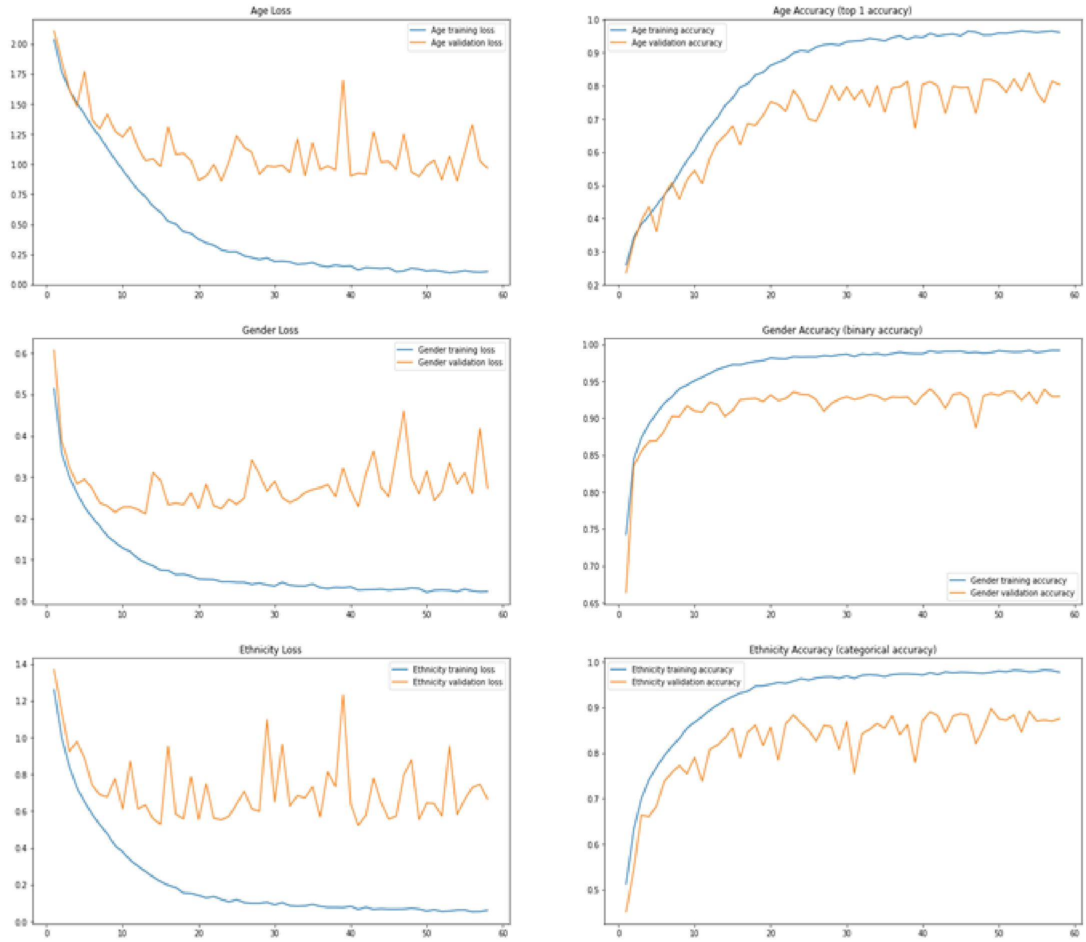

Figure 4.13: Loss and Accuracy graph for DenseNet-169 Model (Loss vs Epoch)

The figure 4.13 The Validation loss shown by orange line and Blue line shown by training loss. In the graph Age Training Loss decreases from 2 to 0.15 as the epoch increases and Age Accuracy validation loss remains constant. The graph shows the gender training loss decreases 0.5 to 0.05 from the gender accuracy rises from 0.65 to 0.95. The ethnicity loss decreases from 1.2 to 0.1 but the training loss for ethnicity validation 1.4 to 0.6 remains the same.

4.5 Multi Label CNN Model Implementation:

In our multi label CNN approach we have used both CNN and DNN to build the model. Here's the model summary. Multi label CNN is used for image classification and prediction. We have used this model for better image prediction for gender, age and ethnicity prediction. The layers used so far consist of different parameters to get better classification and prediction of the image. The dataset is used and trained under different conditions to increase accuracy and get optimum prediction results. In figure 4.14, the model's CNN and DNN layer summary has been described.

```
----------------------------------------------------------------
        Layer (type)               Output Shape         Param #
================================================================
            Conv2d-1           [-1, 32, 48, 48]             320
         MaxPool2d-2           [-1, 32, 24, 24]               0
         LeakyReLU-3           [-1, 32, 24, 24]               0
            Conv2d-4           [-1, 64, 24, 24]          18,496
         MaxPool2d-5           [-1, 64, 12, 12]               0
         LeakyReLU-6           [-1, 64, 12, 12]               0
            Conv2d-7          [-1, 128, 12, 12]          73,856
         MaxPool2d-8            [-1, 128, 6, 6]               0
         LeakyReLU-9            [-1, 128, 6, 6]               0
           Conv2d-10            [-1, 256, 6, 6]         295,168
        MaxPool2d-11            [-1, 256, 3, 3]               0
AdaptiveAvgPool2d-12            [-1, 256, 1, 1]               0
           Linear-13                  [-1, 128]          32,896
        LeakyReLU-14                  [-1, 128]               0
           Linear-15                   [-1, 64]           8,256
        LeakyReLU-16                   [-1, 64]               0
           Linear-17                   [-1, 32]           2,080
        LeakyReLU-18                   [-1, 32]               0
           Linear-19                   [-1, 12]             396
           Linear-20                    [-1, 5]             165
           Linear-21                    [-1, 2]              66
================================================================
Total params: 431,699
Trainable params: 431,699
Non-trainable params: 0
----------------------------------------------------------------
Input size (MB): 0.01
Forward/backward pass size (MB): 1.57
Params size (MB): 1.65
Estimated Total Size (MB): 3.23
----------------------------------------------------------------
```

Figure 4.14: The model summary of Multi Label CNN

We have used pyTorch Argmax, which is an operation that finds the argument that gives the maximum value from our targeted function. Argmax is used for finding the class with the most multitude number's predicted probability. In our case, the largest probability of age, gender and ethnicity classes has been defined and trained with keeping the number epoch = 100 unlike previous models and the model is defined to stop training when the accuracy is at 0.93 and loss = 0.2. The reason multi label CNN model's callback has not been defined like the sequential models is because it is a classification problem. We found that in (Figure 4.15) a multi label CNN model can keep the learning rate curve downwards by following each epoch and achieve the set accuracy within 78 epochs.

```
Epoch : 76/100,   loss : 0.20286,    acc : 0.93090
100% 186/186 [00:13<00:00, 14.10it/s]
Epoch : 77/100,   loss : 0.27188,    acc : 0.90579
100% 186/186 [00:13<00:00, 14.66it/s]
Epoch : 78/100,   loss : 0.19389,    acc : 0.93548
early stop
```

Figure 4.15: The training cell of multi label CNN

4.6 Resnet-50 Model Implementation

Residual Neural Network (Resnet-50) is a convolutional architecture model which is used to optimize prediction and training results for better accuracy results for image classification and prediction. In our model we have proposed an age estimation, gender and ethnicity prediction in convolution neural network (CNN) using the deep residual network (Resnet-50) model. The Figure:4.16 [7] shows the block diagram of Residual Network Network (Resnet).

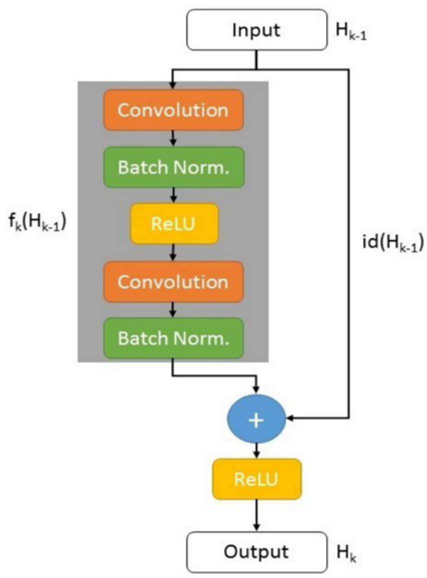

Figure 4.16: Block diagram of residual block in ResNet

Resnet-50 model is used for image classifications and prediction from our UTKFace dataset and helps to get attributes such as age, gender and ethnicity for better accuracy results. UTKFace dataset is used to evaluate the performance of residual networks for age estimation of the range 1-100 years old people. The model is trained on our dataset for ethnicity and gender on this range of people to get accurate prediction results.We select our dataset and divide it into 90% training data and 10% test data as the model itself requires more params than any other models. The dataset is divided into parts and result in easier to train our model. The data set used provides a benchmark for evaluating the optimum performance of the age estimation algorithm on ethnicity and gender prediction. The attributes like gender,age and ethnicity are trained together and we have achieved the results on those attributes and have calculated both the accuracy and loss for those attributes. The epoch is set to 25 for the training and we have achieved our desirable result at epoch 20. From the Figure 4.17, a training result with iteration considering the number of epochs we train for better training accuracy and decrease training loss. The interaction is the number of training epochs which result in better training accuracy and decreasing the loss from scale 4.0 to 0.0. There is a sharp rise in accuracy as the learning rate is increasing. The graph is being trained for all the soft biometric attribute's predictions. From epoch 112 out of 200 we get the training loss 0.23, the accuracy 80.25% from the training of our dataset on age prediction.

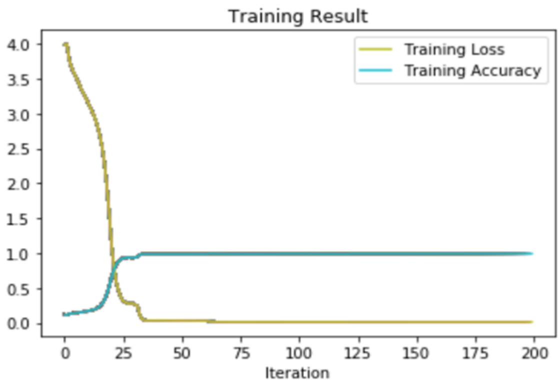

Figure 4.17: Training result for prediction with ResNet-50 (Loss vs Iteration)

From the Figure: 4.18 shows a training result with iteration considering the number of epochs we train for better training accuracy and decrease Training loss. The interaction is the number of training epochs which result in better training accuracy and decreasing the loss from 120 to 0. After training the result we have found out that better results estimation compared to linear regression. Overall, our graph shows that the residual network classification is appropriate to solve the problem related to gender prediction and leads to better accuracy after further training. From 762 out of 1000 epochs we get the training loss 0.22, the accuracy 92% from the training of our dataset on gender prediction.

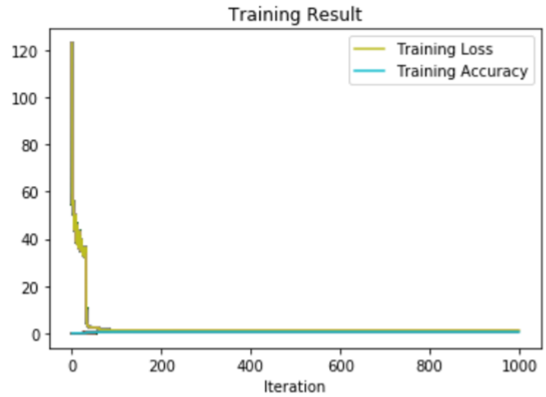

Figure 4.18: Training result for prediction with ResNet-50 (Loss vs Iteration)

4.7 Result Analysis:

By comparing our dataset of age, gender, ethnicity for different parameters and models via modifying convolutional layers we have achieved different levels of accuracy (Table 4.1) after running and training our dataset.
After comparing our dataset with different models we can notice that our CNN has better accuracy and training loss than the multi headed CNN model. Whereas, the results of the both sequential models can not compete with the results of the DenseNet-169 model. The multi label CNN is a different case as the model can train all the labels as a whole and can generate an average training loss and accuracy of all the classes at the same time. The multi label CNN classification model has achieved accuracy of 93% by keeping a training loss rate at 0.2 for all the soft biometric traits. For multi label CNN, we do not know the accuracy of each class (age, gender, ethnicity) as the classes were trained altogether but it can be trained for a long time to achieve the accuracy we want. On the ResNet-50 it has given us a result that is taking more params than the other models but taking a longer time to train than the other models. The ResNet-50 can be trained with more epochs for a longer time and it can achieve a higher accuracy. In our model, we have gained 92% accuracy for 762 out of 1000 epochs. After exploring our CNN models we can see DenseNet-169 can achieve greater accuracy in all the layers even after passing a very low amount of params. The ResNet-50 promises to show higher accuracy but the training loss and time is much higher than the DenseNet-169 model. The ResNet-50 model also shows a higher mean absolute error than any other models for age detection. Therefore, we can conclude that CNN model is giving us a better result after the DenseNet-169 model.

Implemented Models	Training Loss	Accuracy(%)	Mean Absolute Error (%)
CNN (Keras Sequential)	Age : 0.467 Ethnicity:0.3 Gender:0.14	Age: N/A Ethnicity:89.1% Gender:93%	5.47
Multi-Headed CNN (Keras Sequential)	Age:0.709 Ethnicity:0.72 Gender:0.34	Age: N/A Ethnicity:72.9% Gender:84.6%	6.1
DenseNet-169	Age:0.10 Ethnicity:0.06 Gender:0.02	Age: 96.16% Ethnicity: 97.74% Gender: 99.2%	4.92
Multi Label CNN	0.1	93%	7.3
ResNet-50	0.23	80.25%	8.74

Table 4.1: Model Comparison

Chapter 5

Conclusion

The soft biometrics information maintained the stability of the face recognition performance while also boosting prediction performance utilizing the face detection dataset, according to the findings. While biometric data is typically only used to identify a person, developments in machine learning have made it feasible to extract additional information from biometric data, including age, gender and ethnicity prediction. Thinking outside of the traditional recognition system and indulging in the soft biometrics revolution can bring more efficiency in our IoT based security. During the whole research process, we tried to mitigate the training loss and increase the accuracy so that it could be more reliable than the traditional recognition system. In our research, we have used various models to check our dataset and get better accuracy and more precise results. Training the dataset over different models has helped us to get more accurate predictions for age, ethnicity and gender as we approached forward towards coming to a conclusion. Throughout the research our main objective was to identify the optimum prediction model for the dataset and to enhance new research factors in this field to create a way for other researchers to progress with the results[2]. It was quite a challenge for us to gather a vast amount of data to train our model. But, our findings in this research has given us satisfactory outcome in terms human prediction. We believe, if we work on these models with more efficient dataset then our model can be more efficient and accuracy will reach close to the traditional systems. There are more physical soft biometric traits like hair color, walking patterns and facial unique attributes which can be trained to make a larger scale of training model. Nevertheless, there are more CNN models being introduced with a higher accuracy for classification problems. We believe our research, for now, is utilizing the current resources and models utilizing our work which will help other researchers to enhance their research. In our future work, we can add more classification and prediction models to improve our result as this will reduces the search space for biometric identification, gender or age-specific prediction and enhancing the biometric system's recognition accuracy, and so on. We can also add the scheduled learning rate to develop the training and testing performance of our dataset. We can also bring more variation among the attributes and thus train the model for larger dataset. With these strategies, the estimating error will be minimum and will help to get the best out of these models and get more accuracy.

Bibliography

[1] G. Tutz, "Sequential models in categorical regression," *Computational Statistics & Data Analysis*, vol. 11, no. 3, pp. 275–295, 1991.

[2] A. Dantcheva, C. Velardo, A. D'angelo, and J.-L. Dugelay, "Bag of soft biometrics for person identification," *Multimedia Tools and Applications*, vol. 51, no. 2, pp. 739–777, 2011.

[3] N. Kumar, A. Berg, P. N. Belhumeur, and S. Nayar, "Describable visual attributes for face verification and image search," *IEEE Transactions on Pattern Analysis and Machine Intelligence*, vol. 33, no. 10, pp. 1962–1977, 2011.

[4] H. H. K. Tin, "Gender and age estimation based on facial images," *International Journal: ACTA TECHNICA NAPOCENSIS Electronics and Telecommunications*, vol. 52, no. 3, 2011.

[5] ——, "Subjective age prediction of face images using pca," *International Journal of Information and Electronics Engineering*, vol. 2, no. 3, p. 296, 2012.

[6] P. Tome, J. Fierrez, R. Vera-Rodriguez, and M. S. Nixon, "Soft biometrics and their application in person recognition at a distance," *IEEE Transactions on information forensics and security*, vol. 9, no. 3, pp. 464–475, 2014.

[7] S. Kumra and C. Kanan, "Robotic grasp detection using deep convolutional neural networks," in *2017 IEEE/RSJ International Conference on Intelligent Robots and Systems (IROS)*, IEEE, 2017, pp. 769–776.

[8] V. Mirjalili and A. Ross, "Soft biometric privacy: Retaining biometric utility of face images while perturbing gender," in *2017 IEEE International joint conference on biometrics (IJCB)*, IEEE, 2017, pp. 564–573.

[9] K. Zhang, M. Sun, T. X. Han, X. Yuan, L. Guo, and T. Liu, "Residual networks of residual networks: Multilevel residual networks," *IEEE Transactions on Circuits and Systems for Video Technology*, vol. 28, no. 6, pp. 1303–1314, 2017.

[10] A. González-Briones, G. Villarrubia, J. F. De Paz, and J. M. Corchado, "A multi-agent system for the classification of gender and age from images," *Computer Vision and Image Understanding*, vol. 172, pp. 98–106, 2018.

[11] R. Yamashita, M. Nishio, R. K. G. Do, and K. Togashi, "Convolutional neural networks: An overview and application in radiology," *Insights into imaging*, vol. 9, no. 4, pp. 611–629, 2018.

[12] B. De Carolis, N. Macchiarulo, and G. Palestra, "Soft biometrics for social adaptive robots," in *International Conference on Industrial, Engineering and Other Applications of Applied Intelligent Systems*, Springer, 2019, pp. 687–699.

[13] A. Fariza, A. Z. Arifin, *et al.*, "Age estimation system using deep residual network classification method," in *2019 International Electronics Symposium (IES)*, IEEE, 2019, pp. 607–611.

[14] P. Feng, Y. Lin, J. Guan, Y. Dong, G. He, Z. Xia, and H. Shi, "Embranchment cnn based local climate zone classification using sar and multispectral remote sensing data," in *IGARSS 2019-2019 IEEE International Geoscience and Remote Sensing Symposium*, IEEE, 2019, pp. 6344–6347.

[15] A. V. Savchenko, "Efficient facial representations for age, gender and identity recognition in organizing photo albums using multi-output convnet," *PeerJ Computer Science*, vol. 5, e197, 2019.

[16] S. Lim, "Estimation of gender and age using cnn-based face recognition algorithm," *International journal of advanced smart convergence*, vol. 9, no. 2, pp. 203–211, 2020.

[17] C. H. Nga, K.-T. Nguyen, N. C. Tran, and J.-C. Wang, "Transfer learning for gender and age prediction," in *2020 IEEE International Conference on Consumer Electronics-Taiwan (ICCE-Taiwan)*, IEEE, 2020, pp. 1–2.

[18] P. Choudhury, R. T. Allen, and M. G. Endres, "Machine learning for pattern discovery in management research," *Strategic Management Journal*, vol. 42, no. 1, pp. 30–57, 2021.

[19] Y. He, B. Nie, J. Zhang, P. M. Kumar, and B. Muthu, "Fault detection and diagnosis of cyber-physical system using the computer vision and image processing," *Wireless Personal Communications*, pp. 1–20, 2021.

[20] A. M. Saleh and T. Hamoud, "Analysis and best parameters selection for person recognition based on gait model using cnn algorithm and image augmentation," *Journal of Big Data*, vol. 8, no. 1, pp. 1–20, 2021.