Human Face Analysis using Transfer Learning Approach

MSc Research Project
Data Analytics

Gaurav Singh
Student ID: X21136921

School of Computing
National College of Ireland

Supervisor: Prof. Jorge Basilio

National College of Ireland
Project Submission Sheet
School of Computing

Student Name:	Gaurav Singh
Student ID:	X21136921
Programme:	Data Analytics
Year:	2022
Module:	MSc Research Project
Supervisor:	Prof. Jorge Basilio
Submission Due Date:	15/12/2022
Project Title:	Human Face Analysis using Transfer Learning Approach
Word Count:	7956
Page Count:	26

I hereby certify that the information contained in this (my submission) is information pertaining to research I conducted for this project. All information other than my own contribution will be fully referenced and listed in the relevant bibliography section at the rear of the project.

ALL internet material must be referenced in the bibliography section. Students are required to use the Referencing Standard specified in the report template. To use other author's written or electronic work is illegal (plagiarism) and may result in disciplinary action.

Signature:	
Date:	31st January 2023

PLEASE READ THE FOLLOWING INSTRUCTIONS AND CHECKLIST:

Attach a completed copy of this sheet to each project (including multiple copies).	☐
Attach a Moodle submission receipt of the online project submission, to each project (including multiple copies).	☐
You must ensure that you retain a HARD COPY of the project, both for your own reference and in case a project is lost or mislaid. It is not sufficient to keep a copy on computer.	☐

Assignments that are submitted to the Programme Coordinator office must be placed into the assignment box located outside the office.

Office Use Only	
Signature:	
Date:	
Penalty Applied (if applicable):	

Human Face Analysis using Transfer Learning Approach

Gaurav Singh
X21136921

Abstract

The person's face is among the most significant body parts for transmitting key features since it is so prominently shown. People of varying ages, sexes, and races each have a unique set of facial characteristics. Numerous studies have indicated that smokers have a significantly increased risk of experiencing substantial alterations in the middle and lower thirds of their faces. This research investigates the use of five distinct pre-trained - DCNN(deep convolutional neural network) learning models for identifying human face traits by making use of three different datasets and doing a comparative analysis. It is possible for governments, schools, insurance companies, and universities to leverage automation of large-scale BMI imputations, as well as age, gender, ethnicity, and smoking status, in order to collect data on a range of social concerns and make more informed decisions. Multiple studies have been done, with each one concentrating on a limited number of attributes and making use of a more condensed data set. In this particular investigation, pre-defined transfer learning networks such as EfficientNetB7, DenseNet201, InceptionResNetV2, VGG16, and ResNet50V2 were chosen over conventional-CNN (convolutional neural networks) in order to compare the results of the pre-trained models with one another. This comparison was carried out in order to determine which of the models performed the best. This was accomplished by employing a two-way comparison strategy. This study aims to quantify essential human face attributes and evaluate the findings based on parameters such as precision, accuracy, recall, RMSE, and MAE using a variety of methods that are considered to be state-of-the-art in the field.

1 Introduction

1.1 Background & Motivation

The human face is a reliable indicator of one's overall state of health. It is possible to acquire certain hidden traits from the human face, which may subsequently be utilized to create accurate forecasts of the individual's health. In addition to a person's ethnicity, age, and gender(male or female), a lot of additional factors have a significant part in determining their face health. These traits include: The individual's smoking history and their projected body mass index are also examples of these characteristics which can be useful to avoid hospital disparities in Mays et al. (2003). The human face is complicated in part because it contains so many varied traits, each of which contributes to the overall complexity of the face. The dynamic skin tone, the variety in the shapes of the nose, the variety in the structure of the mouth and lips, and the variety in the structure of the cheeks, which can be oval, round, square, diamond, heart, pear, or oblong are some of the features that contribute to the complexity of the human face. In addition to these features, there are a great number of other aspects that play a role in the creation of this complexity.

Gender of a person(male or female), age, and ethnicity or race estimation is one of the well-known deep learning applications in today's data-driven society in Srinivas et al. (2017). This application came about as a consequence of the rising rate at which photos are posted to the internet, which led to the widespread use of deep learning. When it comes to establishing ages, humans continue to have a lot of trouble, despite the fact that they are naturally quite good at recognizing one another, making rapid judgments about one another's country, detecting gender in one another, and doing so quickly. One must realize that the aging of the face is not only determined by inherited factors, but is also influenced by environmental factors, expression, lifestyle choices, and even a person's smoking habits, regardless of whether or not the individual is a chain smoker. Author Mays et al. (2003) came to the conclusion that ethnicity categorization may be used to categorise individuals into broad and distinctive populations or groups based on heritable and phenotypic features, geographical origins, personal appearance, ethnicity, and social status. Similarly, Cigarette smoking poses an alarming threat to the general public's health, since it may often lead to lung cancer in addition to other dangerous conditions. A nicotine test is given to people in order to establish whether or not they are smokers. The measurement of an individual's nicotine consumption during this test is a labor-intensive operation that often takes place in medical facilities and may take quite some time. The use of tobacco results in the deaths of around 8-10 million people throughout the world every year, which in turn has an effect on the people who are in their local settings. People living in contemporary times are placing an ever-increasing emphasis on the importance of their health. It seems that being either overly heavy or too thin is a problem that affects not just one's looks but also their health. Several different indicators that may assist us in measuring and monitoring our bodies have been presented in order to help us become more aware of the danger of being unwell in a more timely manner. As described in Klontz and Jain (2013), the Body Mass Index is fast becoming one of the most frequently used health indicators (BMI) that may be utilized in police files, and crime reports, unlike conventional biometrics which may be inadequate since we can get height and weight characteristics once we know a person's BMI. The Body-Mass-Index (BMI) of a person is calculated by dividing their body weight by the square of their height. People are able to determine if they are underweight, proper-weight, overweight, or obese based on their BMI score, as well as the degree to which their weight deviates from what would be considered optimal for their height. As a result of its ease of use, the BMI has been extensively used for many decades; yet, it is still rather uncomfortable to measure and compute, particularly for those who are unable to walk or stand, as it needs individuals to consciously measure their weights and heights.

One of the aspects that go into image processing is examining a variety of different human faces to establish a person's age and then using that information to generate predictions about that person's age. This is one of the components. This is another area of study that sees a lot of activity. In the past few decades, face analysis and facial variation analysis have attracted a lot of interest in scientific and industrial applications as a result of the potential need for applications such as intelligence, privacy, security systems, legislation, and monitoring. This interest has been driven by the potential need for these applications. Transfer learning is a more modern kind of deep learning that was developed more recently. This type of deep learning does not need the same distribution of training and test data. Although transfer learning is able to address the most significant flaw in the conventional method of machine learning for solving prediction problems, it does so at the expense of generating a number of new and challenging challenges. These challenges include limited interchangeability and even negative transfer. Author Gupta et al. (2022) arrived at the logical conclusion that since since transfer learning algorithms were first developed, they have been put to the test in a wide range of research endeavours. Evidently, tremendous progress has been made in transfer learning across a wide range of different activities and circumstances. Author Padmavathi et al. (2022) laid forth several advantages of employing Transfer Learning pre-trained models for the benefit of identifying brain cancers using MRI images, to separate the segmented images and images using pre-trained models. Transfer Learning is becoming more and more popular.

Predicting human health from a person's face using transfer learning pre-trained models in educational institutions like schools, colleges, and universities, security surveillance systems, and the insurance industry may help create a more user-friendly cloud-based healthcare system that can be used by various healthcare institutions. This would help create a user-friendly, multi-institutional healthcare system. Private enterprises benefit most. These technologies will benefit schools and colleges by automating the computation of human traits, as well as replacing the old way of computing age, gender, and ethnicity. Such systems will replace generations-old methods.

The ability to determine whether or not a person is a smoker is important in many industries, including those linked to insurance, the medical profession, and military medicine. In addition, it is helpful to have knowledge about a person's body-mass-index (BMI), race or ethnicity, gender, and age in a number of contexts, including the check-in process at an airport, insurance companies, the medical sector, and even security systems. A proper health plan may be offered to an individual by insurance companies in the event of any shortfall using the individual's age and body mass index in order to combat any ailment. Additionally, the utilization of facial photographs of human faces is indeed helpful in a variety of domains, including criminal discovery on the basis of ethnic background Belcar et al. (2022), legislation and jury instances, Cctv footage, promotion, and social network identification. These are just a few of the situations in which it could be beneficial to make use of pictures of people's faces. In circumstances when humans are involved in the process, such as when it is impossible to identify a person's age, gender, ethnicity, or body-to-mass ratio in a short amount of time, an automated method is required. This may help in figuring out certain circumstances more quickly and accurately, which would be beneficial.

1.2 Research Goal

A primary objective of my study is to develop automated state-of-the-art systems that can successfully replace the manual calculations that have been done in the past to determine an individual's age, gender(male or female), ethnicity or race, body-mass-index, and smoking status(smoker or non-smoker). Using this study, we are giving more weight to previous research that has been done on extracting human face features using deep learning with the assistance of transfer learning methods that are trained on pre-trained architectures such as EfficientNet, DenseNet, InceptionResNetV2, VGG16, and ResNet50V2. In this study, human facial characteristics were extracted using deep learning and transfer learning techniques that had already been developed on pre-trained architectures.

Through the use of transfer learning networks like EfficientNet, DenseNet, Inception-ResNetV2, VGG16, and ResNet50V2, this goal is to develop a human face analyzer system that can analyse faces more quickly than existing legacy image processing networks.

1.3 Research Objective

To use transfer learning frameworks to extract valuable human face information, compare the performance of the five distinct transfer learning pre-trained models, and choose the best one.

1.4 Report Structure

Sections and subsections separate the report into several parts. Following part 1, section 2 focuses on relevant research in the subject of transfer learning, the benefits of image processing for obtaining meaningful human knowledge, and the ways in which my use case is best suited for transfer learning. I'll talk about the methodology I used for my study, the datasets I utilised, the data cleaning and augmentation methods I employed, and how I came to comprehend the datasets in the Methodology part that follows. The Design Specification follows, in which we describe the design architecture and pre-established transfer models for the project. Part 5 is the next section, and it discusses how the experiments were implemented or used, as well as which technology was most effective. The assessment and experiments were included in section 6 of the paper. In section 7, we have provided a comprehensive conclusion as well as goals for future work.

2 Related Work

In light of the increasing amount of image data that is made available on the internet and the growing popularity of image processing that makes use of deep learning, numerous authors and researchers have published a large number of articles and papers in an effort to develop a method for determining a person's age, gender, ethnicity, and body mass index. This is being done in response to the expanding number of image data that is being made available on the internet. In addition, determining whether or not a person is a smoker, determining the race of a person, predicting the bmi of a person is a topic that has not yet been fully explored and has seen very little research done on it. Although this can be accomplished with the help of image processing and photographs of human faces, the datasets that can be utilized for this purpose are quite limited.

In the parts that follow, I will examine transfer learning and its benefits, technologies, and implementations, as well as the evolution of new methods and techniques throughout time.

2.1 Deep Neural Network in Image Processing

A human face may convey a tremendous amount of information about the individual to whom it relates. Recent research has shown a strong relationship between a person's body-mass-index(BMI) and their facial features. The varied features of the human face may be significantly influenced by a person's ethnicity or race, gender, place of birth, age of a person, smoking habits, and many other factors.

Additionally, there are a variety of other factors that contribute to the formation of the face. These factors include: People who have slim faces have a greater likelihood of having a lower body mass index, and the converse is also true. Obese people often exhibit broader facial features, especially in the centre and bottom thirds of the face. In the vicinity of the ears, this is very obvious. It will not be feasible for a person to determine their body mass index if they do not have access to a measuring tape and a scale (BMI). In recent years, there have been several advancements achieved in the area of deep learning. As a result of these advancements, models are now able to extract essential information from photographs. By applying these methods and analyzing a person's facial characteristics, we are able to calculate a person's body mass index (BMI). As a result, the purpose of this study was to develop a technique for assessing body mass index (BMI) only based on face characteristics using human individuals. It is probable that using this strategy will make it simpler for health insurance companies to maintain track of the medical histories of their respective clientele. In addition, the government may maintain a record of the medical history of a particular region in order to establish rules based on the outcomes of this research.

A statistical method was suggested in Syrowatka et al. (2017) in order to investigate the connection between body mass index (BMI) and waist-to-hip ratio (WHR), as well as face shape and facial texture. With the use of the WindowsOS application named as TPSDig, the authors were able to identify 119 anatomical landmarks as well as semilandmarks. In order to determine the precise locations of the semilandmarks, they used a sliding landmark technique. In their investigation, they used 49 standardized pictures of women whose BMI fell anywhere between the ranges of [17.0 and 35.4] and whose WHR fell anywhere between [0.66 and 0.82]. The RGB (red, green, and blue) values of the standard photos indicate the texture of the face, while the Multiresolution shape coordinates show the size and form of the face. In order to investigate the pertinent links, multivariate linear regression was used. The BMI was more predictable than the WHR based on face characteristics, with 25% of the variance explained by facial shape and 3–10% explained by facial textures.

Savchenko (2019) employed photo-algorithms and video archives in a later study to automatically extract people and the characteristics that set them apart (such gender and birth year). In this article, the author Savchenko (2019) shows that, despite the recommended approach's implementation being significantly more computationally affordable, the effectiveness of face segmentation for the created network is equivalent to the state-of-the-art results produced by deep neural networks.

2.2 Transfer Learning in Image Processing

Learning that is transferable not only boosts performance but also reduces the amount of time and effort required. Transfer learning is very useful for analysing unlabeled datasets since it makes model creation go more quickly and produces more accurate results. A study where a bespoke end-to-end CNN network was suggested by Siddiqui et al. (2020) as a method for predicting BMI was conducted. TFurthermore, the author retrieved characteristics from face pictures using pre-trained CNN models such as Densenet, MobileNet, ResNet, VGG19, and LightCNN. These characteristics were then sent to SVR and RR for final forecasting. They obtained a Mean Absolute Error 5 (MAE) in the range of

[1.04, 6.48] using the VisualBmi, VIP attribute, and Bollywood Datasets. When used in conjunction with Ridge Regression, the DenseNet and ResNet models provided superior results. The performance of the final CNN model was equal to or slightly below that of the pre-trained models.

In addition, in Gao et al. (2020), a study was carried out in which a model for the identification of Chinese facial ethnicity (CFER) was created. This model was based on transfer learning from deep convolutional networks. In this study, we focused on five separate Chinese ethnic groups in order to develop a face dataset that contained information relevant to ethnicity. This was done in order to better understand facial similarities across the groups. Next, it utilized CFER in order to identify characteristics that are exclusive to Chinese ethnicity, and it utilized the 10-fold cross-validation method in order to primarily estimate the accuracy rate of the model. Both of these methods were utilized in order to determine the accuracy rate of the model. The model has an average recognition rate of 80.5%, and it also has a reasonable performance in terms of its ability to generalize. These two indicators are top-notch in their respective categories. It has been shown that the deep learning algorithm is capable of accurately determining the ethnicity of a person based on their facial features.

According to the findings of a different study by Narang and Bourlai (2016), a CNN can be used to categorise the data into categories of ethnicity and gender class when employing both restricted and unconstrained face datasets. This was shown by using both bound and unbound face datasets throughout the demonstration. Second, based on the findings of the face recognition test, we have reached the conclusion that the application of soft biometric traits, such as ethnicity and gender, has the potential to improve the rank-1 identification rate of our FR system. This was determined after analysing the results of the face recognition test. When assessed at about 60 metres, this results in an improvement of 45percent for the Caucasian class and 26percent for the Asian class in the Trans scenario. In the not-too-distant future, we anticipate making the categorization findings even more accurate. As a result, we want to include more datasets in order to train our deep learning model and explore various CNN architectural configurations.

In addition, Pundhir et al. (2022) employed a technique called deep learning to extract relevant parts from an image and determine the smoking behaviour of the persons in the photograph. The author of this article used the pre-defined models Resnet18, Resnet34, and YOLOv3 to categorise the photographs depending on whether or not the subjects were smokers. In this particular instance, the accuracy of the author was 96.74 percent. The dataset that was used in this study is the same one that I used in my own research; thus, it was utilised in this study as well. It will be helpful for my inquiry since its architecture incorporates an aggregate of the ResNet18 and ResNet34 models together with the pre-trained weights for each of those models.

Alzheimer's disease is a rare ailment characterized by the gradual decline of cognitive ability. Alzheimer's disease therapy is too expensive and uncomfortable for the afflicted individual. To circumvent this, Sethi et al. (2022) established a framework for the categorization of Alzheimer's disease utilising the EfficientNet model architecture and the financial instability of the illness's treatment to get past this problem. Here, the architectural paradigm of EfficientNet and transfer learning are integrated with the preset concepts from the ImageNet Dataset. The performance of the model was evaluated using the overall accuracy and the area under the ROC curve (AUC). Compared to conventional deep learning convolutional neural networks, the achieved accuracy was closer to

92 percent and AUC values were comparable to 83 percent.

2.3 Literature Summary

The scant attention paid to the issue is further evidenced by the scanty study that has been done using a combination of evaluating human facial traits. There have been many research in image processing employing cutting-edge techniques, yet the results are only generally adequate. Therefore, this study seeks to use numerous pre-defined state-of-the-art models that would fundamentally aid in measuring and forecasting various parameters for assessing human vitals and to attain greater performance that can be applied in a wide range of contexts.

3 Methodology

3.1 Introduction

Human faces are analysed using image processing and pre-trained transfer learning models EfficientNet, Densenet-201, VGG16, InceptionResNetV2, and ResNet50V2. EfficientNet, Densenet-201, VGG16, InceptionResNetV2, and ResNet50V2 will train each dataset. Knowledge Discovery in Databases (KDD) applies "high-level" data mining techniques. It emphasizes data discovery. This subject will interest AI, machine learning, pattern recognition, database administration, statistics, knowledge acquisition for expert systems, and data visualization researchers. Figure 1 shows the Deep Learning Pipeline research steps. All the steps involved in this research for Deep Learning Pipeline are shown in Figure 1.

3.2 Data Understanding

This section explains the procedure for acquiring the majority of the data. In this investigation, I have made use of three distinct datasets that are in no way connected to one another. I employed three different datasets in my research study: the UTKFace dataset Zhang and Qi (2017), the Mendeley Smoking, and non-smoking dataset Khan (2020), and the VIP attribute Dataset Dantcheva et al. (2018). In the VIP attribute dataset, some demo images are shown in figure 2. There are a total of 1026 pictures in this collection, with an equal number of men and women included (513 each). The frontal stance is the format used for each and every one of the photographs that have been presented. The VIP Attribute dataset is one of a kind since it includes photographs of famous people together with information on their height and weight that can be used to calculate their body mass index.

Following this, in Figure 3 images from the Mendeley dataset have been provided. The dataset includes a total of 2400 raw photos, 1200 of which are classified as belonging to the smoking (smokers) group, while the remaining 1200 images are classified as belonging to the non-smoking (non-smokers) category. Photos uploaded to the Mendeley dataset for smokers are organized into separate folders, one labeled" smoker" and the other "non-smoker," respectively.

The UTKFace dataset makes available some metadata information, including age, gender, and ethnicity of the different age groups of people. The UTKFace dataset is a large-scale face dataset that spans a significant number of years (ranging from 0 to 116

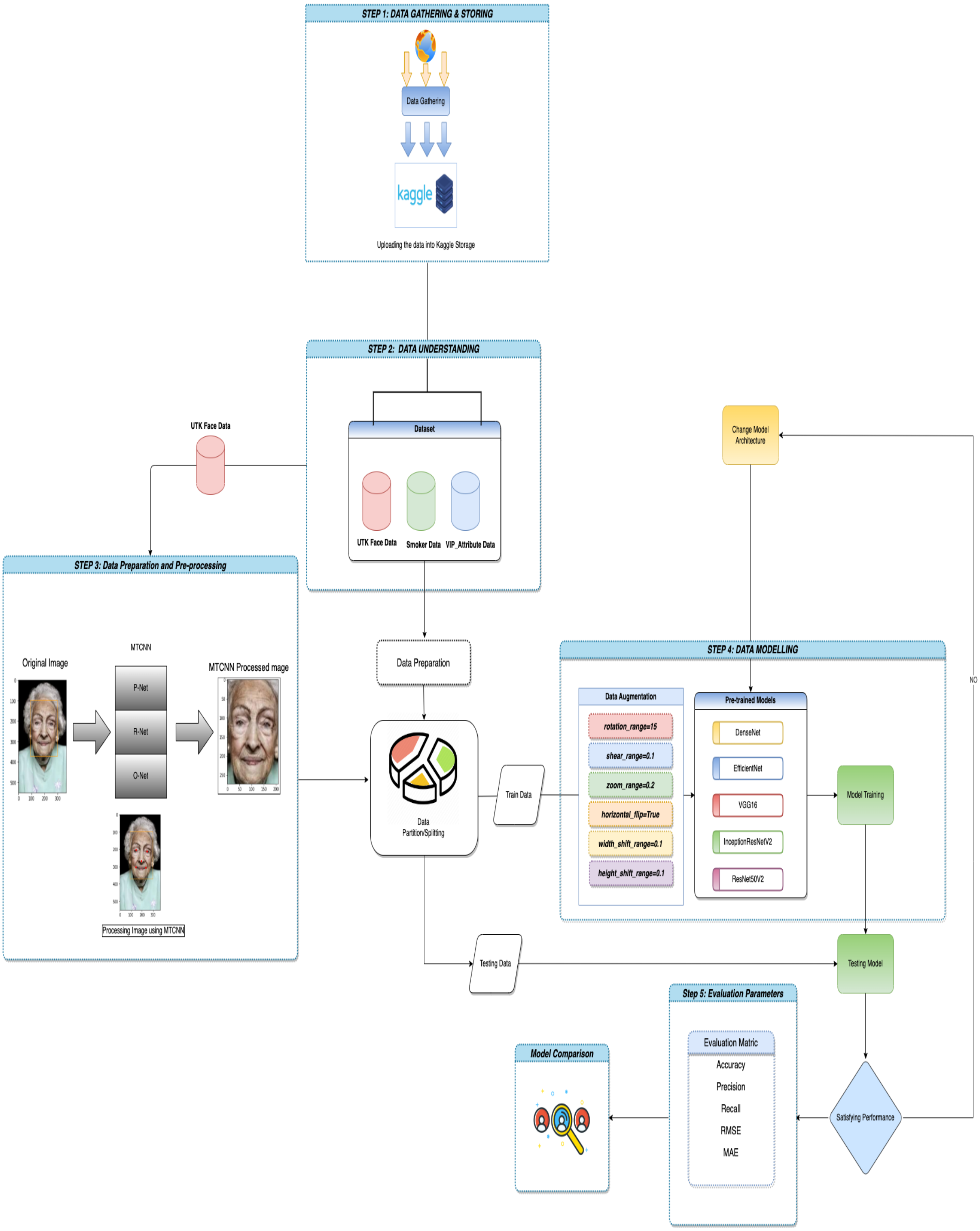

Figure 1: Design Framework

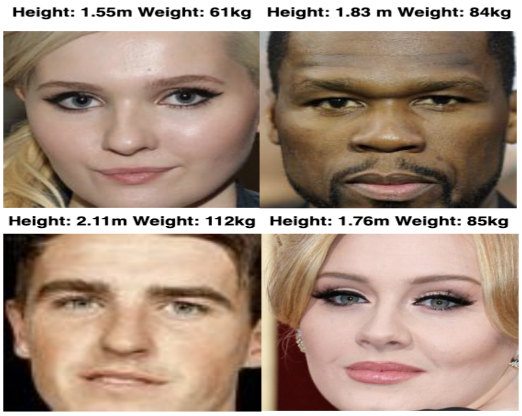

Figure 2: VIP Dataset Images

Figure 3: Smoking Dataset

years old). The collection contains approximately 20,000 face photos, each of which is annotated with information about the subject's age, gender, and ethnicity. The photographs include a wide range of variety in terms of position, facial expression, lighting, occlusion, resolution, and other characteristics. This dataset can potentially be used for a wide range of problems, including, but not limited to, face recognition, age estimation, age progression and regression, landmark localization, and others.

3.3 Data Preparation

3.3.1 Data Cleaning

Each of the three datasets was obtained by downloading it from its corresponding URL. Each piece of information associated with the datasets is examined to see whether or not it is missing any values and whether or not the directory contains any files.

3.3.2 Data Augmentation

As three datasets have been employed in this research, a common data preparation technique has to be used so that a generic approach is achieved. For this problem, the images were converted into an input shape of 224 * 224 * 3 with global average pooling since it calculates the mean for the parameters of height and width over all channels. The photos were then improved using 256 neurons in a separate dense layer with the ReLU activation feature. To avoid over-fitting issues, a dropout rate of 0.25 was used. Since I had more photographs, I used batch normalization as suggested in ŞEN and ÖZKURT (2020), which helps to speed up learning and utilize higher learning rates, which makes learning easier and cuts down on training time. Then, data augmentation techniques including rotation, shearing, and horizontal flip were utilized to produce more number pictures with various rotational patterns as shown in figure 4. Eventually, we utilise the Adam optimizer with pre-default settings, and Mean Square Error is employed as the loss function. As recommended by this study, this optimization technique has been created for deep learning methodologies.

As classifying the smoker is a classification problem. In order to evaluate the evaluation metrics, I have used accuracy, precision, and recall as evaluation parameters. The data available is very accurate as different images have been provided for testing, training and validating the smoker. After examining the dataset's balancing properties, it

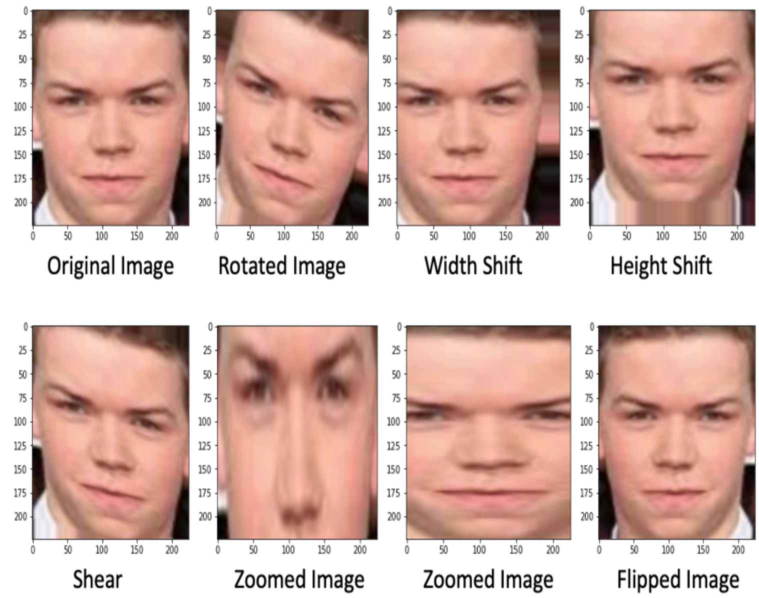

Figure 4: Data Augmentation

seems to have an equal amount of smokers images and non-smoker images. Initially, the input image size is set to 224,224,3 for the image input parameter. Image Enhancement techniques such as sigmoid activation functions were used to provide more details to the images.

Also, BatchNormalization is used as it increases the training process Ioffe and Szegedy (n.d.). In addition, the Adam optimizer approach is used ŞEN and ÖZKURT (2020) since it is advantageous when dealing with deep learning issues due to its quicker calculation time and fewer tuning parameters.

For face recognition, the MTCNN model was chosen after a thorough assessment of processing time and data volume. Through MTCNN analysis, facial characteristics including the position of the eyes, nose, and mouth are learned. The face of a person may be recognized by this model even if it is partly hidden, and it can do so from a variety of angles. After using MTCNN, the input pictures are initially processed as 224 × 224 x 3 input size. The current CNN layer is then supplemented with a functional dense layer. Before the dropout layer, a 2-D global average pooling layer is inserted because it accomplishes down-sampling by calculating the mean of the input image's height and width dimensions. On the previously processed photos, batch normalization is utilized to speed up processing and learning. In order to defeat the binary cross entropy loss function later on, we use the Adam optimizer.

3.3.3 Exploratory Data Analysis

In this part, we analyzed and performed exploratory data analysis and check for imbalance.

For the smoker dataset, the dataset is balanced as it contains an equal number of smoker and non-smoker images. The below figure shows the distribution of smokers/non-smokers. For the BMI calculation, the VIP attribute dataset is employed. This dataset is also balanced and does not contain any null values. The left graph shows an equal number of females and males. Due to this, the gender play will have a pivot role as the dataset

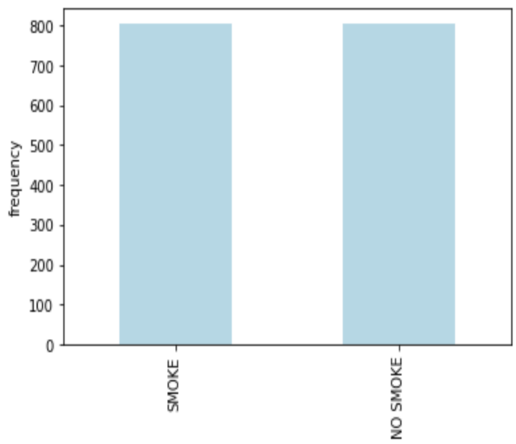

Figure 5: Smoking Dataset

is balanced based on gender which will be helpful in determining the weighted BMI. For the right graph, As seen, we divided the BMI category into 4 different categories namely: Healthy Weight, Overweight, Underweight, and Extreme Obesity in OuYang (2018).

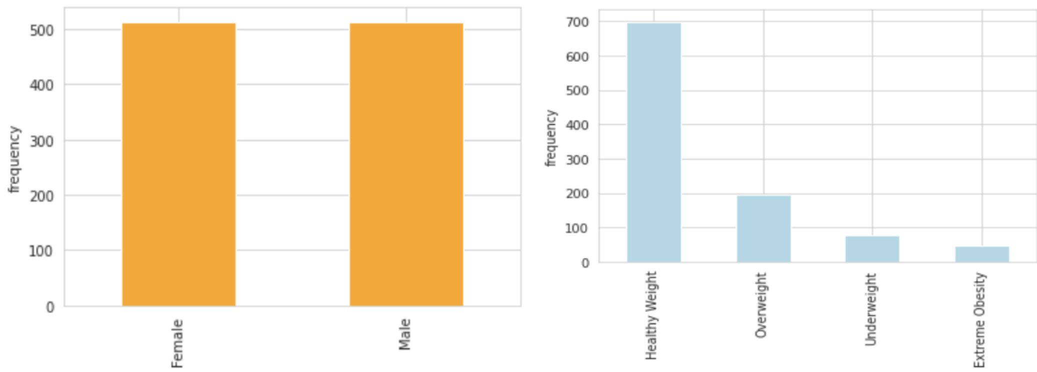

Figure 6: BMI Dataset

As seen from the density plot in figure 6, the number of people having a healthier BMI is more compared to other BMI categories. As, determining the BMI is a regression problem, having different BMI categories won't affect the model performance.For the third dataset which is UTKFace data, first we plotted the gender distribution plot which is shown in Figure 7. The data is almost balanced for gender distribution, whereas for ethnicity distribution is unbalanced.

4 Design Specification

4.1 Design Structure

The general modelling framework is shown in Figure 5. The approach begins with doing some preparatory work on each of the datasets, which are referred to in the following paragraphs as VIP Attribute, Smoker Data, and UTK Face Data, respectively. This is the first phase in the process. The process of preparing the data will start when this has been completed. Because the data had an excessive amount of noise, we used MTCNN

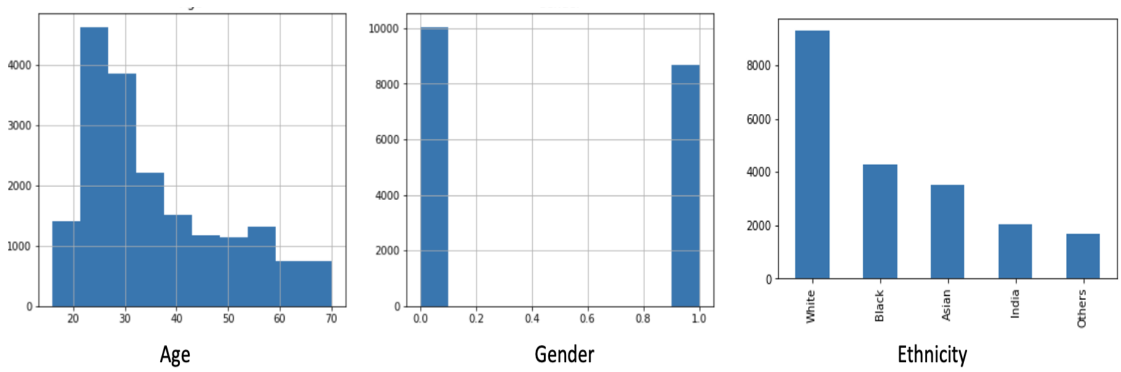

Figure 7: UTK Dataset: Age, Gender and Ethnicity

in Kangwanwatana and Sucontphunt (2022) to analyse it. This was done for the UTK dataset.

This research article applies image preprocessing to the supplied photos in order to improve the face verification rate. Some of the image preprocessing techniques that are used include employing MTCNN to filter out the face, face alignment, and brightness correction. As a result of the fact that MTCNN is a cascaded convolutional neural network, it is made up of a total of three distinct neural networks. The output of one network is copied and pasted into the input of the following network. The findings of the first network are used as input for the second network, which functions as a filter to get rid of the majority of erroneous detections and aggregate bounding boxes. The final network is used to enhance the predictions and adds in facial landmarks predictions. Before the data may be used as an input feature, it is first partitioned into testing data and then training data in the proportion 80:20. Then, the corresponding CNN Models were applied, and afterward.

Due to the fact that I worked with three distinct datasets, I decided to adopt a generic design framework for each of the datasets. After that, the data that will be used for training will be sent to five distinct pre-trained models, namely EfficientNet, DenseNet, InceptionResNetV2, VGG16, and ResNet50V2.In the future, the procedure will be repeated for each and every use case. Following that, the models are going to be evaluated for a number of different epoch values. In the end, different model results will be compared across CNN models such as EfficientNet, DenseNet, InceptionResNetV2, VGG16, and ResNet50V2 in order to determine how BMI, age, gender, and ethnicity can be identified, as well as how smokers may be categorised.

4.2 Model Overview

For BMI detection in Figure 8, then, the weights of these models have been kept (frozen Imagenet weights), and after global average pooling to mitigate the issue of overfitting, only dense layers with the 'ReLu' activation function have been added to the pre-trained base model layer. The last layer of categorization with activation is then applied based on the kind and nature of the classifications. Since the BMI identification problem is regression, the final output layer was added since values with a linear activation function are unbounded.

As illustrated in Figure 9, classifying a smoker is a binary classification issue, and the outcome will be either Smoker: 0 or No Smoker: 1. In addition, Global Average

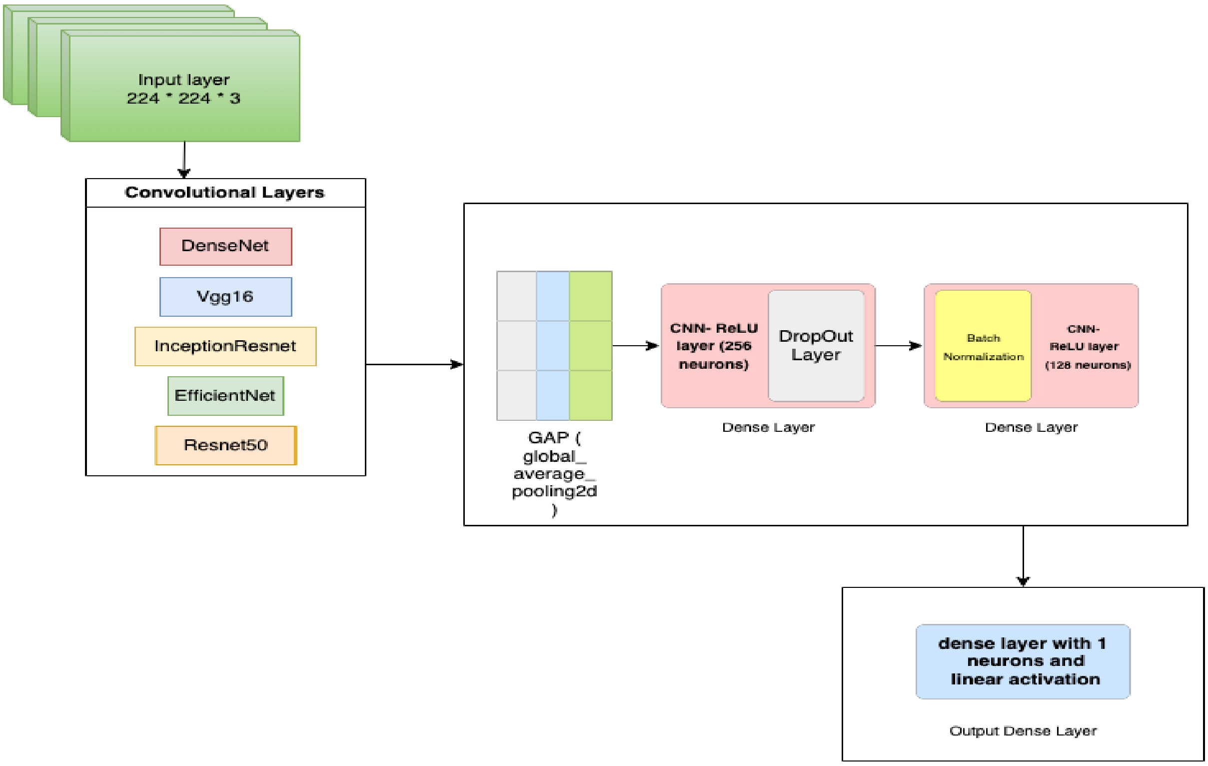

Figure 8: BMI Design Framework

Pooling (GAP) is utilized to address the overfitting problem. The pretrained base model layer only has dense layers with the 'ReLu' activation function, which calculates the spatial average of a feature map and can be used instead of fully connected layers in the network's penultimate layers, as pointed out in this research by Koffas et al. (2022). Due to the limited nature of the input values, the activation function is sigmoid because it is a classification issue and the output is combined with the sigmoid function.

The first step of the procedure involves face capture using MTCNN, which stands for Multi-Cascade Neural Network, and is helpful in capturing the source face from a list of faces that consists of background noise. Figure 10 depicts the overall architecture of MTCNN. The MTCNN Kangwanwatana and Sucontphunt (2022) is a Cascaded Network consisting of three different CNNs. The input for the first stage is an image pyramid, which is a collection of duplicates of the original picture with varying magnifications. This not only gives the model a broad variety of window sizes from which to select, but it also contributes to the model's ability to remain scale-invariant. The CNN Refine Network constitutes the second step (R-Net). It cuts down on the number of boxes even further and combines candidates who overlap using a method called non-maximum suppression (NMS). In the third step, the Output Network performs more of the tasks that are performed by R-Net. Additionally, it adds the 5-point landmark consisting of the eyes, nose, and mouth to the final bounding box that contains the recognized face.

This step is necessary for determining ethnicity, age, and gender using the final dataset which is UTK face data. Later on, for the regression problem of determining the age of a person, as a result of this research, the activation function used for the output function in the neural layer is linear. However, for the problem of determining the ethnicity of a person, the activation function used is sigmoid because it is a multi-class problem. Finally, for the problem of determining the gender of a person, the function was sigmoid,

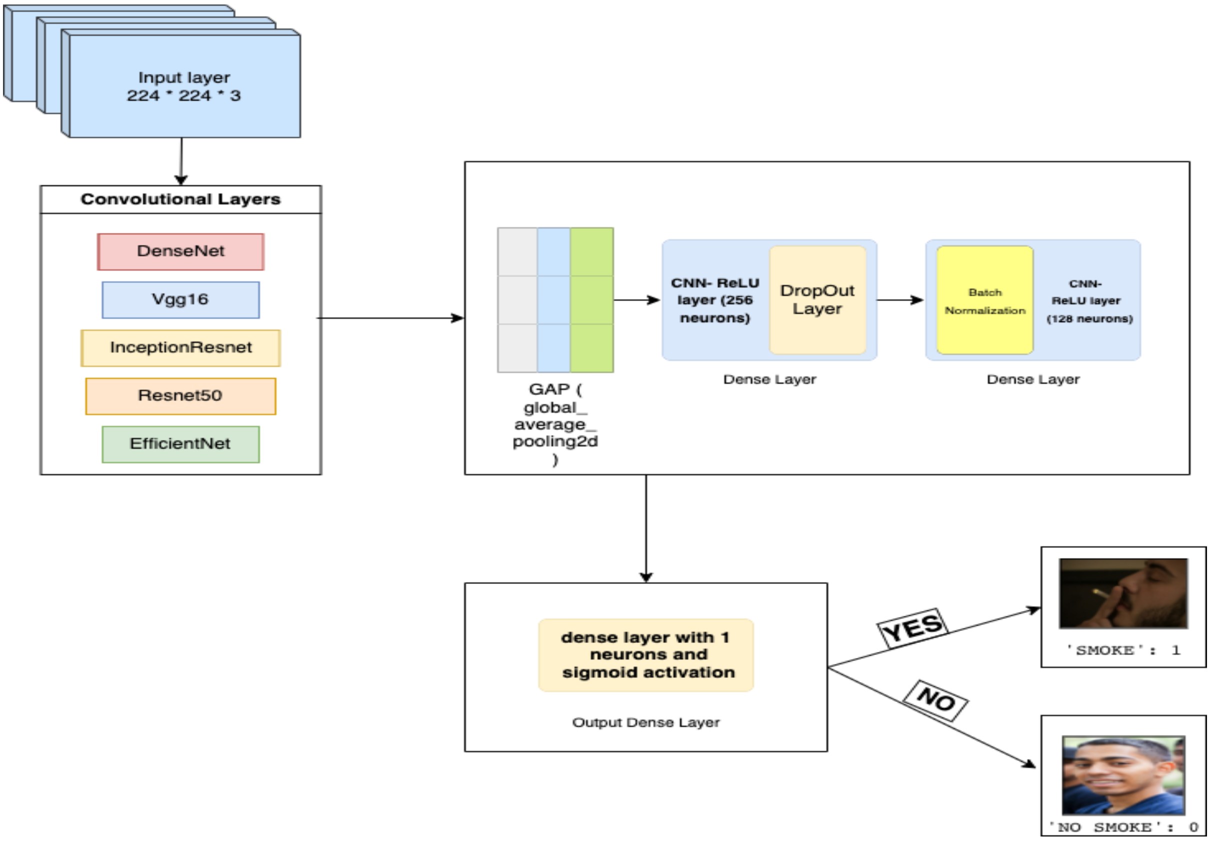

Figure 9: Smoke Design Framework

similar to what we had for the smoke classification problem.

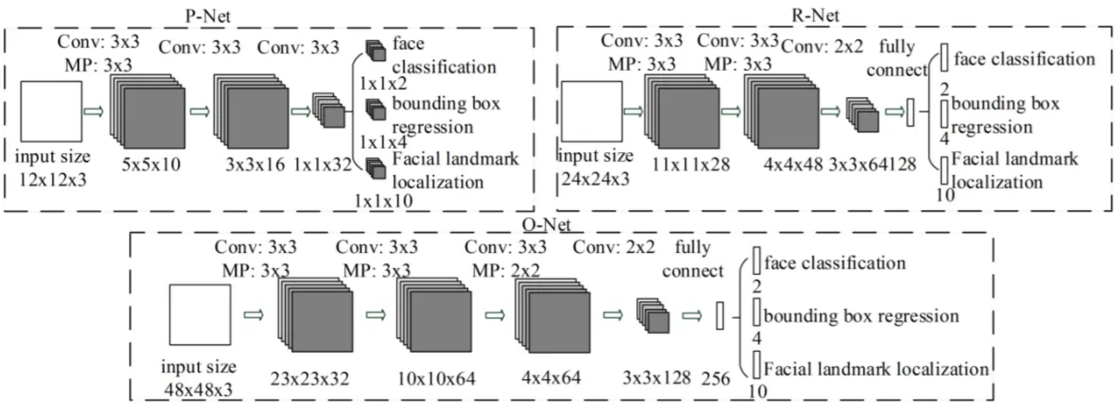

Figure 10: MTCNN Design Framework

4.3 Pre-trained Model Architecture

I employed transfer learning-based DNN architectures EfficientNet, Densenet-201, VGG16, InceptionResNetV2, and ResNet50V2 for this study. Pre-trained networks were trained on millions of photos using better hardware and procedures, yielding better results. After the pretrained base model layer was flattened, only dense layers with the "ReLu" activation function were inserted. In conclusion, a linear activation final classification layer is built according to the classifications in conjunction with classifications (age, gender, ethnicity, smoker or non-smoker, bmi).

4.3.1 EfficientNet

EfficientNet is pre-trained on the ImageNet dataset in Selim et al. (2022). EfficientNet's core component is the Mobile Inverted Bottleneck MBConv with squeeze and excitation optimization. EfficientNet family MBConv blocks vary. From EfficientNet B0 through B7, there are 8 models. These models vary in height, breadth, depth, and resolution. The number of estimated parameters does not rise greatly as the model's number increases, but its accuracy does. The best model, EfficientNetB7, is 6.1x quicker and 8.4x smaller than the best CNN. ImageNet accuracy is 84.4% for EfficientNetB7. EfficientNet B7's picture classification accuracy was so high that we decided to use it in the suggested model. As shown in Fig. 11, EfficientNet B7 may be divided into seven blocks depending on stride, filter size, and channel count.

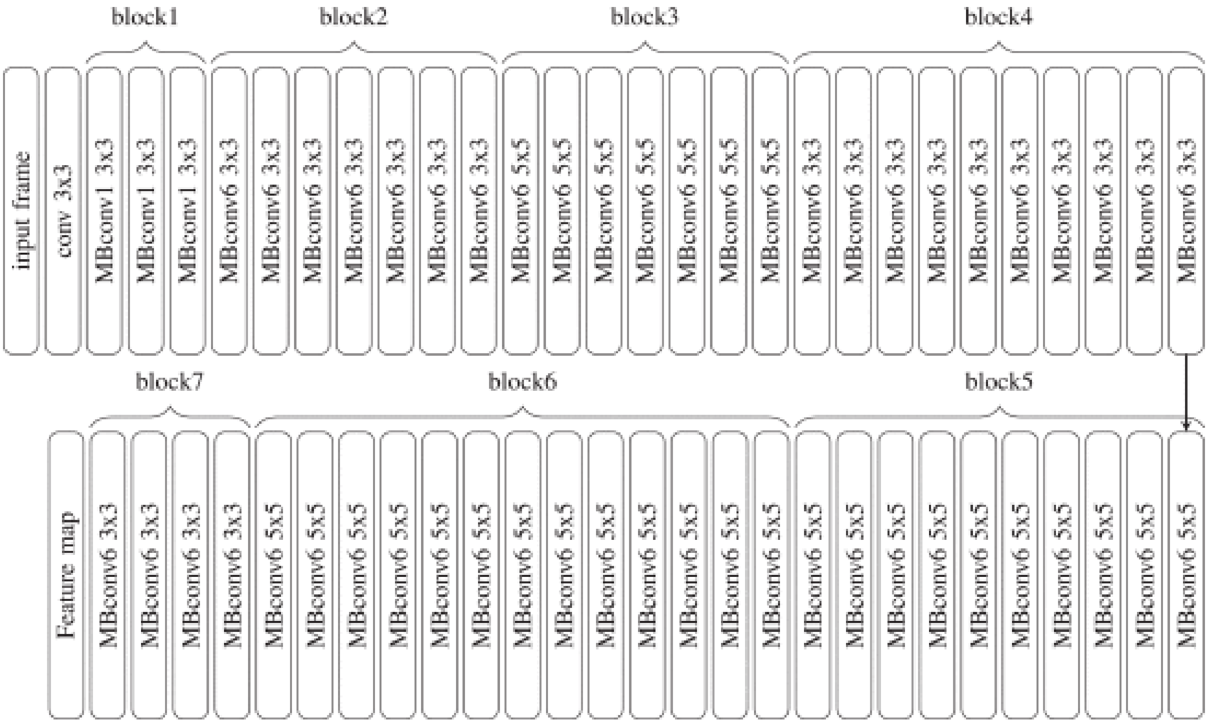

Figure 11: EfficientNetB7 Architecture

4.3.2 Densenet-201

In this research, a pre-trained DenseNet201 neural networks is used for classification in the proposed model. Each layer in DenseNet Godlin Jasil and Ulagamuthalvi (2021) takes inputs from all previous layers and its own feature-maps. The feature map of all preceding layers reduces the network's channels. 3×3 Conv with the growth rate k as extra channels for Batch Norm (BN) and ReLU layers. 2487 dermoscopic pictures from seven classes trained DenseNet201. The architecture's last layer in figure 12 is changed by eliminating the higher layer and replacing it with a softmax layer that classifies seven groups. The input layer receives 224×224 pixel pictures.

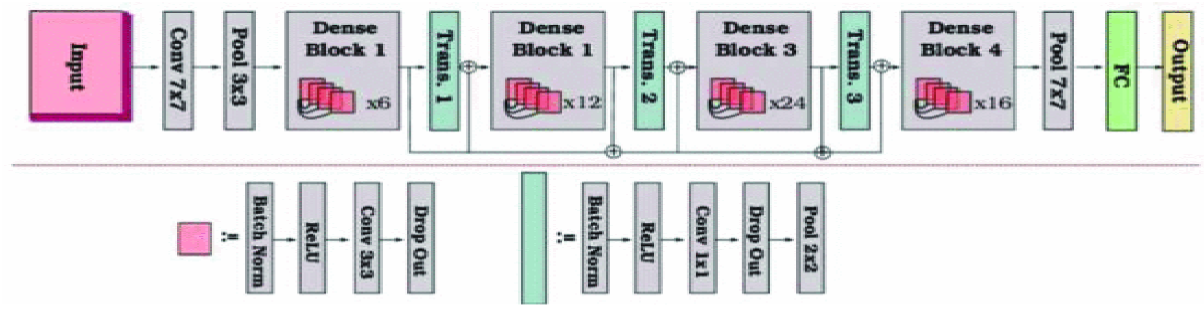

Figure 12: DenseNet201 Architecture

4.3.3 VGG16

As illustrated in Figure 13, The VGG16 spiral neural network has 13 layer layers, 3 completely connected layers, numerous layer layers, and pond layers in Yi et al. (2022). Pooled layers are not included in total layers since they are weightless. Collapsed layers and bow layers are input picture extraction processes that combine multilayer collapsed layers to provide the network a broader field of vision, reduce network parameters, activate ReLU functions to diversify linear changes, and enhance. Softmax activation function may classify samples across all binding and output layers to generate a probabilistic distribution of the sample in distinct categories.

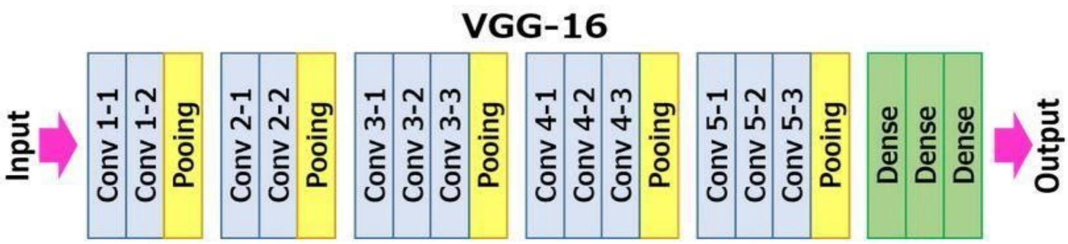

Figure 13: VGG16 Architecture

4.3.4 InceptionResNetV2

Inception-ResNet-v2 trained on over a million ImageNet pictures. The 164-layer network classifies photos into 1000 categories in Guefrechi et al. (2022), including keyboards, mouse, pencils, and animals. Thus, the network has learned many rich aspects from many photographs. Initial structure and residual connections support it. Inception resnet blocks use multi-scale convolutional filters and residual connections. Residual connections address deep structural deterioration and cut training time in half in figure 14 depicts Inception-Resnet v2's network. ResNet and Inception have led recent image recognition advances with great results at minimal computing cost. Inception-ResNet combines residual connections with Inception.

4.3.5 ResNet50V2

The literature review reveals that deep convolutional neural networks (DCNNs) provide a significant contribution to AI applications for image identification and image categorization. Multiple years of study have resulted in the evolution of the deep learning model's

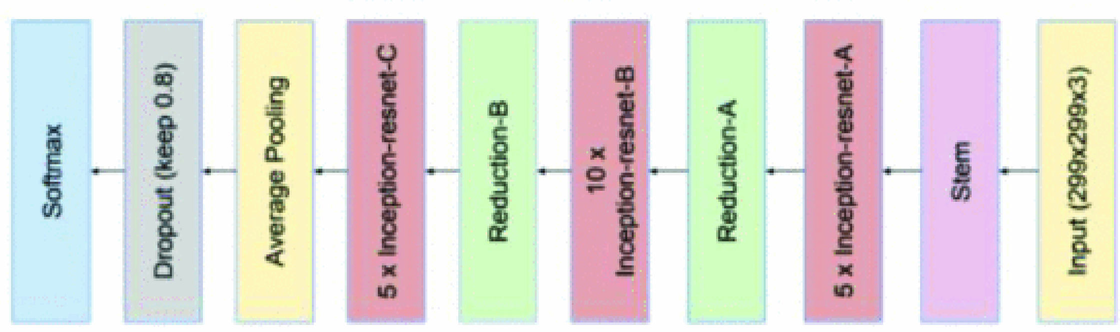

Figure 14: InceptionResNetV2 Architecture

layers, resulting in a more complicated network. This complexity further necessitates that the network improves its robustness. With the addition of more architectural layers in Raje and Jadhav (2022), training the network is tough. This diminishes the model's precision, rendering it totally saturated. This necessitated the development of residual networks ResNets. ResNet architecture is composed of blocks. ResNet's standard design consists of 34 layers and is inspired on the VGG19 model.

5 Implementation

5.1 Environment Setup

This research project is developed with Python 3.6.4 in a Kaggle Notebook environment (RAM: 16GB, GPU:13GB, GPU: P100, CPU: Intel(R) Xeon(R)). All three sets of data were put on the kaggle. The MTCNN pre-processed images were used for the UTK Face dataset and uploaded to the Kaggle drive. TensorFlow[1] and Keras[2] were used to access all of the deep learning models. MTCNN [3] was used for both face alignment and face cropping. At the end, we used scikit-learn [4], Matplotlib [5], and Seaborn[6] to plot graphs and do exploratory data analysis.

5.2 Data Handling

In this research project I perform a thorough examination of the datasets and category-specific data distribution using the meta-data provided. Figures 5, 6, and 7 show the results of my exploratory data analysis for each datasets. Model Training is terminated using Early Stopping Callback to avoid over-fitting if validation loss remains unchanged over 100 iterations.

[1]https://www.tensorflow.org/api_docs/python/tf/keras/layers
[2]https://keras.io/api/applications/
[3]https://github.com/ipazc/mtcnn
[4]https://scikit-learn.org/stable/getting_started.html
[5]https://matplotlib.org/stable/plot_types/index
[6]https://seaborn.pydata.org/api.html

5.3 Model Implementation

- **Model 1: Identification of BMI(Body Mass Index) from face** - Before a model is implemented, custom functions are initially created for data pre-processing, model designs, model stacking, and evaluation. The next step is to set the RGB picture constant to 3, the image size to 224 x 224, the epochs to 100, the trainable to false, and the Train: Test split to 80:20. A batch normalizing layer and a dropout rate of 0.25 were also introduced. The pre-trained models will then be connected by a top layer of the dense architecture. loading of the densenet201 pre-trained model by removing the top layer, which is nothing more than a softmax layer. For pre-trained models, I'll utilize specified weights from the ImageNet competition that attained cutting-edge performance. The optimizer is Adam, the loss function is mean squared error, and the metrics are rmse and mae. As seen in Fig 15, the training and loss data are meticulously recorded and examined for model overfitting.

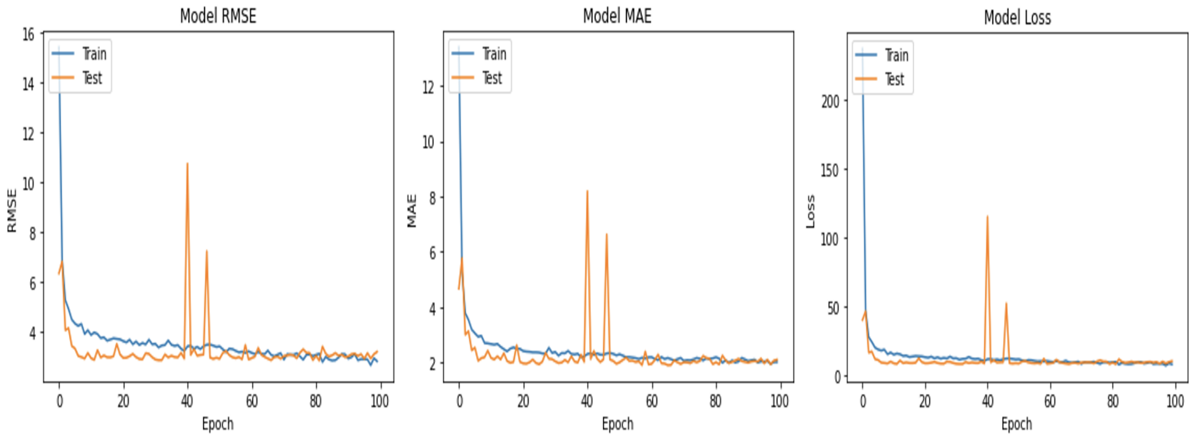

Figure 15: EfficientNetB7 Trained Model for BMI Identification

- **Model 2: Classifying the smoker or non-smoker** - Custom functions are first defined for data pre-processing, model designs, model stacking, and evaluation before a model is implemented. The next step is to declare the constants for the RGB picture as 3, the image size as 224 x 224, the epochs as 100, the trainable as false, and the Train: Test split as 80:20. Additionally, a batch normalizing layer was added, along with a dropout rate of 0.25. Next, a top layer of dense architecture that connects all the pre-trained models will be created. The set includes top=False to remove the top layer of the model and add dense layers with the number of categories to predict. I will use predetermined weights from the ImageNet competition, which achieved state-of-the-art performance, for pre-trained models. Adam is the optimizer while binary cross entropy serves as the loss function. The metrics include accuracy, precision, and recall. The models are then contrasted and the best evaluation metric is used to make the final decision. The traning model accuracy, precision and recall till epoch 100 were calculated and later the results were recorded as shown in Fig 16.

- **Model 3: Determining the Age of a person** - For determining the age, since it is a regression problem, the first was to convert the UTK data set images into preprocessed images using MTCNN. MTCNN consist of three neural network (P-net,O-net and R-net) as defined in Figure 10. Later, before applying the CNN dense

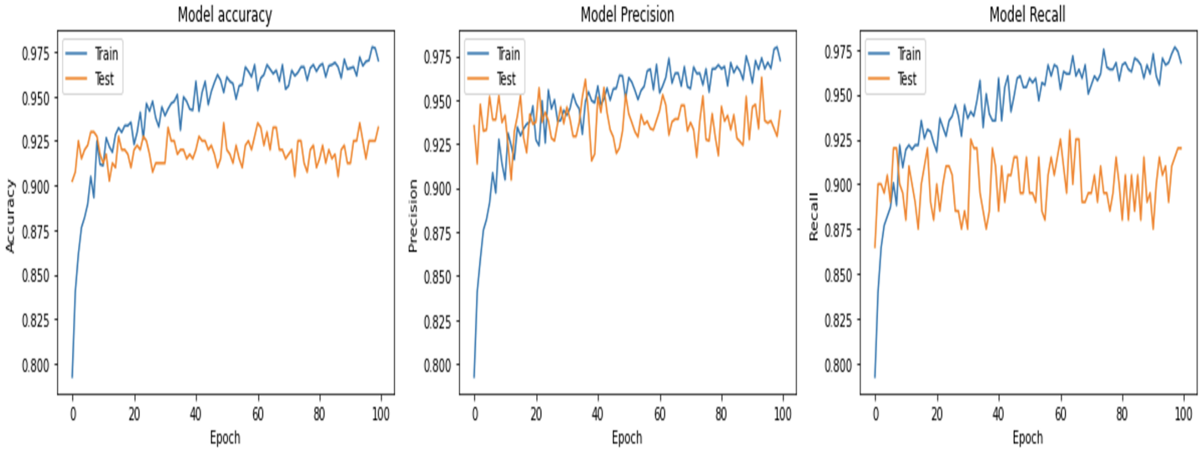

Figure 16: EfficientNetB7 Trained Model for Classifying between smoker or non-smoker

layers, data augmentation like rotation, shearing , horizontal flip, vertical flip were performed using *ImageDataGenerator()* function. Data was split between 80-20 ration. The rotation_range was set to 15, re-scaling was set to 0.0003, zoom_range set to 20% and horizontal_flip was set to True. After this, model.fit_generator() was used to fit the model with callback as defining checkpoints for best epoch model saving and early stopping if there is no improvement in learning. For the evaluation, loss function is mean sqaured error and optimizer is adam , metrics are rmse , mae. The detailed training model metrics are shown in Figure 17.

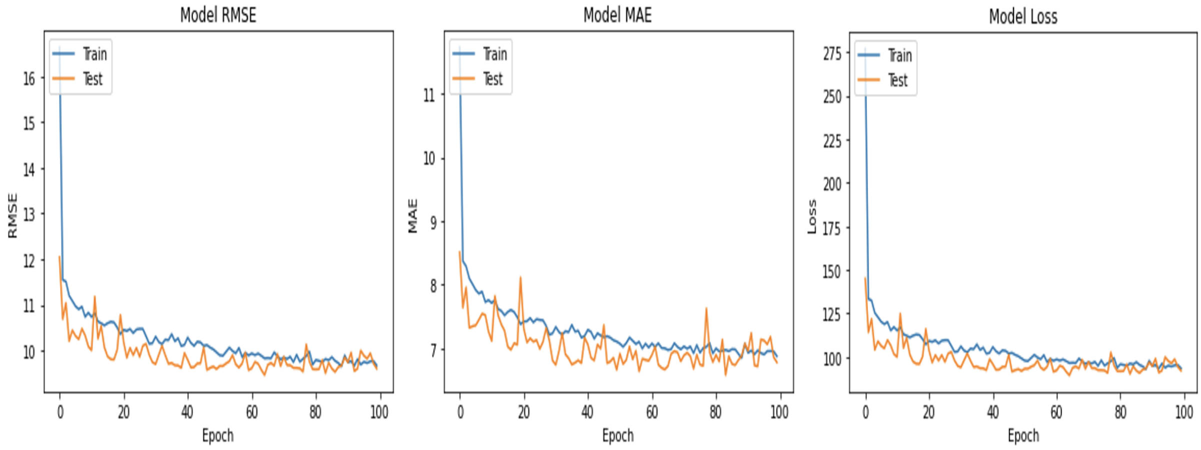

Figure 17: EfficientNetB7 Trained Model for Age

- **Model 4: Determining the Gender of a person** - For determining the gender using UTK Face dataset, MTCNN was used to preprocess UTK data set pictures since there were too much background noise associated with the images. Figure 9 shows MTCNN's three neural networks: P-net, O-net, and R-net. Before adding CNN dense layers, ImageData-Generator() was used to rotate, shear, flip, and rotate. Rotation was 15, re-scaling was 0.0003, zoom range was 20%, and horizontal flip was True. After this, model.fit generator() was used to fit the model with callbacks to provide checkpoints for best epoch model saving and early terminating if learning did not improve. For evaluation, loss function is mean squared error, optimizer is adam, metrics are rmse and mae. The training and loss metrics are carefully noted and checked for model over-fitting as shown in Fig 18.

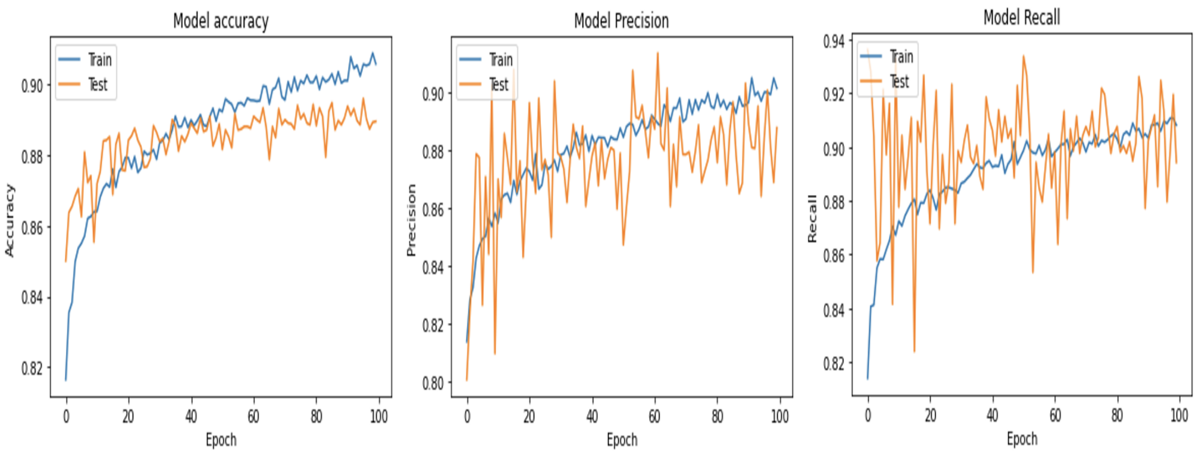

Figure 18: EfficientNetB7 Trained Model Gender Classification

- **Model 5: Determining the Ethnicity of a person** - For determining the ethnicity of a person, which is a multi-class classification problem, firstly as it a from the similar dataset that is UTK face dataset, MTCNN was performed and later various data augmentation attributes were applied like as shown in figure 4. For the dense layer, trainable was set to False, then input layer of image size 224* 224 * 3 was supplied as input to dense layer with Relu Activation function. Later, *GlobalAveragePooling2D()* , *Dropout()* layer with 0,25 dropout rate is added and lastly a *BatchNormalization()* added for fastening the training rate. Finally, one dense layer is added with 128 neurons and it helps in providing more feature extraction for the images and an output layer with softmax as activation is added. Evaluation metrics used were Accuracy, Precision, and Recall and there were recorded for the training model, and the data was recorded as shown in Figure 19.

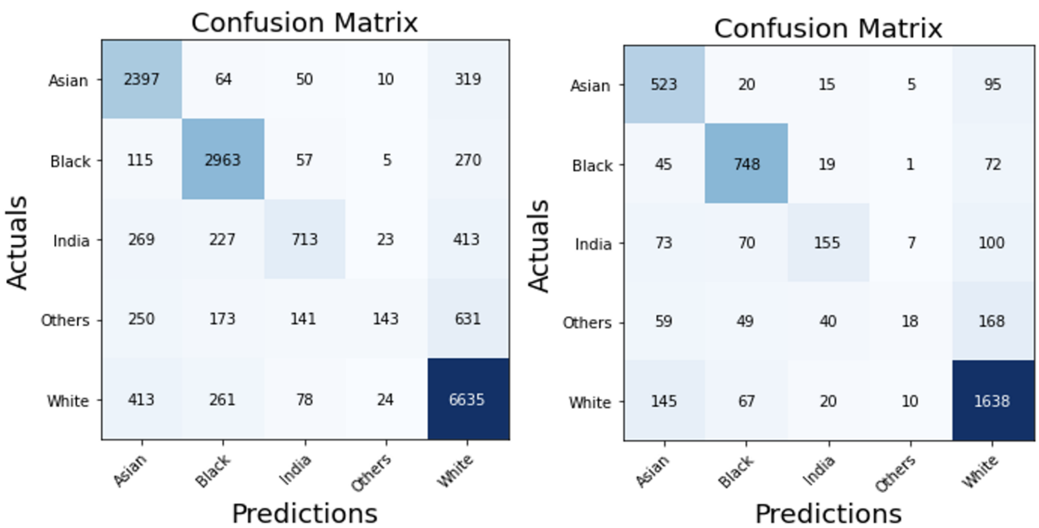

Figure 19: DensetNet201 Trained Model Ethnicity or Region Classification

6 Evaluation

6.1 Experiment 1: BMI Identification

I evaluated models based on bmi identification in the first experiment. Regarding the VIP attribute dataset, EfficientNetB7 scored the greatest RMSE and MAE, 3.18 and 2.09, whereas InceptionResNetV2 fared the poorest and the EfficientNetB7 model suffered the least loss, 10.1%, as shown in Table 1. In addition, the EfficientNetB7 model had the least amount of loss (0.17).

Table 1: BMI results

Model	RMSE		MAE	
	Test	Train	Test	Train
EfficientNetB7	3.1871	2.8063	2.0966	2.0118
DensetNet201	3.3346	3.2611	2.1761	2.2317
VGG16	3.795	3.6466	2.3436	2.3421
InceptionResNetV2	3.1795	3.6867	2.1927	2.3738
ResNet50V2	4.6423	2.9988	3.0637	2.0788

6.2 Experiment 2: Smoker or Non-Smoker

In the second experiment, the identification of smokers which is a binary classification problem was performed. Here Also, out of all the models EfficientNetB7 performed as shown in Table 2 the based with an accuracy of 92.48% on test and 99.75% on training data. The training loss was very negligible for training data for the EfficientNetB7 model whereas the test loss was 0.9. The model which performed the worst was InceptionResNetV with an accuracy of 57.89%.

Table 2: Evaluation of Smoking Based Binary Classification Results

Model	Accuracy		Precision		Recall	
	Test	Train	Test	Train	Test	Train
EfficientNetB7	92.48%	99.75%	92.52%	99.75%	92.48%	99.75%
DensetNet201	87.72	95.02%	88.15	95.02%	87.71	95.02%
VGG16	83.46%	86.82%	83.45%	87.22%	83.49%	86.82%
InceptionResNetV2	57.89%	56.47%	62.93%	61.44%	57.97%	56.47%
ResNet50V2	89.72%	98.01%	90.10%	98.01%	89.71%	98.01%

6.3 Experiment 3: Identification of Age

In the third experiment, which was a regression-based issue and consisted of determining the age of a person based on a continuous variable, Densenet201 outperformed all of the other models with RMSE values of 9.6719 for trained data and 9.6027 for test data. Additionally, MAE scores were 6.7815 and 6.8777 for train and test data respectively. The Densenet201 model saw the lowest amount of data loss during both training and testing. In all models, Resnet50V2 and VGG16 fared the worst shown in Table 3.

Table 3: Evaluation of Age Results

Model	RMSE		MAE	
	Test	Train	Test	Train
EfficientNetB7	20.0096	20.351	15.4478	15.5997
DensetNet201	9.6027	9.6719	6.7815	6.8777
VGG16	12.0148	12.4385	9.0947	9.14
InceptionResNetV2	19.8154	19.1108	16.1248	14.8278
ResNet50V2	19.9665	19.1297	15.5431	14.7911

6.4 Experiment 4: Gender Classification

The fourth experiment consisted on determining the gender of an individual. The best performance was achieved by EfficientNetB7, followed by Densenet201. On training and testing data, EfficientNetB7 obtained an accuracy of 94% and 89%, respectively, with a minimum loss of 0.2231 and 0.2907, respectively. The remaining models fared inadequately, with InceptionResNetV3 doing the worst as shown in Table 4.

Table 4: Gender Based Binary Classification Results

Model	Accuracy		Precision		Recall	
	Test	Train	Test	Train	Test	Train
EfficientNetB7	89.02%	93.99%	89.02%	93.99%	89.02%	93.99%
DensetNet201	87.19%	89.44%	87.25%	89.49%	87.18%	89.47%
VGG16	77.15%	78.04.%	78.09%	79.22%	77.09%	78.17%
InceptionResNetV2	54.76%	55.07%	56.33%	57.49%	54.75%	55.39%
ResNet50V2	71.34%	72.96%	76.08%	77.15%	71.86%	73.08%

6.5 Experiment 5: Ethnicity or Region Based Classification

Densenet201 performed well in the last experiment involving the ethnicity-based multi-class classification issue, achieving an accuracy of 77.21 percent on train data and 74.05 percent on validation data, respectively. All remaining models fared badly, with an accuracy below 50%. The results are shown in Table 5.

Table 5: Multi class Ethnicity classification Results

Model	Accuracy		Precision		Recall	
	Test	Train	Test	Train	Test	Train
EfficientNetB7	43.22%	45.01%	47.01%	48.71%	48.96%	42.97%
DensetNet201	74.05%	77.21%	65.09%	73.71%	58.96%	62.97%
VGG16	49.21%	64.59%	22.37%	75.41	83.57%	52.64%
InceptionResNetV2	47.62%	48.31%	58.20%	58.56%	49.60%	48.90%
ResNet50V2	51.66%	56.78%	61.57%	67.69%	32.34%	40.70%

6.6 Discussion

From the preceding section's findings, it can be observed that all pretrained models performed differently depending on the tasks at hand. The first experiment we conducted was to determine a person's BMI on the VIP Attribute dataset, where the EfficientNetB7 model outperformed all the other models and earned an MAE score of 2.09, outperforming the ResNet50 model results of Dantcheva et al. (2018), which were 2.36, and concluding that BMI prediction does not demonstrate gender bias and the dataset is balanced as well.

For the second experiment, which involved determining whether a person is a smoker or non-smoker utilizing Mendeley Smoker Dataset, in comparison to Pundhir et al. (2022), as seen from the results shown in Table 2, we were able to achieve an accuracy of 99.75% with the EfficientNetB7 model. This meant that our results were superior to those of Pundhir et al. (2022), who achieved an accuracy of 96.74 percent with the Resnet18 and Resnet34 models because first, we choose the latest state of art pre-trained models and also the dataset was balanced as depicted in Figure 5.

For the final three experiments, we choose the UTKFace dataset for age, gender, and ethnicity predictions. Considering that an individual's age is a continuous variable, the third experiment's MAE score was 6.8777 for the train data and 6.7815 for the test data correspondingly. This score was attained by identifying an individual's age. In the beginning, we were unable to acquire decent results since the photos from the UTKFace dataset were too noisy and included low-resolution images. In order to solve this issue, we used MTCCN for its face alignment and face cropping capability as discussed in Section 4 as shown in Figure 10. When it comes to the gender classification, we were able to achieve an accuracy of 93.99% on the test data and 89.02% on the validation data. This is in comparison to Valliappan Raman (2022), who proposed VGG16, Resnet50, and Densenet201 pre-trained models and achieved an accuracy of 80.58%, 77.87%, and 79.31% respectively for gender classification using these models. As can be seen, we fared much better than the earlier research that used the same dataset. This is likely due to the fact that the collection previously had numerous photographs of poor quality. MTCNN, which was utilized to produce quick face identification and face alignment, was the solution that we settled on in order to tackle this challenge. This solution is discussed in the Modelling portion as shown in Figure 9. For the last experiment, which included an ethnicity classification task, the DenseNet201 model outscored all others and obtained a 77.21 percent accuracy. As demonstrated in Figure 7, the dataset was very uneven and skewed, which explains why the accuracy is satisfactory. As seen, Indians may be characterized as either white or black, depending on the location in which they reside.

Using this methodology, the study was able to get improved results for gender classification, BMI identification, smoking status of a person, and age classification, however, ethnicity classification was graded as average. In order to determine the ethnicity of a person, it will be crucial to collect more data points and choose more high-resolution photos.

7 Conclusion and Future Work

As indicated in Section 6.6, the pre-assigned weights and pre-defined architecture of pre-trained model networks allowed for better accuracy, demonstrating the use of state-of-the-art methodologies for assessing human face features using five state-of-the-art pre-trained models. Although the proposed methods enhanced the accuracy of gender prediction, BMI identification, age estimation, and smoking or non-smoking status, the dataset quality was insufficient for ethnicity classification. The EfficientNetB7 model surpassed all others in gender prediction, BMI identification, and identification of smokers or non-smokers, whereas the Densenet201 model outperformed all others in Age and Ethnicity classification. Lastly, in this research, I have used three datasets and developed deep learning models on five different transfer learning models.This study may be used in the insurance industry, schools, universities, and sports academies to expedite the manual calculation of human attributes.

Future research might recover more human characteristics from the human face by incorporating data from other sources, such as video or high-quality images. In addition, we may employ video as a data source and strengthen the capability of extracting human face traits to expand our reach. To further expand its applicability, the model may be evaluated on more heavy datasets. Also, in the future, a method may be devised to further integrate the information and get all the essential features from a single picture or video. Transfer learning, which employs a previously trained version of a popular model, and ensemble networks, which make better predictions and achieve better performance than any single contributing model, will be the focus of future research into improving the network's behavior and the model's performance.

Acknowledgement

I would like to take this opportunity to thank Prof. Jorge Basilio, my supervisor, and mentor, as well as the other members of the NCI faculty, for the information they have provided me with and the direction they have provided for me during the Data Analytics course. In conclusion, I would want to express my gratitude to my family and friends for all of the support and encouragement they have provided over the whole of my academic career.

References

Belcar, D., Grd, P. and Tomičić, I. (2022). Automatic ethnicity classification from middle part of the face using convolutional neural networks, *Informatics* **9**(1).
URL: *https://www.mdpi.com/2227-9709/9/1/18*

Dantcheva, A., Bremond, F. and Bilinski, P. (2018). Show me your face and i will tell you your height, weight and body mass index, *2018 24th International Conference on Pattern Recognition (ICPR)*, pp. 3555–3560.

Gao, S., Zeng, C., Bai, M. and Shu, K. (2020). Facial ethnicity recognition based on transfer learning from deep convolutional networks, *2020 3rd International Conference on Advanced Electronic Materials, Computers and Software Engineering (AEMCSE)*, pp. 310–314.

Godlin Jasil, S. and Ulagamuthalvi, V. (2021). Skin lesion classification using pre-trained densenet201 deep neural network, *2021 3rd International Conference on Signal Processing and Communication (ICPSC)*, pp. 393–396.

Guefrechi, S., Jabra, M. B. and Hamam, H. (2022). Deepfake video detection using inceptionresnetv2, *2022 6th International Conference on Advanced Technologies for Signal and Image Processing (ATSIP)*, pp. 1–6.

Gupta, J., Pathak, S. and Kumar, G. (2022). Deep learning (cnn) and transfer learning: A review, *Journal of Physics: Conference Series* **2273**(1): 012029.
URL: *https://dx.doi.org/10.1088/1742-6596/2273/1/012029*

Ioffe, S. and Szegedy, C. (n.d.). Batch normalization: Accelerating deep network training by reducing internal covariate shift.
URL: *https://arxiv.org/abs/1502.03167*

Kangwanwatana, S. and Sucontphunt, T. (2022). Improve face verification rate using image pre-processing and facenet, *2022 7th International Conference on Business and Industrial Research (ICBIR)*, pp. 426–429.

Khan, A. (2020). Dataset containing smoking and not-smoking images (smoker vs non-smoker), *Mendeley Data*, Vol. V1.

Klontz, J. C. and Jain, A. K. (2013). A case study on unconstrained facial recognition using the boston marathon bombings suspects, *Michigan State University, Tech. Rep* **119**(120): 1.

Koffas, S., Picek, S. and Conti, M. (2022). Dynamic backdoors with global average pooling, *2022 IEEE 4th International Conference on Artificial Intelligence Circuits and Systems (AICAS)*, pp. 320–323.

Mays, V. M., Ponce, N. A., Washington, D. L. and Cochran, S. D. (2003). Classification of race and ethnicity: Implications for public health, *Annual Review of Public Health* **24**(1): 83–110. PMID: 12668755.
URL: *https://doi.org/10.1146/annurev.publhealth.24.100901.140927*

Narang, N. and Bourlai, T. (2016). Gender and ethnicity classification using deep learning in heterogeneous face recognition, *2016 International Conference on Biometrics (ICB)*, pp. 1–8.

OuYang, G. (2018). Design and implementation of body mass index calculator based on android technology, *2018 2nd IEEE Advanced Information Management,Communicates,Electronic and Automation Control Conference (IMCEC)*, pp. 2459–2462.

Padmavathi, K., Thalla, O. S. R. R. Y., Reddy, S. S., Yadlapalli, P., Roshan, T. and Charan, T. (2022). Transfer learning using efficientnet for brain tumor classification from mri images, *2022 International Conference on Computer Communication and Informatics (ICCCI)*, pp. 1–4.

Pundhir, A., Verma, D., Kumar, P. and Raman, B. (2022). Region extraction based approach for cigarette usage classification using deep learning, *in* B. Raman, S. Murala, A. Chowdhury, A. Dhall and P. Goyal (eds), *Computer Vision and Image Processing*, Springer International Publishing, Cham, pp. 378–390.

Raje, N. R. and Jadhav, A. (2022). Automated diagnosis of pneumonia through capsule network in conjunction with resnet50v2 model, *2022 International Conference on Emerging Smart Computing and Informatics (ESCI)*, pp. 1–6.

Savchenko, A. (2019). Efficient facial representations for age, gender and identity recognition in organizing photo albums using multi-output convnet, *PeerJ Computer Science* **5**: e197.

Selim, T., Elkabani, I. and Abdou, M. A. (2022). Students engagement level detection in online e-learning using hybrid efficientnetb7 together with tcn, lstm, and bi-lstm, *IEEE Access* **10**: 99573–99583.

Sethi, M., Ahuja, S., Singh, S., Snehi, J. and Chawla, M. (2022). An intelligent framework for alzheimerx0027;s disease classification using efficientnet transfer learning model, *2022 International Conference on Emerging Smart Computing and Informatics (ESCI)*, pp. 1–4.

Siddiqui, H., Rattani, A., Kisku, D. R. and Dean, T. (2020). Al-based bmi inference from facial images: An application to weight monitoring, *2020 19th IEEE International Conference on Machine Learning and Applications (ICMLA)*, pp. 1101–1105.

Srinivas, N., Atwal, H., Rose, D. C., Mahalingam, G., Ricanek, K. and Bolme, D. S. (2017). Age, gender, and fine-grained ethnicity prediction using convolutional neural networks for the east asian face dataset, *2017 12th IEEE International Conference on Automatic Face Gesture Recognition (FG 2017)*, pp. 953–960.

Syrowatka, C., Windhager, S., Schaefer, K. and Mitteroecker, P. (2017). Bmi and whr are reflected in female facial shape and texture: A geometric morphometric image analysis, *PLoS ONE* **12**: e0169336.

Valliappan Raman, Khaled ELKarazle, P. T. (2022). Gender-specific facial age group classification using deep learning, *Intelligent Automation & Soft Computing* **34**(1): 105–118.
URL: *http://www.techscience.com/iasc/v34n1/47369*

Yi, W., Ma, S., Zhang, H. and Ma, B. (2022). Classification and improvement of multi label image based on vgg16 network, *2022 3rd International Conference on Information Science, Parallel and Distributed Systems (ISPDS)*, pp. 243–246.

Zhang, Zhifei, S. Y. and Qi, H. (2017). Age progression/regression by conditional adversarial autoencoder, *IEEE Conference on Computer Vision and Pattern Recognition (CVPR)*, IEEE.

ŞEN, S. Y. and ÖZKURT, N. (2020). Convolutional neural network hyperparameter tuning with adam optimizer for ecg classification, *2020 Innovations in Intelligent Systems and Applications Conference (ASYU)*, pp. 1–6.