**BMI (BODY MASS INDEX)**

**PREDICTION USING FACE IMAGE**

**PROJECT REPORT**

***Submitted by***

**Aditya Viswabhusan**

## Abstract

Human faces encode plenty of useful information. Recent studies in

psychology and human perception have found that facial features have relations to human weight or body mass index (BMI). These studies focus on finding the correlations between facial features and the BMI. Motivated by the recent psychology studies, we develop a computational method to predict the BMI from face images automatically. We formulate the BMI prediction from facial features as a machine vision problem, and evaluate our approach on a large database with more than 14,500 face images. A promising result has been obtained, which demonstrates the feasibility of developing a computational system for BMI prediction from face images at a large scale.

***Keywords:***  Body mass index (BMI) face image, FaceNet, MTCNN

## CONTENTS

**Title Page Number**

[**Abstract** **i**](#_Toc29712)

[**Contents** **ii**](#_Toc29713)

[**List of Figures iii**](#_Toc29714)

[**List of Tables** **iv**](#_Toc29715)

**CHAPTER 1: Introduction**

1.1 Introduction to BMI

1.2 Advantages and Disadvantages

1.3 Motivation and Objective

1.4 CNN

1.5 MTCNN

**CHAPTER2: Literature Survey**

2.1 Introduction

2.2 Comparisons Table

2.3 Conclusions

**CHAPTER3: Deep Learning**

3.1 Introduction

3.2 Problem Definition

3.3 FaceNet

3.4 Deep Learning

3.5 Image Augmentation

3.6 Block Diagram

3.7 Result Verification

3.8 Results and Discussions

### CHAPTER4: Conclusions

4.1 Conclusions

4.2 Future Scope

**References**

## LIST OF FIGURES

**Title** **Page Number**

|  |  |
| --- | --- |
| Fig.1.1 CNN |  |
| Fig.1.2 Image matrix multiples kernel or filter matrix |  |
| Fig.1.3 Feature Selection  Fig.1.4 Image pyramid  Fig.1.5 P-Net  Fig.1.6 R-Net  Fig.1.7 O-Net  Fig.3.1 Deep Learning  Fig.3.2 CNN layers  Fig.3.3 FaceNet takes an image of a face as input and output the embedding  Fig.3.4 Image Transformation of Training Data  Fig.3.5 Classification  Fig.3.6 BMI chart  Fig.3.7 predicted values from our dataset  Fig.3.8 Experimental outcome |  |

## LIST OF TABLES

|  |  |
| --- | --- |
| **Title** | **Page Number** |
| Table.2.1 Comparison Table |  |

### Chapter 1

**Introduction**

1

**Chapter 1**

### Introduction

#### 1.1 Introduction to BMI

**Body mass index** (**BMI**) is a value derived from the mass (weight) and height of a person. The BMI is defined as the body mass divided by the square of the body height, and is expressed in units of kg/m2, resulting from mass in kilograms and height in metres.

The BMI may be determined using a table or chart which displays BMI as a function of mass and height using contour lines or colours for different BMI categories, and which may use other units of measurement (converted to metric units for the calculation).

The BMI is a convenient rule of thumb used to broadly categorize a person as *underweight*, *normal weight*, *overweight*, or *obese* based on tissue mass (muscle, fat, and bone) and height. Commonly accepted BMI ranges are underweight (under 18.5 kg/m2), normal weight (18.5 to 25), overweight (25 to 30), and obese (over 30).

BMIs under 20 and over 25 have been associated with higher all-causes mortality, with the risk increasing with distance from the 20–25 range.

Body Mass Index (BMI) is a well-utilized measure to generally describe weight status based on the ratio between an individual’s height and weight. It is used to ultimately classify an individual as underweight, normal weight, overweight, or obese. Within dissemination and implementation research in particular, measuring BMI can be an obstacle that interferes with obtaining outcome data. First, reaching participants to physically measure BMI is unlikely when aiming to disseminate a program nationally due to limited time available and the resources needed by facilitators to accurately measure weight and height (a research quality stadiometer and scales) as well as carrying out the intervention. Further, relying on individuals to perform self-measurements may not be realistic or reliable, and measurements would often rely on corrective equations. Nevertheless, self-reported measurements are typically used as a secondary measure when physical measurements are not feasible. Previous research has shown that participants tend to overestimate their height and underestimate their weight, resulting in inaccurate BMI data.

Technologies 2018,6, 83 2 of 8adults tended to overestimate their height while underestimating weight. Similar findings from the National Health and Nutrition Examination Survey (NHANES) and the National Institutes of Health (NHI) data found younger (18–42 years) and older (>55 years old) adults inaccurately reported BMI, with over-reported height and under-reported weight. A solution to this problem may be the utilization of participant facial photographs to assess BMI. With today’s ever evolving technology, ‘selfies’, iPhone facial recognition, and social media sites enable users to easily take and display self-portraits. Recently, life insurance companies have begun offering the option of sending in a selfie to assess health and weight status for eligibility. When performing studies among a population that has grown up in a technology-based atmosphere, such as current college students, the ease of taking a portrait may entice them to participate in health interventions. This type of facial recognition software has been used for various reasons including by the Department of defence, and for criminal investigations, protection, and emotional recognition. However, increases and decreases in weight in a single participant can make it difficult for facial recognition technology to accurately

identify an individual.

***1.2 Advantages and Disadvantages***

#### Advantages

One advantage of predicting BMI from face images is that, the approach is non-invasive. There is no need to measure an individual's height and weight in order to compute his or her BMI. This is a nice property for some practical uses, such as in on-line photos or surveillance videos of faces, where it is impossible to use traditional measures of weight and height for BMI calculation. Recently, there are more and more online dating or friend search sites (e.g., http:// www.onlinedatingsites.net/), where possibly only face photos are shown for each individual. The automated prediction of BMI from face photos can be useful to judge bodily attractiveness and health.

#### Disadvantages

An obvious shortcoming of the model is poor performance when evaluating images taken from angles other than from the front of the subject.

Another possible shortcoming of the model is the inaccurate prediction of the image when the subject is in a dark environment and is illuminated by a concentrated source of light. An intense concentration of light on a subject’s face might make a difference in how salient certain features are to the model —shadows are exaggerated, one side of the face might be lighter than the other, and curvatures and subtleties in skin appearance or bone structure might be blotted out.

#### 1.3 Motivation and Objectives

#### Motivation

Obesity is an important public health concern, and an understanding of the difficulty in reducing obesity in individuals despite the seemingly simple energy balance equation that underlies all weight gain remains elusive. Excessive weight has been associated with obesity, diabetes, and cardiovascular diseases.

Limited attention has been given to the connection between the human face and body characteristics such as body height and weight and even less so on the automatic extraction of such. Estimating body height, weight and the associated BMI is warranted for several reasons.

Firstly, height and weight are attributes frequently used in surveillance, forensics, as well as reidentification applications and image retrieval systems.

Secondly, height and weight are primary and obvious attributes used by humans to verbally describe a person often used in police reports, unlike traditional biometrics which may be insufficient, as this was argued, for example, by Klontz and Jain in the case of the 2013 Boston bombings.

Thirdly, body weight and height have been proposed as soft biometric traits in automated biometric systems.

#### Objectives

* To apply data and image pre-processing operations.
* To apply Facial Recognition algorithms.
* To predict BMI of an individual using his/her face photograph.

#### 1.4 CNN

Convolutional Neural Network is one of the main categories to do image classification and image recognition in neural networks. Scene labeling, objects detections, and face recognition, etc., are some of the areas where convolutional neural networks are widely used.

CNN takes an image as input, which is classified and process under a certain category such as dog, cat, lion, tiger, etc. The computer sees an image as an array of pixels and depends on the resolution of the image. Based on image resolution, it will see as h \* w \* d, where h= height w= width and d= dimension. For example, An RGB image is 6 \* 6 \* 3 array of the matrix, and the grayscale image is 4 \* 4 \* 1 array of the matrix.

In CNN, each input image will pass through a sequence of convolution layers along with pooling, fully connected layers, filters (Also known as kernels). After that, we will apply the Soft-max function to classify an object with probabilistic values 0 and 1.

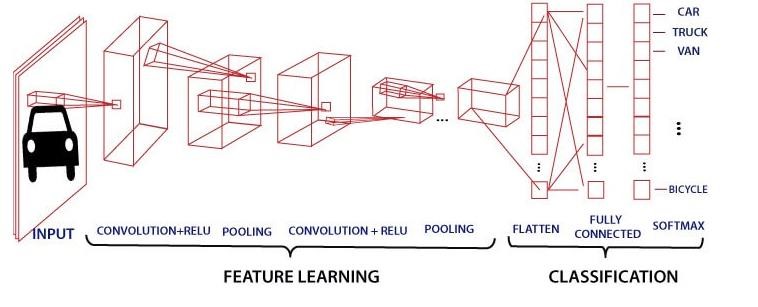


Fig1.1 – CNN

#### Convolution Layer

Convolution layer is the first layer to extract features from an input image. By learning image features using a small square of input data, the convolutional layer preserves the relationship between pixels. It is a mathematical operation which takes two inputs such as image matrix and a kernel or filter.

o The dimension of the image matrix is **h×w×d**. o The dimension of the filter is **fh×fw×d**. o The dimension of the output is **(h-fh+1)×(w-fw+1)×1**.

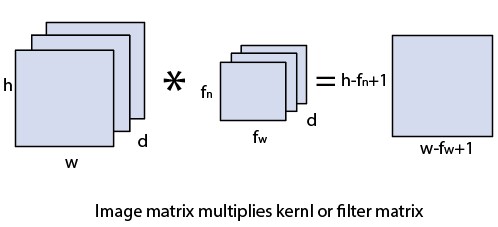
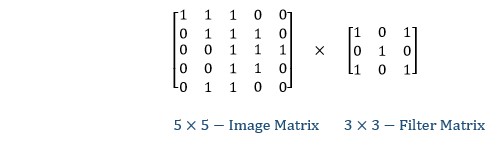
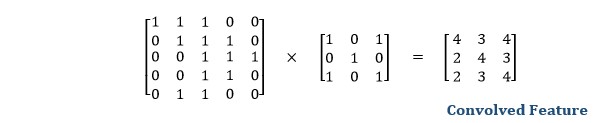


Fig1.2

Let's start with consideration a 5\*5 image whose pixel values are 0, 1, and filter matrix 3\*3 as:



The convolution of 5\*5 image matrix multiplies with 3\*3 filter matrix is called "**Features Map**" and show as an output.



Convolution of an image with different filters can perform an operation such as blur, sharpen, and edge detection by applying filters.

#### 1.4 MTCNN

MTCNN or Multi-Task Cascaded Convolutional Neural Networks is a neural network which detects faces and facial landmarks on images. MTCNN is one of the most popular and most accurate face detection tools today. It consists of 3 neural networks connected in a cascade.

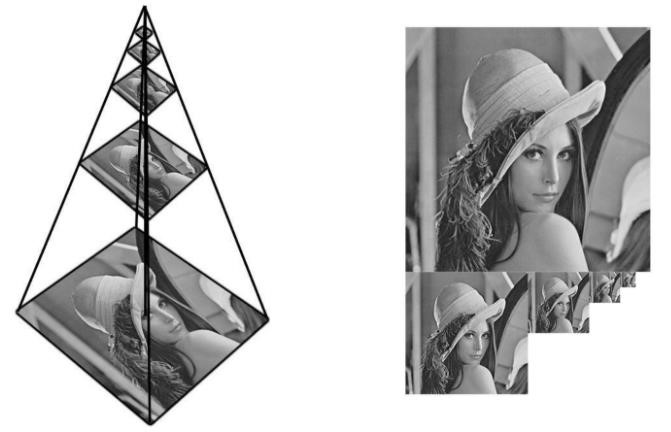


Fig 1.3: Feature selection

**The Three Stages of MTCNN:**

The first step is to take the image and resize it to different scales in order to build an image pyramid, which is the input of the following three-staged cascaded network.

Input image is resized to different scales to build an image pyramid



**Fig1.4:- image pyramid**

#### Stage 1: The Proposal Network (P-Net)

This first stage is a fully convolutional network (FCN). The difference between a CNN and a FCN is that a fully convolutional network does not use a dense layer as part of the architecture. This Proposal Network is used to obtain candidate windows and their bounding box regression vectors.

Bounding box regression is a popular technique to predict the localization of boxes when the goal is detecting an object of some pre-defined class, in this case faces. After obtaining the bounding box vectors, some refinement is done to combine overlapping regions. The final output of this stage is all candidate windows after refinement to downsize the volume of candidates.

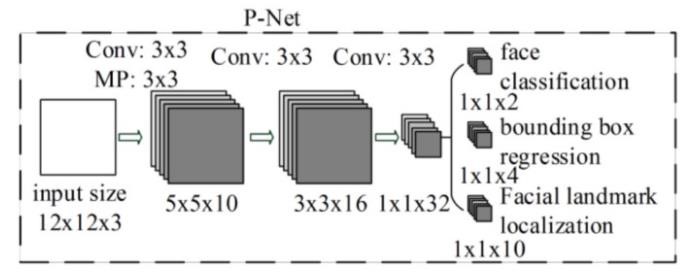
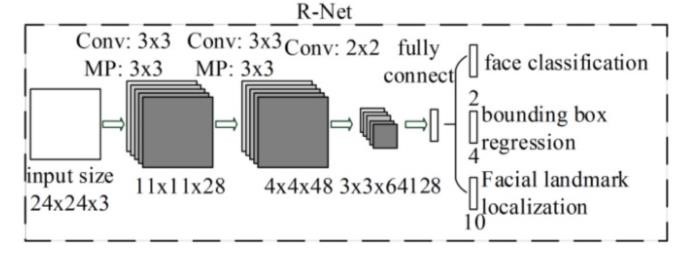


Fig 1.5 - P-Net

#### Stage 2: The Refine Network (R-Net)

All candidates from the P-Net are fed into the Refine Network. Notice that this network is a CNN, not a FCN like the one before since there is a dense layer at the last stage of the network architecture. The RNet further reduces the number of candidates, performs calibration with bounding box regression and employs non-maximum suppression (NMS) to merge overlapping candidates.

The R-Net outputs whether the input is a face or not, a 4 element vector which is the bounding box for



the face, and a 10 element vector for facial landmark localization.

Fig1.6: R-Net

#### Stage 3: The Output Network (O-Net)

This stage is similar to the R-Net, but this Output Network aims to describe the face in more detail and output the five facial landmarks’ positions for eyes, nose and mouth.

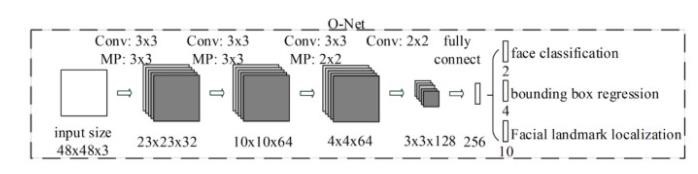


Fig 1.7 : O-Net

### Chapter 2

**Literature Survey**

12

#### Literature Survey

**Chapter 2**

### Literature Survey

#### 2.1 Introduction

Body-mass index (BMI) is the amount of mass per area of a person, and it is an important indicator of the weight status. From health industry till the social media applications, there are many areas where BMI data is used. Various machine learning techniques are developed for BMI prediction only from a face image without any information about the weight and height of a person. Making this kind of predictions is a regression problem. In this study, a deep network based BMI predictor tool is designed and its performance is compared with the existing methods from previous studies.

#### 2.2 Comparisons Table

|  |  |  |  |
| --- | --- | --- | --- |
| **Author** | **Title** | **Proposed**  **Technique** | **Journal, Year** |
| Gülpınar  Bölükbas,  Emrah  Basaran and  Mustafa  Ersel  Kamasak . | Estimating  Body Mass  Index from  Facial Images BMI Prediction  From Face  Images. | Deep Learning, Transfer learning  Support Vector  Machine(SVM) | 2019 27th Signal Processing and Communications Applications  Conference (SIU) |
| Lingyun  Wen, | A  computational | Machine vision | 2013 Elsevier B.V |

13

#### Literature Survey

|  |  |  |  |
| --- | --- | --- | --- |
| Guodong  Guo | approach to body mass  index prediction from face images. |  |  |
| Habib  Bipembi , J.  B. Hayfron-  Acquah,  Joseph K.  Panford,  Obed  Appiah | Calculation of  Body Mass  Index using  Image Processing  Techniques. | Deep Learning, Transfer learning  Support Vector  Machine(SVM) | International Journal of Artificial Intelligence and Mechatronics  Volume 4, Issue 1, ISSN 2320 –  5121 |
| Makenzie L.  Bar,  Guodong  Guo, Sarah E. Colby, Melissa D.  Olfert | Detecting Body  Mass Index from a Facial Photograph in  Lifestyle  Intervention | Correlation, regression function, | Technologies 2018, 6, 83; doi:10.3390/technologies6030083 |

#### 2.3 Conclusions

We studied a lot of IEEE papers where we got to know that SVM and MTCNN is the technique which we are using in our project. MTCNN is the best Technique as per the above IEEE papers.

14

### Chapter 3

**Deep Learning**

#### Chapter 3 Deep Learning

##### 3.1 Introduction

Deep Learning is at the cutting edge of what machines can do, and developers and business leaders absolutely need to understand what it is and how it works. This unique type of algorithm has far surpassed any previous benchmarks for classification of images, text, and voice.

It also powers some of the most interesting applications in the world, like autonomous vehicles and real time translation. There was certainly a bunch of excitement around Google’s Deep Learning based Alpha Go beating the best Go player in the world, but the business applications for this technology are more immediate and potentially more impactful. This post will break down where Deep Learning fits into the ecosystem, how it works, and why it matters.

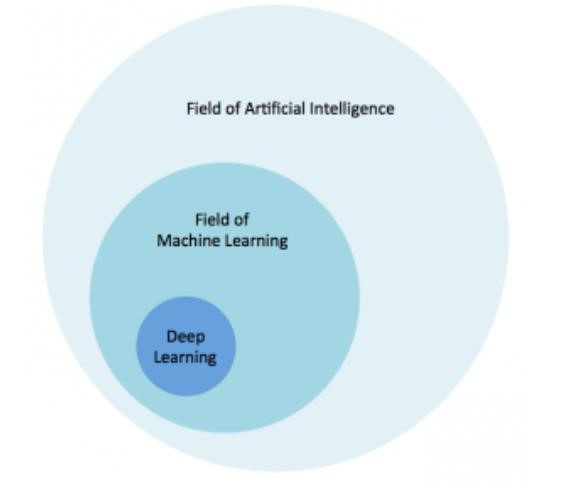


Fig3.1:

16

Deep learning is a specific subset of Machine Learning, which is a specific subset of Artificial Intelligence.

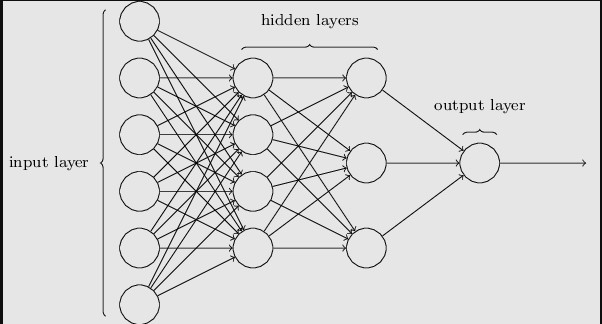
For individual definitions:

* *Artificial Intelligence is the broad mandate of creating machines that can think intelligently*
* *Machine Learning is one way of doing that, by using algorithms to glean insights from data (see our gentle introduction here)*
* *Deep Learning is one way of doing that, using a specific algorithm called a Neural Network* Don’t get lost in the taxonomy – Deep Learning is just a type of algorithm that seems to work really well for predicting things. Deep Learning and Neural Nets, for most purposes, are effectively synonymous. If people try to confuse you and argue about technical definitions, don’t worry about it: like Neural Nets, labels can have many layers of meaning.

Neural networks are inspired by the structure of the cerebral cortex. At the basic level is the perceptron, the mathematical representation of a biological neuron. Like in the cerebral cortex, there can be several layers of interconnected perceptions.

Input values, or in other words our underlying data, get passed through this “network” of hidden layers until they eventually converge to the output layer. The output layer is our prediction: it might be one node if the model just outputs a number, or a few nodes if it’s a multiclass classification problem.

The hidden layers of a Neural Net perform modifications on the data to eventually feel out what its relationship with the target variable is. Each node has a weight, and it multiplies its input value by that weight. Do that over a few different layers, and the Net is able to essentially manipulate the data into something meaningful. To figure out what these small weights should be, we typically use an algorithm called back propagation.



##### Fig3.2-CNN

###### 3.2 Problem Definition

Obesity is an important public health concern, and an understanding of the difficulty in reducing obesity in individuals despite the seemingly simple energy balance equation that underlies all weight gain remains elusive. Excessive weight has been associated with obesity, diabetes, and cardiovascular diseases. Limited attention has been given to the connection between the human face and body characteristics such as body height and weight and even less so on the automatic extraction of such. Estimating body height, weight and the associated BMI is warranted for several reasons. Firstly, height and weight are attributes frequently used in surveillance, forensics, as well as re-identification applications and image retrieval systems. Secondly, height and weight are primary and obvious attributes used by humans to verbally describe a person often used in police reports, unlike traditional biometrics which may be insufficient, as this was argued, for example, by Klontz and Jain in the case of the 2013 Boston bombings. Thirdly, body weight and height have been proposed as soft biometric traits in automated biometric systems.

18

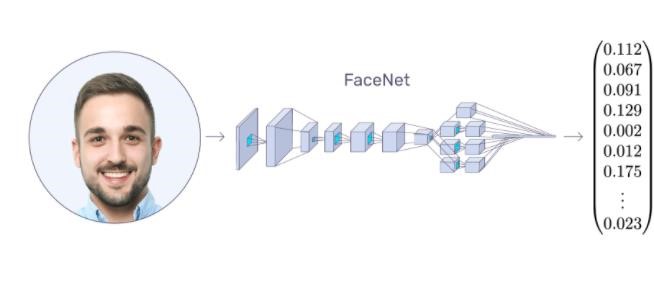
###### 3.3 Proposed Technique name

*MTCNN: pre-trained Convolution Neural Network to extract faces from images*

*FaceNet (InceptionResnetV1): pre-trained Convolution Neural Network To extract features (512) from images*

*SVR: Support Vector Machine FOR REGRESSION*

**FaceNet** is a deep neural network used for extracting features from an image of a person’s face.



**Fig3.3:** *FaceNet takes an image of a face as input and output the embedding factor.*

FaceNet takes an image of the person’s face as input and outputs a vector of 128 numbers which represent the most important features of a face. In machine learning, this vector is called embedding. Why embedding? Because,all the important information from an image is *embedded* into this vector. Basically,

FaceNet takes a person’s face and compresses it into a vector of 128 numbers. Ideally, embedding of similar faces are *also* similar.

**SVM** is a binary classifier based on supervised learning which gives better performance than other classifiers. **SVM** classifies between two classes by constructing a hyper plane in high-dimensional feature space which can be used for classification. SVM is fundamentally a binary classification algorithm. It falls under the umbrella of machine learning.

Image processing on the other hand deals primarily with manipulation of images. For example, image filtering, where an input image is passed through a laplacian filter to be sharpened.

If you want to relate the two, an SVM might be used to perform image classification. For example, given an input image, the classification task is to decide whether an image is a cat or a dog. The image, before being input into the SVM might have gone through some image processing filters so that some features might be extracted such as edges, color and shape.

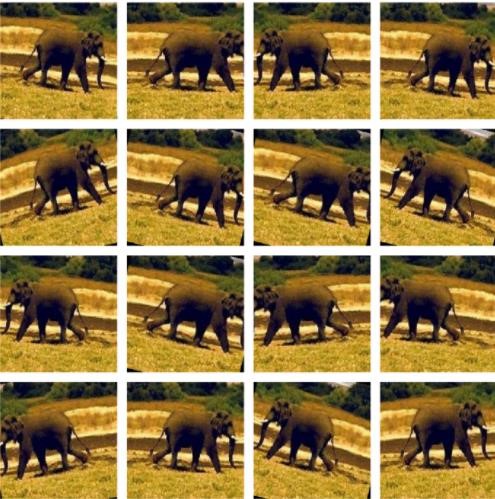
In a nutshell, both are different concepts from two different fields of study (SVM being from Machine Learning and Image Processing a subset if Signal Processing). Image Processing is mostly done prior to SVM.

**3.4 Algorithms**

###### Data Augmentation

The idea of data augmentation is to artificially increase the number of training images our model sees by applying random transformations to the images. For example, we can randomly rotate or crop the images or flip them horizontally. We want our model to distinguish the objects regardless of orientation and data augmentation can also make a model invariant to transformations of the input data.

An elephant is still an elephant no matter which way it’s facing!



**Fig3.4: Image Transformation of training data**

Augmentation is generally only done during training. Each epoch — one iteration through all the training images — a different random transformation is applied to each training image. This means that if we iterate through the data 20 times, our model will see 20 slightly different versions of each image. The overall result should be a model that learns the objects themselves and not how they are presented or artifacts in the image.

3.5 Block diagram

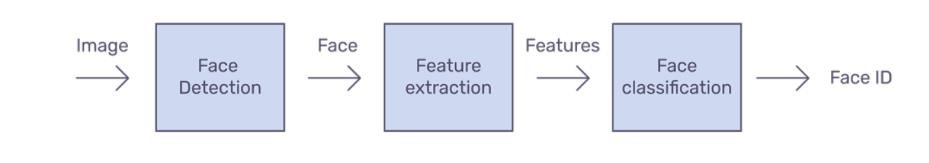
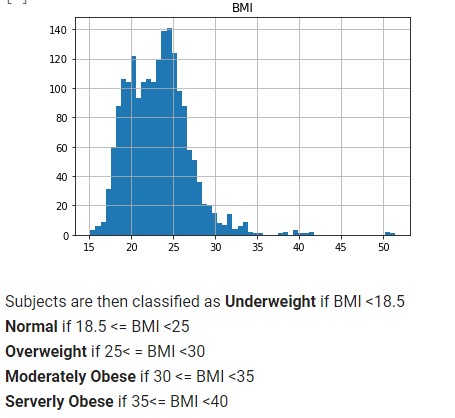


Fig 3.5 Classification

**Face detection** is a computer technology being used in a variety of applications that identifies human faces in digital images. Face detection also refers to the psychological process by which humans locate and attend to faces in a visual scene.

**Feature extraction** is a process of dimensionality reduction by which an initial set of raw data is reduced to more manageable groups for processing. A characteristic of these large data sets is a large number of variables that require a lot of computing resources to process.

3.6 Result Verification



##### Fig3.6:- BMI chart

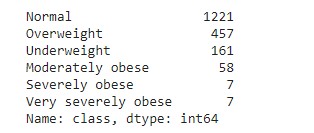


Fig3.7:- predicted values from our dataset

###### 3.7 Result and Discussions

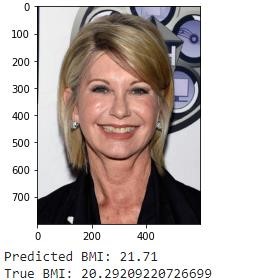
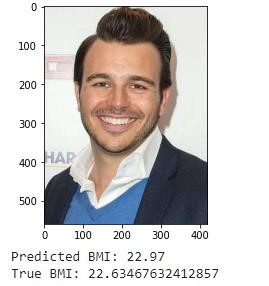
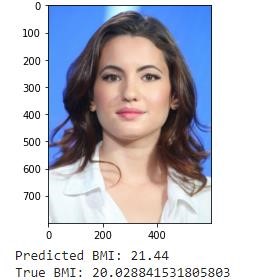
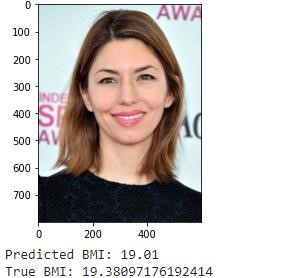


Fig3.8: Experimental outcome

From this we conclude that we are getting a accuracy of 95% .

### Chapter 4

**Conclusions**

**Conclusion**

Chapter 4

#### 4.1 Conclusion

This project presented a novel approach for estimating height, weight and BMI from single-shot facial images, based on regression models. We have developed an automated, computational system for body mass index prediction from face images. We found that MTCNN is a very useful technique to handle facial image problems. While we couldn’t reach out goal of 100% accuracy, we did end up creating a system that can get very close to that goal.

#### 4.2 Future Scope

In future work, we will explore more facial features for BMI prediction, and study age, gender, or ethnicity group-specific features to better characterize the facial appearance for the purpose of BMI prediction.

Also, this model can be improvised to make the BMI prediction more accurate so that it can be implemented in medical diagnostics which will make it easy for the doctors to treat the patients.

**References**

References

1. Guodong Guo “Body Weight Analysis From Human Body Images”IEEE Transactions on

Information Forensics and Security ( Volume: 14, Issue: 10, Oct. 2019)

1. Adolphe Quetelet (1796–1874)—the average man and indices of obesity.
2. Al‐Lawati JA, Jousilahti P. Body mass index, waist circumference and waist‐to‐hip ratio cut‐off points for categorisation of obesity among Omani Arabs. Public Health Nutrition, 2008, 11(1):102108.
3. Keys et al (1972). Indices of Relative Weight and Obesity. Pergamon Press, Britain. J Chron Dis 1972. Vol. 25, pp. 329- 343.
4. Ian Janssen, Peter T Katzmarzyk, and Robert Ross(2004). Waist Circumference and Not Body

Mass Index Explains Obesity Related Health Risk. American Journal of Clinical Nutrition. Retrieved March 4, 2014 cited in Google.com,

1. Shrikant, J. Honade (2013). ―Height, Weight and Body Mass Index Measurement Using

MATLAB.‖ International Journal of Advanced Research in Engineering and Technology, Vol. 4, Issue

5, July – August 2013, pp.35-45. Retrieved August 31, 2013 cited in Google.com

27