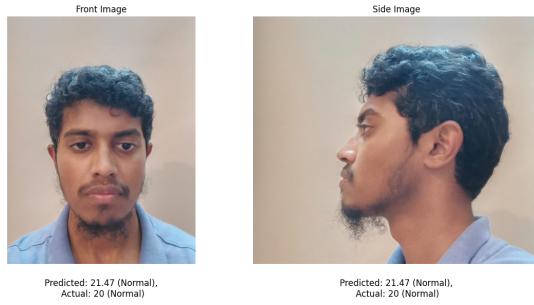# PRML Project Report

Gender Classification and BMI Prediction from facial features

Ajmal   - CS22B2046
Ashiq Irfan   - CS22B2021
Abhishek Chakravarthy   - CS22B2054

Date of Submission - December 5, 2024

The Body Mass Index (BMI) is calculated using the following formula:

$$BMI = \frac{weight(kg)}{height^2(m^2)}$$

Figure 1: Prediction results for Ajmal

1 Data Preprocessing

1.1 Outlier Removal

Outliers in the dataset can distort the results and lead to inaccurate model predictions. To detect and remove outliers, we performed the following steps:

- **BMI Outliers**: We computed the BMI for each individual and identified values that are outside the normal human BMI range (e.g., BMI below 10 or above 50). These extreme values were removed from the dataset.

- **Valid Image**: We inspected the images to identify corrupted files, duplicates, or those which have different number of channels and so on and removed them.

- Resized the image to 128x128

- Used grayscaled image

Added target (1 or 0) for gender Or calculated BMI using above given formula.

2 Feature Extraction Using CNN

To make predictions from images, we first extracted features using cv2 and pre-trained CNN models.

2.1 VGG16

We used VGG16 for feature extraction from both the front and side images. These pre-trained models, commonly used for image classification tasks, were adapted to extract high-level features from the images. we got about 30000 features overall (from 60000).

2.2 FaceNet

we have also used FaceNet which gave 2000 features, 1000 each for front and side images. We concatenated the features extracted from front and side face and added the target. It is then fed into cnn for feature subset extraction.

3   Model Training

We trained a linear regression model to predict BMI, logistic regression for gender. The following steps were followed:

3.1   Data Splitting

The dataset was split into training and testing sets, with 80% of the data used for training and 20% reserved for testing.

```
logistic regression dataset size:  61110      Training:48888 Testing:12222
linear regression dataset size:  59985        Training:47988 Testing:11997
```
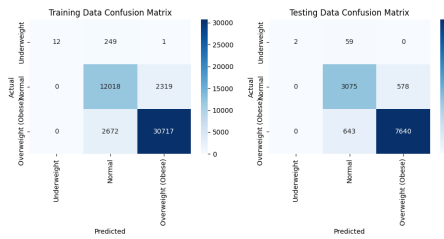
4   Results

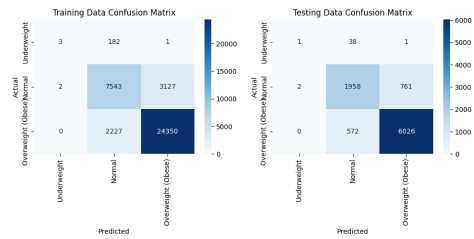Figure 2: Gender prediction using logistic regression

Figure 3: BMI classification prediction, Training: 89.08% Testing:89.33%

Metrics:

Metric	Value
Mean Absolute Error (MAE)	0.0741
Mean Squared Error (MSE)	0.0325
R² Score	0.8700
Pearson Correlation Coefficient	0.9329

Figure 4: Gender Prediction Model Performance Metrics

Metric	Training	Testing
MAE	1.243570	1.249775
MSE	2.776962	2.794754
R²	0.897728	0.895518
Pearson Correlation	0.947485	0.946325

Figure 5: Linear Regression Model Performance Metrics
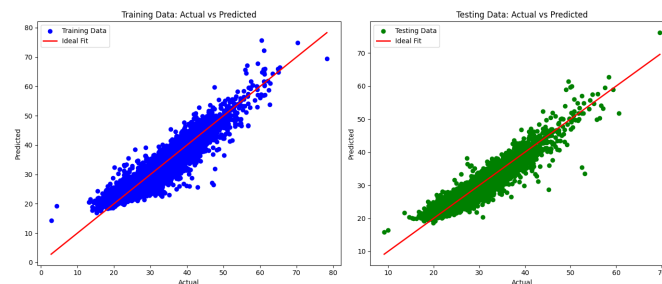
The results of the linear model were as follows:

Figure 6: Linear reg.