BMI Predictor

Use Computer Vision to infer the BMI from face pictures

Machine Learning & Predictive Analytics
Final Project

Submitted By: Manish Kumar
Submission Date: May 22, 2023

1. Introduction

In recent years, the rise of Artificial Intelligence (AI) and Machine Learning (ML) has revolutionized the field of computer vision, enabling breakthroughs in various applications. One area of interest is the analysis of facial features and their potential implications for health assessment.

Facial analysis has garnered attention as a promising avenue for non-invasive health assessment. The human face carries valuable information that can reflect physical characteristics, including body weight and composition. By leveraging AI and ML techniques, facial features extracted from images can be analyzed to predict health-related factors, such as Body Mass Index (BMI), body fat percentage, and potential risk factors for various diseases.

This research project aims to explore the potential of AI in predicting BMI from facial images. Developing an accurate and reliable model offers a convenient and accessible means of weight assessment, with applications in healthcare, fitness monitoring, and personalized recommendations for weight management.

In the following sections, we will delve into the methodology, experiments, and findings of this research project.

2. Dataset Description

The dataset of faces with annotated BMI values was sourced by Kocabey et al[1] from the VisualBMI[2] project. Initial processing steps and dataset information is also sourced from Kocabey et al's work[1].

2.1 Preprocessing

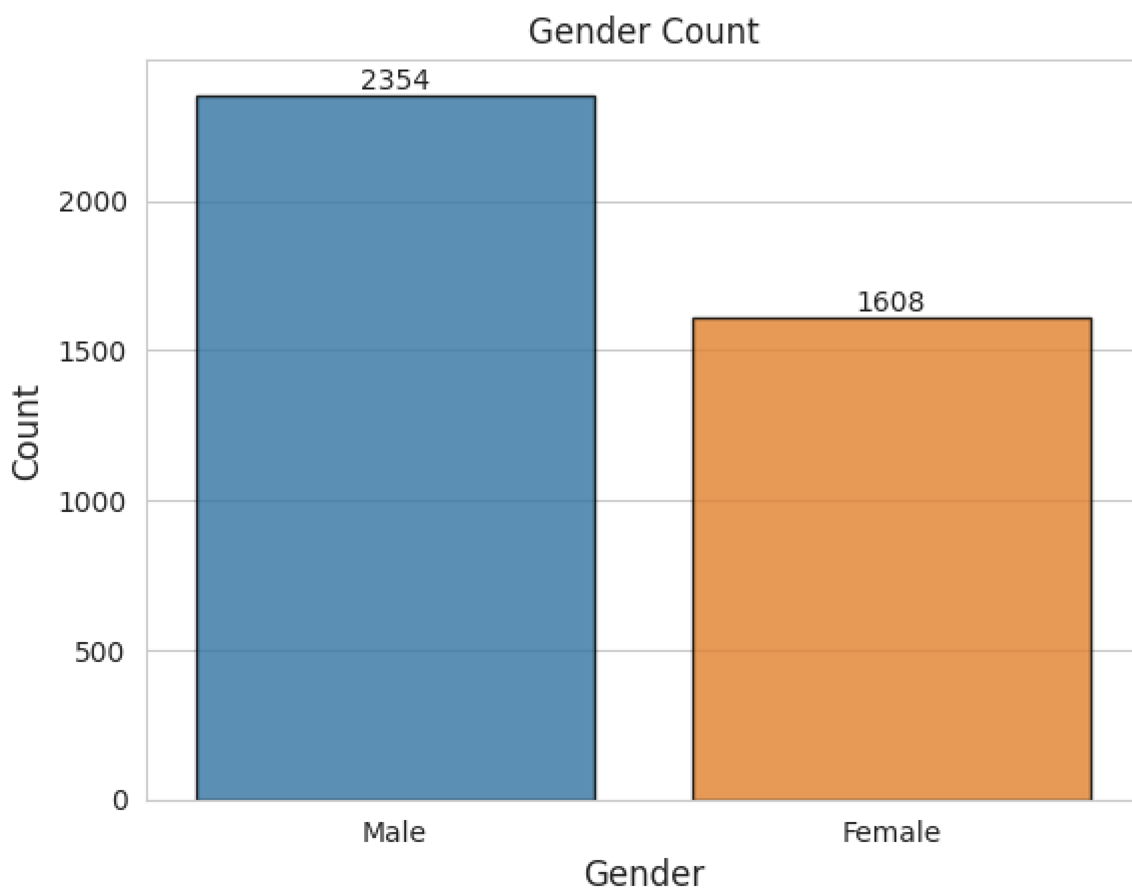
The images in the VisualBMI[2] project were collected from Reddit[3] posts, specifically, from the "progresspics"[4] sub-Reddit where the users post pictures about their body transformations. The VisualBMI[2] dataset originally comprised a total of 16,483 such images containing a pair of "before" and "after" pictures, annotated with gender, height, and previous and current body weights. Kocabey et al[1] manually went through all of the image URLs and cropped the faces. They ignored all the images except the ones with two faces since they only had previous and current body weights.

After the manual cleaning process, the dataset size was reduced to 2103 pairs of faces, with corresponding gender, height, and previous and current body weights. Then for each pair of images, they computed the previous and current BMI. This led to a total of 4206 faces with corresponding gender and BMI information. The dataset available to us had some missing images and we received a dataset of 3962 images.

NOTE: The BMI is defined as (body mass in kg) / (body height in m).

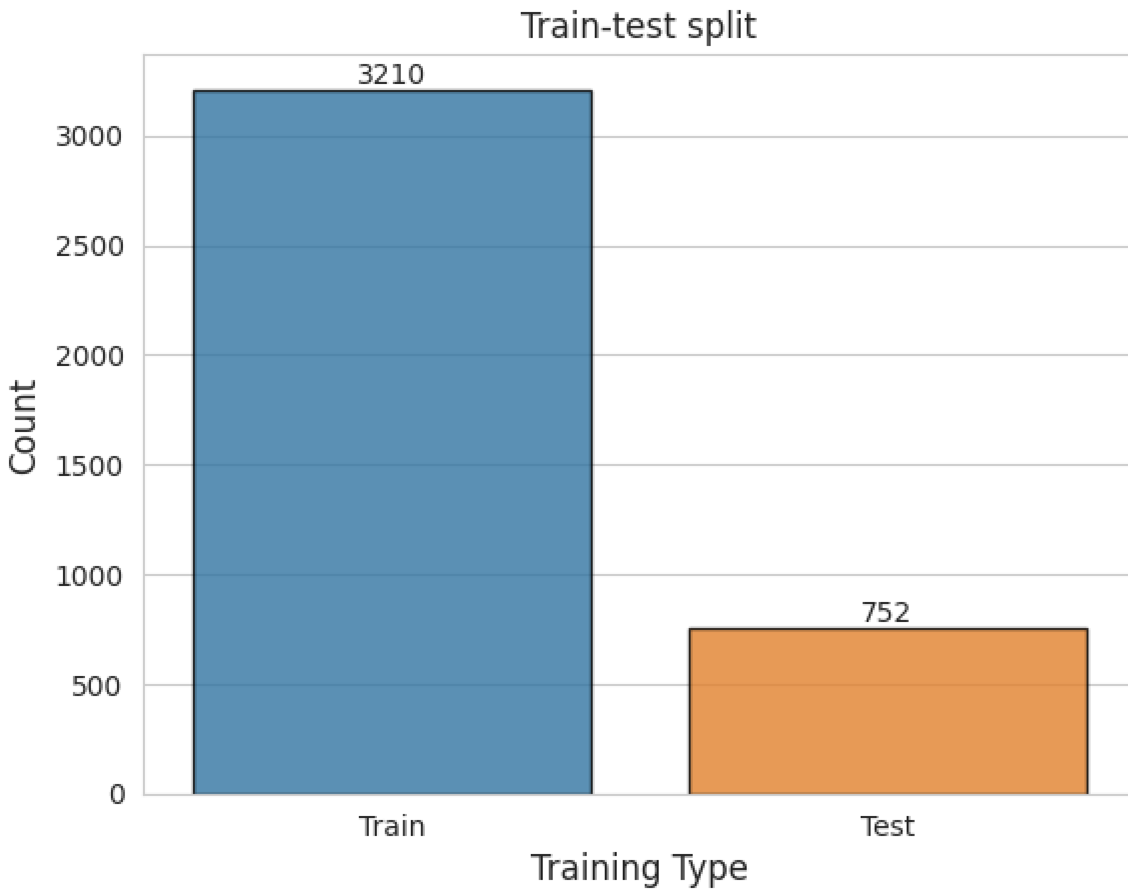
2.2 Exploratory Data Analysis

2.2.1 Gender Analysis



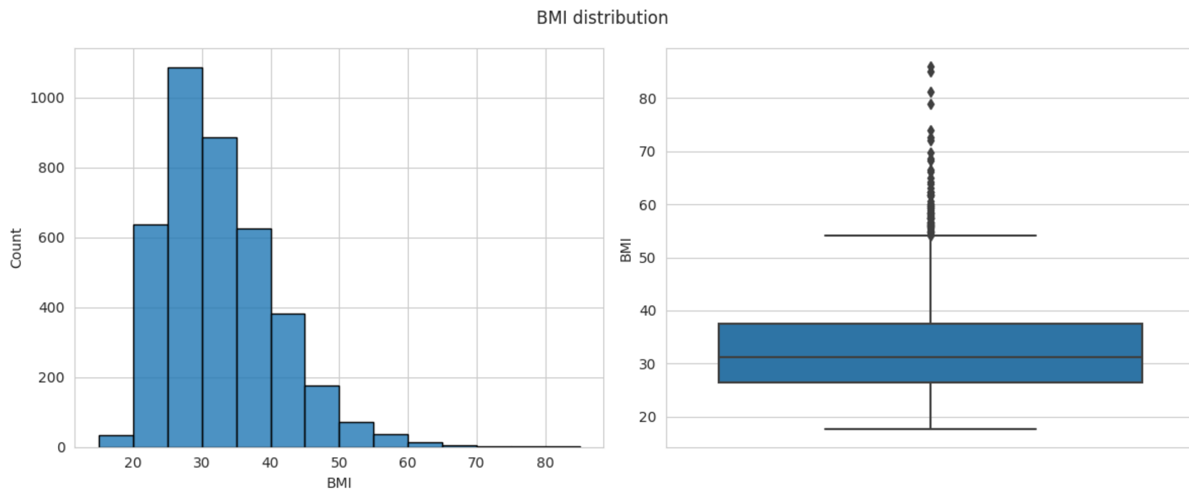
2.2.2 Train-Test Split

Approximately 20% of the dataset was reserved for testing while the remaining 80% was used for training the models. This split was used the same one used by Kocabey et al[1] to make the performance comparable.



2.2.3 BMI Distribution

As shown in the below figure, BMI distribution is right-skewed. Very few data points are available for BMI > 50 or BMI < 20. The mean, median, and standard deviation for BMI are 32.67, 31.17, and 8.27, respectively.

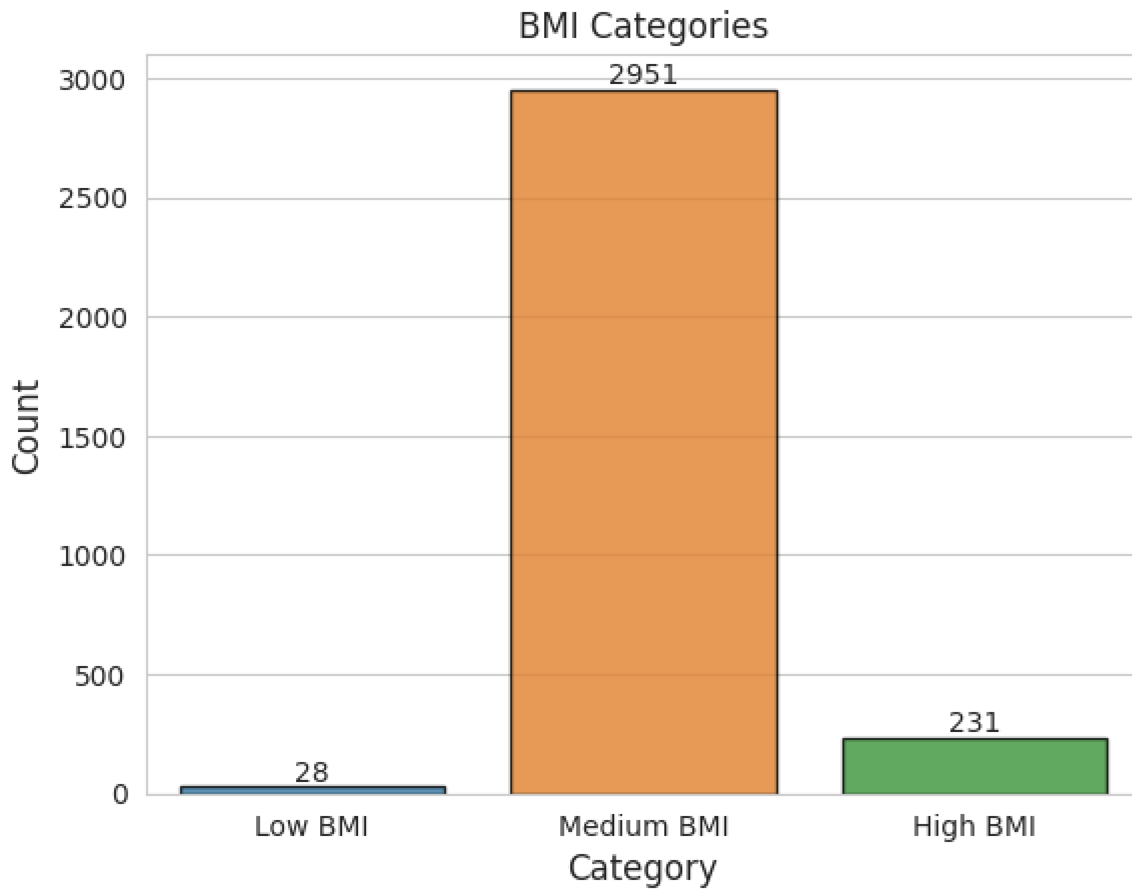


3. Data Augmentation

As we saw in the previous section, the dataset is highly skewed with very few high or low BMI data points. We attempted data augmentation strategies to generate more images for such cases. The dataset was split into three categories:

1. Low BMI (≤ 20)
2. Medium BMI (20 < BMI < 45)
3. High BMI (BMI ≥ 45)

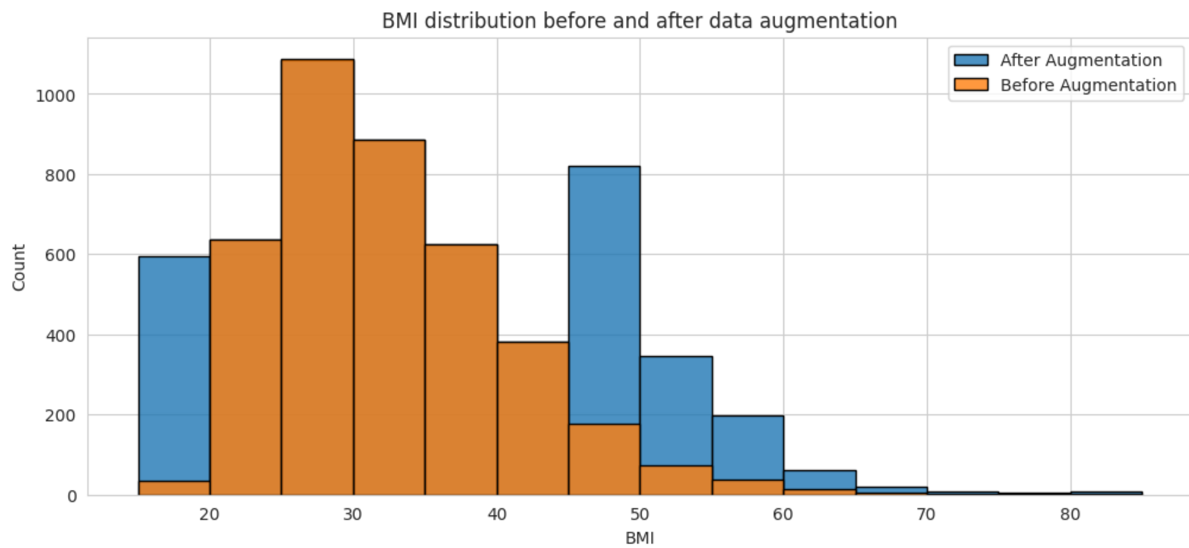
This categorization is not based on a real-world perspective but rather chosen to represent skewness in the dataset. For example, A BMI of 30 would be considered higher in the real world but here marked as a medium since it's fairly well represented by the number of available data points. The below figure shows the distribution among the three classes.



To augment the data, we used the **`imgaug`**[5] library available in Python which applies a series of transformations to a given list of images to generate new images. Here's a list of transformations applied in a random order to each image:

1. Flip the image with a probability of 0.5
2. Crop 0 to 10% of the pixels
3. Apply Gaussian Blur with a probability of 0.5
4. Add image contrast
5. Add Gaussian Noise
6. Multiply all pixels with a random value
7. Apply affine transformations like translation, rotation, scaling, and shearing

Only the images in the training set were augmented. Each of the low BMI images was augmented 20 times, bringing their total to 588. On the other hand, high BMI images were transformed 5 times bringing the total to 1386. Medium BMI images were not augmented since they're already fairly well represented.



After augmentation, the training size was increased from 3210 to 4925. We'll discuss the effect of these augmentations in the result section.

4. Model Architecture

Our dataset is still very small and it is difficult to train deep learning models on this dataset. Transfer learning is one approach that can help in this type of situation. Our BMI prediction system is composed of three stages:

1. Face detection using a pretrained face detection deep-learning model
2. Image feature extraction using a pretrained deep-learning model, and
3. BMI prediction using a regression model

4.1 Face Detection

In our training dataset, images are already cropped to reflect just the faces of people but this might not be the case in the real world. When a user clicks a photo, they might have various backgrounds and other objects in the image which might confuse the model unnecessarily. To deal with this issue, we use the Multi-Task Cascaded Convolutional Neural Networks (MTCNN)[6] to detect and crop the faces. Only these cropped images are passed as inputs to the downstream pipeline.

4.2 Embeddings Extraction

To get the image embeddings, we used the Inception Resnet (V1)[7] model architecture and experimented with two pretrained models trained on

1. VggFace2[8] and
2. CASIA-Webface[9]

4.3 BMI Prediction

Once the embeddings are extracted, the next task is to predict the BMI. For this task, we tried a feed-forward neural network as well as various Machine Learning models such as Random Forest, Support Vector Machines, Linear Regression, and XGBoost.

4.3.1 Hyperparameter Tuning

To get the best possible performance, we used the GridSearch module from the Scikit-Learn library and tried a combination of hyperparameters. A 5-fold cross-validation was used in the GridSearch module.

4.3.2 Evaluation

Root Mean Square Error (RMSE), Mean Absolute Error (MAE), and Pearson Correlation (r) were used to compare the performance across the models.

5. Results

The below table summarizes the performance across all combinations of dataset type, image embeddings, regression models, and hyperparameter tuning. From this table, we see that the combination of the non-augmented dataset for training, InceptionResnetV1 model architecture pretrained on the casia-webface dataset for image feature extraction, and SVM regression model for final BMI prediction gave the least MAE (5.07) and highest Pearson correlation coefficient (0.65). We used this combination in the final deployment app discussed in the next section.

Dataset Type	Embeddings	Regression Model	RMSE	MAE	Pearson R
Original	vggface2	Random Forest	7.34	5.30	0.63
Original	vggface2	XGBoost	7.28	5.25	0.62
Original	vggface2	SVM	7.23	5.15	0.64
Original	vggface2	NN	7.92	5.69	0.55
Original	vggface2	Linear Regression	**7.11**	5.20	0.64
Original	casia-webface	Random Forest	7.49	5.42	0.61
Original	casia-webface	XGBoost	7.41	5.25	0.60
Original	casia-webface	SVM	7.19	**5.07**	**0.65**
Original	casia-webface	NN	8.21	5.94	0.52
Original	casia-webface	Linear Regression	7.24	5.36	0.63
Augmented	vggface2	Random Forest	7.26	5.41	0.62

Augmented	vggface2	XGBoost	7.36	5.42	0.61
Augmented	vggface2	SVM	7.86	5.84	0.55
Augmented	vggface2	NN	8.42	6.35	0.57
Augmented	vggface2	Linear Regression	7.74	5.87	0.61
Augmented	casia-webface	Random Forest	7.32	5.48	0.61
Augmented	casia-webface	XGBoost	7.41	5.52	0.60
Augmented	casia-webface	SVM	7.37	5.43	0.60
Augmented	casia-webface	NN	8.78	7.76	0.56
Augmented	casia-webface	Linear Regression	7.74	5.93	0.60

The parameters used for the SVM models are described in the below table.

Parameter	Value
kernel	rbf
C	5
epsilon	0.2

6. Deployment

To make the model accessible and user-friendly, we developed a web application using the Streamlit[10] library in Python. This application integrated the trained model, allowing users to click a photo and receive instant BMI predictions. By deploying the application on platforms like HuggingFace[11] we aimed to facilitate widespread accessibility and usability. The app is available for the public here:

https://huggingface.co/spaces/kmnis/BMI-Predictor

7. Conclusion and Future Scope

To conclude, in this research project, we explored the feasibility and effectiveness of predicting Body Mass Index (BMI) from facial images using machine learning and deep learning techniques. The objective was to investigate whether facial features extracted from images could serve as reliable predictors of BMI, offering a non-invasive and accessible means of weight assessment. Our trained models gave promising results showing that facial analysis holds promise as a viable approach for BMI prediction.

We saw that augmented datasets are not always helpful in making the right predictions. In this work, we noticed a slight drop in performance after augmenting the dataset. Further, by leveraging the Inception Resnet (V1) model trained on the VGGFace2 datasets, we successfully extracted facial embeddings that captured essential facial characteristics. These embeddings provided valuable information for predicting BMI, showcasing the potential of facial features in reflecting weight-related attributes.

Future work in this field may involve expanding the scope of facial analysis to predict other health-related factors beyond BMI. Exploring the potential of facial features in estimating body fat percentage, metabolic disorders, or disease risk factors could further advance personalized healthcare and well-being.

8. References

1. Kocabey, E., Camurcu, M., Ofli, F., Aytar, Y., Marin, J., Torralba, A., & Weber, I. (2017). Face-to-BMI: Using Computer Vision to Infer Body Mass Index on Social Media. Proceedings of the International AAAI Conference on Web and Social Media, 11(1), 572-575. https://doi.org/10.1609/icwsm.v11i1.14923.
2. http://www.visualbmi.com (website not active anymore)
3. https://www.reddit.com
4. https://www.reddit.com/r/progresspics
5. https://imgaug.readthedocs.io
6. J. Xiang and G. Zhu, "Joint Face Detection and Facial Expression Recognition with MTCNN," 2017 4th International Conference on Information Science and Control Engineering (ICISCE), Changsha, China, 2017, pp. 424-427, doi: https://doi.org/10.1109/ICISCE.2017.95.
7. https://github.com/timesler/facenet-pytorch
8. https://www.robots.ox.ac.uk/~vgg/data/vgg_face2
9. Dong Yi, Zhen Lei, Shengcai Liao, Stan Z. Li. Learning Face Representation from Scratch. Computer Vision and Pattern Recognition. arXiv:1411.7923. https://doi.org/10.48550/arXiv.1411.7923.
10. https://streamlit.io
11. https://huggingface.co