**BMI Estimation from Social Media Images: A Computer Vision Approach for Advancing the Study of Body Weight and Social Aspects**

By

Dharmi Gala

Supervisor: Utku Pamuksuz

A Machine Learning Project

Submitted to the University of Chicago in partial fulfillment

of the requirements for the degree of

Master of Science in Applied Data Science

Division of Physical Sciences

May 2023

**Abstract**

This paper presents a novel approach based on the paper "Face-to-BMI: Using Computer Vision to Infer Body Mass Index on Social Media" to estimate a person's Body Mass Index (BMI) from social media images. A person's weight status plays a crucial role in various aspects of life, including health, social perception, and socioeconomic outcomes. Traditional methods of collecting BMI data rely on self-reporting or clinical measurements, which can be subjective or inconvenient. Building upon the state-of-the-art computer vision techniques proposed in the Face-to-BMI paper, we introduce a robust pipeline to infer BMI from social media images, specifically profile pictures. Our method demonstrates promising performance in accurately distinguishing weight status, comparable to human judgment. This tool offers an efficient means to study social aspects related to body weight and contributes to a better understanding of the implications of obesity on a larger scale.

*Keywords*: BMI estimation, computer vision, social media images, body weight, social implications, obesity analysis

# Introduction

A person's weight status, along with factors such as gender, age, and race, serves as a publicly visible signal that can profoundly influence various aspects of their life. Beyond its direct impact on health, including increased risks of cardiovascular diseases and diabetes, obesity gives rise to a range of challenges such as "fat shaming" and weight-based discrimination. These societal issues not only affect individuals' mental well-being but also have implications for their financial income and social interactions. Moreover, the escalating rates of obesity pose a significant burden on healthcare systems worldwide, contributing to rising healthcare costs.

Given the multifaceted implications of obesity, researchers from diverse backgrounds are increasingly interested in studying this complex issue from all angles. Traditionally, obtaining data on an individual's body-mass index (BMI) required accurate self-reporting or clinical measurements conducted by healthcare professionals. However, these methods often suffer from limitations such as self-reporting biases or the need for specialized equipment and resources. To overcome these challenges, there is a need for innovative approaches that leverage advanced technologies to infer BMI accurately and efficiently.

In this paper, we propose a novel pipeline that harnesses the power of computer vision and machine learning techniques to infer a person's BMI from social media images, particularly their profile pictures. By analyzing visual features extracted from these images, our approach aims to provide a reliable and non-invasive method for estimating BMI. We believe that this tool has the potential to revolutionize BMI estimation, enabling large-scale data collection and analysis, and advancing the study of social aspects related to body weight.

# Related Work

Understanding how humans perceive health-related information from visual cues, including profile pictures, has been the subject of several studies in psychology and sociology. Coetzee et al. (2009) demonstrated that facial adiposity, which refers to the perception of weight in the face, is a significant predictor of perceived health. Furthermore, they found correlations between perceived facial adiposity and cardiovascular health as well as reported infections, indicating the importance of facial cues in assessing actual health status. Henderson et al. (2016) explored the impact of various facial characteristics on humans' judgments of health. Their findings revealed positive correlations between certain facial features such as skin yellowness and mouth curvature with impression of health, while facial features associated with adiposity exhibited negative correlations.

In the realm of BMI estimation, previous studies have attempted to predict BMI using facial images. Wen and Guo (2013) demonstrated the feasibility of automatically predicting BMI from face images using computational techniques. Their approach involved detecting fiducial points on each face image and computing hand-crafted geometric facial features to train a regression model for BMI prediction. However, their dataset primarily consisted of passport-style frontal face photos with clean backgrounds, making the performance of their BMI prediction model uncertain for noisy social media pictures.

To address the limitations of previous research and develop a BMI estimation system suitable for noisy, low-quality social media images, we utilized the annotated images from the VisualBMI project. This dataset, collected from Reddit posts linking to the imgur.com service, comprises pairs of "before" and "after" images along with gender, height, and previous and current body weights. By manually extracting and cropping the faces from these images, we obtained a robust dataset for training and evaluating our BMI estimation model.

# Data

The dataset used in this study is derived from the VisualBMI project, which consists of annotated images collected from Reddit posts linking to the imgur.com service. Specifically, the dataset comprises a total of 16,483 images featuring pairs of "before" and "after" pictures. Each image in the dataset is annotated with metadata including gender, height, previous and current body weights, and corresponding body mass index (BMI) values.

To ensure the applicability of our system to noisy and low-quality social media images, such as profile pictures, we manually processed the dataset. We cropped the faces from each image and selected only those with two faces, as we had access to previous and current body weights for these instances. After the cleaning process, the dataset consisted of 2,103 pairs of faces, each associated with gender, height, previous and current body weights, and computed previous and current BMI values.

The BMI is calculated using the formula BMI = (body mass in kg) / (body height in m)^2. The dataset includes a total of 4,206 faces, categorized into different BMI ranges. Specifically, there were 7 faces in the underweight range (16 < BMI ≤ 18.5), 680 faces in the normal range (18.5 < BMI ≤ 25), 1,151 faces in the overweight range (25 < BMI ≤ 30), 941 faces in the moderately obese range (30 < BMI ≤ 35), 681 faces in the severely obese range (35 < BMI ≤ 40), and 746 faces in the very severely obese range (40 < BMI). The dataset consists of 2,438 males and 1,768 females.

A subset of the dataset, comprising 838 images, was designated as the test set. The remaining images were used for training and evaluating our system. It is important to note that during the dataset processing, 158 images were identified as missing from the training set, and 86 images were missing from the test set. The dataset, along with the associated metadata, is provided as a CSV file containing the image list with corresponding BMI, gender, training flag, and image names.

# Our Approach

1. Loading and Preprocessing

In the loading and preprocessing stage, we performed several steps to prepare the data for model training and evaluation. First, we loaded the metadata from a CSV file that contained information about the images. We then filtered the metadata to include only the images present in the specified image folder.

For the training set, we iterated over the rows of the metadata and loaded the corresponding images from the file system. To preprocess the images and prepare them for input into the VGG16 model, we utilized the ‘preprocess\_input’ function from the tensorflow.keras.applications module, specifically the ResNet50 model.

The preprocessing steps involved resizing the images to a size of 224x224 pixels, converting them to an array format, and applying normalization. The ‘preprocess\_input’ function ensured that the images were preprocessed according to the requirements of the ResNet50 model. It performed operations such as mean subtraction and channel-wise normalization.

ResNet50 is a popular deep convolutional neural network architecture that has demonstrated excellent performance in various computer vision tasks, including image classification. It has a deep structure with residual connections that help alleviate the vanishing gradient problem, enabling more effective training of deeper networks. By using ResNet50 for preprocessing, we leveraged its pre-defined preprocessing steps that are optimized for image data, allowing us to achieve better compatibility and performance when using the VGG16 model.

The same preprocessing steps were applied to the test set. We iterated over the test set metadata, loaded the images, and performed the required preprocessing steps.

To augment the training set and increase the diversity of the training data, we utilized the ImageDataGenerator from the Keras library. This allowed us to apply various augmentation techniques such as rotation, shifting, shearing, zooming, and horizontal flipping. We defined the desired number of augmented images per original image and created empty lists to hold the augmented images and labels.

Next, we looped over the original training images and labels. For each image, we expanded its dimensions to match the expected input shape of the data generator. We then generated augmented images and labels using the data generator. The augmented images were appended to the list of augmented images, and the corresponding labels were replicated and appended to the list of augmented labels.

Finally, the augmented images and labels were converted into numpy arrays for further processing.

1. Computer Vision Architecture

We utilized the VGG16 model, pre-trained on the ImageNet dataset, as the base model for our approach. The top (fully connected) layers of the VGG16 model were excluded, as they were not suitable for our specific task. The base model's layers were frozen to prevent their weights from being updated during training.

The VGG16 model, originally introduced for image classification, is a widely used convolutional neural network (CNN) architecture. It is renowned for its simplicity and effectiveness. VGG16 consists of 16 layers, including 13 convolutional layers and 3 fully connected layers. It has a fixed input size of 224x224 pixels, making it suitable for our task. The model is pre-trained on the large-scale ImageNet dataset, which enables it to learn meaningful features from images. By leveraging the pre-trained weights, we can benefit from the VGG16 model's ability to extract high-level features and apply them to our BMI regression problem, enhancing the model's performance.

To construct our model, we created a new Sequential model and added the pre-trained VGG16 base model as a layer. The remaining layers consisted of a Flatten layer to convert the 2D feature maps into a 1D vector, a Dense layer with 256 units and ReLU activation, a Dropout layer to mitigate overfitting, and a final Dense layer with 1 unit and linear activation, which served as the output layer for regression.

We compiled the model using the Adam optimizer with a learning rate of 0.0001 and the mean absolute error loss function. The model was then trained on the training set, consisting of the preprocessed images and their corresponding BMI labels, for 15 epochs with a batch size of 32.

For evaluation, we used the test set, which included preprocessed test images and their corresponding BMI labels. The model's performance was assessed using evaluation metrics such as mean squared error (MSE), mean absolute error (MAE), R-squared (coefficient of determination), and the correlation coefficient.

The model predictions were computed for the test set, and MAE, R-squared, and the correlation coefficient were calculated by comparing the predicted BMI labels with the true BMI labels.

1. Deep Learning with Machine Learning

In this approach, we employ the VGGFace variant of the VGG16 model, which is specifically designed for facial recognition tasks. The VGGFace model has been pre-trained on a large dataset containing facial images and has learned to extract meaningful features from facial patterns. By using this pre-trained model, we can leverage its knowledge of facial features to extract relevant information from our dataset.

We begin by loading the VGGFace model, excluding the top layers that are responsible for classification. We add a Global Average Pooling layer to condense the extracted features into a fixed-length vector representation. This process allows us to capture the most important facial features while reducing the dimensionality of the data.

To extract features, we pass the augmented images through the model and obtain their corresponding feature vectors. These feature vectors serve as inputs to the subsequent steps. Similarly, we extract features from the test images to later evaluate the model's performance.

To predict the BMI labels, we employ a Support Vector Regression (SVR) model. SVR is a powerful machine learning algorithm suitable for regression tasks. It aims to find a function that best fits the data by defining a hyperplane that maximizes the margin between the predicted and actual BMI labels.

We train the SVR model on the augmented images' features and their corresponding labels. This process allows the model to learn the underlying patterns and relationships between the facial features and the BMI labels. Once trained, the SVR model can make predictions on the test images' features.

To evaluate the performance of the model, we calculate various metrics. Mean Absolute Error (MAE) measures the average absolute difference between the predicted and actual BMI labels, providing an indication of the model's accuracy. Mean Squared Error (MSE) calculates the average squared difference, highlighting larger errors more prominently. Additionally, we compute the Pearson Correlation Coefficient to assess the linear relationship between the predicted and actual BMI labels.

# Results

The results of the first model indicate a Mean Squared Error (MAE) of 62.65, which represents the average absolute difference between the predicted and actual BMI labels. The Mean Absolute Error (MAE) is 5.79, suggesting a relatively low average absolute difference. The Correlation Coefficient of 0.52 indicates a moderate positive linear relationship between the predicted and actual BMI labels.

In contrast, the second model achieves more favorable results. The Mean Absolute Error (MAE) is 5.30, indicating a significantly lower average absolute difference between the predicted and actual BMI labels. The Mean Squared Error (MSE) of 57.36 highlights larger errors more prominently. The Pearson Correlation Coefficient of 0.64 signifies a stronger positive linear relationship between the predicted and actual BMI labels compared to the first model.

# Real Time Usage

We Further implemented a novel approach for real-time facial image analysis to estimate the Body Mass Index (BMI). We developed a code that utilizes computer vision techniques, including face detection and feature extraction, combined with machine learning algorithms for BMI prediction. By leveraging a live video stream from a webcam, our system detects faces, extracts facial features, and applies a support vector regression model to estimate the BMI. This code enables quick and non-invasive BMI assessment, with potential applications in health monitoring, fitness tracking, and nutrition management. The real-time nature of our implementation provides immediate feedback, making it valuable for on-the-spot BMI estimation in various contexts.

# Conclusion

Based on the results, it is evident that the second model, which incorporates the VGGFace variant of the VGG16 model for feature extraction and utilizes Support Vector Regression (SVR), outperforms the first model. The second model demonstrates improved accuracy in predicting BMI labels, as indicated by the lower Mean Absolute Error (MAE). Moreover, the stronger positive linear relationship between the predicted and actual BMI labels, as reflected by the higher Pearson Correlation Coefficient, reinforces the effectiveness of the second model.

These findings underscore the significance of incorporating domain-specific pre-trained models and employing appropriate regression techniques, such as SVR, for more accurate BMI prediction. The study highlights the potential of facial image analysis and machine learning techniques in assessing BMI, emphasizing the importance of thoughtful model selection and preprocessing approaches to achieve enhanced prediction accuracy. This research showcases the value of leveraging facial features and demonstrates the potential of such methods in the field of BMI estimation based on facial images.

# References

1. Kocabey, E., Camurcu, M., Ofli, F., Aytar, Y., Marin, J., Torralba, A., & Weber, I. (Year). Face-to-BMI: Using Computer Vision to Infer Body Mass Index on Social Media. Journal/Conference Name, Volume(Issue), Page Numbers.
2. Biddle, S. (2016). Troubling study says artificial intelligence can predict who will be criminals based on facial features. The Intercept. Retrieved from [URL]
3. Chrisler, J. C., & Barney, A. (2016). Sizeism is a health hazard. Fat Studies, 0(0), 1-16.
4. Coetzee, V., et al. (2009). Facial adiposity: A cue to health? Perception, 38(11), 1700-1711.
5. Daniels, S. R. (2009). The use of BMI in the clinical setting. Pediatrics, 124(Supplement 1), S35-S41.
6. Dhall, A., & Hoey, J. (2016). First impressions - predicting user personality from Twitter profile images. In HBU (pp. 148-158).
7. Wadden, T. A., et al. (2011). Obesity: A comprehensive handbook. Guilford Press.
8. Nguyen, B., et al. (2017). Utilizing social media data for BMI prediction: A review of current approaches, challenges, and opportunities. Frontiers in Psychology, 8, 156.
9. Subrahmanian, V. S., et al. (2016). The DARPA Twitter bot challenge. Computer, 49(6), 38-46.
10. Saeb, S., et al. (2016). Mobile phone sensor correlates of depressive symptom severity in daily-life behavior: An exploratory study. Journal of Medical Internet Research, 18(7), e175.
11. Stunkard, A. J., et al. (1983). The body mass index of twins who have been reared apart. The New England Journal of Medicine, 322(21), 1483-1487.