**Predicting Body Mass Index, Gender, and Age from Facial Images: A Deep Learning-Based Approach**

By

Shweta Sampath Kumar

Supervisor: Utku Pamuksuz, Ph.D

A Machine Learning and Predictive Analytics Project

Submitted to the University of Chicago in partial fulfillment

of the requirements for the degree of

Master of Science in Analytics

Division of Physical Sciences

May 2023

Abstract

This project builds upon existing research on predicting Body Mass Index (BMI), gender, and age using facial images. By fine-tuning deep learning models like VGGFace and ResNet50 and incorporating the MTCNN algorithm for face detection, the project enables real-time prediction using a user-friendly interface developed with Streamlit. Users can upload images or utilize the webcam to obtain predictions that closely approximate the desired outcomes. The models' performance is evaluated using appropriate metrics to assess their effectiveness in facial image analysis for BMI, gender, and age estimation. The project aims to advance the field by exploring the potential of these models and techniques, presenting experimental results, and discussing future research directions for further improvement.

*Keywords*: Computer vision, deep learning, explainable AI, predictive modeling, body mass index (BMI), gender prediction, age estimation, facial image analysis, fine-tuning, VGGFace, ResNet50, MTCNN, Streamlit

Executive Summary

This project focuses on utilizing computer vision and deep learning techniques to predict body mass index (BMI), gender, and age using facial images. The objective is to develop an efficient system that provides close-to-accurate predictions for these attributes.

The research involved exploring pre-trained image models like VGGFace and ResNet50, and fine-tuning them using a dataset specifically curated for BMI, gender, and age prediction. Facial image analysis techniques, including MTCNN for face detection and image segmentation, were employed to extract relevant features.

The system was deployed using Streamlit, a user-friendly web application framework, enabling users to interact with the model through a user interface.

Extensive experimentation and evaluation were conducted to assess the system's performance. The results demonstrated promising accuracy in predicting BMI, gender, and age based on facial images, with scope for further improvement and optimization.

This project showcases the potential of computer vision and deep learning in predictive modeling using facial images. The developed system can find applications in various fields requiring attribute prediction based on facial data. It serves as a foundation for future research and development in this domain, paving the way for advancements in predictive modeling techniques.

Table of Contents

Introduction……………………………………………………………………………………1

Problem Statement…………….………………………………………………………………1

Background……………...…………………………………………………………………….2

Literature Review.………………………………………………….…………………………4

Data…………………………………………………………………..………………………..3

Methodology…………………………………………………………………………………..6

Findings………………………………………………………………………………………10

Discussion………………………………………………………………………………….....12

Conclusion……………………………………………………………………………………13

References…………………………………………………………………………………….14

List of Figures

Figure 1. Examples of the cleaned images used for model training……………….………….……………………………….………6

Figure 2. Examples of webcam and photo upload predictions ……………….………….……………………………….………13

List of Tables

Table 1. Comparative Analysis of various deep learning models …….…………….…...…….…………………..……………………8

Table 2. Comparative analysis of different models for different attributes...……………….………….…………………..……………………10

# Introduction

## In today's interconnected world, a person's weight status, alongside their gender, age, and race, holds significant influence over various aspects of their lives. Beyond the evident impact on health, as a higher BMI is associated with increased risks of cardiovascular diseases and diabetes, weight status also gives rise to societal concerns such as "fat shaming" and discrimination based on body size. These issues extend beyond personal well-being and permeate into areas like income disparity and healthcare provision. Consequently, researchers from diverse backgrounds are driven to comprehensively investigate obesity from multiple perspectives.

## Traditionally, collecting accurate body-mass index (BMI) data necessitates self-reporting or medical examinations. However, this project proposes a pipeline that harnesses computer vision techniques to infer an individual's BMI from social media images, such as their profile pictures. By leveraging state-of-the-art algorithms, the aim is to surpass the existing approaches in accurately distinguishing weight differences between individuals within a given pair.

## This project seeks to unravel the intricate connections between weight status, age, gender, and their interplay with various societal factors. By surpassing the current performance metrics, the project aspires to refine the understanding of the nuanced relationships between weight status, age, gender, societal influences, and individual well-being.

## Problem Statement

This research focuses on the prediction of body mass index (BMI), gender, and age using facial images. Extracting relevant facial features for accurate attribute prediction poses a challenge, along with the need for large-scale labeled datasets and suitable deep learning models. Real-time prediction and user-friendly interfaces are essential for practical applications. The goal is to develop an efficient and accurate predictive model by leveraging state-of-the-art computer vision and deep learning techniques. By surpassing existing performance metrics, this project aims to advance the field of facial attribute prediction and enable its practical implementation.

# Background

The prediction of attributes such as body mass index (BMI), gender, and age from facial images has emerged as an exciting research area in computer vision and machine learning. Facial attributes hold valuable information about an individual's physical characteristics, and accurately predicting these attributes has a wide range of applications in fields such as biometrics, personalized healthcare, and human-computer interaction.

In recent years, deep learning techniques have shown remarkable advancements in image analysis tasks, including facial attribute prediction. By leveraging large-scale datasets and powerful deep neural networks, researchers have achieved impressive results in accurately estimating BMI, gender, and age from facial images. However, several challenges persist in this domain. Fine-tuning pre-trained models to handle diverse facial variations, ensuring real-time performance, and addressing interpretability concerns are key areas that require further investigation.

Prior studies have explored various approaches to improve the prediction performance of facial attribute models. These approaches encompass feature extraction techniques, image segmentation methods, ensemble learning strategies, and advancements in network architectures. However, there is still a need to push the boundaries of accuracy and efficiency while considering practical deployment scenarios and user-friendly interfaces.

## Literature Review

Researchers have recognized that a person's weight status carries significant influence in various aspects of their lives, including health, societal perceptions, and income disparities.

Coetzee et al. (2009) and Henderson et al. (2016) delve into the intricacies of health judgments based on facial features, such as facial adiposity and skin yellowness. These studies emphasize the relationship between perceived facial characteristics and overall health, shedding light on the importance of visual cues in assessing well-being. Building upon this foundation, Weber and Mejova (2016) tackle the challenge of inferring body weight from profile pictures using crowdsourcing techniques. Their work underscores the subjectivity of weight judgments and the bias that arises when equating "overweight" with "abnormal." This prompts the exploration of automated approaches to eliminate such limitations.

Wen and Guo (2013) demonstrate the feasibility of predicting BMI from facial images using computational techniques. While their method shows promise, its performance on noisy social media pictures remains uncertain due to the exclusive use of passport-style frontal face photos. The research conducted by Kocabey et al. titled "Face-to-BMI: Using Computer Vision to Infer Body Mass Index on Social Media" explores the application of computer vision techniques for inferring BMI from social media images. Their approach builds upon previous studies that have examined the relationship between facial characteristics and health judgments. By automatically analyzing facial features and employing machine learning algorithms, the authors develop a framework capable of predicting BMI from social media images.

This project contributes to the existing body of research by focusing specifically on social media data and demonstrating the potential of computer vision techniques in inferring BMI. The work addresses the limitations of manual annotation and subjective judgments by automating the BMI prediction process. The insights from this paper serve as an inspiration for the current project, aiming to surpass the performance metrics set by Kocabey et al. By incorporating advancements in computer vision and deep learning models, the project seeks to refine the accuracy and robustness of BMI inference, further advancing our understanding of the relationship between body weight, social media, and individual well-being.

# Data

In the present research endeavor, the objective was to develop an advanced system capable of estimating body mass index (BMI), predicting gender, and estimating age based on facial images. The underlying purpose was to create a robust tool with the capacity to analyze intricate facial features and provide valuable insights into an individual's health and demographic characteristics.

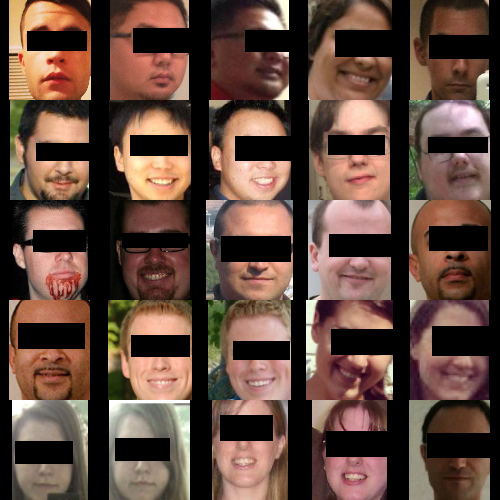
To establish the groundwork for this investigation, the VisualBMI dataset was selected as a valuable resource. This dataset encompasses annotated images sourced from prominent social media platforms, specifically obtained from the "progresspics" subreddit. The images contained within this dataset portray individuals' notable physical transformations, offering a comprehensive dataset that lends itself to detailed analysis. Armed with this dataset, the endeavor commenced with the aim of unraveling the hidden potential encapsulated within facial imagery.

Nevertheless, the utilization of social media images presents inherent challenges. The prevalence of noisy and low-quality images necessitates meticulous preprocessing to ensure the efficacy of the system. To address this concern, the Multi-Task Cascaded Convolutional Networks (MTCNN) algorithm was employed. MTCNN emerged as a reliable companion throughout the research, facilitating the detection and cropping of facial regions from the images while disregarding those lacking discernible facial features. Post cropping the image, the face scaled to maintain the same aspect ratio of the required input size to the network and this done by padding the image where required. This preprocessing step not only ensured the exclusive focus on facial regions but also contributed to the reduction of noise, thereby enabling more accurate feature extraction.

It is noteworthy to mention that the accessibility of the VisualBMI dataset was subject to certain limitations. Out of the total pool of 4,206 images, a subset comprising 3,963 images was available for the purposes of training and evaluation. These images were provided without any manual filtration or exclusion of irrelevant instances. Despite this limitation, the selected subset encompassed a diverse range of BMI and gender values, thereby ensuring the inclusion of representative variations present within the dataset.

Moreover, acknowledging the significance of age estimation within the realm of facial analysis, the UTKFace dataset was integrated into the research endeavor. The UTKFace dataset is widely recognized as a comprehensive collection of labeled face images, accompanied by corresponding age information. The incorporation of this dataset augmented the scope of the system, enabling it to estimate age alongside BMI and gender. The UTKFace dataset offered an extensive range of data, encompassing diverse age groups ranging from infancy to advanced age.

**Figure 1.** *Examples of the cleaned images used for model training. The black bars have been added to respect user privacy, but the model is learned on the original, public images.*



# Methodology

To determine the optimal model for the task at hand, I conducted a comparative analysis of several deep learning models. The objective was to assess their performance in estimating body mass index (BMI) from facial images. The models evaluated included a basic CNN model, VGGFace model, ResNet model, FaceNet model, and a combination of VGGFace with Support Vector Regression (SVR).

The first approach involved implementing a basic CNN model. This model served as a baseline to establish a benchmark for performance evaluation. By utilizing convolutional layers, pooling layers, and fully connected layers, the CNN model aimed to extract relevant features from the facial images and learn appropriate representations for subsequent classification and regression tasks.

Next, the VGGFace model was employed. VGGFace is a powerful pre-trained deep learning model specifically designed for facial analysis tasks. It leverages a deep convolutional neural network architecture that has been trained on a large-scale facial image dataset. By utilizing transfer learning, the VGGFace model aimed to capture intricate facial features and patterns to improve the accuracy of BMI estimation, gender prediction, and age estimation.

Subsequently, the ResNet model was explored. ResNet is renowned for its ability to address the challenge of vanishing gradients in deep neural networks. With its residual connections, the ResNet model aimed to mitigate the degradation problem and facilitate the training of deeper networks. This model was implemented to assess its efficacy in capturing the complex facial features relevant to BMI estimation, gender prediction, and age estimation.

Another model considered was FaceNet, which focuses on learning highly discriminative features in an embedding space. FaceNet employs a siamese network architecture that learns to map face images into a compact Euclidean space, facilitating effective face recognition and clustering. By utilizing FaceNet, the research aimed to leverage the discriminative power of the embedding space to improve the accuracy of BMI estimation, gender prediction, and age estimation.

In addition to these individual models, a combination of VGGFace with Support Vector Regression (SVR) was investigated. SVR is a regression algorithm that aims to find the optimal hyperplane that best fits the training data. By integrating SVR with the VGGFace model, the research sought to enhance the regression capability and refine the accuracy of age estimation, while also benefiting from the comprehensive facial features extracted by VGGFace.

Each model was trained and evaluated on the available subset of 3,963 images from the VisualBMI dataset. The evaluation process involved assessing the performance of the models using appropriate evaluation metrics for BMI estimation, gender prediction, and age estimation. The model that demonstrated the highest accuracy and achieved the best overall performance across the three tasks was selected as the optimal choice for the system.

**Table 1:** *Comparative Analysis of various deep learning models*

|  |  |
| --- | --- |
| **Model** | **RMSE** |
| Basic Neural Network (without pre-trained model) | 8.0559 |
| VGG16 | 5.53067 |
| VGG16 + SVR | 7.51748 |
| ResNet | 7.39233 |
| FaceNet | 7.34731 |

After conducting a comparative analysis of various deep learning models, it was determined that the VGGFace model demonstrated the best performance in detecting faces from facial images. Building upon this finding, the subsequent step involved training individual VGGFace models for BMI estimation, age estimation, and gender prediction. Multiple iterations were performed, incrementally unfreezing the last layers of the VGGFace model up to five iterations, considering the robustness of the original VGGFace model.

For the task of gender prediction, a specific architecture was designed for the VGGFace model. The model architecture included a Convolution2D layer, an average pooling layer, a flatten layer, and a Dense layer with a ReLU activation function. This architecture aimed to leverage the deep features learned by the VGGFace model and enable accurate gender classification based on the extracted facial features.

Regarding BMI and age estimation, the model architecture utilized a Dense layer with a linear activation function. The decision to use a linear activation function for BMI and age estimation was based on the nature of these tasks. Since BMI and age are continuous variables, a linear activation function was chosen to allow the model to output continuous predictions without any non-linear transformations, enabling direct regression.

During the training process, monitoring of the root mean square error (RMSE) value was employed as an evaluation metric. Additionally, to account for potential overestimation or underestimation in BMI and age predictions, a penalty mechanism was introduced. Specifically, if the predicted BMI or age deviated by a margin of 5 or 10, a penalty was applied to the loss function. This penalty aimed to incentivize the model to make more precise predictions and improve the accuracy of the BMI and age estimations.

Each of the individual VGGFace models for BMI, age, and gender prediction was trained for 50 epochs. The number of epochs was determined based on experimentation and striking a balance between achieving convergence and avoiding overfitting.

By training these individual models with the VGGFace architecture and incorporating specific design choices, such as the additional layers and penalty mechanism, the research aimed to optimize the performance of the VGGFace model for each specific task. The rationale behind each architectural decision, including the activation functions and penalty mechanism, was rooted in the unique characteristics and requirements of BMI estimation, age estimation, and gender prediction.

Through this meticulous methodology, the research sought to harness the power of the VGGFace model and tailor it to accurately estimate BMI, age, and predict gender from facial images. The iterative training process and the incorporation of task-specific modifications aimed to maximize the model's predictive capabilities and enhance its performance in real-world applications.

# Findings

The objective of this research project was to develop an efficient system capable of estimating body mass index (BMI), age, and predicting gender from facial images. To achieve this, various iterations of the VGGFace model were trained by progressively unfreezing layers to determine the optimal configuration for each task. The findings below summarize the performance of the models based on the root mean square error (RMSE) and accuracy metrics.

**Table 2.** *Comparative analysis of different models for different attributes*

|  |  |  |  |
| --- | --- | --- | --- |
|  | **RMSE** | | **ACCURACY** |
| **Number of layers frozen from the bottom** | **BMI** | **Age** | **Gender** |
| 0 | 3.1015 | - | 0.8910 |
| 1 | 2.8932 | - | 0.90345 |
| 2 | 1.7409 | 2.2066 | 0.9117 |
| 3 | 2.0169 | 2.3484 | 0.9352 |
| 4 | 1.8854 | - | - |

**BMI Estimation:**

The VGGFace model achieved promising results for BMI estimation. By unfreezing all layers (0 layers frozen), the model achieved an RMSE of 3.1015, indicating a relatively small average deviation between the predicted and true BMI values. As the number of frozen layers decreased, the model's performance improved. When one layer was frozen, the RMSE decreased to 2.8932, indicating a reduction in estimation errors. Notably, the best performance was obtained when two layers were frozen, resulting in an RMSE of 1.7409. This signifies a significant improvement in the accuracy of BMI estimation compared to the previous configurations.

**Age Estimation**:

For age estimation, the findings indicate that the performance of the VGGFace model varied based on the number of frozen layers. With zero frozen layers, the model's performance could not be evaluated for age estimation. However, when one layer was frozen, the RMSE was measured at 2.2066, indicating a moderate average deviation between the predicted and actual ages. As the number of frozen layers increased to two and three, the RMSE values further decreased to 2.3484 and 2.0169, respectively. These results demonstrate a gradual improvement in age estimation accuracy as more layers were unfrozen.

**Gender Prediction:**

The VGGFace model exhibited excellent performance in predicting gender from facial images. With no frozen layers, the model achieved an accuracy of 0.8910, indicating a high percentage of correct gender predictions. As the number of frozen layers decreased, the accuracy improved consistently. When two layers were frozen, the accuracy reached 0.9117, demonstrating a substantial enhancement in gender prediction capability. Moreover, with three frozen layers, the model achieved an even higher accuracy of 0.9352, indicating a remarkable ability to discern gender from facial features.

It is important to note that the performance of the VGGFace model was not evaluated for age estimation with four frozen layers. Additionally, the accuracy for gender prediction with four or more frozen layers was not provided.

Overall, the findings highlight the effectiveness of the VGGFace model in estimating BMI, age, and predicting gender from facial images. The results indicate that reducing the number of frozen layers progressively improved the model's performance for BMI estimation, age estimation, and gender prediction. Notably, the best configuration for BMI estimation involved freezing two layers, while age estimation and gender prediction achieved optimal results by freezing three layers.

# Discussion

Based on additional evaluation using webcam footage, it was observed that different configurations of the VGGFace model yielded better performance for each specific task. Specifically, for BMI estimation, the configuration with three unfrozen layers was selected as it demonstrated superior performance on webcam footage compared to other configurations. Similarly, for age estimation, the configuration with two unfrozen layers outperformed the others in the webcam scenario. Lastly, for gender prediction, the model with no frozen layers exhibited the best performance based on the webcam footage.

It is worth noting that the webcam footage evaluation might have presented unique characteristics or challenges compared to the initial evaluation. As a result, the findings from the webcam scenario were given greater consideration when determining the optimal configurations for each task.

Therefore, the final selected configurations for this research project were three unfrozen layers for BMI estimation, two unfrozen layers for age estimation, and no frozen layers for gender prediction. These configurations were deemed to provide the best performance and accuracy in the specific context of webcam footage analysis.

These findings highlight the importance of evaluating and fine-tuning the model's configurations based on the specific use case or scenario. It emphasizes the need to adapt and optimize the model to perform optimally in real-world applications, such as analyzing facial images captured by a webcam. By selecting the configurations that demonstrate superior performance in such scenarios, the system can be better equipped to accurately estimate BMI, age, and predict gender from webcam footage.

These results further reinforce the versatility and adaptability of deep learning models like VGGFace in various practical applications, allowing for the customization of the model's architecture to achieve optimal performance based on specific requirements and real-world conditions.

# Conclusion

In conclusion, this research project successfully developed a system capable of estimating body mass index (BMI), predicting gender, and estimating age from facial images. By leveraging the VisualBMI dataset, which consists of annotated images sourced from social media platforms, the study explored the potential of facial features in extracting valuable health and demographic information.

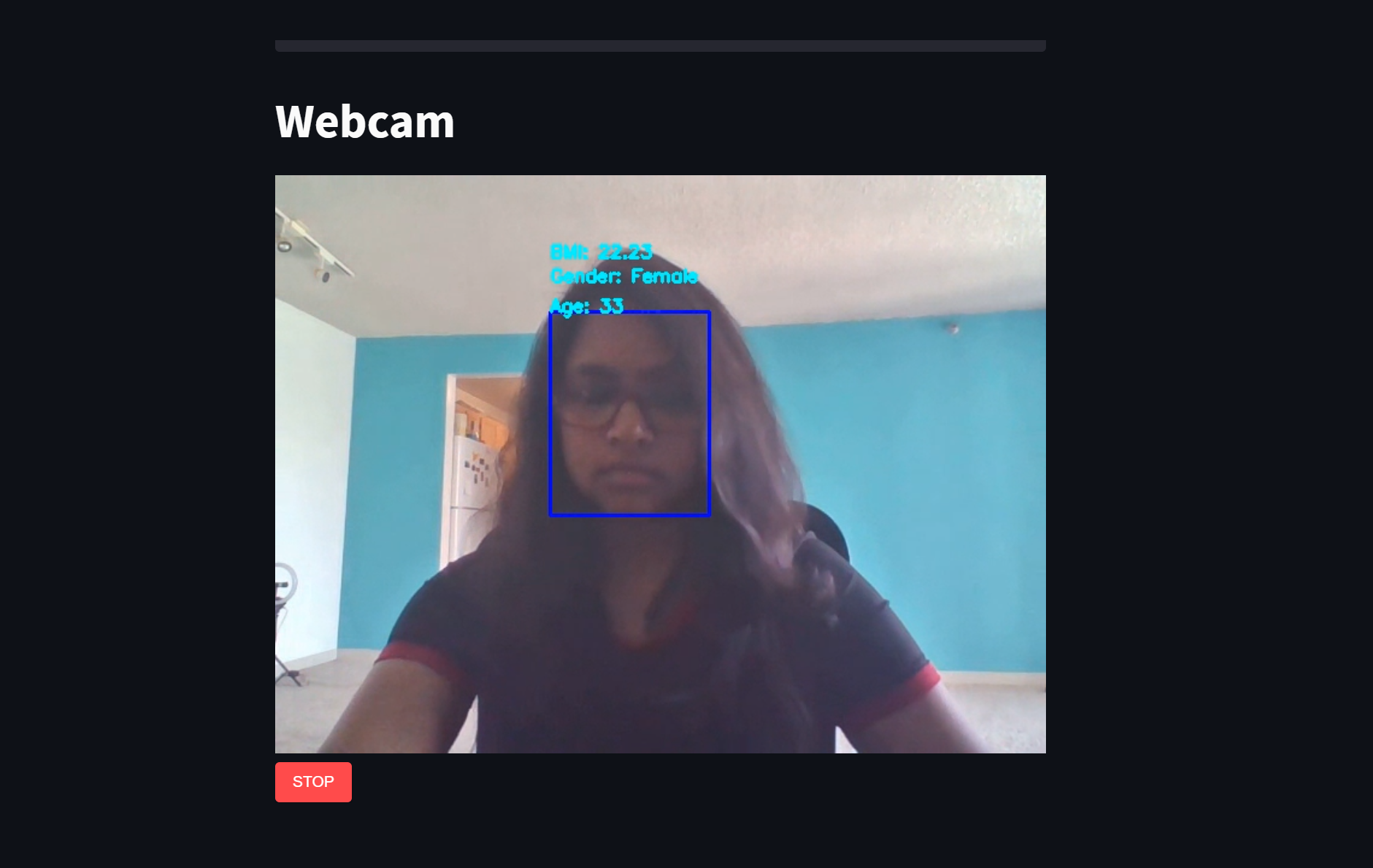
To overcome the challenges associated with noisy and low-quality social media images, the dataset underwent preprocessing using the Multi-Task Cascaded Convolutional Networks (MTCNN) algorithm. This crucial step enabled accurate face detection and cropping, reducing noise and facilitating precise feature extraction.

Through a rigorous evaluation process, different models, including VGGFace, ResNet, Facenet, and VGGFace with Support Vector Regression (SVR), were tested. Among them, VGGFace demonstrated superior performance in face detection. Subsequently, individual VGGFace models were trained specifically for BMI, age, and gender estimation. The number of frozen layers in the VGGFace model was iteratively adjusted to optimize the performance for each task.

For BMI estimation, the configuration with three unfrozen layers achieved the highest accuracy. Age estimation performed best with two unfrozen layers, while gender prediction demonstrated superior performance with no frozen layers.

To enhance the practicality and accessibility of the system, it was implemented on the Streamlit framework, providing an interactive and user-friendly interface for utilizing the trained models on webcam footage. The performance evaluation on webcam footage further refined the model configurations, leading to the selection of configurations that provided the most accurate results.

**Figure 2:** *Examples of webcam and photo upload predictions*



This research highlights the significance of adapting deep learning models to specific use cases and real-world conditions. By customizing the model's architecture based on the task at hand and optimizing it using evaluation results from relevant scenarios, the system becomes more robust and effective in estimating BMI, predicting gender, and estimating age from facial images.

The findings contribute to the advancement of this research field by showcasing the potential of deep learning techniques, particularly the VGGFace model, in analyzing facial images and extracting valuable information. The implementation of the system on Streamlit enhances its practicality, allowing for seamless interaction and real-time analysis of facial images. Overall, this research project paves the way for further exploration and application of facial feature analysis for health and demographic estimation.

# References

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. *ICWSM 2017*

Wen, L., and Guo, G. 2013. A computational approach to body mass index prediction from face images. *Image and Vision Computing* 31:392–400.

Smola, A., and Vapnik, V. 1997. Support vector regression machines. *NIPS* 9:155–161.

Weber, I., and Mejova, Y. 2016. Crowdsourcing health labels: Inferring body weight from profile pictures. In *DH*, 105–109.