Digital Scale: Open-Source On-Device BMI Estimation from Smartphone Camera Images Trained on a Large-Scale Real-World Dataset

Frederik Rajiv Manichand[1,2,4], Robin Deuber[1,2], Robert Jakob[1,2], Steve Swerling[4], Jamie Rosen[4], Elgar Fleisch[1,2,3], Patrick Langer[1,2]

[1]Centre for Digital Health Interventions, ETH Zurich
[2]Agentic Systems Lab, ETH Zurich
[3]Centre for Digital Health Interventions, University of St. Gallen
[4]WayBetter

rmanichand@ethz.ch, rdeuber@ethz.ch, rjakob@ethz.ch, steve@waybetter.app, jamie@waybetter.app, efleisch@ethz.ch, planger@ethz.ch

Abstract

Estimating Body Mass Index (BMI) from camera images with machine learning models enables rapid weight assessment when traditional methods are unavailable or impractical, such as in telehealth or emergency scenarios. Existing computer vision approaches have been limited to datasets of up to 14,500 images. In this study, we present a deep learning-based BMI estimation method trained on our WayBED dataset, a large proprietary collection of 84,963 smartphone images from 25,353 individuals. We introduce an automatic filtering method that uses posture clustering and person detection to curate the dataset by removing low-quality images, such as those with atypical postures or incomplete views. This process retained 71,322 high-quality images suitable for training. We achieve a Mean Absolute Percentage Error (MAPE) of 7.9% on our hold-out test set (WayBED data) using full-body images, the lowest value in the published literature to the best of our knowledge. Further, we achieve a MAPE of 13% on the completely unseen (during training) VisualBodyToBMI dataset, comparable with state-of-the-art approaches trained on it, demonstrating robust generalization. Lastly, we fine-tune our model on VisualBodyToBMI and achieve a MAPE of 8.56%, the lowest reported value on this dataset so far. We deploy the full pipeline, including image filtering and BMI estimation, on Android devices using the CLAID framework. We release our complete code for model training, filtering, and the CLAID package for mobile deployment as open-source contributions.

Code — https://github.com/im-ethz/DigitalScale

1 Introduction

A high Body Mass Index (BMI) is associated with an increased risk of chronic conditions, including cardiovascular disease, certain cancers, and metabolic disorders such as diabetes (Larsson and Burgess 2021). Body weight and its fluctuations serve as important indicators of health status and disease risk. For example, in individuals with overweight, a 5% reduction in body weight can substantially lower the risk of type 2 diabetes and cardiovascular disease (Horn, Almandoz, and Look 2022). BMI is calculated using weight and height, typically measured with a scale and stadiometer. However, in situations where such equipment is unavailable or impractical, such as remote settings or emergency care, BMI estimation from smartphone camera images offers a digital, real-time alternative. By analyzing images of individuals, deep learning models can estimate BMI without requiring active user input. This may enable integration into clinical and telehealth services, supporting passive BMI tracking, population-level monitoring, early risk screening, and behavioral health interventions. Such methods may also aid weight-based drug dosing when patients are unresponsive, disoriented, or unable to report their weight, addressing a well-documented need for rapid weight estimation in emergency care (Wells et al. 2023; Wells, Goldstein, and Bentley 2017).

Previous studies have investigated using deep learning models for estimating body weight from images, yielding promising results. Jin et al. (2022) reported state-of-the-art performance, achieving a Mean Absolute Percentage Error (MAPE) of 9.37% for BMI estimation from full-body images. Other studies have focused on weight estimation using facial images alone (Mirabet-Herranz, Mallat, and Dugelay 2023; Dantcheva, Bremond, and Bilinski 2018), reporting a Mean Absolute Error (MAE) of 2.3.

However, most existing work is limited by relatively small datasets—typically comprising no more than 5,900 full-body images (Jiang and Guo 2019)—and often focuses on a single image perspective per dataset, hindering comparisons across different viewpoints. To the best of our knowledge, no prior study has systematically evaluated BMI estimation from full-body, torso, and facial images using data collected from the same individuals. Furthermore, although previous research has demonstrated the feasibility of image-based BMI estimation, none has presented a fully deployable solution, such as models optimized for smartphone execution or code supporting on-device deployment.

We address these limitations with WayBED (WayBetter BMI Estimation Dataset), our large-scale, real-world dataset comprising over 84,000 full-body images from 25,353 individuals, collected as part of the WayBetter weight-loss program. Our objective is to develop a robust on-device BMI estimation system—one that generalizes well to unseen individuals, can be fine-tuned to new data distributions, and operates efficiently on mobile devices.

To achieve this, we first identify the most suitable image perspective using WayBED by comparing three distinct viewpoints: full-body, torso-up, and face-only. Based on this comparison, we select the best-performing model and evaluate its generalization on the unseen VisualBodyToBMI (Jiang and Guo 2019) dataset. We further fine-tune the model on VisualBodyToBMI to assess its adaptability to new domains, specifically its performance in out-of-distribution scenarios. The final model is deployed in an Android application developed using the CLAID framework, demonstrating its feasibility for real-world use. All code for dataset preparation, model training, mobile deployment, and integration with CLAID is released as open-source contribution (see Table 1 for an overview of available resources).

Languages	Kotlin, Python
Dataset	WayBED dataset, available on request
ML Models	PyTorch, ExecuTorch models, trained on public dataset, or on WayBED on request
License	Apache 2.0
Artifacts	Code for training and deployment, Android App, CLAID package
Sources	https://github.com/im-ethz/DigitalScale

Table 1: Overview of available resources.

The remainder of this work is structured as follows: Section 2 reviews related work on BMI and weight estimation from images. Section 3 outlines our methodology, including dataset curation, model development and deployment. Section 4 presents experimental results and Section 5 discusses the findings and outlines limitations.

2 Related Work

Prior research has investigated a range of approaches for estimating body weight and BMI using machine learning models applied to images captured from various perspectives. Early work primarily focused on facial images (Wen and Guo 2013; Tai and Lin 2015), while more recent studies have extended to full-body images (Altinigne, Thanou, and Achanta 2020; Jiang and Guo 2019; Jin et al. 2022; Kim, Lee, and Lee 2023; Debnath et al. 2024) and images captured via smartphone cameras (Majmudar et al. 2022b; Farina et al. 2022). The literature has employed various target variables, including BMI (Pantanowitz et al. 2021; Jin et al. 2022; Kim, Lee, and Lee 2023; Debnath et al. 2024), body weight (Majmudar et al. 2022b; Wen and Guo 2013; Mirabet-Herranz, Mallat, and Dugelay 2023), and body fat metrics (Farina et al. 2022; Majmudar et al. 2022b). Table 2 summarizes selected studies on image-based body weight estimation, including the datasets used and reported estimation performance.

Datasets Prior research has employed annotated images from a variety of datasets with differing sizes and characteristics. For full-body images, dataset sizes range from around 150 subjects (Majmudar et al. 2022b; Farina et al. 2022; Pantanowitz et al. 2021) to nearly 6000 subjects (Jiang and Guo 2019). For facial images, dataset sizes range from 50 individuals (Tai and Lin 2015) to 14,500 images (Wen and Guo 2013). Various authors have proposed public benchmark datasets for body weight and BMI estimation, such as the Visual-Body-To-BMI dataset (Jiang and Guo 2019) and the Celeb-FBI dataset (Debnath et al. 2024), both containing full-body images sourced from social media and celebrities, respectively. The VIP-attribute dataset (Dantcheva, Bremond, and Bilinski 2018) proposes a benchmark dataset for facial images, containing images of 1026 celebrities.

Existing machine learning approaches Early approaches for BMI estimation from images often relied on manually engineered features to extract relevant information prior to model training. Examples include the use of body keypoints (Jiang and Guo 2019; Altinigne, Thanou, and Achanta 2020), facial landmarks (Tai and Lin 2015), and silhouette masks capturing body contours (Pantanowitz et al. 2021). These features served as inputs to conventional machine learning models such as Support Vector Machines (Wen and Guo 2013; Jiang and Guo 2019) and random forests (Tai and Lin 2015). More recent work has adopted deep learning models to learn features directly from images (Jin et al. 2022; Dantcheva, Bremond, and Bilinski 2018; Mirabet-Herranz, Mallat, and Dugelay 2023). A variety of Convolutional Neural Network (CNN) architectures have been explored, including ResNet (Mirabet-Herranz, Mallat, and Dugelay 2023; Dantcheva, Bremond, and Bilinski 2018) and VGG (Debnath et al. 2024; Pantanowitz et al. 2021). Jin et al. (2022) evaluated multiple CNN architectures on the Visual-Body-To-BMI dataset (Jiang and Guo 2019), identifying a DenseNet architecture that outperformed ResNet and VGG models of comparable size. Their work also introduced an attention mechanism to focus on the most informative regions of the image, further enhancing model performance.

Full-body and face-only perspectives Prior work has investigated various perspectives of images for BMI estimation. The primary distinction has been made between full-body images and facial images. Using full-body images from the VisualBodyToBMI dataset (Jiang and Guo 2019), Jin et al. (2022) achieved an MAE of 3.03. Facial images have also been employed for BMI prediction, achieving an MAE of 2.3 on the VIP-attribute dataset (Dantcheva, Bremond, and Bilinski 2018). In addition, regarding full-body images, work by Kim, Lee, and Lee (2023) explored the effect of combining three different perspectives (front, side, and back) in a single model for BMI estimation, achieving an MAE of 3.12 on their own dataset.

3 Methodology

In this section, we present our approach for estimating BMI from real-world smartphone camera images. We first describe the curation of the WayBED dataset, introducing our procedure for excluding unsuitable images from the real-world dataset (see Section 3.1). Next, we describe our computer vision model for BMI estimation in Section 3.2. Fi-

Paper	Dataset	Estimation Performance[1]
Jiang and Guo (2019)	'VisualBodyToBMI': 5900 images, 2950 subjects	MAE: 3.76 (BMI); MAPE: 12.50% (BMI)
Altinigne, Thanou, and Achanta (2020)	4,400 images sourced from social media	MAE: 9.80 (kg)
Pantanowitz et al. (2021)	161 individuals, photos as silhouettes	MAE: 1.20 (BMI)
Majmudar et al. (2022b)	134 photos of individuals taken by smartphone	MAE: 2.16 (body fat%)
Jin et al. (2022)	'VisualBodyToBMI' + 4190 social-media images	MAE: 3.03 (BMI) MAPE: 9.37% (BMI)
Kim, Lee, and Lee (2023)	1,000 individuals (front, side, back)	MAE: 3.12 (BMI)
Debnath et al. (2024)	Front-facing images of 7,211 celebrities	Accuracy 85.60% on 5-kg weight buckets
Wen and Guo (2013)	Morph-II: 14,500 (Ricanek and Tesafaye 2006)	MAE: 3.14 (BMI)
Tai and Lin (2015)	50 individuals in three weight classes	Not reported
Dantcheva, Bremond, and Bilinski (2018)	VIP-attribute dataset: 1026 images of celebrities	MAE: 2.3 (BMI)
Mirabet-Herranz, Mallat, and Dugelay (2023)	VIP-attribute dataset + 400 images of prisoners	MAE: 7.54 (kg)

Table 2: Related work and commonly used datasets.

Related work in the upper part is on full-body images, work in the lower part on facial-only images.
[1]When multiple performances are reported, the best overall value on the dataset is shown.

nally, we describe our experiments and evaluation strategy in Section 3.3.

3.1 Real-world dataset curation: WayBetter BMI Estimation Dataset (WayBED)

The dataset was collected as part of the WayBetter weight loss program, in which participants were instructed to capture front-facing, full-body photographs at multiple stages of their fitness journey, primarily for progress tracking purposes. A subset of all participants consented for their data to be used in further research, as we describe in Section 4.1.

Quality assurance Along with each photo, users were asked to report their current weight, while height was self-reported at the beginning of the program. For each body photo submitted, participants were also required to upload a photo of their scale displaying their weight, accompanied by a randomly assigned word written on paper, used as a timestamping mechanism to verify the timing and authenticity of submissions ("weigh-in words"). Each submission was reviewed by a team of human referees, who verified the authenticity and adherence to photographic guidelines, accepting or rejecting images as necessary. To maintain data integrity and deter cheating in the program, WayBetter implemented a system for evaluating both the submitted images and associated weigh-in-words along with a defined set of acceptance criteria for images. Further details of the dataset are provided in Section 4.1.

Automatic filtering steps Images captured in real-life scenarios often exhibit substantial variation in factors such as image quality, lighting conditions, and subject posture. Some images - particularly those taken from uncommon angles (e.g., selfies) or featuring atypical postures - may be unsuitable for accurate BMI estimation. To ensure the quality and consistency of our dataset, we implement an automated filtering procedure to curate the collected images. This process aims to exclude images that do not (clearly) depict a human body, those captured from extreme camera angles (e.g., excessively high or low viewpoints), and images with outlier postures.

For the filtering process, we apply a series of steps based on human bounding box and body keypoint estimations, both obtained using the Detectron2 framework (Wu et al. 2019). Detectron2 provides models for human bounding box estimation, including confidence scores, as well as skeleton estimation comprising 17 body keypoints. Using the estimated bounding boxes and keypoints, we implement three filtering steps:

1. Person detection: Images with a human bounding box confidence score below a threshold are removed.
2. Person-to-background ratio: Images with a bounding-box-to-image-area ratio below a predefined threshold are excluded, filtering out cases where the person occupies too little of the image (see Section 4.1).
3. Posture clustering: Images with outlier body keypoints, identified through clustering, are excluded.

Each filtering step is applied independently, and the images that meet one or more exclusion criteria are removed.

Posture clustering For each image, we extract COCO-17 skeleton keypoints (Lin et al. 2014) and normalize them by the image's width and height to account for variations in scale and position. This results in 34 normalized values per image (17 keypoints with x and y coordinates).

Since not all keypoints contribute equally to assessing pose quality, we apply Principal Component Analysis (PCA) to reduce redundancy and emphasize the most informative patterns. We retain the principal components that preserve 95% of the total variance, thereby reducing noise while maintaining relevant pose information. Subsequently, we perform K-means clustering on the PCA-transformed data, testing values of k from 1 to 10, and selecting the optimal k using the elbow method. This clustering groups similar poses and enables the identification of undesirable ones—such as distorted limb positions or elevated camera angles (e.g., selfies). Finally, members of WayBetter visually inspected each image cluster to identify the types of images it represented, retaining clusters with suitable poses and discarding those deemed unsuitable, e.g., selfies.

3.2 Mobile-capable computer vision model for BMI estimation

We adopt the architecture proposed by Jin et al. (2022) for BMI estimation from 2D body images. The model is built upon a DenseNet CNN (Huang et al. 2018). Jin et al. (2022) demonstrated that this architecture yields superior performance compared to other architectures in BMI prediction tasks. While their implementation used a DenseNet-121 backbone, we utilize DenseNet-201 to leverage its increased capacity for learning complex features from larger datasets. An overview of the architecture is shown in Figure 1.

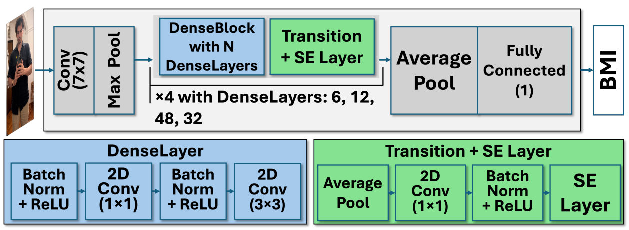

Figure 1: DenseNet-201 architecture with Squeeze-and-Excitation blocks adapted from Jin et al. (2022).

Attention mechanism Following Jin et al. (2022), the architecture incorporates Squeeze-and-Excitation (SE) blocks (Hu et al. 2019) after each transition layer. These blocks implement channel-wise attention by first compressing spatial information through global average pooling and then applying two 1×1 convolutions, separated by a bottleneck layer, to compute channel weights. The resulting weights are used to rescale the feature maps, allowing the model to prioritize more informative channels. For additional details, we refer to Jin et al. (2022).

Mobile deployment To enable BMI estimation in real-world settings, we deploy the full pipeline, from image capture to filtering and inference, directly on mobile devices. To compose and manage this pipeline on-device, we use the CLAID framework (Langer et al. 2024), which facilitates the loading, configuration, and combination of modular processing components. We employ existing CLAID modules for image capture and preprocessing, and introduce three custom modules for subsequent image analysis:

1. PersonFilter: Filters bounding boxes to ensure a sufficient person detection confidence and an adequate person-to-background ratio.
2. PoseFilter: Removes outlier poses based on body keypoints.
3. BMIEstimator: Applies the trained BMI estimation model locally on filtered images.

The BMIEstimator executes our BMI estimation model on-device. We reimplemented the model proposed by Jin et al. (2022) in PyTorch (Ansel et al. 2024) and converted it for mobile inference using ExecuTorch. PersonFilter and PoseFilter use EfficientDet Lite4 and MoveNet Thunder models, respectively, based on TensorFlow Lite.

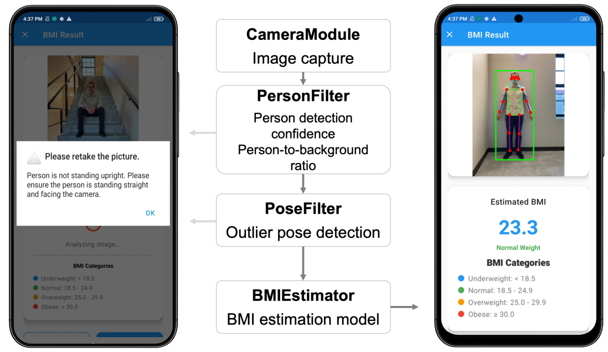

Figure 2: BMI estimation mobile App.

To facilitate reuse in custom applications, we packaged our model and inference code as a CLAID module (Langer et al. 2024). Figure 2 illustrates our BMI estimation app, showcasing an example result.

3.3 Evaluation

To derive a robust machine learning model for BMI estimation suitable for real-world deployment, we follow a multi-step evaluation procedure. We first assess the effectiveness of different image perspectives (full-body, torso-up, face-only) on the WayBED dataset to identify which yields the most accurate predictions. Next, we then evaluate the generalization ability of the best-performing model and perspective on the unseen Visual-Body-To-BMI dataset, which includes new individuals. Since real-world deployment often entails distribution shifts, we further investigate whether our pretrained model—trained on the large WayBED dataset—can be effectively finetuned on a small subset of Visual-Body-To-BMI for domain adaptation.

Performance with different image perspectives on WayBED dataset To identify the most suitable perspective for BMI estimation among those commonly used in the literature, i.e., full-body, torso-up, and face-only, we extract all three crops from the original full-body image for each individual. We divide the dataset into training, validation, and test sets in a 70/15/15 ratio. To ensure subject-disjoint splits, we group images by individuals. The users are randomly shuffled and then sequentially assigned to the split (training, validation, or test) that currently has the largest gap between its target and actual number of images. Importantly, the individuals per splits are kept identical across all perspectives to ensure a fair comparison using the same set of test subjects. We generate face-only crops using the RetinaFace library (Serengil and Ozpinar 2020). For torso-up crops, we use keypoint annotations obtained during the image filtering process (see Section 3.1), specifically shoulder and eye coordinates. The crop width is defined as the horizontal distance between the shoulders, and the height spans vertically from the lowest shoulder to the highest eye keypoint. To ensure full coverage of torso and head, we add a margin equal to 50% of the crop's width and height. Figure 3 shows examples for each perspective derived from one image.

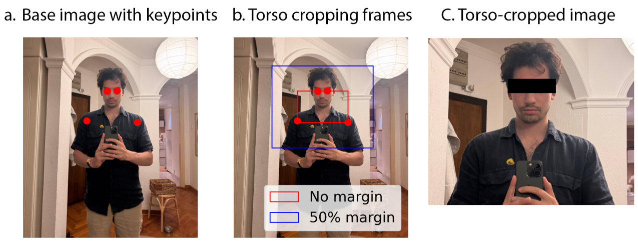

Figure 3: Example of torso-up cropping strategy (private photo with consent).

Evaluation on unseen Visual-Body-To-BMI dataset Using the best-performing image perspective, we evaluate our best-performing model **trained on the WayBED dataset** on the **unseen Visual-Body-To-BMI dataset** (Jiang and Guo 2019) to assess generalization to unseen data. The Visual-Body-To-BMI dataset consists of 5900 images, and we use 30% for the evaluation following prior work (Jin et al. 2022).

Finetuning on VisualBodyToBMI Real-world deployment may involve distribution shifts, e.g., new individuals, environments, or image conditions relative to the training data. To maintain performance without full retraining, the model must be adaptable to new datasets. We evaluate this by finetuning our pretrained model on new data using three strategies: (1) fully unfrozen, all layers updated; (2) partially unfrozen, starting from the last dense block; and (3) only the last layer is retrained.

4 Results

In this section, we present the results of our experiments. Section 4.1 provides an overview of the WayBED dataset and its curation. In Section 4.2, we report model performance across different image perspectives (full-body, torso-up, and face-only). Finally, Section 4.3 presents the model's generalization performance on the VisualBodyToBMI dataset using the best-performing perspective.

4.1 WayBED dataset preparation

The WayBED dataset contains 84,963 images from 25,353 individuals, collected over more than a decade via the WayBetter fitness app. Images were captured by users on a wide range of Android and iOS devices. Participation in the WayBetter fitness programme and sharing of data was entirely voluntary. There was no compensation for data sharing. In total, over 900,000 users submitted more than 5 million images, of which 25,353 individuals provided explicit consent for research use under the condition that their images would not be publicly released and could not be reverse-engineered or inferred from any published results. The dataset includes 49% self-identified female and 9% male participants; 42% did not report gender.

Filtering Applying the filtering procedure from Section 3.1, 13,642 images (16.2%) were removed prior to training. Specifically, 12,402 images were excluded due to too small bounding boxes, 513 due to low bounding box confidence, and 2,052 due to unsuitable poses. Figure 4 visualizes the filtering process, and Figure 5 shows an image being rejected due to the person being too small.

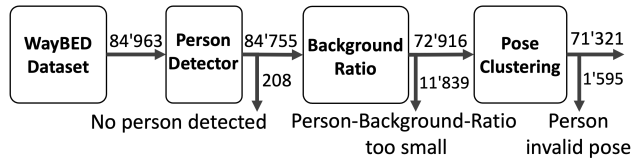

Figure 4: Images filtered out by each step.

Figure 5: Example of person-to-background ratio filtering (private photo with consent)

Posture clustering The optimal number of clusters was determined to be 4 using the elbow criterion . Clusters 1 and 2, comprising 43,192 and 39,719 images respectively, depict individuals standing upright in a frontal pose—distinguished primarily by arm position (arms by the sides vs. raised arms to hold a camera). Clusters 3 and 4, containing 2,051 and 1 image, represent outlying poses and a non-human image, often with alternative camera angles. These two clusters were excluded from the dataset. Between clusters 1 and 2, the difference in arm position is visible in the positioning of the wrist key points. Cluster 3 shows no discernible human pose.

4.2 Estimation performance with different perspectives on WayBED dataset

Table 3 reports MAPE and MAE for different image perspectives on the WayBetter dataset. Full-body perspectives achieve the lowest BMI error (MAPE 7.9%, MAE 2.56), followed by torso-up (9.1%, 2.97) and face-only (11.1%, 3.66). The same pattern holds for absolute weight error, with full-body perspective yielding the lowest MAE (7.29 kg), compared to torso-up (8.41 kg) and face-only (10.38 kg).

4.3 Performance analysis of full-body images

Table 4 reports MAPE for BMI estimation using full-body (best perspective) images with DenseNet-121 and

	Full-Body	Torso-Up	Face-Only
MAPE (%)	**7.9**	9.1	11.1
MAE (BMI)	**2.56**	2.97	3.66
MAE (kg)	**7.29**	8.41	10.38

Table 3: WayBED BMI estimation results by perspective.

DenseNet-201 across the training and evaluation settings described in Section 3.3. When trained and evaluated on WayBED, DenseNet-201 achieved a MAPE of 7.90%. Training only on WayBED and evaluating on the unseen test-set of M201 reaching 13.38%. When trained and evaluated on VMBI, DenseNet-121 reached 9.28% MAPE and DenseNet-201 achieved 8.75%. Fine-tuning DenseNet-201 on VMBI reduced the error to 8.56%.

Model	Trained on	Eval on	MAPE
DenseNet-121	WayBED	WayBED	11.7
DenseNet-121	WayBED	VBMI	23.10
DenseNet-121	VBMI	VBMI	9.28
DenseNet-121	WayBED, ft. VBMI	VBMI	10.72
DenseNet-201	WayBED	WayBED	**7.90**
DenseNet-201	WayBED	VBMI	13.38
DenseNet-201	VBMI	VBMI	8.75
DenseNet-201	WayBED, ft. VBMI	VBMI	**8.56**

Table 4: Results of BMI estimation.

Note: ”ft. VBMI” means finetuned on VisualBodyToBMI; finetuning results are for fully unfrozen model;

5 Discussion

In this section, we discuss the practical implications of our findings and outline the study’s limitations.

Principal findings Our model shows strong generalization to unseen data, achieving a MAPE of 13.38% on the test portion of the VisualBodyToBMI dataset (full-body), despite not being included during training. On WayBED, full-body images yielded the best performance (MAPE 7.9%), highlighting the importance of context and background for BMI estimation. The torso-up perspective also performed well (MAPE 9.1%), suggesting it may serve as a viable alternative when full-body images are unavailable. Face-only inputs resulted in the lowest performance (MAPE 11.1%).

Comparison to prior work To the best of our knowledge, we present the first evaluation of BMI estimation on a large-scale, real-world dataset comprising 84,963 images—substantially larger than prior datasets, which included 14,500 facial or 5,900 full-body images (e.g., VisualBodyToBMI (Jiang and Guo 2019)). Without training on Visual-Body-To-BMI, our model achieves a MAPE of 13.38%, closely matching the original 12.50% (Jiang and Guo 2019) reported with in-dataset training. Fine-tuning on VisualBodyToBMI after training on WayBED reduces the MAPE to 8.56%, setting a new state of the art. This also constitutes the first cross-dataset evaluation of a pre-trained BMI model on an entirely unseen dataset. Additionally, we present the first direct comparison of facial and full-body perspectives within a single dataset, with our full-body model achieving a MAPE of 7.9%, improving over the previously reported 9.6% by Jin et al. (2022).

Practical implications A MAPE of 7.9% in BMI estimation aligns with similar tools for assessing body composition and body fat, where a MAPE between 6–8% is generally considered good (Majmudar et al. 2022a). This performance suggests potential for applications where direct measurements are impractical, such as large-scale population studies, where individuals tend to underreport their BMI (Mourouti et al. 2025), or telehealth applications. While the accuracy declines for torso-only or face-only perspectives, these may still hold value in telehealth contexts to enable BMI estimation during video consultations. For clinical use, however, the achieved MAE of 2.56 BMI might lead to misclassification across critical BMI thresholds (e.g., shifting from “overweight” to “obese”), which might trigger clinical evaluation or interventions (Jan and Weir 2019). BMI estimation for clinical applications should be as accurate as possible (Wenhold and Nel 2022) and thus requires further research. To support such efforts, we fully release the code for replicating and deploying our approach. While the WayBED-trained model cannot be publicly released due to privacy requirements, we may share it upon request. We also provide code for training and deploying models on custom datasets. All models and filtering modules for rejecting unsuitable images are available via the CLAID (Langer et al. 2024) framework for mobile deployment.

Limitations Based on our experiments, we identify the following limitations of our study:

1. *Estimation from a single image:* Our model estimates BMI from a single image; incorporating images from different perspectives or time points could reduce error, particularly for tracking relative BMI changes.
2. *Lack of scale reference:* Images were captured at varying distances, resulting in inconsistent scales of the person within each image. Introducing reference objects (e.g., a paper sheet or credit card) during image capture could potentially reduce estimation error.
3. *Self-reported height and weight:* Labels were self-reported by participants which introduces potential inaccuracies in the labels and may affect reliability.
4. *Gender imbalance:* The dataset shows a gender imbalance: 49% female, 9% male, and 42% not reported, which may affect performance and introduce bias across genders.

References

Altinigne, C. Y.; Thanou, D.; and Achanta, R. 2020. Height and Weight Estimation from Unconstrained Images. In *ICASSP 2020 - 2020 IEEE International Conference on Acoustics, Speech and Signal Processing (ICASSP)*, 2298–2302.

Ansel, J.; Yang, E.; He, H.; Gimelshein, N.; Jain, A.; Voznesensky, M.; Bao, B.; Bell, P.; Berard, D.; Burovski, E.; Chauhan, G.; Chourdia, A.; Constable, W.; Desmaison, A.; DeVito, Z.; Ellison, E.;

Feng, W.; Gong, J.; Gschwind, M.; Hirsh, B.; Huang, S.; Kalambarkar, K.; Kirsch, L.; Lazos, M.; Lezcano, M.; Liang, Y.; Liang, J.; Lu, Y.; Luk, C.; Maher, B.; Pan, Y.; Puhrsch, C.; Reso, M.; Saroufim, M.; Siraichi, M. Y.; Suk, H.; Suo, M.; Tillet, P.; Wang, E.; Wang, X.; Wen, W.; Zhang, S.; Zhao, X.; Zhou, K.; Zou, R.; Mathews, A.; Chanan, G.; Wu, P.; and Chintala, S. 2024. PyTorch 2: Faster Machine Learning Through Dynamic Python Bytecode Transformation and Graph Compilation. In *29th ACM International Conference on Architectural Support for Programming Languages and Operating Systems, Volume 2 (ASPLOS '24)*. ACM.

Dantcheva, A.; Bremond, F.; and Bilinski, P. 2018. Show me your face and I will tell you your height, weight and body mass index. In *2018 24th International Conference on Pattern Recognition (ICPR)*, 3555–3560.

Debnath, P.; Rifa, U. A.; Rafa, B. K.; Akib, A. H. T.; and Rahman, M. A. 2024. Celeb-FBI: A Benchmark Dataset on Human Full Body Images and Age, Gender, Height and Weight Estimation using Deep Learning Approach. arXiv:2407.03486.

Farina, G. L.; Orlandi, C.; Lukaski, H.; and Nescolarde, L. 2022. Digital Single-Image Smartphone Assessment of Total Body Fat and Abdominal Fat Using Machine Learning. *Sensors (Basel)*, 22(21).

Horn, D. B.; Almandoz, J. P.; and Look, M. 2022. What Is Clinically Relevant Weight Loss for Your Patients and How Can It Be Achieved? A Narrative Review. *Postgraduate Medicine*, 134(4): 359–375.

Hu, J.; Shen, L.; Albanie, S.; Sun, G.; and Wu, E. 2019. Squeeze-and-Excitation Networks. arXiv:1709.01507.

Huang, G.; Liu, Z.; van der Maaten, L.; and Weinberger, K. Q. 2018. Densely Connected Convolutional Networks. arXiv:1608.06993.

Jan, A.; and Weir, C. 2019. BMI Classification Percentile And Cut Off Points.

Jiang, M.; and Guo, G. 2019. Body Weight Analysis From Human Body Images. *IEEE Transactions on Information Forensics and Security*, 14(10): 2676–2688.

Jin, Z.; Huang, J.; Xiong, A.; Pang, Y.; Wang, W.; and Ding, B. 2022. Attention Guided Deep Features for Accurate Body Mass Index Estimation. *Pattern Recognition Letters*, 154: 22–28.

Kim, S.; Lee, K.; and Lee, E. C. 2023. Multi-View Body Image-Based Prediction of Body Mass Index and Various Body Part Sizes. In *2023 IEEE/CVF Conference on Computer Vision and Pattern Recognition Workshops (CVPRW)*, 6034–6041.

Langer, P.; Altmüller, S.; Fleisch, E.; and Barata, F. 2024. CLAID: Closing the Loop on AI & Data Collection — A cross-platform transparent computing middleware framework for smart edge-cloud and digital biomarker applications. *Future Generation Computer Systems*, 159: 505–521.

Larsson, S. C.; and Burgess, S. 2021. Causal role of high body mass index in multiple chronic diseases: a systematic review and meta-analysis of Mendelian randomization studies. *BMC Medicine*, 19(1): 320.

Lin, T.-Y.; Maire, M.; Belongie, S.; Hays, J.; Perona, P.; Ramanan, D.; Dollár, P.; and Zitnick, C. 2014. Microsoft COCO: Common Objects in Context. volume 8693. ISBN 978-3-319-10601-4.

Majmudar, M.; Chandra, S.; Yakkala, K.; Kennedy, S.; Agrawal, A.; Sippel, M.; Ramu, P.; Chaudhri, A.; Smith, B.; Criminisi, A.; Heymsfield, S.; and Stanford, F. 2022a. Smartphone camera based assessment of adiposity: a validation study. *npj Digital Medicine*, 5: 79.

Majmudar, M. D.; Chandra, S.; Yakkala, K.; Kennedy, S.; Agrawal, A.; Sippel, M.; Ramu, P.; Chaudhri, A.; Smith, B.; Criminisi, A.; Heymsfield, S. B.; and Stanford, F. C. 2022b. Smartphone camera based assessment of adiposity: a validation study. *npj Digital Medicine*, 5(1): 79.

Mirabet-Herranz, N.; Mallat, K.; and Dugelay, J.-L. 2023. New Insights on Weight Estimation from Face Images. In *2023 IEEE 17th International Conference on Automatic Face and Gesture Recognition (FG)*, 1–6.

Mourouti, N.; Kantaras, P.; Mouratidou, T.; Karaglani, M.; Giménez-Legarre, N.; De Miguel-Etayo, P.; Rurik, I.; Torzsa, P.; Iotova, V.; Tankova, T.; Bazdarska, Y.; Wikstrom, K.; Cardon, G.; Liatis, S.; Makrilakis, K.; and Manios, Y. 2025. Determinants of BMI underreporting in adults from families at high risk for type 2 diabetes in Europe: The Feel4Diabetes study. *Journal of Public Health*, 1–9.

Pantanowitz, A.; Cohen, E.; Gradidge, P.; Crowther, N.; Aharonson, V.; Rosman, B.; and Rubin, D. 2021. Estimation of Body Mass Index from photographs using deep Convolutional Neural Networks. *Informatics in Medicine Unlocked*, 26: 100727.

Ricanek, K.; and Tesafaye, T. 2006. MORPH: a longitudinal image database of normal adult age-progression. In *7th International Conference on Automatic Face and Gesture Recognition (FGR06)*, 341–345.

Serengil, S. I.; and Ozpinar, A. 2020. LightFace: A Hybrid Deep Face Recognition Framework. In *2020 Innovations in Intelligent Systems and Applications Conference (ASYU)*, 23–27. IEEE.

Tai, C.-H.; and Lin, D.-T. 2015. A Framework for Healthcare Everywhere: BMI Prediction Using Kinect and Data Mining Techniques on Mobiles. In *2015 16th IEEE International Conference on Mobile Data Management*, volume 2, 126–129.

Wells, M.; Goldstein, L. N.; Alter, S. M.; Solano, J. J.; Engstrom, G.; and Shih, R. D. 2023. The accuracy of total body weight estimation in adults - A systematic review and meta-analysis. *Am J Emerg Med*, 76: 123–135.

Wells, M.; Goldstein, L. N.; and Bentley, A. 2017. The accuracy of emergency weight estimation systems in children-a systematic review and meta-analysis. *Int J Emerg Med*, 10(1): 29.

Wen, L.; and Guo, G. 2013. A computational approach to body mass index prediction from face images. *Image and Vision Computing*, 31(5): 392–400.

Wenhold, F.; and Nel, S. 2022. When is the estimation of weight and height good enough ? A life cycle view. *South African Journal of Clinical Nutrition*, 35: i–ii.

Wu, Y.; Kirillov, A.; Massa, F.; Lo, W.-Y.; and Girshick, R. 2019. Detectron2. https://github.com/facebookresearch/detectron2.