

https://dspace.org/

Estimation of Body Mass Index from photographs using deep Convolutional Neural Networks

Adam Pantanowitz; Emmanuel Cohen; Philippe Gradidge; Nigel Crowther; Vered Aharonson ...

https://hdl.handle.net/10539/36541
SCOPUS

Downloaded from a DSpace repository. https://dspace.org/

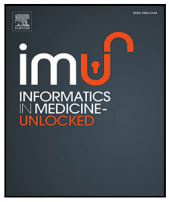
Informatics in Medicine Unlocked 26 (2021) 100727



Contents lists available at ScienceDirect

Informatics in Medicine Unlocked

journal homepage: www.elsevier.com/locate/imu



Estimation of Body Mass Index from photographs using deep Convolutional Neural Networks☆

A. Pantanowitz [a,*], E. Cohen [b,c,d], P. Gradidge [e], N.J. Crowther [f], V. Aharonson [a], B. Rosman [g], D.M. Rubin [a]

[a] Biomedical Engineering Research Group, School of Electrical & Information Engineering, University of the Witwatersrand, Johannesburg, South Africa
[b] CNRS, UMR 7178 «Institut Pluridisciplinaire Hubert CURIEN », Université de Strasbourg, Strasbourg, France
[c] MRC/Wits Developmental Pathways for Health Research Unit, Department of Paediatrics, Faculty of Health Sciences, University of the Witwatersrand, Johannesburg, South Africa
[d] CNRS, UMR 7206 «Eco-anthropologie », Muséum National d'Histoire Naturelle, Paris, France
[e] Centre for Exercise Science and Sports Medicine, Faculty of Health Sciences, University of the Witwatersrand, Johannesburg, South Africa
[f] Department of Chemical Pathology, National Health Laboratory Service and University of the Witwatersrand, Johannesburg, South Africa
[g] School of Computer Science and Applied Mathematics, University of the Witwatersrand, Johannesburg, South Africa

ARTICLE INFO

Keywords:
Anthropomorphism
Body Mass Index
Computer vision
Deep Convolutional Neural Networks
Machine learning

ABSTRACT

Obesity is an important concern in public health, and Body Mass Index is one of the useful, common and convenient measures. However, Body Mass Index requires access to accurate scales and a stadiometer for measurements, and could be made more convenient through analysis of photographs. It could be applied to photographs comprising more than one individual, leading to population screening. We use Convolutional Neural Networks to determine Body Mass Index from photographs in a study with 161 participants. The relatively low number of participants in the data, a common problem in medicine, is addressed by reducing the information in the photographs by generating silhouette images. We successfully determine Body Mass Index for unseen test data with high correlation between prediction and actual values, with correlation measurements of greater than 0.93 and a mean absolute error of 1.20.

1. Introduction

Obesity results in increased morbidity and mortality [1] due to its association with a number of diseases such as hypertension, type 2 diabetes, coronary artery disease, dyslipidaemia, various cancers, asthma, and non-alcoholic steatohepatitis [2]. Obesity is therefore a major public health concern, with an increasing prevalence in both low-and-middle income and high income countries [3]. The accurate assessment of obesity levels in populations is therefore vital for monitoring prevalence and incidence and for the assessment of population-level interventions. It will also allow for the allocation of funds and required infrastructure to areas with high levels of obesity and associated co-morbidities.

The most commonly used measure of total body fat mass is body-mass index (BMI). This is a convenient measure which uses body weight and height and is expressed as kg/m^2, with obesity being defined as a BMI $\geq$ 30 kg/m^2. Current methods of obtaining this data require appropriately calibrated measurement equipment and can be time-consuming when applied to large numbers of participants. There is therefore a need for alternative, rapid, and accurate techniques for measuring BMI that can be easily applied to large population surveys.

A possible alternative to the physical measurement of weight and height would be the assessment of these measures from photographs. This is likely to be more convenient (as indicated in [4]), allowing individuals to use cell phone-based cameras for obtaining images for further processing. The aim of the current study was therefore to generate BMI data from simple photographic images using machine learning.

2. Background

Body mass index (BMI) is one of the easiest and most commonly used methods for assessing obesity despite some shortcomings [5]. Other anthropomorphic measurements, such as waist circumference, is another important indicator of disease and is of clinical importance when considering cardiovascular outcomes [6,7]. Other studies have

☆ Institutional Review Board approval by the University of the Witwatersrand Human Research Ethics Committee (Medical) on 11 July 2018.

* Corresponding author.
E-mail addresses: adam.pantanowitz@wits.ac.za (A. Pantanowitz), emmanuel.cohen@wits.ac.za (E. Cohen).

https://doi.org/10.1016/j.imu.2021.100727
Received 17 May 2021; Received in revised form 3 September 2021; Accepted 3 September 2021
Available online 15 September 2021
2352-9148/© 2021 Published by Elsevier Ltd. This is an open access article under the CC BY-NC-ND license (http://creativecommons.org/licenses/by-nc-nd/4.0/).

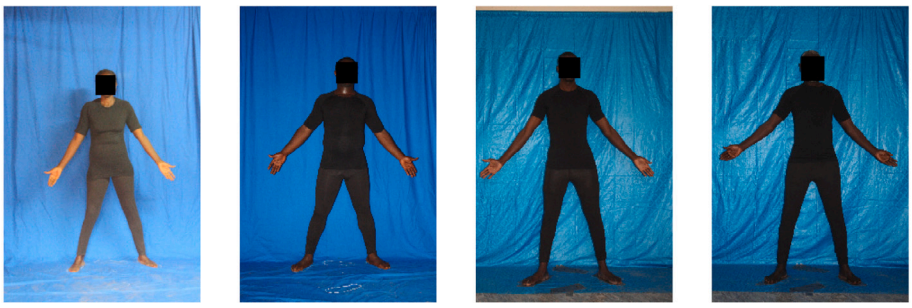


Fig. 1. An arbitrary set of four anonymised full colour photographs from the 161 study participants.



Fig. 2. The preprocessing performed on the hand-generated image set.

shown that neck circumference is another useful important anthropomorphic measurement [8–11]. Nonetheless, BMI remains a widespread and important clinical indicator despite the shortcomings [12], and is simple to calculate [13].

Population surveys assessing the prevalence of obesity measure BMI using calibrated weight scales and stadiometers, or participants are contacted telephonically and asked about their weight and height [14]. These methods may be relatively costly and are time consuming and prone to error, particularly the latter. There is therefore a need for alternative methods of measuring BMI that use techniques that are accurate and quick and can be applied to large sample sizes. One such method would be extracting anthropometric data from photographic images. In this regard, one study has used human silhouettes generated from dual-energy X-ray absorptiometry (DEXA) whole body scans to calculate total body fat mass using active shape modelling (ASM) techniques [15]. In addition, methods using 3D whole body surface scanners and the generation of multiple point cloud images have been used to calculate various anthropometric measures [16].

A similar study used a Kinect system to obtain data [17]. These procedures require specialised imaging technology however, there is a study using face photography as a proxy to determine BMI with deep learning [18]. Recent improvements on these results from full body photographs have been obtained [4], but this method uses machine learning in the form of body contour generation and skeleton markers, rather than learning from the source image alone.

3. Methodology

We utilise silhouette images to determine BMI of participants using imagine processing and machine learning (Convolutional Neural Networks).

Deep learning generally requires large data sets in order to learn effectively [19]. In this study, however, data was constrained due to the number of participants. To enhance the efficacy of the small data set, silhouette images were generated from the original photographs. This substantially reduces the information and complexity of the images, allowing for a binary image as the basis for generation.

In this section, we consider the data set and approach taken in order to generate an estimation of BMI from silhouette images. We then consider the machine learning approach used to generate the BMI estimate.

3.1. Participants

Anthropometric data and full colour images were gathered from 161 (79 female) Cameroonians and Senegalese participants. Participants were selected who were of reproductive age, but not pregnant, and were chosen from two ancestory groups based on differing macroscopic phenotypes based on environment, and differing genotypes [20].

Several full colour front-facing photographs were available for each participant. Fig. 1 presents a sample of the source photographs used in the study. While attempts were made to achieve uniformity for the photographs, variations in lighting, colour, position, and other differences are evident. For this study, only single front-facing images were used.

3.2. Image preprocessing and silhouette generation

Owing to the low number of participants, it was necessary to further reduce complexity by standardising the silhouette images as far as possible, as a preprocessing step to machine learning. This process simplifies the machine learning problem as there are fewer parameters to learn, thus enabling us to leverage a smaller dataset more effectively.

Images of the outlines of the 161 participants were manually generated by marking images by hand (white background, black outline of participant). These hand generated images were non-uniform and contained artefacts (such as the tape markings on the ground; variation in participant shadows; the edge of blue floor boundary; and

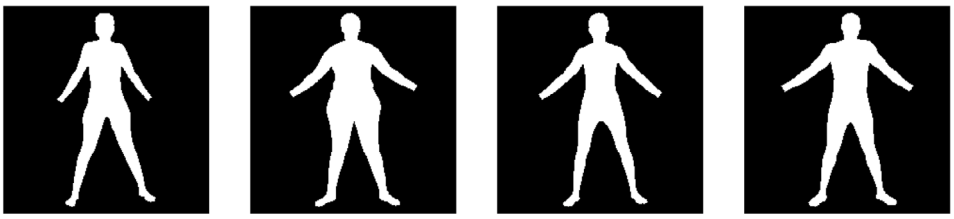


Fig. 3. A random sample of four silhouette images from the 161 study participants.



Fig. 4. A random sample of four preprocessed, automatically generated silhouette images from *s2* (slightly scaled up, actually 64 × 64 pixels).

the shadows in the curtains). Despite the controlled conditions in which the images were taken, segmentation proved challenging due to the boundary between clothes and skin; the pattern of the backdrop; and the aforementioned artefacts. There was also variability between photography which were taken under various lighting conditions, with differing zoom, and blurriness. There is also variations in the clothing, positioning, and lighting of a participant during photography. Furthermore, participants were photographed in two controlled environments which differed from each other. As the images were from an existing study, repeating the imaging under more uniform conditions was not possible. Critically, due to the constrained number of participants, we did not wish to discard any of the source data.

Our method ultimately allowed for the extraction of useable data despite all of the variability described above. In order to achieve this, the images were thus manipulated as described below. This process is presented in Fig. 2. Note that the images do not reflect the size/detail and are shown for illustrative purposes (for instance, artefacts cannot be observed).

The images of all participants, male and female, were combined together. Images were converted to PNG format at a uniform resolution of 150 × 150 (the lowest common resolution in the set), and were thresholded at 20% to reduce artefacts. The images were converted to greyscale with *average* and *separate* filters. The grey channel images were then negated to invert them (black background, white silhouette). Four randomly chosen silhouette images are shown as examples in Fig. 3.

Owing to artefacts, the images were then cropped (in a script) to remove surrounding empty space. Following this, images were modified to a fixed height of $1864 pixels$ (the lowest common height in the images) while maintaining the aspect ratio. Finally, the resulting uniform, greyscale images were padded with black to produce an image sized 2000 × 2000 which has a bounding box. Finally, these images were reduced to 64 × 64 to generate the set used, s. A sample of four silhouette images selected randomly from s is presented in Fig. 4.

3.3. Composition of the dataset

In many machine learning applications, outliers are removed [21] or replaced [22] to produce better predictive results. This is not done in the current study for three reasons: first, outliers in this case are valid in the domain (in fact, one can argue it is more valuable to detect a high or low BMI clinically than one falling into the statistically "normal" range); second, there are not many images available; and third, great care was taken in validating the source data collected [20]. For these reasons, data are not manipulated or pre-processed to improve predictive performance, apart from standardisation with the extremes intact.

The data used in this study were collected for the purpose of generating body scales, and so many participants were required with a BMI between 20 and 30. There is inherent bias in the dataset, in that the BMI data follows a skewed distribution that is reflective of BMI within this general population [23]. In order to illustrate the bias in the dataset, Fig. 6 presents a distribution plot of the BMI data which shows that the majority of the participants had BMI within the normal range (18.8–25.0). Fig. 5 presents a cumulative distribution plot of the full data set (with the full 160 participants). The plot shows a sharp

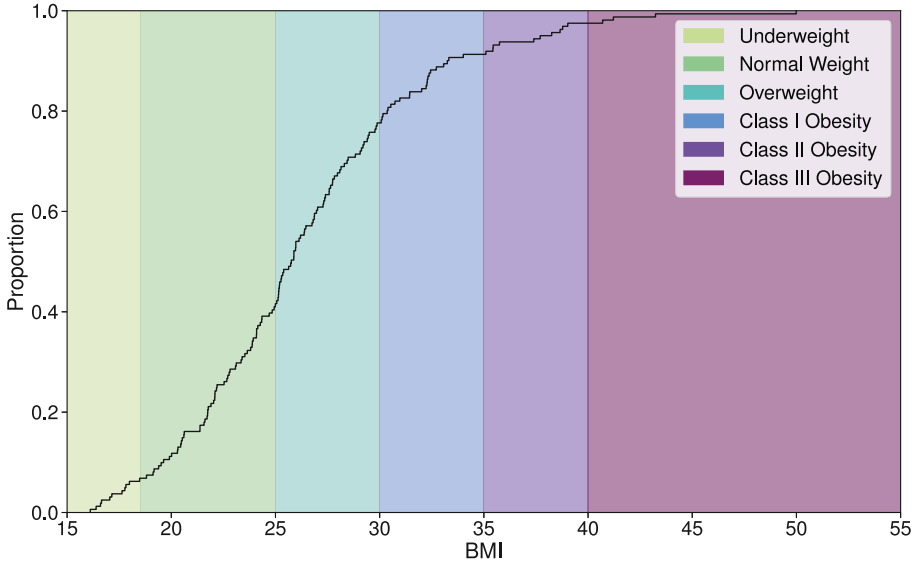


Fig. 5. Cumulative Distribution Plot with highlighted BMI regions as per the six categorical definitions in Table 6.

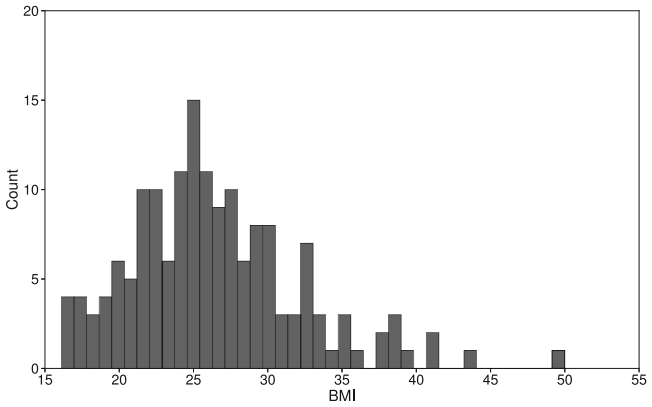


Fig. 6. Histogram representing the distribution of BMI in the dataset.

gradient (faster rise) in regions where there are more data. The plot therefore shows that the majority of the participants' BMIs fall between 20 and 30 (approximately 85%). There are approximately 20% of cases with BMI > 30, and fewer than 10% of the examples with a BMI < 20.

3.4. Machine learning

The processed silhouette images and ground-truth BMI data were imported as the working dataset with the former being the input into the system, and the latter being the output (target).

Deep learning generally requires large amounts of data. In recent years, image data has been used to generate Convolutional Neural Network (CNN) classification algorithms (such as in [24,25]). Creating binary silhouette images dramatically reduces the information content required for processing.

Optimisation was performed to maximise the test set while maintaining a validation set size sufficient to allow for early stopping and for sufficient testing. For the purposes of presenting results, an appropriate heuristic split (with a random seed) is applied with 60% (96 participants) of the data being split for training, 20% (32 participants) for validation, and 20% (33 participants) for testing.

Multiple variations of this split were also tested. There is a variance of results dependent on the random subset sample. The minimum training size with reasonable performance (based on a validation error of less than 0.2) were found, at minimum, to be approximately 50 participant images for training (participant to the drawn sample being representative), and the remainder making up the validation set.

The BMI data were normalised using standard scaling ($\mu = 26.40, \sigma = 5.75$), which was observed empirically to show improved performance over a robust scaler and regular normalisation.

3.5. Feature selection with classical machine learning

Machine learning models are built with differing feature selection algorithms - a Gabor filter (Gabor), and a Histogram of Gradients (HOG). Both feature sets are then used to train and test a Random Forest (RF) [26], XGBoost (XGB) [27], and Support Vector Machine (SVM) [28] regressors. This produces a permutation of six models which are evaluated, namely Gabor-RF, HOG-RF, Gabor-XGB, HOG-XBG, Gabor-SVM, and HOG-SVM. This approach makes use of more classical machine learning, and the feature extraction through Gabor and HOG generates a feature vector which is useful for the model. Gabor filters are commonly used for feature extraction in various computer vision applications [29–32], while HOG has shown great utility in human detection [33], and are useful in medical image processing applications [34].

Each of the parameters used for feature generation were chosen by generation, inspection, and selection. We used a fairly wide variety of variables in the case of the Gabor filter to attempt to generate optimal features as part of the set. The following sections describe the parameters for feature generation for each.

Table 1
Gabor filter parameters for feature generation for Gabor-RF.

Parameter	Values
ksize	[2, 4, 6, 8]
σ	[2, 5, 8]
θ	$[0, \frac{\pi}{4}, \frac{\pi}{2}, 3\frac{\pi}{4}]$
λ	$[0, \frac{\pi}{4}, \frac{\pi}{2}, 3\frac{\pi}{4}]$
γ	[0.05, 0.5]

Table 2
HOG parameters for feature generation for HOG-RF.

Parameter	Values
Orientations	4
Pixels per cell	(4, 4)
Cells per block	(1, 1)

3.5.1. Gabor filters for feature generation

The first considered model consists of the Gabor Filter (applied to the original images for feature generation).

The Gabor Filter was implemented by generating training, validation, and test sets, in a single feature vector, with the full set of parameters in Table 1 generated for each image.

3.5.2. Histogram of gradients for feature generation

Parameters used for HOG feature generation are given in Table 2.

3.5.3. Optimisation

The respective feature vector generated by Gabor and HOG were used to train and test a different regressor for each of the six permutations of the models. An optimisation process leveraged a randomised search on the hyperparameters through between one and five rounds of optimisation for each model resulting in the parameters for each model. In each optimisation, 40 iterations ran with 2 cross validations for each. During each round of optimisation, the hyperparameter search space was narrowed to find the most optimal hyperparameters for each model. Each model was optimised using the validation data set, so as not to bias the comparative analysis in Section 4 by exposing one model to more data than the others. The model was then trained using the training data set, and tested with unseen test data.

3.5.4. Convolutional neural network

A two dimensional Convolutional Neural Network (CNN) is implemented in Keras using TensorFlow to perform deep learning on the model.

We use a modified MiniVGGNet, which is a smaller version of VGGNet, originally proposed by [35]. This is based on the implementation in [36]. VGGNet and MiniVGGNet have been successful in a variety of image processing applications [37] including medical [38–40].

Through empirical measurement of validation performance some changes were made to the architecture presented in [36]. These changes were:

- reducing the number of filters from three to two, and reducing the filter size to (4, 8)
- modifying the output layer to include an additional four node linear layer prior to the single output node
- modification to perform the regression task by collapsing the dense nodes to a single output (from which the regression is taken). As most applications of CNNs are classification tasks with a softmax classifier [41], this modification as per [36] allows regression from this architecture.

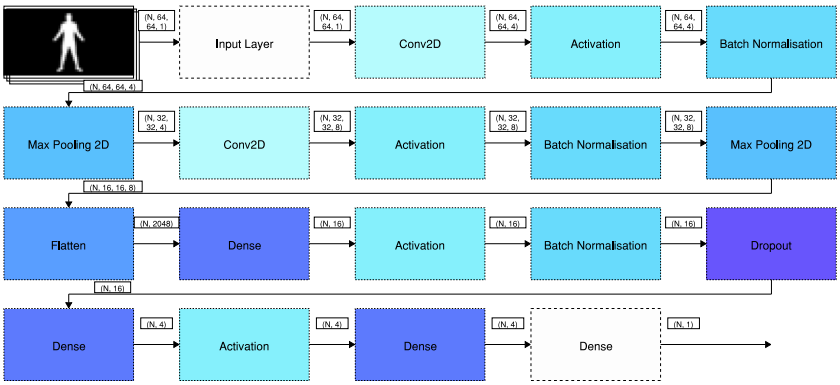


Fig. 7. Model architecture.

Table 3
Comparisons between the various models tested for the purposes of predicting BMI on unseen (test) data.

Metric	Gabor-RF	HOG-RF	Gabor-XGB	HOG-XGB	Gabor-SVM	HOG-SVM	CNN
Mean Squared Error	15.35	21.97	19.43	24.55	12.21	12.28	7.35
Maximum Error	13.65	15.73	11.85	13.39	11.77	10.35	9.15
Co-efficient of Determination (R^2)	0.66	0.52	0.57	0.46	0.73	0.73	0.70
Explained Variance	0.66	0.52	0.59	0.46	0.73	0.73	0.73
Pearson's Correlation Coefficient	0.86	0.75	0.77	0.68	0.88	0.86	0.88

The architecture leverages two CONV => RELU => BN => POOL layer following the initial input filters. The implementation also differs from the original VGGNet [35] due to the addition of the Block Normalisation layers advocated for by [36]. The architecture then flattens the volume, followed by FC => RELU => BN => DROPOUT, followed by another FC => RELU layer. Finally, the model caters for the custom two-stage regression used in this implementation. The CNN architecture and model structure export is presented in Fig. 7. The model accepts images as inputs (of size 128 × 64).

Variations on this final architecture (such as changing the size of convolution windows, the number of dense nodes, and dropout ratios) were evaluated. Sufficient performance was achieved and model structure search was therefore non-exhaustive. The model uses a Mean Squared Error (MSE) loss function.

GridSearch allows for the optimal selection of hyper-parameters for training, and the learning and decay rate make a substantial difference to whether or not the algorithm converges. The optimal parameters were found to be the Adam algorithm with a learning rate of 0.01, and decay of 1×10^{-5}.

Training was performed using source images and the target of normalised BMI between zero and one.

3.5.5. Data augmentation

The CNN is trained using a flow of generated images (which mix in original images and augmented images), and no strict ratios were enforced in terms of the split between original and augmented images. Augmented images are generated in real-time during training, with no upper limit on how many images are generated: training ceases when the mean squared error validation loss is minimised for sufficient training epochs (in this case, the patience criterion is set at 30 epochs). Training was done with a batch size of 80, steps per epoch of 0.25× the length of the training set, and a maximum of 200 epochs. However, training terminates earlier due to early stopping based on the validation of data performance.

Data augmentation parameters were optimised using the Basin-hopping algorithm [42] so as to maximise the usefulness of the data augmentation process. The Basin-hopping algorithm was setup using at first a wide array of parameters, which were gradually decreased through rounds of testing. The algorithm called an optimisation fitness function which varied the rotation and width shift range with the objective of minimum validation loss. The fitness function was repeatedly called with new values, and the CNN was trained and validation loss measured and optimised. This process resulted in low parameters of rotation (2°) and translations in width 0.02 (these values happen to be similar to variations in the original image set), and horizontal image reflections. Augmentations larger than a 3° of rotation, or augmentations such as height changes, reduce performance.

Model convergence does occur even in the absence of augmentation. However, due to the paucity of data, the implementation of augmentation results in variable, but usually marginal improvement in error, depending on the data split.

4. Results

We implement and test the classical machine learning model permutations and CNNs to compare the ability of each machine learning pipeline to represent the data.

4.1. Comparing the models

The results in Table 3 demonstrate that the CNN is the optimal model of those implemented and tested. It performs better by a factor of more than 2x on the Mean Squared Error, produces a lower maximum error, and performs better on the fit metrics such as the Co-efficient of Determination (R^2). As such, the rest of the study leverages the CNN model for the analysis.

4.2. K-fold cross validation

We implement K-fold Cross Validation [43] in order to test the models utilising the full dataset, rather than a held back subset, which is considered in Section 4.3. The results present the Mean Squared Error (MSE) averaged across the folds utilised. As shown in Table 4, Gabor features slightly out-performs HOG features on average, but the results are similar. Of the classical models tested, SVM performs best,

Table 4
K-fold Cross Validation results for the various considered models.

Model	5 Fold Cross Validation Mean Error	10 Fold Cross Validation Mean Error
Gabor-RF	0.29	0.26
HOG-RF	0.38	0.37
Gabor-XGB	19.78	10.27
HOG-XGB	7.56	0.66
Gabor-SVM	0.20	0.26
HOG-SVM	0.29	0.32
CNN	0.017	0.021

Table 5
Results of BMI prediction performance over 100 training and test cycles with sets split with the same random seed (30).

Metric	Mean result	Best result
Validation Loss	0.124	0.0625
Pearson's Correlation Co-efficient	0.928	0.976
Mean absolute error (MAE)	1.66	1.20
Root Mean Square Error (RMSE)	2.16	1.52

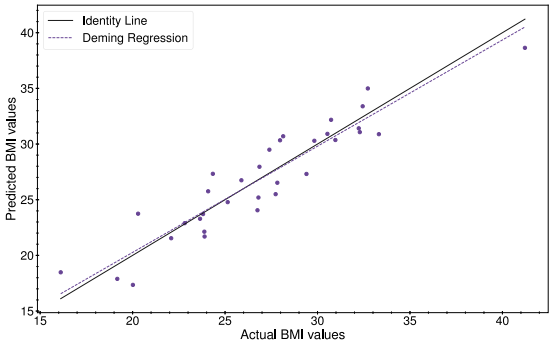


Fig. 8. Scatter Plot of Predicted vs Actual BMI for unseen data.

with the best of the classical machine learning models being Gabor-SVM. XGB performs poorly due to one to two very large deviations in regression outputs which impact the performance results. CNN greatly outperforms the other models by approximately an order of magnitude. Therefore the CNN approach generates significant performance improvement over the other approaches, and thus CNN forms the basis for further investigation.

4.3. Detailed CNN results

The system allows BMI to be predicted with a Pearson Correlation Coefficient of above 0.928 on average across 100 runs between the test values and the predicted test values, and is higher than this for many training/testing cycles. Sufficiently low validation performance (0.2 and below) is a signal of model convergence, and robust prediction results follow. On some runs, we achieved a validation performance below 0.05, and particularly when more training data are allocated. This, of course, represents a trade-off with the number of validation and test data samples. Aggregated performance results across a few metrics is presented in Table 5.

Fig. 8 presents the predictive performance of the model.

Fig. 9 presents the mean difference plot for the actual vs. predicted BMI data, depicting that the mean difference is −0.05 (below one in the majority of test runs), though the 1.96 standard deviation bands differ depending on the run.

As a test of validity, we performed arbitrary rotations of test data for three randomly chosen participants, and plotted the predicted model output for the rotated inputs for three participants. The results of the test (unseen) image rotations are presented as Fig. 10, and indicates the actual participant BMI as a straight line. The estimated BMIs match the

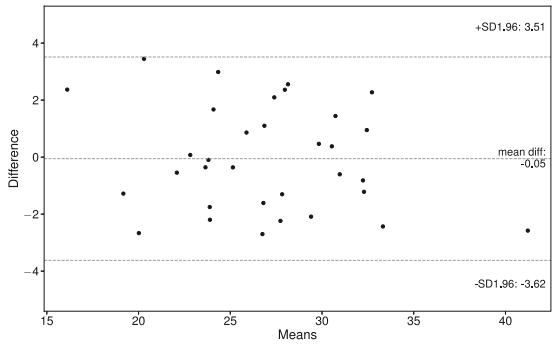


Fig. 9. Mean-difference plot for given model's results and actual data. The identity line represents perfect agreement between predicted and actual BMI values, and the Deming Regression is determined from results.

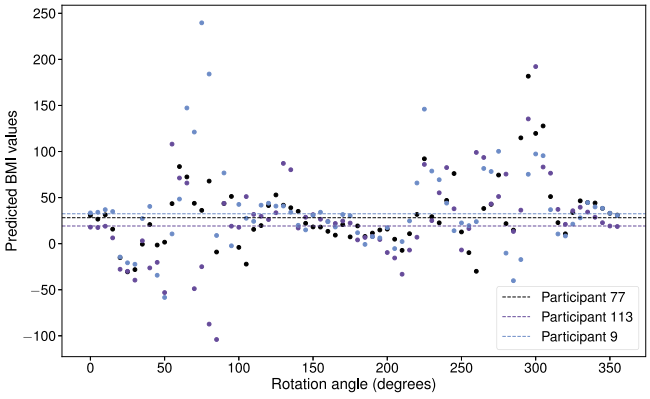


Fig. 10. Model's predicted output for rotations of test input images with 5° rotations.

Table 6
Categories of obesity as per [44].

Category	BMI range	Definition
1	BMI < 18.5	Underweight
2	18.5 ≥ BMI < 25	Normal Weight
3	25 ≥ BMI < 30	Overweight
4	30 ≥ BMI < 35	Class I Obesity
5	35 ≥ BMI < 40	Class II Obesity
6	40 ≥ BMI	Class III Obesity

actual data well at 0° and 360° (as expected). Outside of these, there is generally poor output outside of the ordinary range of reasonable BMI values, where the silhouette appears at some angle. There is a region around 180° where the values normalise back to a reasonable range (as these are likely recognised as broader participants). Interestingly, the rotations show an increased expected BMI as an image rotates, a finding that is consistent with intuition of BMI appearing to increase with rotation.

The definitions of the six categories used is this categorical analysis is as per [44] and is presented in Table 6. A confusion matrix for the test data for a given run is presented in Table 7. Most of the values fall on the diagonal, with some spill-over into adjacent categories. The estimates are close, however, and we see this given that the category six analysis utilises firm numerical boundaries.

4.4. Male & female analysis

The CNN regression performance is considered by grouping male and female participants by running 100 trials. In each trial, training and test data are randomly split with no seed (in the same proportions, 60% training, 20% validation, 20% testing). A new CNN is trained

Table 7
Confusion matrix for prediction test data for the given modelling run.

Prediction \ Category	1	2	3	4	5	6
1	1	0	0	0	0	0
2	2	7	2	0	0	0
3	0	2	7	3	0	0
4	0	0	0	8	0	0
5	0	0	0	0	0	0
6	0	0	0	0	1	0

Table 8
CNN regression performance separated by male and female participants averaged across 88 valid trials.

Metric	Female participants	Male participants
Number of participants	16.10	16.90
Pearson's Correlation Coefficient	0.90	0.85
Explained Variance	0.77	0.68
Maximum Error	7.51	5.68
Mean Squared Error	10.33	7.44
Co-efficient of Determination (R^2)	0.73	0.61

until the validation error is minimised, with a patience of 30. The results across a number of metrics are obtained and averaged over the 100 trials. Averaging the results is necessary due to the relatively low number of instances, as an outlier in either group in a trial creates a significant impact on the results. Models which produced R^2 values < 0 are deemed invalid and excluded as these models are arbitrarily worse and indicated a failure during the training process. 88 valid runs remain, and the average values of the metrics evaluated are presented in Table 8. It is evident that performance is relatively similar, but the model's predictive performance for males is slightly worse across a number of metrics, including Pearson's Correlation Coefficient, Explained Variance, and the R^2 value. The error metrics are higher in females. We hypothesise this set of results can be explained by the variance, as in the full dataset, the BMI variance for females is 40.26, whereas for males it is 23.16. Therefore, the error potential is larger as there is greater spread, but the predictive ability is greater as the model is more able to represent the greater variance across BMI represented in images.

5. Discussion

We present an approach based on images of individual participants that were taken in a somewhat standardised environment. We anticipate that it may be possible to adapt these techniques for large groups of people in a single photograph, and for individuals who are imaged in non-ideal and settings which are not as standard (such as simple photographs from a cell phone). Furthermore, if participants were not in the uniform tightly adhering stocking clothing, it would obfuscate elements of their anatomy, and thus make BMI prediction/inference a harder problem. This problem would require the use of many more images, so that the variation in clothing can be taken into account to the extent possible. The controlled nature of the participants' dress in this study was an enabler of simplicity in BMI detection using photographs. Further study into this is required, such as the study across a much wider dataset of real-world images [4]. With enough variability in the input data, and with utilising information in the face of participant (as is utilised in the study in [18]), the problem of clothing is likely to be soluble albeit perhaps with less accuracy.

With large data sets and modern image processing, it may be feasible to extend these principles into the general public health domain. Looking at multiple participants in a single photograph, the ability to rapidly measure the extent of obesity in populations would constitute a valuable tool for public health research.

Other anthropometric measures, such as, but not limited to, waist circumference, can also be inferred from photographs, as some of these measures are better predictors of disease than BMI [45].

5.1. Recommendations for future work

The number of participants for which photographic images were available was low, resulting in a limited test set. This is consistent with the aforementioned challenge of obtaining data. Therefore, work needs to be done to obtain more test images, or alternatively to leverage the data in such a way as to generate more training data, and therefore free up legitimate data for testing. Data augmentation does achieve this to a marginal extent, improving the validation error somewhat and therefore improving the BMI prediction performance.

The study performed is specifically scoped to BMI. There is a lot of opportunity to extend the work to other anthropometric measures [16], but this proof-of-concept work indicates that the requirement for specialist equipment used in the field of study at present [15,16] may not be necessary.

The source data images would ideally be re-taken while reducing artefacts (such as the tape markings on the ground; ensuring consistency with or absence of shadows; the edge of blue floor boundary; and the shadows in the curtains), under more standard conditions (more consistent lighting; the same background). The images did present some challenges in image processing due to the variability and issues described. Correcting these would enable a single segmentation algorithm to more easily be applied to identify the outline of each participant.

There is a level of error associated with estimation, which does mean that misclassification of the BMI categories shown in Table 2, can occur. With more data, however, the system can be made more robust to variation and we can expect the error to decrease. In future work, adjustments can be made for well-characterised error such as systemic error.

6. Conclusion

The ability to rapidly measure the extent of obesity in large populations would constitute a valuable tool for public health research and in public health screening.

We have successfully implemented a system to predict BMI from a small data set of photographic images of participants. We reduce complexity in the images by creating silhouettes from photographs, and this enables us to obtain accurate prediction of BMI using CNNs in deep learning from a relatively small dataset. This approach may lead to a public health screening tool to assist in supporting/measuring health initiatives in areas of endemic obesity or malnutrition. The use of silhouettes in our process offers the additional benefit of anonymity.

There is increased interest in utilising alternative sources of information for convenient, rapid, and inexpensive assessment of anthropometric variables. We have demonstrated the use of photographs to determine BMI using machine learning, and believe that this method can be easily applied to other body measurements.

Declaration of competing interest

The authors declare that they have no known competing financial interests or personal relationships that could have appeared to influence the work reported in this paper.

Acknowledgement

The authors would like to thank Amazon Web Services for the AWS compute credits which partially enabled this research.

References

[1] Di Angelantonio E, ShN B, Wormser D, Gao P, Kaptoge S, Berrington de Gonzalez A, Cairns B, Huxley R, Joshy G, Lewington S, et al. Global BMI mortality collaboration body-mass index and all-cause mortality: individual-participant-data meta-analysis of 239 prospective studies in four continents. Lancet 2016;388(10046):776–86, [Online] Available: https://doi.org/10.1016/S0140-6736(16)30175-1.

[2] Haslam D, James W. Obsesity. Lancet 2005;366(9492):1197–209, [Online] Available: https://doi.org/10.1016/S0140-6736(05)67483-1.

[3] NCDRisk Factor Collaboration and others. Trends in adult body-mass index in 200 countries from 1975 to 2014: a pooled analysis of 1698 population-based measurement studies with 19 · 2 million participants. Lancet 2016;387(10026):1377–96, [Online] Available: https://doi.org/10.1016/S0140-6736(16)30054-X.

[4] Jiang M, Guo G. Body weight analysis from human body images. IEEE Trans Inf Forensics Secur 2019;14(10):2676–88, [Online] Available: https://doi.org/10.1109/TIFS.2019.2904840.

[5] Adab P, Pallan M, Whincup PH. Is BMI the best measure of obesity?. 2018, [Online] Available: https://doi.org/10.1136/bmj.k1274.

[6] Freiberg MS, Pencina MJ, D'agostino RB, Lanier K, Wilson PW, Vasan RS. Bmi vs. waist circumference for identifying vascular risk. Obesity 2008;16(2):463–9, [Online] Available: https://doi.org/10.1038/oby.2007.75.

[7] Walls HL, Stevenson CE, Mannan HR, Abdullah A, Reid CM, McNeil JJ, Peeters A. Comparing trends in BMI and waist circumference. Obesity 2011;19(1):216–9, [Online] Available: https://doi.org/10.1038/oby.2010.149.

[8] Assyov Y, Gateva A, Tsakova A, Kamenov Z. A comparison of the clinical usefulness of neck circumference and waist circumference in individuals with severe obesity. Endocrine Res 2017;42(1):6–14, [Online] Available: https://doi.org/10.3109/07435800.2016.1155598.

[9] Joshipura K, Muñoz Torres F, Vergara J, Palacios C, Pérez CM. Neck circumference may be a better alternative to standard anthropometric measures. J Diabetes Res 2016;2016. [Online] Available: https://doi.org/10.1155/2016/6058916.

[10] Hingorjo MR, Qureshi MA, Mehdi A. Neck circumference as a useful marker of obesity: a comparison with body mass index and waist circumference. JPMA-J Pak Med Assoc 2012;62(1):36.

[11] Vallianou NG, Evangelopoulos AA, Bountziouka V, Vogiatzakis ED, Bonou MS, Barbetseas J, Avgerinos PC, Panagiotakos DB. Neck circumference is correlated with triglycerides and inversely related with HDL cholesterol beyond BMI and waist circumference. Diabetes/Metabolism Res Rev 2013;29(1):90–7, [Online] Available: https://doi.org/10.1002/dmrr.2369.

[12] Daniels SR. The use of BMI in the clinical setting. Pediatrics 2009;124(Supplement 1):S35–41, [Online] Available: https://doi.org/10.1542/peds.2008-3586F.

[13] Blackburn H, Jacobs Jr. D. Commentary: Origins and evolution of body mass index (BMI): continuing saga. Int J Epidemiol 2014;43(3):665–9, [Online] Available: https://doi.org/10.1093/ije/dyu061.

[14] Brettschneider A-K, Schienkiewitz A, Schmidt S, Ellert U, Kurth B-M. Updated prevalence rates of overweight and obesity in 4-to 10-year-old children in Germany. Results from the telephone-based KiGGS wave 1 after correction for bias in parental reports. Eur J Pediatrics 2017;176(4):547–51, [Online] Available: https://doi.org/10.1007/s00431-017-2861-8.

[15] Xie B, Avila JI, Ng BK, Fan B, Loo V, Gilsanz V, Hangartner T, Kalkwarf HJ, Lappe J, Oberfield S, et al. Accurate body composition measures from whole-body silhouettes. Med Phys 2015;42(8):4668–77, [Online] Available: https://doi.org/10.1118/1.4926557.

[16] Heymsfield SB, Bourgeois B, Ng BK, Sommer MJ, Li X, Shepherd JA. Digital anthropometry: a critical review. Eur J Clin Nutri 2018;72(5):680–7, [Online] Available: https://doi.org/10.1038/s41430-018-0145-7.

[17] Nahavandi D, Abobakr A, Haggag H, Hossny M, Nahavandi S, Filippidis D. A skeleton-free kinect system for body mass index assessment using deep neural networks. In: 2017 IEEE international systems engineering symposium. IEEE; 2017, p. 1–6, [Online] Available: https://doi.org/10.1109/TIFS.2019.2904840.

[18] Kocabey E, Camurcu M, Ofli F, Aytar Y, Marin J, Torralba A, Weber I. Face-to-BMI: using computer vision to infer body mass index on social media. In: Eleventh international AAAI conference on web and social media; 2017.

[19] Taylor L, Nitschke G. Improving deep learning with generic data augmentation. In: 2018 IEEE symposium series on computational intelligence. IEEE; 2018, p. 1542–7, [Online] Available: https://doi.org/10.1109/SSCI.2018.8628742.

[20] Cohen E, Bernard JY, Ponty A, Ndao A, Amougou N, Saïd-Mohamed R, Pasquet P. Development and validation of the body size scale for assessing body weight perception in african populations. PLoS One 2015;10(11):e0138983, [Online] Available: https://doi.org/10.1371/journal.pone.0138983.

[21] Li W, Mo W, Zhang X, Squiers JJ, Lu Y, Sellke EW, Fan W, DiMaio JM, Thatcher JE. Outlier detection and removal improves accuracy of machine learning approach to multispectral burn diagnostic imaging. J Biomed Opt 2015;20(12):121305, [Online] Available: https://doi.org/10.1117/1.JBO.20.12.121305.

[22] Maniruzzaman M, Rahman MJ, Al-MehediHasan M, Suri HS, Abedin MM, El-Baz A, Suri JS. Accurate diabetes risk stratification using machine learning: role of missing value and outliers. J Med Syst 2018;42(5):1–17, [Online] Available: https://doi.org/10.1007/s10916-018-0940-7.

[23] Tsang S, Duncan GE, Dinescu D, Turkheimer E. Differential models of twin correlations in skew for body-mass index (BMI). PLoS One 2018;13(3):e0194968, [Online] Available: https://doi.org/10.1371/journal.pone.0194968.

[24] Huynh BQ, Li H, Giger ML. Digital mammographic tumor classification using transfer learning from deep convolutional neural networks. J Med Imaging 2016;3(3):034501, [Online] Available: https://doi.org/10.1117/1.JMI.3.3.034501.

[25] Jin KH, McCann MT, Froustey E, Unser M. Deep convolutional neural network for inverse problems in imaging. IEEE Trans Image Process 2017;26(9):4509–22, [Online] Available: https://doi.org/10.1109/TIP.2017.2713099.

[26] Breiman L. Random forests. Mach Learn 2001;45(1):5–32, [Online] Available: https://doi.org/10.1023/A:1010933404324.

[27] Chen T, Guestrin C. Xgboost: A scalable tree boosting system. In: Proceedings of the 22nd Acm sigkdd international conference on knowledge discovery and data mining; 2016. p. 785–94. [Online] Available: https://doi.org/10.1145/2939672.2939785.

[28] Cortes C, Vapnik V. Support-vector networks. Mach Learn 1995;20(3):273–97, [Online] Available: https://doi.org/10.1007/BF00994018.

[29] Shariaty F, Mousavi M. Application of CAD systems for the automatic detection of lung nodules. Inform Med Unlocked 2019;15:100173, [Online] Available: https://doi.org/10.1016/j.imu.2019.100173.

[30] Lee C-J, Wang S-D. Fingerprint feature extraction using Gabor filters. Electron Lett 1999;35(4):288–90, [Online] Available: https://doi.org/10.1049/el:19990213.

[31] Li W, Mao K, Zhang H, Chai T. Selection of gabor filters for improved texture feature extraction. In: 2010 IEEE international conference on image processing. IEEE; 2010, p. 361–4, [Online] Available: https://doi.org/10.1109/ICIP.2010.5653278.

[32] Wang X, Ding X, Liu C. Optimized gabor filter based feature extraction for character recognition. In: Object recognition supported by user interaction for service robots, vol. 4. IEEE; 2002, p. 223–6, [Online] Available: https://doi.org/10.1109/ICPR.2002.1047437.

[33] Dalal N, Triggs B. Histograms of oriented gradients for human detection. In: 2005 IEEE computer society conference on computer vision and pattern recognition, vol. 1. IEEE; 2005, p. 886–93, [Online] Available: https://doi.org/10.1109/CVPR.2005.177.

[34] Guo S, Yang Z. Multi-channel-ResNet: An integration framework towards skin lesion analysis. Inform Med Unlocked 2018;12:67–74, [Online] Available: https://doi.org/10.1016/j.imu.2018.06.006.

[35] Simonyan K, Zisserman A. Very deep convolutional networks for large-scale image recognition. 2014, arXiv preprint arXiv:1409.1556.

[36] Rosebrock A. Deep learning for computer vision with Python: Starter bundle. PyImageSearch; 2017, p. 230–3.

[37] Liu L, Wang P, Lin J, Liu L. Intrusion detection of imbalanced network traffic based on machine learning and deep learning. IEEE Access 2020;9:7550–63, [Online] Available: https://doi.org/10.1109/ACCESS.2020.3048198.

[38] Kranthi Kumar K, Bharadwaj R, Ch S, Sujana S. Effective deep learning approach based on VGG-mini architecture for iris recognition. Ann Roman Soc Cell Biol 2021;4718–26.

[39] Wessels S, van der Haar D. Applying deep learning for the detection of abnormalities in mammograms. In: Inform Sci Appl. Springer; 2020, p. 201–10, [Online] Available: https://doi.org/10.1007/978-981-15-1465-4_21.

[40] Gezahegn YG, Gebreslassie AK, Hagos MA, Ibenthal A, Etsub EA. Classical machine learning algorithms and shallower convolutional neural networks towards computationally efficient and accurate classification of malaria parasites. In: International conference on information and communication technology for development for Africa. Springer; 2019, p. 46–56, [Online] Available: https://doi.org/10.1007/978-3-030-26630-1_5.

[41] LeCun Y, et al. 2015. LeNet-5, convolutional neural networks. 20:5. [Online] Available: http://yann.lecun.com/exdb/lenet.

[42] Wales DJ, Doye JP. Global optimization by basin-hopping and the lowest energy structures of Lennard-Jones clusters containing up to 110 atoms. J Phys Chem A 1997;101(28):5111–6, [Online] Available: https://doi.org/10.1021/jp970984n.

[43] Kohavi R, et al. A study of cross-validation and bootstrap for accuracy estimation and model selection. In: Ijcai. 14, (2):Montreal, Canada; 1995, p. 1137–45.

[44] Nuttall FQ. Body mass index: obesity, BMI, and health: a critical review. Nutri Today 2015;50(3):117, [Online] Available: https://doi.org/10.1097/NT.0000000000000092.

[45] Savva S, Tornaritis M, Savva M, Kourides Y, Panagi A, Silikiotou N, Georgiou C, Kafatos A. Waist circumference and waist-to-height ratio are better predictors of cardiovascular disease risk factors in children than body mass index. Int J Obesity 2000;24(11):1453, [Online] Available: https://doi.org/10.1038/sj.ijo.0801401.