A novel method to estimate Height, Weight and Body Mass Index from face images

Ankur Haritosh
Computer Science Department
Jaypee Institute of Information Technology
Noida, India
ankurharitosh@gmail.com

Ayush Gupta
Computer Science Department
Jaypee Institute of Information Technology
Noida, India
ayushgupta2959@gmail.com

Ekam Singh Chahal
Computer Science Department
Jaypee Institute of Information Technology
Noida, India
ecsingh1997@gmail.com

Ashwin Misra
Mechanical Department
Delhi Technological University
New Delhi, India
ashwin_bt2k16@dtu.ac.in

Satish Chandra
Computer Science Department
Jaypee Institute of Information Technology
Noida, India
satish.chandra@jiit.ac.in

***Abstract*—Body Mass Index (BMI) is the most commonly used tool to evaluate an individual's health. It is used to classify a person as underweight, healthy weight, overweight or obese. BMI is co-related with body fat and is a vital indicator of possible diseases that can transpire with higher body fat ranges. Higher body fat is prevalent these days with a higher calorie diet and a low physical activity lifestyle. On the other end of the spectrum, Adult malnutrition is more common and widespread than we are conscious of these days. The BMI can be used as a measure of adult nutritional status, both of individuals and of communities. Given that people have less time in their busy life and most people dont own a weighing machine and/or a measuring tape, we propose a time and cost efficient method of estimating Height, Weight and BMI from a persons face. In this paper, we propose a novel model using Convolution Neural Networks (CNN) and Artificial Neural Networks (ANN). We start by detecting the face from an image using the Viola-Jones algorithm. The image is fed to the Feature Extractor model. The extracted features are passed to an Artificial Neural Network (ANN) model which gives the predicted Height, Weight and BMI values. We have evaluated our model on the Reddit-HWBMI dataset and Face-to-BMI dataset. We propose a novel dataset, the Reddit-HWBMI dataset which contains 982 subjects with their corresponding Height, Weight, BMI, Gender and Age. The best performance for BMI was given by the XceptionNet model when used as a Feature Extractor. The XceptionNet also performed best for weight, whereas VGG-Face (Resnet model) performed slightly better than XceptionNet for height.**

***Index Terms*—Convolution Neural Networks; Deep Learning; Body Mass Index; XceptionNet**

I. INTRODUCTION

The body mass index or BMI processes the ratio between height and weight. The body mass index is the most basic tool which we use to define overweight and obesity. BMI is commonly regarded as a vital indicator of health. A normal BMI is between 18 and 25 and obesity starts at 30. With the gradual increase of body mass index, we notice a higher probability of cardiovascular diseases such as high blood pressure, diabetes, etc. A number in higher ranges leaves an individual at an exponential risk in health that the person has in addition to their BMI being raised, a waist circumference more than 40 inches. Therefore, classifies them into a greater risk category, so imperatively puts them up a risk category.

Weight issues are directly linked to all chronic diseases. If things proceed the way they are now, it is believed that the next generation may not live as long this current generation. The diet of an average overweight individual is one of the causes of this situation which contains various high-calorie fast food items. Even though after an individual eats these meals and gains energy, the energy lasts a very short period of time and causes hunger much sooner. The second reason is physical inactivity. The increased usage of the latest technology such as television, mobile phones, video games, etc has significantly reduced the physical activity of the average person. The third reason is stress. We are much less likely to exercise with the high amounts of stress and low sleep levels in our busy lives. Additionally, weight gain can be promoted, possibly due to hormonal and metabolic changes. One more reason may be due to the environment we live in and our transportation system, which is overly dependant on cars. In many regions, there is an insufficiency of any form of public transportation systems, sidewalks, community parks, playgrounds and bike paths for recreational activities.

On the other end of the spectrum, Adult malnutrition is more common and widespread than we are conscious of these days. Individuals obtaining a BMI value below 18 are considered underweight. An issue in absorbing nutrients from food or consuming an inadequate diet is the root cause of malnutrition. The reasons for this can be many, consisting of having a low income, a long-term health condition or reduced mobility, etc. This results in a low mood, feeling

978-1-7281-3591-5/19/$31.00 ©2019 IEEE.

tired all the time, weak muscles, slow or impaired growth, etc. Another major concern is that people that are overweight can be undernourished if they consume a high-calorie diet but are low in other essential nutrients.

It becomes really difficult for a common individual to measure their BMI values given that people have less time in their busy life and most people dont own a weighing machine and/or a measuring tape. In order to involve more people to measure their BMI, we propose an exciting measuring process and therein spread more awareness. In this paper, we propose a novel method to calculate the BMI of an individual from their face by the use of Deep Learning models. The prediction of deep learning model is largely considered as a black box process. Some reasons to explain the results can be due to the changing of facial features at varied weight ranges[1]. Overweight individuals have widened mid and lower faces, widened nose and a reduced eye height relatively to underweight individuals, who have an angular face with a pointed chin and relatively narrower cheeks. In addition to calculating the BMI, the height and weight of an individual is also calculated in this paper. This can be helpful for gaining knowledge of the physical appearance of a person, which otherwise may not be possible.

II. RELATED WORK

A study by Wen and Guo [2] obtained MAE for BMI in the range from 2.65-4.29 for different ethnic categories. The study was based on the MORPH-II dataset. Initially, a face and two eyes were detected from an image. Then, they applied normalization on the detected face using the eye co-ordinates and ASM(active shape model) fitting for the purpose of detecting the number of fiducial points in face images. Feature normalization was applied to the seven facial features extracted after ASM fitting. Finally, they predicted BMI using statistical learning which consists of Support Vector Regression (SVR), Least squares regression and Gaussian process regression.

In the study by Enes Kocabey et al[3], the BMI prediction system (Face-to-BMI) is composed of two stages: deep feature extraction and training of a regression model. For feature extraction, they had used VGG-16 and VGG-Face and extracted features after the fc6 layer. For the BMI regression, they have used an epsilon support vector regression model. The VGG-Face model had outperformed the VGG-16 which may be due to VGG-Face being trained on facial images only. The Pearson r correlations on the test set came out to be 0.65. They have evaluated their model on the Face to BMI dataset which consists of 4206 subjects and is readily available online.

In the study by Antitza Dantcheva, Franois Bremond and Piotr Bilinski[4], they have used Resnet-50 for feature extraction on the VIP attribute dataset which contains images of 1026 subjects. Resnet-50 has consistently performed well due to skip connections in its architecture which solves the problem of vanishing gradients. Afterward, they use smooth l1 regression and replaced 1000d fc with 1d fc. Therein, predicted the weight, height and BMI of an individual. The MAE for BMI prediction was 2.36+0.006, MAE(in m) for height was 8.2 and MAE(in Kg) for weight was 8.51 for all participants.

In the study by C.Vishnu Prasad and Dr.D.Gladis[5], they focus on Body Volume Index (BVI) which can find about the fat contents present in the different parts of the human body whereas BMI only takes height and weight into consideration. Their dataset consists of 250 subjects, of which 150 are male and 100 are female and contains their corresponding height and weight values.. Then, an Artificial neural network(ANN) is used for the training process.

In the study by Karin Wolffhechel et al[6], they have tested if a Principal component Analysis (PCA) model can estimate the variation of BMI. Additionally, they have tested if a PCA model that takes face shape and color information together into account is superior than a model that considers them separately. Their dataset contains of 526 subjects with their corresponding BMI values. They had implemented a data-driven method by the use of principal components to build statistical models where objectively defined shape and color characteristics in face images derive the principal components. A linear regression model was used to estimate BMI values. They obtained a result in the range for mean amount of variance in BMI of 5.1% to 8.8%.

In another study by Bum Ju Lee, Boncho Ku, Jun-Su Jang, and Jong Yeol Kim[8], proposed a method for classifying the normal weight or overweight using speech signals in age-and gender-specific groups. They had collected speech signals of 1830 individuals from several hospitals and hence, extracted 65 speech features. The machine learning algorithm of logistic regression was used which produced an accuracy ranged from 60.4 to 73.8% and AUC (Area under ROC curve) from 0.628 to 0.738.

III. DATASET

For this study, we have evaluated our model on two datasets. We received access to the Face to BMI Dataset on request at their page http://face2bmi.csail.mit.edu. The dataset contains face images of 4206 subjects, of which are 2438 male and 1768 are female. They used the set of annotated images from

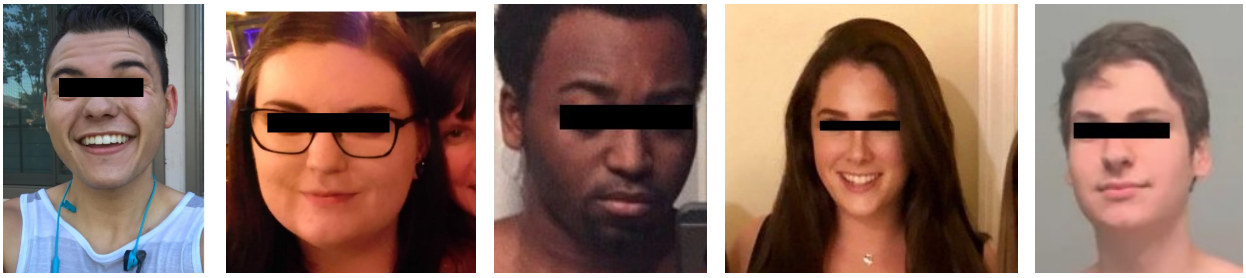

Fig. 1. Example images of 5 individuals in the Reddit-HWBMI dataset. The dataset contains unedited images unlike those shown in the figure. In order to respect individual privacy, we have added black bars.

TABLE I
BMI DISTRIBUTION OF FACE-TO-BMI DATASET

BMI Ranges	No. of Individuals
16<BMI<=18.5	7
18.5<BMI<=25	680
25<BMI<=30	1151
30<BMI<=35	941
35<BMI<=40	681
40<BMI	746

TABLE II
BMI DISTRIBUTION OF REDDIT-HWBMI DATASET

BMI Ranges	No. of Individuals
BMI<=18.5	7
18.5<BMI<=25	241
25<BMI<=30	283
30<BMI<=35	193
35<BMI<=40	120
40<BMI	138

the VisualBMI project which were collected from Reddit posts. The BMI distribution is depicted in Table I.

As the Face to BMI dataset doesnt contain height and weight measurements of an individual, we created Reddit-HWBMI Dataset. The dataset contains 982 subjects, of which 600 are male and 382 are female. Example images of the dataset are shown in figure 1. The Reddit-HWBMI Dataset distribution can be depicted in Table II, Table III, Table IV and Table V. We manually collected images from the progresspics subreddit. The reddit posts contain before and after body transformation images of an individual. The Reddit-HWBMI Dataset is available on request at ankurharitosh@gmail.com.

TABLE III
HEIGHT DISTRIBUTION OF REDDIT-HWBMI DATASET

Height Ranges (m)	Male	Female	All
1.6<H<=1.7	32	166	198
1.7<H<=1.8	154	44	198
1.8<H<=1.9	152	74	226
1.9<H<=201	346	14	360

TABLE IV
WEIGHT DISTRIBUTION OF REDDIT-HWBMI DATASET

Weight Ranges (kg)	Male	Female	All
W<=60	2	15	17
60<W<=85	219	93	312
85<W<=110	240	105	345
110<W<=140	121	58	179
140<W<=170	60	21	81
170<W	42	6	48

IV. EXPERIMENTAL SETUP

We have implement our approach in the Keras deep learning framework (https://github.com/keras-team/keras). For training our model, we are using a single NVIDIA GeForce 1060 with 6 GB memory, i7 8th generation processor and 16GB RAM.

TABLE V
MEAN AND STANDARD DEVIATION FOR MALES AND FEMALES ON THE REDDIT-HWBMI DATASET

	Male	Female
Individuals	1.87 ($\sigma = 1.48$)	1.76 ($\sigma = 0.63$)
Height (m)	110.61 ($\sigma = 4.41$)	99.57 ($\sigma = 3.57$)
Weight (kg)	31.35 ($\sigma = 17.28$)	32.09 ($\sigma = 9.14$)

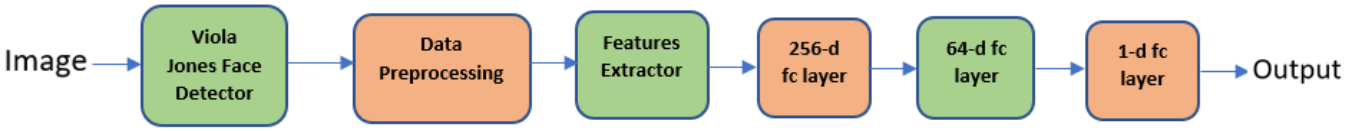

Fig. 2. Block diagram of sequence of steps in our model to predict height, weight and BMI

V. METHODOLOGY

A. Face Detection

In the Face-to-BMI and Reddit-HWBMI datasets, both contain images with close-up shots of individuals. But to make sure our model only considers the facial features of an individual, we apply facial detection beforehand. For this purpose, we use the commonly used Voila-Jones Face Detection[9] algorithm using OpenCV[10].

B. Data Preprocessing

The detected faces are cropped and saved to be images of size 256X256. Additionally, the per-pixel mean has been subtracted for each of the two datasets. Then, a 224X224 crop is randomly sampled from each image or its horizontal flip[11].

C. Feature Extractor

These images are fed to a feature extraction model to extract high-level features before it is put into our custom Artificial Neural Network (ANN) model. We have experimented with various Convolution Neural Network (CNN) models to be used as a feature extractor. Of all the models that we had experimented with, the XceptionNet[12] performed the best for BMI and weight values whereas VGG-Face[13] (with Resnet-50[11] architecture) performed best for height values. The XceptionNet is pre-trained on the ImageNet Large-Scale Visual Recognition Challenge (ILSVRC) 2014 [14] and their weights are readily available. The VGG-Face(Resnet-50 model) is Resnet-50 architecture pre-trained on the VGG-Face dataset.

The better performance of the XceptionNet can be explained by its structure of separating the learning of channel-wise and space-wise features to its logical extreme. Its architecture consists of a pointwise convolution, 1X1 convolutions that computes features which mix together information from the input channels and a depthwise convolution which is a spatial convolution where every input channel is handled separately.

D. Artificial Neural Network

The features extracted from the feature extraction model are fed into an Artificial Neural Network model consisting of 3 layers. The three layers are 256-d fc, 64-d fc and 1-d fc layer respectively. The output is a 1 dimensional floating point value. The loss function used is smooth L1 Loss for this regression problem. Afterward, we fine-tune the finer strides of our model using the Face-to-BMI dataset and Reddit-HWBMI dataset separately.

VI. IMPLEMENTATION DETAILS

The Train-Test split for the Face-to-BMI dataset is the same as in the Face-to-BMI paper. The dataset is divided into 3368 training images and 838 test images with the same individual not occurring in both train and test sets. The Reddit-HWBMI dataset is split into 786 training images and 196 test images. The training set includes 546 male and 240 female images while the test set contains 138 male and 58 female images. The optimal size of layers in the Artificial Neural Network model and the hyper-parameters are decided using the Hyperas library (https://github.com/maxpumperla/hyperas). The Hyperas library is a wrapper around Hyperopt[15] for use in Keras.

VII. RESULTS

This section presents the performances of our model with various feature extractors for Height, Weight and BMI on the Reddit-HWBMI dataset and for BMI on the Face-to-BMI dataset. For the evaluation metrics, we have taken Mean Absolute Error (MAE). MAE is calculated by taking the average of absolute errors predicted labels and actual labels for Height, Weight and BMI.

$$MAE = \frac{1}{N}\sum_{k=1}^{N}|\hat{Y_k} - Y_k| \tag{1}$$

where N is the total count of test images, $\hat{Y}_k$ is predicted labels and Y_k is actual labels for Height, Weight and BMI.

A. BMI

As per the results in Table V, the XceptionNet has performed the best on both the Reddit-HWBMI and Face-to-BMI dataset having MAE for BMI as 4.1 and 3.8 respectively. This is comparable to the results of Wen and Guo et al. [2], which were 2.654.29 and results of Antitza Dantcheva, Franois Bremond, Piotr Bilinski[5] which were 2.3+0.06. The better performances of models on the Face-to-BMI dataset can be explained as it has comparatively more training images.

TABLE VI
MEAN ABSOULUTE ERROR (MAE) VALUES OF BMI FOR VARIOUS FEATURE EXTRACTOR MODELS ON THE REDDIT-HWBMI AND FACE-TO-BMI DATASET

Feature Extractor	Reddit-HWBMI	Face-to-BMI
Xception	4.1	3.8
VGG-Face (Resnet50)	5.7	4.6
Resnet-50	6.4	4.8
VGG-Face	6.4	5.1
Inception Resnet-V2	6.2	5.4

TABLE VII
MEAN ABSOULUTE ERROR (MAE) VALUES OF HEIGHT FOR VARIOUS FEATURE EXTRACTOR MODELS ON THE REDDIT-HWBMI DATASET

Feature Extractor	MAE (in m)
Xception	0.074
VGG-Face (Resnet50)	0.073
Resnet-50	0.080
VGG-Face	0.085
Inception Resnet-V2	0.085

B. Height

As per the results in Table VI, the VGG-Face (Resnet50 model) has performed the best on the Reddit-HWBMI dataset for Height having MAE value of 0.073 with the XceptionNet coming as a close second having a value of 0.074. We outperform the results of Antitza Dantcheva, Franois Bremond, Piotr Bilinski[5], which were 0.077 for male, 0.078 for female and 0.082 for all participants with respect to MAE.

C. Weight

As per the results in Table VII, the Xception model has performed the best on the Reddit-HWBMI dataset for Weight having MAE value of 13.29 which is relatively small compared to the wide weight ranges. Our result is comparable to the results of Antitza Dantcheva, Franois Bremond, Piotr Bilinski[5] which were 8.51 with respect to MAE.

TABLE VIII
MEAN ABSOULUTE ERROR (MAE) VALUES OF WEIGHT FOR VARIOUS FEATURE EXTRACTOR MODELS ON THE REDDIT-HWBMI DATASET

Feature Extractor	MAE (in kg)
Xception	13.29
VGG-Face (Resnet50)	15.96
Resnet-50	17.73
VGG-Face	17.55
Inception Resnet-V2	17.64

VIII. SUMMARY

In this paper, we have presented a novel method to compute the Height, Weight and Body Mass Index of an individual. We have created our own dataset consisting of face images 982 individuals along with their height, weight and BMI. Additionally, we have also evaluated our model on the Face-to-BMI dataset which contains face images of 3368 individuals along with their BMI values. The model process starts with face detection on the preprocessed data using Voila-Jones algorithm. The images are fed to a feature extractor model. The extracted features are put into a custom Artificial Neural Network (ANN) which produces the predicted outputs for our regression problem.

The best performance for BMI was given by the XceptionNet model when used as a Feature Extractor with Mean Absolute Error (MAE) value of 4.1 and 3.8 for Reddit-HWBMI dataset and Face-to-BMI dataset respectively. The XceptionNet also performed best for weight having MAE(in kg) value of 13.29, whereas VGG-Face (Resnet model) performed slightly better than XceptionNet for height having MAE(in m) value of 0.073 on the Reddit-HWBMI dataset.

In our future work, we would work on implementing new ideas to improve our models performance. This paper was motivated by the need to create awareness for health amongst our society, which usually gets neglected in our busy lifestyle.

IX. REFERENCES

[1] C. Mayer, S. Windhager, K. Schaefer and P. Mitteroecker, "BMI and WHR are reflected in female facial shape and texture: a geometric morphometric image analysis", PloS one, vol. 12, no. 1, 2017

[2] L. Wen and G.-D. Guo, "A computational approach to body mass index prediction from face images," Image and Vision Computing, vol. 31,no. 5, pp. 392400, 2013.

[3] E. Kocabey, M. Camurcu, F. Ofli, Y. Aytar, J. Marin, A. Torralba, I. Weber, "Face-to-BMI: Using Computer Vision to infer Body Mass Index on Social Media." Proceedings of the International AAAI Conference on Web and Social Media (ICWSM), pp. 572-575, 2017.

[4] A. Dantcheva, F. Bremond, P. Bilinski, "Show me your face and I will tell you your Height, Weight and Body Mass Index.", International Conference on Pattern Recognition (ICPR), pp. 3555-3560, 2018

[5] C.V. Prasad, D.Gladis, " Assessing Body Volume Index Using Neural Networks " , International Journal of Computer Science and Information Technologies, Vol. 7 (4), pp. 1952-1954, 2016

[6] K. Wolffhechel, C. H. Amanda, H. Jarmer, C. I. Fisher, B. C. Jones, L. M. DeBruine, Testing the utility of a Data-Driven approach for assessing BMI from face images, PloS one, vol. 10, no. 10, 2015

[7] V. Coetzee,J Chen, DI Perrett, ID Stephen, "Deciphering faces: Quantifiable visual cues to weight. Perception", 39(1), pp.51-61, 2010

[8] B. J. Lee, B. Ku, J. S. Jang, J. Y. Kim, "A novel method for classifying Body Mass Index on the basis of Speech Signals for future clinical applications: A pilot study. Evidence-based complementary and alternative medicine", Evidence-Based Complementary and Alternative Medicine, 2013

[9]P. Viola and M. Jones, Robust real-time face detection, in Proceedings of IEEE International Conference on Computer Vision, pp. 137-154, 2001.

[10] G. Bradski, A. Kaehler, "Learning OpenCV: Computer Vision with the OpenCV Library", O'Reilly Media, Inc., 2008.

[11] K. He, X. Zhang, S. Ren, and J. Sun, "Deep residual learning for image recognition," in Proceedings of the IEEE conference on Computer Vision and Pattern Recognition, pp. 770778, 2016

[12] F. Chollet, "Xception: Deep Learning with Depthwise Separable Convolutions.", In Proceedings of the IEEE conference on Computer Vision and Pattern Recognition, pp. 1251-1258, 2017.

[13] O. M. Parkhi, A. Vedaldi, A. Zisserman,Deep Face Recognition ,British Machine Vision Conference, Vol. 1, No. 3, p. 6, 2015

[14] O. Russakovsky, J. Deng, H. Su, J. Krause, S. Satheesh, S. Ma, Z. Huang, A. Karpathy, A. Khosla, M. Bernstein et al., "Imagenet large scale visual recognition challenge", International Journal of Computer Vision, vol. 115, no. 3, pp. 211252, 2015.

[15] J. Bergstra, D. Yamins, and D. D. Cox., "Hyperopt: A python library for optimizing the hyperparameters of machine learning algorithms." In Proceedings of the 12th Python in science conference, pp. 13-20. 2013.