Face To BMI: A Deep Learning Based Approach for Computing BMI from Face

Jiten Sidhpura*
Department of Computer Engineering
Sardar Patel Institute of Technology
Mumbai, India
jiten.sidhpura@spit.ac.in

Rudresh Veerkhare*
Department of Computer Engineering
Sardar Patel Institute of Technology
Mumbai, India
rudresh.veerkhare@spit.ac.in

Parshwa Shah*
Department of Computer Engineering
Sardar Patel Institute of Technology
Mumbai, India
parshwa.shah@spit.ac.in

Prof. Surekha Dholay
Department of Computer Engineering
Sardar Patel Institute of Technology
Mumbai, India
surekha_dholay@spit.ac.in

***Abstract*—Body Mass Index (BMI) is a measure of how healthy a person is with respect to their body weight. BMI has shown a correlation with various factors like physical health, mental health, popularity. BMI calculation often requires accurate height and weight, which would take manual work to measure. Large-scale automation of BMI calculation can be utilized for analyzing various aspects of society and can be used by governments and companies to make effective decisions. Previous works have used only geometric facial features discarding other information, or a data-driven deep learning-based approach in which the amount of data becomes a bottleneck. We used the state of the art pre-trained models such as Inception-v3, VGG-Faces, VGG19, Xception and fine-tuned them on the comparatively large public dataset with discriminative learning. We used the larger Illinois DOC labeled faces dataset for training and Arrest Records, VIP_attribute for evaluation purposes.**

***Index Terms*—Body Mass Index prediction, Face To BMI, Deep Learning, Facial Features, Transfer Learning**

I. Introduction

The BMI(Body Mass Index) of any person is a crucial indicator of health. It checks if the person is underweight, normal, overweight, or obese. In the current scenario, health is one of the most neglected factor. Technology which has more benefits also has some drawbacks. It has made humans lazy and thus reduced their physical activity leading to a sedentary lifestyle and a rise in BMI which adversely affects their health and increases the risk of chronic diseases. The more the BMI, the more is the chance of developing cardiovascular and other harmful diseases. On the other side of the coin, some people have problems like malnutrition and deficiencies. So, BMI can help a person to keep a track record of their health.

According to [1], on average, one out of every three adults is obese, which is about 36% of the population, and by the year 2030, an estimated 20% of the global population would be obese.

Human faces carry a significant amount of information about a person. Recent studies have shown a strong correlation between the human face and the BMI of the person. The people with skinny faces have chances of less BMI and vice versa. Generally, obese people tend to have the middle and lower part of the face wider. It is difficult for the person to calculate BMI if they do not have a measuring tape and weighing machine. Recently there have been many advancements in deep learning where models can extract meaningful features from the images. By utilizing these methods, we can predict the BMI from human faces. So, In this paper, we have proposed a technique to predict BMI from human faces. This system could help health insurance companies to maintain the health records of their customers. Also, the government could track the health records of a particular region and devise policies accordingly.

A. Our Contribution

To improve the facial BMI prediction using a large dataset. The contributions of this paper are as follows:

- To find a correlation between BMI and human faces and to create a technique to predict BMI from human faces using deep learning and transfer learning models such as VGG-Face, Inception-v3, VGG19 and Xception. Outline of our approach is shown in Fig. 1
- Used 3 publicly available datasets containing images of Hollywood celebrities and prisoners. Preprocessing the dataset by aligning the faces to the center. Followed by applying our proposed approach to these datasets and improving the Mean Absolute Error(MAE) score.

The rest of the paper includes the following sections: Section II describes the Literature Survey, where we discussed the existing methodologies. In section III, we have proposed our methodology. Followed by the information on datasets, experimental setup, and evaluation of our methodology in section IV referred to as Experiments. In section V, we have presented our results, followed by a conclusion in section VI.

* These three authors contributed equally to this work

978-1-6654-0670-3/22/$31.00 ©2022 IEEE

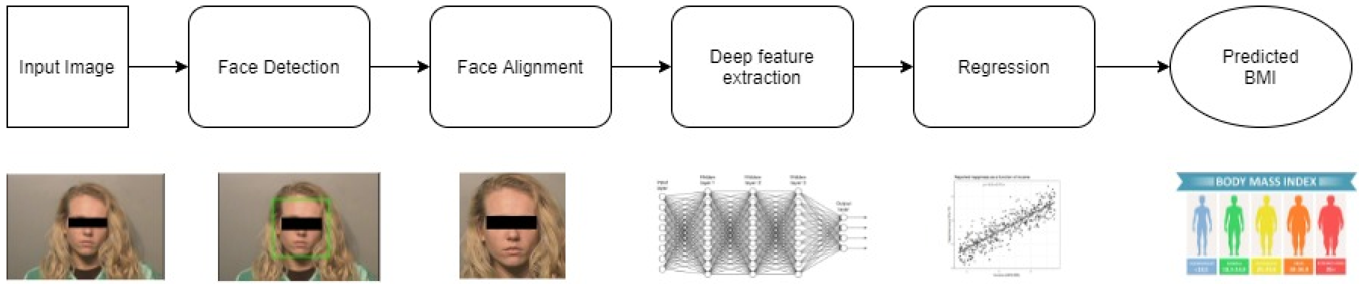

Fig. 1. Flow Diagram

II. Literature Survey

Wen and Guo [2] used the computational approach to calculate the BMI. The authors have extracted facial landmarks from facial images using the Active Shape Model and further processed them to extract seven facial features. These seven features are ES (Eye Size), CJWR (Cheek to Jaw Width Ratio), PAR (Perimeter to Area Ratio), WHR (Width to Upper facial height ratio), FW/FH (lower Face to Face Height Ratio), and MEH (Mean Eyebrow height). Then used Support Vector Regressor (SVR), Least Square Estimation, and Gaussian Process to solve the regression problem. They used Morph II dataset to train and evaluate. Results showed that SVR gave the best results for both sets of data. Using a similar approach, Barr et al. [3] proposed the calculation of Facial BMI (fBMI) from facial images. For evaluation, the correlation between fBMI and BMI showed more correlation for normal and overweight categories but less accuracy for underweight and obese classes. Although results are significant, only facial landmarks are considered for feature extraction. Further improvements can be made by extracting more features from facial images.

Kocabey et al. [4] proposed a computer vision technique to predict a person's BMI from their social media images. They took images from VisualBMI Project. For deep feature extraction, they had used 4206 facial images on VGG-Net and VGG-Face models. For BMI regression, they have used an epsilon support vector regression model. VGG-Faces performed better than the VGG-Net model. The Pearson correlation coefficient on the test set given by the VGG-Face model came out to be 0.71, 0.57, 0.65 for the Male, Female and Overall category respectively. They also presented human vs. machine prediction where humans outperformed the machine in lower BMI and no difference in higher BMI categories.

Ankur Haritosh et al. [5] proposed a novel method to estimate height, weight, and BMI from facial images. They used 4206 images from the VisualBMI project and 982 from the Reddit HWBMI dataset. After applying Voila Jones Face Detection algorithm, the images are cropped to the size 256×256. These images are given to the feature extractor model to extract high-level features and then to the 3-layered ANN model. Taking Face to BMI and Reddit HWBMI dataset, the MAE for BMI by XceptionNet was 4.1 and 3.8, respectively. The MAE was 0.073 for height given by VGG-Face and 13.29 for weight by the XceptionNet model on the Reddit HWBMI dataset.

Christine Mayer et al. [6] proposed a statistical approach to find the relationship of body mass index (BMI) and waist-to-hip ratio (WHR) with facial shape and facial texture. The authors marked 119 anatomical landmarks and semilandmarks with the help of Windows Program TPSDig. They used a sliding landmark algorithm to compute semilandmark's exact positions. Their study consisted of 49 standardized images of women, where BMI was in the range of [17.0, 35.4] and WHR in [0.66, 0.82]. The Procrustes shape coordinates represent the shape of the face, RGB values of the standardized images represent its texture. The multivariate linear regression technique assessed the desired relationships. The BMI was predictable than WHR from facial attributes, 25% of the variation through facial shape and 3-10% through facial textures.

Jiang et al. [7] analyzed geometry-based and deep learning-based approaches for computing visual BMI, authors also discussed the effect of different factors like gender, ethnicity, and head pose on the accuracy of predictions. Deep learning-based approaches gave better results than geometry-based, but the high dimensionality of features has a negative effect as training data is relatively lesser. Also, large head pose changes lead to lower performance. Morph II dataset is used along with FIW-BMI dataset created by authors from social media websites.

Hera Siddiqui et al. [8] proposed a custom end-to-end CNN network to predict BMI. The authors also extracted features from the facial images with the help of pre-trained CNN models such as VGG-19, ResNet, DenseNet, MobileNet, and LightCNN passed them to Support Vector Regressor (SVR) and Ridge Regression (RR) for final predictions. They used VisualBmi, VIP_attribute, and Bollywood Datasets and achieved Mean Absolute Error (MAE) in the range of [1.04, 6.48]. DenseNet and ResNet models gave better performance when used with Ridge Regression. Pretrained Models gave slightly better performance than the end-to-end CNN model.

III. Methodology

A. Data Preprocessing

We are using front-facing images from the dataset. However, some images have tilted head position and are inconsistent with respect to amount of zoom. To make images similar, we used the preprocessing by StyleGan FFHQ Dataset [10], using the DLIB 68 landmark detection [11] model to align the face vertically and then blurred the surroundings while focusing on the face. For training on Tensor Processing Units (TPU) we converted our dataset into Tensorflow Records Dataset (TFRecord) format. Each record consisted of 1024 images and was resized to $256 \times 256 \times 3$, preprocessed according to the backbone network used for transfer learning. The label for the image is the BMI value.

B. Transfer Learning

BMI calculation from facial images is rather complicated so learning all required features from relatively small datasets would be infeasible. Many tasks in computer vision have used transfer learning to boost performance and reduce the training time. Hence we have used state of the art pre-trained models such as:

- **Inception-v3 [12]:** Inception-v3 is a deep convolutional neural network. It is the 3rd edition model of the Inception CNN developed by Google. This model is trained on more than a million images from the popular ImageNet database. It gave a Top-1 accuracy score of 0.779 and a Top-5 accuracy score of 0.937 with around 24 Million parameters only and thus computationally effective when compared to other models.
- **VGG-Face [13]:** VGG-Face model is developed by the researchers of the Visual Geometry Group (VGG) at Oxford. This model uses the VGGFace2 [14] dataset that consists of 3.31 million facial images of 9131 subjects. It was created with the primary intention of training robust face recognition models. We used a pre-trained Resnet-50 model of this dataset for our study.
- **VGG-19 [15]:** VGG-19 model is a deep convolutional neural network, a successor of the VGG-16 model and developed by the researchers of the Visual Geometry Group. Like the Inception-v3 model, it is trained on the popular ImageNet database. It achieved a Top-1 accuracy score of 0.752 and a Top-5 accuracy score of 0.925 with around 143 million parameters.
- **Xception [16]:** Feature extraction base of the Xception model is formed by 36 convolutional layers. It is an extension of the Inception model architecture where it uses depthwise Separable Convolutions in place of the standard inception modules. It was also trained on ImageNet Database and achieved a Top-1 accuracy score of 0.79 and a Top-5 accuracy score of 0.94 with around 22 million parameters.

C. Training

We used the same fully connected layers at the end of all pre-trained models. The layers, we added at the end are shown in Fig. 2. Before feeding the output of the pre-trained model into the fully connected layers, we perform Global Average Pooling [17]. To prevent overfitting, we added one dropout layer [18] with a dropout of 50% into our model architecture. We also used Gaussian Error Linear Unit (Gelu) [19] as an activation function, it combines the properties of the RELU activation function, Dropout, and Zoneout. Due to this, it tends to generalize better when there is more noise in the data so we used it in our models. As we found a comparatively larger dataset for our study we also fine-tuned our models. In deep convolutional networks, layers near the input learn basic features such as edges and corners. As we move towards the output, the layers generally learn advanced features from the images used for training it. We used a higher learning rate for the new fully connected layers and a much lower learning rate for some of the final layers of the pre-trained model so that it extracts more features from the images. Due to these reasons, we used Adam optimizers [20] with decreasing learning rates as we move to deeper layers of the model. We implemented it with the help of the MultiOptimizer [21], [22] provided by TensorFlow Addons.

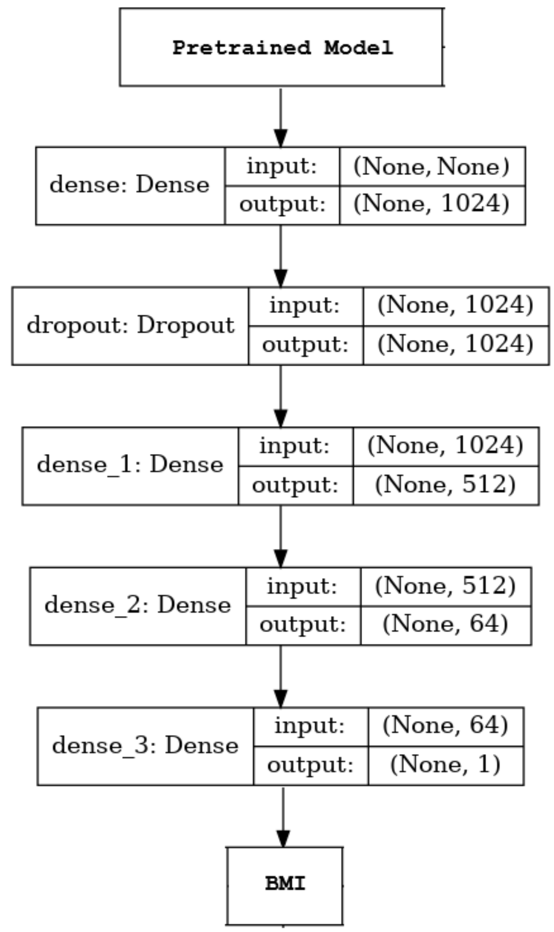

Fig. 2. Model Architecture

IV. EXPERIMENTS

A. Dataset

We evaluated our proposed methodology on three publicly available datasets such as.

- **Illinois DOC labeled faces Dataset [23]:** The source of this dataset is the Illinois Dept. of Corrections. This dataset comprises frontal and side views of 68149 prisoners and has additional information such as gender, height (in inches), weight (in lbs), and date of birth. There were 1365 corrupt images and 7309 images that did not have the height and weight information. They were not included in our research study. The dataset in the end we used had 56200 males and 3649 females with a mean BMI equal to 27.88 and a standard deviation of 5.20. We used this dataset for traning and validation of our models. The sample images from this dataset are shown in Fig. 3.
- **Arrest Records Database [24]:** We had found a dataset of prisoners with their names, height, weight, BMI, and other attributes. It has a total of 1543 images which contained 1243 male and 300 female subjects. The BMI value ranges from 15 to 56. The mean BMI value of the dataset is 26.41, and the standard deviation of BMI values is 5.27. The dataset has an imbalance where male prisoners are four times more than female counterparts. We used this dataset as a test set for our model. The sample images from this dataset are shown in Fig. 4.

Fig. 3. Illinois DOC labeled faces Dataset

Fig. 4. Arrest Records Database

- **VIP-Attribute Dataset [9]:** The VIP Attribute dataset is collected from the world wide web and consists of 513 male and 513 female subjects(mainly singers, actors, and athletes). The mean BMI value of the dataset is 23.04, and the standard deviation of BMI values is 4.24. The BMI values range from 16 to 55. The images had the frontal pose of subjects. There were various covariates like illumination, expression, image quality, and resolution. The authors obtained the height and weight of people from different websites and calculated BMI for making the dataset. Since this is a small dataset, we used it for evaluation purposes only. The sample images from this dataset are shown in Fig. 5

Fig. 5. VIP Attributes Dataset

B. Experimental Setup

We used a Kaggle notebook with RAM of 16GB and ROM of 78GB along with the TPUv3-8 accelerator. With Python 3.7.10, Tensorflow version 2.4.1, Keras version 2.4.0, and batch size of 128. To prevent overfitting, Early Stopping Callback is used to stop training if validation loss does not improve for 5 consecutive epochs.

C. Evaluation Metrics

For our research study, we used Mean Squared Error (MSE) as our regression loss function. As given in (1), it is calculated by computing the summation of squared differences of actual value (y_i) and the predicted value (x_i), divided by the total number of samples (n). Because of squaring the difference, this loss function penalizes the model more heavily.

$$MSE = (\frac{1}{n})\sum_{i=1}^{n}(y_i - x_i)^2 \tag{1}$$

We used Mean Absolute Error (MAE) (2) to evaluate our model's performance. It is calculated by computing the summation of absolute values of differences of actual value (y_i) and predicted value (x_i), divided by the total number of samples (n).

$$MAE = (\frac{1}{n})\sum_{i=1}^{n}|y_i - x_i| \tag{2}$$

V. Results

In this section, we have discussed the performances of all the models on each dataset. For comparing our models there were 2 metrics namely MSE and MAE, but we have used MAE. For Tables I, II, III, the MAE-Overall represents the MAE on the complete dataset, MAE-Male and MAE-Female represent the total MAE for males and females for

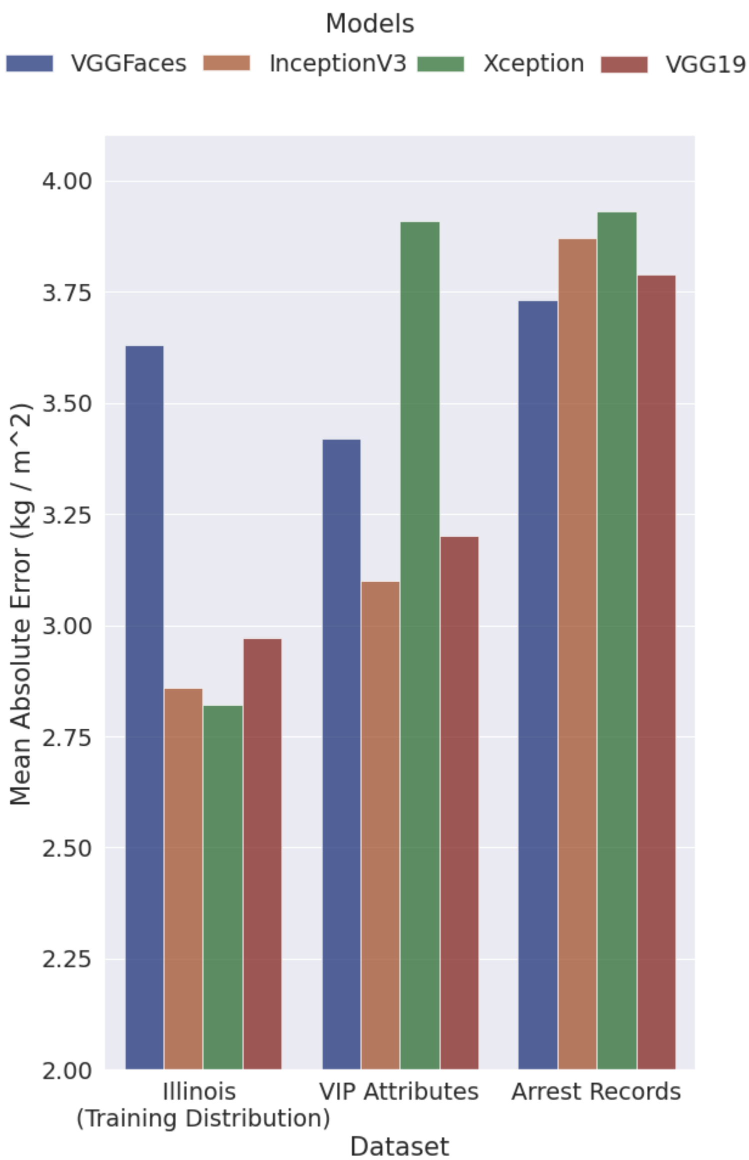

Fig. 6. Overall MAE on all datasets
x-axis: Dataset Names, y-axis: Mean Absolute Error(MAE)

the respective dataset. Fig. 6 represents the graph showing dataset names on the x-axis and Mean Absolute Error (MAE) in kg/m^2 on the y-axis. The four colors represent four transfer learning models.

Table I represents the performance on our training distribution, Illinois DOC labeled faces dataset. The MAE-Overall score for this dataset was in the range of [2.82, 3.63]. The MAE-Male and MAE-Female were in the range of [2.79, 3.61] and [3.54, 4.04] respectively. Xception model gave the best scores on all 3 cases, Inception-v3 gave near about the same results. For all models, the MAE-Male value is slightly lower than the MAE-Overall, the MAE-Female is greater than the MAE-Overall. We saw that for all models, the error rate in females is slightly higher than in the case of males. The differences in error are due to the imbalanced nature of the dataset.

TABLE I
MAE RESULTS ON ILLINOIS DOC LABELED FACES DATASET FOR BMI INFERENCE

	MAE-Overall	**MAE-Male**	**MAE-Female**
VGGFaces	3.63	3.61	3.93
Inception-v3	2.86	2.83	3.59
Xception	2.82	2.79	3.54
VGG19	2.97	2.89	4.04

TABLE II
MAE RESULTS ON VIP_ATTRIBUTE DATASET FOR BMI INFERENCE

	MAE-Overall	**MAE-Male**	**MAE-Female**
VGGFaces	3.42	4.16	2.68
Inception-v3	3.10	3.04	3.17
Xception	3.91	2.78	5.03
VGG19	3.20	3.33	3.07

Table II represents the performance on our testing dataset, VIP_attribute dataset. The MAE-Overall score for this was in the range of [3.10, 3.91]. The MAE-Male and MAE-Female were in the range of [2.78, 4.16] and [2.68, 5.03] respectively. The Inception-v3 model gave the best scores on all 3 cases with very less differences. After Inception-v3, VGG19 performed the best with near about the same results for all 3 cases.

TABLE III
MAE RESULTS ON ARREST RECORDS DATASET FOR BMI INFERENCE

	MAE-Overall	**MAE-Male**	**MAE-Female**
VGGFaces	3.73	3.35	5.09
Inception-v3	3.87	3.57	4.68
Xception	3.93	3.56	5.11
VGG19	3.79	3.75	3.99

Table III represents the performance on our testing dataset, Arrest Records dataset. The MAE-Overall score for this was in the range of [3.73, 3.93]. The MAE-Male and MAE-Female were in the range of [3.35, 3.75] and [3.99, 5.11] respectively. Although VGGFaces gave a better MAE-Overall score, VGG-19 generalized better because of its lesser MAE-Female score.

We observed the same pattern of the MAE-Male value being slightly lower than the MAE-Overall and the MAE-Female being higher than the MAE-Overall like in Table I. This is maybe due to the resemblance of the distribution of images with the Illinois DOC labeled face dataset.

VI. Conclusion

We observed that people with more BMI have a higher risk of developing health issues. We found that there exists a strong association between BMI and the face of a human. So, we proposed an approach to predict BMI from facial images using deep learning. We used 3 publicly available datasets of diverse domains containing images of prisoners and Hollywood celebrities to evaluate our model. We preprocessed the facial data by aligning the faces to the center using the Dlib 68 landmark detection algorithm. For faster processing, we used TPU and created Tensorflow Records for our images of the dataset.

As a part of future work, a more robust model may be obtained by training on a balanced dataset of people of different countries, ethnicity, and age. Federated Learning can be used to train a model on images that are not available publicly. We hope that this study assists companies and government and also help people to be aware of their BMI and maintain their health accordingly.

References

[1] SingleCare Team, “Overweight and obesity statistics 2021” singlecare.com, Jan. 21, 2021. [Online]. Available: https://www.singlecare.com/blog/news/obesity-statistics/. [Accessed Aug.]

[2] Lingyun Wen, Guodong Guo, ”A computational approach to body mass index prediction from face images”, Image and Vision Computing, Volume 31, Issue 5, 2013, Pages 392-400, ISSN 0262-8856.

[3] M. Barr, G. Guo, S. Colby, M. Olfert, Detecting body mass index from a facial photograph in lifestyle intervention, Technologies 6 (3) (2018) 83

[4] 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.

[5] A. Haritosh, A. Gupta, E. S. Chahal, A. Misra and S. Chandra, ”A novel method to estimate Height, Weight and Body Mass Index from face images,” 2019 Twelfth International Conference on Contemporary Computing (IC3), 2019, pp. 1-6, doi: 10.1109/IC3.2019.8844872.

[6] Mayer C, Windhager S, Schaefer K, Mitteroecker P (2017) BMI and WHR Are Reflected in Female Facial Shape and Texture: A Geometric Morphometric Image Analysis. PLoS ONE 12(1): e0169336. doi:10.1371/journal.pone.0169336.

[7] Jiang, Min & Shang, Yuanyuan & Guo, Guodong. (2019). On Visual BMI Analysis from Facial Images. Image and Vision Computing. 89. 10.1016/j.imavis.2019.07.003.

[8] H. Siddiqui, A. Rattani, D. R. Kisku and T. Dean, ”AI-based BMI Inference from Facial Images: An Application to Weight Monitoring,” 2020 19th IEEE International Conference on Machine Learning and Applications (ICMLA), 2020, pp. 1101-1105, doi: 10.1109/ICMLA51294.2020.00177.

[9] A. Dantcheva, P. Bilinski, F. Bremond, ”Show me your face and I will tell you your height, weight and body mass index,” Proc. of 24th IAPR International Conference on Pattern Recognition (ICPR), (Beijing, China), August 2018.

[10] Karras, Tero & Laine, Samuli & Aila, Timo. (2019). A Style-Based Generator Architecture for Generative Adversarial Networks. 4396-4405. 10.1109CVPR.2019.00453.

[11] Davis E. King. Dlib-ml: A Machine Learning Toolkit. Journal of Machine Learning Research 10, pp. 1755-1758, 2009

[12] Szegedy, Christian & Vanhoucke, Vincent & Ioffe, Sergey & Shlens, Jon & Wojna, ZB. (2016). Rethinking the Inception Architecture for Computer Vision. 10.1109/CVPR.2016.308.

[13] Parkhi, O. M., et al. 2015. Deep face recognition. In British Machine Vision Conference.

[14] Cao, Qiong & Shen, Li & Xie, Weidi & Parkhi, Omkar & Zisserman, Andrew. (2018). VGGFace2: A Dataset for Recognising Faces across Pose and Age. 67-74. 10.1109/FG.2018.00020.

[15] K. Simonyan and A. Zisserman, “Very deep convolutional networks for large-scale image recognition,” CoRR, vol. abs/1409.1556, 2014.

[16] 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.

[17] Lin, Min & Chen, Qiang & Yan, Shuicheng. (2014). Network In Network. 10.

[18] Srivastava, Nitish & Hinton, Geoffrey & Krizhevsky, Alex & Sutskever, Ilya & Salakhutdinov, Ruslan. (2014). Dropout: A Simple Way to Prevent Neural Networks from Overfitting. Journal of Machine Learning Research. 15. 1929-1958.

[19] Hendrycks, D., Gimpel, K.: Bridging nonlinearities and stochastic regularizers with gaussian error linear units. CoRR abs/1606.08415 (2016)

[20] Kingma, Diederik & Ba, Jimmy. (2014). Adam: A Method for Stochastic Optimization. International Conference on Learning Representations.

[21] Howard, Jeremy & Ruder, Sebastian. (2018). Universal Language Model Fine-tuning for Text Classification. 328-339. 10.18653/v1/P18-1031.

[22] Jin, Xiaojie & Chen, Yunpeng & Dong, Jian & Feng, Jiashi & Yan, Shuicheng. (2016). Collaborative Layer-Wise Discriminative Learning in Deep Neural Networks. 9911. 10.1007/978-3-319-46478-7_45.

[23] David J. Fisher. (2019, December). Illinois DOC labeled faces dataset, Version 1. Retrieved March 28, 2021 from https://www.kaggle.com/davidjfisher/illinois-doc-labeled-faces-dataset

[24] Domestic Violence Database, “National Violent Offender & Domestic Violence Registry” domesticviolencedatabase.net. [Online]. Available: https://domesticviolencedatabase.net/registry_ia/?co=Polk&abc=R#abc. [Accessed August. 12, 2020].