ELEN4002: Digital Estimation of Body Mass Index

Darrion Singh (1056673)

***Abstract*—The following reports the findings of a study that was used to design and implement an application used to estimate Body Mass Index from a photograph. The total duration of the study, design and implementation of the application was six weeks. The application had a layer that performed various feature extraction on the images, and a layer that estimated BMI using machine learning. The dataset used to train the machine learning layer consisted of 465 anonymised people's photographs. The program detected a reference object of known dimensions located in each photograph, and subsequently estimated the dimensions of the person in the photograph. Dimensions were chosen at the location of common body fat deposits and along with person's estimated silhouette area and estimated height, were used to train three Multilayer Perceptron (MLP) models that estimated BMI – one for the front photograph, another for the side photograph, and finally, one that compensated for errors between the front and side estimations, and was the final output of the program. The results of the final model BMI estimation showed a mean estimation accuracy ranging between 79% and 92% for the various categories of BMI, namely Underweight (< 18.5), Healthy (18.5 – 25), Overweight (25 – 30), or Obese (> 30). Furthermore, the spread of error showed that 60.6% of participants had an error less than 10%, and 89% of participants had an error less than 25% with a mean absolute error of ***, and standard deviation of ***.**

I. INTRODUCTION

The Body Mass Index (BMI) is a screening tool used to indicate whether one's body mass is healthy for their corresponding height [1]. Whilst not an accurate diagnostic tool, the BMI provides a healthy body mass range for a particular height as it takes natural body shape variations into account [1]. Unfortunately, the accurate collection of ones height and mass is time-consuming for large amounts of people. Furthermore, the distance between rural areas and clinics also prevent inhabitants of that area from regular medical checkups, where a medical practitioner might require to measure and track ones height and mass. A solution to both the above mentioned problems would be the creation of a tool that would provide a medical practitioner with this information without the need for the patient to be physically present. The following report presents a study of a technique to estimate BMI from a photograph. It also reports on the design and implementation of program that estimates BMI from a photograph using this techniques, which include a mixture of Computer Vision and Machine Learning. The School of Electrical and Information Engineering (EIE) have collaborated with the Peri-Natal HIV Research Unit (PHRU) for this project.

II. BACKGROUND

A. Project Concept

The aim of this project is to estimate the BMI of a person by creating a program which accepts a front and side photograph of that person. The program will extract body shape data from the photographs with the use of Computer Vision, and estimate their BMI based on their body shape data using Machine Learning.

The purpose of such a tool has vast applications in the medical field, whether it be used as a time-efficient means of tracking the progress of a disease/disorder that affects ones mass, or for a simple indication of ones own health. Furthermore, a possible extension of this tool would be using it to classify which category of BMI one falls into: Underweight (< 18.5), Healthy (18.5 – 25), Overweight (25 – 30), or Obese (> 30).

The success of this project would result in the production of a BMI estimator which precludes the need for a subject to remove their clothes, as well as travel to a clinic for BMI measurement. Furthermore, this solution would not be affected by measurement errors resulting from incorrect calibration of machinery.

B. Assumptions & Constraints

- As to provide the least chance of a false human body outline being detected, we assume that any person being photographed is staged in front of a uniform background, such as a wall whose colour significantly contrasts with that of the patient and their clothes. All photographs used to train the machine learning algorithm should abide by the same assumption.
- In order to provide a measurement reference, the person in each photograph should be standing next to a reference object (RO) of known dimensions, such as a piece of standard A4 paper. The colour and dimensions of the piece of paper should be the same throughout the training dataset.
- This RO is present in the photographs during testing or usage of the application.
- All subjects in the photographs stand alongside and in line with the RO, i.e. if the RO is stuck to a wall, the subject is standing against the wall.
- The distance between the camera and the subjects being photographed is constant and far enough away to capture the subject's full body photograph.
- Subjects have a maximum height of two meters.
- The time frame for this project is six weeks, with a total budget of R1200.

C. Success Criteria

- The program should provide an average accuracy of 75% for it to be classified as a potential viable solution.
- To be considered a working prototype, it should provide an average accuracy of 85%.
- To be considered a completed product, it should provide a minimum average accuracy of 95% across all age, gender, and mass groups.
- BMI estimation can occur with photographs of clothed people.
- The final solution is a robust desktop application that can be easily used by a medical practitioner.
- The solution should be camera resolution-independent.

D. Optional Criteria

- The software and it's trained neural network models are exported to either a mobile application or desktop executable. In this way, one could select a photograph from their cellphone/desktop gallery and the program could perform estimation on the chosen picture.

- The desktop executable can receive pictures from a connected camera.
- A mobile application would access the cellphone's camera and perform a BMI estimation from a photograph taken in-app. The estimation could either be shown to the user or sent to the medical practitioner directly.
- Provided that there is sufficient training data, training two separate models for male and female data points would decrease sex bias in the models. This would be to further improve accuracy by accounting for natural variations of body shape between men and women, with the hope of seeing similar statistical variations in BMI across age and sex as seen in Ref. [2].

III. Literature Review

Before designing the final solution, various literature was consulted. The following literature were those most relevant to the estimation of BMI or relevant to the design choices made, as described in the succeeding sections of this report. Further literature was consulted, the findings of which can be seen in the "Literature Review" section of Appendix 3.

A. Weight and Diameter Estimation of Apples

Comert et al. aims to estimate the mass of an apple through image processing and machine learning techniques [3]. The process involves performing edge detection on the original image using the Prewitt edge extraction algorithm, which creates a binary image containing the edges in the original image. The binary image then undergoes Closing, filtering and Opening.

Closing is the process of performing Dilation followed by Erosion on an image, whilst Opening is the process of performing Erosion followed by Dilation on an image [4], [5]. Dilation is a morphological operation whereby the value (intensity) of a pixel is increased to that of the pixel in its neighbourhood with the highest intensity [6]. Erosion is a morphological operation whereby the value of a pixel is decreased to that of the pixel in its neighbourhood with the lowest intensity [6].

This is done in order to improve the image quality for feature extraction, since Closing an image improves the prevalence of contours, whilst Opening an image improves the uniformity of the spaces between contours. The above operations successfully segmented the image such that the pixels of the apple is white whilst the background is black.

Feature extraction is carried out such that the area and diameter of the apple are extracted. This feature extraction would then be performed on a set of apple photographs whose masses were known. Thereafter, the diameter and area of the apples were related to their recorded masses using a linear regression model. This method reported a 96.5% accuracy of estimation [3].

B. Types of Neural Networks

Mehta discusses the various types of neural networks and what applications in which they prove to be useful [7]. The summary is as follows: Feed-Forward neural networks were useful in situations with minimal data-sets, as well as data-sets that have a substantial amount of noise. The multi-layer perceptron is used when the training data-set is not linearly separable. Networks that make use of back propagation need large data-sets and is generally used in when the network needs to recall or retain information. Convolutional neural networks are used in applications such as object detection and image classification.

C. Estimation of BMI from Photographs Using Deep Convolutional Neural Networks

Pantanowitz et al. have conducted a similar undertaking to that described in this report. The paper discusses the estimation of BMI from a photograph using a Deep Convolutional Neural Network [8]. The analysis was performed on a dataset of 161 photographs that were used to create a set of silhouette images via manual marking as well as automated image processing techniques. The silhouette images were standardised to a particular pixel width without changing aspect ratio, and were thereafter padded appropriately and cropped to a square bounding box. The silhouette images were augmented via minimal rotations and width shifts, with the author stating that the model could only train sufficiently with augmentation. A rather complicated model consisting of 17 layers was used, details of which can be found in Ref. [8]. The results showed the BMI estimations having a correlation of 0.96 with the actual BMI values.

IV. System Design

The proposed solution comprises of two subsystems, namely the Extraction layer and the Machine Learning layer. These subsystems have no dependency on each other and are to be built independently of one another. Figure 1 in Appendix 5 provides a system diagram showing the inputs and outputs of each system, as well as the components of each system. The Extraction layer comprises of the Object Detection, Image Segmentation and Input Extraction components, whilst the Machine Learning layer comprises of the Front View, Side View and Weighted Averaging components.

A. Extraction Layer

A front and side view photograph of a person with the aforementioned RO in the top left corner is the input into the system. For the sake of clarity and simplicity, the RO chosen is a standard A3 cardboard, and its chosen colour is black for the sake of high contrast with the assumed light coloured wall in the background.

The Object Detection component will isolate the two objects in the photograph, namely the person and the RO, and return the bounding box for each. The width of the bounding box surrounding the RO is measured in pixels. Knowing its physical width allows us to calculate the pixels per meter ratio for that photograph. In other words, all objects that exist in line with the RO will be represented by approximately the same number of pixels per meter as that object. This technique was inspired by Ref. [9].

For both the front and side view photographs the Image Segmentation component will receive the locations of the RO and the person in the form of their bounding boxes. Using the bounding boxes, both the RO's and person's pixel mask (all pixels belonging to an object) will be extracted for both the front and side photographs.

Using the pixels per meter found earlier the Input Extraction component performs feature extraction on the person's front and side pixel masks. Six features are extracted from each pixel mask:

- Their approximate height in meters.

- The approximate two dimensional area contained within their shape (silhouette) in squared meters.
- The maximum thickness of their thoracic region in meters.
- The maximum thickness of their abdominal region in meters.
- The maximum thickness of their pelvic region in meters.
- The maximum thickness of their upper leg region (thigh) in meters.

Note that from front images, the 'thickness' is hereafter referred to as 'width' whilst from the side images, 'thickness' is hereafter referred to as 'depth'. The four body regions mentioned above have been chosen strategically as to represent common areas of body fat deposit. In other words, this study explores the use of Computer Vision and Machine Learning to relate the common regions of body fat deposit to BMI by extracting information about these regions from photographs. The output of the Extraction layer is two vectors, namely the front view vector (FVV) and side view vector (SVV). Both FVV and SVV will contain the above mentioned features extracted from their respective photograph.

B. Machine Learning Layer

All the outputs of the Extraction layer mentioned above are provided as inputs into the Machine Learning layer. The FVV and SVV are passed into the Front View and Side View components respectively, both of which are neural network models trained on a dataset of FVV's and SVV's, respectively. Both networks will have been trained to estimate BMI independently until an acceptable accuracy is achieved, or falling short of that, the best possible accuracy is achieved.

The estimated BMI from the Front View and Side View component are then used as inputs into a final neural network, the Error Correction / Compensation model, which performs a weighted averaging between the Front and Side BMI estimations. This network will have been trained to estimate BMI by deciding the contribution that the Front and Side view estimates should make towards a final BMI estimation with improved accuracy.

V. Training Data Collection

As mentioned above, the layer of the system that performs the BMI estimation makes extensive use of neural network models. Thus in order to train these models, a dataset containing photographs of people with their corresponding BMI was required.

A. Ethics Approval

This study was approved by The University of the Witwatersrand Human Research Ethics Committee (Medical), Certificate Number M190539. The committee approved the anonymised collection of the height, mass, age, sex, and a single front and side-facing full-body photograph of 500 fully clothed participants.

B. Collaboration with the PHRU

The PHRU provided clinical research staff who aided in the collection of the above data. Each person was issued an arbitrary subject number that was attached to each anonymised photograph. The dataset was collected at three sites:

- PHRU MDR Tuberculosis (TB) Research Clinic in Klerksdorp.
- PHRU TB Clinic in Chris Hani Baragwanath Academic Hospital in Soweto.
- The Chamber of Mines building at the University of the Witwatersrand.

Data was collected at the last site by the two principal investigators of the study.

C. Data Point Composition

As mentioned above, the data collected was as follows:

- A front and side photograph of the subject. Both photographs had the RO in the top left corner of the image.
- The height and mass of each person. These were captured simultaneously and linked to the persons subject number using the *REDCap* application.
- (Optional) Sex of person in each photograph. Whilst this characteristic is not an input into the system, it was recorded for statistical purposes. Furthermore, to achieve the optional criteria discussed in preceding sections, data points must be labelled as male/female to train separate models.
- (Optional) Age of person in each photograph. This characteristic, despite only being recorded for statistical purposes, is by far the largest contributor to the distribution of BMI, as seen in Ref. [2]. Knowing the distribution of the age of the subjects may provide a range of ages where the program is likely to provide the most accurate estimation.

With the ultimate aim in mind for this program to be a universal BMI estimator, it was encouraged that the data-set of subjects who volunteer to be photographed be spread in terms of sex, ethnicity and age. The final dataset that was used contained 465 photographs.

VI. Methodology

Drawing from the previously discussed literature, the following sections describe methodology employed to achieve each component and subsystem implementation described above. The chosen technologies used for this system is Python, due to its open-source nature, large online community support, and its growing popularity in the Machine Learning field.

A. Extraction Layer

1) Reference Object Detection and Segmentation: The first task for these components is to extract the pixels per meter metric from the photograph, since this metric is used for all subsequent data extraction in the succeeding components such as the person's height, silhouette area, widths and depths. Two methods have been used to extract the RO from the photograph, both of which make extensive use of the OpenCV-Python library. Method One is the primary operation, while Method Two is activated if Method One fails during runtime. Method One is as follows:

- Knowing that the RO is black, the image was filtered of all pixels whose 8-bit colour fell outside of the gray range of 0 – 150, where 0 is pure black and 255 is white, such that the remaining pixels were considered as the 'black' pixels of the image. The value of 150 was chosen experimentally, as to account for lighting in the photographs causing the RO to appear slightly lighter than it should.

- The filtering produced a binary image which showed the positions of black pixels in the image. Edge detection (using the OpenCV implementation of the Canny algorithm) was performed on the filtered image, returning all contours in the image.
- OpenCV's Polygon approximation was then performed on the contours that were returned. This returned all contours in the photograph that have four vertices (rectangular). The contours were ordered from left to right, and the left most contour was assumed to belong to the RO, and returned it.

This method is precise since it searches for a specific type of contour from a specific subset of the pixels in the image, and was partially inspired by Ref. [10]. Unfortunately, whilst this method works for most images there are instances where the lighting in the room caused the pixels which comprise the RO to be so brightened that it fell outside the range of gray specified above. In this case, no rectangular contours would be returned, triggering Method Two as a fail-safe. Method Two is as follows:

- The image is first converted to grayscale and slightly blurred to reduce noise.
- Morphological closing is performed on the image to fill in gaps that may prevent contour detection.
- Edge detection was performed, and returned all contours in the image, ordered from left to right.
- Smaller contours arising from noise were excluded on the basis of the pixel area contained within.

This method (inspired by Ref. [9]), whilst more reliable in finding the contour of the RO, is far more susceptible to noise in the image. As a result, it is susceptible to returning a distorted RO contour. Furthermore, it is also susceptible to non-rectangular objects appearing to the left of the RO being falsely returned as the RO (due to the left most object being assumed as the RO). Using the known dimensions of the RO, the pixels per meter metric is found and returned.

2) Person Detection and Segmentation: The final task for these components is to return the pixel mask of the person in the image and extract that the aforementioned FVV and SVV. Using the key assumption stated earlier that the person in the photograph is standing in front of a uniform background, the following method was developed, partially inspired by Ref. [11]:

- The Mask-RCNN model developed in Ref [12] was used to detect the Region of Interest (ROI) that contained the person – their bounding box. This pre-trained model has been trained on the MS COCO dataset to detect people, among other common objects, and return an approximate (rough) pixel mask of the person [13].
- The bounding box returned from the Mask-RCNN was then used to perform foreground extraction on the corresponding ROI in the original image (using the OpenCV implementation of the GrabCut algorithm), returning the pixel mask of the extracted foreground.
- The foreground-produced mask more accurately defines the silhouette but is susceptible to noise due to nearby objects, whilst the CNN-produced mask provides an approximate shape of the person but is almost unaffected by nearby objects.
- These two pixel masks are combined using the bit-wise AND operation, such that the final pixel mask only contains pixels found in both the foreground-produced and CNN-produced masks.

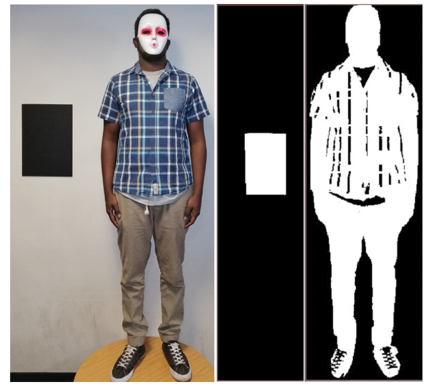

Fig. 1. Example photograph (left), reference object segmentation (middle), person segmentation (right).

The returned pixel mask was rid of zero rows as to crop the top and bottom of the mask. The number of rows in the remaining image was taken as the pixel height of the person, and converted using the pixels per meter metric to a physical height. The pixels per meter metric was used to calculate physical area per pixel. The pixels in the mask were summed and then converted to physical area using this metric.

The image was then approximated into eight sections, to represent the eight distinct regions of the body: Head, Neck, Thorax, Abdomen, Pelvis, Upper Leg, Lower Leg, Foot [14]. The row with the largest number of pixels (maximum thickness) of the Thorax, Abdomen, Pelvis and Upper Leg regions were then extracted and converted to physical dimensions. Altogether, these features made the FVV and SVV. This method of data extraction and composition of FVV and SVV were inspired by Ref. [3]. A visual representation of the entire Extraction Layer can be seen in Figure 1.

B. Machine Learning Layer

1) Model Choice and Evaluation: As per consulting the aforementioned Literature Review, it was found that the simplest type of Neural Network that may account for a noisy, small, and potentially linearly inseparable data-set is the Multilayered Perceptron (MLP). The Front, Side and Compensation models were all MLP's. The models were evaluated by a weighted scoring system that took the maximum deviation from the mean as well as the Mean Absolute Error (MAE) into account. MAE (measured in BMI) was considered as the main metric of model loss, since in this particular application, the accuracy of individual results was more important than evaluating the spread of BMI estimations [15].

2) Training and Final Performance: The program filtered the dataset of features of points that contained errors from the Extraction Layer, leaving 427 points in the feature dataset. Before training, 20% of the feature dataset was removed to be used later as unseen data during model evaluation. The remaining data was used with both the Classical Test-Train split and (Stratified) k-Fold cross validation methods to train the models. The Classical method used a 70/30 train/test split for training, whilst the Cross Validation methods used 10 folds. Different models were trained using different shuffles of data as well as different training parameters such as learning rate, batch size and number of epochs. At the end of each complete training cycle, the models were evaluated by exposing them to the unseen data. The model which had the lowest combination of MAE and maximum deviation when exposed to completely unseen data was saved as having the best performance. The final Model Loss and Training Method for the best performing models were as follows:

- Front: Trained using Stratified Cross Validation using 10 folds, Mean Absolute Error (MAE) — 3.32
- Side: Trained using Cross Validation using 10 folds, MAE — 2.77
- Compensation: Trained using the Classical Test/Train Split, MAE — 2.58

Note that the Front and Side View models were first trained, and the best performing models were used to generate data to train the Compensation model.

The final architecture of the Front model was ******, i.e. four neurons in the input layer, three in the first hidden layer, two in the second hidden layer, and one in the output layer. The final architecture of the Side and Compensation models were ****** and ******* respectively. The full implementation of the system can be found at https://github.com/djsing/ELEN4002LabProject.

VII. Sources of Error

As expected when performing feature extraction on a photographic dataset, the system design either had to account for or circumvent various errors in the dataset. The following errors were circumvented by creating a Method One as described in Section VI-A1:

- Poorly lit photographs, or non-uniform low-contrasting backgrounds that cause the RO to be confused with the background.
- The person being too close to or making contact with the RO.
- Another object being more to the left of the photograph than the RO.

The following errors were circumvented by combining the masks produced by foreground extraction and the Mask-RCNN:

- The person being in contact with another object that would affect normal edge detection.
- The photograph containing other large objects that were in the foreground with the person.

Unfortunately there are some errors that could not be accounted for or circumvented, and must be prevented in any future data collection:

- Taking front and side photographs from an angle such that the RO is seen as behind the person instead of alongside them, creating a distance perspective error.
- A more pronounced distance perspective error in all side photographs due to the RO being a person-width behind the side surface of the person being photographed. This results in the RO dimensions being slightly under-estimated, and the person's side dimensions being slightly exaggerated.

VIII. Findings & Analysis

The models that were described has having the best performance to unseen data in Section VI-B2 was used to make BMI estimations for the entire dataset of 427 people. Figures 2 and 3

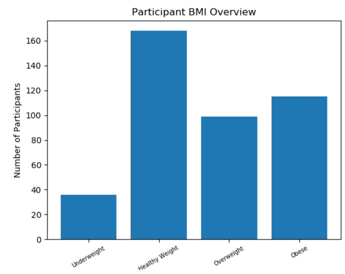

Fig. 2. Number of people in dataset according to BMI categories described in Section II-A).

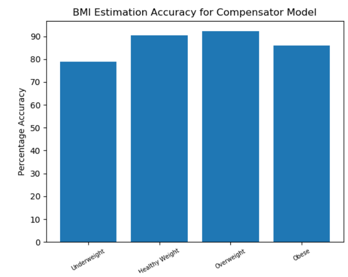

Fig. 3. Average accuracy of Final BMI Model estimations for total dataset according to BMI categories described in Section II-A).

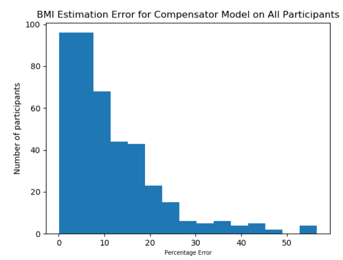

Fig. 4. Frequency of Percentage Error for Final Model BMI Estimations performed on the Full Dataset.

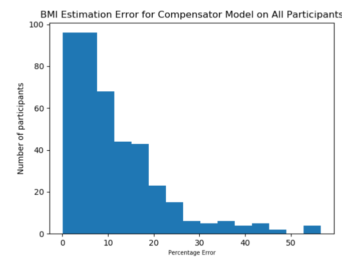

Fig. 5. Distribution of Percentage Error for Final Model BMI Estimations performed on the Full Dataset.

show the BMI category composition of the dataset and average accuracy of the final BMI estimation (i.e. from the Compensator model), respectively, for each category.

The first thing to be noted from Figure 2 is the sheer under-representation of the Underweight population in the dataset. It thus comes as no surprise that the lowest BMI estimation accuracy comes from this group. This is also seen in the Front and Side model accuracies, as seen in Appendix 5.

Despite this, it appears that the results show there is a definitive relationship between the maximum width and depth of the common areas body fat deposit and BMI. A more comprehensive view of the performance of the final model can be seen in the tally of percentage error in Figure 4, and the distribution of error in Figure **??**. This distribution can also be seen for the Front and Side model – Figures 2-5 in Appendix 5. One shortcoming of using these accuracies however, is the fact that the Obese category contains data points with a BMI between 30 and appropriately 55. The reason for this is the drastically diminishing number of data points in existence with a BMI over 40. As expected, these outliers performed poorly in comparison to the categories where the majority of data was located. The performance of these outliers can be easily visualised by the data scatters found in Appendix 5.

Figure 19 in Appendix 5 shows a slight flattening of the Front model estimation scatter, i.e. a constant prediction across a range of inputs. This is most likely due to a phenomenon known as ReLu Neuron Death, which occurs when a ReLu-activated neuron receives an overall negative input, causing it to produce zero [16]. For a constant output to occur, it then follows that multiple neurons from the same layer have died, leaving the constant bias term to make the majority (if not all) contribution to the output. It is worth noting that models trained on a dataset of this small size are particularly susceptible to ReLu death, and must be avoided by increasing dataset sizes in future.

IX. Additional Features

The program allows for height and mass models to be trained, as to provide another means of BMI estimation. The training and evaluation is the same as described above, producing the highest performing model.

All models are automatically indexed by date, time, model type (BMI/Height/Mass), Network architecture. The training also produces an output file listing the parameters used for the cycle, as well as model performance per training type (Classical/Cross Validation).

Finally, the program also provides the user with the option to visualise the data being extracted from the image for the purpose of error detection. This can be extended in future versions of the program to interactively choose a reference object in the photograph.

X. Recommendations

In accordance with Section VII, as to improve the quality of the training data as well as model performance, the following minor changes must be implemented when creating a dataset for this system in future:

- There must be at least half a meter of space to the left of the RO in the photograph, as well as to the right of the person in the photograph. Having other objects vertically below the RO (especially of the same colour) or between the persons head and shoulder in the picture may interfere with detection.
- Consider making the RO a bright, distinctly coloured object worn on the person. This would improve the system's ability to locate the RO by first detecting the person, and using the same ROI to find the RO. This would also significantly reduce the distance perspective error discussed in Section VII.

With regards to the speed of execution and resource usage of the program, the following changes should also be implemented in future:

- The program should be parallelised as far as possible. Processes such as RO detection and person detection are not required to be sequential.
- Instead of using the pre-trained Mask-RCNN model that is trained on 81 classes, train a new model on the same MS COCO dataset but only on two classes: person and background. This would decrease the size of the model, as well increase the speed of person detection.

XI. Conclusion

The results of a study exploring a new technique to estimate BMI from a photograph has been presented, along with the design and implementation of a software system that used this technique for BMI estimation. The technique involved relating BMI to features such as the estimated dimensions of common areas of body fat, silhouette area, and height. The above features were extracted using Computer Vision to perform Object Detection, Image Segmentation, Foreground Extraction and Size Estimation on front and side photographs of people. The features were then used to train three MLP models to perform BMI estimation. The BMI estimation results of the final model showed a mean estimation accuracy ranging between 79% and 92% for the BMI categories. Furthermore, the spread of error showed that 60.6% of participants had an error less than 10%, and 89% of participants had an error less than 25%, with a mean absolute error of ***, and standard deviation of ***.

XII. Acknowledgements

The author would like to extend his gratitude to the following people:

- Sachin Govender (1036148) for his contribution, particularly to the Machine Learning layer of the system. His findings can be seen in Ref [17].
- Prof. Scott Hazelhurst from the University of the Witwatersrand, who supervised this study.
- Dr. Neil Martinson from the Peri-Natal HIV Research Unit. The PHRU have been invaluable to this study due to their contribution to data collection.

References

[1] N. H. S. UK, "What is the body mass index (bmi)?" Jul 2016. [Online]. Available: https://www.nhs.uk/common-health-questions/lifestyle/what-is-the-body-mass-index-bmi/

[2] A. P. De Silva, S. H. P. De Silva, R. Haniffa, I. K. Liyanage, K. S. A. Jayasinghe, P. Katulanda, C. N. Wijeratne, S. Wijeratne, and L. C. Rajapakse, "A cross sectional survey on social, cultural and economic determinants of obesity in a low middle income setting," *International journal for equity in health*, vol. 14, no. 1, p. 6, 2015.

[3] O. CÖMERT, M. HEKİM, and K. ADEM, "Weight and diameter estimation using image processing and machine learning techniques on apple images," *International Journal of Engineering Research and Development*, vol. 9, p. 3, 2017.

[4] R. Fisher, S. Perkins, A. Walker, and E. Wolfart, "Morphology - opening," 2003. [Online]. Available: https://homepages.inf.ed.ac.uk/rbf/HIPR2/open.htm

[5] R. Fisher, S. Perkins, A. Walker, and E. Wolfart, "Morphology - closing," 2003. [Online]. Available: https://homepages.inf.ed.ac.uk/rbf/HIPR2/close.htm

[6] I. The MathWorks, "Understanding dilation and erosion :: Morphological operations (image processing toolbox user's guide)," 2005. [Online]. Available: http://matlab.izmiran.ru/help/toolbox/images/morph3.html

[7] A. Mehta, "A complete guide to types of neural networks," Jan 2019. [Online]. Available: https://www.digitalvidya.com/blog/types-of-neural-networks/

[8] A. Pantanowitz, E. Cohen, P. Gradidge, N. Crowther, V. Aharonson, B. Rosman, and D. M. Rubin, "Estimation of body mass index from photographs using deep convolutional neural networks," *arXiv.org*, Aug 2019. [Online]. Available: http://arxiv.org/abs/1908.11694

[9] A. Rosebrock, "Measuring size of objects in an image with opencv," Mar 2016. [Online]. Available: https://www.pyimagesearch.com/2016/03/28/measuring-size-of-objects-in-an-image-with-opencv/

[10] A. Rosebrock, "Opencv shape detection," Feb 2016. [Online]. Available: https://www.pyimagesearch.com/2016/02/08/opencv-shape-detection/

[11] A. Rosebrock, "Keras mask r-cnn," Jun 2019. [Online]. Available: https://www.pyimagesearch.com/2019/06/10/keras-mask-r-cnn/

[12] K. He, G. Gkioxari, P. Dollár, and R. Girshick, "Mask r-cnn," in *Proceedings of the IEEE international conference on computer vision*, 2017, pp. 2961–2969.

[13] T.-Y. Lin, M. Maire, S. Belongie, J. Hays, P. Perona, D. Ramanan, P. Dollár, and C. L. Zitnick, "Microsoft coco: Common objects in context," in *European conference on computer vision*. Springer, 2014, pp. 740–755.

[14] "Human figure proportions." [Online]. Available: https://anatomymasterclass.com/anatomy-video-lessons/human-figure-proportions

[15] L. Simmons, "Estimating body mass index from face images using keras and transfer learning," Feb 2019. [Online]. Available: https://medium.com/@leosimmons/estimating-body-mass-index-from-face-images-using-keras-and-transfer-learning-de25e1bc0212

[16] D. Liu, "A practical guide to relu," Nov 2017. [Online]. Available: https://medium.com/@danqing/a-practical-guide-to-relu-b83ca804f1f7

[17] S. Govender, "Digital estimation of body mass index," School of Electrical and Information Engineering, University of the Witwatersrand, Johannesburg, 4th Year Project Report 19P44, Sep 2019.

Appendix 1: ELEN4002 Group Reflection

Appendix 2

School of Electrical and Information Engineering
University of the Witwatersrand, Johannesburg
ELEN4002/4012: Project Specification Outline

To be completed by the supervisor	
Assessment:	
☐ Deficient	☐ Acceptable
☐ Good	☐ Excellent

Project Title: Digital Estimation of Body Mass Index

Group Number:	35	Supervisor Name:	Prof Scott Hazelhurst
Student Name A:	Sachin Govender	Student Name B:	Darrion Singh
Student Number A:	1036148	Student number B:	1056673

Ethics:

☐ Request for a waiver (does not involve human participants or sensitive data)
☐ Copy of ethics application attached (Non-medical) – School Committee
✓ Copy of ethics application attached (Medical) – University Committee

____
Supervisor Signature

Project Outline: *(give a brief outline such that ethics reviewers understand what will be done, 100 words maximum)*

There is a need for an accurate way for medical practitioners to document the Body Mass Index (BMI) of a patient without having to set up measurement apparatus or asking that patient to remove their clothes. This is particularly cumbersome in a non-urban clinic environment whereby there is a large volume of people who require consultation and limited time to for practitioners to provide consultation. This project aims to provide a means of accurately estimating the BMI of a person from a photograph of them. The person would not have to remove their clothing, and there would be no need for a measurement related equipment to be set up to take the BMI measurements.

Project Specification:

Overview:

Weight measurement is an effective way of tracking progression or identifying the presence of an illness as well as tracking malnutrition. Physical equipment has their own constraints through accessibility (i.e. patient being physically unable to be assessed on equipment), measurement errors and the time required to take the patient's weight. In situations where these constraints prove to be consequential, a robust solution is required to accurately measure one's weight with a high level of certainty. In efforts to pursue this solution the School of Electrical and Information Engineering alongside the Perinatal HIV Research Unit have requested the development of software that make use of various images of a patient in order to retrieve their weight, height and subsequently their Body Mass Index (BMI) with a high level of accuracy.

Objectives:

- Obtain weight, height and BMI through images of a patient
- Accuracy of BMI must be within 2%
- Data/Image collection of the patient must be done in a minimal time to maintain robustness of solution – a good estimation would be less than 30 seconds
- Should cater to patients who are not minimally dressed or wearing tight fitting clothes
- Camera is only equipment for data collection of patients

Assumptions

- Patient can stand in necessary positions for data collection
- Software does not need to be implemented on an application
- Patient does not need to obtain BMI immediately after image capture
- Patient is willing to remove excess clothing items such as a jacket, jersey and head accessories – provided that the weather conditions permit so
- Cameras on modern smartphones have enough image quality
- The school will provide height and weight measurement equipment when required
- Training data is provided (discussed is methodology) and has a broad spectrum of subjects in terms of gender, age height and weight
- Computer used for image processing requires a suitably high-performance GPU
- Computer will not be sponsored by the university (i.e. required to find a suitable PC or use our own)
- Software used for machine learning development is open-source or has been provided the university.
- Collected data must be confidential and anonymised

Methodology:

- Prof. David Rubin of the School of Electrical and Information Engineering has graciously offered us to use his dataset. This dataset contains silhouettes of people without clothing, and their associated height and weight.
- In addition to that, pending ethics clearance, we hope to collect the photographs, height and weight of at least 40 students within the Faculty of Engineering and the Built Environment. These students will be clothed.
- Various Neural Network (NN) models will be researched and the most appropriate will be used. Now, it appears that the types that have the highest likelihood of success is the Convolutional Neural Network, Feedforward Neural Network, and the Kohonen Self Organizing Neural Network.
- The NN models will be trained with the baseline training data provided by Prof. Rubin, as they will would contain a baseline model of the human body BMI relationship without clothes.
- Once the NN's are trained, further training will be conducted with a portion of the clothed dataset. At this point, testing will be performed with the original dataset and the portion of the dataset just used to train the NN.
- This step will be repeated with new portions of the clothed dataset until the clothed dataset has been used both in training and testing.
- This methodology provides us with the most amount of testing data, which will be tracked between tests.
- In this way, we can track the performance of the various NN models, and therefore choose which model, as well as which version of the model, is the most generalized and therefore the most likely to return the most accurate BMI estimation.

Milestones:

- Ethics application approved
- Receive initial training data
- The neural network implemented and trained with training data
- Collection of data from enough subjects (Range of 40-50 is suitable)
- Initial Assessment of the neural network's performance with collected normal clothing data achieved
- The neural network is further trained with collected data
- Accuracy within 2% is achieved

Preliminary Budget & Resources:

- Two smartphones with cameras with specifications of 20-megapixels and 16-megapixels respectively.
- Two computers with enough capability of image processing:

Specification	PC 1	PC 2
Storage	128 Gb SSD + 1 Tb HDD	500 Gb SSD+ 1 Tb HDD
RAM	8 Gb	16 Gb
GPU	Intel-integrated 620	NVidia GTX-1050 TI
Processor	Intel i5 7200	Intel i7 7700k

- Open-source software (Python, TensorFlow, Keras)
- Fully licensed access to MATLAB and all relevant toolboxes
- Weight and height measurement equipment provided by the school
- Training data provided by Prof. Rubin

Risks / Mitigation:

- Ethics approval could be delayed and affect project progression – this could be mitigated through submitting ethics application as soon as possible.
- 2 % accuracy may be excessive considering the amount of data available – cannot be mitigated unless revised success criteria is considered.
- Insufficient time to complete project – mitigate through finishing data collection and software-related research before allocated 6 weeks.
- Measurement equipment provided may not record data accurately – cannot be mitigated.
- Delay in collecting enough data due to people being uncomfortable with pictures and weight recording – mitigate through explicitly telling subjects that their data will be confidential, and images will be pixelated for anonymity.

Appendix 3: ELEN4002 Project Plan

Darrion Singh (1056673) and Sachin Govender (1036148) – Group 35

***Abstract*—This report contains a project plan to implement an image-based digital Body Mass Index (BMI) estimator. The estimator is to be a robust and accurate alternative BMI estimation to the classical height and mass measurement. The major benefit of the estimator is the neccessary ability to assess large groups of people in a relatively quick manner. Literature related to image processing and neural networks were consulted to propose a system to accurately estimate BMI. The proposed solution consists of taking a front and side view image of participants and using feature extraction methods such as Object Detection and Image Segmentation to obtain various measurements across the body. Thereafter the measurements should be passed into two independent neural networks that estimate BMI based on front view data, separately from side view data. These two independent BMI estimates are passed through a neural network that provides an averaged weighting to obtain a final BMI estimate. In order to create the proposed system, a data-set of photographs, mass, height, and optionally race and sex will be obtained from research clinics in order to train and test the models in the system. Image processing and machine learning will be carried out in Python with the aid of the OpenCV, Tensorflow and Keras libraries. The project is expected to be concluded within six weeks and must aim to meet the success criteria of having a maximum of 2% error. Difficulties are expected in the form of time limitations, data-set inadequacy and potential over-fitting. These issues can be mitigated through extending the data collection time, increasing the size of the data-set and ensuring a relatively equal distribution of participants in terms of age, race and gender.**

I. Introduction

The Body Mass Index (BMI) is a metric used as an indication of whether one's body mass is healthy for their corresponding height [1]. Whilst not an accurate diagnostic tool, the BMI provides a healthy body mass range for a particular height as it takes into account natural body shape variations [1]. Unfortunately, the accurate collection of ones height and mass is time-consuming, especially when considering the time required to collect this data for large amounts of people. This project investigates the potential for a digital estimation of one's BMI from a photograph, using a mixture of Computer Vision and Machine Learning. The following report contains the proposed methods and respective schedules associated with the undertaking of this project.

II. Background

A. Project Concept

The aim of this project is to estimate, to a reasonable degree of accuracy, the BMI of a person from their photograph. This is to be achieved by the creation of a program which accepts a photograph of a person, extracts body shape data from the photograph and with the use of a machine learning algorithm, predict one's BMI based on their body shape data. The purpose of such a tool has vast applications in the medical field, whether it be used as a time-efficient means of tracking the progress of a disease/disorder that increases/decreases ones mass, or for a simple indication of ones own health.

The success of this project would result in the production of a mobile BMI estimator which precludes the need for a subject to remove their clothes, saving practitioners time for BMI measurement. In large rural clinics, where large patient volumes combined with time taken to undress and redress for accurate height and mass measurements, this digital solution may result in a significant reduction in the time taken for consultations. Furthermore, this solution would not be affected by measurement errors resulting from incorrect calibration.

B. Assumptions & Constraints

As to provide the least chance of a false human body outline being detected, we assume that any person being photographed is staged in front of a uniform background, such as a wall of distinct uniform colour. As such, all photographs used to train the machine learning algorithm should abide by the same assumption. In order to provide a distance reference, each photograph should contain a shape of known dimensions. For consistency, the shape should be the same in all pictures, as well as having the same colour and the same dimensions. This reference object must be present both during training and live testing. Furthermore, the distance between the camera and the subjects being photographed should remain static and be far enough away to capture the subject's full body photograph, where the subject is a maximum of two meters tall. The time frame for this project is seven weeks, with a total budget of R1200.

C. Success Criteria

As this solution means to be time efficient, it aims to negate the need for one to remove their clothes as is usual with regular BMI measurement. Thus, the program should be able to operate with pictures of people who are clothed. The estimation should also be possible using pictures of varying resolutions, i.e. the solution must work for multiple cameras. The specified success criteria error has been given as 2%, however the upper limit of the BMI is generally given as 40 [1]. This translates to an absolute error of 0.8 such that for an actual BMI of 20, an estimated value between 19.2 and 20.8 will be considered as accurate enough.

D. Optional Criteria

Since the purpose of digital estimation of BMI is the production of a solution that is mobile and not time-consuming, it would be ideal if the software and it's trained neural network models were exported to either a mobile application or desktop executable. In this way, one could select a photograph from their cellphone/desktop gallery and the program could perform estimation on the chosen picture.

Furthermore, to improve the robustness of the design it would also be most ideal if the application/executable could receive pictures from a connected camera. In the case of the mobile application, it would access the cellphone's camera and perform a BMI estimation from a photograph taken in-app. This would potentially expand the use of the machine learning algorithm to be used by the general consumer with minimal effort.

As seen in Ref. [2], there are significant statistical variations in the BMI of male and females across age groups. Provided that there is sufficient training data, one final manner of decreasing the effects of sex bias in the data from the photographed subjects would be to train two separate models for male and female data points. The exclusion of the male data from the female model and visa-versa will improve accuracy by modelling a specific BMI distribution for each sex, as to statistically correlate with Ref. [2].

III. Training Data Collection

A. Data Types

The data to be collected are as follows:

- A front and side photograph of the subject. Both photographs should have the reference object in the top left corner of the image. As mentioned above, the background of the image should be a uniform surface such as a wall. This is to ensure that the image pre-processing techniques described in the succeeding sections of this report can be carried out without error.
- The height and mass of each person. These must be captured simultaneously and attached to the front and side photographs. Data survey applications such as *REDCap* may be useful in achieving this.
- (Optional) Sex of person in each photograph. Whilst this characteristic is not an input into the system, it should be recorded for statistical purposes for later quantification of algorithmic bias. Furthermore, to achieve the optional criteria discussed above, data points must be labelled as male/female to train separate models.
- (Optional) Race of person in each photograph. This characteristic could also be recorded for statistical purposes. Furthermore, it opens an avenue for further research into how ethnicity can be used to narrow BMI distribution and hence provide more accurate BMI estimations.
- (Optional) Age of person in each photograph. This characteristic, despite only being recorded for statistical purposes, is by far the largest contributor to the distribution of BMI, as seen in Ref. [2]. Knowing the distribution of the age of the subjects will provide a range of ages where the program is likely to provide the most accurate estimation.

With the ultimate aim in mind for this program to be a universal BMI estimator, it is encouraged that the data-set of subjects who volunteer to be photographed be spread in terms of sex, ethnicity and age.

B. Obtaining Data

It is possible to find such images of people on the internet, however it is unlikely for these images to have the corresponding height and mass on record. The chosen method for obtaining the data is to illicit the help of a medical research group who have access to large amounts of people, along with properly calibrated height and mass measuring equipment.

It should be noted that obtaining data in this manner requires prior clearance from a recognised ethics committee, as well as the permission of those in charge of the proposed sites such as research clinics. As with most neural network related projects, it is encouraged for a large data-set to be collected, and it was decided that a reasonable minimum number of subjects would be 100.

IV. Literature Review

Before deciding on the proposed solution described in the succeeding sections, various academic literature regarding image processing and machine learning were consulted. The following provides a summary of findings that were considered when deciding upon the final solution.

A. Weight and Diameter Estimation

Comert et al. aims to estimate the mass of an apple through image processing and machine learning techniques [3]. The process involves using the Prewitt edge extraction algorithm in order to obtain a binary image. The binary image goes through Closing, filtering and Opening.

Closing is the process of performing Dilation followed by Erosion on an image, whilst Opening is the process of performing Erosion followed by Dilation on an image [4], [5]. Dilation is a morphological operation whereby the value (intensity) of a pixel is increased to that of the pixel in its neighbourhood with the highest intensity [6]. Erosion is a morphological operation whereby the value of a pixel is decreased to that of the pixel in its neighbourhood with the lowest intensity [6]. This is done in order to improve the image quality for feature extraction, since Closing an image improves the prevalence of contours, whilst Opening an image improves the uniformity of the spaces between contours. Figure 1 shows the segmentation stage of the apple image.

Feature extraction is carried out such that the area and diameter of the apple are extracted, thereafter the diameter and the recorded masses are used in a linear regression model. Figure 2 shows the diameter extraction from the images.

B. Weight Estimation Using Image Analysis and Statistical Modelling

Similar to the requirements of the BMI system, Kollis et al. discusses a robust image analysis based mass estimation of pigs in large scale livestock farming [7]. The software of choice for image processing is MATLAB in which the Image Processing Toolbox V4.1 is used. Object detection occurred

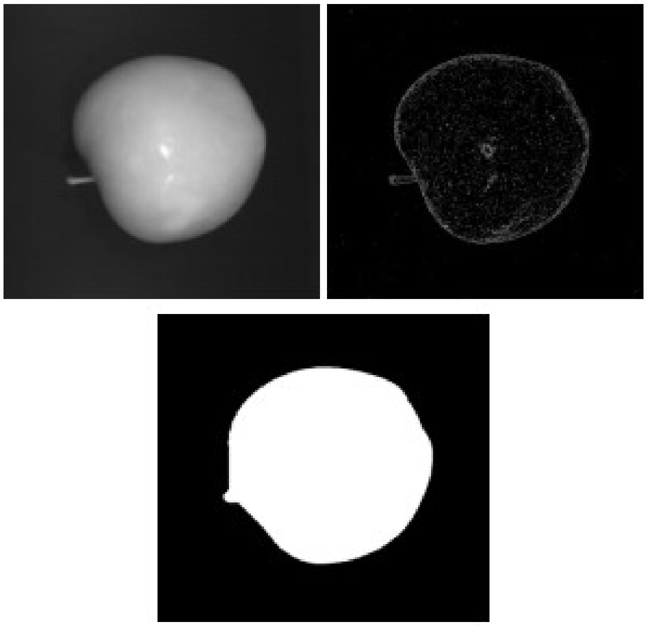

Fig. 1. Segmentation techniques to isolate the apple in the image [3].

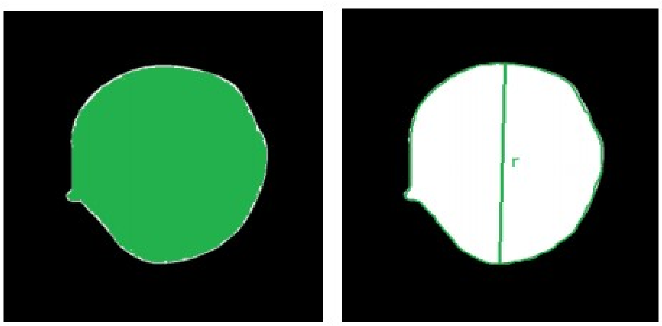

Fig. 2. Apple Area and Diameter Extraction [3].

through comparison between the camera environment without the pig and with the pig. This results in a binary image that undergoes segmentation, filtering and opening to get a clear image for feature extraction.

Feature extraction in this paper is particularly useful as the body shape of the pigs is more complex than the apple described in the previous review. The feature extraction method is comprised of the skeletonisation function in MATLAB which generates a skeletal structure of the object in question. Figure 3 shows the skeletonisation process for the image of the pig.

Using this process the spine (main branch) is used to estimate the length of the pig by counting the pixels which comprise the spine. Thereafter the estimated length and recorded masses were used in a linear regression model to estimate mass.

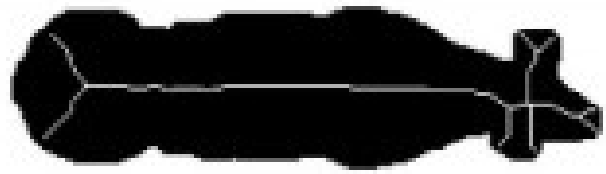

Fig. 3. Skeletonisation function implemented on pig binary image [7].

C. Human Weight from a RGB-D Image

Nguyen et al. explores human mass prediction through a single RGB-D image [8]. The camera in question uses a depth sensor to create a separate depth image. This extra image proves to be useful in object detection due to depth allowing an accurate trace of the human body shape with the aid of the 2D RGB image.

This method is more accurate than the common rectangular object identification method. This silhouette is passed through similar image processing methods used in the aforementioned papers. However, mass prediction is different in this scenario as the author used various mathematical relationships relating body width, height and area with Support Vector Regression – as opposed to machine learning – to predict the mass of the person through the extracted features.

D. Types of Neural Networks

Mehta discusses the various types of neural networks and what applications in which they prove to be useful [9]. In summary: Feed-Forward neural networks were useful in situations with minimal data-sets, as well as data-sets that have a substantial amount of noise. The multi-layer perceptron is used when the training data-set is not linearly separable. Networks that make use of back propagation need large data-sets and is generally used in when the network needs to recall or retain information. Convolutional neural networks are used in applications such as object detection and image classification.

V. System Design

The proposed solution comprises of two subsystems, namely the Extraction layer and the Machine Learning layer. These subsystems have no dependency on each other and are to be built independently of one another. Figure 4 in the Appendix provides a system diagram showing the inputs and outputs of each system, as well as the components of each system. As seen, the Extraction layer comprises of the Object Detection, Image Segmentation and Input Extraction components. The Machine Learning layer comprises of the Front View, Side View and Weighted Averaging components.

A. Extraction Layer

A front and side view photograph of a person with a reference object in the top left corner is the input into the system. Object Detection will isolate the two objects in the photograph, namely the person and the reference object [10]. The dimensions of the bounding box surrounding each object, measured in pixels, will be converted into meters using the known physical dimensions of the reference object. This is done by computing the meters per pixel metric [10]. The vertical dimension of the person's bounding box, i.e. their height, is one of the outputs of the subsystem.

For both the front and side view photographs, each original unprocessed image will undergo Image Segmentation separately. The Image Segmentation process will locate the pixel mask – all pixels belonging to an object – of the two objects in each photograph, although only the masks of the person is required [11], [12]. Using the meters per pixel found in the previous component, different horizontal slices of the person's pixel masks will be extracted as the width (for the front view photograph) and depth (for the side view photograph) of the person at their corresponding heights above the ground. The final number of width/depth measurements as well as the height at which they will be taken is unknown, and it is only through testing that the optimum number of slices/heights can be determined. Finally, using the meters per pixel metric, the pixel masks will be converted into a physical two dimensional area.

The output of the Extraction layer will two vectors, namely the front view vector (FVV) and side view vector (SVV). The FVV will comprise of the person's front view area, and the person's width at various heights. The SVV will comprise of the person's side view area, and the person's depth at various heights. It is important to note that the chosen heights for the width/depth measurement **must** be the same, such that corresponding indices of the FVV and the SVV refer to the same height above the ground.

B. Machine Learning Layer

All the outputs of the Extraction layer mentioned above are provided as inputs into the Machine Learning layer. The front view and side view outputs are passed into the Front View and Side View components independently, both of which are neural networks. Both networks will be trained to estimate BMI independently until an acceptable accuracy is achieved, or falling short of that, the best possible accuracy is achieved.

The estimated BMI from the Front View and Side View component are then used as inputs into a final neural network, the Weighted Averaging component. This network will be trained to estimate BMI by deciding the contribution that the front and side view images should make to obtain a final BMI estimation with improved accuracy.

VI. Methodology

The above processes can be done using multiple image processing libraries with languages such as Python, R, MATLAB etc. However, due to its open-source nature, large online community support, and its growing popularity in the Machine Learning field, this project will be created using Python. Currently, there is no indication that any other language or software package will be required to implement this project, however, any deviations from this statement will be minimised as far as possible and be documented. The following section breaks down the methodology employed to achieve each component and subsystem implementation, as well as the potential tools that may be used.

A. Extraction Layer

1) Object Detection: The ultimate purpose of the Object Detection component is to extract the meters per pixel metric from the photograph, since this metric is used for all subsequent data extraction in the succeeding components such as the person's height, width and depth. The manner in which this is to be performed, using the OpenCV-Python library, is as follows: The image is first converted to grayscale (commonly referred to as a black and white image), and slightly blur it [10]. The purpose of this is to decrease the effects of noise in the picture, which manifests as random sharp changes in the picture such as a scratch on a uniform surface – a distinct line that does not contribute to the contour of any object.

Closing is then performed on the image to remove any gaps in a border line that might prevent contour detection [10]. All contours in the image are then detected and ordered from left to right [10]. As mentioned in the previous sections, the reference object must be placed in the top-left corner – this choice was to facilitate contour detection, and ensuring that the reference contour is always evaluated first. We then discard false contours detected by OpenCV by setting a minimum contour length [10]. This value needs to be sufficiently small to prevent the reference object from being discarded, but large enough that minute details about the person is discarded, such as their eyes and mouth.

The reference object bounding box and person object bounding box is then extracted from the image [10]. Using the known dimensions of the reference object, extract the physical vertical dimension of the person object bounding box, and provide it as an output of the Extraction layer.

2) Image Segmentation: The ultimate purpose of the Image Segmentation component is to extract the width and depth of a person from the front and side view photographs respectively, since these two characteristics (at specified heights) are data points for the FVV and SVV respectively. The manner in which these are extracted, using the OpenCV Python library, is as follows: Using a trained masking model such as ENet or Mask R-CNN, the images will have instance or semantic segmentation performed on them [11], [12]. Due to the assumption that only one person will exist in the photograph, with the rest of the image in the background, there will only ever be one instance of a person in the image, and as such, either semantic or instance segmentation should be sufficient to extract the mask.

Once the pixel mask has been isolated, the pixels along a certain row (specified height) of the mask will be summed and converted to distance, yielding the width/depth for front/side pictures. Finally, a simple pixel to area conversion will yield the mask's corresponding physical area after summing the number of pixels in the entire mask. These widths, depths, and areas will be extracted and provided as outputs of the Extraction layer.

B. Machine Learning Layer

As per consulting the aforementioned Literature Review, it must be noted that the simplest type of Neural Networks that may account for a noisy, small, and potentially linearly inseparable data-set are the Feed Forward or Artificial Neural Network, and the Multilayered Perceptron. These network models are relatively simple to implement due to their lack of kernel requirements or back propagation, and as such are conducive to implementation via a well documented medium such as Python. In particular, libraries such as TensorFlow and its complementary libraries such as Keras are thoroughly documented in the field of Machine Learning. These libraries will be used to create and train the Front View and Side View component models, as well as the Weighted Averaging component model.

The data-set will be split into 70% used for training, and 30% for testing and validation. Training will halt via the early stopping mechanism, which stops a model from training past a certain metric – in this case being the Mean Absolute Error [13]. This early stopping will prevent over-fitting of data points and thus aims to improve generalisation of the model.

Finally, it is important to note that whilst the Front and Side View models are to be trained with the data from the Extraction layer, the Weighted Averaging layer can only be trained once the Front and Side View models are trained. This is due to the fact that the Weighted Averaging model training data-set is to be generated from the output of the Front and Side View models.

VII. Project Management

As stated above, six weeks have been given to deliver a working solution. Due to time being limited, it is imperative to optimise the productivity of the development of the project through scheduling and resource allocation.

The working hours on the project have been summarised as follows:

- The work weeks consist of the five weekdays.
- Working hours lies between 10:30 and 18:00.
- Time spent on the project during weekends and times outside of working hours are considered as overtime.

The resources available for the project are:

- The Primary Investigators.
- Supervisors.
- Clinic Staff.
- Smartphone cameras with a minimum quality of 15 megapixels.
- Three Laptop/Desktop computers with a minimum processing speed of 2.5GHz, 8GB of RAM, two of which have GTX-1050 GPUs.
- Cloud storage sufficient to store photograph backups.

Figure 5 in the Appendix provides a Gantt chart that summarises the work breakdown during the course of the project. The tasks allocated to the primary investigators are as follows: Darrion Singh is allocated the Extraction layer (Image Processing) and Sachin Govender is allocated the Machine Learning layer (Neural Network). The time allocated to the project is split as shown in Figure 5, however there is a possibility of both investigators facilitating each other on their respective tasks due to unforeseen circumstances such as task difficulty or slack in a parallel task.

VIII. Expected Challenges

An neural network based model is only as good as the size and quality of the data set used for training. Due to the limited time available for data collection, there is possibility that the data acquired may not create a model that meets the success criteria. Excluding time sensitivity, the data acquisition can be further hindered by multiple factors such as clinic facilitators/data collectors not following procedure correctly, resulting in the reduced quality and size of the data-set. Preventative measures for data collection errors cannot be mitigated as the primary investigators for this project being unable to provide thorough supervision over the data collection, as per the project schedule not allowing for it.

Another issue relevant to the neural network is over-fitting. There is a possibility of the variety of participants (according to age, gender, race and body type) being has unequally distributed across the total data-set. Thus, over-fitting could occur due to the presence of dominant groups with similar traits. Despite having specified a minimum number of participants for data-collection, there is no finite number of data points that can guarantee good results - in other words, the minimum number of participants may be insufficient for good results.

If training the neural network is to be successful with a lack of participants, a large number of width/depth extractions from the Image Segmentation component may be required. However, knowing that each measurement contains some error, it stands to reason that more width/depth extractions from the same image will result in a higher the overall measurement error.

IX. Recommendations

As stated above this project is conducted in a limited time of 6 weeks, which affects the scheduling of the project in a manner that reduces the time to obtain a sufficient data-set in terms of size and quality. It is recommended that when planning future projects of this nature, that it either be conducted in a longer time frame as to guarantee a sufficiently large and well-spread data-set, or alternatively to only commence on a project of this nature once the data-set has been collected.

Quality of the data set could be improved through ensuring that there is minimal biases in terms of composition of the total data set. Biases come in the form of minimal variations in race, sex and age. Due to the nature of the project being potentially used for clinics in rural areas, there should be a demographic study on the areas that the software will be implemented in to better select the data used for the network.

It is also worth noting that neural networks could also be used to classify participants according to their body type – an attribute that could improve the estimation of a subject's BMI due to having its own BMI distribution. If a larger data-set is available, consideration should be made to incorporate a body type classifier into the BMI estimation.

X. Conclusion

The above report has presented a plan for the implementation of a digital Body Mass Index Estimator. It is intended for this project to be implemented using Python, via the OpenCV, TensorFlow and Keras libraries, among others. The system is divided into a data extraction layer (Extraction) and a neural network layer (Machine Learning). To create data for the training of the Machine Learning layer of the system, a data-set of photographs of a subject labelled with their height, mass and optionally age, sex and race must be pre-processed. The data-set is to be collected from a research facility during the course of the project. This pre-processing can be accomplished by Object Detection and Image Segmentation as shown regarding the Extraction layer above. Training of the various components of the Machine Learning layer is to be controlled by an early stopping mechanism to prevent over-fitting. The largest challenge that may be experienced over the course of the project is the inability to collect a sufficiently good data-set in terms of image quality, volume of data (subjects), and the spread of age, sex and race among the subjects, which may create inaccuracies in the system.

References

[1] N. H. S. UK, "What is the body mass index (bmi)?" Jul 2016. [Online]. Available: https://www.nhs.uk/common-health-questions/lifestyle/what-is-the-body-mass-index-bmi/

[2] A. P. De Silva, S. H. P. De Silva, R. Haniffa, I. K. Liyanage, K. S. A. Jayasinghe, P. Katulanda, C. N. Wijeratne, S. Wijeratne, and L. C. Rajapakse, "A cross sectional survey on social, cultural and economic determinants of obesity in a low middle income setting," *International journal for equity in health*, vol. 14, no. 1, p. 6, 2015.

[3] O. CÖMERT, M. HEKİM, and K. ADEM, "Weight and diameter estimation using image processing and machine learning techniques on apple images," *International Journal of Engineering Research and Development*, vol. 9, p. 3, 2017.

[4] R. Fisher, S. Perkins, A. Walker, and E. Wolfart, "Morphology - opening," 2003. [Online]. Available: https://homepages.inf.ed.ac.uk/rbf/HIPR2/open.htm

[5] R. Fisher, S. Perkins, A. Walker, and E. Wolfart, "Morphology - closing," 2003. [Online]. Available: https://homepages.inf.ed.ac.uk/rbf/HIPR2/close.htm

[6] I. The MathWorks, "Understanding dilation and erosion :: Morphological operations (image processing toolbox user's guide)," 2005. [Online]. Available: http://matlab.izmiran.ru/help/toolbox/images/morph3.html

[7] K. Kollis, C. S. Phang, T. M. Banhazi, and S. J. Searle, "Weight estimation using image analysis and statistical modelling: A preliminary study," *Applied engineering in agriculture*, vol. 23, no. 1, pp. 91–96, 2007.

[8] T. V. Nguyen, J. Feng, and S. Yan, "Seeing human weight from a single rgb-d image," *Journal of Computer Science and Technology*, vol. 29, no. 5, pp. 777–784, 2014.

[9] A. Mehta, "A complete guide to types of neural networks," Jan 2019. [Online]. Available: https://www.digitalvidya.com/blog/types-of-neural-networks/

[10] A. Rosebrock, "Measuring size of objects in an image with opencv," Mar 2016. [Online]. Available: https://www.pyimagesearch.com/2016/03/28/measuring-size-of-objects-in-an-image-with-opencv/

[11] A. Rosebrock, "Semantic segmentation with opencv and deep learning," Sep 2018. [Online]. Available: https://www.pyimagesearch.com/2018/09/03/semantic-segmentation-with-opencv-and-deep-learning/

[12] A. Rosebrock, "Instance segmentation with opencv," Nov 2018. [Online]. Available: https://www.pyimagesearch.com/2018/11/26/instance-segmentation-with-opencv/

[13] L. Simmons, "Estimating body mass index from face images using keras and transfer learning," Feb 2019. [Online]. Available: https://medium.com/@leosimmons/estimating-body-mass-index-from-face-images-using-keras-and-transfer-learning-de25e1bc0212

Appendix

The following pages contain an appendix for the report titled "ELEN4002 Project Plan" by Darrion Singh and Sachin Govender. All images in this Appendix have been referenced in the main report. This Appendix contains the system block diagram of the BMI estimation program, as well as Gantt Chart describing the time allocation to tasks to be undertaken in the project.

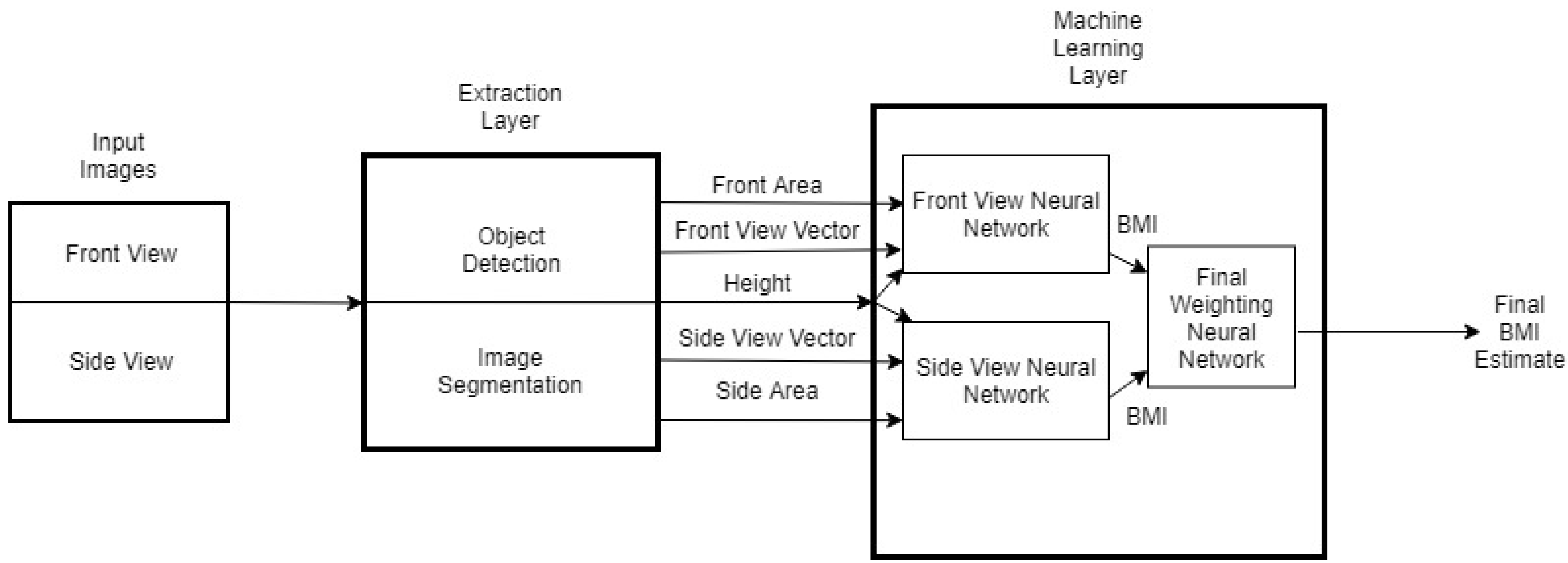

Fig. 4. System Block Diagram

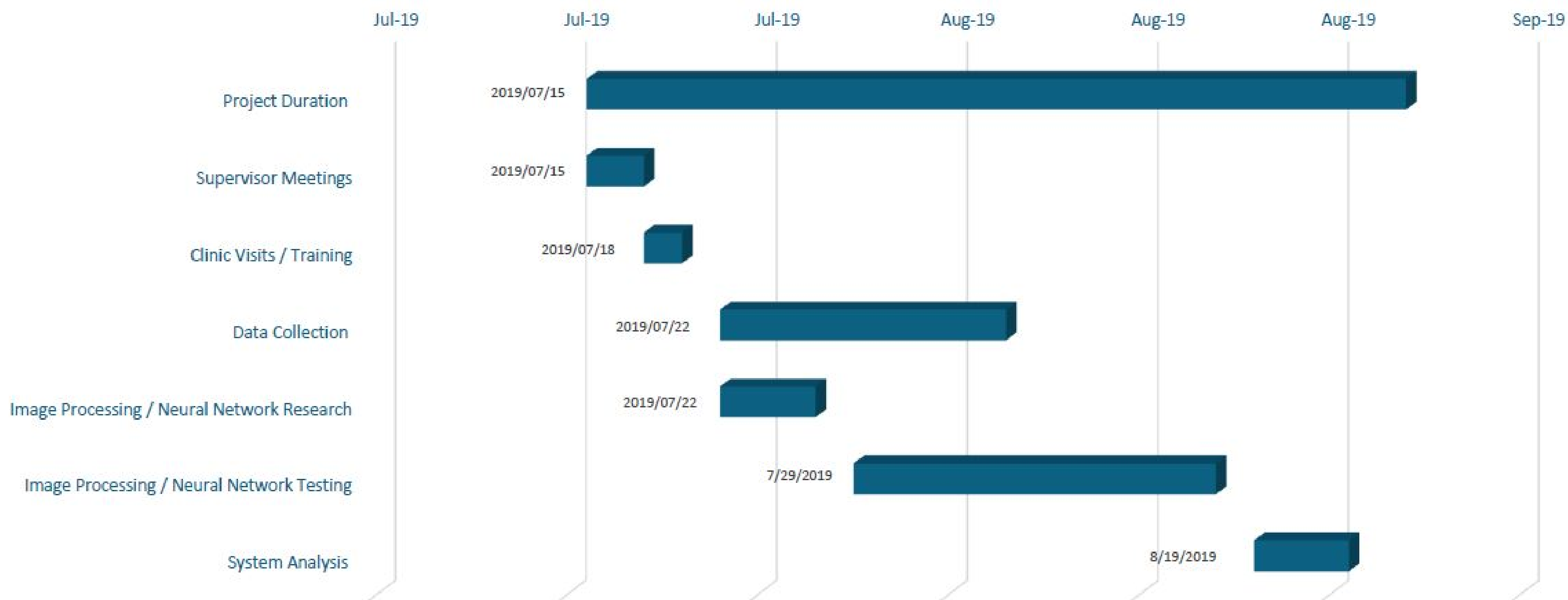

Fig. 5. Gantt Chart showing project task breakdown and time allocation.

Appendix 5: Supplementary Information to ELEN4002 Report

I. Introduction

The following pages contain an appendix for the report titled "ELEN4002: Digital Estimation of Body Mass Index" by Darrion Singh. All images in this Appendix have been referenced in the main report.

II. System Block Diagram

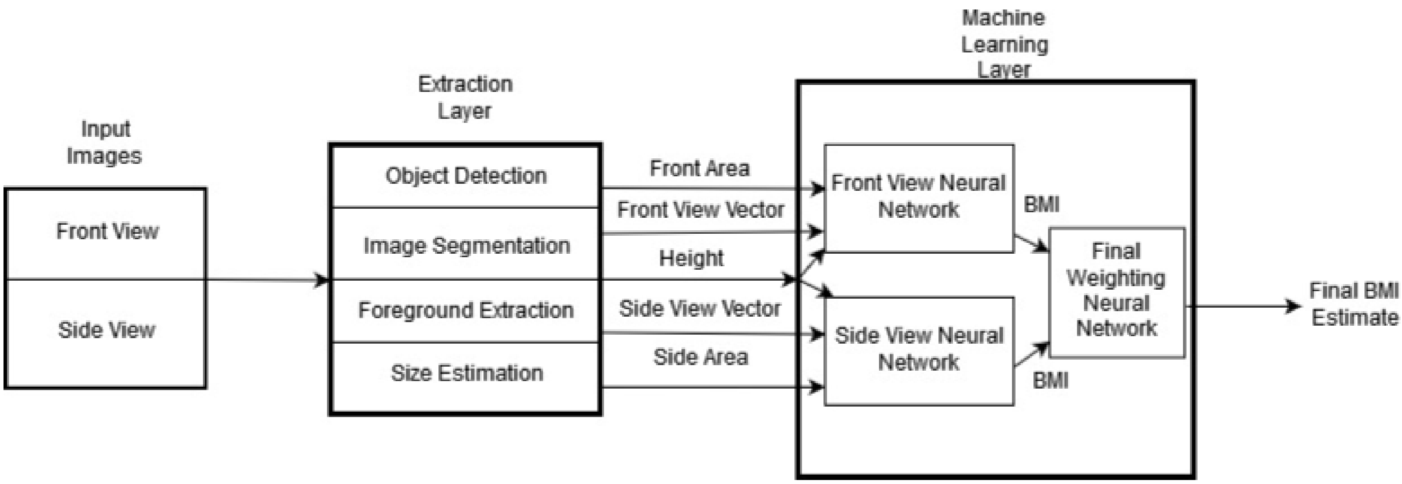

Fig. 1: System Block Diagram showing the processes input images undergo to estimate BMI.

III. Model Estimation Accuracy For Total Dataset

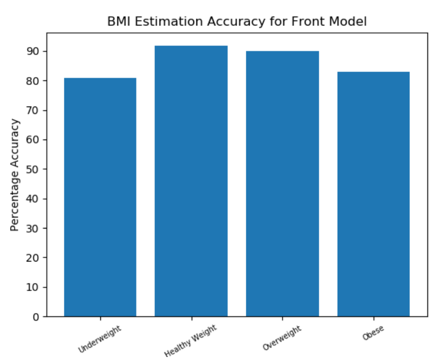

Fig. 2: Estimation Accuracy Across BMI Categories for Front Model.

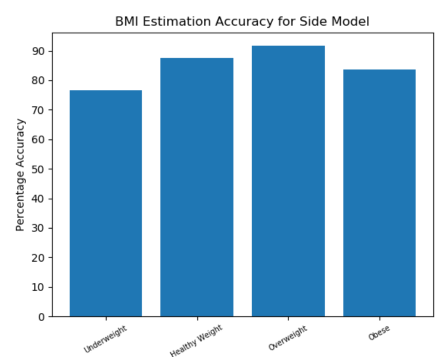

Fig. 3: Estimation Accuracy Across BMI Categories for Side Model.

IV. Model Error Distribution For Total Dataset

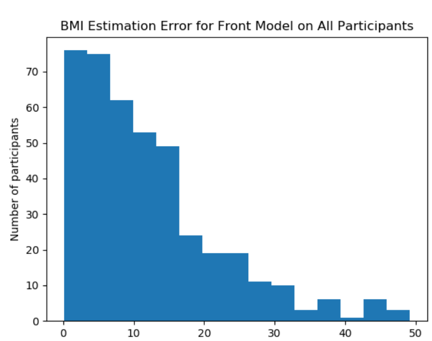

Fig. 4: Error Distribution Across BMI Categories for Front Model.

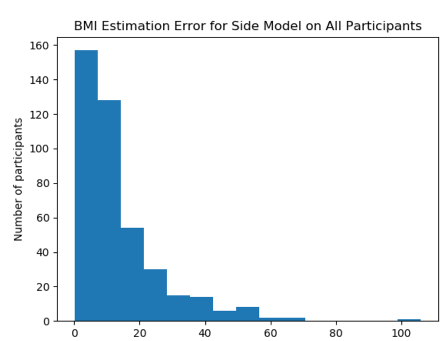

Fig. 5: Error Distribution Across BMI Categories for Side Model.

V. FRONT MODEL ERROR DISTRIBUTION

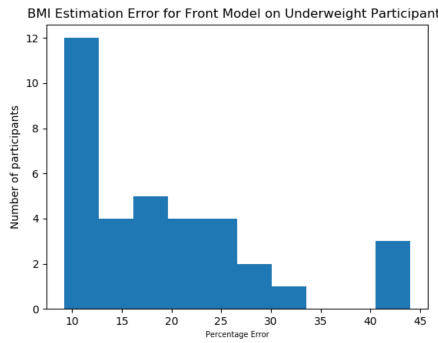

Fig. 6: Distribution of Percentage Error for Front Model BMI Estimations performed on the Underweight portion of the dataset.

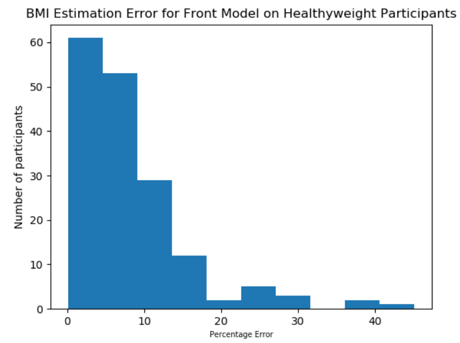

Fig. 7: Distribution of Percentage Error for Front Model BMI Estimations performed on the Healthy portion of the dataset.

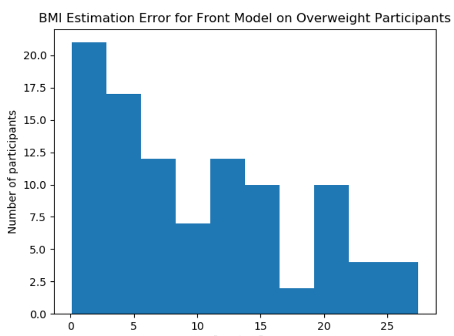

Fig. 8: Distribution of Percentage Error for Front Model BMI Estimations performed on the Overweight portion of the dataset.

Fig. 9: Distribution of Percentage Error for Front Model BMI Estimations performed on the Obese portion of the dataset.

VI. SIDE MODEL ERROR DISTRIBUTION

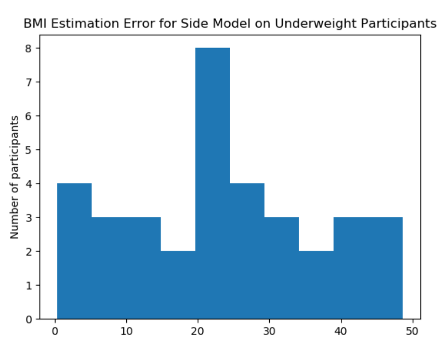

Fig. 10: Distribution of Percentage Error for Side Model BMI Estimations performed on the Underweight portion of the dataset.

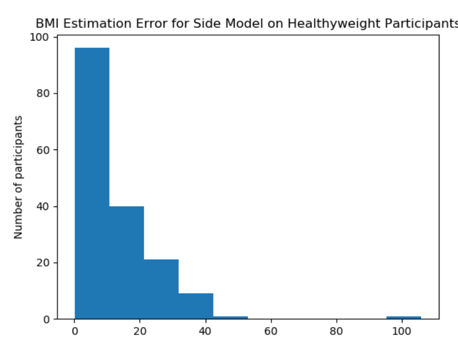

Fig. 11: Distribution of Percentage Error for Side Model BMI Estimations performed on the Healthy portion of the dataset.

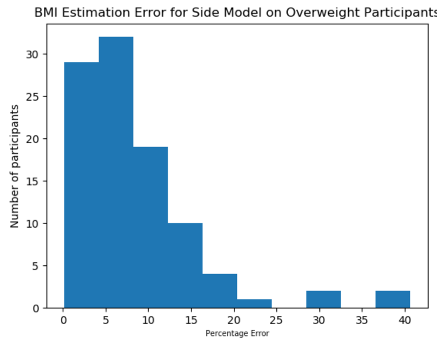

Fig. 12: Distribution of Percentage Error for Side Model BMI Estimations performed on the Overweight portion of the dataset.

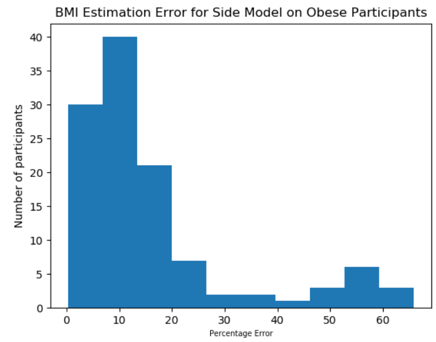

Fig. 13: Distribution of Percentage Error for Side Model BMI Estimations performed on the Obese portion of the dataset.

VII. Final/Compensator Model Error Distribution

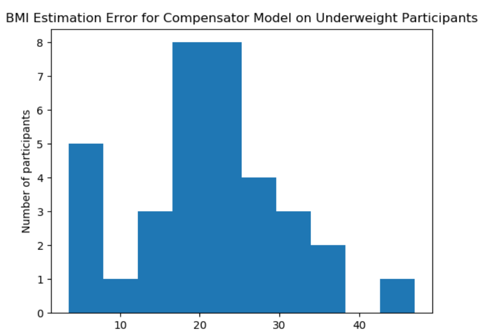

Fig. 14: Distribution of Percentage Error for Final Model BMI Estimations performed on the Underweight portion of the dataset.

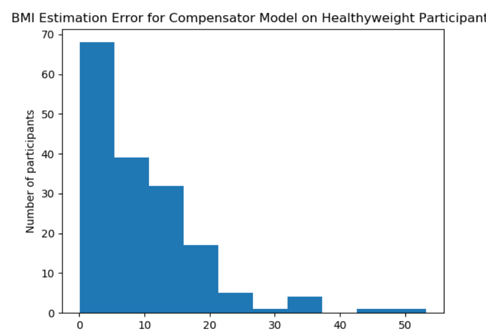

Fig. 15: Distribution of Percentage Error for Final Model BMI Estimations performed on the Healthy portion of the dataset.

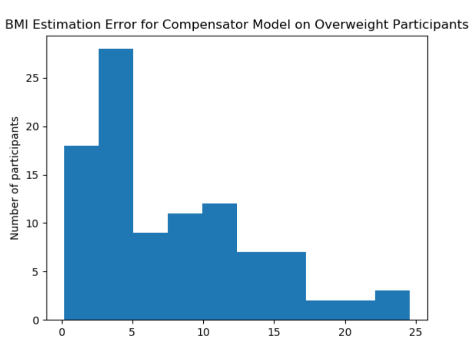

Fig. 16: Distribution of Percentage Error for Final Model BMI Estimations performed on the Overweight portion of the dataset.

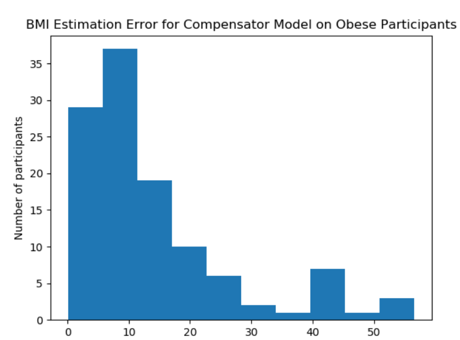

Fig. 17: Distribution of Percentage Error for Final Model BMI Estimations performed on the Obese portion of the dataset.

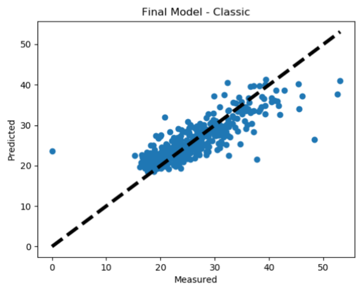

Fig. 18: Final Model Performance for Total dataset.

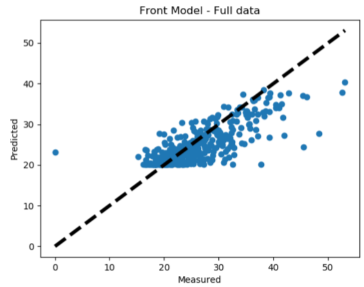

Fig. 19: Front Model Performance for Total dataset.

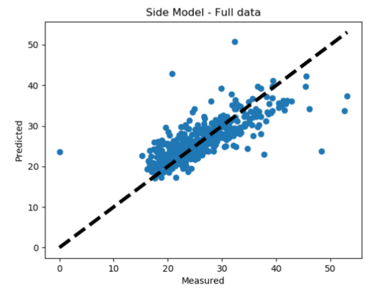

Fig. 20: Side Model Performance for Total dataset.