A Machine Learning Approach to Short-term Body Weight Prediction in A Dietary Intervention Program

Oladapo Babajide[1], Tawfik Hissam[1], Palczewska Anna[1], Gorbenko Anatoliy[1]
Arne Astrup[2], J. Alfredo Martinez[3], Jean-Michel Oppert[4] and Thorkild IA Sørensen[5]

[1]School of School of Built Environment, Engineering & Computing.
O.Babajide5449@student.leedsbeckett.ac.uk

[2]Department of Nutrition, Exercise and Sports, University of Copenhagen, Denmark.
ast@nexs.ku.dk

[3]Centre for Nutrition Research, University of Navarra, CIBERobn Obesity and
IMDEA program on Precision Nutrition, Spain.
jalfmtz@unav.es

[4]Sorbonne University, Department of Nutrition Pitie-Salpetriere Hospital,
Institute of Cardiometabolism and Nutrition (ICAN), Paris, France.
jean-michel.oppert@aphp.fr

[5]Novo Nordisk Foundation Center for Basic Metabolic Research and Department of Public
Health, Faculty of Medical and Health Sciences, University of Copenhagen, Denmark.
tias@sund.ku.dk

Abstract. Weight and obesity management is one of the emerging challenges in current health management. Nutrient-gene interactions in human obesity (NUGENOB) seek to find various solutions to challenges posed by obesity and over-weight. This research was based on utilising a dietary intervention method as a means of addressing the problem of managing obesity and overweight. The dietary intervention program was done for a period of ten weeks. Traditional statistical techniques have been utilised in analyzing the potential gains in weight and diet intervention programs. This work investigates the applicability of machine learning to improve on the prediction of body weight in a dietary intervention program. Models that were utilised include Dynamic model, Machine Learning models (Linear regression, Support vector machine(SVM), Random Forest(RF), Artificial Neural Networks (ANN)). The performance of these estimation models was compared based on evaluation metrics like RMSE, MAE and R2. The results indicate that the Machine learning models(ANN and RF) perform better than the other models in predicting body weight at the end of the dietary intervention program.

Keywords: Weight and Obesity Management, Body weight and Weight-loss prediction, Supervised Machine Learning.

1 Introduction

The main purpose of any weight-loss intervention is to ensure long-term weight loss. The success of this depends on the initial weight loss. There is, therefore, a need to track the progress of the patient in the dietary intervention program. More so, there is a need to understand and consider the end-goal of patient's adaptability and ability to reach any degrees of weight change, which is related to different approaches for weight management. Modifications in lifestyle that is aiming at weight-loss weight loss are

often readily applied as a preferred treatment for overweight and obese patients. A lot of research work has gone into modelling body weight dynamics in humans [1, 2]. Some of the models aggregate observed clinical and laboratory data to make estimates of expected outcomes [3] which can be referred to as the statistical model [4]. The statistical model (1-D model) that predicts long term bodyweight seeks to require less numerous individual parameter estimates (variable inputs) which would be easier to implement in clinical practise [3]. It is known that statistical models aim to identify relationships between variables, but the predictive capabilities (in terms of their accuracies) of these statistical models are low [5].

This work aims to apply supervised machine learning methods, to predict future body weight for individuals attending the dietary intervention program at a lower margin of error. Additionally, we would like to find the best machine learning model and compare it with the statistical and dynamic model for the predictive analysis of body weight. In this study, the dietary intervention project is the Nutrient-gene interaction in human obesity (NUGENOB) [6] which was funded by the EU's 5th framework. It should be noted that since the during of the dietary intervention was in a period of 10 weeks, it is considered as a short-term dietary intervention program.

The paper is organised as follows. In the next section, we discuss related works followed by Data Collection method in section 3. Section 4 presents the methodologies for predictive analysis. We discuss in Section 5, the regression model evaluation metrics. Section 6 presents results, and section 7 presents the discussion, Section 8 discusses the conclusion and future work.

2 Related works

Most of the related works in relation to weight and obesity management in terms of predicting weight relied on statistical/ dynamic models. Therefore, this section will be split into two parts:

2.1 A dynamic model approach for body weight change management

From the mathematical point of view, the human body obeys the laws of energy conservation which is based on the first law of thermodynamics [1]. The body is called an open system because either low or high intake of food will add energy to the process. This energy balance equation is known to take the form of [1] :

$$R = \ I - E \tag{1}$$

Where I is the energy intake and E is the Energy expenditure. R is the rate of kcal/d that are stored or lost [2, 3]. This basis of energy transfer has helped to provide a better perception of how changes in any energy components can affect body weight change. Adopting this view, led to the development of compartmental equations with state variable tracking changes in energy derived from protein, fat, and carbohydrate [4, 5]. The Hall model was able to identify different state of transitions from energy intake to Energy expenditure which in the long run provides essential information into the mechanism behind human body weight change [4].

Forbes model also provides an intuition into how Fat-free fat mass (FFM) and fat mass are companions; i.e. increase/decrease in fat mass will be followed by an increase/decrease in FFM [6]. The equation for women relates to this;

$$FFM_{(t)} \; = \; 10.4 ln\left(\frac{F_{(t)}}{D}\right) \tag{2}$$

Where D $=$ 2.55 and $F_{(t)}$ is the Fat mass

$$FFM_{(t)} \; = \; 10.4 ln\left(\frac{F_{(t)}}{S}\right) \tag{3}$$

Where S $=$ 0.29 and $F_{(t)}$ is the Fat mass

Furthermore, Chow and Hall were able to create a more sophisticated method by coupling the fat-free mass(FFM) model proposed by Forbes equation through a two-dimensional dynamic model [7].

2.2 Machine learning approach for body weight management

As an alternative to the dynamic approach for body weight change dynamics, machine learning has proven to be useful due to its ability to perform predictive analysis and drawing inference on health data [8]. This section discusses various approach by which machine learning has come to play in management overweight and obesity.

Machine learning in medicinal services can be viewed as a type of preventive healthcare. Preventive healthcare services guarantee that measures are taken to forestall disease occurrence, as opposed to disease treatment [9]. There are various degrees of preventive healthcare strategies techniques. The methodology frequently utilized through machine learning system is secondary prevention of health care. This procedure plans to identify, analyze health conditions before the development of the symptoms of complications arising in general wellbeing/ health status of patients.

One of the capabilities of machine learning applications is the ability to identifying patterns in data. Such abilities can be utilized to early diagnosis of diseases and health conditions like cardiovascular diseases. Various techniques like Parameter De-creasing Methods (PDM) and Artificial Neural Network (ANN) were applied to identify variables related to the development of obesity [10]. This strategy distinguished 32 factors, for example, individual data about the way of life containing nutritional habits and genetic profile which are credited to potential elements expanding the danger of cardiovascular illnesses. The utilization of this combined technique yielded an accuracy of 77.89% in the approval tests in characterization assignments identifying with stoutness. The BMI examinations had an accuracy of 69.15% in the forecast of a risk factor for CVD as an independent factor [10]. This method has helped in detecting weight gain at the early stage of development. Early detection of weight gain could be a signal for patients in taking positive action about their lifestyle and simultaneously minimizing health care governmental costs [10, 11].

Furthermore, in the quest to applying secondary preventive health care, a machine learning system was created (a fuzzy logic system), which aimed at predicting degrees of obesity to guide the physician's decision making [12].The system was designed because it is perceived that BMI doesn't evaluate muscle to fat ratio precisely since it does exclude factors like age, sex, ethnicity, and bulk, giving a false diagnosis of body

fatness [13, 14]. Also, another technique that was utilised in predicting overweight/obesity based on body mass index and gene polymorphisms is "Decision trees".

Decision trees were utilised in predicting early healthy and unhealthy eating habits [15]. Decision trees helped in identifying allelic variants associated with rapid body weight gain and obesity [16].

Also, neural networks have enhanced the capacity to predict long-term outcomes [17] from pre-operative data in bariatric surgery patients substantially over linear [18] and logistic regression [19]. Neural networks may have similar potential to amplify predictions of long-term weight loss success during lifestyle interventions and long-term behavioural adherence to physical activity recommendations [20].

From another dimension of using the concept of unsupervised machine learning, analysis of the patient's profile at the entry of the dietary intervention program was carried out by applying K-means clustering [21] on NUGENOB data. It resulted in better the understanding of weight-loss threshold in a dietary intervention program.

3   Data collection

Health data analysis can enhance the efficiency of health policymakers to capture health-related issues [22]. Health data that are utilized in getting more insight into trends in diseases outbreak which can serve as the primary contact to an individual in a relevant population [23]. Concerning weight and obesity management, dietary intervention is a measure aimed at controlling body weight and obesity. Health data are usually generated during health intervention programs. An example of a health intervention program is Nutrient-gene interactions in human obesity (NUGENOB). The objective of this intervention program was, to examine if a 10-week low-fat hypo-energetic diet has a more beneficial effect on body weight, body composition and concentrations of fasting plasma lipids, glucose and insulin than a high-fat hypo-energetic diet [24, 25]. This was achieved by conducting a Randomised intervention trial with obese subjects from eight centres in seven European countries. The samples extracted from the NUGENOB database include attributes from individual subjects in the dietary intervention. These attributes are also referred to as the subject's profile.

The subject's profile is categorised as:

- Subject's Diet's Composition, e.g. High and Low-Fat food content
- Anthropometry Measurement, e.g. height and weight, waist
- Metabolic Rate Measurements
- Body composition measured by Bioimpedance analysis, e.g. fat mass and fat-free mass
- Biochemical Components Measurements, e.g. LDL - and HDL-cholesterol, and Fasting Plasma Insulin

The number of data-features is 25. It should be noted that not all features would be utilized for both machine learning and dynamic approach. Feature selection would be carried out for machine learning algorithms to select the best variable required for the predictive analysis. For better understanding of impact of diets on patients, the body weight percentage change distribution was calculated, and the distribution is shown in Figure 1.
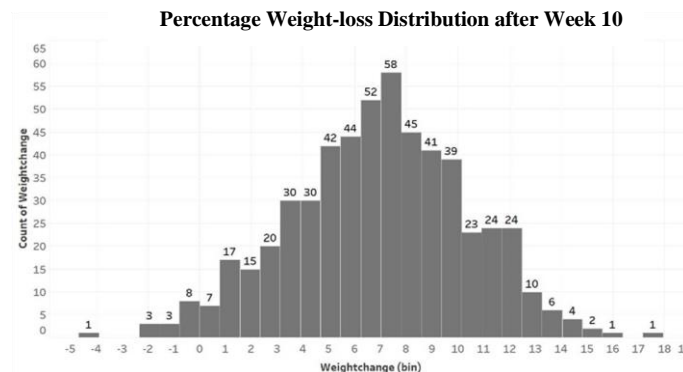
Percentage Weight-loss Distribution after Week 10

Fig 1: Percentage Body-Weight-distribution after Week 10

4   Methodologies

Two methods were applied in predicting body weight at the end of the dietary intervention program. They are:

4.1   Machine learning approach to body weight change dynamics

Machine learning (ML) are computational methods that computers use to make and improve predictions based on data. Machine Learning is a branch of artificial intelligence that can be used for predictive analytics. The use of machine learning in predictive analysis of body weight at the end of the dietary intervention program (week 10) seeks to evaluate the performance and capabilities on how it can provide augmented information for health care providers. In the long run, this is expected to improve the efficiencies of health care providers [26].

A blueprint on how input variables are structured in the machine learning models is shown in Figure 2. Our Machine learning models are implemented using the CARET in R [27]. The process involved in the machine predictive analytics is described in Figure 3. After the acquisition of data, the next phase is the feature selection. Feature selection in machine learning refers to techniques involved in selecting the best variables required for a predictive model.
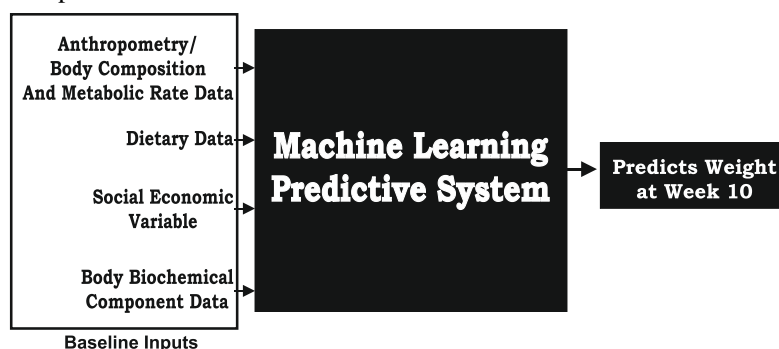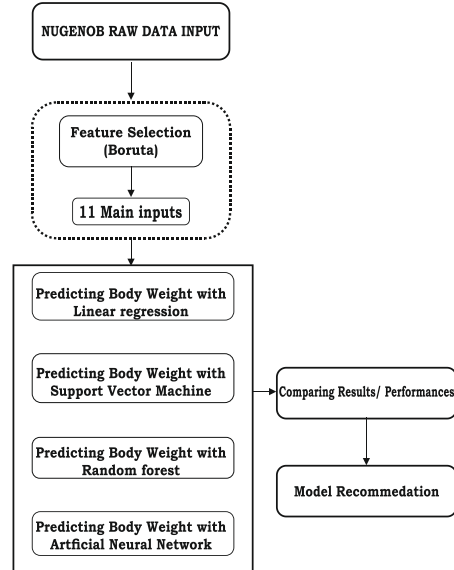

Fig 2: Machine learning blueprint for weight-change predictive analysis

Fig. 3: Machine learning predictive analytics processes

Feature selection is usually carried out to increase the efficiency in terms of computational cost and modelling and mostly to improve the performance of the model. Feature selection for machine learning modelling can be achieved using a method called Boruta. Boruta algorithm is a wrapper built around the Random Forest classification algorithm implemented in the R package RandomForest [28]. It provides a stable and unbiased selection of essential features from an information system. Boruta wrapper is run on the NUGENOB dataset with all the attributes, and it yielded 11 attributes as essential variables. The selected variables are shown in Table 1.

Table 1. Showing features selected variables

Variable Names		
1. Age	2. Body Weight@ Week 0 (Baseline)	3. Gender
4. Mean waist-hip ratio baseline	5. Fat mass baseline	6. Fasting glucose baseline
7. Basic metabolic rate baseline	8. Energy expenditure.T.0	9. Height
10. HOMA Insulin resistance(I0)	11. Fasting insulin baseline	

Machine Learning Models

The two common methods used in machine learning are supervised and unsupervised learning. Supervised learning technique is used when the historical data is available for a particular problem and this deem suitable in our case. The system is trained with the inputs (historical data) and respective responses and then used for the prediction of the response of new data. Conventional supervised approaches include an artificial neural network, Random Forest, support vector machines.

i.   Multivariate Regression Analysis

In multiple linear regression, there is a many-to-one relationship, between a wide variety of independent variables (input) variables and one dependent (response) variable. Including more input variables does not always mean that regression will be better or provide better predictions. In some cases, adding more variables can make things worse as it results in overfitting. The optimal scenario for a good linear regression model is for all of the independent variables (inputs) to be correlated with the output variable, but not with each other. Linear regression can be written in the form:

$$y_i = b_0 + b_1 + b_1 x_{i^1} + \ldots b_p x_{i^p} + e_i \tag{4}$$

Where y_i represents the numeric response for the ith sample, b_0 represents the estimated intercept, b_j represents the estimated coefficient for the jth predictor, e_i represents a Random error that cannot be explained by the model. When a model can be written in the form of the equation above, it is called *linear in the parameters*.

ii.   Support Vector Machine:

Support Vector machines are used for both classification and regression problems. When using this approach, a hyperplane needs to be specified, which means a boundary of the decision must be defined. The hyperplane is used to separate sets of objects belonging to different classes. Support Vector Machine can handle linearly separable objects and non-linearly separable objects of classes.

Mathematically, support vector machines [29] are usually maximum-margin linear models. When there exists no loss of generality that Y = {−1, 1} and that b = 0, support vector machines are works by solving the primal optimisation problem [30]. If there are exists non-linearly separable objects, methods such as kernels (complex mathematical functions) are utilized to separate the object which are members of different classes.

The most commonly used metric to measure the straight-line distance between two samples is defined as follows:

$$\min_{w,\xi} \left\{ \frac{1}{2} ||w||^2 + C \sum_{i=1}^{N} \xi_i \right\}$$
$$\text{Subject to: } \quad y_i(w \cdot x_i) \geq 1 - \xi_i, \ \xi_i \geq 0 \tag{5}$$

For a complete description of the random forest model, we refer the reader to [30].

iii.   Random Forest:

Random Forest is an ensemble learning technique in which numerous decision trees are built and consolidated to get an increasingly precise and stable prediction. The algorithm starts with a random selection of samples with replacement from the sample data. This sample is called a "bootstrapped" sample. From this random sample, 63% of the original observations occur at least once. Samples in the original data set which are not selected are called out of -bag observations. They are used in checking the error rate and used in estimating feature importance. This process is repeated many times, with each sub-sample generating a single decision tree. On the long run, it results in a forest of decision trees. The Random Forest technique is an adaptable, quick, machine learning algorithm which is a mixture of tree predictors. The Random Forest produces good outcomes more often since it deals with various kinds of data, including numerical, binary, and nominal. It has been utilized for both classification and regression. The

reality behind the Random Forest is the combination of Random trees to interpret the model. Furthermore, the Random Forest is based on the utilisation of majority voting and probabilities [31]. Random Forest is also good at solving overfitting issue. A more detailed process of the random forest model is explained in [31].

iv. Artificial Neural Networks:

Artificial neural network mimics the functionality of the human brain. It can be seen as a collection of nodes called artificial neurons. All of these nodes can transmit information to one another. The neurons can be represented by some state (0 or 1), and each node may also have some weight assigned to them that defines its strength or importance in the system. The structure of ANN is divided into layers of multiple nodes; the data travels from the first layer (input layer) and after passing through middle layers (hidden layers) it reaches the output layer, every layer transforms the data into some relevant information and finally gives the desired output [32]. Transfer and activation functions play an essential role in the functioning of neurons. The transfer function sums up all the weighted inputs as:

$$z = \sum_{x=1}^{n} w_i x_i + w_b \tag{6}$$

For a complete description of the artificial neural networks, we refer the reader to [32].

4.2 A dynamic modelling approach to Body weight change dynamics

The utilisation of weight-change models can enable patients to adhere to diets during a calorie restriction program. This is because weight change models generate predicted curves which is a form of diagnostic mechanism to test the difference between the actual predicted weight loss and actual weight loss [27]. There are numerous existing models, but they all require parameter estimates which on the long poses challenges for clinical implementation [27]. A new model was developed by [33], which provided a minimal amount of inputs from baseline variables(age, height, weight, gender and calorie intake. A blueprint on how input variables are structured in the dynamic models is shown in Figure 4.
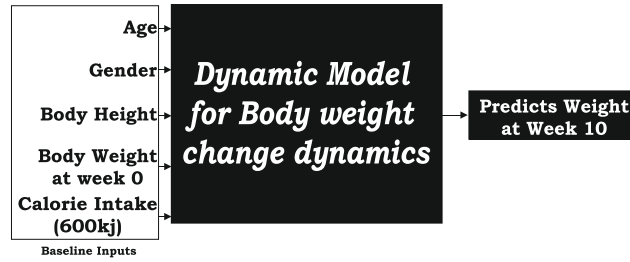

Fig 4: Dynamic model blueprint for weight-change predictive analysis

This new model is an improvement to Forbe's model [6] which takes in a minimal amount of input in calculating the fat-free mass. We refer to the reader to [33] complete biological details on the full model development, and present here is the performance

analysis of the dynamic model with other machine learning models. Dynamic modelling was implemented using the multi-subject weight change predictor simulator software [27].

5   REGRESSION MODEL EVALUATION METRICS

In the process of predictive analysis, errors will be generated. Such errors need to be measured to understand better the accuracies of machine learning algorithms. When the lower error is achieved, the better the predictive performance of the algorithms in term of accuracies. Dataset is split into a training set (80%) and test set (20%). The training set will be used to train the algorithm. The test set is usually used in measuring the performance of the algorithms. The training samples for this study case is 443 (train-set), while the testing samples(test-set) is 107 as depicted in Figure 5. The test set is usually called "the unseen data". The result from the machine learning algorithms (predicted result) will be compared to the test data (Unseen data). Thus, various metrics will be utilised in measuring the degree of error between the actual and predicted results. In this work, we used the Root Mean Square Error (RMSE), the Mean Absolute Error (MAE), the fitness degree R2, A residual plots, will be utilised in measuring the performance of these algorithms. These measures were standard in the literature, generally for prediction analytics tasks. These measures includes:

- **Mean Absolute Error (MAE)**

The MAE is used in measuring the accuracy of continuous variables. The errors generated from prediction analysis using the four selected algorithms are presented in Table III.

- **Root Mean Square Error (RMSE)**

The RMSE is essential to measure the prediction's accuracy because it allows the error to be the same magnitude as the quantity being predicted [34]. For best predictability, lower RMSE is needed.

- **Coefficient of multiple determination R^2 coefficients**

R^2 coefficient also called fitness degree. Better performances are achieved when R^2 values are near 1. Ideally, if $R^2 = 1$, the original series and the predicted one would be superimposed. For best predictability, higher R^2 is needed.
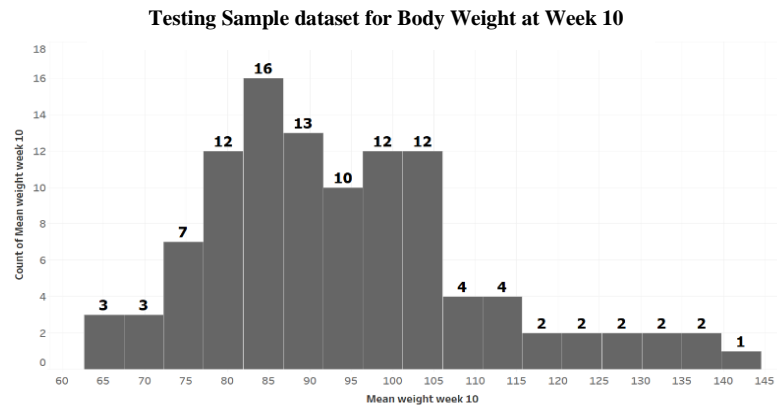
Testing Sample dataset for Body Weight at Week 10

Fig 5: Testset Body-weight at Week 10 Distribution

Table 2. Model comparison performance based on different metrics

	LM	SVM	RF	ANN	Dynamic Model
RMSE	4.309964	4.349037	3.268409	3.55828	4.629934
MAE	3.438419	3.493781	2.64141	2.763006	3.791817
R-SQUARED	0.930191	0.930122	0.960241	0.953391	0.954028

Key: LM- Linear Regression, SVM- Support Vector machine,
RF- Random Forest, ANN- Artificial Neural Network

6 RESULTS

The predictive analysis was carried out on predicting body weight at week 10 using both Dynamic models and machine learning models. Four machine learning algorithms (Linear regression, Support Vector Machine, Random Forest and Artificial neural networks) were used for the predictive analysis of weight at the end dietary intervention program (week 10). The performance of each machine learning algorithms and dynamic model are carried out by calculating the mean absolute error and root mean square based on the test-set are presented in Table 2. The test-set sample distribution is displayed in Figure 5.

It shows that Random Forest has the lowest error in predicting capabilities for weight at week 10. Furthermore, in terms of R-square, as illustrated in Table 2, dynamic models and machine learning models achieved R-square of over 93%. Random Forest has the highest R-square value of 96%, which is a very good fit.

7 Discussion

It should be noted that during the dietary intervention, two types of diets were administered to patients. The two types of diets are low and high hypo-energetic diets, i.e. the total calories in these diets is 600kcal. These types of diets for obesity and weight management in this study assumes that all participants adhere strictly to 600kcal per day in other to have a considerable weight-change or weight-loss. Machine learning and dynamic models come into play in the predictability of future change in the body weight of participants in a dietary intervention. The quest is to either predict actual body weight or to predict weight-loss at the end of the dietary intervention program. In this study, body weight-loss means quantifying by how much participants would lose weight during the diet intervention program.

The results for predicting actual body weight and computing the body weight-loss are explained later on in the text. From the above results (Table II), Random Forest and Artificial neural networks algorithm perform best in predicting body weight, i.e. actual body weight at week 10 (last week of diet intervention). The result is evaluated by calculating their predictive error (Mean Absolute Error). The errors were found to be 2.64 and 2.76kg. Having a lower error is one of the most important factors in predictive analysis.

Linear regression is a form of a statistical model that identify the relationships between variables. However, it comes at the cost of predictive capabilities, which will always result in a higher error [35]. This is reflected in the high mean absolute error

for the linear regression model (±3.438kg). Dynamic models in the context of predicting body weight change, explains energy inflows and outflows in the body, based variables like age, height, weight baseline and gender in relation to time. Utilising dynamic models in predicting short term body-weight produces a high error (Mean Absolute Error) as compared to other Machine learning models with a lower error. However, in terms of achieving a good r-square (variance), both dynamic models and machine learning models are good to explain the variation of the dependent variables from the independent variables(s).Applying Boruta method of feature selection for predicting body-weight at week 10 (last week of diet intervention), variables such as initial body weight at week 0, initial fat-mass at week 0, and energy expenditure and basic metabolic rate play a significant role for predicting body weight at week 10. The technique utilized in identifying variable's relationship with the response as compared to other variables used in the model is Random Forest variable importance.

From the results obtained, a mean absolute error of ±2.6kg is achieved from utilizing Random Forest algorithm. In order to expatiate more on this; assuming we have a 10% reduction of actual body weight (94.23kg) at week 10, it corresponds to 9.42kg weight loss. The model would have predicted 84.81kg ±2.6kg i.e. 3% error in prediction.

Also, when computing weight-loss, comparing the actual percentage weight-loss with the predicted percentage weight-loss, i.e. actual weight loss (9.9%) and predicted average weight-loss (7.23%), the percentage prediction error for the weight-loss would be up to 27.6%. This error incurred in weight-loss prediction analysis is high. It is quite easy from this study to predict the actual short-term body weight than short term body weight loss.

Mean absolute error (MAE) has been utilized to interpret the effect of the error on the models. In order to utilize RMSE, the underlying assumption when presenting the RMSE is that the errors are unbiased and follow normal distribution [36]. Conducting the Shapiro-Wilk normality test on the test-set. We had a p-value of 0.00276 (p-value < 0.05) that shows that the distribution of the residuals are significantly different from normal distribution which is also described in Figure 5. This reflects that the distribution in the test set is not normally distributed. Hence, RMSE cannot be fully relied on. RMSE and the MAE are defined differently, we should expect the results to be different. Both metrics are usually used in assessing the performance of the machine learning model. Various research indicates that MAE is the most natural measure of average error magnitude, and that (unlike RMSE) it is an unambiguous measure of average error magnitude [37]. It follows that the RMSE will increase (along with the total square error) at the same rate as the variance associated with the frequency distribution of error magnitudes increases which in turn will make the RMSE always greater than MAE. Therefore, in this study, MAE would be the main metrics utilized in assessing the model.

8   Conclusion and Future Work

One of the strengths of Random Forest is its ability to perform much better than other machine learning algorithms due to its ability to handle small data sets. Since Random Forest technique is majorly on building trees, it tries to capture and identify patterns with a small dataset and is still able to generate a minimal error. In our case, the training dataset is 443. In contrast to neural networks, it needs more dataset to train with, and

thus with the current dataset, it is expected that the predictive power of neural networks would be much lower than Random Forest.

Comparing the computation performance for other models, it is evident that the Random Forest performs best in the predictive analysis of body weight. Computationally, machine learning models achieve lower predictive error compared to dynamic models in predicting short-term body weight. However, from the clinical point of view, the minimum mean absolute percentage error produced from the discussed model (Random Forest) in predicting body weight-loss is still high. The research work shows the capability of both machine learning models and dynamic model in predicting body weight and weight-loss. Future work includes hybridisation of machine learning and dynamic models in predicting body weight-loss that are represented in terms of classes, i.e. High, Medium and Low weight loss. This approach will provide more solution in body weight-loss predictability. Also, further research work could address the inclusion of dietary type for the predictive analysis, which can also provide information on which diet to recommend under a specific set of conditions.

Acknowledgement

References

[1]   H. S. Fogler, Elements of chemical reaction engineering, p. 876.

[2]   E. Jéquier and L. Tappy, "Regulation of Body Weight in Humans," *Physiological Reviews,* vol. 79, no. 2, pp. 451-480, 1999.

[3]   McArdle, William D & Katch, Victor L & Katch, Frank, Exercise physiology : energy, nutrition, and human performance, Philadelphia: Lippincott Williams & Wilkins, 2001.

[4]   K. D. Hall, "Computational model of in vivo human energy metabolism during semistarvation and refeeding," *American Journal of Physiology-Endocrinology and Metabolism,* vol. 291, no. 1, pp. E23-E37, 1 7 2006.

[5]   K. D. Hall, "Predicting metabolic adaptation, body weight change, and energy intake in humans," *American Journal of Physiology-Endocrinology and Metabolism,* vol. 298, no. 3, pp. E449-E466, 24 11 2009.

[6]   G. B. Forbes, "Lean Body Mass-Body Fat Interrelationships in Humans," *Nutrition Reviews,* vol. 45, no. 10, p. 225–231, October 1987.

[7]   C. C. Chow and K. D. Hall, "The Dynamics of Human Body Weight Change," *PLOS Computational Biology,* vol. 4, no. 3, pp. e1000045-, 28 3 2008.

[8]   M. Mast, *Intel Healthcare-analytics-whitepaper.*

[9]   Leavell, H. R., & Clark, E. G., The science and art of preventing disease, prolonging life, and promoting physical and mental health and efficiency, NY: Krieger Publishing Company., 1979.

[10] Valavanis IK, Mougiakakou SG, Marinos S, Karkalis G, Grimaldi KA, Gill R, et al, "Gene - Nutrition Interactions in the Onset of Obesity as Cardiovascular Disease Risk Factor based on a Computational Intelligence Method," in *IEEE International Conference on Bioinformatics and Bioengineering*, 2008.

[11] Finkelstein EA, Trogdon JG, Brown DS, Allaire BT, Dellea PS, Kamal-Bahl SJ., "The lifetime medical cost burden of overweight and obesity: implications for obesity prevention," *Obesity (Silver),* vol. 16, no. 8, pp. 1843-8, 2008.

[12] Bruna Marmett, Roseana Böek Carvalhoa,Melissa Santos Fortesb, Sílvio César Cazellab, "Artificial Intelligence technologies to manage obesity," *Vittalle – Revista de Ciências da Saúde ,* vol. 30, no. 2, pp. 73-79, 2018.

[13] Heo M, Faith MS, Pietrobelli A, Heymsfield SB, " Percentage of body fat cutoffs by sex, age, and race-ethnicity in the US adult population from NHANES 1999-2004," *Am J Clin Nutr,* vol. 95, no. 3, pp. 594-602., 2012.

[14] Shah NR, Braverman ER. , "Measuring adiposity in patients: the utility of body mass index (BMI) percent body fat, and leptin.," *PLoS One 2012,* vol. 7, no. 4, p. e33308., 2012.

[15] G. Spanakis, G. Weiss, B. Boh, V. Kerkhofs and A. Roefs, "Utilizing Longitudinal Data to Build Decision Trees for Profile Building and Predicting Eating Behavior," *Procedia Computer Science,* vol. 100, pp. 782-789, 1 1 2016.

[16] C. Rodríguez-Pardo, A. Segura, J. J. Zamorano-León, C. Martínez-Santos, D. Martínez, L. Collado-Yurrita, M. Giner, J. M. García-García, J. M. Rodríguez-Pardo and A. López-Farre, "Decision tree learning to predict overweight/obesity based on body mass index and gene polymporphisms," *Gene,* vol. 699, pp. 88-93, 30 5 2019.

[17] M. Diana M. Thomas, PhD; patrick Kuiper, MS; Hinali Zaveri, MD; Amit Surve, MD; and Daniel R. Cottam, "Neural Networks to Predict Long-term Bariatric Surgery Outcomes," *Bariatric Times,* vol. 14, no. 12, p. 14–17, 2017.

[18] Courcoulas AP et al. , "Preoperative factors and 3-year weight change in the Longitudinal Assessment of Bariatric Surgery (LABS) consortium.," *Surg Obes Relat Dis,* vol. 11, no. 5, p. 1109–1118, 2015.

[19] Hatoum IJ et al, "Clinical factors associated with remission of obesity-related comorbidities after bariatric surgery," *JAMA Surg,* vol. 151, no. 2, p. 130–137, 2016.

[20] K. W. Degregory, P. Kuiper, T. Desilvio, J. D. Pleuss, R. Miller, J. W. Roginski and C. B. Fisher, "Obesity / Data Management A review of machine learning in obesity," 2018.

[21] O. Babajide, H. Tawfik, A. Palczewska, A. Gorbenko, A. Astrup, J. A. Martinez, J. Oppert and T. I. Sørensen, "Application of Unsupervised Learning in Weight-Loss Categorisation for Weight Management Programs," in *2019 10th International Conference on Dependable Systems, Services and Technologies (DESSERT)*, 2019.

[22] McGraw-Hill Concise Dictionary of Modern Medicine McGraw-Hill. , 2002.

[23] Charles Safran, MD, MS, Meryl Bloomrosen, MBA, ∗ W. Edward Hammond, PHD, Steven Labkoff, MD, Suzanne Markel-Fox, PHD, Paul C. Tang, MD, f Don E. Detmer, MD, MA, "Toward a National Framework for the Secondary Use of Health Data: An American Medical Informatics Association White Paper," *Journal of the American Medical Informatics Association,* vol. 14, no. 1, pp. 1-9, 2007.

[24] M. Petersen, M. Taylor, W. Saris, C. Verdich, S. Toubro, I. Macdonald, S. Ro¨ssner, V. Stich, B. Guy-Grand, D. Langin, J. Martinez, O. Pedersen, C. Holst, T. Sørensen, A. Astrup and T. N. Consortium, "Randomized, multi-center trial of two hypo-energetic diets

in obese subjects: high- versus low-fat content," *International Journal of Obesity,* vol. 30, pp. 552-560, 2006.

[25] T. I. A. Srensen, P. Boutin, M. A. Taylor, L. H. Larsen, C. Verdich, L. Petersen, C. Holst, S. M. Echwald, C. Dina, S. Toubro, M. Petersen, J. Polak, K. Clment, J. A. Martnez, D. Langin, J.-M. Oppert, V. Stich, I. Macdonald, P. Arner, W. H. M. Saris, O. Pedersen, A. Astrup, P. Froguel and T. N. Consortium, "Genetic Polymorphisms and Weight Loss in Obesity: A Randomised Trial of Hypo-Energetic High- versus Low-Fat Diets," *PLOS Clinical Trials,* vol. 1, no. 2, pp. e12-, 30 6 2006.

[26] A. Bartley, "Predictive Analytics in Healthcare: A data-driven approach to transforming care delivery," Intel.

[27] D. M. Thomas, C. K. Martin, S. Heymsfield, L. M. Redman, D. A. Schoeller and J. A. Levine, "A Simple Model Predicting Individual Weight Change in Humans," *Journal of biological dynamics,* vol. 5, no. 6, pp. 579-599, 2011.

[28] A. Liaw Merck, A. Liaw and M. Wiener, "Classification and Regression by RandomForest," 2002.

[29] C. Cortes, "Support-Vector Networks," 1995.

[30] G. Louppe, "Understanding Random Forests: From Theory to Practice," 28 7 2014.

[31] A. R. Hidalgo-Muñoz, M. M. López, I. M. Santos, M. Vázquez-Marrufo, E. W. Lang and A. M. Tomé, "Affective Valence Detection from EEG Signals Using Wrapper Methods," in *Emotion and Attention Recognition Based on Biological Signals and Images*, InTech, 2017.

[32] T. Kohonen, "An introduction to neural computing," *Neural Networks,* vol. 1, no. 1, pp. 3-16, 1988.

[33] D. M. Thomas, A. Ciesla, J. A. Levine, J. G. Stevens and C. K. Martin, "A mathematical model of weight change with adaptation," *Mathematical biosciences and engineering : MBE,* vol. 6, no. 4, pp. 873-887, 10 2009.

[34] J. Han, M. Kamber and J. Pei, "Data Mining. Concepts and Techniques, 3rd Edition (The Morgan Kaufmann Series in Data Management Systems)," 2011.

[35] M. Mastanduno, "Machine Learning Versus Statistics: When to use each," healthcatalyst, 23 May 2018. [Online]. Available: https://healthcare.ai/machine-learning-versus-statistics-use/. [Accessed 27 December 2019].

[36] T. Chai and R. R. Draxler, "Root mean square error (RMSE) or mean absolute error (MAE)? -Arguments against avoiding RMSE in the literature," *Geoscientific Model Development,* vol. 7, no. 3, pp. 1247-1250, 30 6 2014.

[37] C. J. Willmott and K. Matsuura, "Advantages of the mean absolute error (MAE) over the root mean square error (RMSE) in assessing average model performance," *Research,* vol. 30, no. 1, pp. 79-82, 2005.

[38] Gropper, Sareen S, Smith, Jack L and Groff, James L, Advanced nutrition and human metabolism, CA: Thomson Wadsworth, Belmont,, 2005..

[39] E. Christiansen and L. Garby, "Prediction of body weight changes caused by changes in energy balance," *European Journal of Clinical Investigation,* vol. 32, no. 11, pp. 826-830, 1 11 2002.