Using Genetic Programming to Investigate a Novel Model of Resting Energy Expenditure for Bariatric Surgery Patients

James Alexander Hughes
St. Francis Xavier University
Computer Science
Antigonish, NS, Canada
jhughes@stfx.ca

Ryan E.R. Reid
St. Francis Xavier University
Human Kinetics
Antigonish, NS, Canada
rreid@stfx.ca

Sheridan Houghten
Brock University
Computer Science
St. Catharines, ON, Canada
shoughten@brocku.ca

Ross E. Andersen
McGill University
Kinesiology & Physical Education
Montreal, PQ, Canada
ross.andersen@mcgill.ca

***Abstract*—Traditionally, models developed to estimate resting energy expenditure (REE) in the bariatric population have been limited to linear modelling based on data from 'normal' or 'overweight' individuals — not 'obese'. This type of modelling can be restrictive and yield functions which poorly estimate this important physiological outcome.**

Linear and nonlinear models of REE for individuals after bariatric surgery are developed with linear regression and symbolic regression *via* genetic programming. Features not traditionally used in REE modelling were also incorporated and analyzed and genetic programming's intrinsic feature selection was used as a measure of feature importance.

A collection of effective new linear and nonlinear models were generated. The linear models generated outperformed the nonlinear on testing data, although the nonlinear models fit the training data better. Ultimately, the newly developed linear models showed an improvement over existing models and the feature importance analysis suggested that the typically used features (age, weight, and height) were the most important.

***Index Terms*—Bariatric Surgery; Genetic Programming; Resting Energy Expenditure; Regression Analysis**

I. Introduction

Bariatric (weight loss) surgery is the most effective treatment for severe obesity [3]. After surgery, patients must follow a strict diet to promote optimal weight loss and avoid regaining weight [11]. These dietary instructions are based on an estimation of patient's resting energy expenditure (REE). Traditionally, models developed to estimate REE in the bariatric population (see Equations 1 – 4) have been limited to linear modelling [15]. This type of modelling may be restrictive, and yield functions which poorly estimate this important physiological outcome. Additionally, the original models did not attempt to fit data from obese or post-bariatric surgery individuals.

A more accurate energy prediction equation is required to reduce the risk of weight regain following surgery. In this study, we aim to apply linear regression and symbolic regression to estimate both linear and nonlinear models that will more accurately predict the REE of individuals after bariatric surgery. Although the ultimate goal is to develop new models to be used in a clinical setting, here we focus on feature analysis, model interpretation, and a comparison of the effectiveness of linear and nonlinear models.

In addition to ordinary least squares (OLS) for generating linear models, *Genetic Programming* (GP), a type of computational intelligence, will be used to perform symbolic regression to develop the nonlinear models.

A description of the data is presented in Section II and the model development strategy is outlined in Section III. Multiple sub-experiments are outlined, and their results are presented in Section IV. A discussion of the results are also included in Section IV. Section V contains the findings and outlines a number of future directions the longer term project will take.

II. Data

One-hundred and twenty-seven individuals who had previously undergone bariatric surgery (1 to 17 years prior to this study's assessments) were recruited by telephone. This study was approved by the McGill University Medical Ethics Review Board (A04-M35-11A). Written informed consent was obtained from all participants.

All participants were examined in the morning (8 AM – 11 AM) after an overnight fast. During the in-laboratory assessment, participant height was measured to the nearest centimeter (Seca 216 stadiometer) and weight was assessed to the nearest tenth kilogram (Seca 635 bariatric scale) (Seca, Hamburg, Germany). Body mass index was calculated as body weight/height^2 (kg/m^2). Lightweight clothing and no footwear were worn during all assessments.

Resting energy expenditure (REE) was measured by indirect calorimetry (mREE) using a ventilated hood system (Vmax Encore; CareFusion, Yorba Linda, CA, USA) while participants remained in a lying, resting position for 40 minutes. Oxygen uptake and carbon dioxide production were collected every minute and data were converted to REE. The resting respiratory quotient (RQ) was estimated as the ratio of carbon dioxide production to oxygen uptake. Before each measurement, the gas analyzer was calibrated with standard gas concentration (95% O_2 + 5% CO_2).

TABLE I: Summary Statistics of Subjects. Sample standard deviation is used. A confidence interval of 95% was used.

	Male	Female	Combined
n	29	90	119
Age mean	52.76	50.91	51.36
Age st.dev.	11.80	9.36	9.99
Age CI	± 2.12	± 1.68	± 1.79
Weight (kg) mean	109.10	91.12	95.50
Weight (kg) st.dev.	22.66	27.41	27.46
Weight (kg) CI	± 4.07	± 4.92	± 4.92
Height (cm) mean	172.35	164.19	166.18
Height (cm) st.dev.	6.99	8.63	8.95
Height (cm) CI	± 1.26	± 1.55	± 1.61
REE ($\mathrm{kcal\,d^{-1}}$) mean	1594.14	1339.90	1401.86
REE ($\mathrm{kcal\,d^{-1}}$) st.dev.	330.24	372.50	377.56
REE ($\mathrm{kcal\,d^{-1}}$) CI	± 59.33	± 66.93	± 67.84

TABLE II: Feature/variable labels.

Resting Energy Expenditure	REE
Age	A
Sex	S
Weight/Current Weight	W/W(C)
Before Surgery Weight	W(B)
Height	H
Respiration Quotient	RQ
Years Since Surgery	Y

Table I provides a summary of the main features used in this work. Although additional features are used later, their summary statistics are not presented here.

At various points throughout this paper we refer to features with abbreviations, which are summarized in Table II.

III. Methods

To find the models, multiple forms of regression analysis are performed. The focus of this work is to compare linear and nonlinear models and to investigate their effectiveness.

Models are designed to predict the measured REE (y) based on a number of independent variables, namely, *age*, *weight*, and *height* (X). Additional independent variables (sex, respiration quotient, years since surgery, and weight before surgery) are investigated in Section IV-E. The goal of the regression analysis is to generate some symbolic mathematical equation that describes the relationship between the dependent and independent variables.

A. Linear Regression

Ordinary least squares regression is used for linear regression. This simple form of statistical learning determines the optimal parameters, β, for the independent variables X to predict y; ultimately, it determines $\hat{y} = \beta X$, where $\hat{y}$ is the predicted REE value.

Least absolute shrinkage and selection operator (LASSO) regression is used in Section IV-E as a mechanism to determine feature importance. The effectiveness of the resulting LASSO models are not included in the results comparison as initial tests showed little to no difference between the LASSO generated models and those developed with OLS.

Given the small number of data points, *leave-one-out* (LOO) cross validation is used.

B. Nonlinear Regression

Nonlinear regression was performed with symbolic regression (SR) via GP. GP is a form of evolutionary computation that searches for *programs* to solve a given problem. These programs are *evolved* with an algorithm based on the natural process of evolution [9]. For SR, the programs being evolved are mathematical expressions. In this study, SR is evolving expressions to predict REE, and their effectiveness is calculated by comparing the predicted values to the actual measured REE values for each patient.

SR searches for some function of the independent variables X to predict y; $\hat{y} = f(X)$, where $\hat{y}$ is the predicted REE value. Unlike linear regression, SR is less constrained and requires fewer assumptions. The resulting functions may be nonlinear, require no *a priori* model structure, and can include other basic functions. Ultimately, SR searches for model structure in addition to some parameter values.

1) Genetic Programming System: The GP system used for this work is a specialized system for SR [5]. The system was originally based on Schmidt *et al.*'s work [14] and has since been used in multiple application areas, including finding nonlinear relationships with fMRI data [7], predicting intracranial pressure [8], and modelling Parkinson's Disease patient gait [6].

The system incorporates many improvements, including *fitness predictors* to address overfitting and reduce the computation required for fitness evaluation [13]. Fitness predictors also provide a pseudo model validation throughout the evolutionary search since at any given time, the search is only fitting to a relatively small subset of data points. The system also uses an *acyclic graph representation*, which reduces bloat, overfitting, increases search effectiveness, and allows the search to reuse subexpressions [12]. For more information on these improvements, please see their original sources. Table III provides a summary of the GP system settings. These values were determined empirically.

Given the stochastic nature of SR, 100 nonlinear models were generated for each set of data to give the algorithm a better chance of producing effective results. Further, having more models allows for better statistical comparisons.

Each nonlinear model took roughly $1\mathrm{min}$ to generate on a desktop computer.

Given the computational cost of generating the nonlinear models compared to the linear models, no validation or sophisticated model selection strategy was implemented for the nonlinear models. Simply, the best performing nonlinear model on the training data would be selected for testing where applicable. Note that this strategy will select biased models overfit to the training data, especially considering the small number of data points; however, it is sufficient given a major focus of this work is on the model analysis and the constraints,

TABLE III: Genetic Programming system parameters.

Elitism	1 (Single top candidate solution)
Population	101
Subpopulations	7
Generations	100,000 (1,000/migration)
Migrations	100
Crossover	80%
Mutation	10% (x2 chances)
Fitness Metric	Mean Squared Error: $\frac{1}{n}\sum_{i=1}^{n}(\hat{y_i} - y_i)^2$
Language	$+, - ,* , /, exp, sin, cos, tan$
Max # Graph Nodes	24
Predictors	10
Predictor Pop. Size	25% of data
Trainers	8

such as the small sample size. Once more data is available, a more rigorous model selection strategy will be employed for a more robust comparison.

IV. Results and Discussion

A. Existing Models

The popular existing models within the literature are *Harris-Benedict* [4] and *Mifflin St. Jeor* [10]. These are shown as Equations 1 and 2 for male and female subjects respectively (Harris-Benedict), and Equations 3 and 4 for male and female respectively (Mifflin St. Jeor).

$$f(A, W, H) = -6.755 * A + 13.75 * W + 5.003 * H + 66.50 \tag{1}$$

$$f(A, W, H) = -4.676 * A + 9.563 * W + 1.850 * H + 655.1 \tag{2}$$

$$f(A, W, H) = -5 * A + 10 * W + 6.25 * H + 5 \tag{3}$$

$$f(A, W, H) = -5 * A + 10 * W + 6.25 * H - 161 \tag{4}$$

The mean absolute error (MAE) values obtained by these equations when applied to the data recorded from the post bariatric surgery patients are found in Table IV. Within this paper we make no attempt for a rigorous comparison of the new models to the existing ones as this work focuses on the generation of linear and nonlinear models fit to post bariatric surgery patients and their comparison to one another. Further, the summary results in Table IV clearly show the new models (discussed below) outperforming the existing ones on this post bariatric surgery data. A more thorough analysis and comparison of the new models to the existing will be performed in future work.

B. Fitting All Data

The first experiment was performed by simply fitting models to all the data available. This is not ideal since one would want to validate and test the results, however these results will serve as a baseline for further experiments. Moreover, the amount of data currently available is minimal.

TABLE IV: Mean Absolute Error (MAE) values obtained from the two existing resting energy expenditure models when applied to post bariatric surgery and MAE values obtained by two newly generated models for the same data — linear model (Ordinary Least Squares) and a nonlinear model (Symbolic Regression).

Model	MAE Male	MAE Female	MAE Combined
Harris-Benedict	495.95	289.46	339.78
Mifflin-St.Jeor	349.50	234.39	262.44
Ordinary Least Squares	169.93	167.00	167.71
Symbolic Regression	135.21	151.06	147.20

The linear model obtained with OLS for the male patients is presented in Equation 5 and the model for the female patients is presented in Equation 6.

$$f(A, W, H) = -13.527 * A + 5.998 * W + 10.385 * H - 136.31 \tag{5}$$

$$f(A, W, H) = -6.16 * A + 9.695 * W + 5.27 * H - 95.25 \tag{6}$$

Although 100 nonlinear models were generated for both the male and female data, only the single top model based on error was selected. The nonlinear models are not presented here as they are complex. It should be emphasized that although the nonlinear models generated with SR are mathematical expressions and interpretable, especially compared to other forms of sophisticated computational intelligence techniques, they are still much more difficult to interpret in comparison to the simpler linear models.

In addition to the existing models' results, Table IV shows the MAE values obtained by the linear and nonlinear models. The error value in the *combined* column is the error regardless of sex. Immediately one can see that the nonlinear models are more effective at fitting the data, however these results are likely to be overfit.

Figure 1's left plot presents a visualization of the error between the linear model's predicted REE values and the measured REE values for each subject. The predicted REE value is plotted against the measured value. If the predictive model was perfect, each point would be on the $y = x$ line.

One can visually see that the predictive model is reasonably accurate, even for the patients with high REE values. The data points/error values are well distributed above and below the line, indicating a normal distribution of errors. It does appear that the patients with very high measured REE values are the outliers, which is reasonable since there were fewer data points to fit to with these high values.

Similarly, the right plot in Figure 1 shows the error between the nonlinear model's predicted REE value and the measured REE values for each subject. Similar observations can be made here as before with the linear models, however the nonlinear model does appear to be better able to handle the outliers (patients with high REE values).

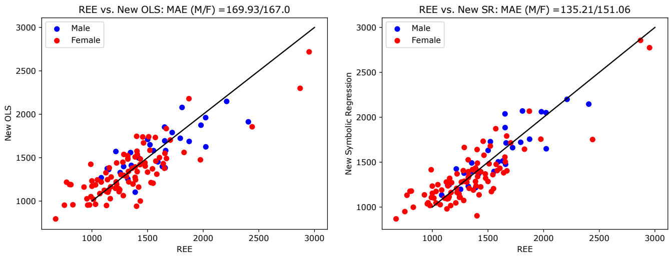

Fig. 1: Scatter plots showing the newly generated OLS model predicted REE (left) and SR nonlinear model predicted REE (right) compared to measured REE of post bariatric surgery patients. Data is separated into male and female.

1) Feature Relationships: In addition to the normally distributed errors, it is important to ensure that the assumptions for ordinary least squares are true. Namely, one must that (a) there are linear relationships between the dependent and independent variables, and (b) the independent variables are not linearly related.

Figure 2 shows relationships between the variables for the male and female patients. Visually, one can see that the relationships between the independent variables (left plots) do not seem to be linear. There could be an argument made for a linear relationship between weight and height, which is reasonable to assume, however the line is nearly horizontal as there was minimal variability in height of the patients when compared to their weights. When investigating the correlation matrices (right plots), the Pearson correlation between the weight and height is still low for male (0.02) and medium for female (0.55).

2) Cross Validation: A simple leave-one-out (LOO) strategy was used to test OLS. The mean training MAE, the mean validation MAE (LOO), and 95% confidence intervals for the linear models generated with OLS are presented in Table V. Column *Best Train* presents the mean absolute error of the model with the lowest training error applied to the left out data, and the column *Best LOO* is the smallest validation error.

The rows labeled *All* contain values corresponding to models generated for patients, regardless of sex (a single model was made for both sexes, discussed in further detail below).

Given the computational cost of generating the nonlinear models with symbolic regression, no explicit validation was performed for the nonlinear models. As outlined in Section III-B1, the GP system does contain a pseudo validation built into evolution through the use of fitness predictors. Further, unlike OLS, symbolic regression does not need LOO for the purpose of generating a set of potential models since the algorithm is stochastic.

When investigating the results, one can see that the results for the linear and nonlinear models are similar, however conclusions cannot be made after this simple analysis and proper testing is required.

C. Train, Validate, and Test

Although there are minimal data points, it is important to perform proper training and testing to better understand the effectiveness of the new models. For both male and female, the data was shuffled and 60% of the data was used for training and the remaining 40% was kept for testing purposes. This resulted in 17 male patients for training and 12 for testing, and 54 female patients for training and 36 for testing.

Linear models generated with OLS for the male and female subjects are presented in Equations 7 and 8 respectively. Note the similarity in the parameters for the measured patient weight.

$$f(A,W,H) = -8.266 * A + 8.527 * W + 20.043 * H - 2315.633 \tag{7}$$

$$f(A,W,H) = -4.121 * A + 8.472 * W + 7.584 * H - 466.239 \tag{8}$$

Figure 3 present the model predicted REE and the measured REE for the linear and nonlinear models. These plots separate the training data and testing data. By chance, the outlier data happened to be in the testing data. Unlike before, the nonlinear models did not seem to perform better on the outliers. Further, one can visually see that the nonlinear model for the male patients performed poorly on the testing data, which is understandable given the very few data points available to fit.

Similar to before, Table VI presents summary statistics of the linear and nonlinear models. The models were fit to all training data points and then applied to the testing data. Again, 100 nonlinear models were generated, but the single top model based on training error was selected. Unsurprisingly, the nonlinear models fit the training data better than the linear, but performed poorly on the testing data, which indicates overfitting. Note however that a large difference between the training and testing results can be seen with the linear models, which emphasizes the importance for a more robust model selection strategy.

1) Cross Validation: For the linear models, LOO validation was done with the training data. This provides a set of linear models to choose from, and then we can simply select the top validated model to apply to the withheld testing data. Table VII contains summary statistics for the results. The *Train* and *Val.* columns are the average MAE values for all models. Like the results in Table V, there are no validation results for the nonlinear models. Again, the model selection strategy was to select the top performing nonlinear model based on training data. This simple strategy is expected to overfit the training data. The *Actual Min* column is the actual best error obtained on the testing data by all generated models.

Within Table VII, the linear models did, on average, fit the training data minimally better than the 100 nonlinear models, however the best nonlinear models were much better than the best linear. When looking at the test error, the linear models performed much better (again, a better selection strategy was used for the linear models). Lastly, note that there were

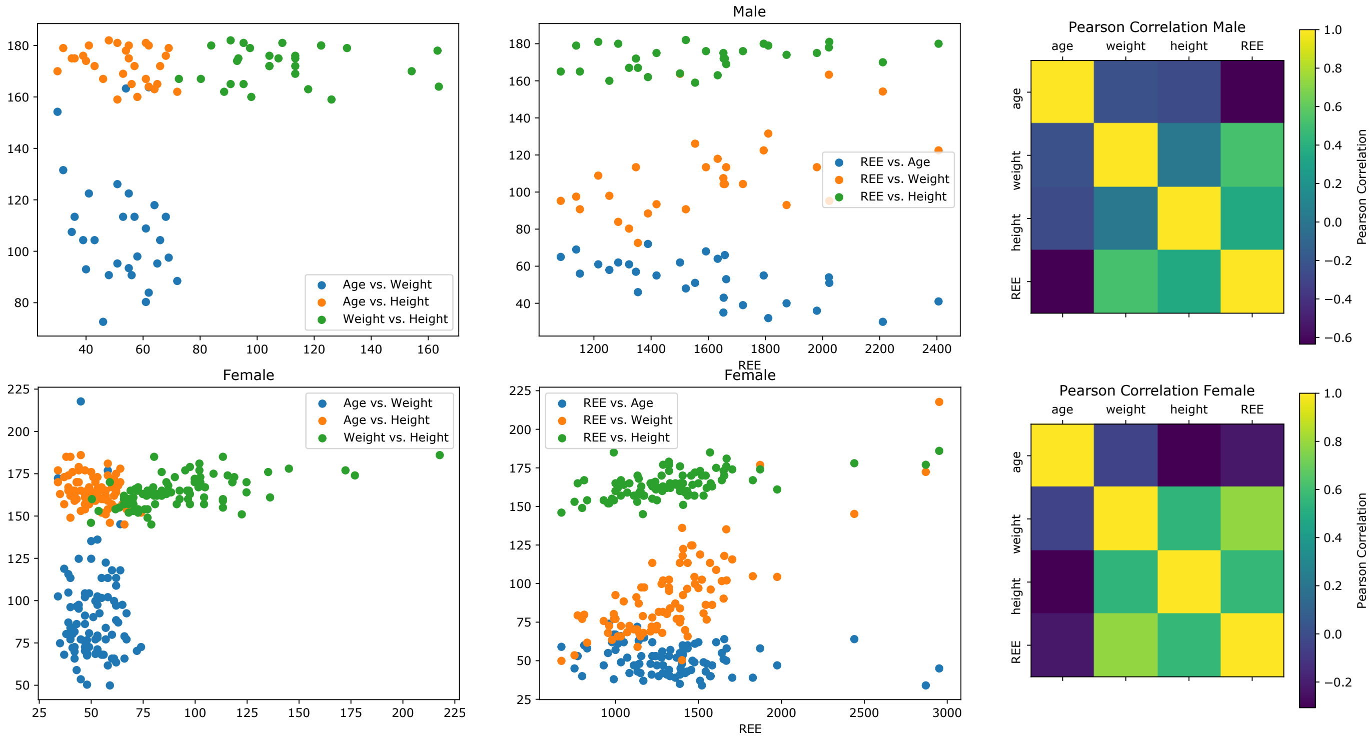

Fig. 2: Visualizations of the linear relationships between the features in the male data (top row) and female data (bottom row). Left plots each feature against each feature and the axes for each point correspond do the legend (eg. Age vs. Weight (blue) has Age on the x-axis and Weight y-axie). Middle plot shows the relationship between each feature and measured REE and y-axis units correspond to the point's legend (eg. REE vs. Age (blue) has age on the y-axis). Right plot presents the correlation matrix comparing each feature against each other feature.

TABLE V: Summary statistics of Bootstrapped results (leave one out) for OLS models and summary statistics for SR models.

Model	Sex	MAE Train	95% CI	MAE LOO	95% CI	Best Train	Best LOO
Ordinary Least Squares	Male	169.00	166.91 – 171.10	197.01	146.35 – 247.66	156.82	33.99
	Female	166.97	166.63 – 167.32	176.57	145.90 – 207.25	159.94	1.10
	All	170.26	170.03 – 170.49	177.16	150.82 – 203.52	165.06	1.42
Symbolic Regression	Male	168.67	164.93 – 172.40			125.09	
	Female	167.03	165.65 – 168.42			148.89	
	All	172.76	171.43 – 174.09			160.95	

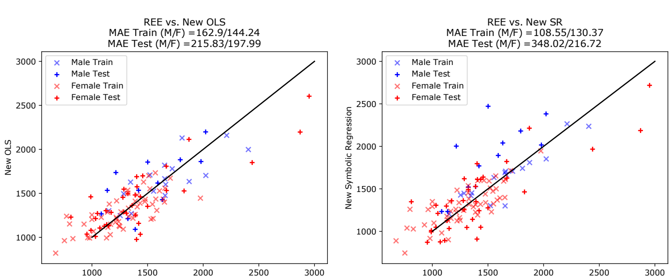

Fig. 3: Scatter plots showing the newly generated OLS model predicted REE (left) and SR nonlinear model predicted REE (right) compared to measured REE of post bariatric surgery patients. Models were fit to training data and applied to training data and testing data. Data is separated into male and female for the right plot.

nonlinear models generated that were able to obtain better errors on the testing data, and perhaps these models could have been found with a more rigorous model selection strategy for the nonlinear models.

Applying all 100 nonlinear models generated to the testing data, the errors obtained were 310.69 with a 95% confidence interval of 289.25 – 332.13 for the male data and 262.56 with a 95% confidence interval of 241.236 – 283.88 for the female data. Models fit to all data, regardless of sex (*All* row within Tables V and VII), are discussed in Section IV-D.

D. *Combined*

Train, validation, and test results are presented in Tables V and VII for models fit to all the data, regardless of the

TABLE VI: Mean Absolute Error (MAE) values obtained from the two newly generated models for resting energy expenditure models when applied to post bariatric surgery data.

Model	Male		Female		Combined	
	Train MAE	Test MAE	Train MAE	Test MAE	Train MAE	Test MAE
Ordinary Least Squares	162.90	215.83	144.24	197.99	148.701	202.45
Symbolic Regression	108.55	348.02	130.37	216.72	125.15	249.55

TABLE VII: Summary statistics for training, validation, and tested results for the OLS models and training and testing summary statistics for the SR models.

Model	Sex	MAE Train	95% CI	MAE Val.	95% CI	Best Train Error	Test Error	Actual Min
Ordinary Least Squares	Male	161.77	156.63 – 166.92	211.29	133.64 – 288.95	141.02	216.13	177.02
	Female	144.08	143.47 – 144.68	155.41	122.47 – 188.35	136.77	198.00	196.21
	All	160.94	160.43 – 161.45	173.28	138.49 – 208.07	153.33	196.42	192.51
	All M. Feat.	173.29	172.73 – 173.85	200.61	159.36 – 241.88	163.62	166.94	159.13
Symbolic Regression	Male	165.26	161.17 – 169.36			108.55	348.02	147.39
	Female	142.34	140.97 – 143.71			130.37	216.72	183.03
	All	158.95	157.35 – 160.54			140.64	213.25	179.87
	All M. Feat.	176.48	174.73 – 178.23			152.96	264.72	159.66

patients' sex (row *All*)[1]. Although the literature has different models for male and female subjects [1], this experiment was done to test if it is necessary to have separate models for post bariatric surgery male and female patients. Note that although there were 60/40 train/test for both males and females in the combined data, the results will be biased towards female patients as there is more data available for them. This difference is representative of the population, as typically 80% of individuals who undergo bariatric surgery are female [2].

In Table VII, which contains the properly trained, validated, and tested results, one can see that these models fit to all data obtained the best testing error values by a small amount.

The linear Equation 9 is obtained when performing no validation, and it obtains a training MAE of 161.06 and a testing MAE of 196.44. An interesting observation here is the difference in the parameters and intercept in this single equation when compared to the separate male and female train test Equations 7 and 8. The weight parameter is not too dissimilar, however the age becomes more negative, height becomes smaller, and the intercept is positive.

The top performing nonlinear model on the training data (not shown) achieved a 140.64 training MAE and 213.25 for testing. If one were to take all 100 nonlinear models generated and apply them to the testing data, an average MAE of 214.85 with a 95% confidence interval of 204.93 – 224.77 is obtained.

$$f(A, W, H) = -10.117 * A + 9.117 * W + 4.259 * H + 345.364 \quad (9)$$

Figure 4 presents the linear and nonlinear models' predicted REE values against the measured REE values.

Although these models did obtain the best results, it is expected that this is a consequence of the models being fit to more data.

[1]Do note that the data in Tables IV and VI under the *Combined* column is different; data under this column is overall average MAE from the male and female patients combined.

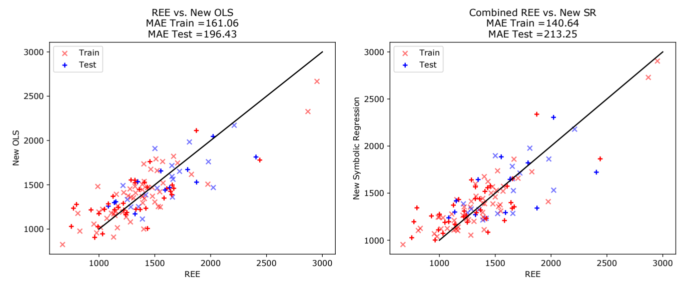

Fig. 4: Scatter plots showing the newly generated OLS model predicted REE (left) and SR nonlinear model predicted REE (right) compared to measured REE of post bariatric surgery patients. In this experiment, male and female data were combined and models were fit to training data and applied to training data and testing data. Data is separated into male and female for the right plot.

E. Combined With More Features

The reason only age, weight and height were selected was because these are the features used by the existing models. Given the success of the models fit to all subjects, and since more features were included with our data, models fit to all subjects with more features were created. All features included in these models were *sex*, *respiration quotient*, *age*, *height*, *years since surgery*, *before surgery weight*, and *current weight*.

The linear Equation 10 contains the model obtained when fit to all training data with no validation. Refer to Table II for the abbreviations of variables. This model results in a MAE of 173.38 for the training data and 166.93 on the testing data.

$$\begin{aligned} f(S, RQ, Y, A, H, W(B), W(C)) = \\ -108.0 * S - 234.193 * RQ - 5.023 * Y \\ -9.807 * A + 6.327 * H - 0.733 * W(B) \\ +9.032 * W(C) + 428.636 \end{aligned} \quad (10)$$

The top performing nonlinear model on the training data (not shown here) obtained a MAE of 152.96 on the training

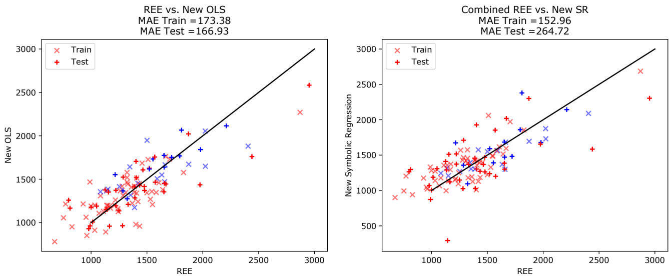

Fig. 5: Scatter plots showing the newly generated OLS model predicted REE (left) and SR nonlinear model predicted REE (right) compared to measured REE of post bariatric surgery patients. In this experiment, male and female data were combined, more features were included, and models were fit to training data and applied to training data and testing data.

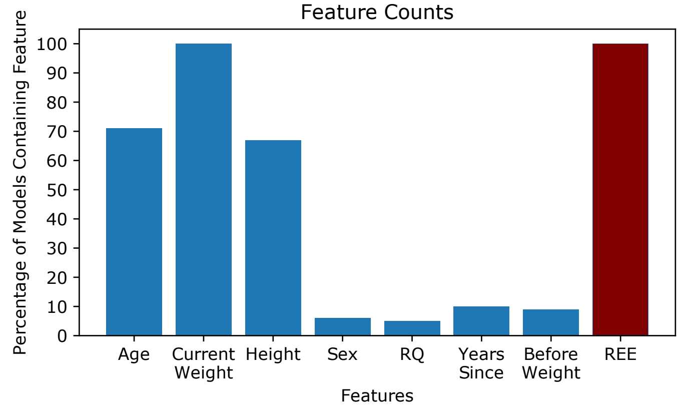

Fig. 6: Percentage of times each feature appeared at least once within each of the 100 nonlinear models generated with SR. Note that REE was required to be within the models as it was the dependent variable.

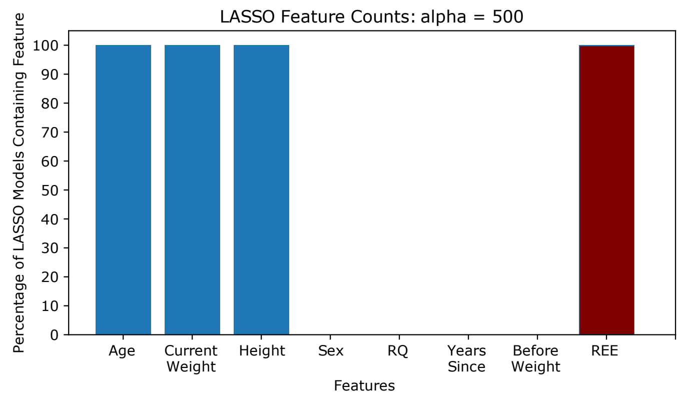

Fig. 7: Percentage of times each feature appeared within each of the LASSO models generated when alpha was set to 500. Note that REE was required to be within the models as it was the dependent variable.

data and 256.72 on the testing data. The average MAE of the 100 nonlinear models on the testing data is 200.98 with a 95% confidence interval of 195.69 – 206.27.

Figure 5 presents the models' predicted REE against the actual measured REE.

Table VII presents the training, validation, and testing results in rows *All M. Feat*. These validated linear models achieved the best testing results of all models generated. The nonlinear models performed poorly, which is likely a result of overfitting the few data points with many features.

1) Feature Analysis: Given the success of the models with more features, an investigation into the new features was performed.

When comparing the features' Pearson correlation, all values are low and suggest no real linear relationship, except for weight before surgery and current weight, which had a Pearson correlation of 0.74.

Figure 6 presents the percentage of the total nonlinear models generated that contained a given feature. Given that SR performs feature selection throughout the evolutionary search, these percentages may provide a *proxy* for how important a given feature is in explaining the underlying system. When investigating these percentages, it is overwhelmingly clear that age, current weight, and height appear much more often than the newly included features, which were almost nonexistent. This observation aligns with the previous models which only use age, current weight, and height [1].

Although the linear models including the new features performed well, based on these results from the nonlinear models, further investigation into the feature importance in the linear models was performed. LASSO regression was performed as it attempts to only select a subset of features to include in the final model.

Figure 7 presents the percentage of the total linear models generated with LASSO regression when run with an alpha[2] value of 500. With an alpha of 1, respiration quotient is the only feature that is removed, but as one increases the alpha value, all of the new features are eliminated and only the original age, weight, and height are included. This corresponds to the observations made with the nonlinear models and the current literature. If one were to increase the alpha value further, only the current weight remains, which also aligns with the nonlinear models' observation of weight being the most commonly used feature.

Although these LASSO results do not necessarily indicate that the new features are not important, it is interesting that they align well with the existing models only containing age, weight, and height.

V. Conclusions and Future Work

Given the need for new models of REE for post bariatric surgery patients, a collection of new linear and nonlinear models were generated. Although no rigorous comparison of the new models to the existing ones was performed, it was clear that the new models obtained much smaller error values

[2]The alpha value is a LASSO parameter one can set. By increasing the alpha value, the regression will become more aggressive in reducing the number of independent variables in the final model.

for the post bariatric surgery patients. A deeper comparison of the new models to the existing ones is being prepared for future work.

The linear models generated outperformed the nonlinear models on testing data, although the nonlinear models fit the training data better. This is likely a consequence of overfitting, the naïve model selection strategy for the nonlinear models, and the fact that the relationships between the features appeared to be linear.

A single model for both male and female subjects was developed to investigate if the post surgery patients require separate models for the sexes. These models obtained high-quality results; however, the authors are cautious to make a case for a single model at this stage as there were very few data points and the models fit to the data from both sexes had an unfair advantage as they were fit to more data.

Additional features were incorporated into the models, namely, sex, respiration quotient, years since surgery, and weight before surgery. This was done in an attempt to develop higher quality models and to investigate if the new features were meaningful for post surgery patients. Ultimately the models developed with the additional features obtained the best results. The authors are again cautious to make a case for the inclusion of these features as there is a current lack of data points and the proxy measures of feature importance suggested that the original age, weight, and height features were the most important.

Although many new models were generated and they all performed well, the best performing validated and tested models are presented in Equations 11 and 12 for male and female patients respectively.

$$f(A, W, H) = -8.287 * A + 8.537 * W + 20.187 * H - 2341.120 \tag{11}$$

$$f(A, W, H) = -4.128 * A + 8.471 * W + 7.584 * H - 465.705 \tag{12}$$

The authors aim to develop more accurate models going forward. The presented results are an improvement over existing models (Harris-Benedict (Equation 1 and 2) & Mifflin-St. Jeor (Equation 3 and 4)); however, we are confident that our models can be improved with additional data. Participants of this study were at least one year post-surgery. New data should include data from both pre- and post-bariatric surgery patients. This would allow for an investigation into the effectiveness of REE models during the initial weight loss period following surgery where physiologic adaptations in REE may be occurring.

A deeper comparison between the newly generated and existing models must be done and a more robust statistical analysis will be performed in a future investigation. Further, a comparison between the model parameters may provide insight into the underlying system. For example, how does age affect REE in healthy individuals versus individuals pre- and post-surgery?

This analysis has established that population-specific models for REE can out-perform traditionally used models, and that linear modelling techniques using relatively easy to obtain measures/features can be sufficient for explaining this physiologic phenomenon (REE).

These newly developed REE models require input from members of a bariatric care team, further analysis of clinically relevant features (i.e. body composition), and longitudinal clinical testing before their impact can be fully understood.

VI. Acknowledgments

This research was supported in part by the Natural Sciences and Engineering Research Council of Canada (NSERC).

The authors would like to thank Dr. Tamara E. Carver, Kathleen M. Andersen, and Dr. Nicolas V. Christou for their contributions to recruitment and data collection. The authors also thank the participants for their time and contributions to the advancement of care for everyone who undergoes bariatric surgery.

References

[1] Bahareh Amirkalali, Saeed Hosseini, Ramin Heshmat, Bagher Larijani, et al. Comparison of harris benedict and mifflin-st jeor equations with indirect calorimetry in evaluating resting energy expenditure. *Indian journal of medical sciences*, 62(7):283–290, 2008.

[2] Henry Buchwald, Yoav Avidor, Eugene Braunwald, Michael D Jensen, Walter Pories, Kyle Fahrbach, and Karen Schoelles. Bariatric surgery: a systematic review and meta-analysis. *Jama*, 292(14):1724–1737, 2004.

[3] Anita P Courcoulas, Nicholas J Christian, Steven H Belle, Paul D Berk, David R Flum, Luis Garcia, Mary Horlick, Melissa A Kalarchian, Wendy C King, James E Mitchell, et al. Weight change and health outcomes at 3 years after bariatric surgery among individuals with severe obesity. *Jama*, 310(22):2416–2425, 2013.

[4] J Arthur Harris and Francis G Benedict. A biometric study of human basal metabolism. *Proceedings of the National Academy of Sciences of the United States of America*, 4(12):370, 1918.

[5] J A Hughes. jGP. https://github.com/jameshughes89/jGP, March 2015. Accessed: February 4, 2020.

[6] James Hughes, Sheridan Houghten, and Joseph Alexander Brown. Models of parkinson's disease patient gait. *IEEE Journal of Biomedical and Health Informatics*, 2019.

[7] James Alexander Hughes and Mark Daley. Generating nonlinear models of functional connectivity from functional magnetic resonance imaging data with genetic programming. In *2019 IEEE Congress on Evolutionary Computation (CEC)*, pages 3252–3261. IEEE, 2019.

[8] James Alexander Hughes, Ethan C Jackson, and Mark Daley. Modelling intracranial pressure with noninvasive physiological measures. In *2017 IEEE Conference on Computational Intelligence in Bioinformatics and Computational Biology (CIBCB)*, pages 1–8. IEEE, 2017.

[9] John R Koza and John R Koza. *Genetic programming: on the programming of computers by means of natural selection*, volume 1. MIT press, 1992.

[10] Mark D Mifflin, Sachiko T St Jeor, Lisa A Hill, Barbara J Scott, Sandra A Daugherty, and Young O Koh. A new predictive equation for resting energy expenditure in healthy individuals. *The American journal of clinical nutrition*, 51(2):241–247, 1990.

[11] Athena H Robinson, Sarah Adler, Helen B Stevens, Alison M Darcy, John M Morton, and Debra L Safer. What variables are associated with successful weight loss outcomes for bariatric surgery after 1 year? *Surgery for Obesity and Related Diseases*, 10(4):697–704, 2014.

[12] M Schmidt and H Lipson. Comparison of tree and graph encodings as function of problem complexity. In *Proc. 9th annual conf. on Genetic and evolutionary computation*, pages 1674–1679. ACM, 2007.

[13] M D Schmidt and H Lipson. Coevolution of fitness predictors. *IEEE Transactions on Evolutionary Computation*, 12(6):736–749, 2008.

[14] Michael Schmidt and Hod Lipson. Distilling free-form natural laws from experimental data. *science*, 324(5923):81–85, 2009.

[15] Britta Wilms, Barbara Ernst, Martin Thurnheer, Sebastian M Schmid, Christina M Spengler, and Bernd Schultes. Resting energy expenditure after roux-en y gastric bypass surgery. *Surgery for Obesity and Related Diseases*, 14(2):191–199, 2018.