A Novel Machine Learning Based Prediction Model for Energy Expenditure in relation to Varying Load, Incline, and Body Composition

Katherine Swanson k8@cs.stanford.edu
Aubrey Gress adgress@cs.stanford.edu
Vlad Andrei vlad.andrei@cs.stanford.edu

Abstract

In this paper we use machine learning techniques to develop a model for the prediction of the energetic cost of walking on a positive incline while carrying a load given the subjects body composition. Previous studies on prediction of energy cost have failed to incorporate load, incline and body composition simultaneously in their predictive models. By leveraging a more robust set of features including mass carried, incline and body composition we provide a more accurate model for predicting energy expenditure and demonstrate novel relationships between body composition and energy expenditure.

I. INTRODUCTION

The standard method of reporting the energetic cost of performing a task is to report the energetic cost in watts, normalized to body mass. Differences in resting energy and energy expenditure have been examined in conjunction with differences in body composition in men and women. There is a relationship between resting energy expenditure and fat free mass and the fat mass [1]. During walking and running, equations have been proposed relating the factors of mass, incline, and speed [2]. However, these studies have yet to investigate any potential relationship between body composition and locomotor efficiency. Thus, the goal of our project is to perform a comprehensive assessment between the variables of load carriage, incline and body composition.

The ultimate goal of our discoveries is to allow optimal distribution of loads among subjects carrying gear and other equipment. We are working with the Human Performance Lab at Stanford under Dr. Thor Bezier in indirectly supplying this data to the Department of Defense.

The energetic cost of an action is linearly related to VO2 (oxygen consumed per unit time – milliliters per minute within this paper).

$$\dot{E}_{\mathrm{metab,gross}} = 16.58\,\frac{\mathrm{W{\cdot}s}}{\mathrm{ml\ O_2}}\,\dot{\overline{V}}_{O_2} + 4.51\,\frac{\mathrm{W{\cdot}s}}{\mathrm{ml\ CO_2}}\,\dot{\overline{V}}_{CO_2} \quad (1)$$

VO2 can be measured by using a breathing apparatus which measures the amount of oxygen being consumed during a task. The model we present in this paper outputs predictions on VO2 directly.

Our research has uncovered two important results. First, we show that body composition is an important factor in predicting VO2 for a given load and incline and we present a small set of biometric features that can be used to predict VO2. Second, we demonstrate that SVMR with a quadratic kernel can accurately predict VO2 given body composition, incline and load.

II. PREVIOUS WORK

The Pandolf Equation (1977) is the most current model relating energetic cost to load and incline and speed [2].

$$\mathrm{EEp} = 4.5M + 6(M+B)(B/M)^2 + \mu(M+B) \times (4.5V^2 + 1.1VG)$$

This model proposes a quadratic relationship between energetic cost and speed, and a linear relationship between energetic cost and incline. While this is the current standard used to predict energy expenditure in research and in the field, the model is insufficient. The data used to build this model did not normalize subject VO2 output to body weight [3]. More importantly, the Pandolf model does not attempt to incorporate data on body composition [3] which we show in this paper is necessary for an accurate model. We have attempted to accurately predict VO2 using the Pandolf model with collected data but have been unable to do so. The Pandolf model performs very poorly on the given data with greater than 50% relative error.

III. PREPROCESSING

The training set for our model consists of nine subjects, both male and female, in good physical shape within the age range of 22 to 44. The data we used was provided by the Human Performance Lab at Stanford. Each subject requires two approximately three hour tests separated by a week interval over which weight and cardiovascular fitness level is required to remain constant. [1]

For each subject we captured body composition data using a Dual-energy X-ray absorptiometry (DEXA) scanner. Data included measurements of bone mineral content (hereafter BMC) (lbs), fat to muscle ratio (FMR) (lbs), lean muscle mass (lbs), and total tissue weight (lbs) for each of five discrete areas of the body: legs, arms, trunk, gynoid region (visceral fat in the chest) and android region (subcutaneous fat located around the hips). The body composition data represents a total of 60 features per subject.

We then collected data for each subject for each of twelve conditions varying load and incline. Each participant walked on a treadmill carrying loads of 0%, 10%, 20% and 30% total body mass and at inclines of 0%, 5% and 10% for 5 minutes each. VO2 was measured during each of the 12 intervals by collecting a datapoint every time the subject exhaled. The order in which load was tested across subjects was randomized to minimize artifacts.

There is a large amount of noise in VO2 values captured during incline changes. VO2 becomes more stable within approximately one minute intervals around incline changes, therefore the first and last minute of VO2 data for each incline/load test is discarded. The remaining data is averaged to give that participants VO2 value for a given load and incline.

Figure 1: Data Capture for varying load and incline.

IV. PREDICTIVE METHODS

We hypothesized that including biometric data in our model will improve the accuracy predicting VO2 significantly. Furthermore, we are working in a very high feature space (62 features per subject) relative to the small number of training examples. Features are not necessarily independent – for example, the feature set includes measurements of arm FMR as well as total limb FMR. Due to the small training set and large amount of features per subject, we hypothesized that models will overfit the data unless we reduce the feature set. Reducing the feature set will also allow us to examine relevant relationships with VO2 output without distributing weight across non-independent features. Finally, we must normalize features so that computed weights have significance for analysis.

[1] While this training set it small, the previous models discussed in this paper used training sets on the same order of magnitude. We are continuing to collect training data to improve our model and predict approx. We expect 20 more subjects within the next three months.

Linear Regression/SVMR with Linear Kernel

We initially model the relationship between energy expenditure, load, incline, and biometric data using Linear Regression and SVMR with a linear kernel. Error was calculated using leave-on-out-cross-validation (LOOCV). Both models performes reasonably well with RMS error of 117.5 and 84.4 ml/second and root relative squared test errors of 26.23% and 27.2%.

Body Composition

We used Linear Regression to examine whether inclusion of body composition information significantly impacts prediction of VO2. These preliminary results show that models trained on features including body composition consistently outperform models which include only load and incline information (see Figures 2 and 3). This result indicates that body composition plays a significant role in determining VO2.

Training Error	LR ALL	LR LI	SVMR LK ALL	SVMR QK ALL	SVMR QK LI	SVMR QK R
RMS	99.17	245.88	101.24	64.92	250.55	64.96
Relative	24.30%	60.2%	24.81%	15.91%	61.40%	15.92%

Figure 3: Training Error Comparison

"LR" = Linear Regression
"LK" = Linear Kernel
"QK" = Quadratic
"ALL" = the full feature set was used
"LI" = Only load and incline were used
"R" = Reduced feature set from backwards search used

SVMR with Quadratic Kernel

Biometric research has indicated that biometric measurements including FMR have a statistically relevant quadratic relationship with energy expenditure [6]. As our preliminary results indicate that biometric data is relevant to the model, we hypothesized that a quadratic kernel may be a better fit when including these additional features. Running SVMR with a quadratic kernel improved RMS error computed with LOOCV by 36.6 ml/sec (Figures 2, 3 and 4). Experimentation with other non-linear models did not produce better results than those given by the quadratic kernel.

Feature Reduction

We have thus far shown an improved model for VO2 prediction by incorporating body composition data; however, as explained above, using this improved model on a 62 dimensional feature set is prone to overfitting. Reducing the set of features should therefore decrease error computed by LOOCV. Furthermore, we would like to determine specifically which features of the body are statistically relevant to the model. In the non-reduced model, weights computed for any relevant set of features are not observable as they are distributed across many dependent biometric features. Therefore, load and incline dominate the feature set in non-reduced feature space.

Test Error	LR ALL	LR LI	SVMR LK ALL	SVMR QK ALL	SVMR QK LI	SVMR QK R
RMS	117.5	252.99	120.6	84.4	264.86	79.0
Relative	26.23%	61.42%	27.2%	18.5%	64.31%	16.7%

Figure 2: Test Error Comparison

We computed a reduced set of features using Backwards Search. Figure 5 shows these relative weights of the reduced feature. [2] As hypothesized, using SVMR with a quadratic kernel on the reduced feature set improves RMS error calculated by LOOCV by 5.5 and relative error reduces from 18.5% to 16.7%.

In the reduced feature set we can now examine the relative weights of relevant features (Figure 5). A preliminary look at the weights indicates that a high BMC in the arms is highly related to energy expenditure. In the reduced feature set arm BMC has a stronger relationship with VO2 than either load or incline.

To determine how these results compare to chance we apply Permutation Analysis. The process is as follows: (1) Create a new feature for the training set by

[2] For visualization purposes, weights shows are calculated using SVMR with a linear kernel.

randomly sampling arm BMC with replacement. (2) Replace arm BMC with the new feature and recomputed the weights of the features. We repeat this process to create a Gaussian distribution with mean μ and variance σ. By computing the value of the cumulative distribution function for the weight of arm BMC we obtain p value = 3.0710e-005. The accepted threshold for discarding a null hypothesis is $p < 0.05$. Therefore, we show that with high probability that the significance of arm BMC in predicting VO2 is not a result of chance. This result is supported by Griffin (2003) who showed that density of the arm region has a large impact on energy expenditure while walking.

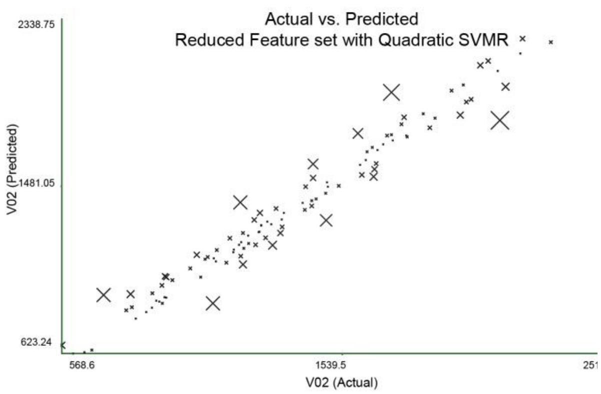

Figure 4: Actual vs Predicted Energy Expendiure

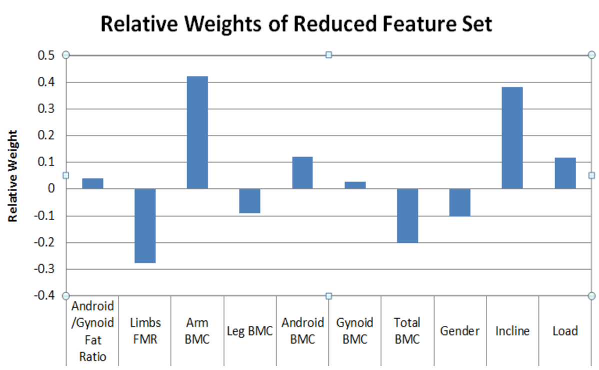

Figure 5: Weights of features selected by backward search

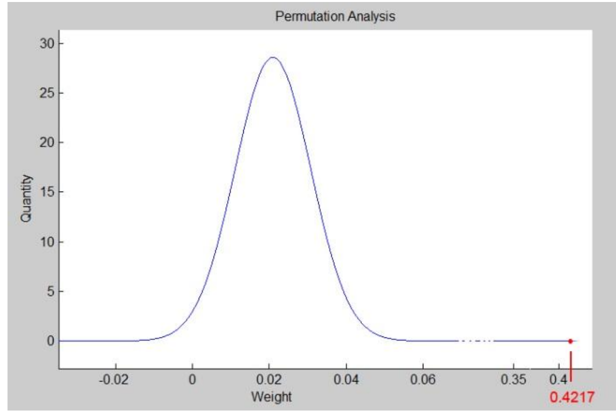

Figure 6: Bootstrapping and Permutation Analysis Distribution

VI. CONCLUSION

By applying Machine Learning to the prediction of energy expenditure we demonstrate two novel results. First, that the relationship of a feature set including body composition features as well as load and incline is best modeled using SVMR with a quadratic kernel. Second, we have shown that arm BMC has a strong relationship with energy expenditure. This result is supported by biometric research [6] which demonstrates that the motion of arms while walking has a strong impact on energy expenditure. It is likely that these two breakthroughs are interrelated; the high relative weights of the reduced features relative to incline and load in our improved model may help explain the need for a quadratic model. Overall, these findings suggest that biometric data is highly relevant in developing predictive models for energy expenditure. Given the small size of our current training set, more exploration should be done to determine precise relationships to allow for better prediction for VO2.

VIII. ACKNOWLEDGEMENTS

This project was undertaken with the guidance of Amy Slider and Thor Bezier of the Stanford Human Performance Lab. They provided us with data as well as given us guidance throughout the project. We look forward to continuing research with them on this project.

BIBLIOGRAPHY

[1] Cunningham, JJ. -“*Body Composition as a determinant of energy expenditure: A synthetic review and a proposed general prediction equation.”* (1991).

[2] Pandolf KB et al. - *Predicting energy expenditure with loads while standing or walking very slowly”* (1977).

[3] Goldman – *“Energy Cost of Load Carriage”* (1962).

[4] Kramer – “*The Effect on Energy Expenditure of Walking on Gradients or Carrying Burdens” (*2010)

[5] Weir, J.B. de V – “*New methods for calculating metabolic rate with special reference to protein metabolism”* (1949).

[6] Griffin – *“Metabolic cost of generating muscular force in human walking: insights from load-carrying and speed experiments”*(2003).