USING PARTICIPANT DESCRIPTIVE CHARACTERISTICS, HEART RATE AND INERTIAL SENSORS TO ESTIMATE RUNNING ECONOMY

Krista Vohlakari[1], Sailee Sansgiri[1], Esa Hynynen[2], Timo Rantalainen[1], Neil Cronin[1,3]

[1] Neuromuscular Research Centre, Faculty of Sport and Health Sciences, University of Jyväskylä, Jyväskylä, Finland
[2] Finnish Institute of High Performance Sport KIHU, Jyväskylä, Finland
[3] School of Sport and Exercise, University of Gloucesterhire, UK

In this study, we aimed to develop a machine learning algorithm to estimate running economy. Inertial sensor, heart rate and participant descriptive characteristics were used to estimate running economy by training and comparing five different machine learning models. Fourteen subjects performed a VO_2max running test on a treadmill with four-minute stages. Submaximal running speeds (lactate level < 4 mmol/l) were used to train the models. The best-performing model was a k-nearest neighbour regressor, which achieved average root mean square error of 0.097 ± 0.059 kcal/kg/km and average mean absolute percentage error of 8.7 ± 5.8 % compared to ground truth running economy data. Despite reasonably accurate running economy estimates, the model is currently not very generalisable, probably due to the small dataset used for training.

KEYWORDS: wearable devices, acceleration, angular velocity, machine learning.

INTRODUCTION: Running economy (RE) is one of the most important determinants of running performance (Moore, 2016). RE describes the energy cost at steady-state running velocities and can be measured via oxygen consumption (VO_2) or energy expenditure (EE) (Van Hooren et al., 2024). Despite the importance of RE in running, there is limited understanding of how it is affected by spatiotemporal, kinematic, kinetic, and neuromuscular factors (Moore, 2016). Daniels (1985) noted that age, sex, air resistance, body temperature, stride length, weight, and running experience all contribute to RE. According to a new meta-analysis, running biomechanics can explain 4–12% of the between-individual variation in RE (Van Hooren et al., 2024). However, it is not possible to identify a single most economical running pattern that applies to all individuals (Patoz et al., 2022).
In recent years, novel technologies such as wearable sensors have been used to accurately predict several RE-related biomechanical factors (e.g., Ahamed et al., 2019). A general rule for prediction tasks is to start with a linear model and if it cannot accurately fit the data, to use a non-linear model. Non-linear models may be more suitable for large datasets with a large number of data points (observations) and few features (input variables) (Strang et al., 2018).
In this work, the aim was to estimate RE using a combination of variables derived from wearable inertial sensors, heart rate sensor and participant descriptive characteristics. Since we had a small dataset, we tested four linear supervised models (linear regression, stochastic gradient descent regressor, support vector regression, k-nearest neighbour regressor (KNR)) and a neural network (Chen & Billings, 1992) to find the most optimal linear model for RE estimation, and to determine whether a more complex non-linear model was needed.

METHODS: Participants with different sport backgrounds and running experience (including casual runners) volunteered for this study (62 overall). Here we used data from 14 randomly-selected participants (22–44-years-old; 7 males) to test five machine learning models. All participants were tested in laboratory conditions. Inclusion criteria were being healthy and no injuries or chronic disorders that might affect running technique within 6 months of the measurements. The weekly number of kilometres run by participants varied between 0–70 km (18.1 ± 22.2). Maximal oxygen consumption (VO_2max) was 25.8–52.5 ml/kg/min (39.9 ± 6.8). Participants classified themselves as novice (4), enthusiast (6), or goal-oriented enthusiast runners (4). Nobody described themselves as a competitive runner.

Participants visited the laboratory twice: familiarisation session and VO_2max test. The familiarisation session included running 3x4min with the same equipment as in the VO_2max test. In the VO_2max session, body mass and height were measured, and participants sat quietly for two minutes while resting heart rate (HR) was measured (Polar H10; Polar Electro Oy, Kempele, Finland). Five inertial sensors (Xsens Dot, Movella, CA, USA) were attached to the sacrum, both shanks and both shoes. Sensor data (3-axis acceleration and angular velocity) were sampled at 120 Hz via a Bluetooth connected Android mobile phone. A gas analysis mask was placed over the participant's face to measure breath-by-breath VO_2 and production of carbon dioxide (VCO_2) (JAEGER Vyntus CPX; Vyaire Medical Inc., Illinois, USA). Next, a five-minute warm-up was performed at self-selected speed. The VO_2max test was performed with 4-minute stages at 1 % gradient and incremental speeds (+1 km/h after each stage). After each stage the treadmill (OJK-1, Telineyhtymä Kotka, Finland) was stopped for 20–40 sec, and a fingertip blood sample was taken (Biosen C-Line -lactate analyser; EKF Diagnostic, Madgeburg, Germany). The test was continued until the participant chose to stop or couldn't maintain the speed. Participants estimated their weekly running kilometres by answering the following question: "On average, how many kilometres do you run in a week?"
RE was estimated only for submaximal speeds between the aerobic and anaerobic thresholds of each participant (Daniels, 1985; Van Hooren et al., 2024). Thresholds were determined based on the lactate values measured during the VO_2max test. Aerobic threshold was set at 0.3 mmol/l above the lowest lactate value during the test (Vesterinen et al., 2016). For anaerobic threshold (Faude et al., 2009), 4 mmol/l was used for all participants. Across all participants, aerobic threshold was 8.9 ± 1.7 km/h and anaerobic threshold was 11.4 ± 1.7 km/h. Thus, speeds in the range 7–14 km/h were used. Each speed was normalised to the speed of anaerobic threshold (Fletcher et al., 2009). In accordance with the findings of Daniels (1985) model inputs included speed, sex, age, body mass, height, and weekly running kilometres. Resting heart rate (HR) and HR were also used as input variables indicative of training status (Buchheit, 2014). To incorporate technique information, we also included cadence (Van Hooren et al., 2024) and peak, minimum and standard deviation of 3-axis acceleration and 3-axis angular velocity signals of all inertial sensors. The predicted target variable was RE expressed as kcal/kg/km.
To calculate RE, the gas exchange data were first averaged from the last minute of each running stage (this method was also used to average HR) (Robergs et al., 2010). Then EE (kcal/min) was estimated as: VO_2 (l/min)*(1.2064*RER+3.8455), where RER=VCO_2/VO_2 (McArdle et al., 2001, pp. 187-200). Finally, RE was calculated as EE*60/body mass/speed. Resting HR was calculated by finding the lowest 15 second average during sitting. Acceleration and angular velocity signals (one minute from the end of each stage after deleting the last 10 seconds to avoid errors caused by participants preparing to stop) were filtered using a fourth order low-pass Butterworth filter with 16 Hz cut-off frequency. Data were divided into individual stride cycles based on right foot initial contact events (IC). The ICs were identified by finding peaks in the resultant acceleration using the following rules: a minimum resultant acceleration of 50 m/s2 and a minimum duration of 500 ms between estimated consecutive ICs (Donahue & Hahn, 2022). The peak, minimum and standard deviation values for each signal and each sensor were calculated from the average stride cycle signals. Right and left foot IC's were used to determine cadence (strides/min): number of IC's-1/duration between first and last IC in signal. The average value of right and left foot cadence is reported.
Python3 with sklearn-library (Pedregosa et al., 2011) was used to build linear models and Keras 2.7 deep learning API (Chollet et al. 2015) was used to build a multiple layer neural network to estimate RE. Data included 43 observations from 14 participants (2–4 observations each, speeds between aerobic and anaerobic thresholds). Data were divided into train and test sets using leave-one-subject-out (14 rounds, 14 trained models), so that each participant's data only featured in the train or the test set at one time. Input data were normalised between 0 and 1 using the MinMaxScaler from Python's sklearn-library (Pedregosa et al., 2011). After manually testing the different hyperparameters, the default options were used for linear models. For the neural network, the following parameters were used: optimizer=adam, loss=mae, metrics=mae, layers=3, batch sizes=512,512,1, activation=relu. Root mean square

error (RMSE) was used to indicate the absolute error of each model and mean absolute percentage error (MAPE) was used to indicate relative error.

RESULTS: Across all models, average RMSE and MAPE varied between 0.097–0.203 kcal/kg/km and 8.7–17.0 % respectively. The lowest RMSE and MAPE were found for KNR, with an average RMSE of all leave-one-subject-out submodels (Table 1) of 0.097 ± 0.059 kcal/kg/km, and average MAPE of 8.7 ± 5.8 %.

Table 1: The accuracy of the separate KNR submodels. For each submodel, 13 participants' data were used for training and one participant's data for testing.

Submodel	RMSE (kcal/kg/km)	MAPE (%)
1	0.054 ± 0.004	4.7 ± 0.3
2	0.100 ± 0.030	8.6 ± 2.7
3	0.114 ± 0.040	10.3 ± 4
4	0.148 ± 0.062	14.0 ± 6.6
5	0.201 ± 0.032	15.6 ± 2.3
6	0.094 ± 0.010	8.9 ± 0.8
7	0.029 ± 0.017	2.4 ± 1.7
8	0.047 ± 0.028	3.9 ± 2.9
9	0.161 ± 0.039	12.9 ± 3.2
10	0.055 ± 0.021	4.7 ± 2.0
11	0.062 ± 0.002	6.0 ± 0.1
12	0.049 ± 0.009	4.6 ± 0.9
13	0.173 ± 0.041	19.2 ± 5.6
14	0.063 ± 0.008	6.0 ± 0.8

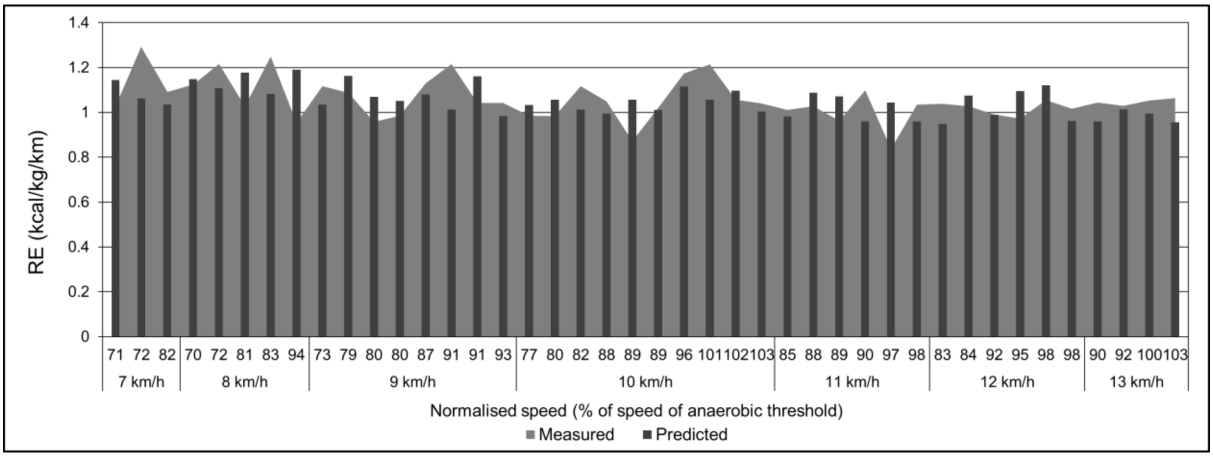

Figure 1: Measured and estimated RE (with the KNR model) for all participants and speeds.

DISCUSSION: KNR was the most accurate of the tested supervised-learning models. However, the RMSE and MAPE showed large variation between the different submodels, implying that the model cannot yet be generalised across runners of different levels. Nonetheless, the average RMSE and model behaviour across different speeds (Figure 1) suggest that KNR provides reasonably accurate estimates of RE.

Compared to using only body mass, the use of RER, speed, and body mass to express RE (together with VO_2) results in models that are less sensitive to varying speeds and energy substrate use. In future, one clear improvement would be to subtract resting or standing energy expenditure from the RE values before training the model. During data preprocessing and model training, it was evident that small changes in input data affected the accuracy ranking of different types of algorithms. This may be because of the simplicity of the tested approaches. We also found that alterations made during data preprocessing affected the resulting model accuracy, which may indicate overfitting, likely due to the dataset being too small.

The study includes some limitations. Firstly, since the default options for hyperparameters were used, with optimum hyperparameter tuning, the best performing model could change.

Predictive models need information about the running speed at anaerobic threshold, which we determined at a lactate level of 4 mmol/l. When using 4-minute running stages, it is possible that individual maximal steady state lactate levels are higher than 4 mmol/l (Faude et al., 2009). In addition, we did not verify whether steady state HR, VO_2 and lactate levels were reached at all stages that were used to train and test the models. Finally, gas analysis devices inherently include some measurement error. These factors may all have contributed to errors in model predictions.

CONCLUSION: Considering the small dataset used here, our results suggest that KNR is a promising approach for predicting RE when combined with physiological and movement data. Although model accuracy was very sensitive to changes in input features, the addition of more data may lead to further improvements. From a practical perspective, the developed algorithm could potentially be a tool for runners and coaches to estimate RE at constant speeds, allowing training to be monitored without expensive equipment.

REFERENCES

Ahamed, N. U., Benson, L. C., Clermont, C. A., Pohl, A. J., & Ferber, R. (2019). New considerations for collecting biomechanical data using wearable sensors: How does inclination influence the number of runs needed to determine a stable running gait pattern? *Sensors*. 19 (11), E2516, https://doi.org/10.3390/s19112516

Buchheit, M. (2014). Monitoring training status with HR measures: do all roads lead to Rome? *Frontiers in physiology*. 5 (73), https://doi.org/10.3389/fphys.2014.00073

Chen, S. & Billings, S. A. (1992). Neural networks for nonlinear dynamic system modelling and identification. *International Journal of Control*. 56 (2), 319–346, https://doi.org/10.1080/00207179208934317

Chollet, F. et al. (2015). Keras. GitHub. Retrieved from https://github.com/fchollet/keras

Daniels, J. T. (1985). A physiologist's view of running economy. *Medicine and Science in Sports and Exercise*. 17 (3), 332–338, https://doi.org/10.1249/00005768-198506000-00006

Donahue, S. R. & Hahn, M. E. (2022). Validation of running gait event detection algorithms in a semi-uncontrolled environment. *Sensors*. 22 (9), 3452, https://doi.org/10.3390%2Fs22093452

Faude, O., Kindermann, W. & Meyer, T. (2009). Lactate threshold concepts: how valid are they? *Sports Medicine*. 39 (6), 469–490, https://doi.org/10.2165/00007256-200939060-00003

Fletcher, J. R., Esau, S. P., & Macintosh, B. R. (2009). Economy of running: beyond the measurement of oxygen uptake. *Journal of Applied Physiology*. 107 (6), 1918–1922. https://doi.org/10.1152/japplphysiol.00307.2009

McArdle, W. D., Katch, F. I. & Katch, V. L. (2001). Physiology: nutrition energy and human performance (5th ed). Lippincott Williams & Wilkins: Philadelphia, PA, USA.

Moore, I. S. (2016). Is there an economical running technique? A review of modifiable biomechanical factors affecting running economy. *Sports Medicine*. 46 (6), 793–807, https://doi.org/10.1007/s40279-016-0474-4

Patoz, A., Lussiana, T., Breine, B., Gindre, C., & Hebert-Losier, K. (2022). There is no global running pattern more economic than another at endurance running speeds. *International Journal of Sports Physiology and Performance*. 101276430, 1 – 4, https://doi.org/10.1123/ijspp.2021-0345

Pedregosa, F. et al. (2011). Scikit-learn: Machine learning in Python. *Journal of Machine Learning Research*. 12, 2825–2830, https://doi.org/10.48550/arXiv.1201.0490

Robergs, R. A., Dwyer, D. & Astorino, T. (2010). Recommendations for improved data processing from expired gas analysis indirect calorimetry. *Sports Medicine*. 40 (2), 95–111. https://doi.org/10.2165/11319670-000000000-00000

Strang, B., van der Putten, P., van Rijn, J. N. & Hutter, F. (2018). Don't rule out simple models prematurely: A large scale benchmark comparing linear and non-linear classifiers in OpenML. In: Duivesteijn, W., Siebes, A., Ukkonen, A. (Eds), *Advances in Intelligent Data Analysis XVII, IDA 2018, Lecture Notes in Computer Science*. Vol 11191, 303–315. Springer, Cham. https://doi.org/10.1007/978-3-030-01768-2_25

Van Hooren, B. et al. (2024). The relationship between running biomechanics and running economy: A systematic review and meta-analysis of observational studies. *Sports Medicine.* https://doi.org/10.1007/s40279-024-01997-3

Vesterinen, V., Nummela, A., Heikura, I., Laine, T., Hynynen, E., Botella, J. & Häkkinen, K. (2016). Individual endurance training prescription with heart rate variability. *Medicine & Science in Sports & Exercise*. 48 (7), 1347–1354, https://doi.org/10.1249/MSS.0000000000000910