



Sensitivity of movement features to fatigue during an exhaustive treadmill run

Christos Chalitsios, Thomas Nikodelis, Vasileios Konstantakos & Iraklis Kollias

To cite this article: Christos Chalitsios, Thomas Nikodelis, Vasileios Konstantakos & Iraklis Kollias (2021): Sensitivity of movement features to fatigue during an exhaustive treadmill run, European Journal of Sport Science, DOI: [10.1080/17461391.2021.1955015](https://doi.org/10.1080/17461391.2021.1955015)

To link to this article: <https://doi.org/10.1080/17461391.2021.1955015>



Published online: 25 Jul 2021.



Submit your article to this journal [↗](#)



Article views: 32



View related articles [↗](#)



View Crossmark data [↗](#)



Sensitivity of movement features to fatigue during an exhaustive treadmill run

Christos Chalitsios ^a, Thomas Nikodelis ^a, Vasileios Konstantakos^b and Iraklis Kollias^a

^aBiomechanics Laboratory, Department of Physical Education and Sports Sciences, Aristotle University of Thessaloniki, Thessaloniki, Greece;

^bElectronics Laboratory, Physics Department, Aristotle University of Thessaloniki, Thessaloniki, Greece

ABSTRACT

The objective of the present study was to examine the sensitivity of several movement variables during running to exhaustion adopting a cross-sectional design. Thirteen recreational runners, that systematically trained and competed, performed an exhaustive running protocol on an instrumented treadmill. Respiratory data were collected to establish the ventilatory threshold in order to obtain a reference point regarding the gradual accumulation of fatigue. A machine learning approach was adopted to analyse a set of 29,650 data points (individual steps) of kinetic and kinematic data, using a random forest classifier for the region pre and post the ventilatory threshold. The overall accuracy of the model was 0.914 (95% CI: 0.907–0.919). The four most important variables, and more sensitive in predictive ability, as it was concluded from the variable importance procedure and the partial dependence (PD), were the angular range in AP axis of upper trunk C7, the maximum loading rate, the angular range in LT axis of the C7 and the maximum value of the ground reaction force. Two-dimensional PD revealed considerable interactions for certain areas of the joint distributions between kinetic and kinematic data. These results provide a direction towards understanding the interconnections of kinetics and kinematics of the torso to maintain the coordinated running pattern under fatigue conditions.

Highlights:

- Trunk frontal plane kinematics is the most sensitive parameter to fatigue. Practitioners should consider this finding during endurance training.
- Kinetics exhibit a stable linear increase in mean values but a non-linear increase in variance during an exhaustive incremental treadmill run. This may affect training at a submaximal fatigued state.
- Specific areas in the joint distributions of kinetics and kinematics during treadmill running exhibit increased sensitivity in predicting fatigue state.

KEYWORDS

Endurance; data; biomechanics; methodology

1. Introduction

Resistance to fatigue is a key issue in endurance running for both recreational and elite runners. The term fatigue is often used to define the decline in various objective measures of performance over a discrete period of time (Abbiss, Peiffer, Meeusen, & Skorski, 2015; Enoka & Duchateau, 2016). Running in a fatigued state introduces technique alterations that in the long term may lead to musculoskeletal injury (Clansey, Hanlon, Wallace, & Lake, 2012). Fatigue identification is usually performed by monitoring various physiological indices. Numerous fatigue thresholds have been proposed in the literature such as the ventilatory threshold, the respiratory compensation point, the heart rate deflection point, the critical power physical working capacity at the fatigue threshold and the electromyographic fatigue threshold. All these parameters can estimate

physical exertion, indicating fatigue and non-fatigued work production (Devries et al., 1987; Kumagai et al., 1982). Fatigue detection, along with the study of the key parameters of performance that are affected during endurance running, are of paramount importance for the planning of training.

Duration of effort and intensity are the two main effectors modulating the various metabolic processes that take place during continuous prolonged exercise. Metabolic transitions from aerobic energy production to more anaerobic denotes the start of a state where the accumulation and detection of fatigue are more likely to be observed (Wasserman, Whipp, Koyl, & Beaver, 1973). This is a multifactor phenomenon (Stirling, Tschanner, Fletcher, & Nigg, 2012), not only metabolic or neuromuscular but also mechanical in nature, as the presence of fatigue modifies running biomechanics (Derrick, Dereu, & McLean, 2002). Many movement

variables change, but not all of them exhibit the same sensitivity when running in a fatigued condition (Martens, Deflandre, Schwartz, Dardenne, & Bury, 2018).

Fatigue is an influential factor for lower extremity and trunk mechanics during running. With fatigue imposing adverse effects in neuromuscular function, an expected reduction in the transfer of mechanical energy during the function of the stretch-shortening cycle can occur (Mizrahi, Verbitsky, & Isakov, 2000a, 2000b) along with a decrease in muscle reaction times (Mizrahi, Verbitsky, & Isakov, 2001). Moreover, fatigue effects on trunk kinematics were found in the literature. Trunk flexion and extension in the sagittal plane was different in a group of recreational runners pre and post the implementation of fatiguing running protocol (Koblbauer, van Schooten, Verhagen, & van Dieën, 2014). Fatigue obscures the athletes' effort to maintain optimal angular displacements, as the stance phase becomes more variant during exhaustive running (García-Pinillos et al., 2020).

Several studies determined fatigue during incremental running protocols as the point that breaks the linearity between the dependent and the independent variable, using physiological or EMG data. Researchers used the integrated EMG and exercise intensity, (Nagata, Muro, Moritani, & Yoshida, 1981) or tried to establish a fatigue threshold based on a significant increase in the magnitude of the integrated EMG at a given running intensity (Hanon, Thépaut-Mathieu, & Vandewalle, 2005). Moreover, recent studies tried to identify physiological or electrophysiological thresholds using machine learning (Miura et al., 2020; Zignoli et al., 2019).

Nevertheless, little is known about how the gradual increase in exercise intensity and consequently fatigue affects running mechanics and if any interactions are developed in order to maintain the co-ordinated running pattern. It is possible that kinematic features, like segmental angular velocities or displacements, and force characteristics like rate of force development of impact forces, are sensitive and may interact as a runner shifts away from the steady metabolic state.

It is logical to expect that increasing running intensity will affect various biomechanical parameters. Indeed, the pattern of running mechanics that takes place at faster or slower speeds interacts with the impact forces, fatigue and metabolic performance of the motion (Schubert, Kempf, & Heiderscheit, 2014). Yet, it is unknown to what extent or direction (specific pattern) this is happening.

The primary purpose of the current study was to map the structure of change in the biomechanical characteristics of running, using a physiological threshold as a criterion to identify the decline in measures of

performance during increasing running intensity. The secondary aim was to create a model that can accurately classify biomechanical parameters in a fatigue class and account for possible interactions between the predictor variables. Such a model could be a valuable tool for athletes and coaches to effectively guide training.

2. Methods

Thirteen male recreational runners (age = 37.84 ± 4.53 years, height = 178.15 ± 5.37 cm, weight = 78.85 ± 6.89 kg) voluntarily participated in the study. Participants were healthy and free of any neuromuscular or musculoskeletal disorders and had at least three years of systematic training and participation in races longer than 10 km. The research was approved by the university's Ethics Committee (EH-12/2020).

Kinematics, kinetics and gas exchange were collected during an incremental running to exhaustion test on a ramp treadmill (Impulse RT700, UK). Measurements took place between 14:00 pm and 18:00 pm. The day before testing, participants performed a five-minute slow run wearing the measuring equipment for familiarisation purposes. Prior to testing, an eight-minute light warm-up at a self-comfort velocity followed by 5 min of dynamic stretching took place. The main running test started with a velocity equivalent to 85% of each athlete's 10 km tempo, which was approximately between 2.5 and 3.61 m/s for all individuals. After the first stage, an increase in the workload of 0.28 m/s was performed every 3 min until the participant reached exhaustion, becoming unable to keep the increase in workload and voluntarily interrupted the procedure. Respiratory data were obtained from a portable breath-by-breath gas analyzer (PNOE, ENDO Medical, Palo Alto, CA).

The treadmill was fixed on top of a dual force plate system (k-Delta, K-Invent Biomechanique, Montpellier, France) that measured the ground reaction force data (vGRF) with a sampling frequency of 516 Hz. Also, a pair of USB connected, 6 DoF IMU_s (k-sens, K-Invent Biomechanique, Montpellier, France) with 218 Hz sampling frequency, provided kinematic information about the torso. The minimum detectable step (raw accuracy) for the IMU sensors was 4 mg/LSB (Least Significant Bit) for the accelerometer, and $0.06^\circ/\text{s}$ for the gyroscope. One sensor was mounted over C7 and the other over L5. Both systems (force-plates and IMU_s) were internally synchronised.

A second-order low pass Butterworth filter with cut-off frequency of 30 and 15 Hz was used for filtering the kinetic and kinematic data respectively. All kinematic and force data variables represented discrete values extracted from every step (Table 1). Respiratory data

Table 1. List of extracted features.

Abbreviation	Feature	Unit
GRF _{peak}	Peak force value per contact	N·kg ⁻¹
RT _{LTu} , RT _{MDu} , RT _{APu}	Angular range around the three axes of motion (C7) RT _{LTu} = vertical rot, RT _{MDu} = flex/ext, RT _{APu} = lateral flex/ext	deg
<i>I</i> _{total}	Total impulse	N·s
RT _{LTlo} , RT _{MDlo} , RT _{APlo}	Angular range around the three axes of motion (PSIS) RT _{LTlo} = rotation-vertical, RT _{MDlo} = flex/ext, RT _{APlo} = lateral flex/ext	deg
RFD _{maxD}	Maximum value of rate of force development until GRF _{peak}	kN·s ⁻¹
RFD _{avgD}	Average value of rate of force development until GRF _{peak}	kN·s ⁻¹

(O₂, CO₂ and VE curves) were smoothed using a moving average filter with a window of 11 breaths. Two experienced independent researchers examined the VE/VO₂ and VE/VCO₂ curves to identify the ventilatory threshold that is closely related to the increase of anaerobic processes. Specifically, an over proportional increase of VE vs. VCO₂ output (VE/VCO₂) was used to define the ventilatory threshold (anaerobic threshold) (Meyer, Lucía, Earnest, & Kindermann, 2005). Their mean estimation was set as the time point that divides the data into two conditions: before and after fatigue initiation. Twenty seconds were removed pre and post the selected point. Since different number of stages were completed from every runner, a backward selection of the last five stages was adopted for everyone, so all stages represented the same magnitude of physical exertion (Figure 1(a,b)).

Afterwards, the dataset was searched for outliers. A point was considered as an outlier if its value fell out of ± 3 standard deviations away from the local mean over a length specified by a three-minute window, symmetrically expanding from both sides of it. In such case, the point's value was replaced with the nearest non-outlier point. For all the participants, the percentage of the outliers was less than 1% of their respected dataset. Finally, a 29,650×10 matrix was created with rows representing every individual step for each of the participants and columns the measured features.

An ensemble robust against over-fitting classification algorithm, called Random Forest (RF; Breiman, 2000), was built to create a classification model that would distinguish instances before and after the point of a ventilatory threshold. RF creates multiple decision trees that are trained on random subsets selected with bootstrap aggregation from the training dataset, a procedure that diversifies the trees. During the model built up, each time a split in a tree is computed, a random sample of *m* predictors is selected from the whole set of predictors (independent variables or features). This loop ends when out-of-bag (OOB) error is minimised. Typically, this kind of classifier produces better results in terms of accuracy among other machine learning algorithms, with reasonable computational time reduction. RF also accounts for complex non-linear interactions among predictors and returns a variable

importance score which can be used to rank the features according to their contribution over predicting the response variable. The RF algorithm was developed and deployed with the “randomForest” library (Liaw & Wiener, 2002).

Validation of the results was checked by splitting the data into training (70%) and testing (30%) sets. The “mtry” (*m* predictors randomly sampled as candidates at each split) parameter was set to the default value ($\sqrt{m}$). The “ntree” (number of trees to be grown by the model) parameter was set to 150 as this value was low enough not to overgrow the forest and, at the same time, provided a low OOB error. Nevertheless, before going forward with the selected values for the parameters “mtry” and “ntree”, multiple “for” loops were built to exhaustively search for different parameter results and validated against grid search as described in the “caret” R package (Kuhn & Johnson, 2013).

To interpret the model results, the following method was used. First, the feature importance was calculated, a procedure which is defined as the increase in the model's prediction error after we permuted the values of the features (break the relationship between the feature and the outcome) (Breiman, 2000). A feature is considered important if permuting its values will end up increasing the misclassification error (degrade the performance). The built-in function *importance* was used to assess how important each feature of the model was in predicting the outcome.

Secondly, partial dependence plots (PDP) were used to show the marginal effects of each feature on the predicted target variable (Friedman, 2001). Also, Friedman's *H*-statistic (Friedman & Popescu, 2008) was computed to assess the strength of interactions between predictors. Finally, two-dimensional PDP was used for visualising interactions. Statistical analysis and RF implementation were performed with R v3.6.1 (R Foundation for Statistical Computing, Vienna, Austria).

3. Results

Model's performance was assessed using accuracy, sensitivity, specificity and Cohen's kappa coefficient (Kuhn & Johnson, 2013). All 10 predictors passed on the final model. The model reached a testing set accuracy value

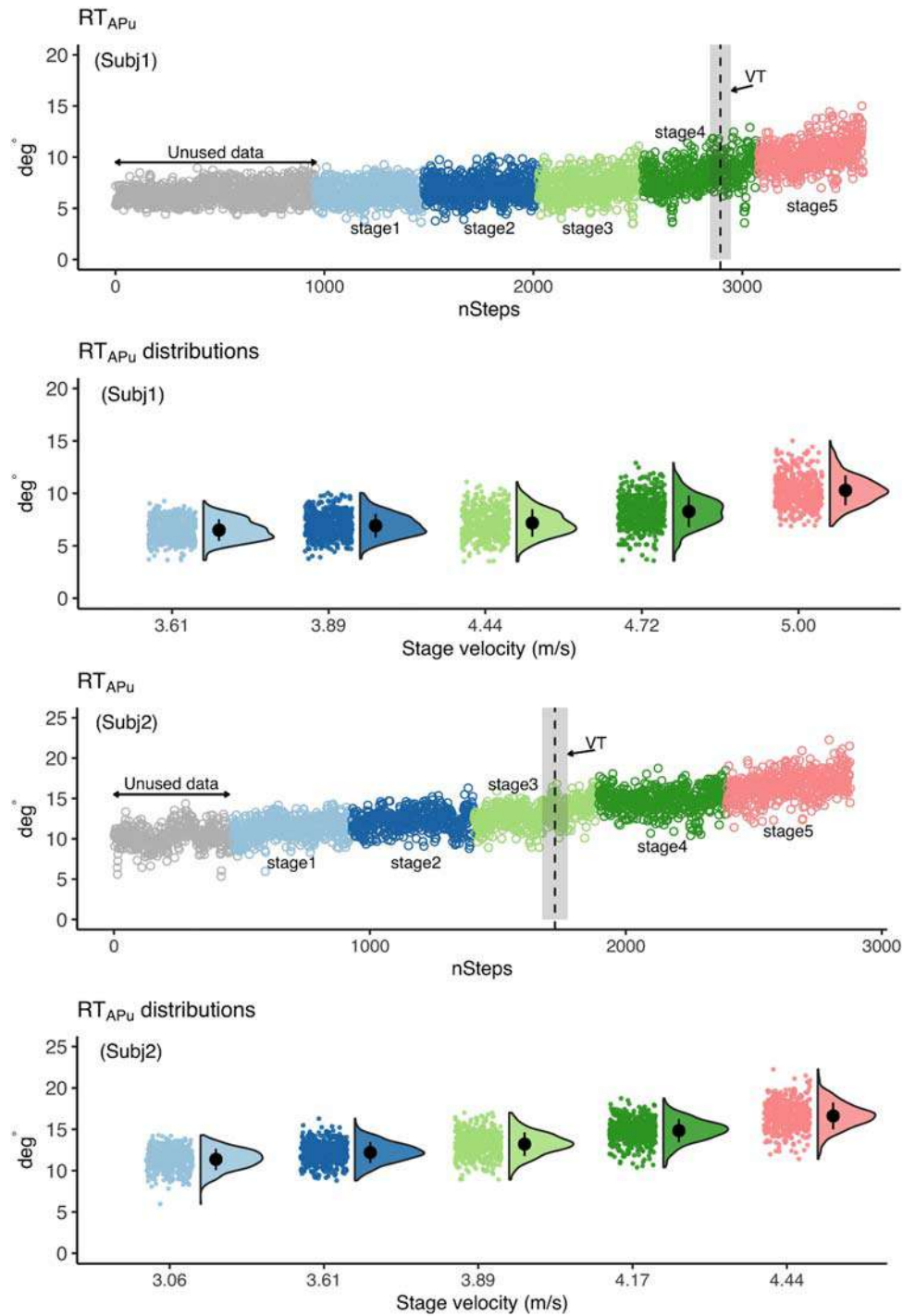


Figure 1. (a) RT_{APu} data as function of step count (upper panel) and the data distribution for each stage (lower panel) for a single individual (1). (b) RT_{APu} data as function of step count (upper panel) and the data distribution for each stage (lower panel) for a single individual (2).

of 0.914 (95% CI: 0.907–0.919) with sensitivity, specificity and kappa values at 0.93, 0.89 and 0.82, respectively (Table 2).

Variable importance ranked the selected features according to their value for the model, to predict the state above the ventilatory threshold. The four most influential variables were RT_{APu} , RFD_{maxD} , RT_{LTu} and

GRF_{peak} (average reduction in accuracy: 0.142, 0.102, 0.087 and 0.074, respectively).

PDPs were used after the identification of the most relevant features to understand the underlying relationship between each variable and the predicted outcome of the model (Figure 2). The black line in Figure 2 shows the partial dependence (PD) for the post ventilatory

Table 2. Descriptive statistics, mean-SD and median-IQR for every stage.

	Stage 1		Stage 2		Stage 3		Stage 4		Stage 5	
	Mean ($\pm 1SD$)	Median (IQR)	Mean ($\pm 1SD$)	Median (IQR)	Mean ($\pm 1SD$)	Median (IQR)	Mean ($\pm 1SD$)	Median (IQR)	Mean ($\pm 1SD$)	Median (IQR)
GRF _{peak} (N·m ⁻¹)	25.1(3.23)	24.5(4.21)	26(3.5)	25.2(4.99)	26.8(3.75)	25.9(5.51)	27.9(4.49)	26.6(5.93)	29.8(5.76)	27.9(8.4)
RFD _{maxD} (kN·s ⁻¹)	60.07 (16.16)	59.35(23.4)	65.27 (17.66)	65.19 (25.94)	71.45(19.8)	71.38 (30.35)	78.71(22.9)	76.81 (36.31)	91.02 (29.85)	85.53 (47.54)
RT _{APu} (deg)	14.7(5.43)	13.8(8.28)	16.1(5.65)	15.9(8.56)	17.9(5.94)	18.3(9.21)	19.1(6.01)	19.7(8.58)	19.6(5.87)	19.8(7.47)
RT _{LTu} (deg)	23.6(4.38)	24(5.86)	25(4.04)	25.3(4.35)	26.2(4.29)	26.2(4.69)	27.5(4.62)	27.6(5.02)	28.7(4.96)	29.1(5.45)

threshold defined class, for each of the top four features. For the RT_{APu} the plot shows that angular range above $\sim 10^\circ$ is associated with a steady linear increase in the probability of an observation to belong in the later stages of the test (Figure 2(I)). Similarly, for RFD_{maxD} on average, the probability increases slowly for values less than $\sim 50 \text{ kN s}^{-1}$, plateaus until it reaches 75 kN s^{-1} and then increases rapidly until $\sim 115 \text{ kN s}^{-1}$ and then remains stable until $\sim 150 \text{ kN s}^{-1}$ (Figure 2(II)). RT_{LTu} is relatively stable until $\sim 25^\circ$ (Figure 2(III)) when there is a shift towards higher probability. For the GRF_{peak}, probability increases from ~ 22 to $\sim 0.38 \text{ N kg}^{-1}$ and then remains stable (Figure 2(IV)). For the RT_{LTu}, the probability increases slightly from the minimum value of the distribution up until $\sim 23^\circ$.

The results of interaction strength test (*H*-statistic) showed that especially the interacting pairs RT_{APu}–RFD_{maxD}, RT_{LTu}–RT_{APu} and GRF_{peak}–RT_{LTu} had a considerably strong effect in predicting the fatigue state (larger than ~ 0.4). These effects are visualised as 2D PDP plots (Figure 3).

4. Discussion

The results demonstrated that RF-approach was a robust method for classifying the post fatigue condition based on the present data. Classification metrics displayed that the model was accurate, sensitive and specific in its predictions. Furthermore, variable importance showed that certain variables were more sensitive in predicting the outcome variable. Various approaches were considered to interpret and confirm that the model produces meaningful results, independently of how accurate it might be. Upon them, it can be inferred that the emerged variables are carrying adequate information to explain the studied phenomenon.

To further support this result, their mechanical function in running needs to be explained. Few studies have investigated trunk kinematics during running. Frontal plane kinematics (RT_{APu}) were identified as critical for balance control due to the trajectory of the centre of mass which is medial to the base of support (Winter, 1995). Flexing the upper trunk laterally closer towards

the supporting leg has been proposed as a strategy assisting hip abductors to oppose the adduction torque mainly induced by GRF during running (Kulmala et al., 2017).

RFD_{maxD} has been thoroughly investigated because it was early identified as a very important parameter for quantifying loading of the lower extremities (Hreljac, Marshall, & Hume, 2000). The present study showed that although there is an effect of velocity (Keller et al., 1996), feature variability should also be considered in order to understand its effect. Similar increases in mean values over the same velocity ranges as those previously reported (Brughelli, Cronin, & Chaouachi, 2011) were found, yet with a considerable increase in variability measures such as standard deviation. Comparatively, in the present study, according to Table 2, the increase in standard deviation from minimum to maximum velocity was $\sim 84.7\%$, whereas in a previous study (Hinrichs, 1987) it was only $\sim 36.6\%$ for the equivalent velocity increase. There was an average increase in the difference of the standard deviation between stages 1–4 at $\sim 45\%$. Particularly, the difference in standard deviation from fourth (22.96 kN) to fifth (29.85 kN) stage was 6.89 kN where from third (19.8 kN) to fourth (22.96 kN) was 3.16 kN representing $\sim 118\%$ increase. This phenomenon is explained by the inherent development of fatigue, since the present measurements of variability are increased, in contrast with findings that reported increasing stability with increases in velocity in non-fatigue running set-ups (Brughelli et al., 2011).

Vertical axial rotation of the upper body in the RT_{LTu} axis is a consequence of arm swinging and is considered as a mechanism for compensating the free rotational moments (i.e. torque around RT_{LTu} axis) produced by the movement of the lower extremities. Several authors provided evidence that during running, the horizontal angular momentum of the upper and lower body is of equal or almost equal in magnitude and in opposing directions, resulting in a net angular momentum near zero for the entire body (Hinrichs, 1987). Thus, they concluded that arm moments serve to cancel lower limb moments about the body's vertical axis. In the present study, RT_{LTu} was also rather important for

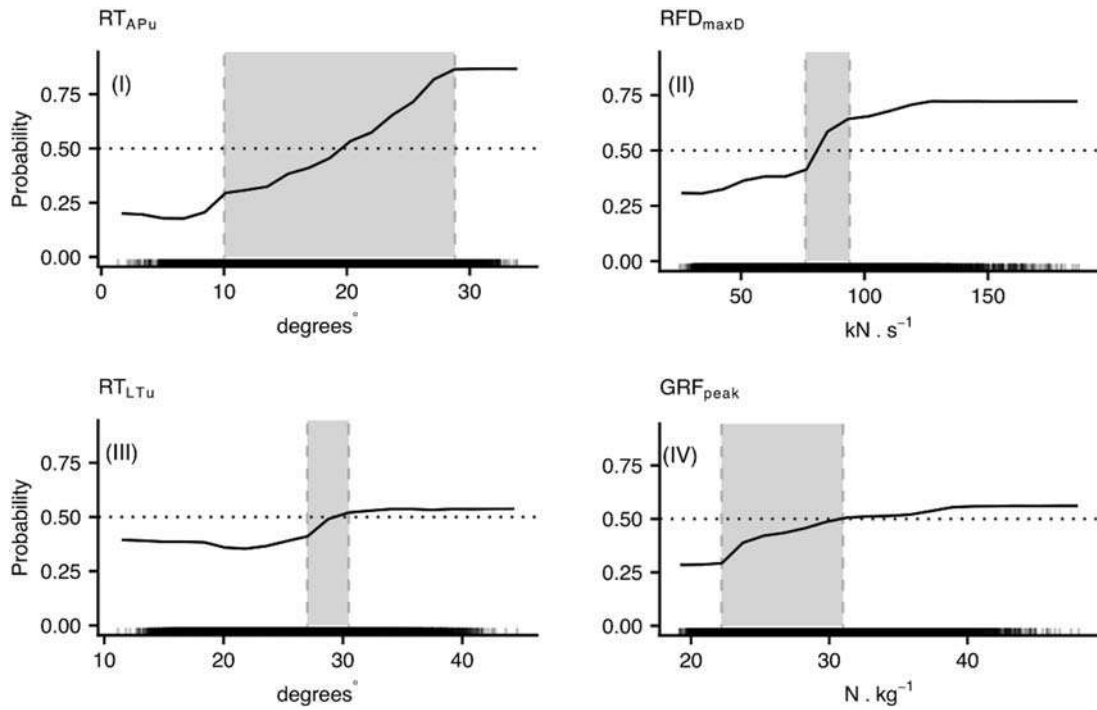


Figure 2. The PD plot for the top four (4) input features. Shaded area shows how the modifications of the values of each feature (taken separately whereas the others stay unchanged) affect the classification. The closer the values of probability on the y-axis get to 0, the more likely those values are to be in the pre ventilatory threshold region; the closer the line is getting to 1, the more likely the values are to post ventilatory threshold region. The rug lines at the bottom of each plot describe the distribution of the data by visualising individual points.

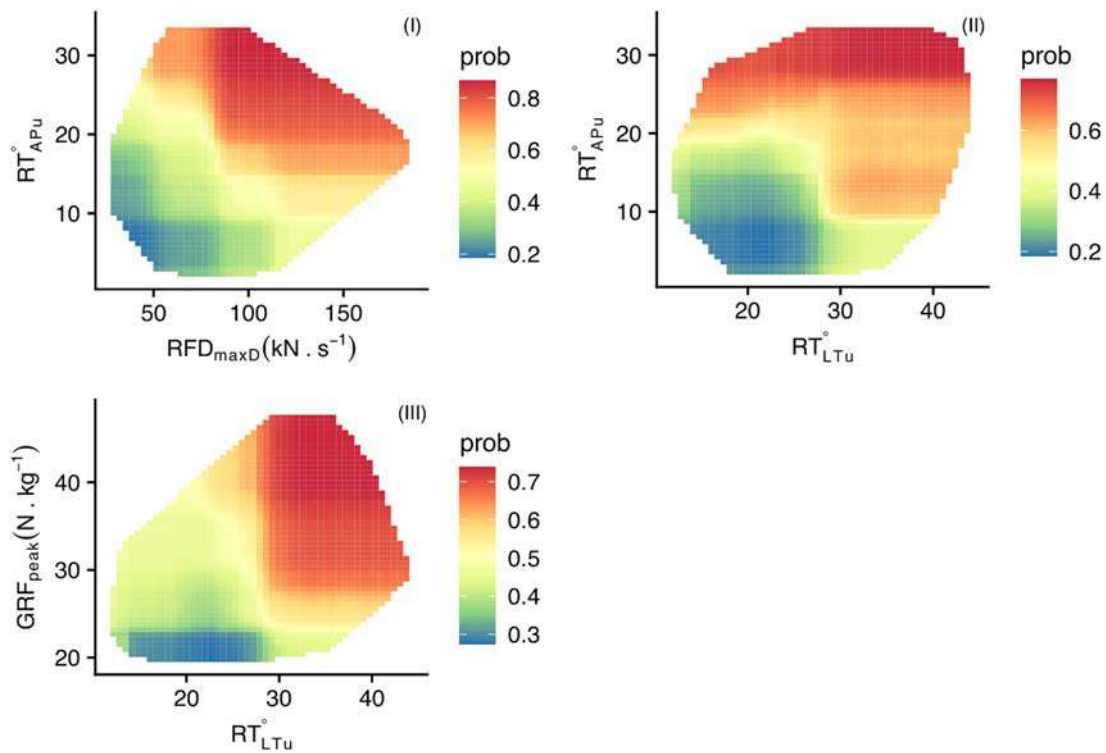


Figure 3. Illustration of PDP in 2 dimensions (2D). These plots can capture the way features interact and the effect of these interactions on the prediction. Various combinations of two features can be computed. In this graph we present the interactions among the top 4 features as they also had the greatest values of H -statistic for interaction strength.

the accurate outcome prediction. RT_{LTu} angular range was trending upwards with increasing velocity but the standard deviation and inter-quartile range were stable across the different stages (SD: $\pm 4.04^\circ$ – 4.96° and IQR: 4.35° – 5.86°) as imprinted in Table 2.

During normal running at medium velocities, such as those that are usually achieved at endurance running, it has been found that the upper body (trunk and arms) provides a critical amount of the angular impulse needed from the lower extremities for alternating their strides (Hinrichs, 1987). Thus, increases in the amplitude of the angular range and a low variability across the different stages of the test are very likely to be a compensation for the decreased performance of the lower body caused by fatigue. Without the upper body rotating around the vertical axis, although the legs would still manage to produce a flight phase by themselves, it would not be possible to provide the necessary torque to reverse their own angular momentum, which is a prerequisite for preparing the next contact with the ground (Hinrichs, 1987).

GRF_{peak} has been studied extensively in running, due to the association of impact forces with repeated loading and particular types of overuse injuries or syndromes of musculoskeletal system (Cavanagh & LaFortune, 1980; Nigg, Bahlens, Luethi, & Stokes, 1987). GRF_{peak} was also important for the model. This feature exhibited a similar behaviour with RFD_{maxD} , with a stable increase over the first three stages of the test and more variable for the last two. Also, a large increase in standard deviation was evident with the difference between the first and last stage reaching $\sim 78\%$ on average where other studies (Keller et al., 1996) reported increases of about 37.5%. Particularly, in the present data, the standard deviation of stages 1–3 was increased $\sim 16.22\%$ while for stages 3–5 $\sim 53.3\%$ (Table 2).

The PDPs indicated that there are limits for predictors within which they provide value for the model, but beyond them no usable information is present. PDP depict how the different values that a feature could take, affect and regulate model predictions. The practice of using PDP to explain the results of a machine learning model can provide the researcher with cut-off points beyond which the prediction rate becomes sufficiently acceptable and stable. In other words, the PDP analyses gave an idea of the values above or below which the model can classify observations with the greatest possible certainty (see Figure 2). PDP are not a measure of correct classification. Actually, what can be inferred from them is when the model starts to favour the prediction of one class versus the other in a probabilistic way. PDP analysis in machine learning models and especially in highly non-linear complex algorithms such as RF can also provide some idea about the pathways that features

interact with each other, which is not entirely possible in traditional statistical analyses (Zhao & Hastie, 2019). The use of the 2D PDP plots showed diverse combinations of features (Figure 3). More specific, 2D PDP plots revealed that there are certain areas in the joint distributions that are more sensitive in predicting fatigue state. RT_{APu} – RFD_{maxD} and GRF_{peak} – RT_{LTu} have a very distinct pattern where both variables have to surpass a certain magnitude threshold in order to produce meaningful predictive information. On the other hand, a more complex relationship was evident for RT_{APu} – RT_{LTu} . It appears that RT_{APu} interacts with RT_{LTu} only in the high range of its distribution, whereas RT_{LTu} can interact in a broader range of values. This is likely referring to the fact that RT_{LTu} is more sensitive to technique variations in contrast with RT_{APu} , which is rather robust to technique variations because of its strong association with the maintenance of balance. More specific, rotation around the anterior-posterior axis (lateral flexion) is where the trunk moment of inertia has its highest value (Whitsett, 1963) and thus, even small deviations in angular ranges could produce large undesirable torques that compromise balance.

Although there was a very large number of data points for the model to consider, the studied sample was small to account for the substantial inter-individual variations. Future studies should try to incorporate larger sample sizes and possibly a single subject analysis perspective.

5. Conclusion

In conclusion, results supported the machine learning approach using a RF classifier for the classification of group-based running patterns before and after the development of fatigue. Also, RT_{APu} , RFD_{maxD} , RT_{LTu} and GRF_{peak} were identified as the most important variables in the classification model. Nevertheless, interpretation of algorithms that can take into account various complex non-linear relationships in the feature space is a demanding task and an ongoing field of research.

Acknowledgements

The authors would like to thank Professor Georgios Mavromatis, Associate Professor Argyris Toubekis for their consulting, Mr Georgios Manakis for his help during the testing processes, and the participants for taking part in the study.

Disclosure statement

We wish to confirm that to our knowledge there are no conflicts of interest related to this study that could have influenced its outcome.

ORCID

Christos Chalitsios  <http://orcid.org/0000-0002-0886-5936>
 Thomas Nikoladis  <http://orcid.org/0000-0002-5305-1472>

References

- Abbiss, C. R., Peiffer, J. J., Meeusen, R., & Skorski, S. (2015). Role of ratings of perceived exertion during self-paced exercise; what are we actually measuring? *Sports Medicine*, 45, 1235–1243.
- Breiman, L. (2000). Random forests. *Machine Learning*, 45, 5–32. doi:10.1023/A:1010933404324
- Brughelli, M., Cronin, J., & Chaouachi, A. (2011). Effects of running velocity on running kinetics and kinematics. *Journal of Strength and Conditioning Research*, 25(4), 933–939. doi:10.1519/JSC.0b013e3181c64308
- Cavanagh, P. R., & LaFortune, M. A. (1980). Ground reaction forces in distance running. *Journal of Biomechanics*, 13(5), 397–406. doi:10.1016/0021-9290(80)90033-0
- Clansey, A. C., Hanlon, M., Wallace, E. S., & Lake, M. J. (2012). Effects of fatigue on running mechanics associated with tibial stress fracture risk. *MSSE*, 44, 1917–1923.
- Derrick, T. R., Dereu, D., & McLean, S. P. (2002, June). Impacts and kinematic adjustments during an exhaustive run. *Medicine & Science in Sports & Exercise*, 34(6), 998–1002.
- Devries, H. A., Tichy, M. W., Housh, T. J., Smyth, K. D., Tichy, A. M., & Housh, D. J. (1987). A method for estimating physical working capacity at the fatigue threshold. *Ergonomics*, 30, 1195–1204.
- Enoka, R. M., & Duchateau, J. (2016). Translating fatigue to human performance. *Medicine & Science in Sports & Exercise*, 48(11), 2228–2238. doi:10.1249/MSS.0000000000000929
- Friedman, J. H. (2001). Greedy function approximation: A gradient boosting machine. *Annals of Statistics*, 29(5), 1189–1232.
- Friedman, J. H., & Popescu, B. E. (2008). Predictive learning via rule ensembles. *The Annals of Applied Statistics*, JSTOR, 2(3), 916–954.
- García-Pinillos, F., Cartón-Llorente, A., Jaén-Carrillo, D., Delgado-Floody, P., Carrasco-Alarcón, V., Martínez, C., & Roche-Seruendo, L. E. (2020, February). Does fatigue alter step characteristics and stiffness during running? *Gait & Posture*, 76, 259–263. doi:10.1016/j.gaitpost.2019.12.018
- Hanon, C., Thépaut-Mathieu, C., & Vandewalle, H. (2005). Determination of muscular fatigue in elite runners. *European Journal of Applied Physiology*, 94, 118–125. doi:10.1007/s00421-004-1276-1
- Hinrichs, R. (1987). Upper extremity function in running. II. Angular momentum considerations. *International Journal of Sport Biomechanics*, 3, 242–263.
- Hreljac, A., Marshall, R. N., & Hume, P. A. (2000). Evaluation of lower extremity overuse injury potential in runners. *MSSE*, 32(9), 1635–1641.
- Keller, T. S., Weisberger, A. M., Ray, J. L., Hasan, S. S., Shiavi, R. G., & Spengler, D. M. (1996). Relationship between vertical ground reaction force and speed during walking, slow jogging, and running. *Clinical Biomechanics*, 11(5), 253–259. doi:10.1016/0268-0033(95)00068-2
- Koblbauer, I. F., van Schooten, K. S., Verhagen, E. A., & van Dieën, J. H. (2014, July). Kinematic changes during running-induced fatigue and relations with core endurance in novice runners. *Journal of Science and Medicine in Sport*, 17(4), 419–424. doi:10.1016/j.jsams.2013.05.013. Epub 2013 Jun 19. PMID: 23790535.
- Kuhn, M., & Johnson, K. (2013). Measuring performance in classification models - evaluating predicted classes. In *Applied predictive modeling* (pp. 254–261). New York, NY: Springer.
- Kulmala, J. P., Korhonen, M. T., Kuitunen, S., Suominen, H., Heinonen, A., Mikkola, A., & Avela, J. (2017, September). Whole-body frontal plane mechanics across walking, running, and sprinting in young and older adults. *Scandinavian Journal of Medicine & Science in Sports*, 27(9), 956–963. doi:10.1111/sms.12709
- Kumagai, S., Tanaka, K., Matsuura, Y., Matsuzaka, A., Hirakoba, K., & Asano, K. (1982). Relationships of the anaerobic threshold with the 5, 10 km, and 10-mile races. *European Journal of Applied Physiology and Occupational Physiology*, 49, 13–23.
- Liaw, A., & Wiener, M. (2002). Classification and regression by randomForest. *R News*, 2(3), 18–22.
- Martens, G., Deflandre, D., Schwartz, C., Dardenne, N., & Bury, T. (2018). Reproducibility of the evolution of stride Biomechanics during exhaustive runs. *Journal of Human Kinetics*, 64, 57–69. Published 2018 October 15. doi:10.1515/hukin-2017-0184
- Meyer, T., Lucía, A., Earnest, C. P., & Kindermann, W. (2005). A conceptual framework for performance diagnosis and training prescription from submaximal gas exchange parameters—theory and application. *International Journal of Sports Medicine*, 26(Suppl 1), S38–S48. doi:10.1055/s-2004-830514
- Miura, K., Goto, S., Katsumata, Y., Ikura, H., Shiraishi, Y., Sato, K., & Fukuda, K. (2020, October 29). Feasibility of the deep learning method for estimating the ventilatory threshold with electrocardiography data. *NPJ Digital Medicine*, 3, 141. doi:10.1038/s41746-020-00348-6
- Mizrahi, J., Verbitsky, O., & Isakov, E. (2000a). Shock accelerations and attenuation in downhill and level running. *Clinical Biomechanics*, 15, 15–20.
- Mizrahi, J., Verbitsky, O., & Isakov, E. (2000b). Fatigue-related loading imbalance on the shank in running: A possible factor in the stress fractures. *Annals of Biomedical Engineering*, 28, 463–469.
- Mizrahi, J., Verbitsky, O., & Isakov, E. (2001, March). Fatigue-induced changes in decline running. *Clinical Biomechanics (Bristol, Avon)*, 16(3), 207–212. doi:10.1016/s0268-0033(00)00091-7
- Nagata, A., Muro, M., Moritani, T., & Yoshida, T. (1981). Anaerobic threshold determination by blood lactate and myoelectric signals. *The Japanese Journal of Physiology*, 31, 587–597.
- Nigg, B. M., Bahlens, H. A., Luethi, S. M., & Stokes, S. (1987). The influence of running velocity and midsole hardness on external impact forces in heel-toe running. *Journal of Biomechanics*, 20(10), 951–959. doi:10.1016/0021-9290(87)90324-1
- Schubert, A. G., Kempf, J., & Heiderscheid, B. C. (2014, May). Influence of stride frequency and length on running mechanics: A systematic review. *Sports Health*, 6(3), 210–217. doi:10.1177/1941738113508544
- Stirling, L. M., Tschanner, V. V., Fletcher, J. R., & Nigg, B. M. (2012). Quantification of the manifestations of fatigue

- during treadmill running. *European Journal of Sport Science*, 12(5), 418–424. doi:[10.1080/17461391.2011.568632](https://doi.org/10.1080/17461391.2011.568632)
- Wasserman, K., Whipp, B. J., Koyl, S. N., & Beaver, W. L. (1973). Anaerobic threshold and respiratory gas exchange during exercise. *Journal of Applied Physiology*, 35, 236–243.
- Whitsett, C. E. (1963). Some dynamic response characteristics of weightless Man. *AMRL Technical Report*, 63–18. Wright-Patterson Air Force Base, OH.
- Winter, D. A. (1995). Human balance and posture control during standing and walking. *Gait & Posture*, 3, 193–214.
- Zhao, Q., & Hastie, T. (2019). Causal interpretations of black-Box models. *Journal of Business & Economic Statistics*, 39(1), 272–281. doi: [10.1080/07350015.2019.1624293](https://doi.org/10.1080/07350015.2019.1624293)
- Zignoli, A., Fornasiero, A., Stella, F., Pellegrini, B., Schena, F., Biral, F., & Laursen, P. B. (2019, October). Expert-level classification of ventilatory thresholds from cardiopulmonary exercising test data with recurrent neural networks. *European Journal of Sport Science*, 19(9), 1221–1229. doi:[10.1080/17461391.2019.1587523](https://doi.org/10.1080/17461391.2019.1587523)