SUPPORT VECTOR MACHINES CAN CLASSIFY RUNNER'S ABILITY USING WEARABLE SENSOR DATA FROM A VARIETY OF ANATOMICAL LOCATIONS

Josh Carter, Adrian Rivadulla, and Ezio Preatoni

University of Bath, UK

We developed and tested an algorithm to automatically classify twenty runners as novice or experienced based on their technique. Linear accelerations and angular velocities collected from six common wearable sensor locations were used to train support vector machine classifiers. The model using input data from all six sensors achieved a classification accuracy of 98.5% (10 km/h running). The classification performance of models based on single sensor data showed a 56.3-94.5% accuracy range, with sensors from the upper body giving the best results. Comparisons of kinematic variables between the two populations confirmed significant differences in upper body biomechanics throughout the stride, thus showing applied potential when aiming to compare novice runner's technique with movement patterns more akin to those with greater experience.

KEYWORDS: running biomechanics, machine learning, inertial measurement unit, gait analysis

INTRODUCTION: For runners, coaches, and running technology manufacturers, there is great interest in exploring how running performance can be optimised whilst guaranteeing healthy participation. Alongside physiological factors, running technique is known to be determinant of running performance and injury risk (Moore, 2016). Running technique is an important focus of training, with multiple studies showing how training programmes can be effectively implemented to optimise lower body biomechanics (Napier et al., 2015) for better performance and lower injury risk (Crowell & Davis, 2011). However, thorough running technique analyses, to identify areas for improvement, can be costly and inaccessible to most runners.

In contrast to lab-based techniques, wearable technologies are more accessible and allow uninterrupted gait datasets to be collected in a 'real world' environment. However, there is reduced control over measurement conditions when using wearables. Machine learning has proven an effective technique to analyse these higher noise datasets. (Halilaj, 2018). For instance, Clermont et al. (2019) used a Support Vector Machine model (SVM) to identify runners as belonging to a 'competitive' or 'recreational' group, using three dimensional accelerations from a single Inertial Measurement Unit (IMU) attached to the sacrum. This study reported a maximum classification accuracy of 82.6% and 80.4% for male and female groups respectively. Such a classifier could be used to track a runner's technique development over time. Whilst commercially available wearables could be placed near the sacrum, this location is not popular in the consumer technology market. A network of sensors embedded in the devices currently used by runners could be a more accessible solution, offer improved biomechanical insights, and provide more information about a runner's technique.

The purpose of this study was to develop a SVM classification algorithm, which could successfully distinguish between experienced and novice runners using wearable sensor data to assess their running technique. Multiple sensor locations and combinations were analysed with the aim to minimise hardware requirements whilst still achieving high levels of classification accuracy.

METHODS: Twenty healthy males participated in this study and were allocated to the experienced (10) or novice (10) runners group, based on their recent 10 km race times and training volumes (Table 1). This study was approved by the University Research Ethics Committee for Health, and participants signed informed consent prior to data collection.

Six Delsys Trigno units (Delsys, Massachusetts, USA) were securely attached to the posterior right wrist, lateral right upper-arm, posterior T10 of the spine, sacrum, proximal tibial tuberosity, and lateral aspect of the right foot using medical adhesive spray and double-sided tape. These landmarks were chosen to replicate the location of widespread consumer technology with embedded IMUs (e.g. smartphones, smartwatches, shoe sensors) or the most common areas where wearable sensors have been used in previous running research. All participants completed three running bouts of four minutes (10, 11 and 12 km/h) on a treadmill (Powerjog JX200, Ultimate Fitness, Leeds) at a 1% gradient, with one-minute standing rest between each bout. Three-dimensional linear accelerations and angular velocities were logged at 370.37 Hz.

Data processing: Data were low-pass filtered (zero-lag 4^{th} order Butterworth - 20 Hz cut-off, Clermont et al., 2019). Filtered accelerations from the foot sensor were used to identify right foot-strikes through a validated gait event detection algorithm (Benson et al., 2019) and segment continuous raw kinematic data into strides. Each stride was time registered to 300 data points and every five consecutive strides were averaged to form a single, more consistent, waveform (Benson et al., 2018) that was labelled as belonging to an experienced or novice runner.

Data analysis: Experienced and novice runners were randomly paired, to create ten approximately equal folds for 10-fold cross-validation. For each cross-validation iteration, data in the training set were standardised (z-scores) and Principal Component Analysis was applied to reduce the dimensionality of the data from each sensor. The minimum number of principal components accounting for 90% variance within all chosen sensors formed the features of the training dataset. The validation set was transformed using the standardisation scaling factors and projected onto the principal components extracted from the training set. Support vector classifier models with a linear kernel were then trained with the standardised principal component values to differentiate between experienced and novice runners using multiple combinations of sensors. Average accuracy across the ten folds was calculated. Additionally, we further investigated the biomechanical differences between the two populations. Specifically, data from the individual sensors that provided the best classification accuracies were evaluated.

Statistical Parametric Mapping analysis (SPM) was used to identify at which point within the stride there were statistically significant differences between the average novice and experienced runner movement patterns. Open-source Python code (Pataky et al., 2016) was used to perform a 1-dimensional independent two-tailed t-test between selected acceleration or angular velocity waveforms. The input for these analyses were an average waveform for each participant running at 10 km/h. Significance between groups was accepted when the SPM{t} value exceeded the critical threshold ($\alpha = 0.05$) at any of the normalised time points within the full gait cycle, meaning that identically smooth random 1D data would produce clusters of that breadth with a probability of $p < 0.05$.

RESULTS AND DISCUSSION: On average, novices were 11 years younger ($P = 0.007$) and 11.1 kg heavier ($P = 0.003$) than the experienced runners, but there were no significant differences between the two sub-groups height ($P = 0.578$) (Table 1).

Table 1. Participant characteristics, values reported as mean ± one standard deviation. Criteria for group allocation: Experienced – < 40 min in 10 km races; > 25 km/week training distance; Novice – no regular running activity. P value reported is the result from an independent t-test (†) or Mann-Whitney U test (‡).

	Experienced (n = 10)	**Novice** (n = 10)	*P*
Age	35 ± 10	24 ± 4	0.007*‡
Mass (kg)	69.2 ± 6.6	80.3 ± 7.6	0.003*†
Height (m)	1.79 ± 0.07	180.4 ± 6.8	0.578†
Weekly distance (km)	46 ± 25	na	
10km Time	36:32 ± 2:18	na	

Classification accuracy: Multiple combinations of sensors often achieved classification accuracies over 95% (Figure 1), which is in line with and improves that reported in previous studies (Clermont et al., 2017; Clermont et al 2019). More specifically, the classifier using data from six sensors (10 km/h) could correctly identify running experience with an accuracy of 98.5%. Removal of the sacrum and tibia data from the models had minimal influence on classification accuracy. Combining data only from the T10 and upper-arm sensors resulted in the highest accuracy at an individual running speed, achieving 99.1% at 11 km/h.

As may be expected, several of the single sensor location models achieved lower accuracies than those within the combined sensor models. However, using data exclusively from the upper-arm still returned classification accuracies (average accuracy of 94.5% across the three running speeds) comparable to those reported by the combined sensor models. Classification performance was worse when data from lower body sensors were used in isolation (average classification accuracies of 60.9% and 56.3% from tibia and foot, respectively).

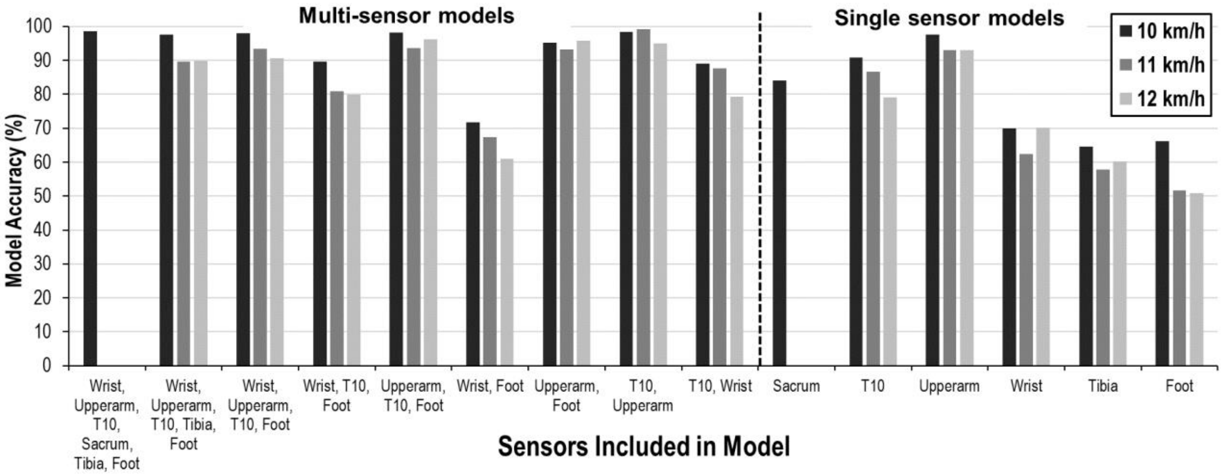

Figure 1. Classification accuracy for each sensor combination tested. Sensor combinations including the sacrum sensor were not testable at 11 and 12 km/h due to the sensor falling off.

Biomechanical Differences: Data collected from the lateral aspect of the upper-arm was analysed further, as this was the highest performing single-sensor classifier. The experienced runners elicited greater levels of linear acceleration along the anteroposterior axis (Figure 2). These findings seem to agree with previous research that conclude increased arm swing is an influential factor for a more efficient running technique, by maintaining a more constant horizontal velocity, and reducing 'unwanted' movement of the centre of mass and rotation of the upper body (Arellano & Kram, 2014). The experienced runners also elicited greater angular velocity around this same axis. In comparison to the novice runners, the experienced runners also showed greater inter-population consistency of arm abduction movements during the stance phase.

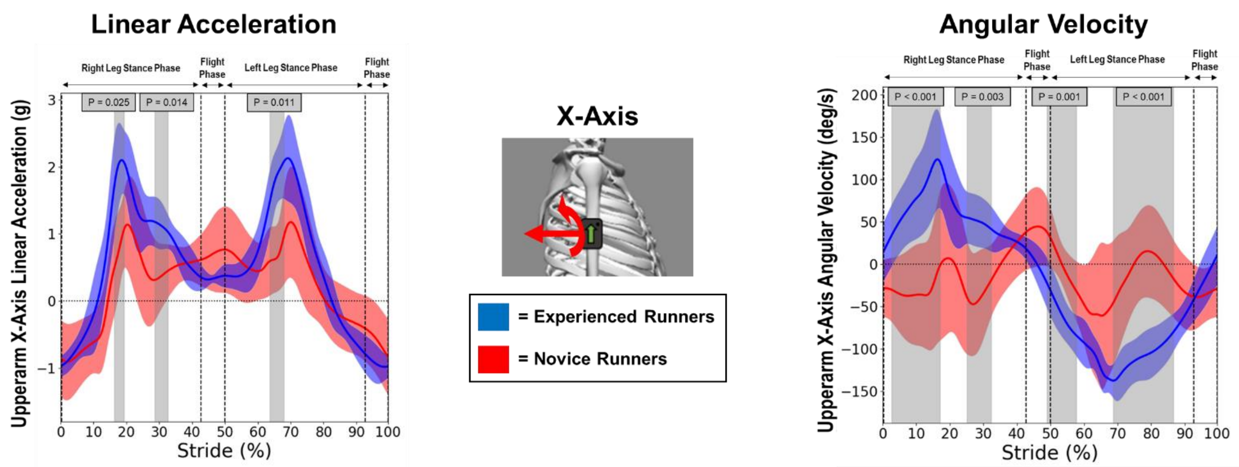

Figure 2. Upper-arm average (± one standard deviation) linear acceleration and angular velocity (10 km/h running). The grey bands indicate stride phases in which significant differences were found (SPM unpaired t-test) and the correspondent P-values are reported.

CONCLUSION: Linear support vector machine algorithms can successfully identify individuals belonging to a novice or experienced runner sub-group by utilising waveform data collected from multiple IMU sensor locations. Results suggest that upper body biomechanics can be most clearly differentiated between individuals of differing running experience. If the methodology presented in this study were to be implemented in commercially available wearables, it would have the potential to help novice runners gradually shift their technique towards that which is more characteristic of an experienced runner. Equally, it could be used to identify technique regression and consistency within experienced runners. This study serves as a preliminary methodological investigation that could be developed on in future studies with more ecologically valid environments and greater sample sizes.

REFERENCES:

Arellano, C. J., & Kram, R., (2014). The metabolic cost of human running: is swinging the arms worth it? *Journal of Experimental Biology, 217*(14), 2456-2461. 10.1242/jeb.100420.

Benson, L. C., Clermont, C. A., Osis, S. T., Kobsar, D., & Ferber, R. (2018). Classifying running speed conditions using a single wearable sensor: optimal segmentation and feature extraction methods. *Journal of Biomechanics, 71*(1), 94-99. 10.1016/j.biomech.2018.01.034.

Benson, L. C., Clermont, C. A., Watari, R., Exley, T., & Ferber, R. (2019). Automated accelerometer-based gait event detection during multiple running conditions. *Sensors, 19*(7), 1483-1501. 10.3390/s19071483.

Clermont, C. A., Benson, L. C., Osis, S. T., Kobsar, D., & Ferber, R. (2019). Running patterns for male and female competitive and recreational runners based on accelerometer data. *Journal of Sports Sciences, 37*(2), 204-211. 10.1080/02640414.2018.1488518.

Clermont, C. A., Phinyomark, A., Osis, S. T., & Ferber, R. (2017). Classification of higher- and lower-mileage runners based on running kinematics. *Journal of Sport and Health Science, 8*(3), 249-258. 10.1016/j.jshs.2017.08.003.

Crowell, H. P., & Davis, I. S. (2011). Gait retraining to reduce lower extremity loading in runners. *Clinical Biomechanics, 26*(1), 78-83. 10.1016/j.clinbiomech.2010.09.003

Halilaj, E. (2018). Machine learning in human movement biomechanics: best practices, common pitfalls, and new opportunities. *Journal of Biomechanics, 81*(1), 1-11. 10.1016/j.jbiomech.2018.09.009.

Moore, I. S., Jones, A. M., & Dixon, S. J. (2012). Mechanisms for improved running economy in beginner runners. *Medicine & Science in Sports & Exercise, 44*(9), 1756-1763. 10.1249/MSS.0b013e318255a727

Napier, C., Cochrane, C. K., Taunton, J. E., & Hunt, M. A. (2015). Gait modifications to change lower extremity gait biomechanics in runners: a systematic review. *British Journal of Sports Medicine, 49*(21), 1382-1388. 10.1136/bjsports-2015-095507.

Pataky, T. C., Vanrenterghem, J., & Robinson, M. A. (2016). The probability of false positives in zero-dimensional analyses of one-dimensional kinematic, force and emg trajectories. *Journal of Biomechanics, 49*(9), 1468-1476. 10.1016/j.jbiomech.2016.03.032.