LABORATORY VERSUS ECOLOGICAL RUNNING: A COMPARISON OF FOOT STRIKE ANGLE AND PATTERN ESTIMATION

Stephanie R. Moore[1], Christina Kranzinger[2], Gerda Strutzenberger[3], Magdalena Taudes[4], Aaron Martinez[1], Hermann Schwameder[1], Josef Kröll[1]

Department of Sport and Exercise Science, University of Salzburg, Austria[1]
Salzburg Research Forschungsgesellschaft m.b.H., Salzburg, Austria[2]
Motion Analysis Zurich, Department of Orthopaedics, Balgrist University Hospital, Children's Hospital, University of Zurich, Switzerland[3]
University of Applied Science Technikum Wien, Vienna, Austria[4]

The purpose of the current study was to evaluate the ecological validity of two previously developed laboratory-based random forest machine learning models and train two new ecologically valid models for the 1) prediction of foot strike angle (FSA) and 2) classification of foot strike pattern (FSP) from wearable insoles during running. The original models performed worse with track-surface running data inputs than in their original validation (prediction RMSE = 6.84° vs. 3.65°, classification accuracy = 79.5% vs. 94.1%). The new models, trained using track-surface data, improved the estimation of FSA (RMSE = 4.10°) and FSP (accuracy = 84.8%). To ensure estimation accuracy, future models should be trained with respect to the environment/conditions in which they will be implemented.

KEYWORDS: track running, random forest, pressure insoles

INTRODUCTION: The use of wearable sensors has enabled the collection of immense biomechanical information of sporting activities in their natural environments. Further, the employment of robust machine learning prediction and classification techniques on wearable sensor data can diversify the information collected from these sensors. Machine learning models work by capitalizing on distinctive features of large-scale data sets, thus the model training inputs should be accurate and reliable (Alpaydin, 2014). Training such models using data collected in the laboratory setting is beneficial because the environment can be controlled and gold-standard equipment employed. However, these benefits come with some challenges, in that biomechanically equipped laboratories are often limited in space, which may affect the nature of movement. This is especially relevant for movements like over ground running. In these cases, although lab-based models can be trained with high quality data inputs, the resulting model may not be fully applicable to more ecologically valid implementations. Alternatively, models trained with high ecological validity may lack gold-standard references for accurate model training due to equipment and logistical limitations. Ultimately, the accuracy of modelling techniques should be investigated specific to practical, in-field applications.
Thus, the purpose of the current study was to evaluate the accuracy and precision of two previously trained and validated laboratory-based random forest machine learning models (Moore et al., 2020) for the 1) prediction of foot strike angle (FSA) and 2) classification of foot strike pattern (FSP) when data from running on a track-surface was used in the model execution. An additional purpose of the current study was to train two new models using the independent and dependent variables from track-surface running as data inputs for comparison to the original lab-based models.

METHODS: The original models were trained and tested using independent variables collected from two-part wearable pressure insoles (separated by anterior/posterior sections) and a dependent, criterion measure of FSA or FSP from 3D marker-based kinematics (Moore et al., 2020). The current data collection closely resembled that of the original data collection in order to create comparable independent variable inputs for use in the model (Moore et al., 2020). To explore the whole spectrum foot strike patterns, researchers in the previous study investigated six types of FSPs: participants' natural pattern (NA), rear foot (RF), mid foot (MF), fore foot (FF), extreme-RF, and extreme-FF patterns. However, it was reported that participants found it difficult to execute MF steps (ultimately limiting the amount of MF steps in model training) and that "extreme" conditions felt very unnatural to them. Therefore, the current study collected an additional MF trial and removed the "extreme" conditions to account for these issues. As a result, in the current study, 13 recreational male and female runners (8 men, 5 women; Mean ± SD; 1.75 ± 0.09 m, 70.9 ± 11.7 kg, 31.3 ± 7.5 yr) were asked to complete five trials with the NA condition performed first, followed by trials of the RF, MF_1, MF_2, and FF strike patterns in a randomized order. Each trial consisted of 3-4 100 m loops on an indoor track surface (to ensure 20 steps were collected per condition per participant) at a self-selected, comfortable speed (3.34 ± 0.63 $m \cdot s^{-1}$). The capture volume was a 15 m straight phase, allowing anywhere between 4 and 14 steps to be collected per lap (dependent on the use of one or two working insoles). In total, after data loss and processing, 1,244 steps with the following pattern distribution were included in the current analyses: RF = 671, MF = 436, FF = 137.

Participants were equipped with Loadsol^{TM} wearable pressure insoles (Novel GmbH; Munich, Germany) over the insole of their personal running shoes. The insole system recorded at its maximum sampling rate (100 Hz). Three-dimensional (3D) motion capture of both feet was collected using a 10-camera Qualysis motion capture system (2020.2, Göteborg, Sweden). Retroreflective markers were attached at the medial and lateral malleoli, the head of the 2nd metatarsal, the heel (placed at the same height as the 2nd metatarsal) and the lateral side of the 1st and 5th metatarsals. Marker trajectories were captured at 200 Hz and filtered using a 15 Hz low-pass filter (> 99% of signal power retention; Moore et al., 2020). The foot-segments were modelled in Visual 3D X64 Professional (v6.03.06; Germantown, MD, USA) and virtual foot-segments were modelled ensuring that the shoe-elicited angulation was negated (C-Motion, Inc., 2017). The resulting FSA was extracted relative to the ground surface.

The subsequent processing steps were performed using a custom code in MATLAB (R2020a; The MathWorks, Inc., Natick, MA, USA). First, the kinematic and insole data were synchronized using a stomp event performed before and after each trial. The instant of ground contact of the stomps was found via the force threshold defined by Seiberl and colleagues (2018) for the insoles, and the peak positive acceleration of the foot segment for the kinematic data. Due to the difference in sample rate between the systems, the kinematic data was down-sampled (using linear interpolation) to match the Loadsol^{TM} data length (at 100 Hz).

Initial contact and toe off events for each running step were determined using the same thresholds used for stomp detection (Seiberl et al., 2018). The ten time and force related variables (i.e. independent variables; including impulse ratio, peak force, and peak rate of force development during the first 33% or 100% of the stance phase) in the original model were computed from the Loadsol^{TM} sensors between the initial contact and toe off events (Moore et al., 2020). Finally, the angle of the foot segment at initial contact (i.e. FSA) was extracted for each step performed within the capture volume as a criterion measure.

Following the methods of the original study, the Loadsol^{TM} independent variables of the 1,244 track-surface steps collected were fed into the original random forest models in the manner of a validation set (Moore et al., 2020). One random forest model was used to predict the FSA ($FRST_{PRED}$), while the second served to classify the FSP ($FRST_{CLASS}$). Therefore, the models estimated the FSA and FSP using the insole data. To assess the accuracy and precision of the models' use on the ecological (track-surface) running data, the estimates were compared to the kinematic ground truth measure. The accuracy of the $FRST_{PRED}$ was reported as the root mean squared error (RMSE) and mean absolute error (MAE) of the estimated vs. criterion FSA. The precision was quantified as the range of the Bland-Altman 95% limits of agreement (Bland & Altman, 2010). The criterion measure of the FSP was pre-classified as RF (FSA > 8.0°), MF (−1.6° ≤ FSA ≤ 8.0°), or FF (FSA < −1.6°) using the recommendations of Altman and Davis (2012). Subsequently, the $FRST_{CLASS}$ was assessed using confusion matrices and model accuracy, classifier recall, and classifier precision were reported in the same manner as the original assessment of the models (for a detailed explanation see Moore et al., 2020).

Finally, two new random forest models were trained and assessed using the same methods as the previous model-development (Breiman, 2001), however the track-surface running data was used as the input for the model training. As a result, 70% of the track-surface steps (n = 870) were used for model training, while 30% were used for model validation (n = 374). The aims of the models were the same: FSA prediction and FSP classification (henceforth referred to as ECO_FRST_{PRED} and ECO_FRST_{CLASS}, respectively). The most important independent variables for the new models were determined by ranking those with the highest “mean decrease Gini” (Calle & Urrea, 2011).

RESULTS: The accuracy and precision of the $FRST_{PRED}$ and $FRST_{CLASS}$ models with track-surface running inputs are presented in Table 1. When fed into the original $FRST_{PRED}$ model, the track-surface data had worse performance when compared to the original validation set (greater RMSE, MAE, bias and precision; Table 1). Similarly, the accuracy of the $FRST_{CLASS}$ for predicting FSP during track-surface running was 14.6 percentage points worse than the accuracy of the original validation set (Table 1). The performance of the new models (ECO_FRST_{PRED} and ECO_FRST_{CLASS}) are also presented in Table 1. Track-surface running FSA was predicted with less error and improved Bland-Altman bias and precision when using the new model. All classification performance metrics were greater when the track-surface running was assessed with the ECO_FRST_{CLASS} (as opposed to the original model). The most important variable (highest mean decrease Gini) was the peak rate of force development ratio between the aft insole sensor region and the total foot (Peak RFD_Aft) for both new models.

Table 1. The performance metrics of two previously published random forest models ($FRST_{PRED}$ and $FRST_{CLASS}$) are presented with track-surface running inputs. For ease of comparison, the accuracy and precision of the models using the original laboratory data (validation set) is also included. Further, the performance of the validation set of two new random forest models (ECO_FRST_{PRED} and ECO_FRST_{CLASS}) using the track-surface data set are also presented.

Model		**Original Lab Models***		**New Ecological Models**
Data set used for accuracy/precision assessment		Original Lab Validation Set (steps = 1,047)	Track-Surface Validation Set (steps = 1,244)	30% Track-Surface Validation Set** (steps = 374)
FSA Prediction model		*$FRST_{PRED}$*	*$FRST_{PRED}$*	*ECO_FRST_{PRED}*
RMSE		3.65	6.84	4.10
MAE		2.69	4.78	2.98
B-A Bias		-0.11°	-1.42°	-0.36°
B-A Precision		14.30°	26.23°	16.02°
FSP Classification model		*$FRST_{CLASS}$*	*$FRST_{CLASS}$*	*ECO_FRST_{CLASS}*
Model Accuracy (%)	ALL	94.1	79.5	84.8
Classifier Recall (%)	RF	96.4	87.6	90.4
	MF	76.7	78.1	80.7
	FF	96.3	44.5	70.3
Classifier Precision (%)	RF	97.0	86.3	87.7
	MF	74.8	69.6	79.6
	FF	95.9	83.6	89.7

*the original models used were consistent with those published by Moore et al., 2020; ** 70% of the data were used for model development; FSA = foot strike angle; FSP = foot strike pattern; B-A = Bland-Altman; RF = rear foot, MF = mid foot, FF = fore foot

DISCUSSION: All accuracy and precision metrics for the prediction of FSA and classification of FSP indicated inferior performance of laboratory-trained random forest models when implemented with track-surface running data inputs (Table 1). Further, when new models with the same purpose were trained using the track-surface running data, the models performed better to predict and classify foot strike from track-surface running than the models developed in the lab. However, the original lab-based models and lab-based validation data set (Moore et al., 2020) boasted better performance than the new models (i.e. when track-surface running

was used in model training). Importantly, the original models ($FRST_{PRED}$ and $FRST_{CLASS}$) were trained using a larger and more diverse data set (participants = 30, cases used for model training = 2,442 steps) than the newly trained models (ECO_FRST_{PRED} and ECO_FRST_{CLASS}). However, the original data set had a low percentage of MF steps in the validation set, which also had the lowest recall and precision of the $FRST_{CLASS}$. Therefore, the increased number of MF steps included in the track-surface running likely influences the overall model accuracy (as they are most likely to be mis-classified/predicted). Further, the average speed of the participants had a greater variability when track running was assessed (standard deviation of the average running velocity = 0.63 $m \cdot s^{-1}$ vs. 0.40 $m \cdot s^{-1}$ during the original laboratory trials). Therefore, the better accuracy and precision of the original models may be due to the greater number of data inputs in model training, the lower percentage of MF strikes included in the original validation, and the more uniform speed of the participants.
The most important variable (Peak RFD_Aft) of the ECO_FRST_{PRED} and ECO_FRST_{CLASS} were consistent with each other and that of the original $FRST_{PRED}$/$FRST_{CLASS}$ models (Moore et al., 2020). This indicates that the foot strike of both the laboratory and track-surface running can be distinguished from similar variables. However, because there was more predictive error of $FRST_{PRED}$ and lower classification accuracy of $FRST_{CLASS}$ when implemented with track-surface data, there is likely differences in the features of the independent variables. Therefore, future models should be trained with respect to the conditions (i.e. track-surface or laboratory) where the models will be implemented. Importantly, the insole-based independent variables may also be altered with running velocity differences, shoe midsole design, or running style (Seiberl et al., 2018) however further investigations are needed to determine their influence on the efficacy of random forest models to predict and classify foot strike during running.
Although the ECO_FRST_{CLASS} had a lower overall accuracy than $FRST_{CLASS}$ (84.8% vs. 94.1%, respectively), the new model was able to classify MF strike patterns with greater recall (+ 4.0 percentage points) and precision (+ 4.8 percentage points). This suggests that the larger number of MF strikes in the training data set may have improved the MF model performance.

CONCLUSION: The current ecological (track-surface) validation of two previously-developed random forest machine learning models proves the robustness of the modelling technique in that the main variables for estimation of both types of running are similar. However, to obtain the highest accuracy and precision during implementation, future models should be trained with respect to the environment in which they will be implemented.

REFERENCES

Alpaydin, E. (2014). *Introduction to Machine Learning* (3rd ed.). Massachusetts Institute of Technology.

Altman, A. R., & Davis, I. S. (2012). A kinematic method for footstrike pattern detection in barefoot and shod runners. *Gait & Posture*, *35*(2), 298–300.

Bland, J. M., & Altman, D. G. (2010). Statistical methods for assessing agreement between two methods of clinical measurement. *International Journal of Nursing Studies, 47*(8), 931–936. https://doi.org/10.1016/j.ijnurstu.2009.10.001

Breiman, L. (2001). Random forests. *Machine Learning*, *45*(1), 5.

Calle, M. L., & Urrea, V. (2011). Letter to the Editor: Stability of Random Forest importance measures. *Briefings in Bioinformatics*, *12*(1), 86–89. https://doi.org/10.1093/bib/bbq011

C-Motion, Inc. (2017). *Visual3D Wiki Documentation*. https://c-motion.com/v3dwiki/index.php?title=Main_Page

Moore, S., Kranzinger, C., Fritz, J., Stöggl, T., Kröll, J., & Schwameder, H. (2020). Foot strike angle prediction and pattern classification using LoadsolTM wearable sensors: A comparison of machine learning techniques. *Sensors*, *20*, 6737. https://doi.org/10.3390/s20236737

Seiberl, W., Jensen, E., Merker, J., Leitel, M., & Schwirtz, A. (2018). Accuracy and precision of loadsol® insole force-sensors for the quantification of ground reaction force-based biomechanical running parameters. *European Journal of Sport Science*, *18*(8), 1100–1109. https://doi.org/10.1080/17461391.2018.1477993

ACKNOWLEDGEMENTS: This work was partly funded by the Austrian Federal Ministry for Climate Action, Environment, Energy, Mobility, Innovation and Technology, the Austrian Federal Ministry for Digital and Economic Affairs, and the federal state of Salzburg.