Journal of Biomechanics 185 (2025) 112663

Contents lists available at ScienceDirect

Journal of Biomechanics

journal homepage: www.elsevier.com/locate/jbiomech

Neural networks can accurately identify individual runners from their foot kinematics, but fail to predict their running performance

Patrick Mayerhofer [a,*], Christopher Napier [a,b], David C. Clarke [b], Ivan Bajić [c], J. Maxwell Donelan [a,b]

[a] WearTech Labs, Simon Fraser University, Burnaby, Canada
[b] Department of Biomedical Physiology and Kinesiology, Simon Fraser University, Burnaby, Canada
[c] Department of Engineering Science, Simon Fraser University, Burnaby, Canada

ARTICLE INFO

Keywords:
Running Form
Neural Networks
Foot Kinematics
Spectrogram
Running Performance
Image Recognition
Running Biomechanics

ABSTRACT

Athletes and coaches may seek to improve running performance through adjustments to running form. Running form refers to the biomechanical characteristics of a runner's movement, and can distinguish individual runners as well as groups of runners, such as long-distance and short-distance runners. Yet, in long-distance running it is still unclear whether certain running forms lead to better performance. In this study, we used a neural network to test the extent to which individual running forms, measured from foot kinematics, exist within long-distance runners and whether running forms can predict performance. To accomplish this goal, 119 participants ran on a treadmill at three different speeds and overground at a self-selected sub- maximal speed while we collected data from insole-embedded Inertial Measurement Units (IMUs) mounted in both shoes. Participants reported their personal best 10 km run times. We used these data to train the neural network to identify individual runners from their running data. Then, we trained the same neural network architecture to predict the runners' performance. With enough data, the neural network was successful in identifying individual runners, but was comparable to a random coin flip (57 % accuracy) in predicting whether an individual runner is slow or fast. We interpret the success of the model to identify runners, but the subsequent failure of the same model to predict running performance as evidence that individual running form measured from foot kinematics contains insufficient information about a runner's performance.

1. Introduction

Running form, which refers to biomechanical characteristics of a runner's movement, varies between individuals and groups. Runners have individual running forms that algorithms can detect using either laboratory measurements like 3D motion analysis or wearable technology like Inertial Measurement Units (IMUs) (Hoitz et al., 2021; Horst et al., 2023; Weich et al., 2020). Running forms also vary between different groups of runners, including across age groups, distance specializations, and shoe models (Cunningham et al., 2013; Eskofier et al., 2012; Fukuchi et al., 2011; Horst et al., 2023; Patoz et al., 2022b; Phinyomark et al., 2014). Running forms can also contain multiple layers of information from a singular type of data. Using ground reaction force, an algorithm can separate individual runners, regardless of changing footwear, while using the same data the same algorithm can also separate the different footwear, regardless of the runner (Horst et al., 2023).

Despite the detectable differences in running forms, the association between a type of running form and long-distance performance remains uncertain. While good technique in sprinting can improve performance, the situation is less clear for long-distance running (Mero et al., 1992). Moore reviewed factors affecting running economy in distance runners and concluded that, while evidence is not strong, biomechanical variables during the stance phase had the strongest associations with running economy (Moore, 2016). Other studies have sought evidence for good and bad running forms. Agresta et al. found no relationship between years of running experience and kinetic and kinematic motion analysis data, while Patoz et al. found no relationship between running form and running economy (Agresta et al., 2018; Patoz et al., 2022a). Subjectively, even running coaches are unable to tell whether a specific running form is good or bad (Cochrum et al., 2021). In contrast to the above studies that were unable to identify performance from running form, Clermont et al. distinguished between recreational and

* Corresponding author at: 9639 – 137A Street, Surrey, BC V3V 0C6, Canada.
E-mail address: pmayerho@sfu.ca (P. Mayerhofer).

https://doi.org/10.1016/j.jbiomech.2025.112663
Accepted 27 March 2025
Available online 2 April 2025

competitive runners with 3D motion analysis independently of their sex as well as with IMU data collected at the center of mass when divided into their respective sexes (Clermont et al., 2019a; Clermont et al., 2017).

Neural networks have the potential to detect differences in the performance of running forms, beyond that which has been possible with more conventional statistics and machine learning methods. Traditionally, running science has used more conventional machine learning algorithms, such as support vector machines, for classifying movements (Clermont et al., 2019a, 2019b, 2017; Cust et al., 2019; Fukuchi et al., 2011; Horst et al., 2023; Moghadam et al., 2023; Nigg et al., 2012; Phinyomark et al., 2018). To classify data with more conventional machine learning algorithms, scientists typically extract features from the raw data before training the classification algorithm (Phinyomark et al., 2018). In running, these features could, for example, include stride time, coefficient of variation, or minimum and maximum joint angles (Benson et al., 2020, 2018; Fukuchi et al., 2011; Phinyomark et al., 2014). Compared to more conventional machine learning algorithms, neural networks often achieve higher classification accuracy in biomechanics (Barshan and Yüksek, 2014; Cust et al., 2019; Halilaj et al., 2018; Hoitz et al., 2021; Moghadam et al., 2023; Ngoh et al., 2018; Nigg et al., 2012; Nweke et al., 2018; Tabrizi et al., 2020). Neural networks can learn from complex, raw datasets without the need to extract features, ameliorating the risk of losing important information in the data, and — because of their adaptability to new data — can learn from an array of different datasets. Therefore, if detectable differences in running forms for different performances exist, neural networks may be able to better detect them than more conventional machine learning algorithms.

In this study, our primary aim was to test whether a neural network could predict long-distance running performance based on running form measured from foot kinematics. To accomplish this aim, 119 participants reported their running performance. They then ran on a treadmill and on overground while we collected foot kinematics measured as linear acceleration and angular velocity from IMUs inserted in the insoles of both shoes. We chose to use raw linear acceleration and angular velocity from the foot because they capture detailed biomechanical aspects of foot kinematics and are standard variables measured by popular commercial sensors. Compared to extracted features like stride time or peak acceleration, these raw variables reflect both the dynamic and rotational characteristics of running form. We next trained a neural network model to accurately identify individual runners (initial model output) from the IMU-measured foot kinematics (model input), thus confirming that there was sufficient information in the foot kinematics, and capability in our model architecture, to identify individual differences in running form. We then retrained the neural network and tested whether it could use the individual differences in running forms, based on the same model input (IMU-measured foot kinematics), to predict runner performance (new model output) rather than runner identity (initial model output). Failure of this model in predicting performance after success in classifying identity we consider as strong evidence that individual running forms do not contain enough useful information about a runner's performance. Our second aim for this study was to determine how the model's accuracy in identifying individual differences in running forms would scale with an increase in the number of runners or an increase in the amount of data for each runner. We designed this second aim to test the feasibility of using running forms as digital signatures for identification or verification of individual athletes.

2. Methods

2.1. Participants and Setup

This analysis is part of a larger study conducted with 188 healthy recreational runners (Napier et al., 2022). For our analysis, we excluded 25 participants that did not provide their 10 km personal best time, three because of a pain rating of over 2 out of 10, 24 because there was no overground data available, and 17 because measurements exceeded the sensor's capability to record them, leading to an incomplete capture of the data. Therefore, we included 119 runners for the study (female: n = 58; male: n = 61; age: 43 ± 11 years; body mass: 69.9 ± 11.3 kg; height: 175 ± 10 cm; US shoe size: 9.2 ± 1.6; running experience: 17 ± 13 years; average distance per week: 36 ± 23 km; mean ± std). Out of the 119 participants who completed one session, 37 repeated the experiment one week later, and 24 repeated the experiment a third time one month after the initial visit. To measure foot kinematics, we equipped participants with two IMUs (Plantiga Technologies Inc., Vancouver, Canada) mounted in each shoe insole. The IMU collected both acceleration and angular velocity at a sampling rate of 416 Hz and up-sampled its output to 500 Hz. We did not further filter this data, following established best practices in deep learning (Goodfellow et al., 2016; He et al., 2016; Krizhevsky and Sutskever, n.d.; Zhao et al., 2019). The accelerometer's range was +/-16 G and the gyroscope's range was +/-2000 degrees/second. Fig. 1a illustrates the coordinate system of the accelerometer of the insole. These sensors have been validated for their accuracy in measuring the kinematics of the foot and have been used in other studies (Gaiesky et al., 2023; Mitchell, 2023; Napier et al., 2022; Napier et al., 2021; Slattery et al., 2022).

2.2. Experimental Design

Prior to data collection, participants completed an eligibility questionnaire and intake form, including previous injury history and 10 km personal best time. We used the latter to quantify runners' performance. Runners then warmed up for 5 min to become familiar with the treadmill (NordicTrack C700, NordicTrack, Logan, USA). Participants wore two insole-embedded IMUs in standardized neutral cushioned footwear (New Balance 880, New Balance, Boston, USA) and ran for one minute each at three fixed speeds (2.5, 3.0, and 3.5 m/s) in a randomized order (Fig. 1b). Participants then ran ~90 m overground in one direction, turned around, and ran the same distance in the opposite direction at their preferred speed.

2.3. Data Preparation

To ensure steady state running on the treadmill, we cut five seconds at the beginning and end of the one-minute trial. Similarly in overground running, we excluded speed changes in the beginning, during turn-around, and at the end of the running phase. To identify steady-state, we low-pass filtered the raw data with a 4th-order Butterworth filter with a cutoff frequency of 1 Hz. We then detected steady-state running when the filtered data no longer increased over a one-second period. We then split the data into data windows of 20 s or 10,000 data points to capture over 95 % of the variability in the data (Riazati et al, 2019; Belli et al., 1994). To augment our dataset, we use a sliding windows approach, where each new window shared 9,000 data points with the previous one (Dehghani et al., 2019). This method effectively increases the number of data samples available for analysis or training models.

We then transformed the linear acceleration and angular velocity data from each window into spectrograms (Fig. 1c–d). Using spectrograms instead of raw acceleration and angular velocity data has three main advantages. First, image recognition algorithms are computationally more efficient than methods that process data point-by-point over time, such as Recurrent Neural Networks (RNNs) (Gehring et al., 2017). Second, using spectrograms enabled us to use well-established, gold standard image recognition algorithms that scientists have already successfully applied in many different areas (Alnuaim et al., 2022; Koh et al., 2019; Tripathi et al., 2019). And lastly, using spectrograms allows us to leverage novel algorithms that can interpret the decision-making of an image recognition neural network, creating possibilities for better interpretation of the results (Kim et al., 2021; Selvaraju et al., 2017; Yoo and Jeong, 2022). For each window, we created

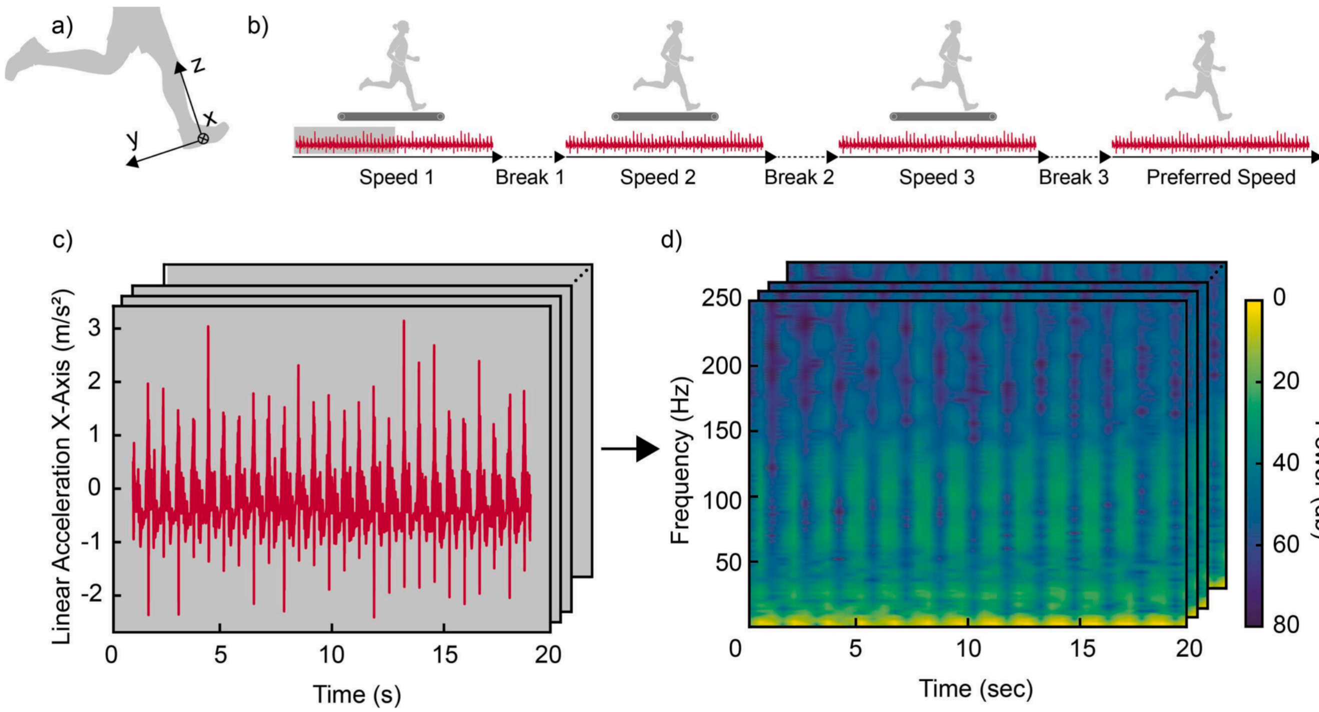

Fig. 1. A) illustrates the linear acceleration orientations of the sensor in the shoe. b) illustrates the order in which we conducted the experiment. participants first ran with three different speeds (randomized order) on the treadmill, followed by their preferred speed on overground. between every running trial, participants took a break. the red line illustrates an example measurement from the imu sensors, and the grey box at speed 1 illustrates an example of a 20 s data window. c) illustrates 20 s of data from 12 channels (3-axis linear acceleration x 3-axis angular velocity x 2 feet), also highlighted as the grey box in b, and d) illustrates the subsequent spectrograms for each channel. (For interpretation of the references to colour in this figure legend, the reader is referred to the web version of this article.)

one spectrogram image per sensor using the spectrogram function from the signal library in python (Python Software Foundation), with a segment length of 250 samples (0.5 s) and no overlap. In pilot tests, these settings had the best performance when compared to longer and shorter segment lengths and different overlap lengths. Each spectrogram consisted of 126 frequencies on the y-axis and 40 time data points on the x-axis. To decrease the difference between high and low frequencies, we converted the power spectrogram into decibel units using the *power_to_decibel* function from the librosa library. This step made the spectrogram visually more interpretable, but did not have an affect on the actual data analysis. Each input for the neural network comprised twelve spectrogram images: six representing the three acceleration axes per foot and six representing the three angular velocity axes per foot. In the following, we explain how the image recognition algorithm uses the spectrograms as its input, to either predict the individual runner or their running performance as the output.

2.4. Data analysis

We first optimized our model to predict individual runners. Here, we used the spectrograms as an input to the model, and the individual runner ID as an output of the model. To learn the running form of each individual runner in this study, the model used 100 % (119/119 runners) of the treadmill data for training, and 80 % (95/119 runners) of the overground data for validation. We left out 20 % (24/119 runners) of the overground data for testing the accuracy of the trained model (Fig. 2a). This methodology ensured that during testing, the model identified individual runners using overground data it had never encountered during training. Due to computational reasons, we did not do k-fold cross-validation. During training, we evaluated the model's accuracy after

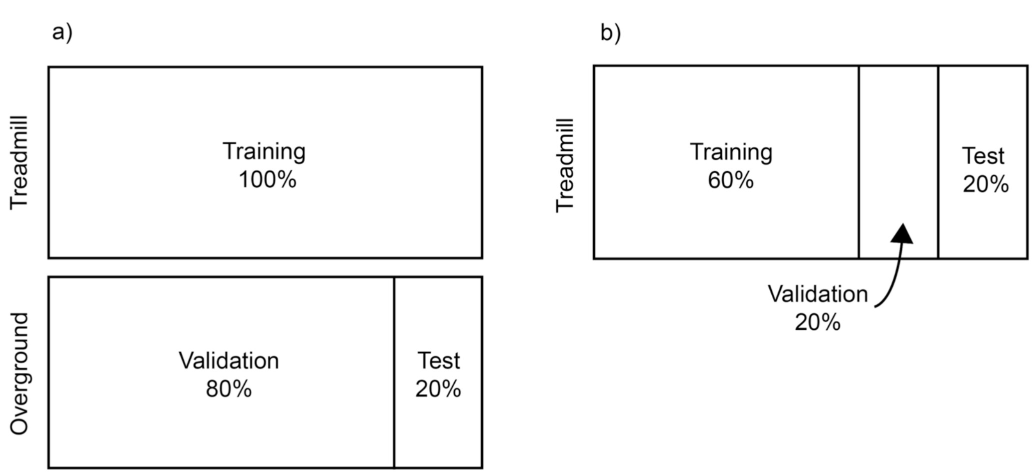

Fig. 2. a) illustrates the data separation for predicting individual runners and b) illustrates the data separation for predicting running speed.

each epoch using the separate validation dataset. During testing, the model had to correctly predict each of these 24 individual runners out of 119 possible runners. We also tested the model on the runners that completed the experiment two or three times, using the same ratios between training, validation, and test set. To quantify the performance of the model during training, validation, and testing, the model calculated accuracy, which is the ratio between the number of correctly identified runners and the total number of runners. Being able to accurately identify individual runners would demonstrate that the neural network can effectively learn meaningful patterns from this dataset, confirming that the input data contains sufficient individual-specific information to potentially predict performance.

For our second aim, we tested how the number of runners and the amount of training data per runner affected the accuracy of the model predictions. To do so, we trained the model with treadmill data from one session from 10 runners and used the overground data for validation. We did this test with an increasing number of runners, in increments of 10 [10, 20, 30, …, and 119]. We then repeated the test but reduced the amount of training data from each participant to 67 %. For both tests, we used overground data for validation, and did not reserve a separate test set. This allowed us to still have an acceptable amount of data to test the model's accuracy when evaluating a lower number of runners. For example for 10 runners, using 80 % of the overground data for validation and 20 % of the overground data for testing would only leave overground data from 2 runners for the final test set, risking non-representative results. Our solution for this was to use 100 % of overground data (10 runners in this example) for validation, and use the validation result as the final metric in this test.

We then optimized the model to predict individual 10 km personal best times (model output) from the spectrograms (model input). We used treadmill data from 60 % of the runners for training (71/119) and 20 % for validation (24/119). We left out treadmill data from 20 % of the runners (24/119) for testing the accuracy of the model (Fig. 2b). To prevent the model from deriving a runner's 10 km personal best time from their preferred overground speed, we excluded the overground data from this part of the analysis. We trained and tested the model to predict the runners' 10 km personal best time in seconds, and quantified the model's performance during training, validation, and testing by calculating the root mean square error between the predicted and the actual 10 km personal best time. Here, we also calculated the mean 10 km personal best time of all runners in the training dataset, and used this number as a benchmark prediction for every runner in the test set, and compared the result to the performance of our model's predictions. To simplify the task, we also created a fast and slow group, with the median 10 km personal best time separating both groups into equal sizes, and tested the model's prediction performance in separating fast from slow runners. Here the spectrograms were the input to the model and the group ID (slow or fast) was the output of the model. To quantify the model's performance during training, validation, and testing, we calculated prediction accuracy, which is the ratio between the number of runners with a correctly predicted group (fast/slow) and the total number of runners.

Before training and testing the models, we chose an image recognition algorithm and optimized the hyperparameters. We used ResNet50V2, a Convolutional Neural Network (CNN), which is freely available in the keras library. ResNet50V2 is composed of 50 layers and designed to solve the vanishing gradient problem through residual learning. To best understand the complex architecture of this model, it is best to read the original paper (He et al., 2016). This architecture enables deep networks to train effectively and significantly improves performance in tasks like image recognition and spectrogram classification across various fields (Chao et al., 2023; Kim et al., 2021; Koh et al., 2019; Setiawan and Lin, 2021; Zhang et al., 2021). The network processes input data hierarchically. Earlier layers extract low-level features, such as edges and textures, while subsequent layers integrate these into progressively complex, domain-specific patterns. This hierarchical processing is supported by shortcut connections, which efficiently propagate critical information across layers, mitigating training challenges associated with vanishing gradients. In its final dense layers, the network consolidates these refined representations into precise classification outputs, enabling robust generalization across a variety of tasks. We did all the calculations in Google Colab (Alphabet Inc., Mountain View, USA). Based on pilot analysis, we used a batch size of 32, did 5,000 epochs, used the Adam optimization algorithm with a learning rate of 10^5, and automatically re-initialized the weights if the performance of the validation set did not increase for more than 1,000 epochs. To test the algorithm's final performance, we used the model with the best accuracy on the validation set.

We used ResNet50V2, a Convolutional Neural Network (CNN), which is freely available in the keras library. ResNet50V2 is composed of 50 layers, designed to solve the vanishing gradient problem through residual learning. To best understand the complex architecture of this model, it is best to read the original research paper (He et al., 2016). This architecture enables deep networks to train effectively and significantly improves performance in tasks like image recognition and spectrogram classification across various fields (Chao et al., 2023; Kim et al., 2021; Koh et al., 2019; Setiawan and Lin, 2021; Zhang et al., 2021). The network processes input data hierarchically. Earlier layers extract low-level features, such as edges and textures, while subsequent layers integrate these into progressively complex, domain-specific patterns. This hierarchical processing is supported by shortcut connections, which efficiently propagate critical information across layers, mitigating training challenges associated with vanishing gradients. In its final dense layers, the network consolidates these refined representations into precise classification outputs, enabling robust generalization across a variety of tasks.

All our code is publicly available on GitHub (Mayerhofer, 2024). To get access to the data, please reach out to the authors.

3. Results

The model accurately identified individual runners from their running form. When training and testing the model with all runners that had data from at least one experimental session, the model had a test accuracy of 84 %. To arrive at this single session value, we used the treadmill data from all 119 runners to train the model, overground data from 95 of these 119 runners for validation during training, and the remaining 24 runners' overground data for testing. It is for these 24 test runners that the model has never had access to their overground data before, and assigned the overground data to the correct runner, while having to choose from all 119 possible runners. The test accuracy for the models that trained on data from 2 or 3 sessions was 89 % and 95 %, respectively. For these tests, the models had more data from each runner to train on and fewer runners to distinguish between. Table 1 presents the accuracy of these different tests.

For our second aim, we tested how the amount of available training data per runner affects the model's accuracy, and found that the accuracy of the model decreased with a greater number of runners and less data per runner. We designed this second aim to test the feasibility of using running forms as digital signatures for identification or verification of individual athletes. Fig. 3 illustrates the validation accuracies with an increasing number of participants for 100 % and 67 % of data per runner, respectively. Averaging across all calculations, the

Table 1
The number of total runners, test runners, and the test accuracy for all runners that completed one, two, or three sessions.

	1 session	2 sessions	3 sessions
Total runners	119	37	24
Test runners	24	7	5
Test accuracy	84 %	89 %	95 %

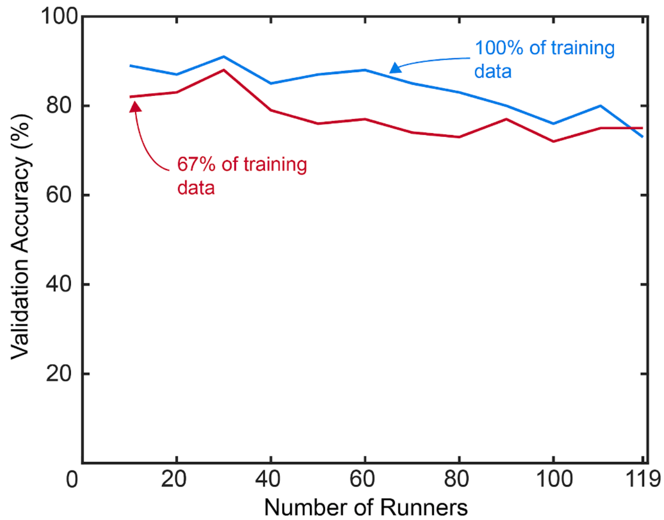

Fig. 3. The blue line illustrates the validation accuracy with an increase in the number of runners when using 100% of the runners' data for training. The red line illustrates the validation accuracy with an increase in the number of runners when using 67% of the runners' data for training. (For interpretation of the references to colour in this figure legend, the reader is referred to the web version of this article.)

validation accuracy with 67 % of the training data was 6.1 $\pm$ 4.1 % lower than with 100 % of the training data (i.e., the average of the red line versus the average of the blue line). Here, we used all overground data for validation, and did not reserve any test data throughout the experiment. Specifically for a lower number of runners (i.e, 10 runners), this allowed us to still have an acceptable amount of data on which to test the model's accuracy.

The model did not accurately predict running performance from individual running form. When predicting an individual runner's 10 km personal best time, the model's root mean square error of the test set was 473 s (~8 min). To arrive at this value, we trained the model using treadmill data from 71 runners, validated it during training with data from additional 24 runners, and tested it with data from another 24 runners. To evaluate whether the model could predict 10 km personal best times from individual running form, the model was never exposed to any data from the test runners during training or validation. For comparison, we also calculated the mean 10 km personal best time of all runners in the training dataset (2759 s or ~46 min) and used this number as a benchmark prediction for every runner in the test dataset, which resulted in a root mean square error of 486 s (~8 min). With a root mean square error of 473 s, the neural network was not much better than simply using the mean of all the training runners' 10 km personal best times as the same prediction for each individual runner. To test whether this approach would perform better on a simpler task, we created a fast and slow group, with the median 10 km personal best time separating both groups into equal sizes. We then tested the model's accuracy in predicting fast and slow runners. Here, the model was not able to accurately separate fast from slow runners based on their running forms, with a prediction accuracy of 57 % (only 7 % better than a coin toss).

4. Discussion

While our neural network can accurately identify individual runners from their foot kinematics, it fails to predict running performance. We collected foot kinematics from insole-embedded IMUs during both treadmill and overground running. We transformed acceleration and angular velocity time series running data into spectrograms and trained a state-of-the-art image recognition neural network to identify individual runners from their running data. To ensure the neural network's predictive ability generalized to new environments, we trained the model with treadmill data, and tested its accuracy with overground data. We then trained the same neural network to predict the runners' performance based on their personal best 10 km run time. With enough data, the neural networks were successful in identifying the individual runner, but failed to predict their performance, even when we simplified the task to only separate fast from slow runners.

Our finding that running form does not predict running performance is consistent with some studies, but inconsistent with others. Agresta et al. found no association between years of running experience and trunk and lower extremity kinematics and kinetics (Agresta et al., 2018). Patoz et al. found no association between running economy and whole body kinematics (Patoz et al., 2022a). In contrast, Clermont et al. were able to accurately classify recreational and competitive runners from their running form, using an IMU at the lower back (Clermont et al., 2019a) and lower limb 3D motion analysis (Clermont et al., 2017). While there was not enough information about running performance in the foot kinematics data that we collected, the studies of Clermont and colleagues suggest that the movements of other body parts contain additional relevant information about performance.

Our results are limited to IMU-measured foot kinematics, which do not provide a complete biomechanical assessment. Specifically, a single IMU embedded in the shoe insole does not capture all aspects of foot movement. There is much more kinetic or kinematic data that one could collect in a laboratory setting. For example, force plates could collect impact forces or 3D motion analysis could provide a clearer picture of the movement of different parts of the foot. We focused on IMUs embedded in insoles due to their commercial popularity and ability to capture real-world movement patterns, and because this analysis is part of another study that only involved foot kinematic data collection (Grand View Research, 2023; Napier et al., 2022; Parnell, 2023) Thus, incorporating sensors at other body locations, either independently or in combination with foot sensors, could enhance the classification of running performance, but this must be balanced against the need for commercial practicality.

Choosing an image recognition neural network to detect running performance has both advantages and disadvantages. Neural networks, unlike more conventional machine learning algorithms such as Support Vector Machines, can learn directly from complex, raw datasets without feature extraction. They also adapt well to new data, making them promising for detecting running performance from foot kinematics (Barshan and Yüksek, 2014; Cust et al., 2019; Halilaj et al., 2018; Hoitz et al., 2021; Moghadam et al., 2023; Ngoh et al., 2018; Nigg et al., 2012; Phinyomark et al., 2018; Tabrizi et al., 2020). However, due to the use of raw data instead of prepared features, neural networks are harder to interpret in their decision making process compared to traditional methods. Neural networks also require larger datasets to perform optimally, whereas more conventional methods often work well with smaller datasets due to manual feature extraction. To ensure sufficient high-quality data for our neural network, we took several steps. First, we selected a sensor that accurately measures foot kinematics (Gaiesky et al., 2023; Mitchell, 2023; Napier et al., 2022, 2021; Slattery et al., 2022). Second, we collected data from 119 runners, exceeding typical sample sizes in biomechanical running studies (Agresta et al., 2018; Clermont et al., 2019a, 2017; Patoz et al., 2022a; Weich et al., 2020). Finally, to also confirm the presence of enough valuable information about running form in our dataset, we successfully trained the model to identify individual runners based on their running form. By combining the neural network with spectrogram-based analysis, a large dataset, and high-quality sensors, we have demonstrated its strengths and weaknesses in identifying individual runners, while also highlighting the challenges it faces in predicting performance.

We found that an individual's running form translates from treadmill running to overground running at changing speeds. Most researchers in this area have either trained and tested their algorithm with treadmill data or trained and tested their algorithm with overground data (Hoitz et al., 2021; Weich et al., 2020). Their prediction accuracies are > 95 %. Only Horst et al. switched between multiple different shoes and found that their algorithm could predict the individual runner with an accuracy of 80 % (Horst et al., 2023). This is referred to as domain shift or dataset shift, and it shows how generalizable an algorithm is to changing environments (Quinonero-Candela et al., 2008). In our study, we trained the neural network on treadmill data with three fixed speeds, and tested its accuracy on overground data with the runners' preferred speed. That is, the neural network had to identify runners on a different surface and with different speeds than what it trained on. Our model was able to accurately identify runners in different running environments (accuracies between 84–95 %) and adds to the findings of Horst et al. where they successfully identified runners with different footwear (Horst et al., 2023).

While we and others have used running form to accurately identify an individual from within a small pool of runners, we suspect that this accuracy may not readily scale with increasing pool size. Past studies could accurately identify individual running forms (Hoitz et al., 2021; Horst et al., 2023; Weich et al., 2020). Here, we had a bigger sample than these earlier studies and found that, as we increased the number of runners in the analysis, the identification accuracy decreased, because the network had to choose among a larger number of alternatives (Fig. 3). These results suggest that if we keep increasing the number of different runners, the accuracy of the neural network in identifying individual runners would eventually be indistinguishable from 0. It is unlikely, for example, that one could accurately identify an individual runner from within the pool of ~25,000 runners that complete the Boston Marathon each year. Decreasing the amount of training data for each runner by $^1/_3$ (i.e., using 67 % of their available training data) had a similar effect to increasing the pool size. This suggests that by increasing the amount of training data per runner, we could simultaneously increase the pool size from within which an individual runner could be identified accurately. For example, 10 participants with 67 % of the available training data had a similar validation accuracy as 80 participants with all available training data (Fig. 3). It's well established that increasing the amount of data per class increases the overall classification accuracy of a model (Çolak, 2021; Lei et al., 2019; Rajput et al., 2023; Ramezan et al., 2021). Our study adds to this knowledge as it helps understand the effects of increasing the number of classes while keeping the amount of data the same.

Alternatively, rather than identify an individual, our results suggest that it may be more effective to use running form to verify an individual's identity. This verification is like how signatures are used—an individual's measured running form is compared against a previously provided template running form. Verification is a simpler task than identification. That is, the question "Are you who you say you are?" is easier to answer than "Who are you?". This is because the algorithm verifies an individual's data against their own data rather than trying to identify who an individual is from within an existing pool of data (Ali et al., 2016; Crosswhite et al., 2018). Being able to verify runners by measuring their running form will be useful for virtual sport companies and events, to minimize fraud. But, using running form to verify identity would not work if running form was inconsistent, varying within an individual across running conditions. Our results suggest that this may not be an issue because there is some inherent part of individual running form that generalizes across running speeds, and between treadmill and overground running. Using running form as a digital signature may prove useful in virtual sports environments to verify competitor identity.

Ethics Information

The Office of Research Ethics at Simon Fraser University approved the study and all participants provided written and verbal informed consent before participating in our study.

Funding Details

This work was funded by an NSERC Discovery Grant to JMD. This work was supported by Mitacs [IT24201]. CN is supported by Michael Smith Health Research BC [HPI-2020-0719].

CRediT authorship contribution statement

Patrick Mayerhofer: Writing – review & editing, Writing – original draft, Visualization, Validation, Software, Project administration, Methodology, Formal analysis, Conceptualization. **Christopher Napier:** Writing – review & editing, Supervision, Methodology, Formal analysis, Data curation, Conceptualization. **David C. Clarke:** Writing – review & editing, Formal analysis, Conceptualization. **Ivan Bajić:** Writing – review & editing, Methodology, Formal analysis, Conceptualization. **J. Maxwell Donelan:** Writing – review & editing, Supervision, Project administration, Methodology, Funding acquisition, Conceptualization.

Declaration of competing interest

The authors declare that they have no known competing financial interests or personal relationships that could have appeared to influence the work reported in this paper.

References

Agresta, C.E., Peacock, J., Housner, J., Zernicke, R.F., Zendler, J.D., 2018. Experience does not influence injury-related joint kinematics and kinetics in distance runners. Gait Posture 61, 13–18.

Ali, M.M.H., Mahale, V.H., Yannawar, P., Gaikwad, A.T., 2016. Fingerprint Recognition for Person Identification and Verification Based on Minutiae Matching. In: 2016 IEEE 6th International Conference on Advanced Computing (IACC). IEEE, pp. 332–339.

Alnuaim, A.A., Zakariah, M., Shashidhar, C., Hatamleh, W.A., Tarazi, H., Shukla, P.K., Ratna, R., 2022. Speaker gender recognition based on deep neural networks and ResNet50. Proc. Int. Wirel. Commun. Mob. Comput. Conf. https://doi.org/10.1155/2022/4444388.

Barshan, B., Yüksek, M.C., 2014. Recognizing daily and sports activities in two open source machine learning environments using body-worn sensor units. Comput. J. 57, 1649–1667.

Benson, L.C., Clermont, C.A., Ferber, R., 2020. New considerations for collecting biomechanical data using wearable sensors: the effect of different running environments. Front. Bioeng. Biotechnol. 8, 86.

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. J. Biomech. 71, 94–99.

Chao, Q., Wei, X., Tao, J., Liu, C., Wang, Y., 2023. Cavitation recognition of axial piston pumps in noisy environment based on Grad-CAM visualization technique. CAAI Trans. Intell. Technol. 8, 206–218.

Clermont, C.A., Benson, L.C., Osis, S.T., Kobsar, D., Ferber, R., 2019a. Running patterns for male and female competitive and recreational runners based on accelerometer data. J. Sports Sci. 37, 204–211.

Clermont, C.A., Osis, S.T., Phinyomark, A., Ferber, R., 2017. Kinematic gait patterns in competitive and recreational runners. J. Appl. Biomech. 33, 268–276.

Clermont, C.A., Phinyomark, A., Osis, S.T., Ferber, R., 2019b. Classification of higher- and lower-mileage runners based on running kinematics. J. Sport Health Sci. 8, 249–257.

Cochrum, R.G., Conners, R.T., Caputo, J.L., Coons, J.M., Fuller, D.K., Frame, M.C., Morgan, D.W., 2021. Visual classification of running economy by distance running coaches. EJSS 21, 1111–1118.

Çolak, A.B., 2021. An experimental study on the comparative analysis of the effect of the number of data on the error rates of artificial neural networks. Int. J. Energy Res. 45, 478–500.

Crosswhite, N., Byrne, J., Stauffer, C., Parkhi, O., Cao, Q., Zisserman, A., 2018. Template adaptation for face verification and identification. Image Vis. Comput. 79, 35–48.

Cunningham, R., Hunter, I., Seeley, M.K., Feland, B., 2013. Variations in running technique between female sprinters, middle, and distance runners. Int. J. Exerc. Sci. 6, 6.

Cust, E.E., Sweeting, A.J., Ball, K., Robertson, S., 2019. Machine and deep learning for sport-specific movement recognition: a systematic review of model development and performance. J. Sports Sci. 37. https://doi.org/10.1080/02640414.2018.1521769.

Dehghani, A., Sarbishei, O., Glatard, T., Shihab, E., 2019. A quantitative comparison of overlapping and non-overlapping sliding windows for human activity recognition using inertial sensors. Sensors 19. https://doi.org/10.3390/s19225026.

Eskofier, B.M., Kraus, M., Worobets, J.T., Stefanyshyn, D.J., Nigg, B.M., 2012. Pattern classification of kinematic and kinetic running data to distinguish gender, shod/barefoot and injury groups with feature ranking. Comput. Methods Biomech. Biomed. Eng. 15, 467–474.

Fukuchi, R.K., Eskofier, B.M., Duarte, M., Ferber, R., 2011. Support vector machines for detecting age-related changes in running kinematics. J. Biomech. 44, 540–542.

Gaiesky, S.K.T., Fridman, L., Michie, T., Blazey, P., Tran, N., Schneeberg, A., Napier, C., 2023. The one-week and three-month reliability of acceleration outcomes from an insole-embedded inertial measurement unit during treadmill running. Sports Biomech. 1–15.

Gehring, J., Auli, M., Grangier, D., Yarats, D., Dauphin, Y., 2017. Convolutional sequence to sequence learning. ICML abs/1705.03122, 1243–1252.

Goodfellow, I., Bengio, Y., Courville, A., Bengio, Y., 2016. Deep learning. MIT press Cambridge.

Grand View Research, 2023. Smart Shoes Market Size, Share & Trends Analysis Report.

Halilaj, E., Rajagopal, A., Fiterau, M., Hicks, J.L., Hastie, T.J., Delp, S.L., 2018. Machine learning in human movement biomechanics: Best practices, common pitfalls, and new opportunities. J. Biomech. 81, 1–11.

He, K., Zhang, X., Ren, S., Sun, J., 2016. Deep residual learning for image recognition. In: 2016 IEEE Conference on Computer Vision and Pattern Recognition (CVPR). Presented at the 2016 IEEE Conference on Computer Vision and Pattern Recognition (CVPR). IEEE, pp. 770–778.

Hoitz, F., von Tscharner, V., Baltich, J., Nigg, B.M., 2021. Individuality decoded by running patterns: movement characteristics that determine the uniqueness of human running. PLoS One 16, e0249657.

Horst, F., Hoitz, F., Slijepcevic, D., Schons, N., Beckmann, H., Nigg, B.M., Schöllhorn, W. I., 2023. Identification of subject-specific responses to footwear during running. Sci. Rep. 13, 11284.

Kim, J., Oh, J., Heo, T.-Y., 2021. Acoustic scene classification and visualization of beehive sounds using machine learning algorithms and Grad-CAM. Math. Probl. Eng. 2021. https://doi.org/10.1155/2021/5594498.

Koh, C.-Y., Chang, J.-Y., Tai, C.-L., Huang, D.-Y., Hsieh, H.-H., Liu, Y.-W., 2019. Bird Sound Classification Using Convolutional Neural Networks, in: CLEF (Working Notes). ceur-ws.org.

Krizhevsky, A., Sutskever, I., n.d. Imagenet classification with deep convolutional neural networks.

Lei, S., Zhang, H., Wang, K., Su, Z., 2019. How training data affect the accuracy and robustness of neural networks for image classification.

Mayerhofer, P., 2024. Neural networks can accurately identify individual runners from their foot kinematics, but fail to predict their running performance - GitHub Repository [WWW Document]. GitHub. URL https://github.com/patmorli/running_project_jBiom (accessed .03.2).

Mero, A., Komi, P.V., Gregor, R.J., 1992. Biomechanics of sprint running. A review. Sports Med. 13, 376–392.

Mitchell, C., 2023. Using Inertial Foot Pods to Develop a Gait Protocol to Assist with Concussion Diagnosis and Monitoring (Master of Sport, Exercise and Health). Auckland University of Technology.

Moghadam, S.M., Yeung, T., Choisne, J., 2023. A comparison of machine learning models' accuracy in predicting lower-limb joints' kinematics, kinetics, and muscle forces from wearable sensors. Sci. Rep. 13, 5046.

Moore, I.S., 2016. Is there an economical running technique? A review of modifiable biomechanical factors affecting running economy. Sports Med. 46, 793–807.

Napier, C., Fridman, L., Blazey, P., Tran, N., Michie, T.V., Schneeberg, A., 2022. Differences in peak impact accelerations among foot strike patterns in recreational runners. Front Sports Act Living 4, 802019.

Napier, C., Willy, R.W., Hannigan, B.C., McCann, R., Menon, C., 2021. The effect of footwear, running speed, and location on the validity of two commercially available inertial measurement units during running. Front Sports Act Living 3, 643385.

Ngoh, K.-J.-H., Gouwanda, D., Gopalai, A.A., Chong, Y.Z., 2018. Estimation of vertical ground reaction force during running using neural network model and uniaxial accelerometer. J. Biomech. 76, 269–273.

Nigg, B.M., Baltich, J., Maurer, C., Federolf, P., 2012. Shoe midsole hardness, sex and age effects on lower extremity kinematics during running. J. Biomech. 45, 1692–1697.

Nweke, H.F., Teh, Y.W., Al-garadi, M.A., Alo, U.R., 2018. Deep learning algorithms for human activity recognition using mobile and wearable sensor networks: State of the art and research challenges. Expert Syst. Appl. 105, 233–261.

Parnell, K., 2023. How Footpods Can Take Your Running to the Next Level [WWW Document]. trinewbies.com. URL https://trinewbies.com/how-footpods-can-take-your-running-to-the-next-level/#:~:text=Running%20footpods%20can%20help%20triathletes,race%20or%20a%20training%20run. (accessed 9.18.24).

Patoz, A., Lussiana, T., Breine, B., Gindre, C., Hébert-Losier, K., 2022a. There is no global running pattern more economic than another at endurance running speeds. Int. J. Sports Physiol. Perform. 17, 659–662.

Patoz, A., Lussiana, T., Breine, B., Gindre, C., Malatesta, D., Hébert-Losier, K., 2022b. Examination of running pattern consistency across speeds. Sports Biomech. 1–15.

Phinyomark, A., Hettinga, B.A., Osis, S.T., Ferber, R., 2014. Gender and age-related differences in bilateral lower extremity mechanics during treadmill running. PLoS One 9, e105246.

Phinyomark, A., Petri, G., Ibáñez-Marcelo, E., Osis, S.T., Ferber, R., 2018. Analysis of big data in gait biomechanics: current trends and future directions. J. Med. Biol. Eng. 38, 244–260.

Quinonero-Candela, J., Sugiyama, M., Schwaighofer, A., Lawrence, N.D., 2008. Dataset shift in machine learning. MIT Press.

Rajput, D., Wang, W.-J., Chen, C.-C., 2023. Evaluation of a decided sample size in machine learning applications. BMC Bioinf. 24, 48.

Ramezan, C.A., Warner, T.A., Maxwell, A.E., Price, B.S., 2021. Effects of training set size on supervised machine-learning land-cover classification of large-area high-resolution remotely sensed data. Remote Sens. (Basel) 13, 368.

Selvaraju, R.R., Cogswell, M., Das, A., Vedantam, R., Parikh, D. and Batra, D., 2017. Grad-cam: Visual explanations from deep networks via gradient-based localization. In Proceedings of the IEEE international conference on computer vision (pp. 618-626).

Setiawan, F., Lin, C.-W., 2021. Identification of neurodegenerative diseases based on vertical ground reaction force classification using time–frequency spectrogram and deep learning neural network features. Brain Sci. 11, 902.

Slattery, P., Eduardo Cofré Lizama, L., Wheat, J., Gastin, P., Dascombe, B., Middleton, K., 2022. The agreement between wearable sensors and force plates for the analysis of stride time. Presented at the ISEA 2022 – The Engineering of Sport 14, Purdue University.

Tabrizi, S.S., Pashazadeh, S., Javani, V., 2020. Comparative study of table tennis forehand strokes classification using deep learning and SVM. IEEE Sens. J. 20, 13552–13561.

Tripathi, S., Kumar, A., Ramesh, A., Singh, C., Yenigalla, P., 2019. Focal Loss based Residual Convolutional Neural Network for Speech Emotion Recognition. arXiv [eess.AS].

Weich, C., Vieten, M., 2020. The gaitprint: identifying individuals by their running style. Sensors 20. https://doi.org/10.3390/s20143810.

Yoo, Y., Jeong, S., 2022. Vibration analysis process based on spectrogram using gradient class activation map with selection process of CNN model and feature layer. Displays 73, 102233.

Zhang, T., Feng, G., Liang, J., An, T., 2021. Acoustic scene classification based on Mel spectrogram decomposition and model merging. Appl. Acoust. 182, 108258.

Zhao, R., Yan, R., Chen, Z., Mao, K., Wang, P., Gao, R.X., 2019. Deep learning and its applications to machine health monitoring. Mech. Syst. Sig. Process. 115, 213–237.