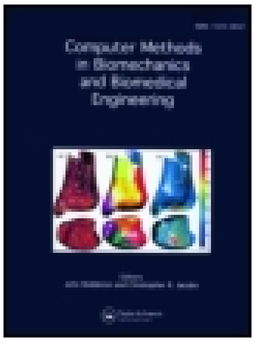

Taylor & Francis
Taylor & Francis Group

Computer Methods in Biomechanics and Biomedical Engineering

ISSN: (Print) (Online) Journal homepage: https://www.tandfonline.com/loi/gcmb20

Deep learning approach for prediction of impact peak appearance at ground reaction force signal of running activity

Anastasiia Girka , Juha-Pekka Kulmala & Sami Äyrämö

To cite this article: Anastasiia Girka , Juha-Pekka Kulmala & Sami Äyrämö (2020): Deep learning approach for prediction of impact peak appearance at ground reaction force signal of running activity, Computer Methods in Biomechanics and Biomedical Engineering, DOI: 10.1080/10255842.2020.1786072

To link to this article: https://doi.org/10.1080/10255842.2020.1786072

Published online: 09 Jul 2020.

COMPUTER METHODS IN BIOMECHANICS AND BIOMEDICAL ENGINEERING
https://doi.org/10.1080/10255842.2020.1786072

Deep learning approach for prediction of impact peak appearance at ground reaction force signal of running activity

Anastasiia Girka[a], Juha-Pekka Kulmala[b] and Sami Äyrämö[a]

[a]Faculty of Information Technology, University of Jyvaskyla, Jyvaskyla, Finland; [b]Motion Analysis Laboratory, Children's Hospital, University of Helsinki and Helsinki University Hospital, Helsinki, Finland

ABSTRACT

Protruding impact peak is one of the features of vertical ground reaction force (GRF) that is related to injury risk while running. The present research is dedicated to predicting GRF impact peak appearance by setting a binary classification problem. Kinematic data, namely a number of raw signals in the sagittal plane, collected by the Vicon motion capture system (Oxford Metrics Group, UK) were employed as predictors. Therefore, the input data for the predictive model are presented as a multi-channel time series. Deep learning techniques, namely five convolutional neural network (CNN) models were applied to the binary classification analysis, based on a Multi-Layer Perceptron (MLP) classifier, support vector machine (SVM), logistic regression, *k*-nearest neighbors (kNN), and random forest algorithms. SVM, logistic regression, and random forest classifiers demonstrated performances that do not statistically significantly differ. The best classification accuracy achieved is 81.09% ± 2.58%. Due to good performance of the models, this study serves as groundwork for further application of deep learning approaches to predicting kinetic information based on this kind of input data.

ARTICLE HISTORY
Received 23 May 2018
Accepted 18 June 2020

KEYWORDS
CNN; binary classification; running gait analysis; risk assessment; force platform

1. Introduction

Every second runner suffers from injuries on a yearly basis (Van Mechelen 1992; Taunton et al. 2002; Ristolainen et al. 2009). Since recreational running has obvious health benefits, it is important to avoid negative effects, such as injuries. The risk of injuries can be assessed by running gait analysis. Nowadays that implies 3D motion analysis along with utilization of force platforms including analysis of ground reaction force (GRF) signal over time, among other measured signals. GRF provides crucial information on a lower limb's loading. In particular, a protruding impact peak (Figure 1) is one of the features of vertical GRF that is indirectly related to injury risk while running. Protruding impact peak is often related to heel striking, which gained a lot of interest in research on running-related sport injuries (Kulmala et al. 2013; Knorz et al. 2017).

The motivation for this research is that GRF is measured by force platforms, which makes conducting GRF measurements inconvenient. Firstly, the force platforms are relatively expensive. Usually, there are a limited number of them; thus, it is possible to collect data only for a limited number of running cycles (strides). Secondly, experiments with force platforms require a solid platform, where force platforms can be mounted into the floor. In other words, it requires a restricted laboratory environment compared to measurements with only a motion capture system that can be conducted without binding to any specific place. Thus, it would be much easier to organize and cheaper to collect the signals with only a motion capture system while a subject is running on a treadmill and predict kinetic variables of interest by a machine-learning model. Also, a cheaper motion capture system with lower sampling frequency could be utilized in that case. Although integrated treadmills with force plates enable collecting a large number of strides, they are also expensive and limited to a laboratory environment. In the real world environment, determination of vertical GRF without direct measurement can be done via video or other motion capture techniques. Therefore, by predicting GRF from raw kinematics signals, we exclude force platforms from the measurements and make having GRF

CONTACT Anastasiia Girka angirka@jyu.fi; anastasiia.v.girka@student.jyu.fi

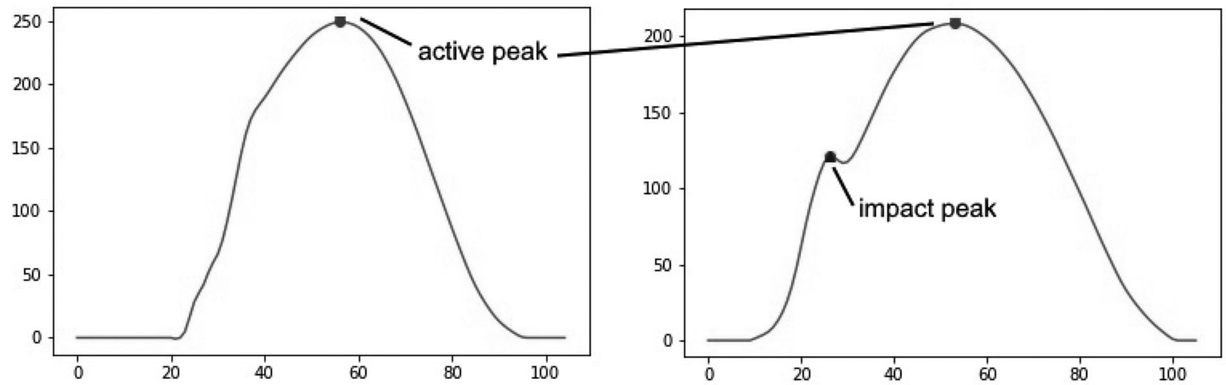

Figure 1. Vertical component of a GRF signal. Ordinate corresponds to newtons normalized by weight of the runner. The active peak corresponds to the value of 250%, where 100% means the body weight of the runner. Abscissa corresponds to frames in a way that there is one data point per frame, 300 frames per second. On the left plot GRF peak without protruding impact peak is presented. On the right plot is GRF peak with protruding impact peak.

data possible whenever kinematics signals from a motion capture system are available.

The objective of this study was to predict an appearance of an impact peak on a vertical GRF signal using raw kinematics time series, since impact peak is considered a key biomechanical feature related to the risk of injury while running (van der Worp et al., 2016; Davis et al., 2016).

The novelty of this research consists of employing raw time series signals alternatively to hand-crafted features and thus applying a deep learning approach to predicting GRF information for running gait analysis.

This paper is organized as follows: Section 2 is devoted to related research and state of the art; Section 3 discusses data collection and preprocessing; Section 4 describes methods and includes a description of five deep learning models (i.e., five convolutional neural networks (CNNs) with identical convolutional parts, but different classifier parts: one with fully-connected layers as a classier or MLP, in other words, then with SVM, logistic regression, kNN, and random forest as classifiers); Section 5 describes results; Section 6 provides the conclusion and future work suggestions.

2. Related research

Among related studies there are several dedicated to a prediction of GRF using kinematic data and performing inverse dynamics-based simulations, for example, Fluit et al. (2014). Some of the research assumes utilizing additional devices, as in Jung et al. (2016). Their estimation of the vertical GRF from body kinematics was based on utilizing an array of purposely developed smart force elements. Their approach was tested with four gait speeds from 1 m/s (slow walking) up to 3 m/s (slow running). The model has a limitation: there is a deviation of up to 10% of predicted vertical GRF when increasing the speed from walking to running. Other research applied an artificial neural network (ANN) in addition to inverse dynamics-based simulation and developed a hybrid method. For example, in Oh et al. (2013), application of an ANN was limited to a prediction of GRF for a double support phase of the gait cycle. However, none of the research mentioned considered a running case.

In another study, a two-mass model was developed in order to analyze running at different speeds (Clark et al. 2017). However, approach presented has a limitation, since it employs contact time value, which must be estimated based on motion capture data or video data unless it is measured with force platforms. In this case, contact time estimation is the main source of error in the model, as it is, for example, in the research on active peak force prediction (Niemelä et al. 2017).

A neural network approach to this problem is still novel and also promising. There are a number of studies that made a significant contribution to the deep learning approach for biomedical time series. In one of those studies a novel deep learning framework was developed for classification of multivariate time series heart rate data (Zheng et al. 2016). Also, such studies as Bashivan et al. (2015) are dedicated to an application of recurrent CNN on electroencephalogram data or deep CNN application on multi-channel time series data for human activity recognition (Yang et al. 2015).

One study details example of walking gait pattern classification based on twenty handcrafted features extracted from three components of GRF (Andrade

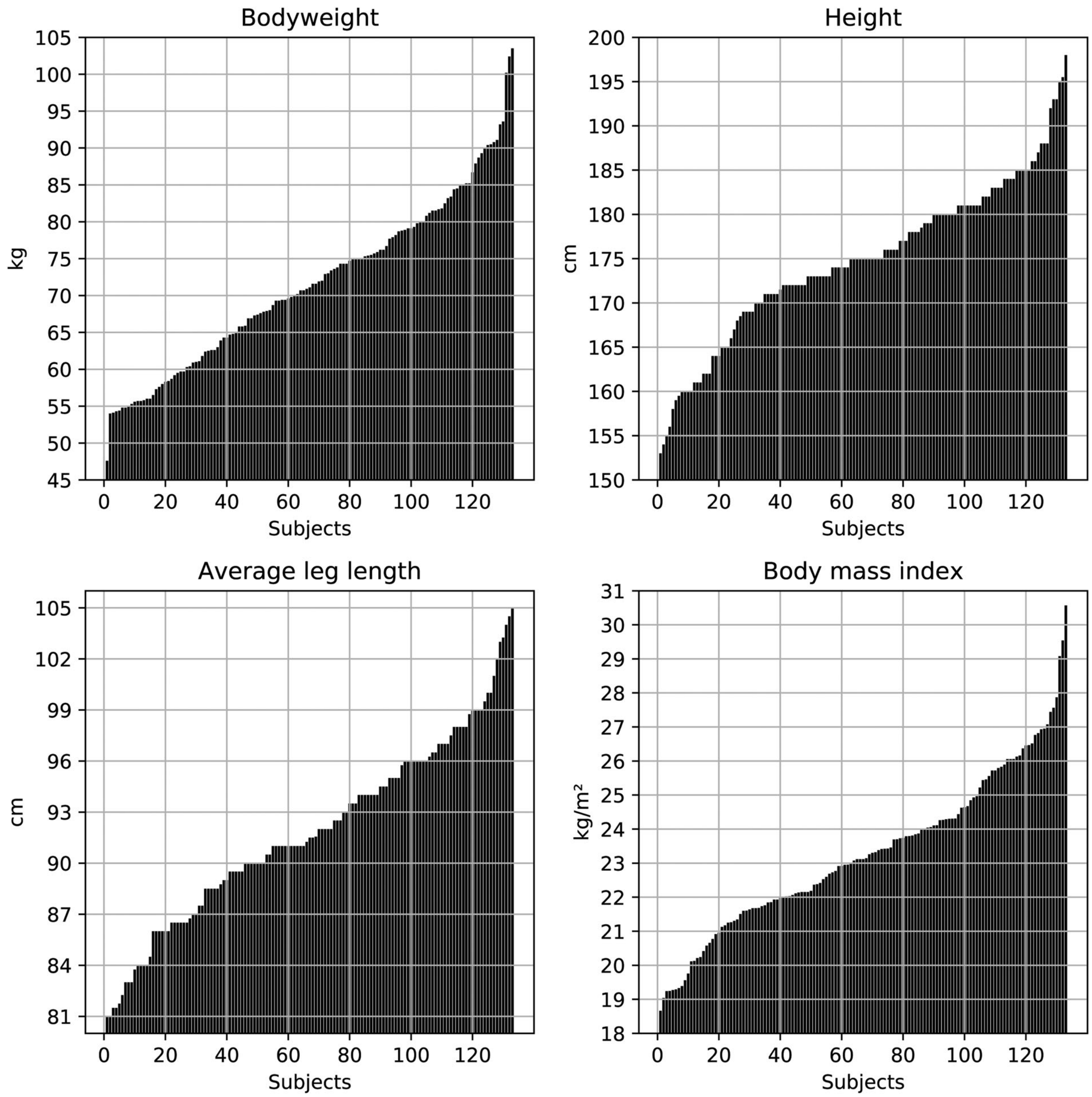

Figure 2. Anthropometric data of the subjects. In the first row: bodyweight (kg), height (cm); in the second row: average leg length (cm), body mass index (kg/m^2). The values on each plot are sorted from the smallest to the greatest independently from the other plots after the calculation of body mass index.

et al. 2015); another includes an example of utilizing of motion analysis data to characterize the differences between osteoarthritic and normal knee function data (Jones et al. 2008). Both used GRF-related data as an input. Another example of the neural network approach is extracting walking gait parameters with a deep CNN (Hannink et al. 2017). Still, none of the above-mentioned studies considered running.

There are not many running-related studies. One of them is a recent paper (Niemelä et al. 2017) related to the problem of the present study that includes prediction of the maximum value of GRF (active peak force) with MLP for running based on kinematic data; however, hand-crafted features (not raw signals) were employed, which is a time-consuming approach that does not utilize all the data. Nevertheless, this study reveals the possibility of using a machine-learning approach to analyze kinematic data in order to predict vertical GRF in running activity. One study (Johnson et al. 2018) with the same motivation as the present study also utilizes raw marker trajectories from a motion capture system beside the subject's mass, sex, and height (but different from the present study: six markers from both feet, pelvis back center point between LPSI and RPSI, and vertebra prominens; cf. Section 3). Nevertheless, Johnson et al. have

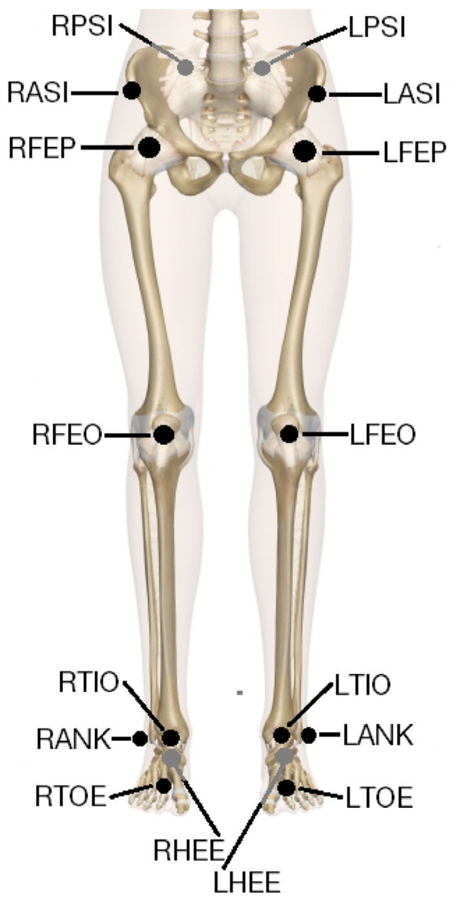

Figure 3. Location of the markers employed (according to standard Vicon plug-in gait lower body model[1]). RPSI, LPSI, RHEE, and RHEE markers' location is on the back side of the body, thus they are marked with gray color. Credit (Taylor 2012).

predicted three components of GRF and moments, based on the sidestep movement, not running. In the recent paper Johnson et al. reported on prediction of GRF and moments for the sidestep movement with pre-trained CNN models (Johnson et al. 2019).

3. Data collection and preprocessing

The kinematic data in three dimensions were recorded by the Vicon motion capture system (Vicon T40, Vicon, Oxford, UK), which includes eight infrared cameras with a sampling frequency of 300 Hz. GRF data were recorded by five force platforms (AMTI BP6001200, AMTI, Watertown, USA) synchronously with the kinematic data at a sampling frequency of 1500 Hz.

Vicon Nexus (v. 1.7.1) software was used to collect kinematic and GRF data in the C3D file format. The standard Vicon plug-in gait lower body model[1] was applied. Healthy subjects ($n = 135$) including both elite and recreational runners participated in the measurements. Gender information was not tracked in this research, because the goal was not to develop gender-specific models. Subjects were running at self-selected speeds (average speed: 3.71 ± 0.77 m/s). Each subject made from 1 to 24 trials; therefore, the collected data set contains 1196 trials altogether.

Anthropometric data collected from the subjects are presented in Figure 2. The average body weight was 71.6 ± 11.4 kg, the average height was 174.8 ± 9.0 cm, the average leg length was 91.8 ± 5.4 cm. The average length of the two legs was taken, because some of the subjects had a difference between left and right leg lengths. From those data, body mass index (BMI) was calculated by dividing body weight in kilograms by height squared in meters. The average BMI was 23.3 ± 2.3 kg/m^2.

For data analysis eight skin markers were selected empirically in order to reduce the number of inputs for the ANN. The skin markers were located on (R/L means right and left side) (Figure 3):

- pelvis front side – (R/L)ASI;
- pelvis back side – (R/L)PSI;
- heels – (R/L)HEE;
- ankles – (R/L)ANK;
- toes – (R/L)TOE.

In addition, virtual joint center markers calculated from the skin markers and anthropometric measures were employed:

- hip joint – (R/L)FEP;
- knee joint – (R/L)FEO;
- ankle joint – (R/L)TIO.

Only one dimension was taken into account to reduce the number of inputs for the ANN – namely vertical axis, since we are interested in vertical GRF in this research.

The labeling of data was performed by seeking the impact peak on the GRF time series with the derivative. A Python function for signal processing, namely SciPy.signal.argrelextrema, was employed for it. A local maximum was accepted as an impact peak when it was higher than 90% of global maximum (active peak) and was not located too far away from the active peak (>50 data points) or too close to it (<10 data points). Time series were chopped on strides taking into account center of mass (COM); therefore the stride in this research is defined as a phase between

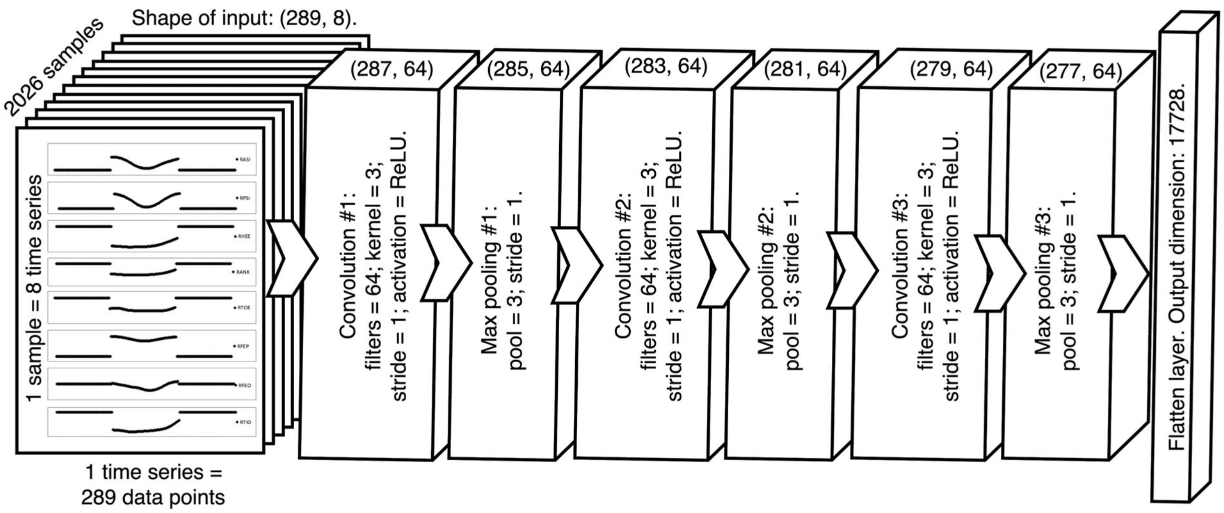

Figure 4. A scheme of a convolutional part of the neural network models.

Table 1. A list of classifiers employed and their training parameters (default if not mentioned).

Classifier	Imported class	Parameters
kNN	sklearn.neighbors.KNeighborsClassifier	Number of neighbors = 7; weight function = distance; algorithm used to compute the nearest neighbors = auto.
MLP	sklearn.neural_network.MLPClassifier	3 layers × 16 units; activation function = ReLU; optimizer = RMSProp; 25 epochs of training; loss function = binary cross entropy.
SVM	sklearn.svm.SVC	Linear kernel; C = 85 000.
Logistic regression	sklearn.linear_model.LogisticRegression	C = 135 000; solver = newton-cg; multi_class = ovr; tolerance for stopping criteria = 0.0001.
Random forest	sklearn.ensemble.RandomForestClassifier	Number of trees in the forest = 15; the function to measure the quality of a split (criterion) = entropy; the number of features to consider when looking for the best split = $\sqrt{number\ of\ features}$.

the two maximum points of the COM position, where the COM time series were obtained as an average of the time series recorded from pelvis markers (RASI, LASI, RPSI and LPSI). Thereafter, the data set at our disposal consisted of 4098 samples (strides). However, there were only 967 samples without the impact peak. It was decided to take only a part of samples with the impact peak. Consequently, every third sample randomly selected among the samples with the impact peak was included into the input data set, since otherwise the data were imbalanced. Hence, the data set used for ANN model creation consisted of 2026 samples: 967 without the impact peak and 1059 with it.

Also, preprocessing of the data included zero padding in a time scale and z-score normalization of signal amplitude. Zero-padding was required by the procedure of the convolutional layers (Hannink et al. 2017). It was performed in such a way that zero values were added on both sides of the time series sample. Strides have different durations from trial to trial. After zero-padding all the samples had an equal number of data points, specifically 289, because of added zero values. The relative duration of the contact phase was left unchanged.

4. Methods

All five models developed are based on a CNN, because the objective is to employ raw kinematics time series and CNN is a technique for cases when the temporal or spatial structure of the data is meaningful. The models were developed with Python 2.7 and Keras library (v2.1.2) for deep learning with Tensorflow framework (v1.41) as a backend. The architecture of the convolutional layers includes three similar convolutional layers, each followed by a similar spatial pooling layer (Figure 4). Parameters of the convolutional layers were as follows: number of filters = 64; kernel size = 3; kernel stride length = 1; activation function = ReLU; kernel initializer = zeros. The spatial pooling layers were represented by the max pooling layer with the following parameters: pool size = 3; pool stride length = 1.

All five models have identical convolutional parts, but the difference between the models is in the classifier part: five different classifiers were employed. In Table 1 a list of classifiers can be found along with parameters that were employed for the training. The parameters for the classifiers were defined with a grid

Table 2. Performance of the models.

	Accuracy	Precision	Sensitivity	Specificity	*F*-measure
kNN	70.08% ± 3.04% $^{\#\#\#}$	71.18% ± 2.93%	71.96% ± 3.95% $^{\#\#\#}$	68.03% ± 4.08%	71.53% ± 3.04% $^{\#\#\#}$
MLP	77.29% ± 4.12% ***	73.57% ± 5.23%	89.61% ± 3.76% ***	63.80% ± 10.89%	80.59% ± 2.63% ***
SVM	80.35% ± 2.97% ***	80.58% ± 3.61% $^{***\#}$	82.43% ± 3.10% ***	78.07% ± 4.84% $^{**\#}$	81.44% ± 2.65% ***
Logistic regression	80.65% ± 2.71% ***	80.98% ± 3.03% $^{***\#}$	82.43% ± 3.33% $^{***\#}$	78.69% ± 4.11% $^{***\#}$	81.66% ± 2.58% ***
Random forest	81.09% ± 2.58% ***	81.55% ± 3.31% $^{***\#}$	82.72% ± 3.17% $^{***\#\#}$	79.30% ± 4.92% $^{***\#}$	82.07% ± 2.31% ***

*Difference versus kNN; $^{\#}$ difference versus MLP.
$^{*\#}p < 0.05$, $^{**\#\#}p < 0.01$, $^{***\#\#\#}p < 0.001$.

search (sklearn.model_selection.GridSearchCV(cv = 5)). All the classifiers were imported from Scikit-learn library (v0.19.0) (Pedregosa et al. 2011). The CNN part is connected with a classifier in such a way that a model is created right after the convolutional part – let us call it an intermediate model – and output is obtained from the intermediate model. This intermediate output is employed as an input for a classifier.

In order to assess the generalization ability of the models, the performance metrics for the classification task were estimated by ten-fold cross-validation so that the test fold was not present in the training phase. Performances of the classifiers were compared using an ANOVA test with a post-hoc Bonferroni test. The normality was tested preliminarily with the Shapiro–Wilk test. A performance of the classifiers was measured in terms of accuracy, precision, sensitivity, specificity, and *F*-measure (Hossin and Sulaiman 2015).

5. Results and discussion

The classifiers listed in Table 2 are ordered according to their performance, from the worst to the best. The worst performance was demonstrated by the kNN algorithm, the second worst was MLP, then SVM, logistic regression, and random forest had demonstrated similar performance.

The performed tests revealed that there is no statistically significant difference for any metric between three classifiers: SVM, logistic regression, and random forest (Table 2). Besides, the MLP classifier's performance has no statistically significant difference from the group of three classifiers mentioned above for most metrics, excepting precision and specificity (Table 2).

Thereafter, according to the present results with the present configuration of the neural network (Figure 4 and Table 1) and with this kind of kinematic time series as inputs (Section 3), the accuracy of the binary classification can gain 81.09% ± 2.58% as the random forest classifier demonstrated. However, the kNN classifier fits the problem worse than the other tested classifiers. The performance achieved is good enough, taking into account that on the one hand, an impact peak appearance is not always clear even for an expert, and on the other hand, the method of impact peak detection employed was not honed to perfection. The approach presented could be, perhaps, improved by the further

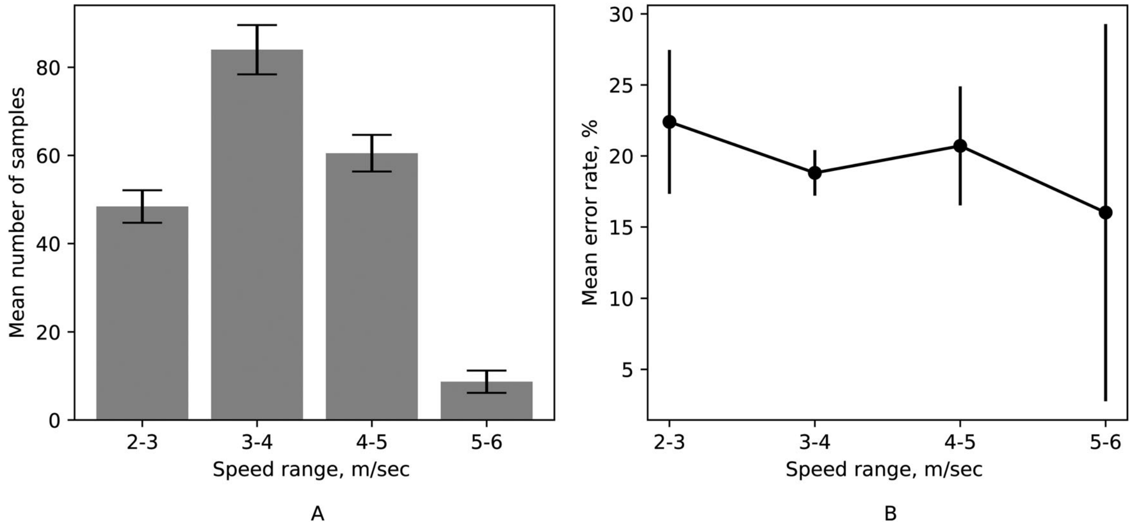

Figure 5. A bar plot A represents mean number of samples over the cross-validated test folds against speed ranges. A plot B represents mean error rate values over the cross-validated test folds for each range of speed. Error rate was calculated as a sum of false positive and false negative predictions divided by the total amount of the predictions. The random forest model was employed to obtain predictions.

improvement of the convolutional part and/or by a different selection of predictors.

The gait speed represented in the data is mostly higher than Jung et al. (2016) had in their research. The majority of the gait speed values are between 2 and 5 m/s in the current study. In Figure 5 the distribution of the gait speeds is represented (plot A) along with the error rates (plot B), which were calculated as the sum of false positive and false negative predictions divided by the total amount of the predictions. The predictions were obtained using the random forest model. This shows that the method developed is applicable to different running speeds. Significant differences were not observed between the four groups of error rate values.

6. Conclusion and future work

In the present research, the prediction analysis of GRF impact peak appearance was accomplished with CNN (Figure 4) and five different classifiers (Table 1). The models were validated with the ten-fold cross validation technique. The performance was estimated on a testing data set that was not presented to a model during the training phase. The highest achieved accuracy and *F*-measure metrics are 81.09% ± 2.58% and 82.07% ± 2.31% correspondingly (Table 2) with the random forest classifier, which is a promising result, but there is still room for improvement. Having a bigger data set may help to train models better. Also, a more systematic approach to the selection of the predictors could yield model inputs with greater predictive power.

The present research has successfully served as an initial testing of applicability of a deep learning approach for this kind of data, namely raw kinematic signals for running. It forms the basis for future research on predicting other features of vertical GRF, and the GRF signal itself, subsequently. Future development may include predictive models for other kinetic features such as joint forces and moments. Further development may also enable vertical GRF predictions via wearable inertial sensors, which in turn, would make an impact load analysis during running even more convenient.

This will lead to a new approach for risk assessment of injuries while running based on predictive analysis and will make the arrangement of measurements less complicated and cheaper due to force platforms exclusion from it. Since running gait analysis belongs to biomedical tasks, the research may have a variety of applications: sports, health care, and rehabilitation, to name a few.

Note

[1] Vicon documentation. Lower body modeling with Plug-in Gait. https://docs.vicon.com/display/Nexus25/Lower+body+modeling+with+Plug-in+Gait Date accessed: February 11, 2018.

Disclosure statement

No potential conflict of interest was reported by the authors.

ORCID

Anastasiia Girka https://orcid.org/0000-0002-6960-7420
Juha-Pekka Kulmala https://orcid.org/0000-0003-1881-4689
Sami Äyrämö https://orcid.org/0000-0002-7532-2771

References

Andrade A, Costa M, Paolucci L, Braga A, Pires F, Ugrinowitsch H, Menzel HJ. 2015. A new training algorithm using artificial neural networks to classify gender-specific dynamic gait patterns. Comput Methods Biomech Biomed Eng. 18(4):382–390.

Bashivan P, Rish I, Yeasin M, Codella N. 2015. Learning representations from EEG with deep recurrent-convolutional neural networks. arXiv preprint arXiv:1511.06448.

Clark KP, Ryan LJ, Weyand PG. 2017. A general relationship links gait mechanics and running ground reaction forces. J Exp Biol. 220(Pt 2):247–258.

Davis IS, Bowser BJ, Mullineaux DR. 2016. Greater vertical impact loading in female runners with medically diagnosed injuries: a prospective investigation. Br J Sports Med. 50(14):887–892.

Fluit R, Andersen MS, Kolk S, Verdonschot N, Koopman HF. 2014. Prediction of ground reaction forces and moments during various activities of daily living. J Biomech. 47(10):2321–2329.

Hannink J, Kautz T, Pasluosta CF, Gasmann K-G, Klucken J, Eskofier BM. 2017. Sensor-based gait parameter extraction with deep convolutional neural networks. IEEE J Biomed Health Inform. 21(1):85–93.

Hossin M, Sulaiman MN. 2015. A review on evaluation metrics for data classification evaluations. Int J Data Min Knowl Manage Process. 5(2):01–11.

Johnson WR, Alderson J, Lloyd D, Mian A. 2019. Predicting athlete ground reaction forces and moments from spatio-temporal driven CNN models. IEEE Trans Biomed Eng. 66(3):689–694.

Johnson WR, Mian A, Donnelly CJ, Lloyd D, Alderson J. 2018. Predicting athlete ground reaction forces and moments from motion capture. Med Biol Eng Comput. 56(10):1781–1712.

Jones L, Holt CA, Beynon MJ. 2008. Reduction, classification and ranking of motion analysis data: an application to osteoarthritic and normal knee function data. Comput Methods Biomech Biomed Eng. 11(1):31–40.

Jung Y, Jung M, Ryu J, Yoon S, Park S-K, Koo S. 2016. Dynamically adjustable foot-ground contact model to estimate ground reaction force during walking and running. Gait Posture. 45:62–68.

Knorz S, Kluge F, Gelse K, Schulz-Drost S, Hotfiel T, Lochmann M, Eskofier B, Krinner S. 2017. Three-dimensional biomechanical analysis of rearfoot and forefoot running. Orthop J Sports Med. 5(7).

Kulmala J-P, Avela J, Pasanen K, Parkkari J. 2013. Forefoot strikers exhibit lower running-induced knee loading than rearfoot strikers. Med Sci Sports Exerc. 45(12): 2306–2313.

Niemelä M, Kulmala J-P, Kauppi J-P, Kosonen J, Äyrämö S. 2017. Prediction of active peak force using a multilayer perceptron. Sports Eng. 20(3):213–219.

Oh SE, Choi A, Mun JH. 2013. Prediction of ground reaction forces during gait based on kinematics and a neural network model. J Biomech. 46(14):2372–2380.

Pedregosa F, Varoquaux G, Gramfort A, Michel V, Thirion B, Grisel O, Blondel M, Prettenhofer P, Weiss R, Dubourg V, et al. 2011. Scikit-learn: machine learning in python. J Mach Learn Res. 12:2825–2830.

Ristolainen L, Heinonen A, Turunen H, Mannström H, Waller B, Kettunen JA, Kujala UM. 2009. Type of sport is related to injury profile: a study on cross country skiers, swimmers, long-distance runners and soccer players. A retrospective 12-month study. Scand J Med Sci Sports. 20(3):384–393.

Taunton JE, Ryan MB, Clement D, McKenzie DC, Lloyd-Smith D, Zumbo B. 2002. A retrospective case-control analysis of 2002 running injuries. Br J Sports Med. 36(2): 95–101.

Taylor T. 2012. Skeletal system. HowToMedia, Inc. http://www.innerbody.com/image/skelfov.html. Accessed 11.02. 2018

van der Worp H, Vrielink JW, Bredeweg SW. 2016. Do runners who suffer injuries have higher vertical ground reaction forces than those who remain injury-free? A systematic review and meta-analysis. Br J Sports Med. 50(8): 450–457.

Van Mechelen W. 1992. Running injuries. Sports Med. 14(5):320–335.

Yang J, Nguyen MN, San PP, Li X, Krishnaswamy S. 2015. Deep convolutional neural networks on multichannel time series for human activity recognition. In: 24th IJCAI, Buenos Aires, Argentina. p. 3995–4001.

Zheng Y, Liu Q, Chen E, Ge Y, Zhao JL. 2016. Exploiting multi-channels deep convolutional neural networks for multivariate time series classification. Front Comput Sci. 10(1):96–112.