Applied Ergonomics 96 (2021) 103414

Contents lists available at ScienceDirect

Applied Ergonomics

journal homepage: http://www.elsevier.com/locate/apergo

Applying deep neural networks and inertial measurement unit in recognizing irregular walking differences in the real world

B. Hu [a,b,*], S. Li [a], Y. Chen [a], R. Kavi [c], S. Coppola [b]

[a] Department of Industrial and Systems Engineering, University of Florida, Gainesville, FL, USA
[b] Department of Environmental Health, Harvard T.H. Chan School of Public Health, Boston, MA, USA
[c] West Virginia University, Morgantown, WV, 26505, USA

ARTICLE INFO

Keywords:
Uneven surface
Gait
Convolutional neural network
Inertial measurement units
Artificial intelligence

ABSTRACT

Falling injuries pose serious health risks to people of all ages, and knowing the extent of exposure to irregular surfaces will increase the ability to measure fall risk. Current gait analysis methods require overly complicated instrumentation and have not been tested for external factors such as walking surfaces that are encountered in the real-world, thus the results are difficult to extrapolate to real-world situations. Artificial intelligence approaches (in particular deep learning networks of varied architectures) to analyze data collected from wearable sensors were used to identify irregular surface exposure in a real-world setting. Thirty young adults wore six Inertial Measurement Unit (IMU) sensors placed on their body (right wrist, trunks at the L5/S1 level, left and right thigh, left and right shank) while walking over eight different surfaces commonly encountered in the living community as well as occupational settings. Three variations of deep learning models were trained to solve this walking surface recognition problem: 1) convolution neural network (CNN); 2) long short term memory (LSTM) network and 3) LSTM structure with an extra global pooling layer (Global-LSTM) which learns the coordination between different data streams (e.g. different channels of the same sensor as well as different sensors). Results indicated that all three deep learning models can recognize walking surfaces with above 0.90 accuracy, with the Global-LSTM yielding the best performance at 0.92 accuracy. In terms of individual sensors, the right thigh based Global-LSTM model reported the highest accuracy (0.90 accuracy). Results from this study provide further evidence that deep learning and wearable sensors can be utilized to recognize irregular walking surfaces induced motion alteration and applied to prevent falling injuries.

1. Introduction

Falling injuries are a serious health risk for anyone, regardless of their age (Verma and others (2016)). Twenty five percent of adults over 65 experience falls annually, with the occurrence of a first fall increasing the likelihood of future fall events (Stevens and others (2012); O'Loughlin and others (1993)). Even though older adults tend to experience falls more frequently, younger adults and children are prone to falling incidents as well. Previous literature has reported that the rate of falls increased with age from 18% in youth, to 21% in middle-aged, and 35% in elders (Talbot and others (2005)). In addition, unintentional falls are the leading cause of nonfatal accidental injury for all ages (CDC (2013)). As such, unintentional falls are a widespread public health issue to address, and a better understanding of fall exposure risk is necessary to design successful interventions.

Fall risk is associated with ground surface conditions, and uneven outdoor surfaces may pose greater risks (Schepers and others (2017); Oxley and others (2018); Menz and others (2003); Su and Dingwell (2007)). Gait adaptations are necessary to navigate on uneven surfaces, but research on real-world outdoor surfaces has been limited. In order to design better outdoor walkways and successful interventions, it is necessary to understand gait biomechanics associated with different surfaces and fall risk in outdoor real-world environments. In addition, real-time recognition of the gait alteration caused by irregular walking surface will also help in timely prevent falling events. However, previous studies in this domain typically involve complex equipment setups that are only practical in laboratory settings such as motion tracking through multiple cameras or ground reaction force sensing devices (Chen and others (2016); Muro-De-La-Herran and others (2014); Tedesco and others (2017)). Although these studies have contributed to a better

* Corresponding author. 1949 Stadium Road, University of Florida, Gainesville, FL, 32606, USA.
E-mail addresses: boyhu@ise.ufl.edu (B. Hu), rahulkavi@gmail.com (R. Kavi).

https://doi.org/10.1016/j.apergo.2021.103414
Received 20 September 2020; Received in revised form 13 January 2021; Accepted 5 March 2021
Available online 1 June 2021
0003-6870/

understanding of how human gait and motor control interact with the environment, the complexity of the equipment setup has restricted the scalability of the implementation and the generalizability of the results. Recently, the incorporation of wearable devices in human motion studies has proven to be a powerful tool and has shown promising results (Chen and others (2016); MuroDe-La-Herran and others (2014); Tedesco and others (2017); Tao and others (2012); Kobsar and others (2014); Schall Jr and others (2016)). Inertial Measurement Unit (IMU) sensors collect data on an individual's gait, providing useful indirect information into users' fatigue levels, physical conditions, and more (Lara and Labrador (2012); Tao and others (2012); Norris and others (2014); Reenalda and others (2016); Shimazaki and Murata (2015)). These devices are a good candidate in reducing the equipment burden without sacrificing too much measurement quality.

Gait and motion analysis has also greatly benefited from the use of machine learning and data analytic methods. Previous literature has shown that machine learning algorithm is capable of recognizing human activity and distinguishing individuals from one another (Lim and D'Souza (2019); Muller and others (2020)). Earlier works have usually utilized intense feature-engineering operations to train classifiers to recognize simple activities with obvious spatial pattern differences. Classifiers that were used in these studies include fuzzy-basis-function-based (FBF-based) classifier (Kao and others (2009); Chen and others (2008)), decision trees (Jatoba and others (2008); Maurer and others (2006)), Bayesian (Tapia and others (2007)) and vanilla neural networks (Randell and Muller (2000)). Although in many cases, these relatively simple and straightforward approaches are adequate to render accurate prediction. However, more elaborate gait applications which are commonly seen in healthcare and behavioral science pose new challenges calling for innovative solutions. A majority of these earlier applications focused on recognizing simple human actions such as walking and running, but the focus has been shifting towards the recognition of complex human activities with subtle differences using advanced machine learning methods. For example, deep learning networks have been applied to classify different Parkinson Disease states (Hammerla and others (2015)), different material-handling models in occupational lifting tasks (Mehrizi and others (2019)), and to predict pedestrian trajectory in varied social interactions (Alahi and others (2016); Gupta and others (2018); Sadeghian and others (2019)). It is worth noting that most of these deep learning applications utilize image data while time series spatial and temporal information collected from wearable IMUs are much infrequently used. One motivation of the current work was to expand the deep learning application spectrum with richer data modality. Our previous works have used deep learning networks and a simple IMU sensor setup (1 or 2 sensors) to recognize walking/running surfaces with different levels of irregularity (Hu and others (2018); Dixon and others (2019)). These studies were successful in distinguishing between different surfaces and age-related traits while conceptually demonstrating the feasibility of wearable sensors in this context, however, those studies only included limited (i.e. no more than 3) experimental walking surfaces. There are numerous categories of irregular surfaces commonly seen in the real-world that can significantly affect human gait performance (Dixon and Pearsall (2010); Damavandi and others (2012)). Another limitation of our previous studies is the possibility of over fitting caused by the restrained sample size. Thus, larger data sets of more terrain surfaces are needed in order to validate the scalability of the algorithms. Research using wearable sensors has also emphasized the importance of collecting data outside of laboratory settings. Real-life environments are complex, and there are various conditions that can affect gaits such as clothing, footwear, load carrying, walking surfaces, and the inclination of the ground (Sprager and Juric (2015)). Although lab-based systems can collect highly accurate human movement data, they are relatively expensive and require expert operator (Simon (2004)). Furthermore, they are restricted to laboratory settings and thus the information derived may not reflect gait in real-world contexts. Testing in real-world scenarios allow better external validity but can come at a cost of decreased accuracy when analyzing such data, raise the demand for better analysis technique (Brodie and others (2016); Alsheikh and others (2016); Khandelwal and Wickstrom (2017); Weiss and others (2011, 2013)). Therefore, the primary aim of this study was to investigate if outputs from wearable IMUs coupled with a deep convolutional neural network could detect motion pattern alteration caused by real-world irregular walking surfaces. Our first goal was to characterize the precision and recall performance with which deep learning algorithms can detect walking surface categories using IMU data. The second goal was to compare the performance from different sensor placements and neural network layouts.

2. Methods

2.1. Participants

Thirty young adults volunteered for this study (14 females, 16 males, 23.5 ± 4.3 years, 169.3 ± 21.9 cm, 71.2 ± 14 kg). All participants had normal or corrected to normal vision. Participants had no reported neurological or musculoskeletal conditions that affected their gait or posture and no history of falling injuries in the previous two years. The Harvard and Northeastern Institutional Review Boards approved this study and all participants provided written consent.

2.2. Data collection on the irregular walking surfaces

Participants performed 42 walking trials (7 different surfaces $\times$ 6 repetitions) while wearing six IMU sensors set (MTw Awinda, Xsens, the Netherlands). Each sensor included a tri-axial accelerometer, gyroscope, and magnetometer with respective ranges of ± 160 m/s2, ± 2000 deg/s and ± 1.9 Gauss. The sampling frequency was 100 Hz. The locations for the sensors were: 1) right wrist; 2) the mid-line of the lower-back (over the L3-L5 spinous processes of the trunk); 3–4) left and right thigh; 5–6) left and right shank (Fig. 1). The researchers palpated participants' bodies to place the sensors. The seven walking surfaces were: 1) flat uneven cobble stone (26 $\times$ 18 cm blocks) pavement); 2–3) bank right and 4) bank left; 4–5) slope up and slope down; 6–7) upstairs and downstairs (Fig. 2). The eighth condition was a static standing without locomotion activity.

There was no inclination in any of the surfaces except the slope condition. Participants walked around 15 m for each trial with small variances between each condition. All testing surfaces were in the real-world and outside of the lab which guarantees the external validity of the study. Data collection occurred on a busy university campus with foot, vehicle, and train traffic which tests whether the algorithm will be robust to external influences in the data. These irregular surfaces were presented in a randomized order and rest between surfaces was provided to minimize fatigue. The overall protocol required less than 2 h in order to reduce fatigue. All data collection was conducted on days without precipitation and the walking surfaces were dry and clear. All experiments were finished during daylight hours and the same surfaces were used for each participant. It is worth noting that footwear attributes (comfortness, sole pattern, etc.) can also influence the wearer's walking pattern. Thus, on the lab visit day participants were informed in advance that they wore their daily sport sneakers. Before formal data collection, the experimenter manually checked the wear and tear of the thread on the sole (more than 1/16″). More detailed descriptions of this data set can be found in our recent publication (Luo and others (2020)).

2.3. Initial data processing

For each trial, all 36 data channels (3D accelerometer and 3D gyroscope from each of the 6 sensors) were scaled from 0 to 1 to ensure equal weightings across channels and improve model performance. A sliding window of 4 s width (400 data points) that captured 2–3 full gait

Fig. 1. Sensor setup demonstration.

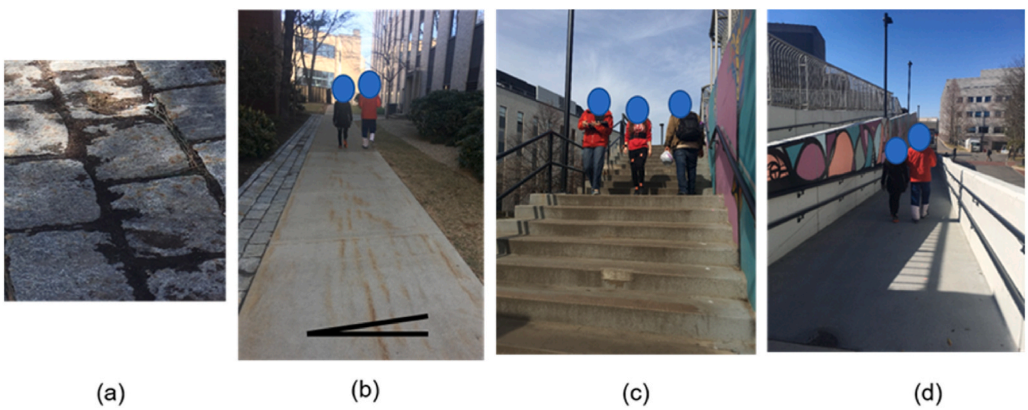

Fig. 2. Irregular surfaces tested in the study: a) uneven cobble stone; b) flat bank surface; c) stairs; d) slope surface.

cycles and a fixed step length of 1 s was applied on each trial to segment and augment the data, respectively (Xun and others (2016); Hu and others (2018)). After the data segmentation procedure, the whole data set included 49420 time series. The detailed category distribution is shown in Table 1.

2.4. Network architecture

Three different variations of different deep learning networks were generated in this study: 1) Convolutional neural networks (CNN) are feed-forward that are different from fully connected multi-layer networks as they include one or more convolutional layers. This method has the advantage of significantly reducing the computational complexity with respect to fully connected feed forward neural networks. CNNs have been proven to be excellent feature extractors for motion data (Gadaleta and Rossi (2018)). It has the speciality of solving classification problems of sensor data, some of the previous works on time series signal classification have shown that this model is practicable (Kiranyaz and others (2019); Faust and others (2018); Nweke and others (2018); Strodthoff and Strodthoff (2019)); 2) long short term memory (LSTM) network is a variation of recurrent neural network that can learn the temporal dynamics of sequential data, which is well suited for learning time series data obtained from IMU sensors (Steven Eyobu and Han (2018); Hochreiter and Schmidhuber (1997)); 3) Global-LSTM: despite its advantages in learning temporal features, LSTM networks have a weakness in terms of capturing dependencies and interactions between multiple correlated time series. In the context of human activity recognition, different channels of the same IMU sensor (3 axis acceleration and 3 axis angular velocity) and outputs from different sensors all carry critical information which reveals segments' spatial coordination and motion patterns. To address this limitation, inspired by the recent success of Social-LSTM for pedestrian walking trajectory regression (Alahi and others (2016)), in this work we proposed a new approach - LSTM with a global pooling mechanism (i.e. Global-LSTM) that is able to learn the correlation between different parallel time series. First, the CNN model started with the data input layer, and data from the IMU sensor were used to train the model. Therefore, each training data can be treated as a matrix of 36×400. The input layer was followed by the first 1D convolution layer which included 100 filters in size of 50 and a batch normalization layer. The structure (100×50 1D convolution + batch normalization) was repeated once (i.e. 2nd convolution layer) plus a drop out layer of 0.2. Subsequently, the structure (100×50 1D

Table 1
Class IDs for each action.

Class ID	Action name
0	Stand
1	Flat uneven cobble stone
2	Bank left
3	Bank right
4	Uphill slope
5	Downhill slope
6	Upstairs
7	Downstairs

convolution + batch normalization) was repeated another time plus a maxpooling layer (size of 3) (Scherer and others (2010)) and a drop out layer of 0.2. The purpose of having three convolution layers was to perform dimensionality reduction, feature extraction, and extract local connectivity (LeCun and others (2015); Gadaleta and Rossi (2018)). The 4th to 6th convolution layers were 50 × 25 1D convolution plus batch normalization. Only the 6th layer had a drop out of 0.2, followed by a 1D global average pooling layer. Finally, a fully connected layer with 8 neurons (i.e. 8 types of walking surface) with Softmax as the activation function completed the CNN architecture. The output of the Softmax function represents the probability of each walking surface in each trial. The network architecture is illustrated in Fig. 3. Furthermore, multiple sub-models that only used a subset of sensors were also trained on the data to investigate the influence of sensor placement and data fusion on the model performance. In terms of the LSTM models, in the current study, the model first started with the input data layer. We added one layer of LSTM unit with 80 filters, with Relu as the activation function. We inputted the whole reshaped data and returned a sequence of data by the LSTM Unit, followed by another same LSTM layer. Dropout 50% of the neural network unit from the model was then applied. Subsequently, we added another two LSTM layers and a dropout layer but with 100 filters. The model ended with an output layer. The network architecture is illustrated in Fig. 4. For the Global-LSTM network, we first create a LSTM cell with 64 embedding dimension, with a linear layer to embed the linear position and the social tensor. Same as the LSTM model, we also use Relu as the activation function and a dropout rate of 50%. Each LSTM cell will have a 128-dimensional hidden layer and mapped to the output. After getting the output from the cells, all correlated time series output will be forwarded to a global pooling layer and be calculated as a global tensor there. Each cell learns the global information by getting the distance from other cells at each timestamp. The network architecture is illustrated in Fig. 5.

2.5. Model implementation and training

The CNN and LSTM models were developed with TensorFlow (Abadi and others (2016)) and Keras (Chollet (2015)) under Python (Python Software Foundation, https://www.python.org/), running on Google's (Google LLC, Mountainview, USA) Colaboratory GPU (GPU: 1xTesla K80, 2496 cores, 12 GB RAM). Global-LSTM model was developed with Pytorch (Paszke and others (2019)) under Python. Models were trained in a fully-supervised way and gradients were backpropagated from the

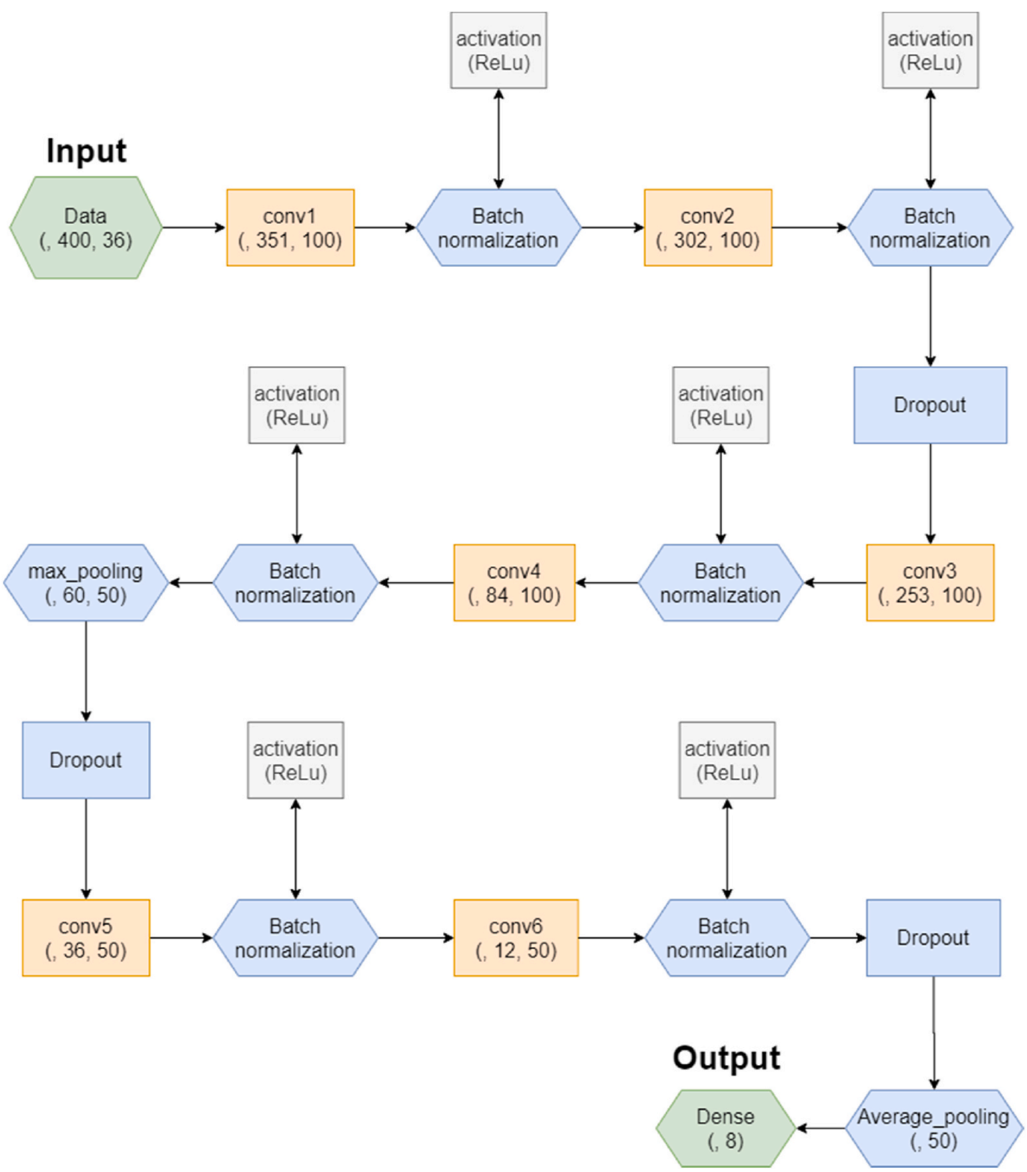

Fig. 3. Convolution Neural Network architecture.

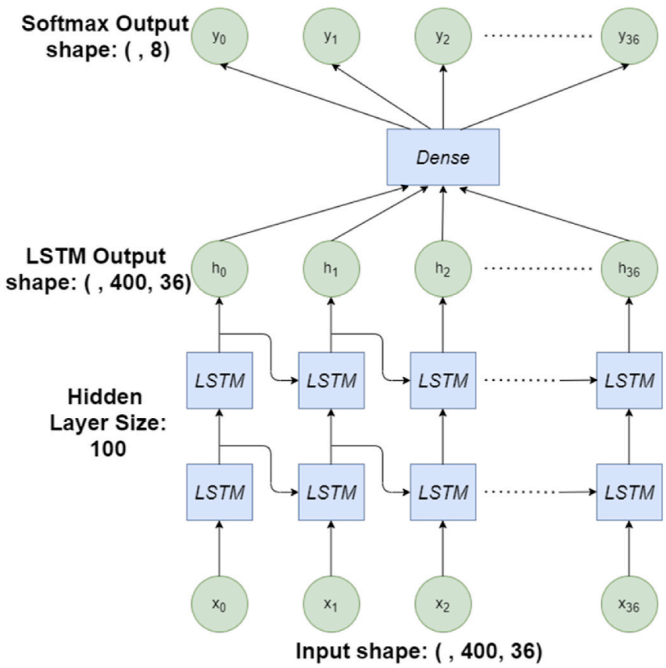

Fig. 4. Long Short Term Memory architecture.

final Softmax layer to the input layer. The 'sparse categorical cross entropy' was applied as the loss function and was optimized using mini-batch gradient descent. Adaptive Moment Estimation (Adam) was used as the update rule due to its optimization convergence rate (Kingma and Ba (2014)). The initialization of model parameters was randomly assigned through the normal distribution. For the activation functions of the network (for the initial layers only), "swish" activation function was chosen. It is a relatively new activation function proposed by (Ramachandran and others (2017)).

2.6. Model performance analysis

The performance of the models was evaluated with the testing data set. Specifically, the following metrics were processed and compared: (1) overall prediction accuracy (the percentage of time series correctly predicted by the model); (2) precision (the number of true positives divided by the sum of true positives plus false positives); (3) recall (the number of true positives divided by the sum of true positives and false negatives); (4) F1-score (harmonic mean of (2) and (3)) (Powers (2011)).

3. Results

In order to measure the performance of the model on the data set and elicit guidance on the sensor placement location for future applications, we created subsets of data to see which sensors contributed the most to the performance of the model. We have taken this approach to see if one of the sensors (Table 2) disproportionately contributed to the performance of the model (CNN: Table 3; LSTM: Table 4; Global-LSTM: Table 5). For the CNN models, each individual sensor averaged between 0.71 and 0.78 (macro weighted avg precision, recall, and F1-score). Sensor 1 and 2 (i.e. wrist and lower back) produced the best model (0.78 accuracy). On the other hand, sensor 3 (i.e. left thigh) had the worst prediction performance (0.72 accuracy). However, sensor fusion may have improved model performance in this case: compared to

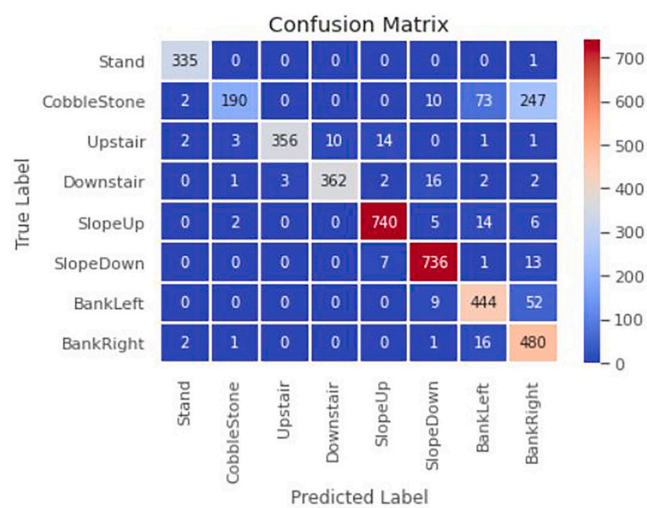

Fig. 6. Confusion Matrix of CNN model.

Table 2
Various sensor subsets for sub-models.

Dataset ID	Sensors	Location
S1	Sensor 1	Wrist
S2	Sensor 2	Lower back
S3	Sensor 3	Left thigh
S4	Sensor 4	Right thigh
S5	Sensor 5	Left Shank
S6	Sensor 6	Right Shank
S7	All 6 sensors	All above

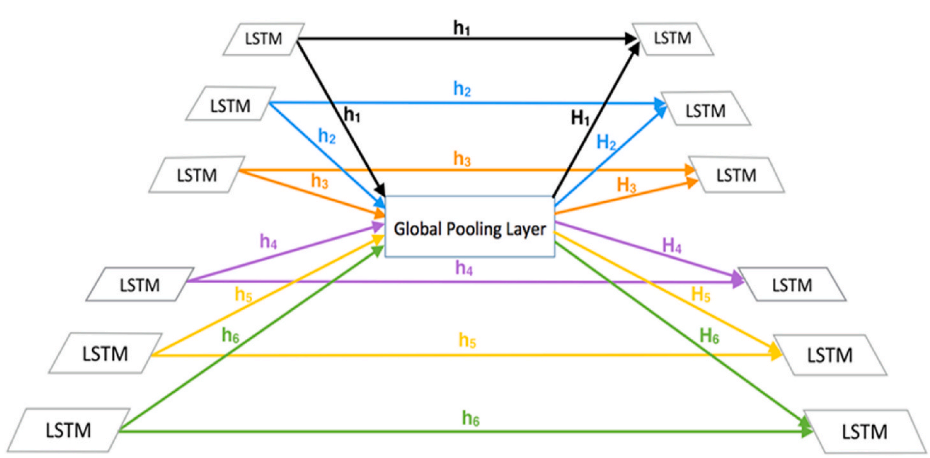

Fig. 5. Global pooling layer architecture of Global LSTM.

Table 3
Sensor sub-model results for the CNN models.

Dataset ID	Avg Precision	Avg Recall	Avg F1-Score
S1	0.78	0.77	0.77
S2	0.78	0.77	0.78
S3	0.72	0.71	0.71
S4	0.77	0.77	0.77
S5	0.75	0.75	0.75
S6	0.77	0.76	0.75
S7	0.91	0.91	0.90

Table 4
Sensor sub-model results for the LSTM models.

Dataset ID	Avg Precision	Avg Recall	Avg F1-Score
S1	0.90	0.89	0.89
S2	0.79	0.78	0.79
S3	0.89	0.87	0.88
S4	0.89	0.88	0.88
S5	0.87	0.88	0.88
S6	0.88	0.87	0.87
S7	0.91	0.90	0.90

Table 5
Sensor sub-model results for the Global-LSTM models.

Dataset ID	Avg Precision	Avg Recall	Avg F1-Score
S1	0.88	0.90	0.89
S2	0.84	0.83	0.84
S3	0.87	0.86	0.87
S4	0.90	0.89	0.89
S5	0.88	0.86	0.87
S6	0.89	0.90	0.90
S7	0.92	0.93	0.92

a single sensor, the combination of all sensors elicited the best classification performance (S7 model, 0.91 accuracy). This result indicated that sensor placement could potentially affect model classification performance. In regard to the LSTM networks, models trained from each sensor data elicited significantly improved results: on average between 0.79 and 0.91 (accuracy). Sensor 1 produced the best performance (0.90 accuracy). Surprisingly, sensor 2 barely benefited from the new model structure: the model produced the worst performance (0.79 accuracy). In order to measure the performance of the model on the data set and elicit guidance on the sensor placement location for future applications, we created subsets of data to see which sensors contributed most to the performance of the model. Finally, for the Global-LSTM networks, the overall accuracy was further improved to 0.92 (i.e. the best in the study). In addition, model trained using sensor 2 data was improved to 0.84 which demonstrated the effectiveness of the global layer on this part of data. Finally, confusion matrices of the best three models (best one from each structure) were presented as Figs. 6–8.

4. Discussions

The purpose of this study was to evaluate if deep learning models on wearable IMU sensor data can detect different real-world irregular walking surfaces during human walking tasks. The results indicate in all three categories, these models can detect different walking surfaces with satisfactory performance. LSTM and Global-LSTM networks yielded improved performance when compared to the baseline CNN models. Furthermore, results from the current study could guide optimal sensor placement and sensor fusion strategy for risk predicting models. Utilizing AI methods to model the human physical motion trajectory has received substantial attention recently. This direction of research has broad implications including but not limited to: autonomous driving and pedestrian safety (Liang and others (2020); Tang and Salakhutdinov

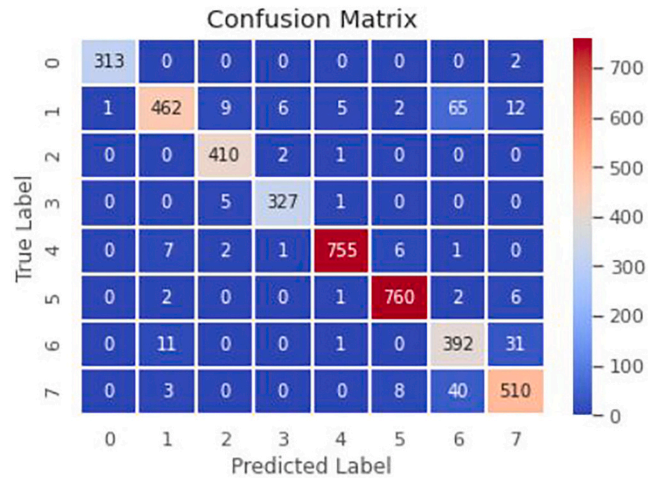

Fig. 7. Confusion Matrix of LSTM model.

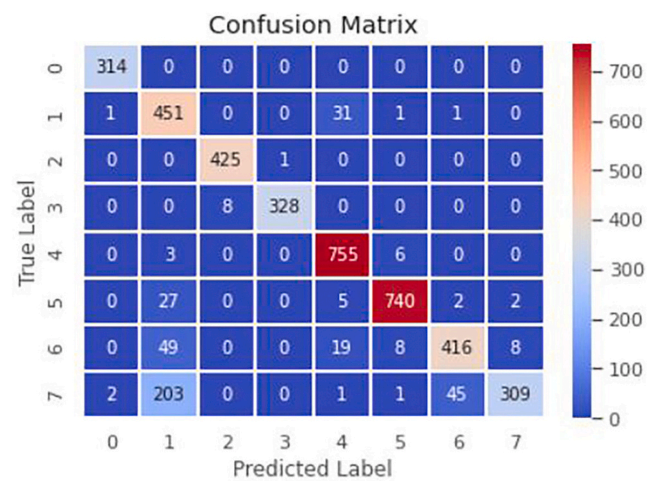

Fig. 8. Confusion Matrix of Global LSTM model.

(2019)), human safety in collaborative human robotics interaction (Kruse and others (2012); Lasota and others (2017)), and public safety surveillance (Fernando and others (2018)). In the current work, we successfully applied multiple deep learning models to detect gait alteration caused by different walking surfaces. Despite its perceived intuitiveness, there are many fundamental challenges in solving the problem of human gait pattern prediction due to its inherent complexity and the stochastic and dynamic interaction between humans and surrounding environments. Since subtle locomotion adaptations to the environment are not easily detectable by human intelligence, this presents significant challenges to AI methods. The CNN model trained with all 6 sensors elicited an overall accuracy of 0.91 in detecting irregular walking surfaces in the real-world environment. Meanwhile, models trained with only one sensor were less accurate, suggesting that sensor fusion is necessary for accurate gait analysis. In our previous laboratory study, one IMU placed on the L5S1 joint combined with a deep learning neural network with LSTM units was able to recognize gait alteration caused by an irregular uneven brick surface at 0.96 accuracy (a binary classifier) (Hu and others (2018)). The study takes a step forward by conducting the experiment in a real outdoor environment and aims to maximize ecological validity. There are several notable benefits obtained by switching the experiment environment that helped to maximize the ecological validity. Firstly, conducting the experiment outdoors introduced multiple unpredictable factors that could affect human behavior such as temperature, wind, and behavior of surround pedestrians.

Secondly, the laboratory study used a designed surface where the magnitude of unevenness between bricks was larger than that found in the outdoor environment. Thus, it may be less challenging for the human motor control system to adapt to and it may cause less gait alteration. Apart from these benefits, we had the opportunity to validate our algorithms for outdoor use, as previous studies have shown that some algorithms trained with well controlled indoor data do not perform as well when tested outdoors (Sprager and Juric (2015)). However, our results (0.92 accuracy) suggest that the algorithm is robust in an outdoor setting. Although the full six sensors elicited the best results, the relatively complex setup may cause some compliance issues during the real-world implementation. Some approaches may need to be taken in mitigating the negative impacts. To begin with, IMU sensors related communication micro processor and battery unit can be integrated into the personal protective devices (PPE) that workers are required to wear per OSHA regulations such as safety vests, helmets, and safety boots. This would enable the system to be carried easily by the users as well as avoiding the potential compliance problem. In addition, we want to emphasize that results from the current study indicated a single sensor can also achieve excellent performance, especially for the GlobalLSTM models. For instance, if a specific application context only includes stairs and a flat surface, then we can look into the results table and select a single sensor that has the best classification accuracy. Among all the six individual sensors, the models that trained with low back sensor outputs showed the lowest overall accuracy except the CNN model. This finding might be explained from the motor control perspective. The foot and distal limb serve as the 'shock absorber' which absorbs, modulates, and controls the foot-ground interaction to maintain a relatively stable upper body. The irregularity of the walking surfaces could still generate subtle but consistent and unique modifications to a person's gait and which would be recognized by the AI models. However, when it comes to the lower back region, the whole body gait and balance control may well mitigate the influence of the walking surfaces and keep the movement pattern of this segment unchanged. This explanation is also supported by the results that Global-LSTM which is able to learn the coordination between different time series channels had the highest accuracy in general. Furthermore, humans rely on both active and passive control to adapt their gaits to terrain variations and tend to move their centers of mass ballistically during walking, utilizing the available mechanical forces and inherent stability to maintain an optimal metabolic energy expenditure status (Matthis and others (2017)). It is possible that in the current experiment, participants' motor control systems absorbed and modulated the irregular surfaces' impact effectively and maintained their gait sta-ble and relatively unchanged at the low back level. Follow up kinematics analysis is needed to evaluate this finding and we will describe the analysis in another manuscript. Advanced machine learning methods are becoming more and more common in activity analysis, and these results suggest that deep learning is a feasible method for conducting real-world gait analysis. Compared to the other previous human activity recognition studies, the current CNN models showed comparably accurate performance (Roggen and others (2010); Lockhart and others (2011); Anguita and others (2012); Bachlin and others (2010); Li and others (2020); Hammerla and others (2016); Jiang and Yin (2015)). It is worth mentioning that these studies analyze more distinguishable activities such as normal walking, eating, etc. For those few studies that included varied irregular surfaces, the magnitude of the unevenness of the walkway used was larger. In contrast, the current study used real-world structures and the unevenness was mainly caused by design and natural deterioration. There remain some technical challenges to using deep learning and wearable IMUs for accurate gait analysis. In many cases, the features derived from deep learning networks might be less discriminating than shallow features that come from a manual feature engineering process. A possible explanation may be the fact that in order to train deep learning networks properly, complex layer structure and a large amount of training data are required to recognize the entire hierarchy of features. Since deep learning is a data driven process by nature, if the input data is not comprehensively represented in all the possible modalities, the trained networks will not be able to generalize these data modalities automatically for the subsequent classification purpose (Ravi and others (2016)). We hope that the results from this study can be used as a guide on how to apply IMUs and deep learning technology to avoid falling injuries and create observable impacts in the applied ergonomics community. A few applicable scenarios are described here: 1) real-time unsafe walking surface detection and warning. People may not be aware of the alteration of the surfaces and related falling risks due to multiple reasons such as their attention is taxed by their work at hand, physical/mental fatigue caused by prolonged work, or complacency issues. This phenomenon is commonly seen in occupational settings as well as our daily lives. For example, the lane departure warning system in modern vehicles is a good example of correcting this type of human errors. Results from the current study will help in the development of similar systems in the pedestrian walking domain; 2) occupational long term risk exposure assessment. In the construction industry, the working environment is highly dynamic and unstructured, traditional ergonomics observation methods will not induce effective and accurate risk exposure evaluation. With our algorithms, workers' detailed falling risk exposure profile throughout the course of their work shift can be obtained. Employers can utilize this piece of information to adjust their safety program, while workplace insurance providers can consider this in their premium determination models. Following the recent advancements in Recurrent Neural Network (RNN) models for sequence data recognition tasks, we proposed a Global-LSTM model which can learn general human movement patterns and predict their future adaptions due to the environmental influences. This is in contrast to traditional approaches that use hand-crafted functions such as classical gait or biomechanics analyses. While LSTMs have the ability to learn and reproduce long time series, they do not capture dependencies between multiple correlated sequences. This weakness is observed in the current study: although single sensor LSTMs outperformed CNNs by a big margin, the full set models did not show a comparable level of improvement because CNNs can take advantage of multi-modal data streams and learn their interconnections while LSTMs lack this capability. Compared to regular LSTM models, we applied an innovative layer (i.e. global layer) to learn the interconnection and coordination between different joints and body segments during the physical motion in 3D space. Specifically, we designed a novel architecture which connects the LSTMs corresponding to nearby sequences (i.e. data streams from different sensors). In particular, we introduce a "Global" pooling layer which enables the LSTMs of proximal time series to share their hidden-states with each other. This new model, which we refer to as the "Global-LSTM", can automatically extract and learn typical interactions among sensors' outputs which coincide in the time domain. Results from the study demonstrate the effectiveness of this new network architecture. Several limitations of this study need to be noted. First, even though interpretable machine learning has received substantial attention recently, CNN models still work in a quasi-black box manner. This makes it difficult to understand exactly which gait features were exploited by the models in the classification task. Follow-up studies with traditional gait analyses are necessary to provide more insight into this question. The participants in this study were all healthy young adults which may limit generalizability to the overall population. In addition, we only collected data during days without precipitation to protect participants as well as the digital devices. However, on some surfaces walking can be significantly altered in case of rain or snow fall, when surfaces become more slippery. Future studies with more safety precaution approaches are required to test the algorithms with wet surfaces. In addition, while we were successful in detecting subtle gait changes due to walking surface changes, more work is necessary to understand which of these gait adaptations are associated with fall risk and potential applications of sensors and machine learning might be applied to real-world fall prevention. Lastly, a systematic feature selection Ribeiro and others (2016) and Lundberg and Lee

(2017), for instance and ranking analysis would shed light on the system optimization, which is lacking in the current study.

5. Conclusion

In conclusion, this study developed different deep learning models to classify different irregular surfaces using data collected from wearable sensors during human walking tasks with satisfactory performance. Different sub-models were also trained to investigate the influence of sensor placement, data fusion, and model structure on model prediction performance. Results indicate the strong potential of using deep learning models and wearable sensors to track human gait and prevent falling injuries in real-world settings.

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.

Acknowledgments

This investigation was made possible by Grant No. T42 OH008416 from the National Institute for Occupational Safety and Health (NIOSH). Its contents are solely the responsibility of the authors and do not necessarily represent the official views of NIOSH. The authors would like to thank Julia Rhyins, Alice Loukianova, and Emily Man for help with the collection of the data. The authors would also like to thank Philipe Dixon, Nicholas Gonzales, and Yue Luo for help with the data processing and cleaning. The authors would also like to thank Mustafa Yerebakan, Nicolas Grimaldi and Samuel Falzone for help with the manuscript development.

References

Abadi, Martin, Agarwal, Ashish, Barham, Paul, Brevdo, Eugene, Chen, Zhifeng, Citro, Craig, Corrado, Greg S., Davis, Andy, Dean, Jeffrey, Devin, Matthieu, others, 2016. Tensorflow: Large-Scale Machine Learning on Heterogeneous Distributed Systems. arXiv preprint arXiv:1603.04467.
Alahi, Alexandre, Goel, Kratarth, Ramanathan, Vignesh, Robicquet, Alexandre, Fei-Fei, Li, Savarese, Silvio, 2016. Social lstm: human trajectory prediction in crowded spaces. In: Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition, pp. 961–971.
Alsheikh, Mohammad Abu, Selim, Ahmed, Niyato, Dusit, Doyle, Linda, Lin, Shaowei, Tan, Hwee-Pink, 2016. Deep activity recognition models with triaxial accelerometers. In: Workshops at the Thirtieth AAAI Conference on Artificial Intelligence.
Anguita, Davide, Ghio, Alessandro, Oneto, Luca, Parra, Xavier, Reyes-Ortiz, Jorge, L., 2012. Human activity recognition on smartphones using a multiclass hardwarefriendly support vector machine. In: International Workshop on Ambient Assisted Living. Springer, pp. 216–223.
Bachlin, Marc, Plotnik, Meir, Roggen, Daniel, Giladi, Nir, Hausdorff, Jeffrey M., Troster, Gerhard, 2010. A wearable system to assist walking of Parkinson s disease patients. Methods Inf. Med. 49 (1), 88–95.
Brodie, Matthew AD., Coppens, Milou JM., Lord, Stephen R., Lovell, Nigel H., Gschwind, Yves J., Redmond, Stephen J., Del Rosario, Benjamin, Michael, Wang, Kejia, Sturnieks, Daina L., Persiani, Michela, others, 2016. Wearable pendant device monitoring using new wavelet-based methods shows daily life and laboratory gaits are different. Med. Biol. Eng. Comput. 54 (4), 663–674.
CDC, 2013. Leading causes of nonfatal injury reports, 2001–2014. URL: https://webappa.cdc.gov/sasweb/ncipc/nfirates2001.html.
Chen, Shanshan, Lach, John, Lo, Benny, Yang, Guang-Zhong, 2016. Toward pervasive gait analysis with wearable sensors: a systematic review. IEEE J. Biomed. Health Informat. 20 (6), 1521–1537.
Chen, Yen-Ping, Yang, Jhun-Ying, Liou, Shun-Nan, Lee, Gwo-Yun, Wang, JeenShing, 2008. Online classifier construction algorithm for human activity detection using a tri-axial accelerometer. Appl. Math. Comput. 205 (2), 849–860.
Chollet, Francois, 2015. Keras: Theano-based deep learning library. Code. https://github.com/fchollet. Documentation. http://keras.io.
Damavandi, Mohsen, Dixon, Philippe C., Pearsall, David J., 2012. Ground reaction force adaptations during cross-slope walking and running. Hum. Mov. Sci. 31 (1), 182–189.
Dixon, P.C., Schutte, K.H., Vanwanseele, B., Jacobs, J.V., Dennerlein, J.T., Schiffman, J. M., Fournier, P.A., Hu, B., 2019. Machine learning algorithms can classify outdoor terrain types during running using accelerometry data. Gait Posture 74, 176–181.
Dixon, Philippe C., Pearsall, David J., 2010. Gait dynamics on a cross-slope walking surface. J. Appl. Biomech. 26 (1), 17–25.
Faust, Oliver, Hagiwara, Yuki, Hong, Tan Jen, Lih, Oh Shu, Acharya, U Rajendra, 2018. Deep learning for healthcare applications based on physiological signals: a review. Comput. Methods Progr. Biomed. 161, 1–13.
Fernando, Tharindu, Denman, Simon, Sridharan, Sridha, Fookes, Clinton, 2018. Soft+hardwired attention: an lstm framework for human trajectory prediction and abnormal event detection. Neural Network. 108, 466–478.
Gadaleta, Matteo, Rossi, Michele, 2018. Idnet: smartphone-based gait recognition with convolutional neural networks. Pattern Recogn. 74, 25–37.
Gupta, Agrim, Johnson, Justin, Fei-Fei, Li, Savarese, Silvio, Alahi, Alexandre, 2018. Social gan: socially acceptable trajectories with generative adversarial networks. In: Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition, pp. 2255–2264.
Hammerla, Nils Yannick, Fisher, James, Andras, Peter, Rochester, Lynn, Walker, Richard, Plotz, Thomas, 2015. Pd disease state assessment in naturalistic environments using deep learning. In: Twenty-Ninth AAAI Conference on Artificial Intelligence.
Hammerla, Nils Y., Halloran, Shane, Plotz, Thomas, 2016. Deep, Convolutional, and Recurrent Models for Human Activity Recognition Using Wearables. arXiv preprint arXiv:1604.08880.
Hochreiter, Sepp, Schmidhuber, Jurgen, 1997. Long short-term memory. Neural Comput. 9 (8), 1735–1780.
Hu, B., Dixon, P.C., Jacobs, J.V., Dennerlein, J.T., Schiffman, J.M., 2018. Machine learning algorithms based on signals from a single wearable inertial sensor can detect surface and age-related differences in walking. J. Biomech. 71, 37–42.
Jatoba, Luciana C., Grossmann, Ulrich, Kunze, Chistophe, Ottenbacher, Jorg, Stork, Wilhelm, 2008. Context-aware mobile health monitoring: evaluation of different pattern recognition methods for classification of physical activity. In: 2008 30th Annual International Conference of the Ieee Engineering in Medicine and Biology Society. IEEE, pp. 5250–5253.
Jiang, Wenchao, Yin, Zhaozheng, 2015. Human activity recognition using wearable sensors by deep convolutional neural networks. In: Proceedings of the 23rd ACM International Conference on Multimedia. Acm, pp. 1307–1310.
Kao, Tzu-Ping, Lin, Che-Wei, Wang, Jeen-Shing, 2009. Development of a portable activity detector for daily activity recognition. In: 2009 IEEE International Symposium on Industrial Electronics. IEEE, pp. 115–120.
Khandelwal, Siddhartha, Wickstrom, Nicholas, 2017. Evaluation of the performance of accelerometer-based gait event detection algorithms in different real-world scenarios using the marea gait database. Gait Posture 51, 84–90.
Kingma, Diederik P., Ba, Jimmy, 2014. Adam: A Method for Stochastic Optimization. arXiv preprint arXiv:1412.6980.
Kiranyaz, Serkan, Avci, Onur, Abdeljaber, Osama, Ince, Turker, Gabbouj, Moncef, Inman, Daniel J., 2019. 1d Convolutional Neural Networks and Applications: A Survey. arXiv preprint arXiv:1905.03554.
Kobsar, Dylan, Olson, Chad, Paranjape, Raman, Hadjistavropoulos, Thomas, Barden, John M., 2014. Evaluation of age-related differences in the stride-to-stride fluctuations, regularity and symmetry of gait using a waist-mounted tri-axial accelerometer. Gait Posture 39 (1), 553–557.
Kruse, Thibault, Basili, Patrizia, Glasauer, Stefan, Kirsch, Alexandra, 2012. Legible robot navigation in the proximity of moving humans. In: 2012 IEEE Workshop on Advanced Robotics and its Social Impacts (ARSO). IEEE, pp. 83–88.
Lara, Oscar D., Labrador, Miguel A., 2012. A survey on human activity recognition using wearable sensors. IEEE Commun. Surv. Tutorials 15 (3), 1192–1209.
Lasota, Przemyslaw, A., Fong, Terrence, Shah, Julie, A., others, 2017. A Survey of Methods for Safe Human-Robot Interaction. Now Publishers.
LeCun, Yann, Bengio, Yoshua, Hinton, Geoffrey, 2015. Deep learning. Nature 521 (7553), 436–444.
Li, Li, Martin, Tara, Xu, Xu, 2020. A novel vision-based real-time method for evaluating postural risk factors associated with musculoskeletal disorders. Appl. Ergon. 87, 103138.
Liang, Junwei, Jiang, Lu, Murphy, Kevin, Yu, Ting, Hauptmann, Alexander, 2020. The garden of forking paths: towards multi-future trajectory prediction. In: Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition, pp. 10508–10518.
Lim, Sol, D`Souza, Clive, 2019. Statistical prediction of load carriage mode and magnitude from inertial sensor derived gait kinematics. Appl. Ergon. 76, 1–11.
Lockhart, Jeffrey W., Weiss, Gary, M., Xue, Jack, C., Gallagher, Shaun, T., Grosner, Andrew, B., Pulickal, Tony T., 2011. Design considerations for the wisdom smart phone-based sensor mining architecture. In: Proceedings of the Fifth International Workshop on Knowledge Discovery from Sensor Data. ACM, pp. 25–33.
Lundberg, Scott M., Lee, Su-In, 2017. A unified approach to interpreting model predictions. In: Advances in Neural Information Processing Systems, pp. 4765–4774.
Luo, Yue, Coppola, Sarah M., Dixon, Philippe C., Li, Song, Dennerlein, Jack, T., Hu, Boyi, 2020. A database of human gait performance on irregular and uneven surfaces collected by wearable sensors. Sci. Data 7 (1), 1–9.
Matthis, Samir, Jonathan, Barton, Sean L., Fajen, Brett R., 2017. The critical phase for visual control of human walking over complex terrain. Proc. Natl. Acad. Sci. Unit. States Am. 114 (32), E6720–E6729.
Maurer, Uwe, Smailagic, Asim, Siewiorek, Daniel P., Deisher, Michael, 2006. Activity recognition and monitoring using multiple sensors on different body positions. In: International Workshop on Wearable and Implantable Body Sensor Networks (BSN'06). IEEE, p. 4.
Mehrizi, Rahil, Peng, Xi, Xu, Xu, Zhang, Shaoting, Li, Kang, 2019. A deep neural network-based method for estimation of 3d lifting motions. J. Biomech. 84, 87–93.
Menz, Hylton B., Lord, Stephen R., Fitzpatrick, Richard C., 2003. Age-related differences in walking stability. Age Ageing 32 (2), 137–142.

Muller, Antoine, Pontonnier, Charles, Robert-Lachaine, Xavier, Dumont, Georges, Plamondon, Andre, 2020. Motion-based prediction of external forces and moments and back loading during manual material handling tasks. Appl. Ergon. 82, 102935.

Muro-De-La-Herran, Alvaro, Garcia-Zapirain, Begonya, Mendez-Zorrilla, Amaia, 2014. Gait analysis methods: an overview of wearable and non-wearable systems, highlighting clinical applications. Sensors 14 (2), 3362–3394.

Norris, Michelle, Anderson, Ross, Kenny, Ian C., 2014. Method analysis of accelerometers and gyroscopes in running gait: a systematic review. Proc. Inst. Mech. Eng. P J. Sports Eng. Technol. 228 (1), 3–15.

Nweke, Henry Friday, Teh, Ying Wah, Al-Garadi, Mohammed Ali, Alo, Uzoma Rita, 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.

O'Loughlin, Jennifer L., Robitaille, Yvonne, Boivin, Jean-Francois, Suissa, Samy, 1993. Incidence of and risk factors for falls and injurious falls among the community dwelling elderly. Am. J. Epidemiol. 137 (3), 342–354.

Oxley, Jennifer, O'Hern, Steve, Burtt, Duane, Rossiter, Ben, 2018. Falling while walking: a hidden contributor to pedestrian injury. Accid. Anal. Prev. 114, 77–82.

Paszke, Adam, Gross, Sam, Massa, Francisco, Lerer, Adam, Bradbury, James, Chanan, Gregory, Killeen, Trevor, Lin, Zeming, Gimelshein, Natalia, Antiga, Luca, Desmaison, Alban, Kopf, Andreas, Yang, Edward, DeVito, Zachary, Raison, Martin, Tejani, Alykhan, Chilamkurthy, Sasank, Steiner, Benoit, Fang, Lu, Bai, Junjie, others, 2019. Pytorch: an imperative style, high-performance deep learning library. In: Wallach, H., Larochelle, H., Beygelzimer, A., dAlché-Buc, F., Fox, E., Garnett, R. (Eds.), Advances in Neural Information Processing Systems, vol. 32. Curran Associates, Inc., pp. 8024–8035

Powers, David Martin, 2011. Evaluation: from Precision, Recall and F-Measure to Roc, Informedness, Markedness and Correlation.

Ramachandran, Prajit, Zoph, Barret, Le, Quoc V., 2017. Searching for Activation Functions. arXiv preprint arXiv:1710.05941.

Randell, Cliff, Muller, Henk, 2000. Context awareness by analysing accelerometer data. In: Digest of Papers. Fourth International Symposium on Wearable Computers. IEEE, pp. 175–176.

Ravi, Daniele, Wong, Charence, Lo, Benny, Yang, Guang-Zhong, 2016. Deep learning for human activity recognition: a resource efficient implementation on low-power devices. In: 2016 IEEE 13th International Conference on Wearable and Implantable Body Sensor Networks (BSN). IEEE, pp. 71–76.

Reenalda, Jasper, Maartens, Erik, Homan, Lotte, Buurke, JH Jaap, 2016. Continuous three dimensional analysis of running mechanics during a marathon by means of inertial magnetic measurement units to objectify changes in running mechanics. J. Biomech. 49 (14), 3362–3367.

Ribeiro, Marco Tulio, Singh, Sameer, Guestrin, Carlos, 2016. "why should i trust you?" explaining the predictions of any classifier. In: Proceedings of the 22nd ACM SIGKDD International Conference on Knowledge Discovery and Data Mining, pp. 1135–1144.

Roggen, Daniel, Calatroni, Alberto, Rossi, Mirco, Holleczek, Thomas, Forster, Kilian, Troster, Gerhard, Lukowicz, Paul, Bannach, David, Pirkl, Gerald, Fer- scha, Alois, others, 2010. Collecting complex activity datasets in highly rich networked sensor environments. In: 2010 Seventh International Conference on Networked Sensing Systems (INSS). IEEE, pp. 233–240.

Sadeghian, Amir, Kosaraju, Vineet, Sadeghian, Ali, Hirose, Noriaki, Rezatofighi, Hamid, Savarese, Silvio, 2019. Sophie: an attentive gan for predicting paths compliant to social and physical constraints. In: Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition, pp. 1349–1358.

Schall Jr., Mark C., Fethke, Nathan, B., Chen, Howard, 2016. Working postures and physical activity among registered nurses. Appl. Ergon. 54, 243–250.

Schepers, Paul, den Brinker, Berry, Methorst, Rob, Helbich, Marco, 2017. Pedestrian falls: a review of the literature and future research directions. J. Saf. Res. 62, 227–234.

Scherer, Dominik, Muller, Andreas, Behnke, Sven, 2010. Evaluation of pooling operations in convolutional architectures for object recognition. In: International Conference on Artificial Neural Networks. Springer, pp. 92–101.

Shimazaki, Yasuhiro, Murata, Masaaki, 2015. Effect of gait on formation of thermal environment inside footwear. Appl. Ergon. 49, 55–62.

Simon, Sheldon R., 2004. Quantification of human motion: gait analysis—benefits and limitations to its application to clinical problems. J. Biomech. 37 (12), 1869–1880.

Sprager, Sebastijan, Juric, Matjaz B., 2015. Inertial sensor-based gait recognition: a review. Sensors 15 (9), 22089–22127.

Steven Eyobu, Odongo, Han, Dong Seog, 2018. Feature representation and data augmentation for human activity classification based on wearable imu sensor data using a deep lstm neural network. Sensors 18 (9), 2892.

Stevens, Judy A., Ballesteros, Michael F., Mack, Karin A., Rudd, Rose, A., DeCaro, Erin, Adler, Gerald, 2012. Gender differences in seeking care for falls in the aged medicare population. Am. J. Prev. Med. 43 (1), 59–62.

Strodthoff, Nils, Strodthoff, Claas, 2019. Detecting and interpreting myocardial infarction using fully convolutional neural networks. Physiol. Meas. 40 (1), 015001.

Su, Jimmy Li-Shin, Dingwell, Jonathan B., 2007. Dynamic Stability of Passive Dynamic Walking on an Irregular Surface.

Talbot, Laura A., Musiol, Robin, J., Witham, Erica K., Metter, E Jeffery, 2005. Falls in young, middle-aged and older community dwelling adults: perceived cause, environmental factors and injury. BMC Publ. Health 5 (1), 86.

Tang, Charlie, Salakhutdinov, Russ R., 2019. Multiple futures prediction. In: Advances in Neural Information Processing Systems, pp. 15424–15434.

Tao, Weijun, Liu, Tao, Zheng, Rencheng, Feng, Hutian, 2012. Gait analysis using wearable sensors. Sensors 12 (2), 2255–2283.

Tapia, Emmanuel Munguia, Intille, Stephen S., Haskell, William, Larson, Kent, Wright, Julie, King, Abby, Friedman, Robert, 2007. Real-time recognition of physical activities and their intensities using wireless accelerometers and a heart rate monitor. In: 2007 11th IEEE International Symposium on Wearable Computers. IEEE, pp. 37–40.

Tedesco, Salvatore, Barton, John, O'Flynn, Brendan, 2017. A review of activity trackers for senior citizens: research perspectives, commercial landscape and the role of the insurance industry. Sensors 17 (6), 1277.

Verma, Santosh K., Willetts, Joanna, L., Corns, Helen, L., Marucci-Wellman, Helen, R., Lombardi, David A., Courtney, Theodore K., 2016. Falls and fall-related injuries among community-dwelling adults in the United States. PloS One 11 (3), e0150939.

Weiss, Aner, Brozgol, Marina, Dorfman, Moran, Herman, Talia, Shema, Shirley, Giladi, Nir, Hausdorff, Jeffrey M., 2013. Does the evaluation of gait quality during daily life provide insight into fall risk? a novel approach using 3-day accelerometer recordings. Neurorehabilitation Neural Repair 27 (8), 742–752.

Weiss, Aner, Sharifi, Sarvi, Plotnik, Meir, van Vugt, Jeroen, P.P., Giladi, Nir, Hausdorff, Jeffrey M., 2011. Toward automated, at-home assessment of mobility among patients with Parkinson disease, using a body-worn accelerometer. Neurorehabilitation Neural Repair 25 (9), 810–818.

Xun, Guangxu, Jia, Xiaowei, Zhang, Aidong, 2016. Detecting epileptic seizures with electroencephalogram via a context-learning model. BMC Med. Inf. Decis. Making 16 (2), 70.