

Optimization of Wearable Sensor's Type and Location for Outdoor Running Terrain Classification

Gabrielle Thibault M.Sc Candidate ©
August 2022

Department of Kinesiology and Physical Education
McGill University
Montréal, Québec, Canada

Table of Contents

Abstract	IV
Abrégé	V
Acknowledgments	VI
Contributions of Authors	VII
List of Figures	VIII
List of Tables	VIII
1. Introduction	1
2. Literature Review	2
2.1 Inertial Measurement Units Sensors in Running	2
2.2 Human Activity Recognition with Wearable Sensors	3
2.3 Deep Learning with Inertial Measurement Units Sensors	4
2.3.1 Inertial Measurement Units Sensors	5
2.3.2 Multi-Channel Inputs	5
2.4 Terrain Classification with Inertial Measurement Units Sensors	6
3. Rational, Objective and Hypotheses	8
4. Methods	10
4.1 Participants	10
4.2 Experimental setting	10
4.3 Data Acquisition	10
A) Anthropometric measures and questionnaires	10
B) Sensor Positioning	11
C) Calibration	12
D) Data Collection	12
4.4 Data Analysis	14
4.4.1 Dataset description	14
4.4.2 Data Pre-processing Steps	14
Step A) Conversion and channel selection	15
Step B) Cut trials and Gait cycle segmentation	15
Step C) Normalization	17
Step D) Table extraction	17
Step E) Reshaping in python	17
4.4.3 Developing the CNN Model	18
Step A) Basic CNN model for sensor type/location testing	18

Step B) Tuning CNN models with validation set	19
5. Results	22
5.1 Impact of pre-processing steps	22
5.2 Impact of sensor combinations	23
5.3 Final CNN model architecture	23
5.4 Performance for Subject-wise and subject-dependent splits	25
5.5 Final testing	25
6. Discussion	26
7. Conclusion	30
8. Appendix	31
Annex A	31
Annex B	32
Annex C	34
9. References	36

Abstract

Understanding training effect in high-level running is important for performance optimization and injury prevention. This includes awareness of different running surface types (e.g. hard versus soft). Recent advances in body worn smart sensors and deep learning have shown great accuracy in classification of human activity recognition using convolutional neural networks (CNN) with accelerometer signals. To date, no study has done a thorough full-body analysis optimizing sensor type and location as the CNN input to classify surfaces. Hence, this study explored the use of seventeen inertial measurement unit (IMU) sensors to classify between grass and asphalt using a deep learning approach. Signals were collected from 40 participants while they were running on grass and asphalt. Multiple variations of CNN models were tested using KerasTuner to optimize the hyperparameters and model architecture for this specific task. Data were collected from casual runners running for 20 m, 80 times on each surface (Fig. 1). The IMU signals (acceleration, gyroscope, and magnetometer) were collected using the Xsens full-body system. Only acceleration and angular velocity were used in the following results. As part of the pre-processing steps, those signals were separated into gait cycles and then reshaped into the appropriate tensor format using MATLAB and python. This study showed the significant impact of trial segmentation into gait cycles before being fed into the CNN model. A 1D-CNN model was developed, and two different protocols were tested for the sub-training, validation, and testing splits. The first one was leave-n-subject-out where the different splits were made of completely different participants. The second one was the subject-dependent approach where all participants trials were shuffled and randomly separated between all splits. A thorough comparison of sensor locations was conducted to identify their optimal position for surface classification tasks. The model developed was able to classify with high accuracy using a subject-dependent approach for the two surfaces. In general, the acceleration combinations gave better classification performance than the angular velocity combinations, and the foot sensor had the best performance-number of sensor ratio (95.5% accuracy). This analysis could provide useful quantitative feedback to athletes and coaches in terms of running technique effort on varied terrain surfaces.

Abrégé

Il est important de comprendre l'effet de nos entraînements de course pour optimiser nos performances et prévenir les blessures. Cela inclut la prise de conscience des différents types de surface de course (par exemple, dure et souple). Les récentes avancées en matière de capteurs intelligents et d'intelligence artificielle ont montré de bonnes performances pour classifier entre différentes activités humaines à l'aide de signaux d'accéléromètres et de réseaux neuronaux convolutifs. À ce jour, aucune étude n'a fait une analyse détaillée sur l'optimisation des combinaisons de capteurs (types/emplacements sur tout le corps) pour classifier entre différentes surfaces de course. Cette étude a donc exploré l'utilisation de dix-sept capteurs d'unité de mesures inertielles pour classifier entre l'herbe et l'asphalte à l'aide d'une approche d'apprentissage profonde. Les signaux ont été recueillis auprès de 40 participants pendant qu'ils couraient sur les deux surfaces. De multiples variations d'architecture de modèle de réseaux neuronaux convolutifs ont été testées en utilisant KerasTuner pour optimiser les paramètres et l'architecture du modèle. Les données ont été recueillies auprès de coureurs occasionnels courant 20 m, 80 fois pour chaque surface (Figure 1). Les signaux d'unité de mesures inertielles (accéléromètre, gyroscope et magnétomètre) ont été collectés à l'aide du système Xsens. Seules l'accélération et la vitesse angulaire ont été utilisées dans les résultats suivants. Dans le cadre des étapes de prétraitement de signaux, ceux-ci ont été séparés en cycles de course, puis remodelés dans le format tensoriel approprié en utilisant MATLAB et python. Cette étude a démontré l'impact significatif de la segmentation des essais de course en cycles de course avant d'être introduit dans le modèle. Un modèle de réseaux neuronaux convolutifs a été développé et testé avec deux protocoles de séparation des trois sous-bases de données (apprentissage, validation et test). Le premier protocole sépare les ensembles de données avec des participants complètement différents. Le second mélangeait tous les cycles de course des participants puis faisait la séparation de manière aléatoire. En utilisant la 2^e approche de séparation, le modèle développé a été capable de classifier avec précision. Une analyse approfondie des emplacements des capteurs a été réalisée afin d'identifier les positions optimales pour classifier les surfaces. En général, les combinaisons d'accélération ont donné de meilleures performances de classification que ceux de la vitesse angulaire, et le capteur de pied s'est avéré être le meilleur capteur pour cette tâche (précision de 95.5 %). Cette analyse pourrait fournir un retour quantitatif utile aux athlètes et aux entraîneurs en termes d'effort de la technique de course sur des surfaces de terrain variées.

Acknowledgments

I would first like to thank my thesis advisor, Dr. David Pearsall for trusting me in a project that was different from what the lab's expertise initially was and giving me the support and guidance to conduct my masters even with the hard condition COVID-19 brought to the research department. I hope you have an amazing retirement. To my co-supervisor, Philippe Dixon, thank you for answering all my machine learning questions and giving me access to resources to develop my project. To Philippe Renaud, thank you for your help in collecting the data and always being available to chat and brainstorm on different issues I was encountering.

To my parents and my boyfriend, thank you so much for supporting me in my studies. The last couple of years have been difficult and having you by my side made all the difference in the world.

Thank you to Jenni, your help in introducing me to EMG sensors was very much appreciated. You were available to help me fix bugs any day even if we were not located in the same lab. Thank you to Vaibhav who helped me so much with the toolbox BiomechZoo, you spent so many hours with me answering all my questions and teaching me how to use the functions and for that I am very grateful. Thank you to my lab mates Laura and Kristin for spending so many hours helping me collect my data and thank you to Sam and Cait for making my experience in the lab much more enjoyable with brainstorming sessions or just with casual chats.

Contributions of Authors

Gabrielle Thibault, MSc candidate, Department of Kinesiology and Physical Education, McGill University, was responsible for the research design, processing and analysis of data, and writing of this thesis. The candidate's supervisor, David J. Pearsall, PhD, Associate Professor, Department of Kinesiology and Physical Education, McGill University, contributed to the research design and analysis of the data. The candidate's co-supervisor, Philippe C. Dixon, PhD, Assistant Professor, School of Kinesiology, Université de Montréal, contributed to the research design and gave support to develop the machine learning models.

Philippe J. Renaud, MSc., Department of Kinesiology and Physical Education, McGill University, helped with data collection, acting as the lab research assistant. Dr. Shawn Robbins, PhD, Assistant Professor, School of Physical and Occupational Therapy, McGill University was a member of the thesis committee.

List of Figures

Figure 1: Flow chart chain of events from initial contact to toe-off adapted from (Nicola & Jewison, 2012)	2
Figure 2: Flow chart of artificial intelligence different subsections.	4
Figure 3: Outdoor set up for data collection.	10
Figure 4: Participant with IMU sensors.	11
Figure 5: The 3 surfaces at park Rutherford: a) Grass b) Asphalt. Cones mark the distance and direction of the running paths followed used by subjects.....	12
Figure 6: Cone set up.	13
Figure 7: MVN software (avatar).	13
Figure 8: Flow chart for pre-processing steps.....	15
Figure 9: Peak Knee flexion for gait event identification.	16
Figure 10: Representative gait cycles for all participants for the acceleration/ angular velocity signals of the a) foot, b) lower leg, c) upper leg and d) pelvis.....	16
Figure 11: Tensor format for CNN input.	17
Figure 12: Model for premilary testing.....	18
Figure 13: 5-fold cross validation split organization.	21
Figure 14: Final CNN architecture flow chart.	24
Figure 15: Confusion matrix for the final CNN model using the feet sensor of a) accelerometer b) angular velocity signals.....	24

List of Tables

Table 1: Sensor combinations tested for acceleration and angular velocity.	19
Table 2: Tuning the CNN model.	20
Table 3: Impact of trial segmentation on signals on accuracy (%).	22
Table 4: Impact of signal normalization using all lower body sensors on accuracy (%).	23
Table 5: Maximum accuracies (%) for different sensor combinations.	23
Table 6: Performance analysis of the final model acceleration for the feet configuration.	24
Table 7: Performance analysis of the final model angular velocity for the feet configuration.	24
Table 8: Accuracies (%) for different splitting protocols.	25
Table 9: 5-fold cross validation using subject-dependent split.....	25
Table 10:Participant's information.....	32

1. Introduction

Running is a popular sport and recreational activity due to its accessibility and health benefits; however, running-related injuries are common. Consequently running gait analysis studies have analyzed biomechanical and physiological measures to identify relations to these common injuries (Benca et al., 2020). Earlier studies had previously been limited to indoor lab environments (treadmill) which do not necessarily represent the runners' natural over ground training and competing environments (Reenalda et al., 2016). Several studies have demonstrated the feasibility of accurate outdoor measures of runners' gait mechanics. In particular, recent advances in wearable smart sensor technologies used in combination with Machine Learning (ML) algorithms hold the promise of accurate, multidimensional, temporal data gait analysis in real-world environments (Abu-Faraj et al., 2015). ML may offer improved modelling of non-linear, biological processes such as step-by-step variations due to adaptations to changes in external terrain orientation, obstacles, surfaces, etc. This is relevant as previous research has found running gait variability in response to different surface terrains (S. J. Dixon et al., 2000). ML may allow us to predict surface conditions and/or irregular terrains while athletes are running.

With recent advancement in wearable and wireless technologies, it is now possible to analyze runners in their natural training and racing environments with the same precision as laboratory-controlled environment (Van Caekenbergh et al., 2013) (García-Pérez et al., 2013) (Reenalda et al., 2016). Furthermore, it is possible for elite to novice runners to use these devices to self-monitor and optimize their training effect (e.g. heart rate, stride length and stride rate, etc.). Another factor affecting training effect is the mechanical response of running terrain, which has a direct impact on the energy cost of a training session (Pinnington et al., 2005). Running terrain classification algorithms have shown to be feasible using acceleration (P. C. Dixon et al., 2019) and angular velocity (Worsey et al., 2021) signals. To date, no research has analyzed the optimal sensor location(s) and type(s) to identify running surface.

The main purpose of this study was to develop a method to classify between grass and asphalt by optimizing signal pre-processing steps, different signal type(s) and location(s) and the deep learning model.

2. Literature Review

2.1 Inertial Measurement Units Sensors in Running

Multiple types of wearable sensors are used in sports performance enhancement and injury prevention (Prosser & Schmidt, 1999) (Aroganam et al., 2019), such as Inertial Measurement Units (IMUs), surface electromyography (sEMG), pressure sensors, heart rate monitors, etc. Many studies have analyzed the running gait of athletes with different types of wearable sensors. The chain of actions in the running gait cycle requires different recruitment muscles, body segments and joints for each sub-phases that can be quantified with these technologies.

IMUs are a combination of a tri-axial gyroscope (angular velocity), accelerometer (linear acceleration) and magnetometer (strength and direction of the magnetic field), allowing for accurate tracking of motion activity and biomechanical analysis of activities that cannot be generalized by indoor laboratory-controlled studies (Hollis et al., 2021). Nicola et al. (2012) describes the various changes throughout the kinetic chain of a running gait cycle from initial contact to mid stance and from mid stance to toe off (Figure 1) (Nicola & Jewison, 2012).

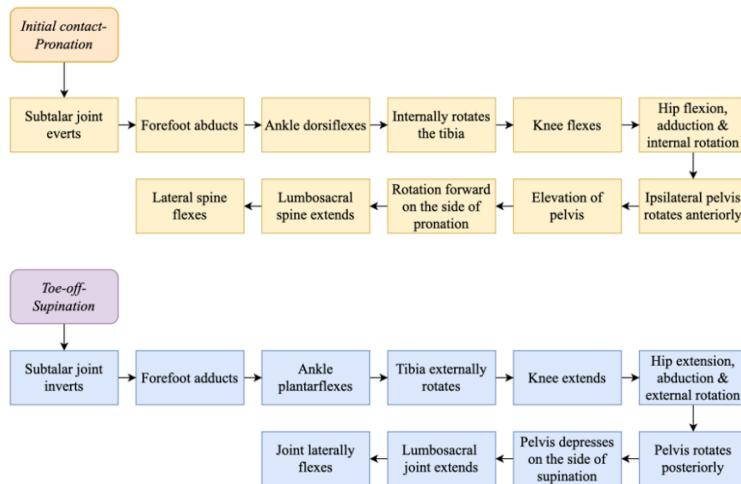


Figure 1: Flow chart chain of events from initial contact to toe off adapted from (Nicola & Jewison, 2012).

2.2 Human Activity Recognition with Wearable Sensors

Artificial intelligence analysis techniques are a very effective tool for bio-signal analysis which is most commonly nonlinear and complex (multidimensional). Machine Learning (ML) algorithms can analyze high-dimensional, temporal data with interdependent variables simultaneously. The advantage of ML models is that they can learn and adapt without following explicit instructions using statistics and probability concepts to map inputs to a subgroup based on previous knowledge.

Precise and accurate gait analysis often requires expensive technologies; however, recent advances in ML models allow for accurate, robust, flexible, and quick model conception for classification of biomechanical and physiological data at a low cost. The development of wearable sensors has facilitated the data collection of sport and outdoor daily activities due to its practicality and simplicity of use. This convenience in data collection allows for large databases to be created and, hence, more data can be fed into classification algorithms for better generalization of the population. Different sports have been using the combination of wearable sensors and artificial intelligence for performance analysis and injury prevention. Previous literature was mostly conducted in a laboratory-controlled context, but recently studies have taken place in different environments (Grzeszick et al., 2017) (Slim et al., 2019).

Many models have been developed to satisfy different classifications and regression problems. Feature-based models, such as decision tree, K-nearest neighbour (KNN), and Support Vector Machine (SVM) have been extensively used in many areas including Human Activity Recognition (HAR). Traditional ML (feature-based models) (Figure 2) requires feature engineering which is often time consuming and requires a good knowledge base on the specific domain of applications. Feature engineering can be very difficult and, in some cases, might not give the best classification performance (Remondini, 2018). Another subsection of machine learning is called deep learning (DL) (Figure 2) and, from previous surveys (Slim et al., 2019), it was concluded that it is now the preferred method for Artificial Intelligence (AI) analysis. Figure 2 summarizes different categories of ML.

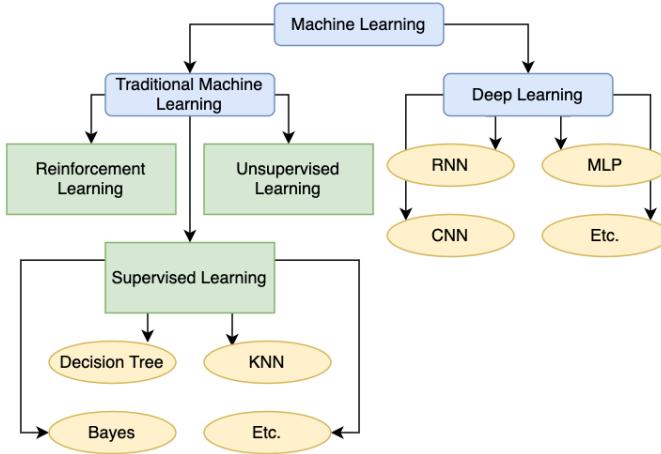


Figure 2: Flow chart of artificial intelligence different subsections.

KNN: K-nearest-neighbour, RNN: Recurrent neural network, CNN: Convolutional neural network, MLP: Multilayer perceptron

One of the DL approaches is called Convolutional Neural Networks (CNN). CNN networks allow multi-channel convolutions to automatically extract features from raw data and classifies the output with different activation functions instead of manually handcrafting features (Negi et al., 2020)(Moya Rueda et al., 2018). CNN is the most popular model for pattern recognition CNNs have shown great performance in HAR tasks by automatically and accurately detecting the patterns from different raw signals (Dehzangi et al., 2017) (Slim et al., 2019). The different layers of operations such as filters and pooling, allows to gain information on non-linear and temporal relations of human movement. The specific combination of convolutions, pooling, rectified linear units (ReLU) and fully connected layers will vary depending on the application (Slim et al., 2019)(Pinzón-Arenas et al., 2019). Convolution layers are made of filters applied on the input data to better understand the relationship between the different cells. Pooling layers are used to reduce the amount of information that the model must learn from, by summarizing the most important features. ReLU is a non-linear activation function that is often used in deep learning (Chollet, 2018).

2.3 Deep Learning with Inertial Measurement Units Sensors

As shown by Luo et al. (2020), with the rise of wearable sensors, it is now possible to collect data in different environments to maximize ecological validity. The data collected can then be used to predict and optimize different tasks in physical activities (Luo et al., 2020). To obtain good performance with deep learning models, it is important to have a large dataset. The greater number of participants the model can learn from, the better the model will be able to generalize to multiple subgroups. Having a dataset coming from different distributions can reduce the proportion of the variability caused by noise and participant differences. A good

classification algorithm will be able to generalize accurately between different variations, such as sessions, postures, subjects and more. Remondini et al. (2018) collected EMG signals from different subjects, on different days, with different postures to increase variability and the robustness of their 1D-CNN model. Single-session trainings were found to give lower validation accuracies compared to intersession (Remondini, 2018).

DL has been used in many different types of tasks already. The shift from feature-based models (traditional ML) to deep learning allow researchers to develop high-performance models without extensive a priori knowledge in the specific area of their classification (Remondini, 2018).

2.3.1 Inertial Measurement Units Sensors

Deep learning has also been used with the different signals from IMU sensors. The combination of acceleration, gyroscopes and magnetometer are often used for sports and daily activity classification tasks. CNNs are the most commonly used model with these sensors. Jang J et al. (2018) was able to classify eight different cross-country skiing techniques with the use of the Xsens full-body system (17 sensors) using ML (Jang et al., 2018).

Usually, IMU signals do not require extensive pre-processing steps except being organized into a specific matrix shape. With deep learning, the raw output signals, from the sensors, can be directly used as input for the models. For repetitive motion activities, such as running, trials can be separated into those repeated motions (i.e. gait cycles) to help the model learn a specific movement. The locations of the sensors are important because the movement analyzed will be different depending on the position of the IMU, which can affect the accuracy of task prediction (Slim et al., 2019).

2.3.2 Multi-Channel Inputs

Previous research in HAR has found that single channel inputs (one sensor) underperform compared to multi-channel data for classification tasks. Most research uses a sensor combination method with multi-channel inputs. The goal of sensor fusion is to combine multiple sources of information generated by different sensor types or locations to increase the classification accuracy. Combinations of sensor types such as IMUs, pressure sensors, EMG and more, are often used. Also, combinations of sensor location have been used with good results (Gao et al., 2020).

For multi-channel classification, often the different types of sensors will have different sample frequencies and they will need to be brought to a common sampling rate to be synchronized and

used as ML inputs. Chen et al. (2017) determined that modifying the sampling frequency did not significantly impact the accuracy of classification. More specifically, for a sampling rate of 400 Hz compared to 1000 Hz the accuracy only dropped 0.43% and 1.5% for 200 Hz with a KNN model (Chen et al., 2017).

2.4 Terrain Classification with Inertial Measurement Units Sensors

Gait analysis is the study of human locomotion. A walking or running gait cycle can be defined as the time between two ipsilateral foot strikes (Saunders et al., 1953). There have been multiple studies published on terrain comparison using the output of different wearable sensors. Previous studies have compared overground, and treadmill running and found several differences in the kinematics of participants; hence the importance of conducting sports studies in the natural competing and training environments of athletes.

Depending on the hardness of the running surface, different amplitude, frequency and velocity of muscle concentric and eccentric contractions will be observed (L. Wang et al., 2014) (Wu et al., 2016). García-Pérez et al. (2013) and Van Caekenberghe et al. (2013) have found several differences between overground and treadmill running; this included stride frequency, ground contact time, kinematics of lower limb joints, muscular recruitment, energy loss, shock absorption, and plantar pressure distribution(García-Pérez et al., 2013)(Van Caekenberghe et al., 2013).

Schütte et al. (2016) studied the variation in dynamic stability and loading between woodchip trails and concrete roads with a 3D accelerometer sensor located at the trunk. They found significant difference between the dynamic stability and loading of runners on both surfaces based on root mean square ratio, step/stride regularity and sample entropy(Schütte et al., 2016). Dixon et al. (2019), showed that raw acceleration data collected from 2 accelerometers (tibia and pelvis) could classify 3 outdoor surfaces (concrete road, synthetic track, woodchip trail) with high accuracies for both feature-based (97.0%) and deep learning models (96.1%) (P. C. Dixon et al., 2019). Finally, using an ankle wore IMU sensor, Worsey et al. (2021) were able to classify with overall accuracy of 96% between athletics track, soft sand, and hard sand using support vector machine with tenfold cross validation (subject dependent). However, when the test/train split was done subject-wise (leave-one-subject-out), the results to differentiate the track and hard sand were poor and the results between the soft sand and the two other hard surfaces were mediocre. The overall accuracy was 61% using this splitting approach. For high classification performance, calibration data should be included in the training data set for each individual athlete (Worsey et al., 2021).

Both single channel and multi-channel models were compared for Gupta R et. al (2019)'s study to determine which combination gives the highest accuracy. They found that multi-channel combinations had better classification accuracy to classify between level ground, ramp ascent/descent, and stair ascent/descent (Gupta & Agarwal, 2019). Similarly, Negi S et al. (2020) used a deep neural network to classify the same terrains using both EMG and IMU sensors on the Tibialis Anterior and gastrocnemius muscle with 97.55% and 91.22% accuracy during stance and swing phase (Negi et al., 2020).

3. Rational, Objective and Hypotheses

Previously, detailed running measurements were limited to indoor lab environments which do not necessarily represent the runners' natural training and competing environments. With the development of wearable sensors, a growing number of studies are conducted outside the lab. With technologies, such as smart watches, being more accessible, it is now possible for novice runners to practise their sport with access to performance enhancement tools. Running terrain classification algorithms have been proven to show great performance with accelerometer sensors (P. C. Dixon et al., 2019) (Worsey et al., 2021).

However, a full analysis of how many accelerometer and angular velocity sensors, and which locations are optimal has not been done. To our knowledge, algorithms in smart watches do not detect surface type while the user is running. From previous research, we know that it is possible to classify different outdoor surfaces using deep learning classification models and accelerometer sensors located at the pelvis and tibia (P. C. Dixon et al., 2019).

If possible, surface classification models could be embedded into wearable devices to monitor surface variations while the athlete is running in real time. Adding this feature to sport watches would increase the precision in the outputted training effect calculated by the device. The degree of surface stiffness affects the athlete's training (e.g. training on soft surfaces requires more effort but is less damaging to muscle and soft tissues while hard surfaces provide a stable transient base of support for the landing and push-off; but too hard may lead to cumulative tissue fatigue). Potentially, by better quantifying our training surfaces, we may better optimize our training by adapting our striding and foot landing technique. Surface terrain classification could give useful feedback for athletes and coaches in terms of selected running technique between varied terrains (P. C. Dixon et al., 2019). These findings may potentially help improve algorithms in smart watches to increase the accuracy of training effects. Better optimization of training could prevent athletes from developing injuries and/or increase performance (Khera & Kumar, 2020).

The main goal of this study was to develop a method to classify between grass and asphalt by optimizing signal pre-processing steps, different signal type(s) and location(s) and the deep learning model. Kinematic data were collected using the Xsens system (Xsens Technologies BV, Enschede, Netherlands) to record 3D body kinematics (acceleration and angular velocity) while participants were running over the two terrains.

I hypothesize that these data will allow deep learning models to identify how kinematics patterns vary due to the different surfaces. I also hypothesize that the acceleration signal at the foot will give the highest classification performance between the tested signals.

4. Methods

4.1 Participants

Forty adults, casual distance runners, were recruited for testing ('Casual' defined as a minimum of 15 km per week of running). All participants were injury free at the time of the trials. Male and female athletes between 18 and 50 years of age were recruited for this research. Subjects were recruited with word of mouth and using social media platform such has Facebook. Emails were also sent to students at the McGill university triathlon club and to running teams in the city of Montreal. From previous research, a minimum of thirty participants was set for this study (P. C. Dixon et al., 2019).

4.2 Experimental setting

This study was approved by the McGill University Ethics Board (REB File # 21-07-023) (Annex A). Testing occurred at the Rutherford Reservoir. Both running tested surfaces were within the same area and flat (0-degree slope), allowing us to keep the same outdoor set up for both surfaces (Figure 3). The Receiver for the Xsens (router) was connected to a laptop and a portable power supply source (Battery Jackery Explorer 1000).

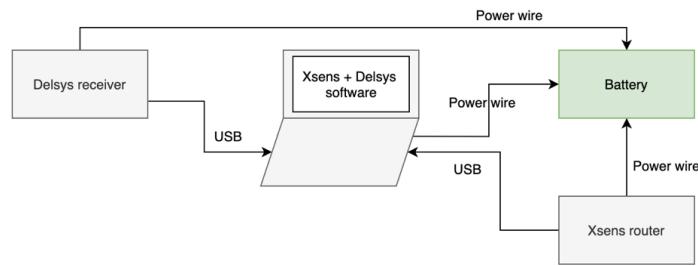


Figure 3: Outdoor set up for data collection.

4.3 Data Acquisition

A) Anthropometric measures and questionnaires

Anthropometric measures were taken before the trials to calibrate the Xsens system. More specifically, the participant's height, hip height/width, ankle width and knee height/width were taken. More details on the exact landmarks can be found in the MVN user manual (Xsens, 2021). A measuring tape was used to take these measures. A measuring tape was used to take these

measures. Additional questions (Annex B) were asked (e.g. gender, age, number of years running, number of runs per week, previous running injuries, type of shoe).

B) Sensor Positioning

The Xsens system's 17 sensors were positioned on the body according to Xsens specifications in the user manual and secured in position with medical tape (Xsens, 2021). The sensors were positioned on the head, sternum, pelvis, shoulders, upper/lower arms, hands, upper/lower leg and feet to collect 3D acceleration and angular velocity data (Figure 4).

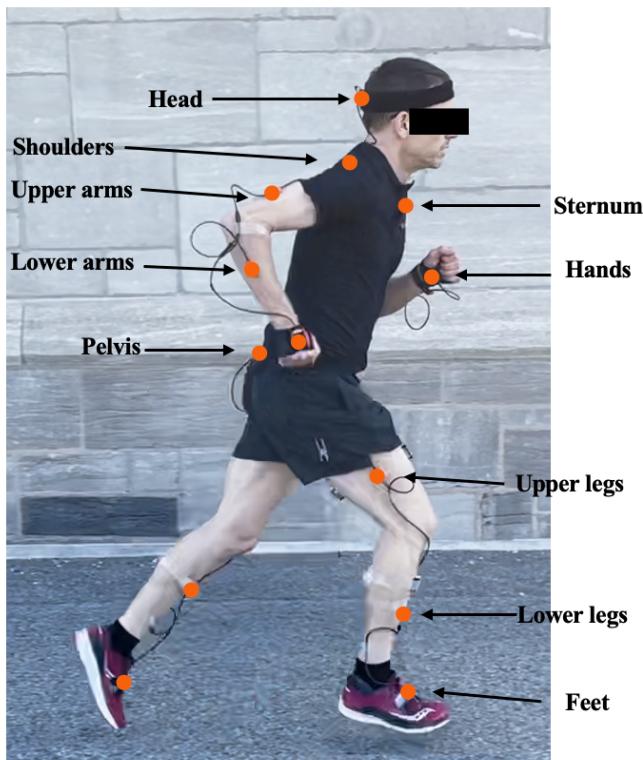


Figure 4: Participant with IMU sensors.

C) Calibration

To calibrate the Xsens system, the participants were asked to stand in a neutral position (i.e. standing with the feet close together and both arms hanging relaxed to each side of the body and the thumbs pointing forward). Then, the participants had to walk four metres forward and backwards and come back to their neutral position.

D) Data Collection

The participants were given time to complete a brief warm-up and, in the meantime, become more familiar with the sensors. All participants were asked to run on the two surfaces with a consistent individual sub-maximal pace (Figure 5a, b). Participants were asked to run at a comfortable pace of around 4.30 to 5.30 min/km (not sprinting).



Figure 5: The 3 surfaces at park Rutherford: a) Grass b) Asphalt. Cones mark the distance and direction of the running paths followed used by subjects.

Each condition was identified with cones (Figure 6). The trials were conducted between two cones separated by 20 m with the receiver/laptop in the middle to wirelessly transmit the data during the tests. A second pair of cones was positioned 2 m further from the ones separated by 20 m (Figure 6) to avoid the acceleration/deceleration phase in the data collected.

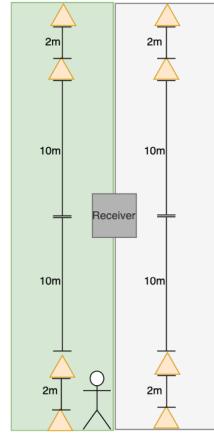


Figure 6: Cone set up.

Each participant had to run 80 times on each surface. The sessions lasted for approximately 1h30. To reduce the effect of fatigue on a specific surface the following protocol was followed:

- Run 40 times on asphalt
- Run 40 times on grass
- Run 20 times on asphalt
- Run 20 times on grass
- Run 20 times on asphalt
- Run 20 times on grass

The order of the surface's evaluation was established randomly. This would need to be studied, in later study, to determine if it affects the algorithm detection accuracy.

The MVN Xsens software was used to collect the IMU data from the Xsens full-body system (Figure 7).

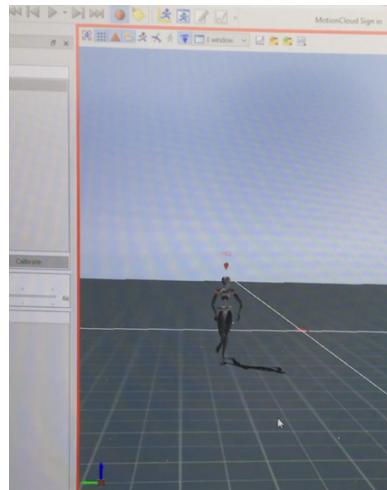


Figure 7: MVN software (avatar).

4.4 Data Analysis

4.4.1 Dataset description

Data were collected from 40 participants. The average age was 27.4 years old (7.8 SD) and the average running year was 10.5 ± 7.3 SD years. The percentage of women in the study was 43.9 % (See Annex B for information about each participant).

The data collected from the Xsens system were triaxial acceleration, triaxial angular velocity and magnetic fields at a frequency of 240 Hz (158 MB/min) with a dynamic accuracy of 1 deg RMS. Each sensor weight 10g and are relatively small (36 x 24.5 x 10 mm). The body joints' angle kinematics were determined based on those signals and used to identify gait cycles. More specifically, the MVN software creates quaternions with the ‘sensor fusion’ algorithm developed by Xsens (Schepers et al., 2018) to determine the orientation and position of each body segment. The joint angles are determined by subtracting the orientation of the distal and proximal segments and then converted to Euler angles (Sinclair et al., 2012).

Only the acceleration and angular velocity were used in this project as inputs for the CNN model. A total of 43,803 running gait cycles were collected divided into terrain surfaces of grass (49.38%) and asphalt (50.62%).

4.4.2 Data Pre-processing Steps

The IMU data collected were processed by the MVN Link system software (Xsens Technologies BV, Enschede, Netherlands). All joint angles, free acceleration and angular velocity data for all participants' trials were exported to spreadsheets. A total of 160 data sheets were exported for each participant.

Pre-processing steps were done with the raw signals using MATLAB (The Mathworks, Inc., Natick, USA) and the biomechZoo toolbox (P. C. Dixon et al., 2017). The main code for step A to D can be found in the function Main_preprocessing_IMU and will be explained below.

The final reshaping of signals into tensors was conducted in Python software (Python Software Foundation, <https://www.pyton.org/>) on Google's Colaboratory Pro+ GPU (GPU: 1xTesla P100, 54.8 GB RAM) (Steps E).

All the functions used in the bellow steps can be found in the GitHub
<https://github.com/Gab2697/Surface-classification-IMU>.

Figure 8 summarizes the pre-processing steps discussed in the next section of this thesis.

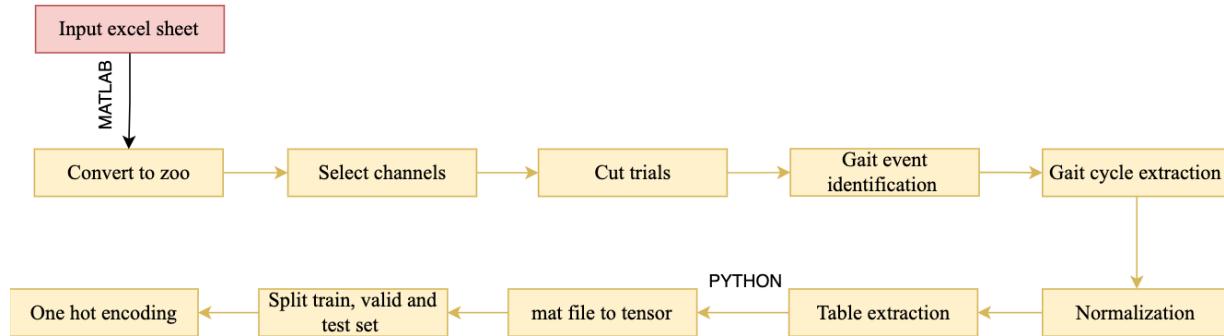


Figure 8: Flow chart for pre-processing steps.

Step A) Conversion and channel selection

The outputted Excel files from the data collection software (Xsens MVN) were converted to zoo files (function: Xsens2zoo) in MATLAB to use the functions in the toolbox BiomecZoo (P. C. Dixon et al., 2017). The acceleration, angular velocity channels that were kept are enumerated in Annex C. For the joint angles, only the right and left knee angle channels were kept.

Step B) Cut trials and Gait cycle segmentation

This was the first element that was tested in the results section of this thesis. Two different signal separation approaches were compared using the lower body sensors combination:

- The first one was a four second section of the trials without the acceleration and deceleration phases.
- The second one was to extract gait cycles from those trials and use them as the inputs for the model.

The separation of the trials into gait cycles was done in two steps using two different functions (function: gait_event_knee, outdoor_gait_cycle_data_Knee). The first function was used to create events to identify knee flexion at foot strike which is the first local minimum between each peak (P. C. Dixon et al., 2017). As it can be observed in Figure 9, peak knee flexions can easily be identified.

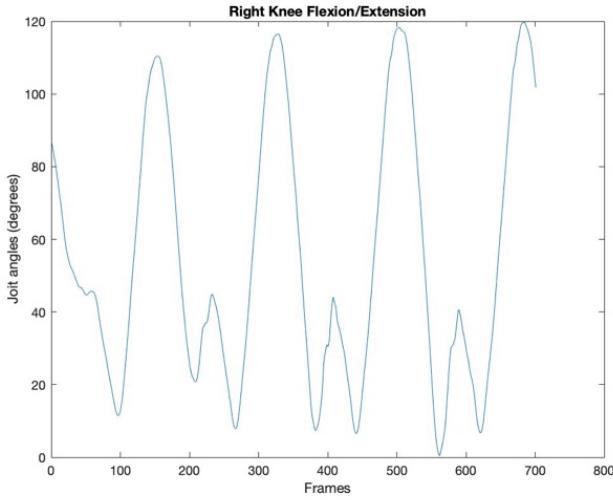


Figure 9: Peak Knee flexion for gait event identification.

The second function segmented the trials into gait cycles.

All gait cycles were time normalized to 101 sample points with a linear interpolation (function: `bmech_normalize`). Most of the gait cycles were down sampled.

To illustrate the similarity between the two surfaces the most representative gait cycle for each participant and each surface were selected. These gait cycles were selected if they had the closest root mean square error to the mean trial. Figure 10 shows the average gait cycle and standard deviation (in gray) for lower body sensors.

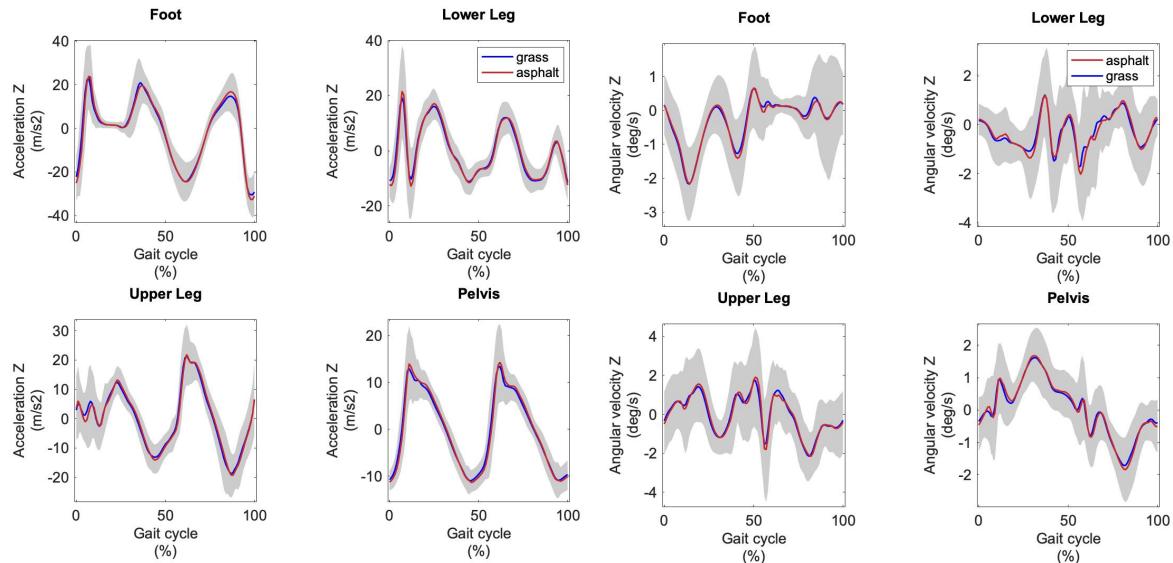


Figure 10: Representative gait cycles for all participants for the acceleration/ angular velocity signals of the a) foot, b) lower leg, c) upper leg and d) pelvis.

Step C) Normalization

The second pre-processing step that was evaluated in the result section is the impact of max-normalization on the input signals. To do so, all signals were normalized using the max value of each gait cycles to bring them to a common scale between minus one and one (function: normMax_data).

Step D) Table extraction

In this last step, two columns were added at the end of the table for the surface type and the participant number, respectively (function: extract_filestruct). Finally, table data were exported to a .mat file (function: table2struct) for further processing in python.

Step E) Reshaping in python

One main function was used for all the steps conducted in python

(CNN_SurfaceClassification_IMU) and this code can be found in the GitHub

<https://github.com/Gab2697/Surface-classification-IMU>.

First the .mat files were loaded into python and converted into the correct tensor shape demonstrated in Figure 11, which is #trials × #frames × # channels (function: mat_to_tensor). The labels were one hot encoded (function: one_hot) which means that instead of trials being linked to the word grass or asphalt they were labelled to zero and one respectively.

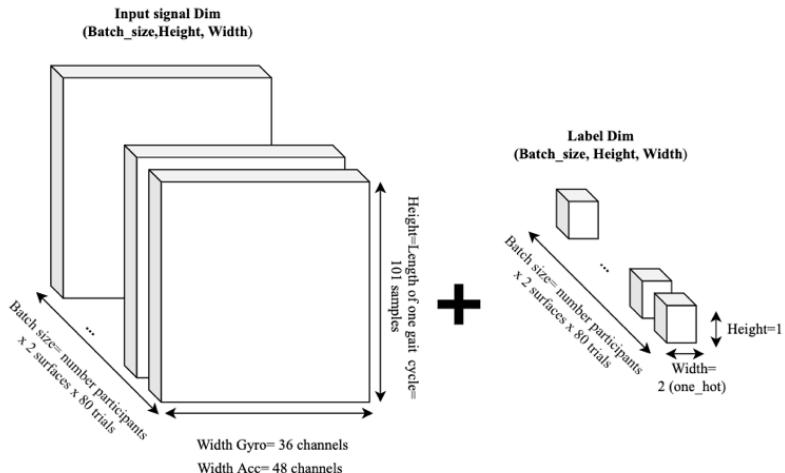


Figure 11: Tensor format for CNN input.

4.4.3 Developing the CNN Model

The following steps were conducted with Python software (Python Software Foundation, <https://www.python.org/>) on Google's Colaboratory Pro+ GPU (GPU: 1xTesla P100, 54.8 GB RAM). Machine learning and data analysis Python packages (e.g., Tensorflow, PyTorch, Numpy, Scipy, Scikit-learn, Pandas) were used for the deep learning task. One main function was used for all the following steps (CNN_SurfaceClassification_IMU) and this code can be found in the GitHub <https://github.com/Gab2697/Surface-classification-IMU>.

Step A) Basic CNN model for sensor type/location testing

Figure 12 demonstrates the initial basic model that was used to determine which sensor combination is optimal for this classification task.

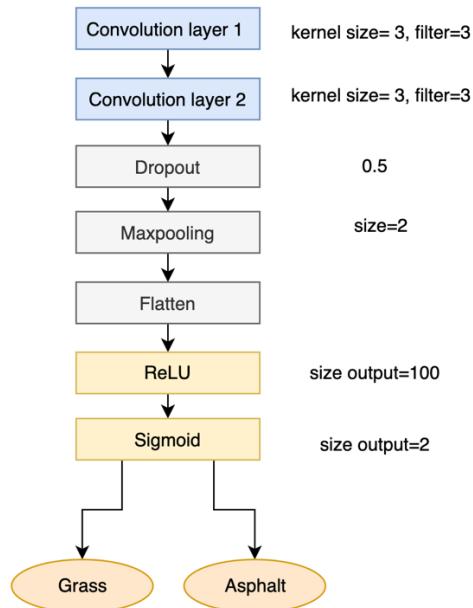


Figure 12: Model for preliminary testing.

Functions were created to test different combinations of sensors. Temporary files were created to save the two main datasets:

- Acceleration (list of all signals in Appendix C)
- Angular velocity (list of all signals in Annex C)

Only the subject-dependent approach was tested for these four sensor combinations (Table 1). All four-sensor combinations were tested for the acceleration and angular velocity signals. The sensor combinations were determined based on initial testing. After obtaining

preliminary results showing a better performance of the model with only the lower body, it was concluded that a deeper analysis would be conducted on the individual sensors of that body segment. Only 12 out of the 17 sensors were used in this analysis for convenience reasons in the pre-processing steps.

Table 1: Sensor combinations tested for acceleration and angular velocity.

Number of triaxial sensors	Combinations
12	Full body (Head, left and right shoulder, left and right upper arm, left and right hand, pelvis, left and right lower leg, left and right foot)
5	Lower body (pelvis, left and right lower leg, left and right foot)
1	Pelvis
2	Feet (left and right foot)

Step B) Tuning CNN models with validation set

First, one to four convolutional layers were tested to determine the optimal general CNN architecture. Then, three different optimizers (Adam, RMSprop and SGD) and different batch sizes were evaluated for this classification task.

Second, the model hyperparameters were tuned, using KerasTuner and a callback function for early stop (using the validation loss with a patience of 50). The learning rate, number of filters, kernel size, dropout and regularization ratio were tuned. The regularization parameter was also initially tested with KerasTuner but then removed due to lower performance on the validation accuracy when included in the model. The following steps were conducted using the optimal combination of sensors found in the previous section. The parameters evaluated can be found in Table 2.

Table 2: Tuning the CNN model.

Hyperparameters	Options
Epochs	Using callback for early stop (patience 50)
Batch size	50, 100, 200 and 300
Optimization function	Adam, RMSprop, SGD
Learning rate	From 0.0001 to 0.01
Model architecture	Options
Number convolutional layers	From 1 to 4
Filter number	From 32 to 256 (step 32)
Kernel size	From 3 to 5 (step 1)
Dropouts	From 0 to 0.5 (step 0.1)
Regularization	L1, L2, L1_L2

Step C) Train, validation, and test split

Two different splitting approaches were tested in this thesis: Subject-wise and Subject-dependent split. The subject-wise split (leave-n-subject-out) separates the datasets with different participants in train, test, and validation sets (inter-subject split). The subject-dependent split shuffled all trials before separating the three dataset subsections (intra-subject split) (Funciton: `subject_wise_split`).

Step D) Final model evaluated with testing set

Both the acceleration and angular velocity were tested with the final optimized model using the best sensor combination from Table 1. Precision (Equation 1), recall (Equation 2) and f1-score (Equation 3) were obtained for both surfaces.

$$\text{Equation 1: } \text{Precision} = \text{True Positive} / (\text{True Positive} + \text{False Positive})$$

$$\text{Equation 2: } \text{Recall} = \text{True Positive} / (\text{True Positive} + \text{False Negative})$$

$$\text{Equation 3: } \text{F1-score} = 2 \times (\text{Precision} \times \text{Recall}) / (\text{Precision} + \text{Recall})$$

The precision demonstrates how precisely the model was able to correctly classify a specific surface. This is often useful when we want to minimize false positive. Recall shows the proportion of times the model accurately classified a surface out of all the positive prediction. This is often useful when we want to minimize false negative. The F1-score is a weighted

average of the precision and recall (Pedregosa et al., 2011). This metric will be useful to quantify the performance of the final model.

The final model with the optimal sensor combination, splitting approach and pre-processing steps, was tested with a 5-fold cross validation approach using the testing set (Figure 13).

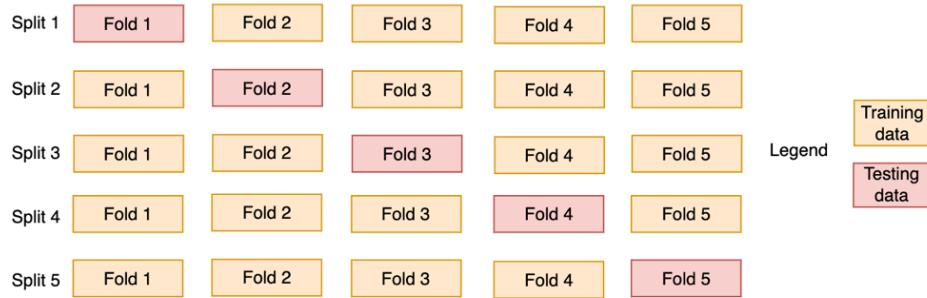


Figure 13: 5-fold cross validation split organization.

5. Results

In the next section, results comparing the impact of different pre-processing steps, sensor combination and dataset splitting approaches are summarized in multiple tables. It was observed that both the acceleration and the angular velocity signal patterns for the two tested surfaces (grass and asphalt) were very similar. The p-values (correlation between waveform) were calculated between the two surfaces, averaging all gait cycles for specific channels. The calculated values were high, which translate in low difference between the two groups means (e.g., for acceleration-Z; Pelvis: 0.9627, Upper leg: 1.000, Lower leg: 0.9926, Foot: 0.9926). If softer (compliant) grass would have been used, the classification model would have probably given even higher results.

5.1 Impact of pre-processing steps

Two different steps of the pre-processing section were evaluated: the impact of *Gait cycle extraction* and *Normalization*. The model in Figure 12 was used to test the different pre-processing steps for both the acceleration and angular velocity signals.

Table 3 compares different signal separation approaches using the lower body sensor combination. The accuracies were obtained using a subject-dependent splitting approach. The *4 sec. trial* method used the full trial (without the acceleration/ deceleration phase) as the model input. The *Gait cycle* method separated those trials into gait cycles and used those signals as the model inputs.

Table 3: Impact of trial segmentation on signals on accuracy (%).

Epoch= 100, batch size=200, lower body sensors, using the validation set with the model from figure 12.

Signal	Input length	Accuracies
Acceleration	4 sec. trial	63.58
	Gait cycle	91.79
Angular velocity	4 sec. trial	56.04
	Gait cycle	87.61

The second preprocessing step evaluated is the impact of normalization of signals using their max value for each gait cycles (Table 4). As it can be observed, this step had a minimal impact on the accuracy (decreased < 2%). Hence, normalization of data was deemed unnecessary for this project.

Table 4: Impact of signal normalization using all lower body sensors on accuracy (%).

Epoch= 100, batch size=200, lower body sensors, using the validation set with the model from figure 12.

Signal	Normalization	Accuracies
Acceleration	With max normalization	89.25
	Without max-normalization	91.83
Angular velocity	With max normalization	87.01
	Without max-normalization	87.61

Given the above results in trial segmentation, only the *Gait cycle* approach was used given its higher classification performance.

5.2 Impact of sensor combinations

The model described in Figure 12 was used to test different sensor combinations for both the acceleration and angular velocity signals. Four-sensor combinations were tested in

Table 5. The accuracies were obtained using a subject-dependent splitting approach.

Table 5: Maximum accuracies (%) for different sensor combinations.

Epoch= 100, batch size=200, using the validation set with the model from figure 12.

	Full-body system	Lower body	Pelvis	Foot
Acceleration	91.10	92.20	79.03	92.07
Angular Velocity	88.26	86.39	78.20	86.72

Hence, moving forward the foot sensors' acceleration and angular velocity signals were used for terrain surface identification.

5.3 Final CNN model architecture

As noted in the data analysis section (Table 2), different parameters were tuned to develop the optimal model for this task using KerasTuner (https://keras.io/keras_tuner/). The final model (Figure 14) was found using a subject-dependent split with the acceleration and angular velocity foot sensors, separately, as input.

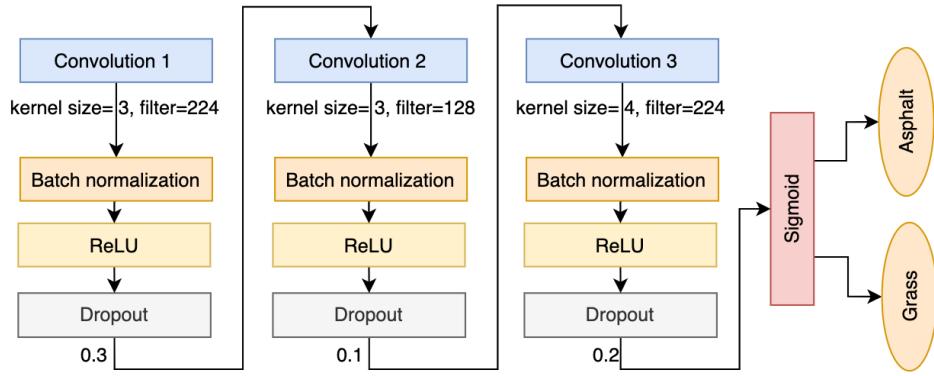


Figure 14: Final CNN architecture flow chart.

Using the final model, different metrics were evaluated for the acceleration and angular velocity feet sensors (Table 6 and

Table 7) from Equation 1, Equation 2, Equation 3.

Table 6: Performance analysis of the final model acceleration for the feet configuration.

Epoch= 500 (callback), batch size=200, feet sensors, using the validation set with the model from figure 14

Surfaces	Support	Precision	Recall	F1-Score
Grass	1429	0.9706	0.9481	0.9481
Asphalt	1400	0.9685	0.9214	0.9444

Table 7: Performance analysis of the final model angular velocity for the feet configuration.

Epoch= 500 (callback), batch size=200, feet sensors, using the validation set with the model from figure 14

Surfaces	Support	Precision	Recall	F1-Score
Grass	1429	0.9382	0.8810	0.9087
Asphalt	1400	0.8857	0.9407	0.9124

In Figure 15, the confusion matrix for both the acceleration and angular velocity (from Table 6 and

Table 7) can be observed.

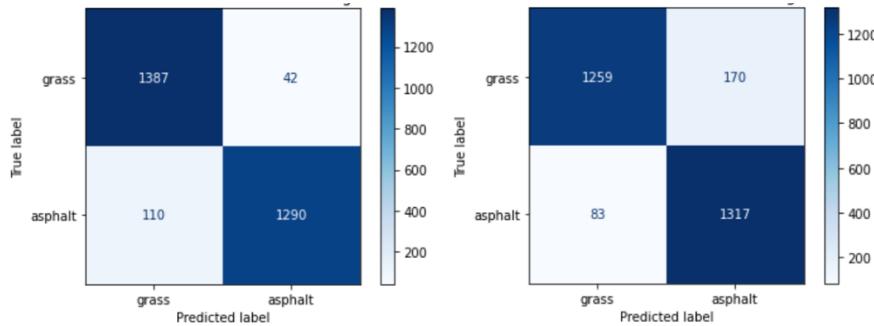


Figure 15: Confusion matrix for the final CNN model using the foot sensors of a) accelerometer b) angular velocity signals.

The maximum accuracies for both signal types were 95.71% (loss of 0.16) for the foot acceleration sensors and 91.91% (loss of 0.23) for the foot angular velocity sensors. Also, all F1-score from table 6 and 7 were higher than 90% which can convert in high model performance.

5.4 Performance for Subject-wise and subject-dependent splits

As previously mentioned, two different splitting protocols were tested. The first one being the subject-wise approach which separated the training, testing and validation sets with completely different participants in each sub-dataset. The second was the subject-dependent approach, which was conducted by shuffling all participants gait cycles before the training, testing and validation splits were done. In

Table 8, a comparison of the performance for the two splitting approaches can be found, using the foot sensors.

Table 8: Accuracies (%) for different splitting protocols.

Epoch= 500 (callback), batch size=200, feet sensors, using the validation set with the model from figure 14.

	Split protocol	Accuracies
Acceleration	Leave-n-subject-out	82.32
	Subject-dependent	95.71
Angular velocity	Leave-n-subject-out	57.58
	Subject-dependent	91.91

As can be observed, some calibration data for each user is necessary to obtain high classification accuracies.

5.5 Final testing

The final model with the optimal pre-processing steps, sensor combination, and splitting approach was tested with a 5-fold cross validation using the testing set (Figure 13). In Table 9, the summary of results for the 5-fold cross-validation can be found.

Table 9: 5-fold cross validation using subject-dependent split.

Epoch= 200, batch size=200, feet sensors, using the testing set with the model from figure 14.

	Accuracy	Loss
Fold 1	94.94	0.144
Fold 2	95.46	0.145

Fold 3	95.29	0.137
Fold 4	94.01	0.195
Fold 5	94.77	0.163

6. Discussion

Two different pre-processing steps and four different sensor combinations were tested for the acceleration and angular velocity signals. Two different splitting approaches, separating the dataset into training, validation, and testing sets, were also compared. As hypothesized, the acceleration and angular velocity signals were able to successfully classify surfaces. In general, the acceleration signals gave better results than the angular velocity for the accuracy, precision, recall and f1-score. The foot-sensor acceleration configuration was found to be the best location/type of sensors and achieved a maximum classification accuracy of 95.5% using a subject-dependent split approach.

The protocol to develop the final model was inspired by Chollet's method (Chollet, 2018) and the optimization of parameters was done using the KerasTuner given the high performance obtained in previous research using this tool (Joshi et al., 2021). Monitoring the training and the validation loss is a very useful tool to determine when overfitting is reached by the model. When the performance of the validation set starts decreasing, the minimum error rate of the validation data is reached, and the model starts overfitting (Chollet, 2018). In the CNN model developed for this project, a callback for early-stop was added to monitor the validation loss (or the testing loss) and prevent overfitting. More specifically, the training stopped when the validation loss did not improve for 50 epochs.

As previously mentioned, machine learning can be separated into two main subsections: deep learning and features based leaning (traditional machine learning). The disadvantage of the traditional machine learning models is the feature engineering task which requires domain specific expertise. A main advantage of deep learning models is that they do not require handcrafted features given that model itself can automatically extract features from the data. Deep learning, more specifically CNN models, has shown great performances in human activity recognition tasks (HAR) (Alharthi et al., 2019) which is why they were selected for this project.

Two pre-processing steps were tested to ensure the highest classification performance. First, separating the trials into gait cycles increased the accuracy of prediction, giving a simpler and more repetitive pattern for the model to learn from (Table 3). The surfaces tested were very similar which increases the difficulty of finding specific patterns for each task. This could explain why separating the trials by gait cycle had such a significant impact on the performance

of the model. Cutting the full trials to four seconds, to remove the acceleration/deceleration phases, gave us input signals starting at different phases of the gait cycle. This makes it harder for the model to learn a pattern that fluctuates greatly over each participant and slightly over each condition (Zheng et al., 2018). Second, it is also a common preprocessing step to standardize the input signals by bringing them to a common scale (amplitude normalization) (Gu et al., 2021). In this thesis, we analyzed the impact of max-normalizing and found that, for this specific task, it did not have a big impact on the accuracy. However, the speed of training decreased slightly when removing this pre-processing step. In fact, the accuracy of prediction decreased by 2% when the signals were normalized. Thus, the input signals were not normalized during pre-processing, but batch normalization was used directly in the final CNN model to stabilize and speed up the training phase.

Afterwards, to develop the optimal model for our classification task, the CNN architecture and multiple hyperparameters were tuned. For the general architecture of the model, one to four layers of convolutions were tested. Reducing the size of the network decreases the number of parameters the model will learn which increases the training speed. However, the more complex the pattern is to train on, the more layers the model might need to perform at full capacity (Chollet, 2018). For this task, three layers gave the best result. Furthermore, different numbers of filters and kernel sizes were tested using KerasTuner to find the optimal numbers for the task. The exact values are presented in Figure 14. Also, three different optimizers were evaluated (SGD, Adam, RMSprop) with different learning rates. The Adam optimizer gave the best results followed by RMSprop. The learning rate was determined using KerasTuner. All tested values can be found in Table 2. When developing a deep learning model, one of the most common methods to prevent overfitting is to add dropout layers and to add regularization on the weights of the model. Adding dropout consists of randomly setting to zero output features after the specified layers while the model is training. This technique introduces noise to remove patterns that are not significant. Dropout rates were determined using KerasTuner with values from 0 to 0.5. Adding regularization puts constraints on the weights of the model, to regularize the standard deviation between them (Chollet, 2018). For this specific task, adding regularization did not increase the validation accuracy, and so it was not included in the convolutional layers.

As mentioned previously, different sensor combinations were tested. First, the all-body sensor configuration and lower-body sensor configuration were compared to determine the impact of removing the upper body sensors. It was concluded that the model performed better when only the lower body sensor data was used vs using both the lower body and upper body sensor data (92.2% and 91.1% for acceleration, respectively). Second, the lower body sensors were

individually tested. The pelvis sensor gave the lowest results and the foot sensors the highest results (92% and 79% for acceleration, respectively) (Table 5). The foot sensors are the closest sensors to the surface tested which makes them more susceptible to small movement caused by terrain variations. Also, signals located on more proximal segments are experiencing damping and amplitude attenuation provided by muscle and distal joint crossings. Dixon et al. (2019) also obtained better results with an accelerometer sensor placed on the tibia, then with a sensor placed on the lower back using a CNN model for running surface classification (P. C. Dixon et al., 2019).

Two different splitting approaches were tested in this thesis: subject-wise (inter-subject) and subject-dependent (intra-subject). Similar to previous research, the results using the subject-dependent split were significantly better (Shah et al., 2022). Worsey et al. (2021) compared athletics track, soft sand and hard sand using an ankle accelerometer and obtained acceptable results using the subject-wise approach. They saw significant improvement when some calibration data was given to the model before evaluating it on the test set (subject-dependent). Using the subject-wise protocol, our model classified between grass and asphalt with an overall accuracy of 81% with the acceleration-foot sensors and Worsey et al. (2021) obtained $\geq 61\%$ wearing an ankle accelerometer. One of the differences between these studies is that we collected from 40 participants and the study conducted by Worsey et al. (2021) only had 6 participants. This may have had an impact on the generalization of the model and thus, the accuracy of prediction. The subject-dependent splitting protocol gave high performance classification results for this project and Worsey et al. (2021)'s study (95.5 % and 96%, respectively). Even using the subject-dependent approach (Worsey et al., 2021) observed that the performance was lower when classifying between the two hard surfaces (athletics track and hard sand). The more similar the patterns that we want to classify, the harder it is for the model to find differences between the subgroups. In this study, the surfaces tested were very similar (e.g., P-value for lower body sensors vertical acceleration was between 0.9627 and 1.000).

One of the strengths of this study was the sample size. Previous work had fewer participants and fewer trials (J. Wang et al., 2019). This dataset provides 80 trials on grass and asphalt for 40 participants (6400 trials) separated into multiple gait cycles (43,803). Also, the gender representation was evenly separated with a percentage of women and men of 43.9 % and 56.1%, respectively. Most of the participants ran on different days, which increased the ability to generalize to different temperatures (heat/cold, wind strength, dry/wet grass, etc.) of the model.

Despite the strong performance of the models, there were some limitations within the current study. First, it did not fully represent a complete running session where an athlete would accelerate/decelerate and switch between surfaces. The short distances of the trials might have

introduced fatigue which could have had an impact on the running patterns. Adding more variations in the patterns would have made it harder for the model to learn the specific patterns for each surface but would have made it more representative of real-world situations. Second, in this study, we only included the two most common running surfaces. However, there are plenty more that could be encountered (snow/ice, sand, gravel, trail, etc.). This research can be used to determine which sensor combination gives the best results to classify between running on grass and asphalt and could possibly be generalized to other surfaces. Third, this research may not generalize to injured people, people running more/less than 15 km/week, or runners younger than 18 or older than 50 years old. Finally, the grass area, used for the trials, was selected to increase the communication range between the surface electromyography (sEMG) Delsys sensors and the receiver by being close to a long building wall. The data collected by the sEMG sensors were not used in this project. After conducting all the trials around the same area, the grass became more compact and harder which possibly made the two surfaces more similar.

7. Conclusion

This study provides a detailed full-body analysis using two different sensor types to optimize the classification between the two most common running surfaces. Different pre-processing steps were tested in this study showing the improvement in accuracy when segmenting trials into gait cycles and the small impact of normalization. Four different sensor combinations were tested with both the acceleration and the angular velocity signals to determine which combination would allow us to predict the type of surface with the highest accuracy. The acceleration combinations gave the best general classification performance, and the foot sensors were found to be the best combination for a performance-number of sensor ratio. Finally, two different training, validation and testing splits approaches were compared to quantify the necessity of some calibration data for each participant before being able to classify with high performance. The model developed was able to classify with high accuracy using a subject-dependent approach for two similar surfaces (hard grass and asphalt).

Future research should confirm that these findings, about the foot sensors being the best location, can be generalized to all other running surfaces. Also, in this research acceleration and angular velocity were compared separately but not combined as the CNN input. It could be interesting to see if better results could be obtained with combination of types and locations of sensors. Similarly, all 3 axes of each sensor were always when inputted into the models. Further analysis could be done to determine if a specific axis could give better classification performance. Furthermore, current work is being done on the use of electromyography sensors as an input signal for this classification task. Comparing the use and combination of even more wearable sensor types could increase our knowledge on surface prediction for wearable technology applications.

8. Appendix

Annex A

**Research Ethics Board Office**

James Administration Bldg.
845 Sherbrooke Street West. Rm 325
Montreal, QC H3A 0G4

Tel: (514) 398-6831

Website: www.mcgill.ca/research/research/compliance/human/

Research Ethics Board 2
Certificate of Ethical Acceptability of Research Involving Humans

REB File #: 21-07-023

Project Title: Optimization of Wearable Sensors combination for Outdoor Running Terrain Classification.

Principal Investigator: Gabrielle Thibault

Department: Kinesiology and Physical Education

Status: Master's Student

Supervisor: Professor David Pearsall

Co-Supervisor: Dr. Philippe Dixon, Université de Montréal

Approval Period: September 8, 2021 – September 7, 2022

The REB 2 reviewed and approved this project by delegated review in accordance with the requirements of the McGill University Policy on the Ethical Conduct of Research Involving Human Participants and the Tri-Council Policy Statement: Ethical Conduct for Research Involving Humans.

Georgia Kalavritinos
Ethics Review Administrator

- * Approval is granted only for the research and purposes described.
- * Modifications to the approved research must be reviewed and approved by the REB before they can be implemented.
- * A Request for Renewal form must be submitted before the above expiry date. Research cannot be conducted without a current ethics approval. Submit 2-3 weeks ahead of the expiry date.
- * When a project has been completed or terminated, a Study Closure form must be submitted.
- * Unanticipated issues that may increase the risk level to participants or that may have other ethical implications must be promptly reported to the REB. Serious adverse events experienced by a participant in conjunction with the research must be reported to the REB without delay.
- * The REB must be promptly notified of any new information that may affect the welfare or consent of participants.
- * The REB must be notified of any suspension or cancellation imposed by a funding agency or regulatory body that is related to this study.
- * The REB must be notified of any findings that may have ethical implications or may affect the decision of the REB.

Annex B

Table 10: Participant's information

No	Gender	Age	Years of Running	Runs/week	Injuries last year	Shoe type	Dominant leg	Weather during trial
1	F	32	14	5	NA	Pegasus nike	R	Cloudy 12°
2	M	46	21	4	NA	Saucony convarez	R	Sunny 9°
3	M	24	5	3	Ankle pain 3 months ago	New balance Freshfoam 880	R	Cloudy 10°
4	F	30	15	7+	NA	Adidas Adizero6	R	Cloudy 15°
5	F	32	10	2	NA	Asics Kayano 27	R	Cloudy 17°
6*	M	28	15	4	Stress fracture foot	NB Freshfoam 1080	R	Sunny 19°
7	M	24	10	3	PFPS 4 months ago	Saucony	R	Sunny 10°
8	F	49	30	3	NA	Asics gel nimbus 19	R	Sunny 21°
9	M	49	35	3	Left calve 6 months ago	Asics	R	Sunny 21°
10	F	23	4	2	NA	Nike zoom	R	Sunny 14°
11	M	26	7	2	NA	ON cloud	R	Sunny 21°
12	F	21	1	2	NA	Saucony	R	Sunny 10°
13	M	46	11	4	NA	sSucony type A	R	Cloudy 15°
14	M	23	18	4	NA	Nike soccer	R	Cloudy 15°
15	M	23	5	3	NA	Mizuno Men's Wave Rider 22	R	Cloudy 19°
16	F	21	11	2	NA	Salomon speed cross 5GTX	L	Cloudy 8°
17	M	20	3	1	Knee cap	Nike zoom	R	Sunny 10 °
18	M	29	15	6	NA	Merrell minimalist	R	Cloudy 10°
19	M	21	10	1	NA	Kalenji	R	Cloudy 10° (raining)
20	M	26	hockey 20 years	5	NA	Nike air	R	Cloudy 9°
21	M	29	10	3	NA	Oka rincon2	R	Sunny 16°
22	M	22	8	3	Knee cap pain (last winter)	Saucony guide 13	L	Cloudy 10°
23	F	24	9	1	NA	Salomon speed cross	R	Cloudy 9°
24	M	28	8	06-Jan	Knee	Saucony fast switch	R	Cloudy 6 °
25	M	24	5	3	Archillis tendon (March 2021)	New balance	R	Sunny 8°
26	F	24	6	3	Tendonitis knee and	New balance prism	R	Sunny 5°
27	F	24	7	3	Knee (6 months ago)	Nike free rn	R	Sunny 9 °
28	F	24	5	1	NA	nike free rn	R	Cloudy 5°
29	M	24	14	5	ITBS	Asics gel hyper speed	R	Cloudy 3° (evening-18h)
30	F	23	6	2	Right calve spring 2020	Asics gel kayano	R	Sunny 4°

31					ITBS, claquage fessier			
32	F	24	8	3	NA	Asics gel cumulus 23	R	Sunny 10°
33	F	23	15	2	NA	Asics GT-1000	R	Sunny 10°
34	F	24	3	1	NA	Nike free tr fit	R	Sunny 10°
35	M	37	25	3	NA	Nike rehax	R	Sunny 10°
36	M	24	14	2	NA	ON cloud	L	Sunny 5°
37	M	34	5	1	NA	Nike free	R	Sunny 4°
	F	22	8	1	right knee	New balance fresh foam 860	R	Cloudy 4°
38	M	22	8	4	fracture feet	Asics gel nimbus 23	R	Sunny 6 °
39	M	19	2	1	NA	New balance fresh foam tempo	R	Sunny 6°
40	F	22	4	1	NA	Brooks Ghost 12	L	Sunny 11°
41	F	33	15	3	NA	Saucony Kivana 12	R	Sunny 11°

*Participant 6 was not used in this project because there was issues with the data collection.

Annex C

1. Xsens MVN full-body sensor – Angular velocity

SegmentAngularVelocity_Pelvis_x	SegmentAngularVelocity_Left_Shoulder_x
SegmentAngularVelocity_Pelvis_y	SegmentAngularVelocity_Left_Shoulder_y
SegmentAngularVelocity_Pelvis_z	SegmentAngularVelocity_Left_Shoulder_z
SegmentAngularVelocity_Right_Shoulder_x	SegmentAngularVelocity_Left_Forearm_x
SegmentAngularVelocity_Right_Shoulder_y	SegmentAngularVelocity_Left_Forearm_y
SegmentAngularVelocity_Right_Shoulder_z	SegmentAngularVelocity_Left_Forearm_z
SegmentAngularVelocity_Right_Forearm_x	SegmentAngularVelocity_Left_Hand_x
SegmentAngularVelocity_Right_Forearm_y	SegmentAngularVelocity_Left_Hand_y
SegmentAngularVelocity_Right_Forearm_z	SegmentAngularVelocity_Left_Hand_z
SegmentAngularVelocity_Right_Hand_x	SegmentAngularVelocity_Left_Upper_Leg_x
SegmentAngularVelocity_Right_Hand_y	SegmentAngularVelocity_Left_Upper_Leg_y
SegmentAngularVelocity_Right_Hand_z	SegmentAngularVelocity_Left_Upper_Leg_z
SegmentAngularVelocity_Right_Lower_Leg_x	SegmentAngularVelocity_Left_Lower_Leg_x
SegmentAngularVelocity_Right_Lower_Leg_y	SegmentAngularVelocity_Left_Lower_Leg_y
SegmentAngularVelocity_Right_Lower_Leg_z	SegmentAngularVelocity_Left_Lower_Leg_z
SegmentAngularVelocity_Right_Foot_x	SegmentAngularVelocity_Left_Foot_x
SegmentAngularVelocity_Right_Foot_y	SegmentAngularVelocity_Left_Foot_y
SegmentAngularVelocity_Right_Foot_z	SegmentAngularVelocity_Left_Foot_z

2. Xsens MVN full-body sensor – Acceleration

SensorFreeAcceleration_Pelvis_x	SensorFreeAcceleration_Head_x
SensorFreeAcceleration_Pelvis_y	SensorFreeAcceleration_Head_y
SensorFreeAcceleration_Pelvis_z	SensorFreeAcceleration_Head_z
SensorFreeAcceleration_Right_Shoulder_x	SensorFreeAcceleration_Left_Shoulder_x
SensorFreeAcceleration_Right_Shoulder_y	SensorFreeAcceleration_Left_Shoulder_y
SensorFreeAcceleration_Right_Shoulder_z	SensorFreeAcceleration_Left_Shoulder_z
SensorFreeAcceleration_Right_Upper_Arm_x	SensorFreeAcceleration_Left_Upper_Arm_x
SensorFreeAcceleration_Right_Upper_Arm_y	SensorFreeAcceleration_Left_Upper_Arm_y
SensorFreeAcceleration_Right_Upper_Arm_z	SensorFreeAcceleration_Left_Upper_Arm_z
SensorFreeAcceleration_Right_Forearm_x	SensorFreeAcceleration_Left_Forearm_x
SensorFreeAcceleration_Right_Forearm_y	SensorFreeAcceleration_Left_Forearm_y
SensorFreeAcceleration_Right_Forearm_z	SensorFreeAcceleration_Left_Forearm_z
SensorFreeAcceleration_Right_Hand_x	SensorFreeAcceleration_Left_Hand_x

SensorFreeAcceleration_Right_Hand_y	SensorFreeAcceleration_Left_Hand_y
SensorFreeAcceleration_Right_Hand_z	SensorFreeAcceleration_Left_Hand_z
SensorFreeAcceleration_Right_Upper_Leg_x	SensorFreeAcceleration_Left_Upper_Leg_x
SensorFreeAcceleration_Right_Upper_Leg_y	SensorFreeAcceleration_Left_Upper_Leg_y
SensorFreeAcceleration_Right_Upper_Leg_z	SensorFreeAcceleration_Left_Upper_Leg_z
SensorFreeAcceleration_Right_Lower_Leg_x	SensorFreeAcceleration_Left_Lower_Leg_x
SensorFreeAcceleration_Right_Lower_Leg_y	SensorFreeAcceleration_Left_Lower_Leg_y
SensorFreeAcceleration_Right_Lower_Leg_z	SensorFreeAcceleration_Left_Lower_Leg_z
SensorFreeAcceleration_Right_Foot_x	SensorFreeAcceleration_Left_Foot_x
SensorFreeAcceleration_Right_Foot_y	SensorFreeAcceleration_Left_Foot_y
SensorFreeAcceleration_Right_Foot_z	SensorFreeAcceleration_Left_Foot_z

9. References

- Abu-Faraj, Z. O., Harris, G. F., Smith, P. A., & Hassani, S. (2015). Human gait and Clinical Movement Analysis. In *Wiley Encyclopedia of Electrical and Electronics Engineering* (pp. 1–34). American Cancer Society. <https://doi.org/10.1002/047134608X.W6606.pub2>
- Alharthi, A., Yunas, S., & Ozanyan, K. (2019). Deep Learning for Monitoring of Human Gait: A Review. *IEEE Sensors Journal*, *PP*, 1–1. <https://doi.org/10.1109/JSEN.2019.2928777>
- Aroganam, G., Manivannan, N., & Harrison, D. (2019). Review on Wearable Technology Sensors Used in Consumer Sport Applications. *Sensors*, *19*(9), 1983. <https://doi.org/10.3390/s19091983>
- Benca, E., Listabarth, S., Flock, F. K. J., Pablik, E., Fischer, C., Walzer, S. M., Dorotka, R., Windhager, R., & Ziai, P. (2020). Analysis of Running-Related Injuries: The Vienna Study. *Journal of Clinical Medicine*, *9*(2). <https://doi.org/10.3390/jcm9020438>
- Chen, H., Zhang, Y., Zhang, Z., Fang, Y., Liu, H., & Yao, C. (2017). Exploring the relation between EMG sampling frequency and hand motion recognition accuracy. *2017 IEEE International Conference on Systems, Man, and Cybernetics (SMC)*, 1139–1144. <https://doi.org/10.1109/SMC.2017.8122765>
- Chollet, F. (2018). *Deep Learning with Python*.
- Dehzangi, O., Taherisadr, M., & ChangalVala, R. (2017). IMU-Based Gait Recognition Using Convolutional Neural Networks and Multi-Sensor Fusion. *Sensors*, *17*(12), 2735. <https://doi.org/10.3390/s17122735>
- Dixon, P. C., Loh, J. J., Michaud-Paquette, Y., & Pearsall, D. J. (2017). biomechZoo: An open-source toolbox for the processing, analysis, and visualization of biomechanical movement

data. *Computer Methods and Programs in Biomedicine*, 140, 1–10.

<https://doi.org/10.1016/j.cmpb.2016.11.007>

Dixon, P. C., Schütte, 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.

<https://doi.org/10.1016/j.gaitpost.2019.09.005>

Dixon, S. J., Collop, A. C., & Batt, M. E. (2000). Surface effects on ground reaction forces and lower extremity kinematics in running. *Medicine and Science in Sports and Exercise*, 32(11), 1919–1926. <https://doi.org/10.1097/00005768-200011000-00016>

Gao, S., Wang, Y., Fang, C., & Xu, L. (2020). A Smart Terrain Identification Technique Based on Electromyography, Ground Reaction Force, and Machine Learning for Lower Limb Rehabilitation. *Applied Sciences*, 10(8), 2638. <https://doi.org/10.3390/app10082638>

García-Pérez, J. A., Pérez-Soriano, P., Llana, S., Martínez-Nova, A., & Sánchez-Zuriaga, D. (2013). Effect of overground vs treadmill running on plantar pressure: Influence of fatigue. *Gait & Posture*, 38(4), 929–933. <https://doi.org/10.1016/j.gaitpost.2013.04.026>

Grzeszick, R., Lenk, J. M., Rueda, F. M., Fink, G. A., Feldhorst, S., & ten Hompel, M. (2017). Deep Neural Network based Human Activity Recognition for the Order Picking Process. *Proceedings of the 4th International Workshop on Sensor-Based Activity Recognition and Interaction*, 1–6. <https://doi.org/10.1145/3134230.3134231>

Gu, F., Chung, M.-H., Chignell, M., Valaee, S., Zhou, B., & Liu, X. (2021). A Survey on Deep Learning for Human Activity Recognition. *ACM Computing Surveys*, 54(8), 177:1–177:34. <https://doi.org/10.1145/3472290>

- Gupta, R., & Agarwal, R. (2019). Single channel EMG-based continuous terrain identification with simple classifier for lower limb prosthesis. *Biocybernetics and Biomedical Engineering*. <https://doi.org/10.1016/J.BBE.2019.07.002>
- Hollis, C. R., Koldenhoven, R. M., Resch, J. E., & Hertel, J. (2021). Running biomechanics as measured by wearable sensors: Effects of speed and surface. *Sports Biomechanics*, 20(5), 521–531. <https://doi.org/10.1080/14763141.2019.1579366>
- Jang, J., Ankit, A., Kim, J., Jang, Y. J., Kim, H. Y., Kim, J. H., & Xiong, S. (2018). A Unified Deep-Learning Model for Classifying the Cross-Country Skiing Techniques Using Wearable Gyroscope Sensors. *Sensors (Basel, Switzerland)*, 18(11). <https://doi.org/10.3390/s18113819>
- Joshi, S., Owens, J. A., Shah, S., & Munasinghe, T. (2021). Analysis of Preprocessing Techniques, Keras Tuner, and Transfer Learning on Cloud Street image data. *2021 IEEE International Conference on Big Data (Big Data)*, 4165–4168. <https://doi.org/10.1109/BigData52589.2021.9671878>
- Luo, Y., Coppola, S. M., Dixon, P. C., Li, S., Dennerlein, J. T., & Hu, B. (2020). A database of human gait performance on irregular and uneven surfaces collected by wearable sensors. *Scientific Data*, 7(1), 219. <https://doi.org/10.1038/s41597-020-0563-y>
- Moya Rueda, F., Grzeszick, R., Fink, G. A., Feldhorst, S., & Ten Hompel, M. (2018). Convolutional Neural Networks for Human Activity Recognition Using Body-Worn Sensors. *Informatics*, 5(2), 26. <https://doi.org/10.3390/informatics5020026>
- Negi, S., CBS Negi, P., Sharma, S., & Sharma, N. (2020). *Electromyographic and Acceleration Signals-Based Gait Phase Analysis for Multiple Terrain Classification Using Deep*

Learning (SSRN Scholarly Paper ID 3657175). Social Science Research Network.

<https://papers.ssrn.com/abstract=3657175>

Nicola, T. L., & Jewison, D. J. (2012). The Anatomy and Biomechanics of Running. *Clinics in Sports Medicine*, 31(2), 187–201. <https://doi.org/10.1016/j.csm.2011.10.001>

Pedregosa, F., Varoquaux, G., Gramfort, A., Michel, V., Thirion, B., Grisel, O., . . . Dubourg, V. (2011). Scikit-learn: Machine learning in Python. *The journal of machine learning research*, 12, 2825-2830.

Pinnington, H. C., Lloyd, D. G., Besier, T. F., & Dawson, B. (2005). Kinematic and electromyography analysis of submaximal differences running on a firm surface compared with soft, dry sand. *European Journal of Applied Physiology*, 94(3), 242–253. <https://doi.org/10.1007/s00421-005-1323-6>

Pinzón-Arenas, J. O., Jiménez-Moreno, R., & Herrera-Benavides, J. E. (2019). Convolutional Neural Network for Hand Gesture Recognition using 8 different EMG Signals. *2019 XXII Symposium on Image, Signal Processing and Artificial Vision (STSIVA)*, 1–5. <https://doi.org/10.1109/STSIVA.2019.8730272>

Prosser, S. J., & Schmidt, E. D. D. (1999). Smart sensors for industrial applications. *Microelectronics International*, 16(2), 20–23. <https://doi.org/10.1108/EUM0000000004551>

Reenalda, J., Maartens, E., Homan, L., & Buurke, J. H. J. (2016). Continuous three dimensional analysis of running mechanics during a marathon by means of inertial magnetic measurement units to objectify changes in running mechanics. *Journal of Biomechanics*, 49(14), 3362–3367. <https://doi.org/10.1016/j.jbiomech.2016.08.032>

Remondini, D. (2018). *SEMG-based Hand Gesture Recognition with Deep Learning*. Universita` di Bologna.

Saunders, J. B., Inman, V. T., & Eberhart, H. D. (1953). The major determinants in normal and pathological gait. *The Journal of Bone and Joint Surgery. American Volume*, 35-A(3), 543–558.

Schepers, H., Giuberti, M., & Bellusci, G. (2018). Xsens MVN: Consistent Tracking of Human Motion Using Inertial Sensing. In *Xsens Technologies* (pp. 1-8).

Sinclair, J., Greenhalgh, A., Edmundson, C. J., Brooks, D., & Hobbs, S. J. (2012). Gender differences in the kinetics and kinematics of distance running: implications for footwear design. *J International Journal of Sports Science Engineering* 6(2), 118-128.

Schütte, K. H., Aeles, J., De Beéck, T. O., van der Zwaard, B. C., Venter, R., & Vanwanseele, B. (2016). Surface effects on dynamic stability and loading during outdoor running using wireless trunk accelerometry. *Gait & Posture*, 48, 220–225.

<https://doi.org/10.1016/j.gaitpost.2016.05.017>

Shah, V., Flood, M. W., Grimm, B., & Dixon, P. C. (2022). Generalizability of deep learning models for predicting outdoor irregular walking surfaces. *Journal of Biomechanics*, 139, 111159. <https://doi.org/10.1016/j.jbiomech.2022.111159>

Shin, S., Baek, Y., Lee, J., Eun, Y., & Son, S. H. (2017). Korean sign language recognition using EMG and IMU sensors based on group-dependent NN models. *2017 IEEE Symposium Series on Computational Intelligence (SSCI)*, 1–7.

<https://doi.org/10.1109/SSCI.2017.8280908>

- Slim, S. O., Atia, A., M.A, M., & M.Mostafa, M.-S. (2019). Survey on Human Activity Recognition based on Acceleration Data. *International Journal of Advanced Computer Science and Applications*. <https://doi.org/10.14569/IJACSA.2019.0100311>
- Van Caekenberghe, I., Segers, V., Willems, P., Gosseye, T., Aerts, P., & De Clercq, D. (2013). Mechanics of overground accelerated running vs. Running on an accelerated treadmill. *Gait & Posture*, 38(1), 125–131. <https://doi.org/10.1016/j.gaitpost.2012.10.022>
- Wang, J., Chen, Y., Hao, S., Peng, X., & Hu, L. (2019). Deep learning for sensor-based activity recognition: A survey. *Pattern Recognition Letters*, 119, 3–11. <https://doi.org/10.1016/j.patrec.2018.02.010>
- Wang, L., Hong, Y., & Li, J. X. (2014). Muscular Activity of Lower Extremity Muscles Running on Treadmill Compared with Different Overground Surfaces. *American Journal of Sports Science and Medicine*, 2(4), 161–165. <https://doi.org/10.12691/ajssm-2-4-8>
- Worsey, M. T. O., Espinosa, H. G., Shepherd, J. B., & Thiel, D. V. (2021). Automatic classification of running surfaces using an ankle-worn inertial sensor. *Sports Engineering*, 24(1), 22. <https://doi.org/10.1007/s12283-021-00359-w>
- Wu, J., Sun, L., & Jafari, R. (2016). A Wearable System for Recognizing American Sign Language in Real-Time Using IMU and Surface EMG Sensors. *IEEE Journal of Biomedical and Health Informatics*, 20(5), 1281–1290. <https://doi.org/10.1109/JBHI.2016.2598302>
- Xsens. (2021). *MVN User Manual*. https://www.xsens.com/hubfs/Downloads/usermanual/MVN_User_Manual.pdf

Zheng, X., Wang, M., & Ordieres-Meré, J. (2018). Comparison of Data Preprocessing Approaches for Applying Deep Learning to Human Activity Recognition in the Context of Industry 4.0. *Sensors*, 18(7), 2146. <https://doi.org/10.3390/s18072146>