

Article

Human Activity Recognition of Individuals with Lower Limb Amputation in Free-Living Conditions: A Pilot Study

Alexander Jamieson ¹, Laura Murray ¹, Lina Stankovic ², Vladimir Stankovic ² and Arjan Buis ^{1,*}

¹ Wolfson Centre, Department of Biomedical Engineering, University of Strathclyde, Glasgow G4 0NW, UK; alexander.jamieson@strath.ac.uk (A.J.); laura.murray.100@strath.ac.uk (L.M.)

² Department of Electronic and Electrical Engineering, University of Strathclyde, Glasgow G1 1XW, UK; lina.stankovic@strath.ac.uk (L.S.); vladimir.stankovic@strath.ac.uk (V.S.)

* Correspondence: arjan.buis@strath.ac.uk

Abstract: This pilot study aimed to investigate the implementation of supervised classifiers and a neural network for the recognition of activities carried out by Individuals with Lower Limb Amputation (ILLAs), as well as individuals without gait impairment, in free living conditions. Eight individuals with no gait impairments and four ILLAs wore a thigh-based accelerometer and walked on an improvised route in the vicinity of their homes across a variety of terrains. Various machine learning classifiers were trained and tested for recognition of walking activities. Additional investigations were made regarding the detail of the activity label versus classifier accuracy and whether the classifiers were capable of being trained exclusively on non-impaired individuals' data and could recognize physical activities carried out by ILLAs. At a basic level of label detail, Support Vector Machines (SVM) and Long-Short Term Memory (LSTM) networks were able to acquire 77–78% mean classification accuracy, which fell with increased label detail. Classifiers trained on individuals without gait impairment could not recognize activities carried out by ILLAs. This investigation presents the groundwork for a HAR system capable of recognizing a variety of walking activities, both for individuals with no gait impairments and ILLAs.

Keywords: human activity recognition; lower limb amputation; physical activity; machine learning; lower limb prosthetics



Citation: Jamieson, A.; Murray, L.; Stankovic, L.; Stankovic, V.; Buis, A. Human Activity Recognition of Individuals with Lower Limb Amputation in Free-Living Conditions: A Pilot Study. *Sensors* **2021**, *21*, 8377. <https://doi.org/10.3390/s21248377>

Academic Editors: Yangquan Chen, Nunzio Cennamo, M. Jamal Deen, Subhas Mukhopadhyay, Simone Morais and Junseop Lee

Received: 27 October 2021
Accepted: 13 December 2021
Published: 15 December 2021

Publisher's Note: MDPI stays neutral with regard to jurisdictional claims in published maps and institutional affiliations.



Copyright: © 2021 by the authors. Licensee MDPI, Basel, Switzerland. This article is an open access article distributed under the terms and conditions of the Creative Commons Attribution (CC BY) license (<https://creativecommons.org/licenses/by/4.0/>).

1. Introduction

Following an amputation procedure, Individuals with Lower Limb Amputation (ILLAs) have been shown to be less physically active than individuals without limb loss [1,2], and there has been extensive research into the physical and socio-economical barriers that prevent ILLAs from performing physical activity, such barriers include existing co-morbidities such as diabetes or chronic obstructive pulmonary disease and a lack of resources or group support [3–5]. Maintaining sufficient levels of physical activity is vital to ILLA's physical and mental well-being: physical activity has evidently improved heart and lung functionality and reduced the effects of chronic lower back pain [4,6]. Physical activity has further given improved perceptions of the individual's quality of life, self-esteem, and body image [4,7]. Therefore, it is imperative that healthcare professionals can monitor and evaluate the physical activity of their clients. Modern day healthcare of ILLAs is limited by a lack of efficient implementation of activity monitoring systems, as evidenced in a recent literature review [8], and is worthy of further research.

By performing automatic recognition and classification of different activities, it is possible to establish the basis of a reliable and low-cost activity monitoring system, in which healthcare professionals that specialise in the care and wellbeing of ILLAs (e.g., physiotherapists) can track how their patients remain active over long periods of time and in free-living conditions. Free-living Human Activity Recognition (HAR) studies are not

uncommon, see e.g., [9–11], and there are publicly available datasets of individuals in free-living [12]. However, there are currently very few studies which have carried out a HAR exercise in free-living conditions with an ILLA population [13]. One of the few examples are the works of Zhang et al. [14,15]; however, their sensor configuration utilised RGB data from a camera as one of the inputs for terrain recognition process, which can have serious ethical and cost considerations if applied in a clinical monitoring context [16]. Additionally, while locomotion activities are very commonly analysed in ILLA-focused HAR studies, these are typically limited only to walking on level ground, ramps, and stairs [13]. In free living conditions, one would be far more likely to traverse hills instead of ramps, and it is worth highlighting the differences between them. Ramps are manmade structures that typically exhibit a fixed incline over a short distance, while hills have variable incline over longer distances. Naturally, the latter can be more challenging for adequate machine learning recognition due to their intrinsic variations.

The primary objective of this pilot study was to investigate the implementation of supervised classifiers and a neural network for the recognition of activities carried out by Individuals with Lower Limb Amputation (ILLAs), as well as healthy individuals with no gait impairment, in free living conditions. Additionally, to the researchers' knowledge, this is one of the first papers that have investigated the potential for terrain "resolution" in HAR for an ILLA population. The investigations carried out in this paper are divided into three experiments: The first step takes an assortment of classifiers and a deep learning network and uses hyperparameter tuning to determine which classifiers were the most suited for recognizing walking activities. Subsequently, the best classifiers from the first experiment were then used in the second step to investigate how the terrain "resolution" impacts on classifier performance. Finally, in the third step, there is an investigation as to whether the data trained on a group of subjects (either ILLAs or healthy individuals with no gait impairment) could recognize activity carried out by an ILLA.

2. Materials and Methods

2.1. Data Collection and Pre-Processing

Ethical approval for the pilot study was granted by the University of Strathclyde's University Ethics Committee prior to conducting the investigation. Participants were recruited between September and December 2020 via social media posters and contacting people who had consented to follow-up participations from previous investigations. Due to lack of recruitment numbers (primarily caused by the Coronavirus), participants were selected based on convenience sampling. All participants had to be at least 18 years of age, be comfortable performing moderate vigorous activity, and not be at risk of life-threatening conditions if infected with the coronavirus. The ILLA volunteers were further required to be able to ambulate with a prosthesis without the use of walking aids and without any comorbidities or musculoskeletal conditions which could potentially impact their ability to ambulate for sustained periods of time. As the study was primarily targeted at monitoring the activity of ILLAs, many of which will have poor levels of mobility, participants were explicitly asked not to carry out vigorous activities such as jogging or running.

Recruited participants were instructed to carry out approximately 140 min of walking over the duration of a week (expecting a minimum of 20 min of walking per day) in the local vicinity of their homes while recording inertial data via a thigh-worn accelerometer. The accelerometer is an ActivPAL, presented in Figure 1 (PAL Technologies, Glasgow, UK). The ActivPAL is 4×2 cm, weighing approximately 0.01kg and is characterized with an operating frequency of 20Hz. The ActivPAL's accuracy has been validated in terms of its ability to measure levels of activity intensity (including being able to distinguish sitting and standing behaviour) as well as accurately measure the step count with an ILLA population [17–19]. Ground truth validation of the activities were acquired by instructing the participant to additionally wear a chest-mounted camera and record elevation data via a GPS recording application (Strava, Strava Inc., CA, USA) operating on an iPhone 6 (Apple Inc., CA, USA) smartphone provided by the researchers. An example of the recording

set-up is shown in Figure 2. Activities were annotated using the Visual object Tagging Tool (VoTT) (Microsoft, WA, USA), using the GPS co-ordinates from the Strava application to aid with labelling uphill and downhill movement; this is elaborated further in Section 2.2. All signal processing and machine learning was handled in a Matlab environment (Matlab 2020b, Mathworks, MA, USA). The accelerometer data were digitally filtered by a linear bandpass Butterworth filter with a preserved frequency range of 0.4 Hz to 3 Hz and order of 3 to remove the gravitational constant and high frequency muscle artifact movements from the raw signal [20,21]. The data were segmented into chunks of 2 s long (or 40 sample width) windows with no overlap, a two second window length would be adequate to capture at least one step, and the choice of no overlap in the sampling window was arbitrary due to the choice being non-consequential compared to other design choices in the machine learning process [22]. Longer sampling windows decreased the number of samples available for training, which was not ideal given the large imbalance of classes in the dataset, see Section 2.3 and Table 1. A list of 243 statistical, frequency and wavelet domain features (refer to Appendix A.1) were calculated for each segment of data and are standardized to have zero mean and a standard deviation of one. To give an abridged justification of feature inclusion, the list of features in this study were chosen as a combination of their widespread inclusion in similar HAR papers that cover exploratory data analysis (e.g., [13,23,24]), or through analysis of the fundamental differences in acceleration between different types of activities. For example, to distinguish terrains with accelerometer data, a change in the ground reaction force when traversing different types of terrains (as evidenced in [25]) will result in a change of the accelerometer waveform. This will have impact in the wavelet domain (a domain that considers both time and frequency), and can therefore be quantifiable with wavelet features. The use of wavelet features for the distinguishment of anatomically similar activities has been further supported by recent literature [26].

Table 1. List of all collected activities in investigation and their total corresponding sample counts.

Terrain Label	Total No. of Samples
Concrete, Flat	12,200
Concrete, Camber, Parallel	3923
Concrete, Downhill	2549
Concrete, Uphill	2427
Grass, Flat	1346
Concrete, Camber, Perpendicular	877
Upstairs	691
Downstairs	656
Concrete, Camber, Downhill, Parallel	572
Concrete, Camber, Uphill, Parallel	540
Stone, Flat	496
Sand, Flat	232
Stone, Uphill	176
Grass, Downhill	151
Concrete, Camber, Downhill, Perpendicular	93
Concrete, Camber, Uphill, Perpendicular	93
Stone, Downhill	80
Grass, Uphill	78
Gravel, Downhill	46
Gravel, Flat	32

Table 1. *Cont.*

Terrain Label	Total No. of Samples
Sand, Uphill	31
Red Ash, Flat	26
Sand, Downhill	12
Gravel, Uphill	3

**Figure 1.** Scaled diagram of an ActivPAL device, with axes of orientation for reference.**Figure 2.** Example of a recording set-up. AP = ActivPAL. The smartphone and ActivPAL would both be worn under clothing.

2.2. Annotation of Slopes

To identify uphill and downhill segments of a walk, GPS data from Strava was utilised. The angle of slope being traversed as a function of time $\theta(t)$ was calculated by approxi-

imating the slope as a straight line and then calculating the inverse tangent of the elevation change ($\Delta Y(t)$) divided by the distance change ($\Delta X(t)$). ΔY was obtained from the Strava elevation readings, while ΔX was calculated using WGS84 ellipsoid distance of latitude and longitude co-ordinates also taken from Strava's GPS data. Elevation data in Strava recordings are generated based on a combination of barometer readings directly extracted from the operating smartphone's sensor and, where applicable, a basemap of elevation data where points of elevation are known [27]. The iPhone 6 contains a commercial BMP280 barometer (Bosch, Gerlingen, Germany) which can measure changes in altitude with a precision of ± 0.12 hPa (or ± 1 m equivalent), operating at a sample frequency of approximately 1 Hz [28]. A study of various commercial GPS devices found the Strava application running on an iPhone 6 device produced the lowest error in elevation and distance, and so was considered the best solution for low-cost GPS acquisition [29]. Since elevation readings were dependent on the accuracy of GPS readings, whose accuracy can fluctuate depending on the strength of the GPS signal among other factors [30], the angular values were given a smoothing average over an empirically determined 10 samples, which gave a reasonable trade-off between "stability" of the angular values plotted over time and the peak magnitude of the angle readings. A failsafe countermeasure was put in place where if there were more than 15 s between two successive Strava readings, the angle would not be calculated and all unclear movement within that timeframe would be given a "null" label, meaning no activities were annotated within this time period. An exemplary angle plot is demonstrated in the Figure 3. Due to the uncontrolled nature of the investigation, the absolute angle of slopes in terrain traversed by the participants were unknown prior to the investigation, making the accuracy of the slope readings not possible to validate. This should be acknowledged as a key limitation of the slope readings, and would require subsequent validation in future studies using an outdoor structure with fixed and known angle of slope.

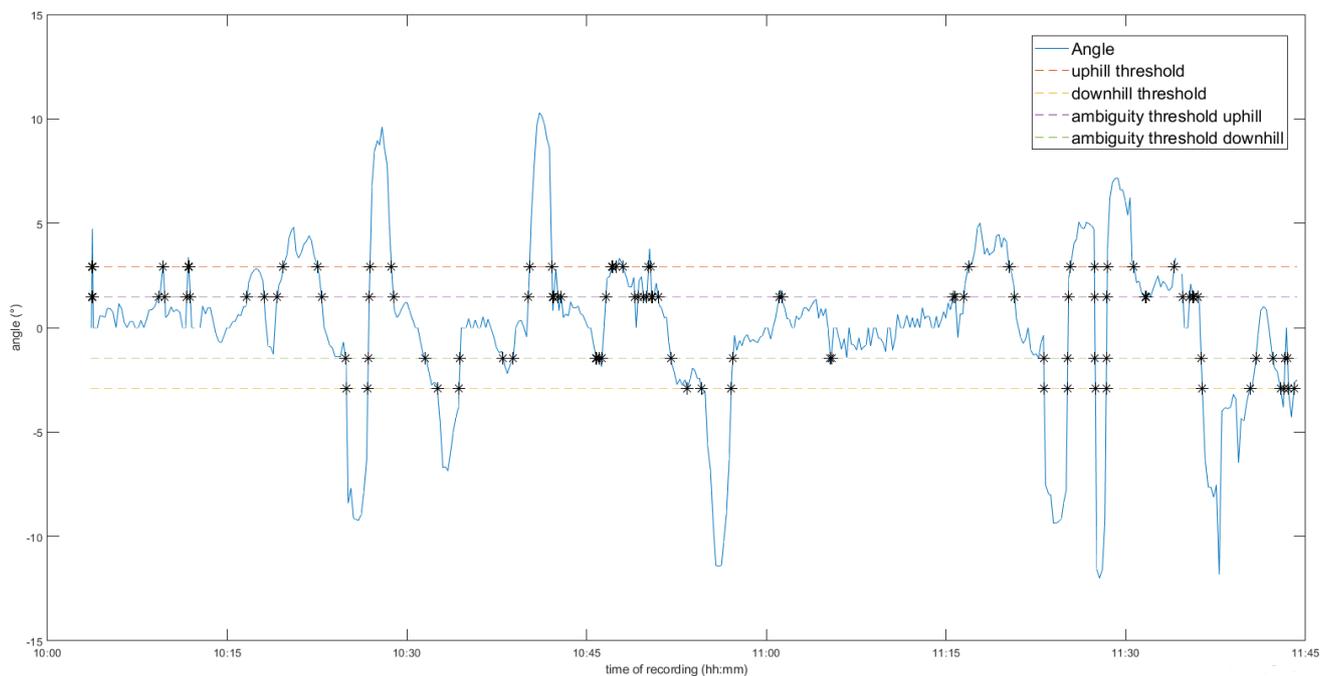


Figure 3. Example of a slope angle plot for one participant's recording.

As annotation of the video progressed, the annotator would periodically check for whenever the thresholds in the slope angle data were crossed. There were four key angular thresholds in the annotation process: a slope would be determined as uphill when the angle of the slope exceeded $+2.9^\circ$, or -2.9° for downhill. This threshold was referenced from the American Disability Association's minimum slope angle requirement for construction of a handrail in a public access building [31]. Due to the imprecision of elevation and

ellipsoid distance readings, an “unconfident” threshold was also constructed between $+1.45$ to $+2.9^\circ$ and -1.45 to -2.9° . Any data that fell within this threshold would be given a null label, the exception being when a participant was traversing stairs. As it could not be confidently ascertained whether readings within this threshold were uphill, downhill, or flat, it would have led to numerous mislabelling instances; thus, readings within these ambiguity thresholds are discarded from the process.

2.3. Introducing “Label Resolution”

In this paper, annotated activities are initially given what is referred to as “full label resolution”, such that each sample is described with its terrain and the condition of the terrain. For example, walking along a flat concrete road is initially annotated as “Concrete, Flat”, while walking uphill on a grassy knoll is annotated as “Grass, Uphill”. The exception to this is stair traversals, which are referred to as “Upstairs” or “Downstairs”. However, at such a level of detail, this resulted in 24 unique activities (as seen in Table 1), thus will adversely impact on the machine learning system. Subsequently, samples undergo a process called “label consolidation”, where similar activities are given the same label. As a result, there are three levels of Label Resolution used in this study:

- Level 1: No Terrain Resolution. Labels only contain the condition of the traversed ground (ground is flat, ground is sloped, or is traversing stairs)
- Level 2: Hard/Soft Terrain Resolution. Terrains are resolved into whether the terrain involved is hard/soft. Hard terrains included concrete, stone, and gravel, while soft terrains included grass, sand and red ash (a type of clay pitch).
- Level 3: Full Resolution: All Labels use their full label resolution.

To further illustrate the consolidation process, a flowchart in Figure 4 has been created to show how the process operates for all activity labels that relate to traversing flat terrain. The process is identical for all uphill and downhill related activities.

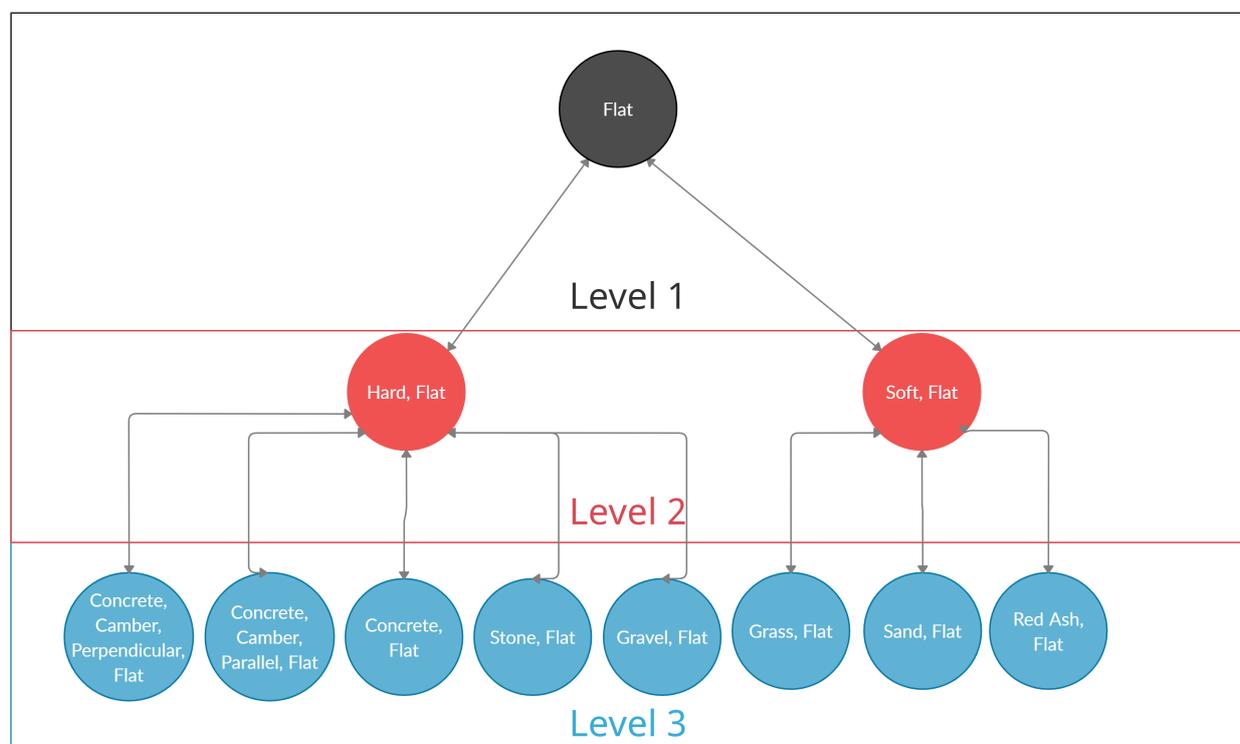


Figure 4. Example of the label consolidation process for “flat” activity labels.

In the first and third experiments, the annotated labels were assigned Level 1 terrain resolution, this meant that there were only five classes for potential classification: flat

walking, uphill walking, downhill walking, upstairs movement, and downstairs movement. This was designed to give an idea of how the classifiers would perform at the simplest level of classification and is comparable to the labels used in other studies with ILLAs [13]. Due to the unstructured nature of the investigation, there was a large imbalance between labels; at Level 1, flat walking comprises approximately 70% of the total dataset, while stair data (up and down) both comprise only around 5% combined. In order to avoid having a highly biased classifier that favours prediction of the majority class [32], classes were rebalanced using Synthetic Minority Oversampling Technique (SMOTE) [33].

2.4. Experiment 1: Comparison of Supervised Classifiers

The objective of the first experiment was to identify the best classifier for supervised classification of the five walking activities (Level 1 resolution). Six off-the-shelf supervised classifiers were tested: Support Vector Machine (SVM), k Nearest Neighbours (kNN), Random Forest, Adaboost, Naïve-Bayes, Linear Discriminant Analysis (LDA) and were compared with an additional Long Short-Term Memory Network (LSTM). The classifiers were chosen based on their frequent appearance in HAR studies (see e.g., [34–39]). After heuristic experimentation with various dimensionality reduction techniques, it was found that application of Principal Component Analysis (PCA) with express preservation of components that explained 95% of the total systematic variance (approximately 65–70 components) was a suitable technique for dimensionality reduction for the off-the-shelf classifiers. The LSTM model constructed for this investigation was a Sequence-to-Label network demonstrated in Figure 5. Experimentation with feature selection indicated that the LSTM network performed best when using each axis of the raw accelerometer signal (post-filtering) as 3 separate channel inputs. Using guidelines set by Bengio [40] and some heuristic experimentation, the most suitable parameters defined for the network were to perform ADAM Optimization [41], set mini batch size to 256, set the initial learn rate to 0.005 with a decrease in the learn rate by a factor of 10 every 2 training epochs, and set the maximum number of epochs to 15. These parameters were kept consistent throughout all stages of analysis because they had minimal influence on training accuracy, and so this avoided cumbersome tuning times for the LSTM network. Raw training data is also balanced via the application of SMOTE.

The six tested supervised classifiers with embedded feature selection, and the LSTM network with raw data, underwent hyperparameter tuning to determine the best classifier. The list of eligible hyperparameters for tuning are listed in Table 2. For the SVM models, Kernel Scale variations and One vs. All training were excluded from tuning as they resulted in training times exceeding an hour per optimization iteration. For the supervised classifiers, the tuning process was performed with Random Search with 20 maximum iterations [42]; this excludes Naïve-Bayes and LDA classifiers which had less than 20 possible input combinations. The LSTM network, due to its long training times, used a simple grid-search approach with discrete values for the dropout factor (0.1 to 0.5) and the number of hidden units (175 to 350).

Table 2. Parameters tuned in Experiment 1.

Classifier	Parameters Tuned
SVM	Kernel Type, Box Size
KNN	Number of neighbours, Distance Metric, Distance Weight
Random Forest	Maximum number of learners, maximum number of splits
AdaBoost	Maximum number of learners, maximum number of splits, learning rate
Naïve-Bayes	Type of distribution, Type of kernel (for kernel distribution)
Discriminant Analysis	Type of discriminant analysis (Linear, Quadratic)
LSTM	Dropout factor, Number of Hidden Units

SVM—Support Vector Machine | KNN—k-Nearest Neighbour | LSTM—Long-Short Term Memory (Network).

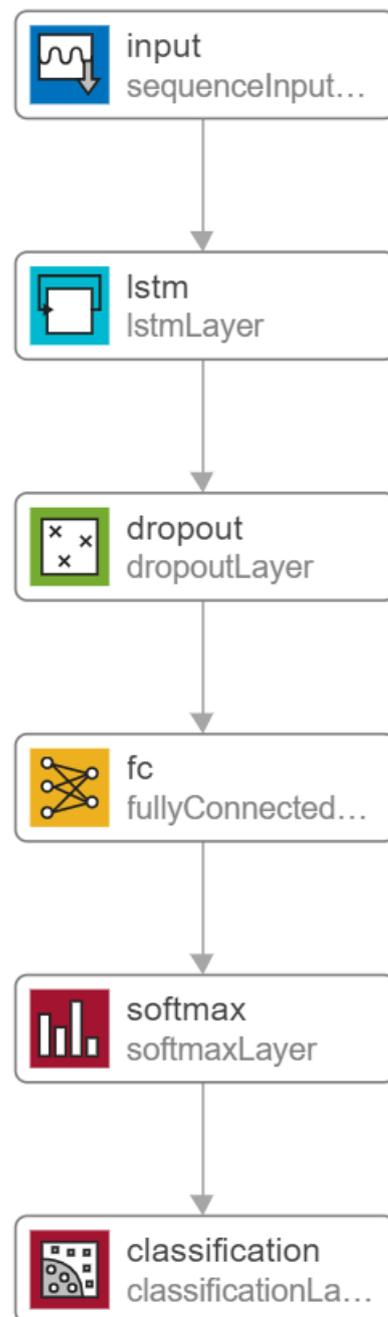


Figure 5. Overview of Long-Short Term Memory (LSTM) Matlab architecture. The input represents a 40 sample segment of a 3 dimensional (triaxial) input of raw accelerometer data.

All classifiers were validated using 5-Fold cross validation of the entire dataset: a combination of all ILLA and non-impaired individuals' data. Though 10 is generally the standard value of the K in cross validation in standard HAR papers, it was found that training times, particularly of the LSTM network, were very long (upwards of 10 min per iteration), and therefore was unsuitable for 10-fold. 5-Fold is generally considered to be the minimal acceptable level of K for machine learning validation [43]. Parameter tuning for the supervised classifiers used a nested cross validation approach [44]. For outer k_o folds, the data is split into training and testing. The training data is further split into an inner sub-fold k_i to tune hyperparameters via k-fold cross validation, with both the inner and outer loop values for k set to five. Thus, for each k_o fold, a set of optimal parameters are obtained, and the classifier is trained with these parameters. The classifier is validated with

the test-set data to get an accuracy, and the overall reported accuracy is the mean value of test accuracy across the k_o folds. While LSTM used 5-folds in the outer cross validation, due to very long training times, the internal validation process for parameter tuning instead used a holdout of 25% of the training data.

2.5. Experiment 2: Label Terrain Resolution

In this experiment, the appropriate level of terrain resolution is determined. There was an expected trade-off between the resolution of detail in the activity and the corresponding classifier performance. This experiment aimed to validate this theory, and to show at what specific label of terrain resolution can an acceptable level of classifier accuracy be obtained. Using the best performing classifiers from experiment 1, the terrain resolution was increased in stages from level 1 to level 3 as per the criteria described in Section 2.3. Each level of resolution is validated using 5-fold cross validation of the entire dataset.

2.6. Experiment 3: Subject Cross-Validation

In the final experiment, the objective was to determine whether a classifier trained exclusively on either healthy or ILLA participants would be robust enough to detect physical activities in a single ILLA participant. Hence, this experiment uses Leave One Subject Out (LOSO) validation. There were two methods of training the classifier:

- Method 1: Healthy Participant Training. For the best performing classifiers, train data on the healthy participants, then use LOSO validation to test the classification accuracy for each of the ILLA participants individually.
- Method 2: Amputee Participant Training. All data from healthy participants are ignored. LOSO validation is again performed with the best classifiers, this time training the data on all-minus-one ILLA participants and testing on the remaining ILLA participant, repeating the process for each ILLA.

In this experiment, the label resolution is kept at Level 1.

3. Results

3.1. Participant Information

A total of eight healthy participants with no gait impairment and four participants with lower limb amputation were recruited for the study. From this point onwards, specific participants are referred to using a letter-number rule. For example, healthy participant #8 is referred to simply as “H8”, and amputee participant #2 is referred to as “A2”. All participants who consented to participate in the study were able to carry out their recording successfully, with zero dropouts. However, two of the recordings from participant H1, and one recording each from participants H4 and H5 had to be discarded from analysis due to improper camera wear or recording at night-time, making analysis of those recordings infeasible. The participants’ primary characteristics are listed in Tables 3 and 4. The demographics match in terms of height and weight, however there were large gender biases in both demographics with the vast majority of participants being male. Additionally, there was a noticeable age bias between healthy and amputee demographics; most of the healthy volunteers were in their mid-20s, while the mean age of the ILLAs is approximately 50. While this could have had adverse impact on machine learning testing accuracies, for research purposes it would be useful to determine if training data on a younger healthy population could still result in the detection of activity for an older population with gait impairments. Despite the relatively small sample size of ILLA participants, there was some interesting variation in the amputation time with 2 long-term experienced ILLAs and two comparatively inexperienced ILLAs, and there was a bilateral amputee to provide a comparison point to unilateral amputees. Contrary to Scottish national statistics [45], all but one of the amputees had vascular-related issues as the primary cause of amputation, with the other three having traumatic-related causes.

Table 3. Characteristics of healthy participants without gait impairment.

Participant	Height (m)	Weight (kg)	Age (Years)	Gender
#1	1.80	84	24	Male
#2	1.65	63	51	Female
#3	1.62	65	18	Female
#4	1.97	99	25	Male
#5	1.92	102	25	Male
#6	1.83	89	24	Male
#7	1.84	88	25	Male
#8	1.78	98	25	Male

Table 4. Characteristics of lower-limb amputated participants.

Participant	Height (m)	Weight (kg)	Age (years)	Gender	Type of Amputation	Type of Prosthesis	Origin	Years since Amputation
#1	1.79	95	55	Male	Unilateral transtibial	TSB socket, ESR foot	Traumatic	3
#2	1.70	86	57	Male	Unilateral transtibial	TSB socket, ESR foot	Traumatic	32
#3	1.72	110	40	Male	Unilateral transtibial	PTB socket, ESR foot	Traumatic	33
#4	1.52	61	48	Female	Bilateral transtibial	TSB socket, ESR foot	Vascular	4

TSB—Total Surface Bearing | ESR—Energy-Storing-and-Returning | PTB—Patellar Tendon Bearing.

3.2. Experiment 1: Classifier Optimization

The results of the first experiment are shown in Table 5, and from these the “best” classifiers were identified as the SVM and LSTM models. Both models were considered to be equally viable due to the high fold-dependent standard deviation in the LSTM model, while SVM had much more consistent recognition accuracy across each fold. The best parameters indicate the hyperparameters that were present in the highest scoring fold of the outer cross validation.

Table 5. Comparisons of Supervised Classifiers under 5-fold Cross Validation and the best suited hyperparameters for each.

Classifier	5-Fold Accuracy (%)	Fold Std. ($\pm\%$)	Best Parameters
SVM	77.22	0.54	Kernel: Gaussian Box Constraint: 193
KNN	75.76	2.26	Num. Neighbours: 12 Distance Metric: Correlation Distance Weight: Squared Inverse
Random Forest	71.84	0.69	Num. Learners: 188 Num. Splits: 4392
Adaboost	69.23	3.18	Num. Learners: 211 Num. Splits: 910 Learn Rate: 0.07
NB	64.40	0.95	Gaussian Kernel Distribution
LDA	73.91	0.93	Quadratic
LSTM	78.46	2.89	Dropout Factor: 0.2 Num. Hidden Units: 190

SVM—Support Vector Machine | KNN—k-Nearest Neighbour | NB—Naïve Bayes | LDA—Linear Discriminant Analysis | LSTM—Long-Short Term Memory (Network).

3.3. Experiment 2: Label Terrain Resolution

All subsequent experiments used SVM and LSTM models with their best parameters from Table 5. The mean classification accuracy results from varying terrain resolution can

be seen in Table 6, which is accompanied with F1 scores for each activity in Tables 7–9. In the base terrain resolution, flat and stair data had notably greater F1 scores compared to hill data, which is to be expected due to stair movement having significantly different kinematic behaviour from standard walking [46]. Uphill movement typically had less confusion than downhill movement, which is supported by Kimel-Naor et al [47], whose findings suggest that downhill movement and flat movement are often very similar due to minimal required changes in gait kinematics for downhill movement. The mean accuracy of both classifiers decreases as terrain resolution (and in turn, the number of activities) increase. While the drop-off in mean classification accuracy for the LSTM model was less severe than the SVM model, the F1 scores at each stage of terrain resolution (Tables 7–9) reveal that many of the individual classes for Level 2 and 3 have poor F1 scores, and the SVM model had larger F1 scores in many class-by-class comparisons. The confusion matrices for the LSTM and SVM models at each level of terrain resolution are contained in Appendix B. By viewing the confusion matrices of the LSTM model (Refer to Figures A4–A6 in Appendix B), the higher accuracies are a result of having high recall for the majority class (“Flat”, “Hard, Flat” or “Concrete, Flat”) indicating that despite balancing the classes with SMOTE, the LSTM model has high model complexity and was overfitting [48]. The SVM models (Figures A1–A3), while having relatively worse performance, had less evidence of underfitting and better precision in the minority classes.

Table 6. Comparison of mean classification accuracy in SVM and LSTM classifiers for varying resolutions of activity label, using 5-fold cross-validation.

Label Resolution Level	1	2	3
SVM accuracy (%)	77.22 ± 0.54	62.60 ± 2.73	32.74 ± 2.46
LSTM accuracy (%)	78.46 ± 2.89	73.77 ± 1.83	41.85 ± 4.19

SVM—Support Vector Machine | LSTM—Long-Short Term Memory (Network).

Table 7. Mean F1 Scores for Level 1 Label Resolution.

Activity	No. of Test Samples	LSTM F1 Scores	SVM F1 Scores
Downstairs	131	80.0	74.6
Upstairs	138	80.1	75.7
Flat	3826	83.8	84.2
Uphill	670	68.7	69.9
Downhill	701	62.8	61.7

SVM—Support Vector Machine | LSTM—Long-Short Term Memory (Network).

Table 8. Mean F1 Scores for Level 2 Label Resolution.

Activity	No. of Test Samples	LSTM F1 Scores	SVM F1 Scores
Downstairs	131	80.3	75.8
Upstairs	138	77.4	75.1
Hard, Flat	3506	80.2	82.3
Hard, Uphill	647	69.9	70.8
Hard, Downhill	668	63.1	61.5
Soft, Flat	321	57.1	61.3
Soft, Uphill	22	27.1	36.2
Soft, Downhill	33	27.6	36.7

SVM—Support Vector Machine | LSTM—Long-Short Term Memory (Network).

Table 9. Mean F1 Scores for Level 3 Label Resolution.

Activity	No. of Test Samples	LSTM F1 Scores	SVM F1 Scores
Concrete, Camber, Downhill, Parallel	114	34.8	38.0
Concrete, Camber, Downhill, Perpendicular	19	3.7	0.0
Concrete, Camber, Parallel	784	50.8	62.3
Concrete, Camber, Perpendicular	176	8.9	4.4
Concrete, Camber, Uphill, Parallel	108	39.3	50.5
Concrete, Camber, Uphill, Perpendicular	18	3.2	0.0
Concrete, Downhill	510	48.1	58.3
Concrete, Flat	2440	42.4	72.5
Concrete, Uphill	485	55.5	69.5
Downstairs	132	71.5	74.9
Grass, Downhill	30	31.5	37.1
Grass, Flat	269	51.0	62.6
Grass, Uphill	16	25.1	29.1
Gravel, Downhill	9	17.8	24.2
Gravel, Flat	6	7.4	0.0
Gravel, Uphill	1	0.0	0.0
Red Ash, Flat	5	17.4	25.0
Sand, Downhill	2	0.0	0.0
Sand, Flat	47	33.3	52.0
Sand, Uphill	6	18.6	27.0
Stone, Downhill	16	15.9	9.4
Stone, Flat	99	18.4	19.4
Stone, Uphill	35	38.3	41.3
Upstairs	139	69.4	73.4

SVM—Support Vector Machine | LSTM—Long-Short Term Memory (Network).

3.4. Experiment 3: Subject Cross-Validation

Tables 10 and 11 present the results of training classifiers by an ILLA or healthy population and testing on one ILLA, using Level 1 or “no terrain” resolution. Clearly, this experiment was unsuccessful; while SVM accuracies are mediocre, the LSTM models failed completely, for some participants having a recognition accuracy lower than random guess. Despite the significantly smaller size of training dataset available, the mean classification performance of the SVM when trained by healthy participants was lower than the performance when the SVM was trained by other ILLAs, indicating that training a classifier via individuals with unimpaired gait are not suitable for classification of activities carried out by ILLAs. A significant confounding factor in this study was the fact that the ILLAs walked over considerably more variations in terrain than the healthy individuals, and even within the ILLA group, some ILLA walked on certain terrains that other ILLAs rarely or never walked over. For instance, participant A1 walked over stony terrain, with no instances of sandy terrain, while the opposite was true for participant A2. Participant A4 being a bilateral amputee, while the other three were unilateral, may have caused further confusion in the models. The LSTM model had wildly inconsistent inter-participant accuracy and high standard deviations, but given the much smaller size of training dataset, particularly when trained with only ILLA participants, this was anticipated. The confusion matrices for each participant (by Classifier and by training method) are located in Appendix B.

Despite the overall poor findings, a positive perspective of these results is that the years of amputation experienced by the amputee (refer to Table 4) do not appear to influence the validation results, judging by the comparable performances of participant A1 (amputation time 3 years) to participants A2 and A3 (amputation time 32 and 33 years respectively). Though, due to the minimal sample size, this is not a generalizable result.

Table 10. Mean Leave-One-Subject-Out (LOSO) accuracies of each of the four lower limb-amputated subjects, using models trained only on individuals without gait impairment.

Participant	A1	A2	A3	A4	Mean	Inter-Subject Std.
SVM accuracy (%)	52.41	60.92	46.44	58.41	54.55	±5.61
LSTM accuracy (%)	24.97	48.71	14.79	25.60	28.52	±12.42

SVM—Support Vector Machine | LSTM—Long-Short Term Memory (Network).

Table 11. Mean Leave-One-Subject-Out (LOSO) accuracies of each of the four lower limb-amputated subjects, using models trained only on lower limb amputated individuals.

Participant	A1	A2	A3	A4	Mean	Inter-Subject Std.
SVM accuracy (%)	58.12	52.99	54.08	61.55	56.68	±3.40
LSTM accuracy (%)	53.97	10.26	43.22	16.96	31.10	±18.06

SVM—Support Vector Machine | LSTM—Long-Short Term Memory (Network).

4. Discussion

4.1. Main Findings and Interpretations

By analysing walking activity data carried out by ILLAs and healthy individuals in free-living settings with supervised classifiers, this research was able to achieve satisfactory results, but with significant room for improvement. The best suited classifiers for the supervised approach were discovered to be an SVM which calculated 243 features in the time, frequency, and wavelet domains, with compressed dimensionality via PCA to 68 principal components, and an LSTM Network which required only the raw sequential triaxial accelerometer data. The LSTM model generally had superior mean classification accuracy but required much longer training times and had perceptible overfitting on the majority class. Given these significant drawbacks, the SVM was overall the more robust classifier and, if based purely on these results, would be the most suitable classifier in a clinical-based activity monitoring system. The LSTM's design is based on a simple preconfigured model available from Matlab's Deep Learning designer application. In cutting-edge deep learning HAR experimentation, LSTM models tend to exhibit complicated designs. Zhao et al. [49] for instance implemented a bidirectional LSTM which learns time dependencies in both directions of the input signal. Using this model, Zhao et al. [49] was able to recognize the five activities listed in Level 1 terrain resolution of the Opportunity dataset [50], acquiring an average of 94% in F1 scores across the five activities for a 70/30% split of training and test subject data. Sun et al. [51] meanwhile used a combination of CNN and LSTM architecture. The CNN component performed feature extraction from convolution of the raw input data, from which the LSTM recognizes the time-dependencies of the CNN-forged features. They performed classification through an Extreme Learning Machine—an alternative to the SoftMax classifier [52], and were able to achieve strong recognition accuracies for high level activities. Implementation of the described LSTM variants from [49–51] may have resulted in improved classification results in this study. Technical limitations of the deep learning networks arose as a result of Matlab's lack of official support for deep learning classifiers; for example, Extreme Learning Machines are not supported in their deep learning classification app and require downloading unofficial community-supported toolboxes in place. Future implementations of the project will utilize Keras in a Python™ (Python Software Foundation, DE, USA) Integrated Development Environment (IDE) for

more refined control of deep learning parameters as well as for accessing a wider range of implementable deep learning libraries [53].

From the terrain-resolution experiments (experiment 2), there is some potential to expand the typical walking activities captured in a HAR investigation to include hard and soft terrains (label resolution: Level 2) as additional outcomes. While “soft, flat” had acceptable F1 scores, “soft, uphill” and “soft, downhill” did not, this is likely due to having much smaller quantities of these activities compared to “soft, flat” (1604 for “soft, flat” vs. 109 and 163 for “soft, uphill” and “soft, downhill” respectively) and the hard terrain activities (which total around 24,000 samples). While the inclusion of SMOTE can help increase performance in these minority classes and prevent overfitting [54], if the initial training data size is too small, the SMOTE process will still be unable to successfully capture the full range of potential feature data that minority class can exhibit because it can only create new samples within a nearest neighbour proximity to that existing data, which can be especially problematic if the data is noisy [55]. Balancing the dataset in the future may require greater degrees of undersampling in the majority class to provide a better initial balance to the data prior to SMOTE. Unfortunately, Level 3 label resolution appears to be impractical to detect accurately, at least within the scope of the data available for the investigation. There would likely require significantly more data from the minority classes and likely require more additional sensors. For example, a second ActivPAL attached to the other leg could use inter-sensor axis correlation coefficients to detect the presence of laterally uneven ground. The presence of “camber movement” in particular may be impossible to detect with consistent accuracies due to its high similarity with “concrete, flat” movement coupled with variations in road designs (for example, having different degrees of slope).

The subject cross-validation experiments (experiment 3) were regrettably completely unsuccessful, even at the most basic level of terrain resolution. As briefly mentioned in the results, this likely was due to the ILLAs carrying out different walking activities, both from the healthy individual group and from each other. Future investigations would need to acquire considerably larger numbers of ILLA volunteers with a diverse range of types of amputation, and work towards avoiding the gender and age biases shown in this dataset which may have had further influence on the results. As it stands, the clinical implications the final experiment is that a classifier for the detection of walking activities in ILLAs cannot be reliably trained on activity data from a population with non-impaired gait, and by comparing the LOSO validation of the bilateral amputee (participant A4) to the others, this experiment also indicates that different classifiers may need to be trained for different types of amputation.

4.2. Comparisons with Other Studies

The recognition accuracies in this investigation were far from ideal: each of the lab-based studies identified in [13] that used only an IMU sensor for data collection had comparable or superior recognition accuracies [56–60]. This pilot study is however does have comparable performances with other free-living HAR investigations that use a healthy gait-unimpaired population. Gyllensten and Bonomi. [61] found that activity data trained in laboratory conditions dropped from an average recognition accuracy of 95.4% to 75.6% with an SVM. Their standard deviations for the SVM model were also considerably higher ($\pm 10.4\%$ compared to 0.54% in this study), though this difference may be explained by measuring different activities and using Leave-One-Subject-Out Cross Validation. Their methodology also indicates that it is better to have the training component in free-living conditions. Other free-living studies have achieved better recognition accuracies but use simple classification problems, for instance Ellis et al. [10] had an average recognition accuracy of 85.6% but only distinguished sedentary, standing and walking movement, and driving. Likewise, Fullerton et al. [11] acquired 96.7% accuracy using broad activity label definitions (e.g., “Self-Care, “Home Activities”). The recognition accuracies in this investigation were superior to Liu et al [62] for a single accelerometer sensor (69.9%),

comparable when multiple accelerometer sensors were used (74%), and inferior when they included an additional ventilation sensor (84.7%). This further reinforces that the performance in this investigation could have been improved with the additional sensors. In addition to neural networks, recognition of a wide variety of activities may be achievable through probabilistic models such as the General Probabilistic model proposed in [63] that build upon the recognition of simple motions (atomic activities) to recognize more dynamic and complex physical activities. Within the context of this pilot study, a probabilistic model could theoretically first recognize the condition of terrain (flat, hill or stairs) followed by the type of terrain traversed to build a complete activity label. Such models are worth investigating in future studies.

Generally, there are not many studies which have attempted to differentiate walking across different terrains using only IMU data collected by humans. The work carried out by Hu et al. [64] appears to be the first study to attempt terrain classification with a single IMU, however this was performed in laboratory conditions and only distinguished between flat and uneven ground. Hashmi et al. [65] was able to acquire very high classification accuracies (87.5% with an SVM classifier) when distinguishing six types of terrain: concrete, tiles, carpet, asphalt, soil, and grass. Likewise, they also found consolidating terrains into hard and soft categories improved the average recognition accuracy to 92.08%, they however only tested flat terrain and did not include hills or stairs in the classification problem. As previously discussed in the methodology, the angle of slopes calculated during the study could not be validated, which may have had an impact on the accuracy of the ground truth, and so will require validation in a future study prior to collection activity data from participants. More recently, Russell et al. [66] carried out HAR in a cross-country trail. While the study acknowledged the presence of terrain, as well as the speed at which the terrain was traversed, their classification problem was simplified to only distinguish between laying, sitting, walking, running, or climbing a fence. Publication bias may explain the apparent absence of these types of studies: experiments which have attempted similar methodology to this investigation have achieved similar or worse results, and so have elected not to publish results.

The poor cross-subject validation results in experiment 3 reflect the findings of O'Brien et al. [67], who tried training data on healthy populations and testing them on individuals who had suffered a stroke. In O'Brien et al.'s findings, [67], the recall rate of classification accuracy was 53%, increased to 75% when trained on patients with only mild gait impairments. They also had the benefit of a significantly higher training population (15 healthy participants and 30 stroke patients), so from their results it is inconclusive whether it was the larger training dataset or the condition of the gait that was the key factor in improving performance. Due to only having four participants, this investigation did not have the capability of being able to investigate different types of severity in gait impairment for the ILLA population, and all four of the participants were considered to only have mild gait impairment. There should also be consideration of the volunteering bias in this study: participants who are willing to volunteer will likely have a healthy lifestyle which will allow them to participate with minimal physical challenge, and the machine learning performance generated from this study may not reflect the performance if applied to the general ILLA population. Though the sample size was small enough that no comment can be made on statistical significance, the majority of recruited amputee participants had traumatic-related amputation, and traumatic amputees typically tend to exhibit greater physical and functional mobility over amputees with a vascular-related amputation [68]. Therefore, this study cannot implicate how effective the system would be on individuals who for instance have severe comorbidities impacting on their gait mobility. In contrast, Vageskar [69] was able to achieve much higher recognition accuracies when training on a healthy population, gaining 86.5% accuracy when trained in free-living conditions, however they were able to do so by simplifying the classification problem; most notably, they relabelled stair movement as just "walking" and did not account for sloping

movement at all. Including those activities as separate labels would likely have caused a significant drop in their classification performance.

5. Conclusions

This pilot study was an investigation of how supervised learning systems can recognize activities carried out by healthy individuals and those with lower limb amputation, finding reasonable machine learning accuracies with room for improvement. SVM and LSTM classifiers were found to be the best off-the-shelf approach to classification of activities, achieving 77% recognition accuracies at a basic level of label resolution. Generally, there was scope to expand walking activities to include hard and soft terrains alongside walking uphill, downhill, upstairs, and downstairs. It was not feasible to try and recognize ILLA activities when the system was trained only on healthy individuals or other ILLAs, though a small recruitment sample and confounding factors on the types of terrains traversed by participants limit the generalizability of these findings. Future endeavours should investigate more intricate classifier designs and include a wider range of ILLA participants with varying degrees of amputation and mobility.

Author Contributions: Conceptualization, A.J.; methodology, A.J.; software, A.J.; validation, A.J.; formal analysis, A.J.; investigation, A.J.; resources, A.J., L.M. and A.B.; data curation, A.J.; writing—original draft preparation, A.J.; writing—review and editing, A.J., L.M., A.B., L.S. and V.S.; visualization, A.J.; supervision, L.M. and A.B.; project administration, A.B.; funding acquisition, A.B. All authors have read and agreed to the published version of the manuscript.

Funding: This research was funded indirectly as part of the author A.J.'s PhD sponsorship. The sponsorship is jointly funded by the EPSRC and PAL Technologies Ltd. The open access publication fees for this paper was funded by the University of Strathclyde.

Institutional Review Board Statement: The study was conducted in accordance with the Declaration of Helsinki. Ethical approval for data collection involving individuals with lower limb amputation was granted by the University of Strathclyde's University Ethics Committee prior to conducting the investigation (Ref: UEC20/55). Ethical approval for data collection involving healthy individuals with no gait impairment was granted by the University of Strathclyde's Departmental Ethics Committee for Biomedical Engineering prior to conducting the investigation (Ref: BioMed/2020/293).

Informed Consent Statement: Informed consent was obtained from all participants involved in the study. Additionally, written informed consent has been obtained from the participants to publish this paper.

Data Availability Statement: The accelerometer data presented in this study are openly available in Alexander Jamieson's PURE account, DOI: 10.15129/ae451315-5258-4a07-8eb4-204e4d2e357f, see Jamieson [70]. The video files of the original recordings are not included in this dataset to protect the privacy of the participants.

Acknowledgments: The authors would like to thank PAL Technologies for the supply of ActivPAL™ devices used in this study.

Conflicts of Interest: A.J. receives grants from PAL Technologies Ltd. as part of his PhD funding, PAL Technologies manufactures the ActivPAL™ which is the primary sensory device used in this investigation. PAL Technologies had no direct involvement in the design of the study itself, they provided the devices to the researcher and gave assistance with technical issues relating to the ActivPAL™.

Appendix A. Feature List

Appendix A.1. Condensed Feature List

This appendix contains details about the features used in "Human Activity Recognition of Individuals with Lower Limb Amputation in Free-Living Conditions: A Pilot Study". A full list of features is available from the lead author's repository [70].

Table A1. Features used in investigation.

Feature	Explanation
Statistical and Time Domain Features	
Mean	$\bar{X} = \frac{\sum_{i=1}^n X_i}{n}$
Median	$Med(X) = \frac{(X[\frac{n-1}{2}] + X[\frac{n+1}{2}])}{2}$
Variance	$\sigma^2 = \frac{\sum_{i=1}^n (X_i - \bar{X})^2}{n-1}$
Root Mean Square	$RMS(X) = \sqrt{\frac{1}{n}(\bar{X})^2}$
Crest Factor	$c(X) = \frac{ X_{peak} }{RMS(\bar{X})}$
L1 Norm	$L1(X) = \sum_{i=1}^n X$
L2 Norm	$L2(X) = \sqrt{\sum_{i=1}^n X ^2}$
Variance of the sample-wise Norm ^a	$\sigma_{norm}^2 = \sigma^2(L2(X_i, Y_i, Z_i)) ; i = 1, 2 \dots n$
Skewness	$SV(X) = \frac{\sum_{i=1}^n (X_i - \bar{X})^3}{(N-1)*\sigma^3}$
Kurtosis	$KV(X) = \frac{\sum_{i=1}^n (X_i - \bar{X})^4}{(N-1)*\sigma^4}$
25th Quartile	$Q_{25}(X) = X[\frac{1}{4(n+1)}]$
75th Quartile	$Q_{75}(X) = X[\frac{3}{4(n+1)}]$
Interquartile Range	$IQR(X) = Q_{75}(X) - Q_{25}(X)$
Maximum	$Max(X)$
Minimum	$Min(X)$
Range	$Max(X) - Min(X)$
Mean Absolute Deviation	$MAD = \frac{1}{n} \sum_{i=1}^n X_i - \bar{X} $
Signal Magnitude Area	$SMA = \frac{1}{n} (\sum_{i=1}^n X_i + \sum_{i=1}^n Y_i + \sum_{i=1}^n Z_i)$
Energy	$E(X) = \sum_{i=1}^n X_i ^2$
Power	$P(X) = \frac{\sqrt{\sum_{i=1}^n X_i ^2}}{n}$
Entropy	$H(X) = \sum_{i=1}^n (P_i) \log_2 (P_i)$ where $P_i = \frac{\frac{X_i}{Max(X_i)}}{\sum_{i=1}^n (\frac{X_i}{Max(X_i)})}$
Integral Features ^b	[23]
Inter-axis Correlation Coefficients	$r = \frac{\sum_{i=1}^n (X_i - \bar{X})(Y_i - \bar{Y})}{\sqrt{\sum_{i=1}^n (X_i - \bar{X})^2 (Y_i - \bar{Y})^2}}$
Eigenvalues of Dominant Direction	[71]
Frequency Features	
Spectral Energy	$E(FFT(X))$
Spectral Centroid	$SC = \frac{\sum_{k=1}^n X(i,k) * f(k)}{\sum_{k=1}^n X(i,k) }$
Spectral Entropy	$H(FFT(X))$
Cepstral Coefficients	[24]

Table A1. Cont.

Feature	Explanation
Wavelet Features ^c	
Tamura Coefficients	[72]
Nyan Coefficients	[73]
Sekine Coefficients	[74]
Wang Coefficients	[75]
Fractal Dimension	[76]
Preece Coefficients	[77]

Abbreviations: X —one dimensional time series input signal of acceleration. n —number of samples within a segment. σ —standard deviation of time series signal. $|X_{peak}|$ —the magnitude of an identified peak within a time series signal. Y —a secondary time series input signal of acceleration that is not the same as the primary input signal (X). Z —a tertiary time series input signal of acceleration that is not the same as the primary input signal (X) or the secondary input signal (Y). $\int_0^T |X| dt$ —trapezoidal integration of time series signal over sampling period T where T is the inverse of the sampling frequency. $FFT(X)$ —Fast Fourier Transformation of signal X . k —frequency bin from a FFT; $f(k)$ —frequency of bin k in Hz, ^a—to elaborate, the variance of the sample-wise norm is calculated by acquiring the L2 norm of the X , Y and Z axis for each sample of the input signal from 1 to n . σ_{norm}^2 is the variance of the calculated norms across the entire segment. ^b—The integral features used from Rosati et al. [23] are: “Mean of integral of Superior-Inferior (SI) acceleration”, “Mean of double integral of SI acceleration”, “Mean of integral of Anterio-Posterior (AP) acceleration”, “Root Mean Square of integral of AP acceleration”, “Root Mean Square of double integral of AP acceleration”. Using the axes of the ActivPAL during gait, SI and AP directions are approximated as the X and Z axis respectively. ^c—All Wavelet Features are utilised from Table 1 of Preece et al. [77] and is the recommended starting point for understanding the wavelet features used.

Appendix B. Confusion Matrices

Appendix B contains the confusion matrices obtained from experiments 2 and 3 for Support Vector Machine (SVM) and Long-Short Term Memory (LSTM) Network classifiers.

Appendix B.1. Experiment 2: SVM and LSTM Confusion Matrices



Figure A1. Support Vector Machine Confusion Chart for Level 1 label resolution.

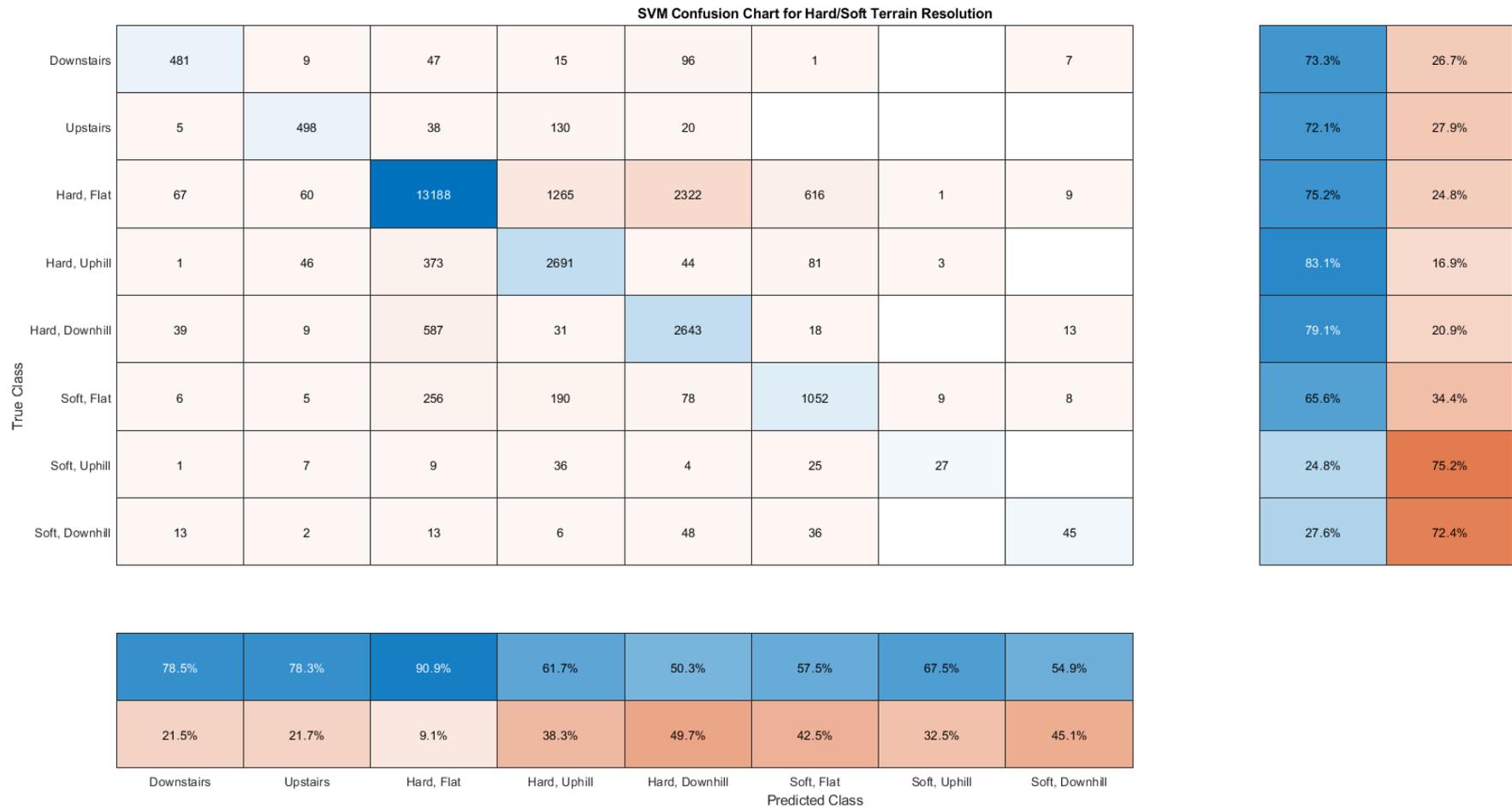


Figure A2. Support Vector Machine Confusion Chart for Level 2 label resolution.

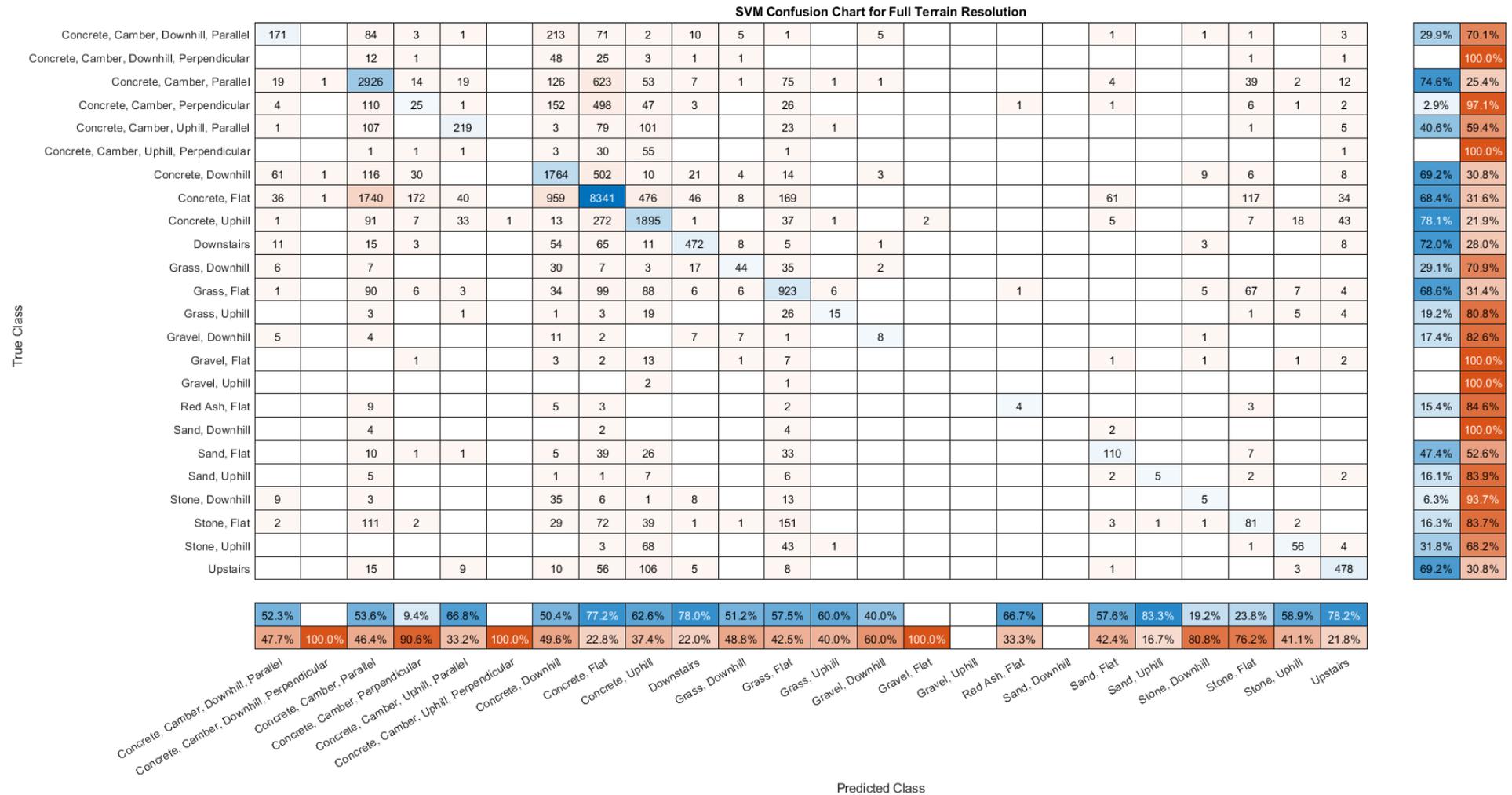


Figure A3. Support Vector Machine Confusion Chart for Level 3 label resolution.

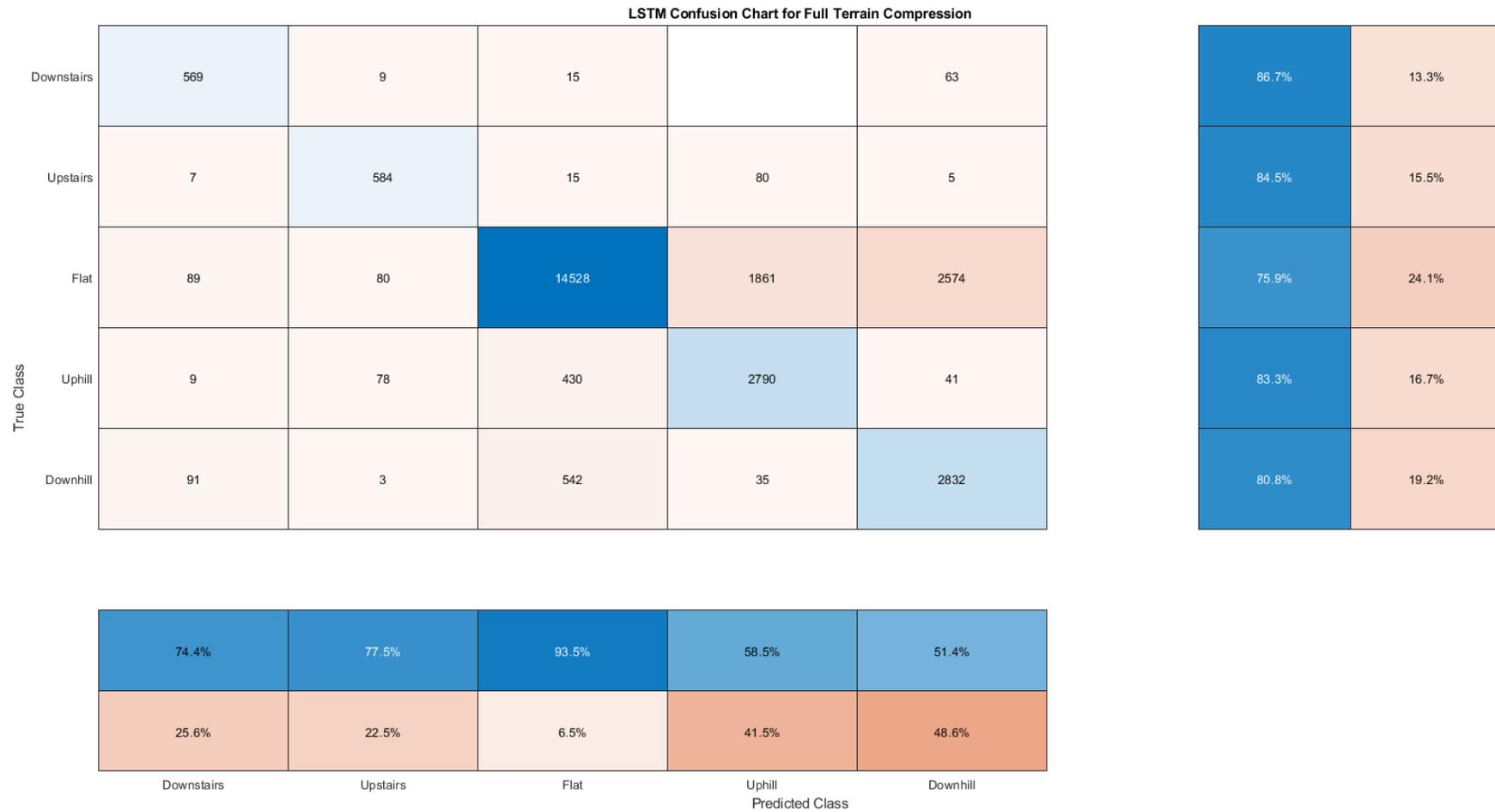


Figure A4. Long Short-Term Memory Network Confusion Chart for Level 1 label resolution.

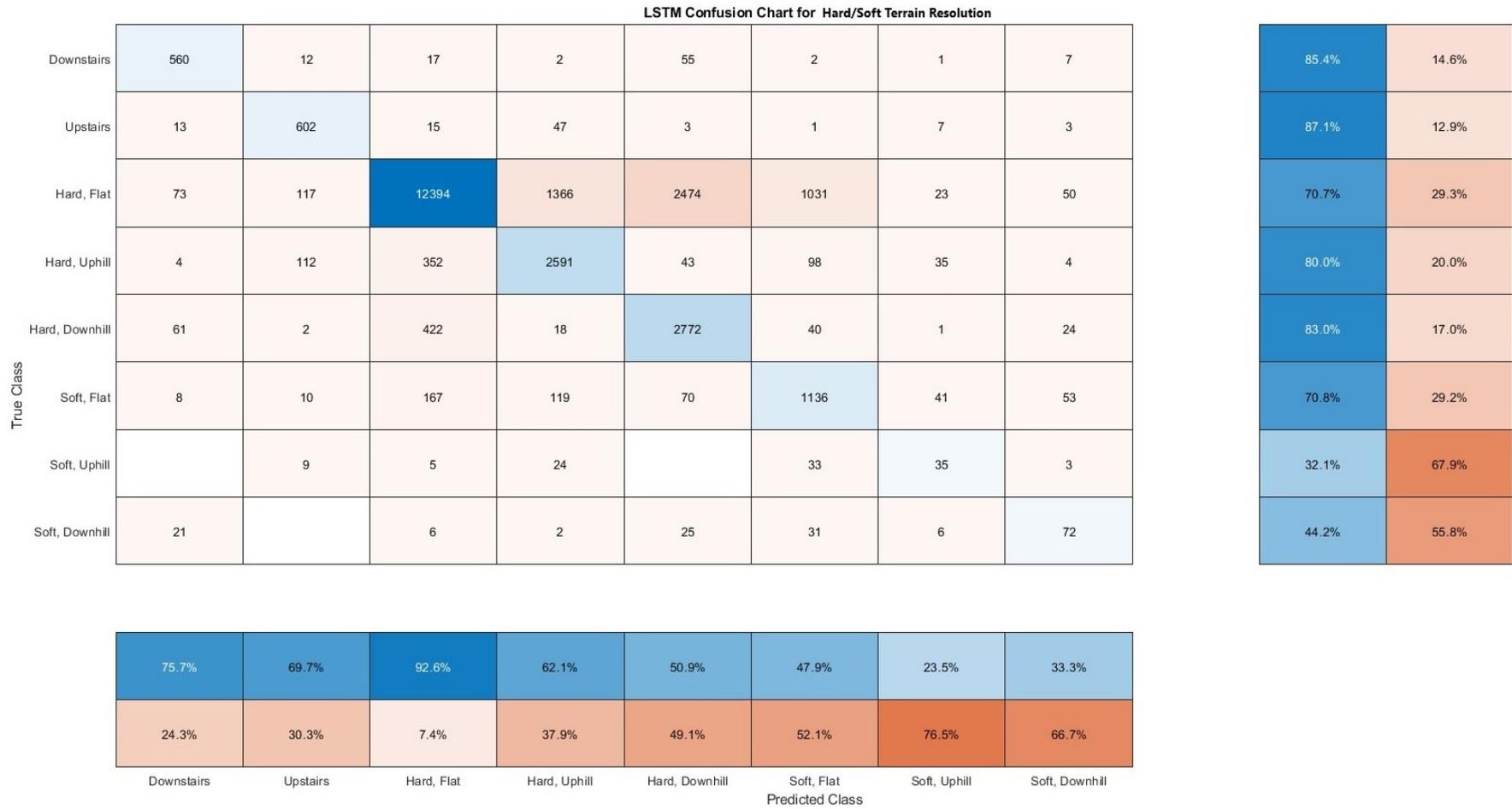


Figure A5. Long Short-Term Memory Network Confusion Chart for Level 2 label resolution.

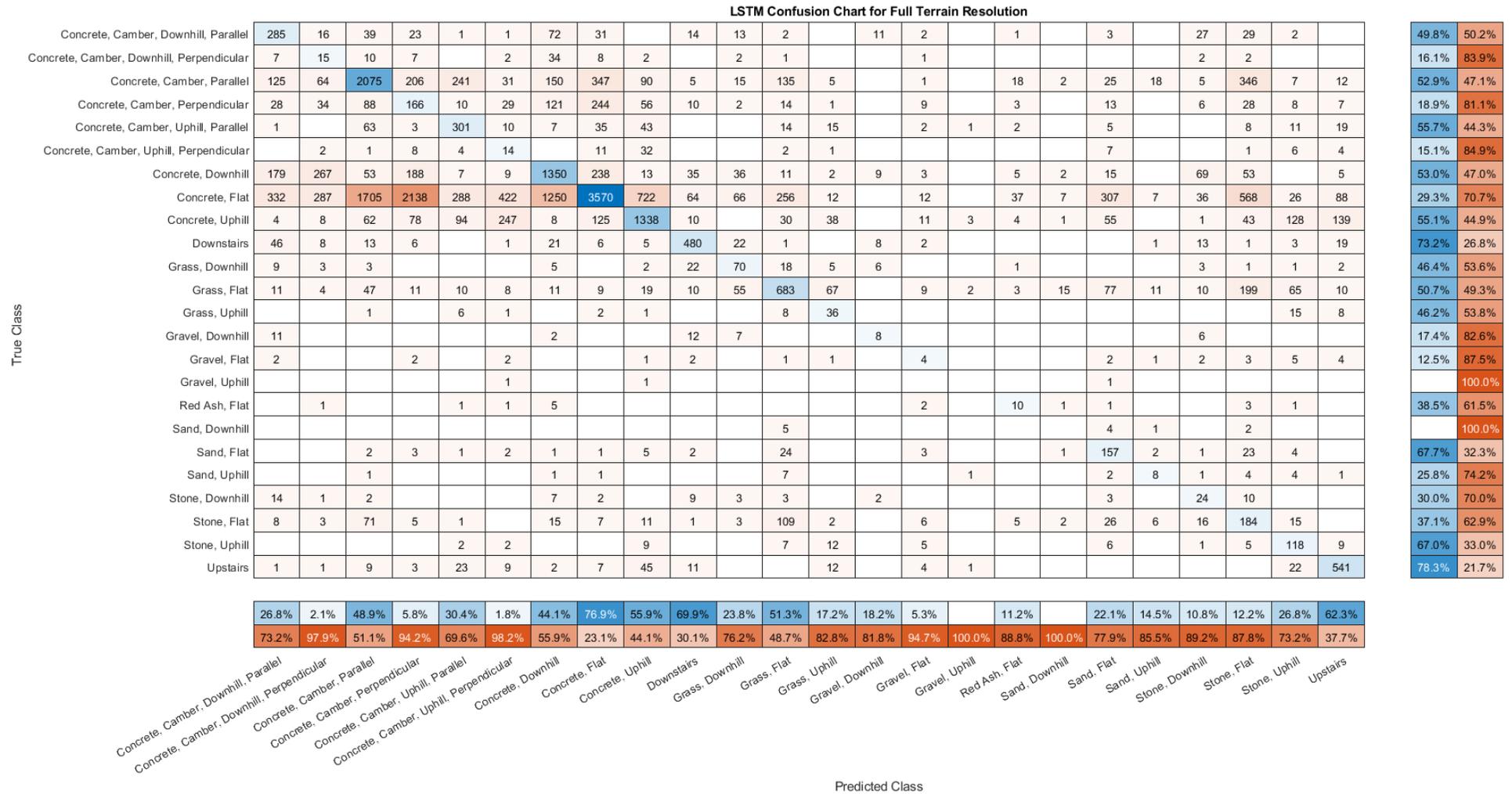


Figure A6. Long Short-Term Memory Network Confusion Chart for Level 3 label resolution.

Appendix B.2. Experiment 3: Confusion Matrices

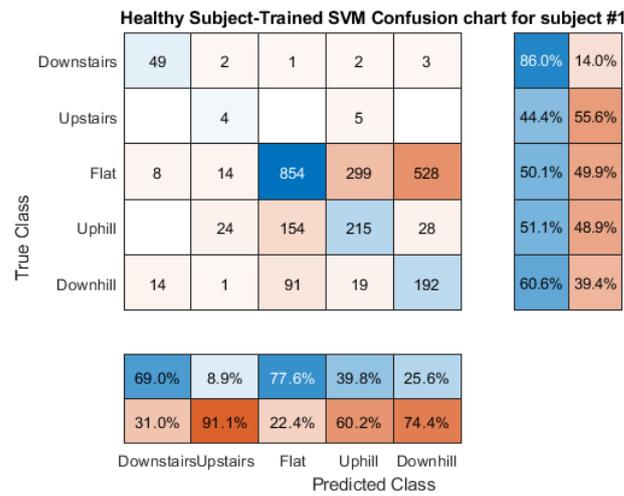


Figure A7. Support Vector Machine Confusion Chart for ILLA participant A1 as trained by healthy individuals.

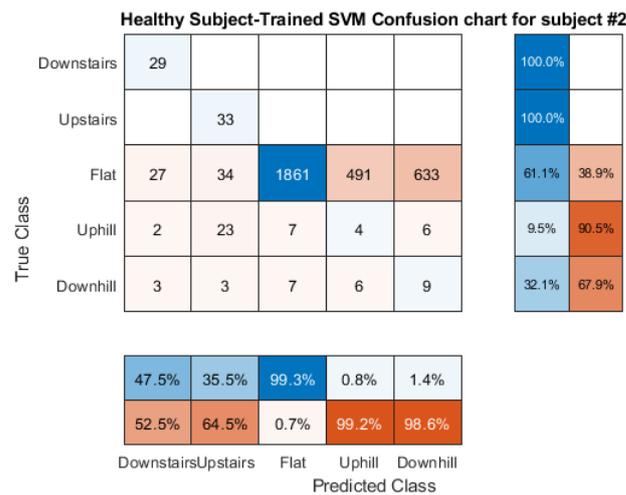


Figure A8. Support Vector Machine Confusion Chart for ILLA participant A2 as trained by healthy individuals.

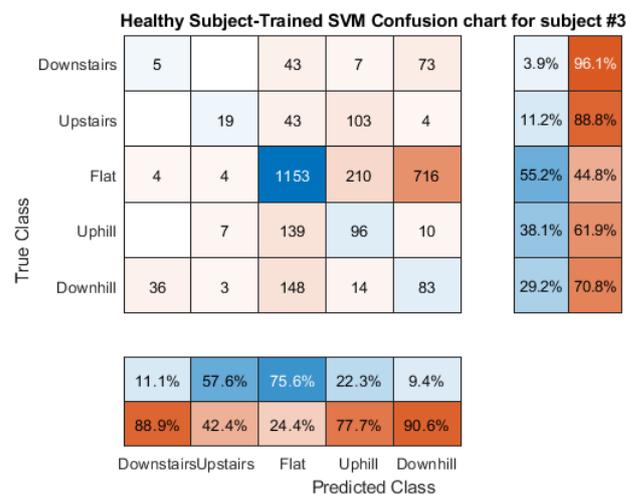


Figure A9. Support Vector Machine Confusion Chart for ILLA participant A3 as trained by healthy individuals.

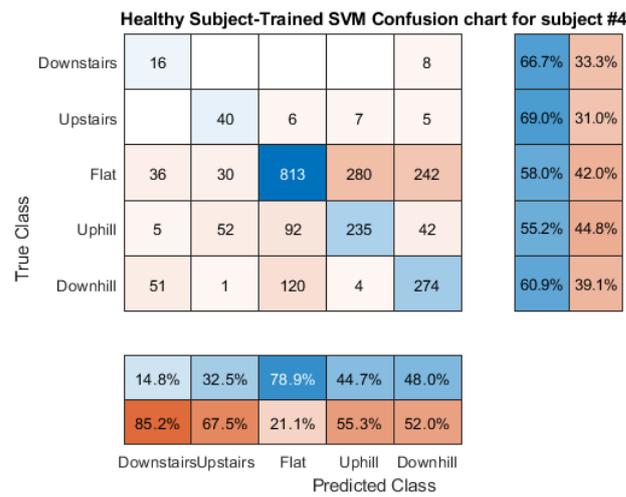


Figure A10. Support Vector Machine Confusion Chart for ILLA participant A4 as trained by healthy individuals.

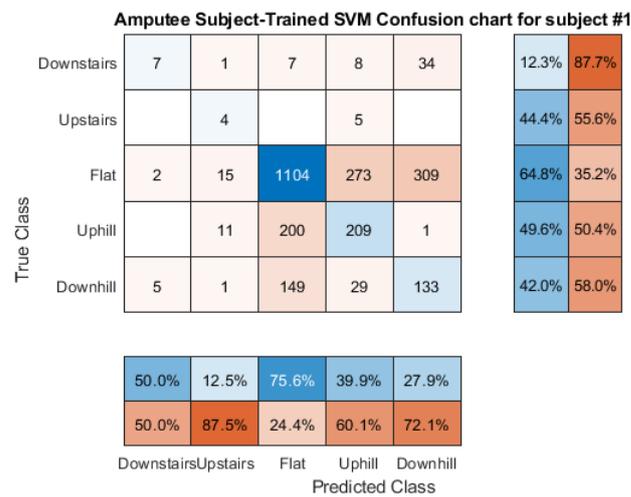


Figure A11. Support Vector Machine Confusion Chart for ILLA participant A1 as trained by all other ILLA participants.

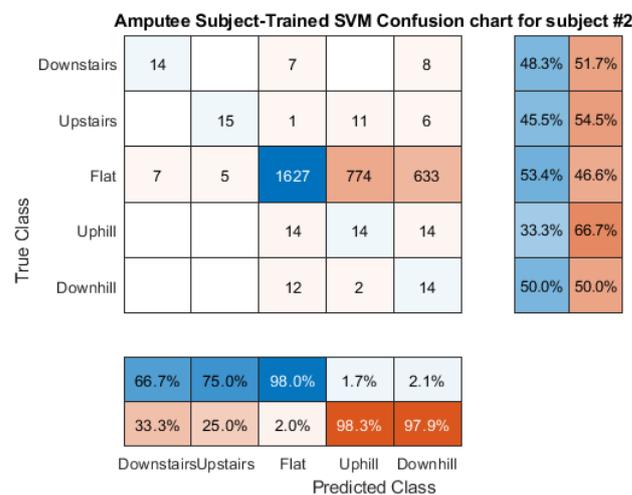


Figure A12. Support Vector Machine Confusion Chart for ILLA participant A2 as trained by all other ILLA participants.

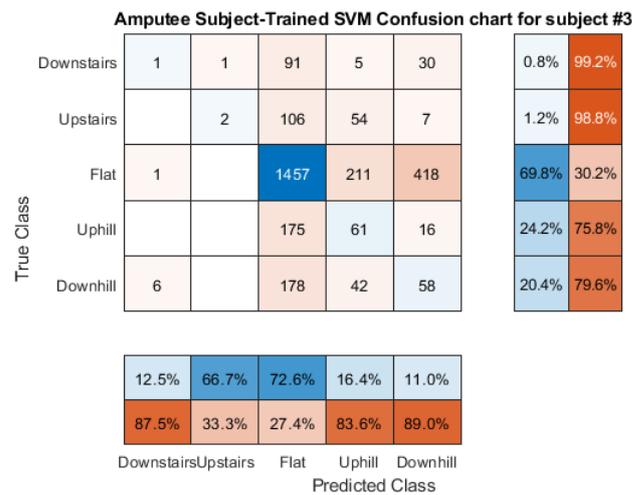


Figure A13. Support Vector Machine Confusion Chart for ILLA participant A3 as trained by all other ILLA participants.

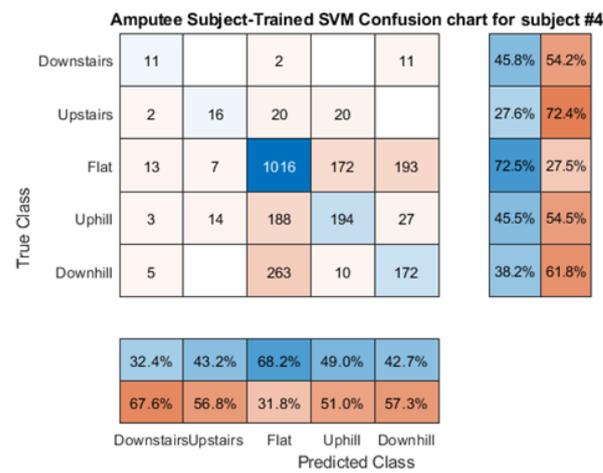


Figure A14. Support Vector Machine Confusion Chart for ILLA participant A4 as trained by all other ILLA participants

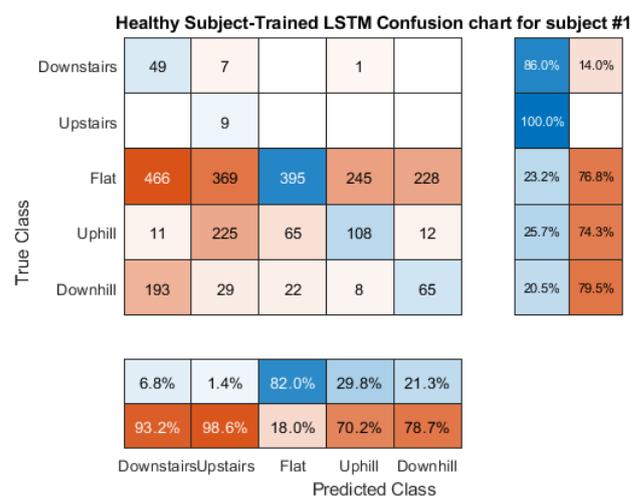


Figure A15. Long Short-Term Memory Network Confusion Chart for ILLA participant A1 as trained by healthy individuals.

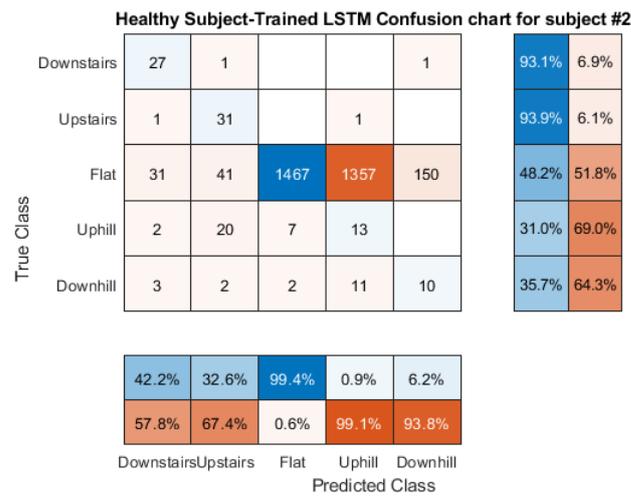


Figure A16. Long Short-Term Memory Network Confusion Chart for ILLA participant A2 as trained by healthy individuals.

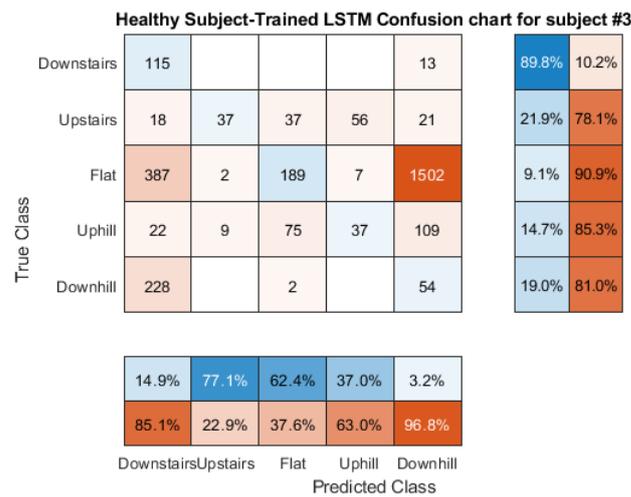


Figure A17. Long Short-Term Memory Network Confusion Chart for ILLA participant A3 as trained by healthy individuals.

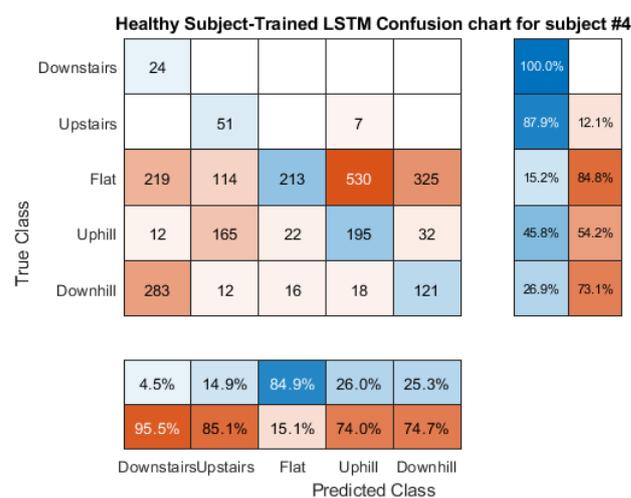


Figure A18. Long Short-Term Memory Network Confusion Chart for ILLA participant A4 as trained by healthy individuals.

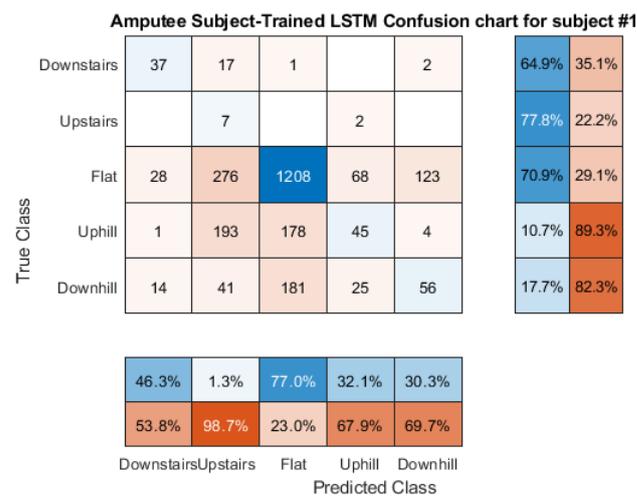


Figure A19. Long Short-Term Memory Network Confusion Chart for ILLA participant A1 as trained by all other ILLA participants.

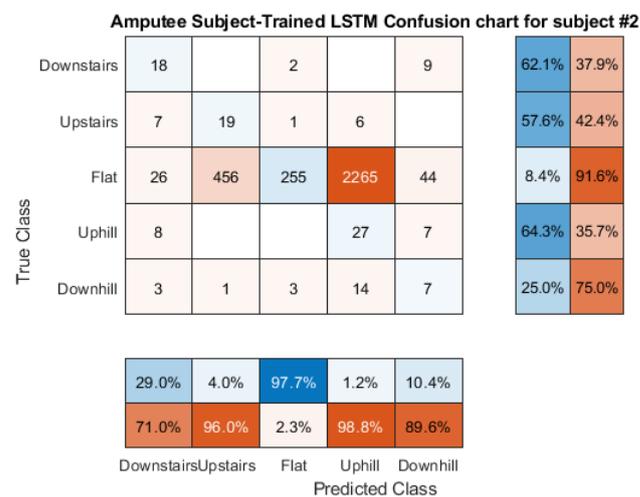


Figure A20. Long Short-Term Memory Network Confusion Chart for ILLA participant A2 as trained by all other ILLA participants.

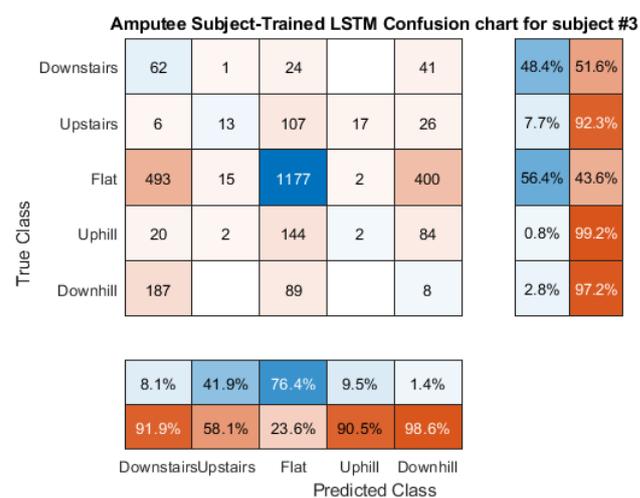


Figure A21. Long Short-Term Memory Network Confusion Chart for ILLA participant A3 as trained by all other ILLA participants.

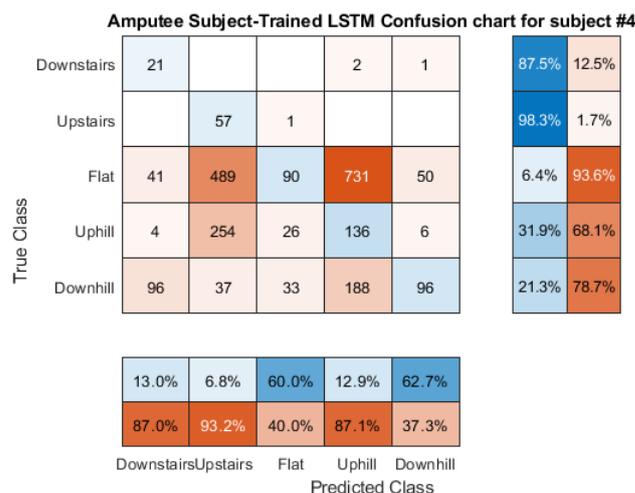


Figure A22. Long Short-Term Memory Network Confusion Chart for ILLA participant A4 as trained by all other ILLA participants.

References

1. Bussmann, J.B.; Grootcholten, E.A.; Stam, H.J. Daily physical activity and heart rate response in people with a unilateral transtibial amputation for vascular disease. *Arch. Phys. Med. Rehabil.* **2004**, *85*, 240–244. [\[CrossRef\]](#)
2. Langford, J.; Dillon, M.P.; Granger, C.L.; Barr, C. Physical activity participation amongst individuals with lower limb amputation. *Disabil. Rehabil.* **2019**, *41*, 1063–1070. [\[CrossRef\]](#) [\[PubMed\]](#)
3. Bragaru, M.; Van Wilgen, C.P.; Geertzen, J.H.B.; Ruijs, S.G.J.B.; Dijkstra, P.U.; Dekker, R. Barriers and Facilitators of Participation in Sports: A Qualitative Study on Dutch Individuals with Lower Limb Amputation. *PLoS ONE* **2013**, *8*, e59881. [\[CrossRef\]](#) [\[PubMed\]](#)
4. Deans, S.; Burns, D.; McGarry, A.; Murray, K.; Mutrie, N. Motivations and barriers to prosthesis users participation in physical activity, exercise and sport: A review of the literature. *Prosthet. Orthot. Int.* **2012**, *36*, 260–269. [\[CrossRef\]](#)
5. Littman, A.J.; Boyko, E.J.; Thompson, M.L.; Haselkorn, J.K.; Sangeorzan, B.J.; Arterburn, D.E. Physical activity barriers and enablers in older Veterans with lower-limb amputation. *J. Rehabil. Res. Dev.* **2014**, *51*, 895–906. [\[CrossRef\]](#)
6. Bragaru, M.; Dekker, R.; Geertzen, J.H.B.; Dijkstra, P.U. Amputees and Sports. *Sport. Med.* **2011**, *41*, 721–740. [\[CrossRef\]](#)
7. Wetterhahn, K.; Hanson, C.; Levy, C. Effect of Participation in Physical Activity on Body Image of Amputees. *Am. J. Phys. Med. Rehabil. / Assoc. Acad. Physiatr.* **2002**, *81*, 194–201. [\[CrossRef\]](#)
8. Chadwell, A.; Diment, L.; Micó-Amigo, M.; Morgado Ramírez, D.Z.; Dickinson, A.; Granat, M.; Kenney, L.; Kheng, S.; Sobuh, M.; Ssekitolesko, R.; et al. Technology for monitoring everyday prosthesis use: A systematic review. *J. NeuroEng. Rehabil.* **2020**, *17*. [\[CrossRef\]](#)
9. Cruciani, F.; Cleland, I.; Nugent, C.; Mccullagh, P.; Synnes, K.; Hallberg, J. Automatic Annotation for Human Activity Recognition in Free Living Using a Smartphone. *Sensors* **2018**, *18*, 2203. [\[CrossRef\]](#) [\[PubMed\]](#)
10. Ellis, K.; Kerr, J.; Godbole, S.; Lanckriet, G. Multi-sensor physical activity recognition in free-living. In Proceedings of the 2014 ACM International Joint Conference on Pervasive and Ubiquitous Computing: Adjunct Publication, Seattle, WA, USA, 14–17 September 2014. [\[CrossRef\]](#)
11. Fullerton, E.; Heller, B.; Munoz-Organero, M. Recognizing Human Activity in Free-Living Using Multiple Body-Worn Accelerometers. *IEEE Sens. J.* **2017**, *17*, 5290–5297. [\[CrossRef\]](#)
12. Cruciani, F.; Sun, C.; Zhang, S.; Nugent, C.; Li, C.; Song, S.; Cheng, C.; Cleland, I.; Mccullagh, P. A Public Domain Dataset for Human Activity Recognition in Free-Living Conditions. In Proceedings of the 2019 IEEE SmartWorld, Ubiquitous Intelligence and Computing, Advanced and Trusted Computing, Scalable Computing and Communications, Cloud and Big Data Computing, Internet of People and Smart City Innovation (SmartWorld/SCALCOM/UIC/ATC/CBDCOM/IOP/SCI), Leicester, UK, 19–23 August 2019. [\[CrossRef\]](#)
13. Labarrière, F.; Thomas, E.; Calistri, L.; Optasanu, V.; Gueugnon, M.; Ornetti, P.; Laroche, D. Machine Learning Approaches for Activity Recognition and/or Activity Prediction in Locomotion Assistive Devices—A Systematic Review. *Sensors* **2020**, *20*, 6345. [\[CrossRef\]](#)
14. Zhang, K.; Xiong, C.; Zhang, W.; Liu, H.; Lai, D.; Rong, Y.; Fu, C. Environmental Features Recognition for Lower Limb Prostheses Toward Predictive Walking. *IEEE Trans. Neural Syst. Rehabil. Eng.* **2019**, *27*, 465–476. [\[CrossRef\]](#)
15. Zhang, K.; Zhang, W.; Xiao, W.; Liu, H.; De Silva, C.W.; Fu, C. Sequential Decision Fusion for Environmental Classification in Assistive Walking. *IEEE Trans. Neural Syst. Rehabil. Eng.* **2019**, *27*, 1780–1790. [\[CrossRef\]](#)
16. Jung, I. A review of privacy-preserving human and human activity recognition. *Int. J. Smart Sens. Intell. Syst.* **2020**, *13*, 1–13. [\[CrossRef\]](#)

17. Grant, P.M.; Ryan, C.G.; Tigbe, W.W.; Granat, M.H. The validation of a novel activity monitor in the measurement of posture and motion during everyday activities. *Br. J. Sport. Med.* **2006**, *40*, 992–997. [CrossRef]
18. Lyden, K.; Keadle, S.K.; Staudenmayer, J.; Freedson, P.S. The activPAL™ Accurately Classifies Activity Intensity Categories in Healthy Adults. *Med. Sci. Sport. Exerc.* **2017**, *49*, 1022–1028. [CrossRef] [PubMed]
19. Buis, A.; Dumbleton, T.; Murray, K.; McHugh, B.; McKay, G.; Sexton, S. Measuring the Daily Stepping Activity of People with Transtibial Amputation Using the ActivPAL™ Activity Monitor. *JPO J. Prosth. Orthot.* **2014**, *26*, 43–47. [CrossRef]
20. Van Hees, V.T.; Gorzelniak, L.; Dean León, E.C.; Eder, M.; Pias, M.; Taherian, S.; Ekelund, U.; Renström, F.; Franks, P.W.; Horsch, A.; Brage, S. Separating Movement and Gravity Components in an Acceleration Signal and Implications for the Assessment of Human Daily Physical Activity. *PLoS ONE* **2013**, *8*, e61691. [CrossRef] [PubMed]
21. Frølich, L.; Dowding, I. Removal of muscular artifacts in EEG signals: A comparison of linear decomposition methods. *Brain Inform.* **2018**, *5*, 13–22. [CrossRef] [PubMed]
22. Dehghani, A.; Sarbishei, O.; Glatard, T.; Shihab, E. A Quantitative Comparison of Overlapping and Non-Overlapping Sliding Windows for Human Activity Recognition Using Inertial Sensors. *Sensors* **2019**, *19*, 5026. [CrossRef]
23. Rosati, S.; Balestra, G.; Knaflitz, M. Comparison of Different Sets of Features for Human Activity Recognition by Wearable Sensors. *Sensors* **2018**, *18*, 4189. [CrossRef] [PubMed]
24. San-Segundo, R.; Montero, J.M.; Barra-Chicote, R.; Fernandez, F.; Pardo, J.M. Feature extraction from smartphone inertial signals for human activity segmentation. *Signal Process.* **2016**, *120*, 359–372. [CrossRef]
25. Abdul Yamin, N.A.A.; Amran, M.; Basaruddin, K.; Salleh, A.; Rusli, W. Ground reaction force response during running on different surface hardness. *ARPN J. Eng. Appl. Sci.* **2017**, *12*, 2313–2318.
26. Nedorubova, A.; Kadyrova, A.; Khlyupin, A. Human Activity Recognition using Continuous Wavelet Transform and Convolutional Neural Networks. *arXiv* **2021**, arXiv:2106.12666.
27. Meg. Elevation on Strava FAQs, 2021. Available online: <https://support.strava.com/hc/en-us/articles/115001294564-Elevation-on-Strava-FAQs> (accessed on 1 September 2021).
28. Bosch. Pressure sensor BMP280. Available online: <https://www.bosch-sensortec.com/products/environmental-sensors/pressure-sensors/bmp280/> (accessed on 1 September 2021).
29. Barberi, J. GPS vs. Smartphone vs. Cyclocomputer (Round 2). Available online: <https://www.singletracks.com/mtb-gear/gps-accuracy-gps-vs-smartphone-vs-cyclocomputer/> (accessed on 1 September 2021).
30. Meg. Why is GPS Data Sometimes Inaccurate? Available online: <https://support.strava.com/hc/en-us/articles/216917917-Why-is-GPS-data-sometimes-inaccurate-> (accessed on 1 September 2021).
31. 2010 ADA standards for accessible design. Available online: https://www.ada.gov/2010ADASTandards_index.htm (accessed on 1 September 2021)
32. Guo, X.; Yin, Y.; Dong, C.; Yang, G.; Zhou, G. On the Class Imbalance Problem. In Proceedings of the 2008 Fourth International Conference on Natural Computation, NW Washington, DC, USA, 18–20 October 2008. [CrossRef]
33. Chawla, N.; Bowyer, K.; Hall, L.; Kegelmeyer, W. SMOTE: Synthetic Minority Over-sampling Technique. *J. Artif. Intell. Res. (JAIR)* **2002**, *16*, 321–357. [CrossRef]
34. Anguita, D.; Ghio, A.; Oneto, L.; Parra, X.; Reyes-Ortiz, J.L. Human Activity Recognition on Smartphones Using a Multiclass Hardware-Friendly Support Vector Machine. In *Lecture Notes in Computer Science*; Springer: Berlin, Germany, 2012; pp. 216–223. [CrossRef]
35. Kaghyan, S.; Sarukhanyan, H. Activity Recognition using K-nearest neighbor algorithm on smartphone with tri-axial accelerometer. *Int. J. Inf. Model. Anal.* **2012**, *1*, 146–156.
36. Xu, L.; Yang, W.; Cao, Y.; Li, Q. Human activity recognition based on random forests. In Proceedings of the 2017 13th International Conference on Natural Computation, Fuzzy Systems and Knowledge Discovery (ICNC-FSKD), Guilin, China, 29–31 July 2017. [CrossRef]
37. Subasi, A.; Dammas, D.H.; Alghamdi, R.D.; Makawi, R.A.; Albiety, E.A.; Brahimi, T.; Sarirete, A. Sensor Based Human Activity Recognition Using Adaboost Ensemble Classifier. *Procedia Comput. Sci.* **2018**, *140*, 104–111. [CrossRef]
38. Gupta, P.; Dallas, T. Feature Selection and Activity Recognition System Using a Single Triaxial Accelerometer. *IEEE Trans. Biomed. Eng.* **2014**, *61*, 1780–1786. [CrossRef]
39. Inoue, M.; Inoue, S.; Nishida, T. Deep recurrent neural network for mobile human activity recognition with high throughput. *Artif. Life Robot.* **2018**, *23*, 173–185. [CrossRef]
40. Bengio, Y. *Practical Recommendations for Gradient-Based Training of Deep Architectures*; Springer: Berlin, Heidelberg, 2012.
41. Kingma, D.P.; Ba, J. Adam: A Method for Stochastic Optimization. *arXiv* **2015**, arXiv:1412.6980.
42. Bergstra, J.; Bengio, Y. Random Search for Hyper-Parameter Optimization. *J. Mach. Learn. Res.* **2012**, *13*, 281–305.
43. Marcot, B.G.; Hanea, A.M. What is an optimal value of k in k-fold cross-validation in discrete Bayesian network analysis? *Comput. Stat.* **2021**, *36*, 2009–2031. [CrossRef]
44. Fearn, T. Double Cross-Validation. *NIR News* **2010**, *21*, 14–15. [CrossRef]
45. Davie-Smith, F.; Heberton, J.; Scott, H. A Survey of the Lower Limb Amputee Population in Scotland 2015 Full Report. Scottish Physiother. Amputee Res. Group. Available online: <https://bacpar.csp.org.uk/publications/sparg-report-2015> (accessed on 1 September 2021)

46. Park, S.E.; Ho, Y.J.; Chun, M.H.; Choi, J.; Moon, Y. Measurement and Analysis of Gait Pattern during Stair Walk for Improvement of Robotic Locomotion Rehabilitation System. *Appl. Bionics Biomech.* **2019**, *2019*, 1495289. [[CrossRef](#)]
47. Kimel-Naor, S.; Gottlieb, A.; Plotnik, M. The effect of uphill and downhill walking on gait parameters: A self-paced treadmill study. *J. Biomech.* **2017**, *60*, 142–149. [[CrossRef](#)]
48. Allamy, H.; Khan, R. Methods to Avoid Over-Fitting and Under-Fitting In Supervised Machine Learning (Comparative Study). In *Computer Science, Communication & Instrumentation Devices*; World Press: Singapore, 2014, pp. 163–172. doi: 10.3850/978-981-09-5247-1_017 [[CrossRef](#)]
49. Zhao, Y.; Yang, R.; Chevalier, G.; Gong, M. Deep Residual Bidir-LSTM for Human Activity Recognition Using Wearable Sensors. *Math. Probl. Eng.* **2017**, *2017*, 7316954. [[CrossRef](#)]
50. Roggen, D.; Calatroni, A.; Rossi, M.; Holleczeck, T.; Förster, K.; Tröster, G.; Lukowicz, P.; Bannach, D.; Pirkl, G.; Ferscha, A.; et al. Collecting complex activity datasets in highly rich networked sensor environments. In Proceedings of the 2010 Seventh International Conference on Networked Sensing Systems (INSS), Kassel, Germany, 15–18 June 2010; pp. 233–240. [[CrossRef](#)]
51. Sun, J.; Fu, Y.; Li, S.; He, J.; Xu, C.; Tan, L. Sequential Human Activity Recognition Based on Deep Convolutional Network and Extreme Learning Machine Using Wearable Sensors. *J. Sens.* **2018**, *2018*, 8580959. [[CrossRef](#)]
52. Huang, Z.; Niu, Q.; You, I.; Pau, G. Acceleration Feature Extraction of Human Body Based on Wearable Devices. *Energies* **2021**, *14*, 924. [[CrossRef](#)]
53. Keras. Available online: <https://keras.io/> (accessed on 1 September 2021).
54. Qazi, N.; Raza, K. Effect of Feature Selection, SMOTE and under Sampling on Class Imbalance Classification. In Proceedings of the 2012 UKSim 14th International Conference on Computer Modelling and Simulation, Cambridge, UK, 28–30 March 2012. [[CrossRef](#)]
55. Flores, A.C.; Icoy, R.I.; Pena, C.F.; Gorro, K.D. An Evaluation of SVM and Naive Bayes with SMOTE on Sentiment Analysis Data Set. In Proceedings of the 2018 International Conference on Engineering, Applied Sciences, and Technology (ICEAST), Phuket, Thailand, 4–7 July 2018. [[CrossRef](#)]
56. Mai, J.; Xu, D.; Li, H.; Zhang, S.; Tan, J.; Wang, Q. Implementing a SoC-FPGA Based Acceleration System for On-Board SVM Training for Robotic Transtibial Prostheses. In Proceedings of the IEEE International Conference on Real-time Computing and Robotics (RCAR), Kandima, Maldives, 1–5 August 2018. [[CrossRef](#)]
57. Mai, J.; Chen, W.; Zhang, S.; Xu, D.; Wang, Q. Performance analysis of hardware acceleration for locomotion mode recognition in robotic prosthetic control. In Proceedings of the 2018 IEEE International Conference on Cyborg and Bionic Systems (CBS), Shenzhen, China, 25–27 October 2018. [[CrossRef](#)]
58. Shell, C.E.; Klute, G.K.; Neptune, R.R. Identifying classifier input signals to predict a cross-slope during transtibial amputee walking. *PLoS ONE* **2018**, *13*, e0192950. [[CrossRef](#)]
59. Stolyarov, R.; Burnett, G.; Herr, H. Translational Motion Tracking of Leg Joints for Enhanced Prediction of Walking Tasks. *IEEE Trans. Biomed. Eng.* **2018**, *65*, 763–769. [[CrossRef](#)] [[PubMed](#)]
60. Su, B.Y.; Wang, J.; Liu, S.Q.; Sheng, M.; Jiang, J.; Xiang, K. A CNN-Based Method for Intent Recognition Using Inertial Measurement Units and Intelligent Lower Limb Prosthesis. *IEEE Trans. Neural Syst. Rehabil. Eng.* **2019**, *27*, 1032–1042. [[CrossRef](#)]
61. Gyllensten, I.C.; Bonomi, A.G. Identifying Types of Physical Activity With a Single Accelerometer: Evaluating Laboratory-trained Algorithms in Daily Life. *IEEE Trans. Biomed. Eng.* **2011**, *58*, 2656–2663. [[CrossRef](#)]
62. Liu, S.; Gao, R.X.; John, D.; Staudenmayer, J.; Freedson, P.S. SVM-based multi-sensor fusion for free-living physical activity assessment. In Proceedings of the 2011 Annual International Conference of the IEEE Engineering in Medicine and Biology Society, Boston, MA, USA, 30 August–3 September 2011. [[CrossRef](#)]
63. Liu, L.; Wang, S.; Hu, B.; Qiong, Q.; Wen, J.; Rosenblum, D.S. Learning structures of interval-based Bayesian networks in probabilistic generative model for human complex activity recognition. *Pattern Recognit.* **2018**, *81*, 545–561. [[CrossRef](#)]
64. Hu, B.; Dixon, P.C.; Jacobs, J.V.; Dennerlein, J.T.; Schiffman, J.M. Machine learning algorithms based on signals from a single wearable inertial sensor can detect surface- and age-related differences in walking. *J. Biomech.* **2018**, *71*, 37–42. [[CrossRef](#)] [[PubMed](#)]
65. Hashmi, M.Z.U.H.; Riaz, Q.; Hussain, M.; Shahzad, M. What Lies Beneath One’s Feet? Terrain Classification Using Inertial Data of Human Walk. *Appl. Sci.* **2019**, *9*, 3099. [[CrossRef](#)]
66. Russell, B.; Mcdaid, A.; Toscano, W.; Hume, P. Moving the Lab into the Mountains: A Pilot Study of Human Activity Recognition in Unstructured Environments. *Sensors* **2021**, *21*, 654. [[CrossRef](#)]
67. O’Brien, C.M.; Duda, J.L.; Kitas, G.D.; Veldhuijzen Van Zanten, J.J.C.S.; Metsios, G.S.; Fenton, S.A.M. Measurement of sedentary time and physical activity in rheumatoid arthritis: An ActiGraph and activPAL™ validation study. *Rheumatol. Int.* **2020**, *40*, 1509–1518. [[CrossRef](#)] [[PubMed](#)]
68. Wezenberg, D.; Van Der Woude, L.H.; Faber, W.X.; De Haan, A.; Houdijk, H. Relation Between Aerobic Capacity and Walking Ability in Older Adults With a Lower-Limb Amputation. *Arch. Phys. Med. Rehabil.* **2013**, *94*, 1714–1720. [[CrossRef](#)]
69. Vageskar, E. Activity Recognition for Stroke Patients. Master’s Thesis, Norwegian University of Science and Technology, Trondheim, Norway, 2017.
70. Jamieson, A. *Dataset for: Construction of a Clinical Activity Monitoring Framework Based on Free-living Investigations of Individuals with Lower Limb Amputation*; University of Strathclyde: Glasgow, UK, 2021. [[CrossRef](#)]

71. Zhang, M.; Sawchuk, A.A. A Feature Selection-Based Framework for Human Activity Recognition Using Wearable Multimodal Sensors. In Proceedings of the 6th International Conference on Body Area Networks, Brussels, Belgium, 7–8 November 2011; pp. 92–98.
72. Tamura, T.; Sekine, M.; Ogawa, M.; Togawa, T.; Fukui, Y. Classification of acceleration waveforms during walking by wavelet transform. *Methods Inf. Med.* **1997**, *36*, 356–359.
73. Nyan, M.N.; Tay, F.E.H.; Seah, K.H.W.; Sitoh, Y.Y. Classification of gait patterns in the time–frequency domain. *J. Biomech.* **2006**, *39*, 2647–2656. [[CrossRef](#)] [[PubMed](#)]
74. Sekine, M.; Tamura, T.; Togawa, T.; Fukui, Y. Classification of waist-acceleration signals in a continuous walking record. *Med. Eng. Phys.* **2000**, *22*, 285–291. [[CrossRef](#)]
75. Wang, N.; Ambikairajah, E.; Lovell, N.H.; Celler, B.G. Accelerometry Based Classification of Walking Patterns Using Time-frequency Analysis. In Proceedings of the 2007 29th Annual International Conference of the IEEE Engineering in Medicine and Biology Society, Lyon, France, 23–26 August 2007. [[CrossRef](#)]
76. Sekine, M.; Tamura, T.; Akay, M.; Fujimoto, T.; Togawa, T.; Fukui, Y. Discrimination of walking patterns using wavelet-based fractal analysis. *IEEE Trans. Neural Syst. Rehabil. Eng.* **2002**, *10*, 188–196. [[CrossRef](#)] [[PubMed](#)]
77. Preece, S.J.; Goulermas, J.Y.; Kenney, L.P.J.; Howard, D. A Comparison of Feature Extraction Methods for the Classification of Dynamic Activities From Accelerometer Data. *IEEE Trans. Biomed. Eng.* **2009**, *56*, 871–879. [[CrossRef](#)] [[PubMed](#)]