

IMU-Based Deep Neural Networks: Prediction of Locomotor and Transition Intentions of an Osseointegrated Transfemoral Amputee

Julian Bruinsma^{ID} and Raffaella Carloni^{ID}, *Member, IEEE*

Abstract—This paper focuses on the design and comparison of different deep neural networks for the real-time prediction of locomotor and transition intentions of one osseointegrated transfemoral amputee using only data from inertial measurement units. The deep neural networks are based on convolutional neural networks, recurrent neural networks, and convolutional recurrent neural networks. The architectures' input are features in both the time domain and the time-frequency domain, which are derived from either one inertial measurement unit (placed above the prosthetic knee) or two inertial measurement units (placed above and below the prosthetic knee). The prediction of eight different locomotion modes (i.e., sitting, standing, level ground walking, stair ascent and descent, ramp ascent and descent, walking on uneven terrain) and the twenty-four transitions among them is investigated. The study shows that a recurrent neural network, realized with four layers of gated recurrent unit networks, achieves (with a 5-fold cross-validation) a mean F1 score of 84.78% and 86.50% using one inertial measurement unit, and 93.06% and 89.99% using two inertial measurement units, with or without sitting, respectively.

Index Terms—Deep neural networks, lower-limb prosthetic.

I. INTRODUCTION

FOR individuals with lower-limb amputation, the need to conveniently perform activities of daily living is critical [1]. A fundamental step in developing active lower-limb prostheses is to achieve an intuitive control, where the intentions of the user should be accurately predicted. To avoid discomfort in using the prosthetic leg and to reduce the cognitive load, the intentions of the user should be predicted and converted to a proper control input to the prosthesis within 300 ms [2].

In the current literature, a variety of data analysis and machine learning techniques has been proposed to translate data from inertial measurement units (IMUs) into locomotion

information in real-time. These pattern recognition techniques can be broadly divided into two categories, namely, methods based on feature engineering [3] and methods based on feature learning [4], either with handcrafted or raw input IMU data.

Feature engineering methods on IMU data have been studied for the recognition and prediction of both locomotor and transition intentions. In [5], handcrafted features in the time domain are extracted from IMUs data to compare different supervised machine learning algorithms (i.e., support vector machine, multi-layer perceptron, random forest, k -nearest-neighbour, discriminant analysis) for the disjoint prediction of locomotion modes and transitions of healthy subjects. In [6], linear discriminant analysis (LDA) is used to predict locomotion modes of transfemoral amputees from handcrafted features in the time domain of one IMU. LDAs are also used to predict both the locomotion modes and transitions of healthy subjects from handcrafted features in the time domain of IMUs combined with pressure sensors in [7], and to predict the locomotion modes of transfemoral amputees from handcrafted features in the time domain of IMUs combined with mechanical sensors in [8]. In [9], a triplet Markov model uses features in the time domain of one IMU for the recognition of both locomotion mode and gait phases of healthy subjects. A binary tree is used on raw data from one IMU to predict locomotion modes and transitions of healthy subjects in [10], while a gradient tree boosting method on several time-domain IMU data in combination with encoders and load cells is proposed in [11] for the locomotion prediction of transfemoral amputees. In [12], different methods are compared for the recognition of locomotion modes and transitions of healthy subjects by using IMUs in combination with several sensors.

Feature learning methods on IMU data, by means of deep neural networks, have also been used for the recognition of locomotion modes and for the prediction of both locomotor and transition intentions, with the main advantage of obtaining higher-level features from IMUs without relying on domain-specific knowledge [4]. Deep belief networks (DBNs) are used on features in the time-frequency domain of a triaxial accelerometer to recognize (loco)motion modes of both healthy and impaired subjects in [13], while a convolutional neural network (CNN) is used on time domain features of a triaxial accelerometer to recognize locomotion modes of healthy subjects in [14]. CNNs are used on raw-data from one IMU placed on the foot to recognize locomotion modes of healthy subjects in [15], on raw-data from multiple IMUs on the lower limbs to predict locomotion modes of healthy

Manuscript received April 13, 2021; revised May 26, 2021; accepted June 2, 2021. Date of publication June 7, 2021; date of current version June 11, 2021. This work was supported by the European Commission's Horizon 2020 Program as part of the Project MyLeg under Grant 780871. (Corresponding author: Raffaella Carloni.)

This work involved human subjects or animals in its research. Approval of all ethical and experimental procedures and protocols was granted by the University of Twente, The Netherlands, under Approval No. NL67247.044.18.

The authors are with the Faculty of Science and Engineering, Bernoulli Institute for Mathematics, Computer Science and Artificial Intelligence, University of Groningen, 9747 AG Groningen, The Netherlands (e-mail: j.bruinsma.6@student.rug.nl; r.carloni@rug.nl).

Digital Object Identifier 10.1109/TNSRE.2021.3086843

TABLE I

STATE OF THE ART OF MACHINE LEARNING TECHNIQUES (FEATURE ENGINEERING AND FEATURE LEARNING) FOR THE RECOGNITION AND/OR PREDICTION OF (LOCO-)MOTION MODE AND TRANSITION INTENTIONS BY MEANS OF IMU DATA OF HEALTHY AND/OR IMPAIRED SUBJECTS

Article	Method	Features	Mean Accuracy	Limb(s)	Motion(s)	Transit.	Subject(s)
Feature Engineering Methods							
[5] 2020	Several methods, prediction	IMU time domain	97.65%	Upper, lower	Locomotion (5 modes)	Yes (9 trans.)	10 Healthy
[6] 2018	LDA, prediction	IMU time domain	96.22%	Lower	Locomotion (5 modes)	No	6 Transtibial amputees
[7] 2014	LDA, prediction	IMU time domain & pressure insoles	99.71%	Lower	Locomotion (6 modes)	Yes (12 trans.)	7 Healthy
[8] 2013	LDA, prediction	IMU time domain & mechanical sensors	93.9%	Lower	Locomotion (5 modes)	No	6 Transfem. amputees
[9] 2019	Markov model, recognition	IMU time domain	99.2%	Lower	Locomotion (4 modes) and gait phases	No	10 Healthy
[10] 2021	Binary tree, prediction	IMU raw data	98.7%	Lower	Locomotion (3 modes)	Yes	10 Healthy
[11] 2020	Gradient tree, prediction	IMUs time domain & mechanical sensors	98.23%	Lower	Locomotion (5 modes)	No	8 Transfem. amputees
[12] 2021	Several methods, recognition	IMUs time domain & several sensors	99%	Lower	Locomotion (4 modes)	Yes	15 Healthy
Feature Learning Methods							
[13] 2016	DBN (5 hidden layers), recognition	Accelerometer time-freq. domain	98.23%	Lower	6 activities	No	29 Healthy
[13] 2016	DBN (5 hidden layers), recognition	Accelerometer time-freq. domain	91.5%	Lower	Locomotion (gait freeze classif.)	No	10 Impaired
[13] 2016	DBN (5 hidden layers), recognition	Accelerometer time-freq. domain	89.38%	Upper	10 activities	No	1 Healthy
[14] 2018	CNN (4 hidden layers), recognition	Accelerometer time domain	91.97%	Lower	6 activities	No	36 Healthy
[15] 2020	CNN (12 hidden layers), recognition	IMU raw data	87.74%	Lower	Locomotion (6 modes)	No	30 Healthy
[16] 2020	CNN (3 hidden layers), prediction	IMU time domain	95.58%	Lower	Locomotion (6 modes)	No	10 Healthy
[17] 2019	CNN (6 hidden layers), recognition	IMU raw data	94.15%	Lower	Locomotion (5 modes)	Yes (8 trans.)	10 Healthy
[17] 2019	CNN (6 hidden layers), recognition	IMU raw data	89.23%	Lower	Locomotion (5 modes)	Yes (8 trans.)	1 Transtibial amputee
[18] 2019	CNN (7 hidden layers), recognition	IMU raw data	n/a	Lower	16 motions	No	19 Healthy
[19] 2017	CNN (10 hidden layers), recognition	IMU time-freq. domain	97.06%	Upper, lower	Locomotion (gait phase classif.)	No	10 Healthy
[20] 2016	CNN (3 hidden layers), recognition	IMU time-freq. domain	97.01%	Lower	Locomotion (6 modes)	No	12 Healthy
[21] 2016	CNN (2 hidden layers), recognition	IMU time-freq. domain	95.1%	Upper, lower	7 activities	No	10 Healthy
[21] 2016	CNN (2 hidden layers), recognition	Accelerometer time-freq. domain	98.2%	Lower	6 activities	No	29 Healthy
[21] 2016	CNN (2 hidden layers), recognition	Accelerometer time-freq. domain	91.7%	Lower	2 activities	No	7 Healthy
[21] 2016	CNN (2 hidden layers), recognition	Accelerometer time-freq. domain	n/a	Upper	10 activities	No	1 Healthy
[22] 2018	RNN (1 hidden layer), recognition	IMU time domain	96.63%	Upper, lower	5 activities	No	11 Healthy
[23] 2017	RNN (3 hidden layers), recognition	IMU time domain	77%	Upper	3 hand motions	No	1 Healthy
[24] 2019	NN (1 hidden layer), recognition	IMU raw data	98.6%	Upper	10 hand motions	No	5 Healthy
[25] 2018	RNN (2 hidden layers), prediction	IMU raw data	96.3%	Lower	Locomotion (walking surface)	No	35 Healthy
[25] 2018	RNN (2 hidden layers), prediction	IMU raw data	94.7%	Lower	Locomotion (age)	No	35 Healthy
[26] 2019	NN (1 hidden layer), prediction	IMU time domain	n/a	Upper, lower	Locomotion (segmental contributions)	No	5 Healthy

subjects in [16], on raw-data from multiple IMUs on the lower limbs to recognize both the locomotion modes and the transitions of healthy subjects and transtibial amputees in [17], on raw-data from multiple IMUs on the lower-limbs and/or torso to recognize locomotion modes of healthy subjects in [18], on IMU features in the time-frequency domain

to recognize (loco)motion modes on healthy subjects in [19]–[21], and on features in the time-frequency domain of an accelerometer to recognize locomotion modes on healthy subjects in [21]. Recurrent neural networks (RNNs) have also been used to learn IMUs features. In [22], a RNN is used on time-domain features of two IMUs on the forearms

to recognize locomotion activities of healthy subjects. To recognize hand motions in healthy subjects, a RNN is used with time domain features from one in-hand IMU in [23], while a neural network (NN) is used with raw data from one in-hand IMU in [24]. In [25], RNNs are used with raw IMU data to predict the walking surface and age of healthy subjects. In [26], a NN is used on time domain IMUs features to predict the contributions of body parts during locomotion. [Table I](#) summarizes the main contributions to the state of the art of machine learning techniques (feature engineering and feature learning methods) for the recognition and/or prediction of (loco-)motion mode and transition intentions by means of IMU data. The table also reports the mean accuracies, whether the techniques were tested on upper or lower limbs, and on healthy or impaired subjects.

This study focuses on the real-time joint prediction of locomotor and transition intentions of one osseointegrated transfemoral amputee by means of deep neural networks on IMUs data. Nine different artificial neural networks, based on CNN, RNN and convolutional recurrent neural networks (CRNNs) have been designed and compared. The inputs to the architectures are features in both the time-domain and the time-frequency domain, which have been extracted from either one IMU (placed above the prosthetic knee), or two IMUs (placed above and below the prosthetic knee). Specifically, the chosen features are: (i) the means of the angular accelerations and angular velocities (i.e., the raw data obtained from the 3-axis accelerometers and the 3-axis gyroscopes of each IMU) computed within a time window; (ii) the corresponding quaternions in the same time window; (iii) the time-localized frequency information of each IMU, calculated using the short-time Fourier transform (STFT) within the same time window. The deep neural networks are designed to predict eight locomotor intentions (i.e., sitting, standing, level ground walking, stair ascent and descent, ramp ascent and descent, walking on uneven terrain) and the twenty-four transitions between these different locomotion modes. This study shows that a RNN, realized with four layers of gated recurrent unit networks, achieves (with a 5-fold cross-validation) a mean F1 score of 84.78% (standard deviation of 1.33) and 86.50% (standard deviation of 0.38) using one IMU, and 93.06% (standard deviation of 1.21) and 89.99% (standard deviation of 5.95) using two IMUs, with or without the sitting mode, respectively. To summarize, the contributions of this paper are:

- To design nine deep neural networks for the real-time prediction of eight locomotion modes and twenty-four transitions intentions.
- To use only IMU data (either from one or two IMUs, placed on the transfemoral prosthesis), from which features in the time domain and the time-frequency domain are learned by the deep neural networks.
- To validate the methods on data collected on one osseointegrated transfemoral amputee.
- To achieve a best F1 score of 93.06% (standard deviation of 1.21) with a RNN (realized with four layers of gated recurrent unit networks) for the real-time prediction.

The remainder of the paper is organized as follows. In Section II, the data-set used in this study is described together with the data processing. Section III presents the

different deep neural networks designed for the prediction of the locomotor and transition intentions. Section IV reports and discusses the results. Finally, concluding remarks are drawn in Section V.

II. MATERIALS

This Section presents the data-set and describes the data processing to obtain the inputs to the deep neural networks for the locomotor and transition intentions prediction.

A. Data-Set

The data-set used in this study has been collected at the Roessingh Research and Development center (Enschede, The Netherlands) on one osseointegrated transfemoral amputee subject (male, 75 years old, weight of 84.1 kg, height of 186.6 cm, left-sided amputation since 45 years, osseointegration since 4 years, functional level K3), wearing a 3R80 Ottobock prosthetic knee (www.ottobockus.com) and a Variflex Össur prosthetic ankle (www.ossur.com). The data were collected from the subject by using wearable electromyographic sensors and eight IMUs as part of the Xsens MVN Link motion capture system (Xsens Technologies B.V., The Netherlands, www.xsens.com).¹ From the data-set, this study only uses data from two IMUs, i.e., one IMU is placed above the prosthetic knee (i.e., on the thigh of the prosthetic leg, which is rigidly connected to the subject's stump thanks to the osseointegration), and a second IMU is placed below the prosthetic knee (i.e., on the shank of the prosthetic leg). The IMU data, collected over a measuring time of ~30 minutes of different activities, were sampled at 1000 Hz, which corresponds to a total of 1.801.775 data points. The data were further filtered to remove interruptions between the trials and/or the activities, which finally correspond to a total of 785.174 data points for a measuring time of ~13 minutes.

The subject was asked to execute eight locomotion modes (i.e., S: Sitting, ST: Standing, LW: Level ground Walking, SA: Stair Ascent, SD: Stair Descent, RA: Ramp Ascent, RD: Ramp Descent, TW: Walking on uneven Terrain) along a circuit, and the twenty-four transitions among these modes. The ramps have a slope of 10° for three meters, and continue on with a slope of 15°. The step-size of the stairs was not provided.

The data labelling has been done manually by analyzing the body segments positions and joint angles in the Xsens MVN motion capture software, and by extracting the gait events (i.e., heel contact, toe off) by using the peak foot accelerations [27]. The transitions were initially labelled as the future mode (e.g., the transition from S to ST was labelled as ST). To include the transitions in the data-set, a window of 500 ms was chosen between two subsequent modes (i.e., 250 ms in the previous mode and 250 ms in the next mode) and labelled with the corresponding transition label (e.g., the transition from S to ST was labelled as S - ST and lasts 500 ms).

B. Data Processing

1) *Features*: The features used in this study are extracted from the raw IMUs data (either one or two IMUs) and are

¹The study, under protocol number NL67247.044.18, was evaluated and approved by the Medical Ethics Review Committee of the University of Twente (The Netherlands) on December 13, 2018.

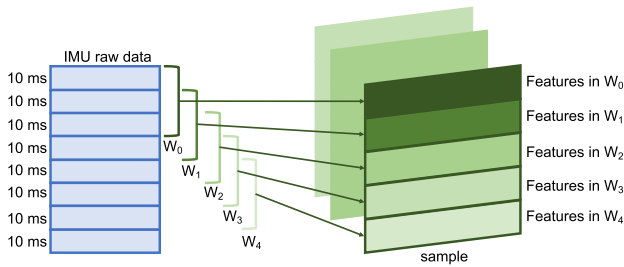


Fig. 1. Sequential samples are the inputs to the deep neural networks. Using time windows $W_i = 30$ ms and time-step of 10 ms, each samples is formed with the features computed in five subsequent time windows.

calculated within a time window W in both the time domain and the time-frequency domain. Specifically, the features are (i) the means of the angular accelerations and angular velocities (i.e., the raw data obtained from the 3-axis accelerometers and the 3-axis gyroscopes of each IMU) computed within the time window W ; (ii) the quaternions calculated on the mean IMU data in the same time window W , by using the filter proposed in [28], with the implementation in [29]; (iii) the time-localized frequency information of each IMU, calculated using the STFT within the same time window W . The time window W has been chosen to be 30 ms with a step length of 10 ms, to leave enough time to process the data, predict the locomotor or the transition intentions in real-time, and convert that into a control input [2].

2) *Inputs*: The inputs to the deep neural networks are sequential samples. Each sample consists of five sets of features, which have been computed within five time windows $W_i = 30$ ms with a step length of 10 ms. Figure 1 shows how each sample is formed from the raw IMU data. Specifically, one sample is built with features from five windows W_i , with $i = 0, \dots, 5$. Each windows W_i contains 30 ms of IMU raw data, with an overlap of 20 ms (i.e., W_1 contains data from 0 to 30 ms, W_2 from 10 to 40 ms, W_3 from 20 to 50 ms, W_4 from 30 to 60 ms, and W_5 from 30 to 70 ms). This implies that, at each run, the neural network receives a sample as input that contains features derived from 70 ms of raw IMU data.

Table II shows the eight locomotion modes, the number of data points in the original data-set (i.e., the IMU raw data in the time domain, sampled at 1000 Hz and filtered) and the number of samples in the processed data-set in the time and time-frequency domain. Table III shows the twenty-four transitions and the number of samples in the processed data-set in the time and time-frequency domain. The total samples (i.e., the sum of the samples of the locomotion modes and the transitions) form the inputs to the deep neural networks.

3) *Scaling*: The data have been standardized within each sample by centering to the mean and by scaling them component-wise to the unit variance.

4) *Data Partitioning*: Using 5-fold cross-validation, 80% of the data was used for training and 20% was used for testing. Within training, 10% of the data was used for validation.

C. Output

The output of the deep neural networks has a dimension equal to the sum of the locomotion modes and the transitions to be predicted. As sitting is a static mode that is

TABLE II

LOCOMOTION MODES. (LEFT) NUMBER OF DATA POINTS IN THE ORIGINAL DATA-SET, SAMPLED AT 1000 HZ AND FILTERED (TIME DOMAIN). (RIGHT) NUMBER OF SAMPLES OF THE LOCOMOTION MODES IN THE PROCESSED DATA-SET (TIME AND TIME-FREQUENCY DOMAIN)

Mode	n. data in the original data-set (time domain)	n. samples in the processed data-set (time and time-freq. domain)
S	187.347	18.133
ST	67.374	5.303
W	154.621	13.812
SA	85.123	7.912
SD	66.875	6.137
RA	54.966	5.047
RD	82.121	7.713
TW	86.747	7.974
Total:	785.174	72.031

TABLE III

TRANSITIONS. NUMBER OF SAMPLES OF THE TRANSITIONS IN THE PROCESSED DATA-SET (TIME AND TIME-FREQUENCY DOMAIN)

Transition	n. samples	Transition	n. samples
S - ST	350	S - W	100
ST - S	50	ST - W	350
ST - SA	50	ST - SD	200
ST - RA	200	ST - RD	150
ST - TW	150	W - S	200
W - ST	200	W - SA	550
W - SD	300	W - RA	200
W - RD	100	W - TW	400
SA - ST	50	SA - W	350
SD - ST	50	SD - W	350
RA - RD	300	RD - W	150
TW - ST	150	TW - W	350
Total:		5300	

also overrepresented, it could add unnecessary complexity by increasing the number of output dimensions, while the difficulty could lie in the distinction between the dynamic locomotion modes and the transitions. Therefore, in this study, two different experimental scenarios are analyzed, i.e., with and without sitting. Consequently, the output dimension is thirty-two when sitting is included (i.e., eight locomotion modes and twenty-four transitions), and twenty-seven when sitting is excluded (i.e., seven locomotion modes and twenty transitions).

III. METHODS

In this study, nine deep neural networks were designed and compared. The architectures are further described in the following subsections and are based on the CNNs, RNNs, and CRNNs, similarly to our previous work [16].

A. Convolutional Neural Networks

Three different CNN architectures (i.e. CNN1D, CNN2D, and WaveNet) have been designed.

1) *CNN1D and CNN2D*: Figure 2a shows both the CNN1D and the CNN2D, which consist of six hidden layers, i.e., four convolution layers and two dense layers. In this study, the convolutional kernel size is set to (1×2) and (2×2) for the CNN1D and CNN2D, respectively. The first four convolutional layers have a filter size of 32, 64, 128, and 256, respectively.

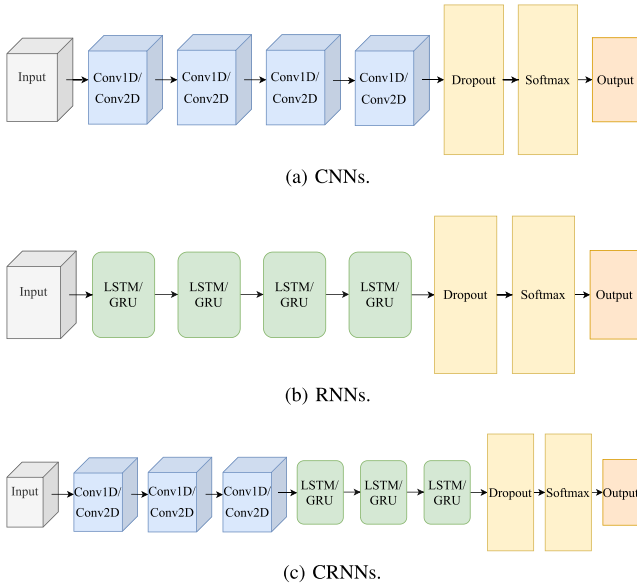


Fig. 2. Three deep neural network architectures. (a) CNNs with four convolutional layers (one- or two-dimensional) and two dense layers. (b) RNNs with four recurrent layers (LSTM or GRU) and two dense layers. (c) CRNNs with three convolutional layers (one- or two-dimensional), three recurrent layers (LSTM or GRU), and two dense layers.

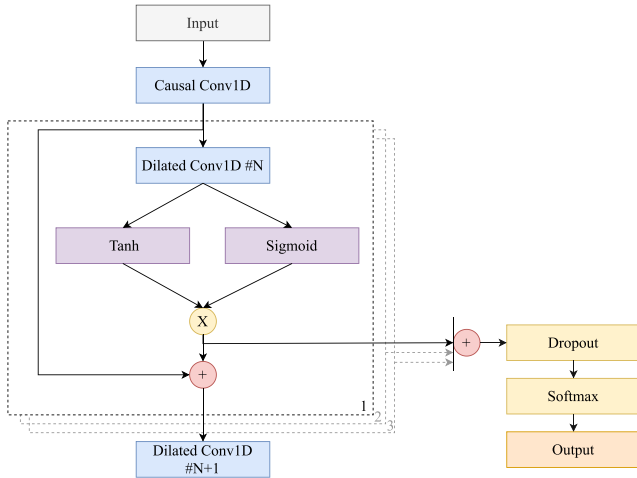


Fig. 3. WaveNet with four convolutional layers, from which three are dilated and one is causal, and two dense layers.

A rectified linear unit is used as activation function in each filter. Finally, two dense layers follow: first a dense layer with 200 units and a dropout of 0.25 and, then, an output layer that has the units equal to the dimension of the output (32 or 27) and a softmax activation function. The most significant difference between the CNN1D and CNN2D is the direction of the convolution kernels, i.e., CNN1D slides only sample-wise (from top to down) while CNN2D slides both sample-wise and column-wise.

2) *WaveNet*: **Figure 3** shows another CNN architecture, i.e., the WaveNet [30], which consists of four convolutional layers. The input is first processed by a causal convolutional layer, consisting of 256 filters with a filter size of 2. Next, the output goes into two ways. In one direction, it is served as an input into a dilated convolutional layer (256 filters with a filter size of 2), which consists of two convolutions

with either a tanh or sigmoid activation function that are combined by using a dot multiplication to the output layer. In the other direction, it skips the dilated convolution and is directly summed up with the output of the dilated convolution, which then serves as an input for the second layer. The output layer consists of a dense layer (200 units, 0.25 dropout) and a dense layer with a softmax activation function, where the number of units is equal to the dimensions of the output.

B. Recurrent Neural Networks

Two different RNN architectures have been designed in this study, as shown in **Figure 2b**. Both RNNs consist of six hidden layers, i.e., four recurrent layers and two dense layers. The recurrent layers are either long short-term memory (LSTM) networks [31] or gated recurrent units (GRU) networks [32]. The first four layers consist of 128 LSTM or GRU units. Then two dense layers follow, one that has 200 units and a dropout of 0.25, and one that serves as an output layer with units equal to the dimension of the output (32 or 27) and a softmax activation function.

C. Convolutional Recurrent Neural Networks

Four different CRNN architectures have been designed in this study, as shown in **Figure 2c**. They consist of eight hidden layers, i.e., three convolutional layers (either one- or two-dimensional), three recurrent layers (either LSTM or GRU), and two dense layers. The first three convolutional layers are similar to that of the CNNs in **Figure 2a** with the only difference that they have filters of size 64, 128, and 256, respectively. The last three RNN layers are equivalent to the layers in the RNNs in **Figure 2b**, as they also have 128 units. The final two layers are both dense layers. One of them has 200 units and 0.25 dropout, and the other has a number of units equal to the dimensions of the output (32 or 27) and a softmax activation function. The only difference between the one- and two-dimensional version of the CRNN is that the output of the CNN2D layers needs to be wrapped together with the time-step to serve as a compatible input to the RNN layers.

D. Evaluation: Performance Metric

Due to the uneven distribution of the data, the deep neural networks are compared based on the F1 scores, a metric that compares both precision and recall and is calculated as:

$$F1 = 2 \cdot \frac{\text{precision} \cdot \text{recall}}{\text{precision} + \text{recall}}$$

where $\text{precision} = tp/(tp + fp)$ and $\text{recall} = tp/(tp + fn)$, with tp being the number of true positive predictions, fp the number of false positives and fn the number of false negatives. The comparisons of the networks are based on this metric.

K-fold cross-validation is used to compare the general effectiveness of the neural networks. In this study, k is set to 5, which means that the data is divided into five subsets. Next, the data is validated on one subset and trained on the remaining four. The validation is done on every subset and,

hence, happens five times in total. Finally, the mean F1 score of all validations is taken to compare the performances. This way, the data gets fully utilized, underfitting is prevented, and the reliability of the evaluation increases as the training and testing data are set differently each time.

E. Hyperparameters

This Section describes the hyperparameters that were used for training. Specifically, the hyperparameter search started from the results obtained in our previous study on the IMU-based locomotor intention prediction of ten healthy subjects [16], the hyperparameter optimization is done empirically, and the hyperparameter final choice is based on the F1 scores achieved by the different neural networks and such that the hyperparameters are the same across the different networks. The training was done on one computer with an NVIDIA GeForce GTX 1060, a quad-core Intel i7 – 6700 processor, and 8 GB RAM.

1) *Learning Rate*: The learning rate is set at 0.001. This value is set as such because using a high learning rate causes the network to never converge, while a lower learning rate would increase the risk of falling into a local minimum.

2) *Optimizer*: The optimizer is chosen to speed up the convergence of the neural networks, by optimizing the gradient descent. The Adaptive Moment Estimation (Adam) has been used in this study [33]. Adam computes individual adaptive learning rates for different parameters.

3) *Batch Size*: The batch size represents how many input data are shown simultaneously to the network before updating its weights. The batch size is chosen to be 512, which is relatively high, but in combination with the number of filters, it has been observed to obtain the highest F1 score. The high batch size does not result in processing more previous data before predicting a locomotion mode or a transition, as the data is shuffled, but may increase the accuracy of the error estimate in training.

4) *Loss Function*: As a loss function, the categorical cross-entropy has been used.

5) *Class Weighting*: The loss function assumes there is an equal distribution among the different classes (i.e., locomotion modes and transitions). However, in this study, the transitions are underrepresented and, to account for this unbalance a weight is added to each class, by exploiting the `sklearn.utils.class_weight` module of the `scikit-learn` Python library [34]. This weight makes the transitions more important for the network and penalizes mistakes made for the transitions more opposed to the locomotion modes, without influencing the number of samples in the data-set.

6) *Shuffling*: The training of the networks is done by feeding the network with the input batch by batch. If the network is fed the data in chronological order, the network would overfit between multiple classes. To avoid this, the sequential samples are shuffled, i.e., the order within each sample remains fixed, but the order of the samples does not.

7) *Epochs*: The data is presented 150 times to the networks during training to optimize data use and avoid under-fitting.

8) *Early Stopping*: The number of epochs is set to a high number to ensure that the data is used enough and is not

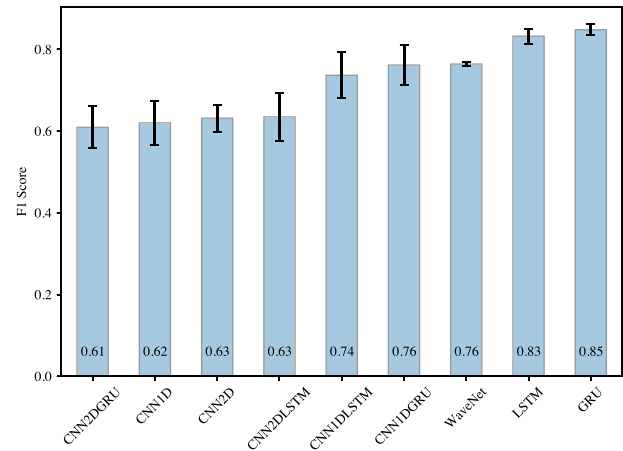


Fig. 4. F1 score (mean and SD), with a 5-fold cross validation, for all the deep neural networks (including the sitting mode). Only features from the IMU above the prosthetic knee are used.

under-fitting. If the number of epochs is too high, the network will start to overfit on the data. Therefore, an early stopping approach is used. If the accuracy on the validation set has not increased for 15 epochs, the network will stop training and will be the final model. This number is empirically set to avoid an increase in validation loss, which is a sign of overfitting.

IV. RESULTS AND DISCUSSION

In this Section, the proposed deep neural networks are compared using the F1 score performance metric. The results are reported separately based upon the features extracted from either one IMU or two IMUs. Additionally, within each subsection, the results for including or excluding the sitting mode from the prediction are reported separately. Finally, the results are discussed and compared to the literature.

A. One IMU (Above the Prosthetic Knee)

Figure 4 shows the F1 scores (mean and standard deviation SD), with a 5-fold cross-validation, of all the deep neural networks when only features from the IMU above the prosthetic knee are used for the prediction of the locomotion modes and transitions, including the sitting mode. It can be observed that the GRU outperforms the other networks with a mean F1 score of 84.78% (SD = 1.33). To verify whether the experiments results are significant, a repeated measures analysis of variance (ANOVA) test has been performed on the three best performing models, together with a Tukey's honestly significant difference (HSD) test. The ANOVA test shows a $F(2, 8) = 93.57\%$, $p = 0.000$. Post-hoc comparisons using the Tukey's HSD test indicates that the WaveNet (mean = 77.56%, SD = 0.64) is significantly different from both the GRU (mean = 84.78%, SD = 1.33) and the LSTM (mean = 83.21%, SD = 1.86), but the LSTM does not differ significantly from the GRU.

Figure 5 shows the F1 scores (mean and SD), with a 5-fold cross-validation, of all the deep neural network when only features from the IMU above the prosthetic knee are used, but the sitting mode and the corresponding transitions are excluded. It can be observed that the GRU outperforms the

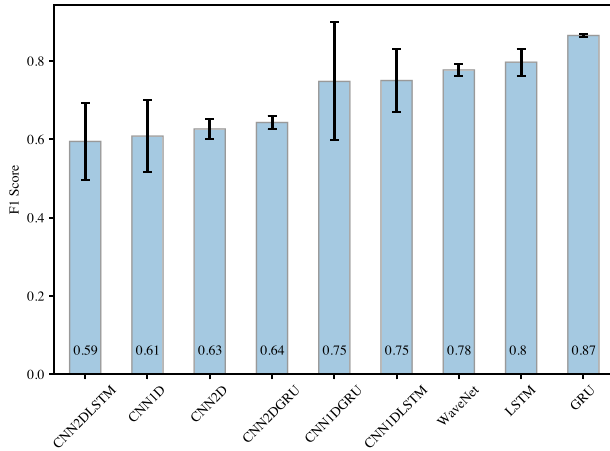


Fig. 5. F1 score (mean and SD), with a 5-fold cross validation, for all the deep neural networks. Only features from the IMU above the prosthetic knee are used. The sitting mode and corresponding transitions are excluded from the experiment.

other networks with a mean of 86.50% (SD = 0.38). The repeated measures ANOVA test shows a $F(2, 8) = 21.12$, $p = 0.0006$. Post-hoc comparisons using the Tukey's test indicates that there is a significant difference between the GRU (mean = 86.50%, SD = 0.38) and the WaveNet (mean = 77.70%, SD = 1.53) and between the GRU and the LSTM (mean = 79.68%, SD = 3.48), but the LSTM and WaveNet do not differ significantly.

B. Two IMUs (Above and Below the Prosthetic Knee)

Figure 6 shows the F1 scores, with a 5-fold cross-validation, of all the deep neural networks when features from the IMUs above and below the prosthetic knee are used for the prediction of the locomotion modes and transitions, including the sitting mode. It can be observed that the GRU outperforms the other networks with a mean of 93.06% (SD = 1.21). The repeated measures ANOVA test shows a $F(2, 8) = 4.86$, $p = 0.042$. Post-hoc comparisons using the Tukey's HSD test indicates that the GRU (mean = 93.06%, SD = 1.21) is significantly different from the WaveNet (mean = 89.30%, SD = 0.95), but the LSTM (mean = 90.11%, SD = 3.08) does not significantly differ from the GRU and WaveNet.

Figure 7 shows the F1 scores, with a 5-fold cross-validation, of all the deep neural networks when features from the IMUs above and below the prosthetic knee are used, but the sitting mode and the corresponding transitions are excluded. It can be observed that the WaveNet, GRU and LSTM have seemingly similar performances. The repeated measures ANOVA test shows a $F(2, 8) = 1.47$, $p = 0.30$. In this case, no Tukey's HSD test is needed as none of these three results are significantly different.

C. Running Time

Table IV shows the running time (in ms) (averaged on 70.000 random data points, i.e., 1000 sequential samples that contain features extracted from 70 ms of raw data) for the prediction of the three outperforming deep neural networks (i.e., WaveNet, GRU, and LSTM). It can be noted that the

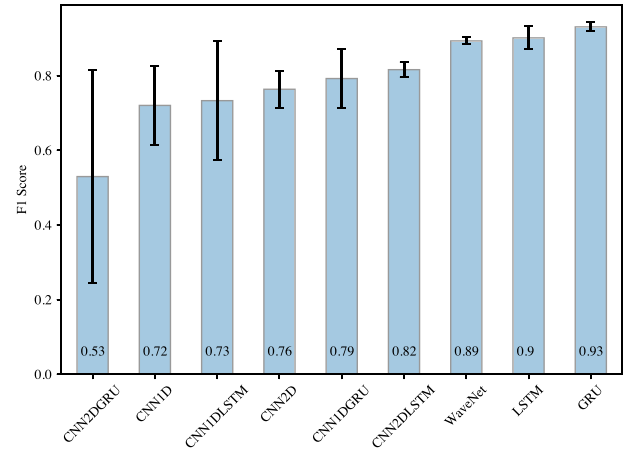


Fig. 6. F1 score (mean and SD), with a 5-fold cross validation, for all the deep neural networks (including the sitting mode). Features from the IMUs above and below the prosthetic knee are used.

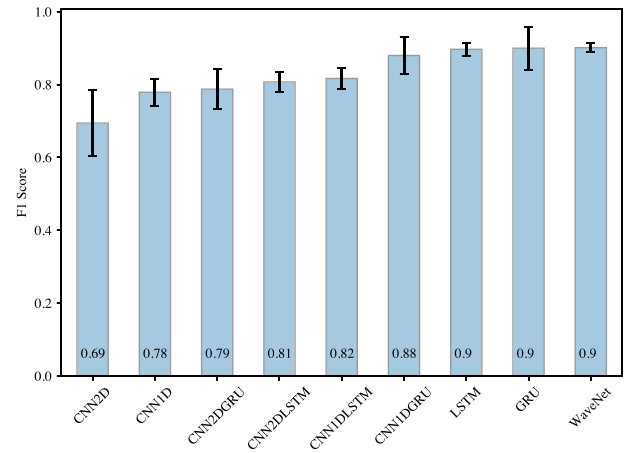


Fig. 7. F1 score (mean and SD), with a 5-fold cross validation, for all the deep neural network architectures. Features from the IMUs above and below the prosthetic knee are used. The sitting mode and the corresponding transitions are excluded from the experiment.

TABLE IV

RUNNING TIME (MEAN AND SD) IN ms FOR THE PREDICTION OF THE THREE OUTPERFORMING DEEP NEURAL NETWORKS, AVERAGED ON 1000 SEQUENTIAL SAMPLES INCLUDING THE SITTING MODE AND THE CORRESPONDING TRANSITIONS

	One IMU (above knee)	Two IMUs (above/below knee)
WaveNet	17.56 ± 5.33 [ms]	16.87 ± 5.00 [ms]
LSTM	17.29 ± 6.02 [ms]	17.33 ± 6.58 [ms]
GRU	17.79 ± 13.59 [ms]	18.17 ± 13.72 [ms]

processing time does not increase when two IMUs are used, while the performance does increase significantly. Moreover, the table shows that the SD of the running time of the GRU is higher than the other two networks.

D. Discussion

Table V summarizes the results of the three best performing deep neural networks in the four experimental scenarios of this study, i.e., one/two IMUs and with/without the sitting mode. It can be noted that the GRU outperforms the other

TABLE V

SUMMARY OF THE BEST PERFORMING DEEP NEURAL NETWORKS IN DIFFERENT EXPERIMENTAL SCENARIOS: ONE IMU (ABOVE THE PROSTHETIC KNEE) OR TWO IMUS (ABOVE AND BELOW THE PROSTHETIC KNEE) AND WITH OR WITHOUT THE SITTING MODE. THE HIGHEST F1 SCORES FOR EACH EXPERIMENTAL SCENARIO ARE IN BOLD

Locomotor and transition intentions prediction for one osseointegrated transfemoral amputee	
Features	106 (for one IMU), 212 (for two IMUs) For each IMU: - 3 acceler. and 3 gyros. (mean values) - 4 quaternions (from the mean) - 96 frequencies (from STFT)
n. of original data points (sampled at 1000Hz and filtered)	785.174 (~13 minutes)
n. of samples (locomotion + transitions)	77.331 (72.031 + 5300)
n. of modes & transitions	32 (with sitting) 27 (without sitting)
Mean F1 score (one IMU) from 5-fold cross-validation	With sitting: WaveNet: 76.56% SD = 0.64 LSTM: 83.21% SD = 1.86 GRU: 84.78% SD = 1.33
	Without sitting: WaveNet: 77.70% SD = 1.53 LSTM: 79.68% SD = 3.48 GRU: 86.50% SD = 0.38
Mean F1 score (two IMUs) from 5-fold cross-validation	With sitting: WaveNet: 89.30% SD = 0.95 LSTM: 90.11% SD = 3.08 GRU: 93.06% SD = 1.21
	Without sitting: WaveNet: 90.14% SD = 1.15 LSTM: 89.68% SD = 1.82 GRU: 89.99% SD = 5.95

networks except when sitting is removed from the data-set and two IMUs are used. However, the running time of the GRU is higher and fluctuates more between different samples than the other two networks, but remains far below 300 ms [2]. Specifically, this study shows that the best performing deep neural network is a RNN, realized with four layers of gated recurrent unit networks, which achieves a mean F1 score of 93.06% (SD of 1.21) by using features from two IMUs for the real-time prediction of eight locomotion modes (including sitting) and twenty-four transitions intentions. As expected, the GRU outperforms the other deep neural networks because, thanks to its memory units, it appears to be better suited for forecasting time-series in the time-frequency domain.

Figure 8 shows the confusion matrix for this best performing experimental scenario. It can be noted that the GRU is able to jointly predict the intentions of both the locomotion modes and the transitions of the osseointegrated transfemoral amputee. The most challenging intentions' predictions (i.e., for which the mean F1 score is below 94%) are: the locomotion mode TW (walking on uneven terrain) and the transitions W - ST (walking to standing), W - SA (walking to stair ascent), W - TW (walking to walking on uneven terrain), TW - ST (walking on uneven terrain to standing), TW - W (walking on uneven terrain to walking).

E. Comparison to the State of the Art

In comparison to the state of the art (see Table I), this study contributes on the aspects detailed hereafter.

1) *Transfemoral Amputees*: This study focuses on feature learning methods (deep neural networks) on data from one or two IMUs placed on the prosthetic leg of an osseointegrated transfemoral amputee. Previous research has build on feature engineering methods for the prediction of locomotion modes of six transtibial amputees [6] (LDAs on IMUs), six transfemoral amputees [8] (LDAs on IMUs and mechanical sensors), eight transfemoral amputees [11] (gradient tree on IMUs and several mechanical sensors), or on feature learning methods on ten impaired subjects (DBN) for the recognition of the freezing of gait [13], and on one transtibial amputees (CNN) [17] for the recognition of locomotion modes and transitions.

2) *Locomotor and Transition Intention Joint Prediction*: This study focuses on the joint intention prediction of eight locomotion modes and twenty-four transitions of one osseointegrated transfemoral amputee. Previous research has been devoted to the disjoint prediction of five locomotion modes and nine transitions of healthy subjects with (several) feature engineering methods [5], to the prediction of six locomotion modes and twelve transitions of healthy subjects with a feature engineering method (LDA) on IMU data combined with pressure insoles [7], to the disjoint prediction of three locomotion modes and four transitions of healthy subjects with a features engineering method on one IMU [10], to the disjoint prediction of four locomotion modes and four transitions of healthy subjects with different features engineering methods on several IMUs and other sensors [12], and to the recognition of five locomotion modes and eight transitions of healthy subjects and transtibial amputees with feature learning methods (CNNs) [17].

3) *Accuracy*: This study shows that a RNN, realized with four layers of gated recurrent unit networks, achieves (with a 5-fold cross-validation) a mean F1 score of 93.06% (standard deviation of 1.21) using time and time-frequency domain features engineered from two IMUs, in the best performing experimental scenarios (eight locomotion modes and twenty-four transitions). Previous research on the prediction of both locomotor and transition intentions has reported 97.65% on healthy subjects (five locomotion modes and nine transitions) with engineered features in the time-domain from seven IMUs [5], 99.71% on healthy subjects (six locomotion modes and twelve transitions) with engineered features in the time-domain from IMUs and pressure insoles [7], 98.7% on healthy subjects (three locomotion and four transitions) with engineered features from raw-data of one IMU [10], 99% on healthy subjects (four locomotion and four transitions) with engineered features from raw-data of several IMUs and several other sensors [12], 94.15% on healthy subjects (five locomotion modes and eight transitions) with features learned from raw-data of three IMUs [17], and 89.23% on transtibial amputees (five locomotion modes and eight transitions) with features learned from raw-data of three IMUs [17].

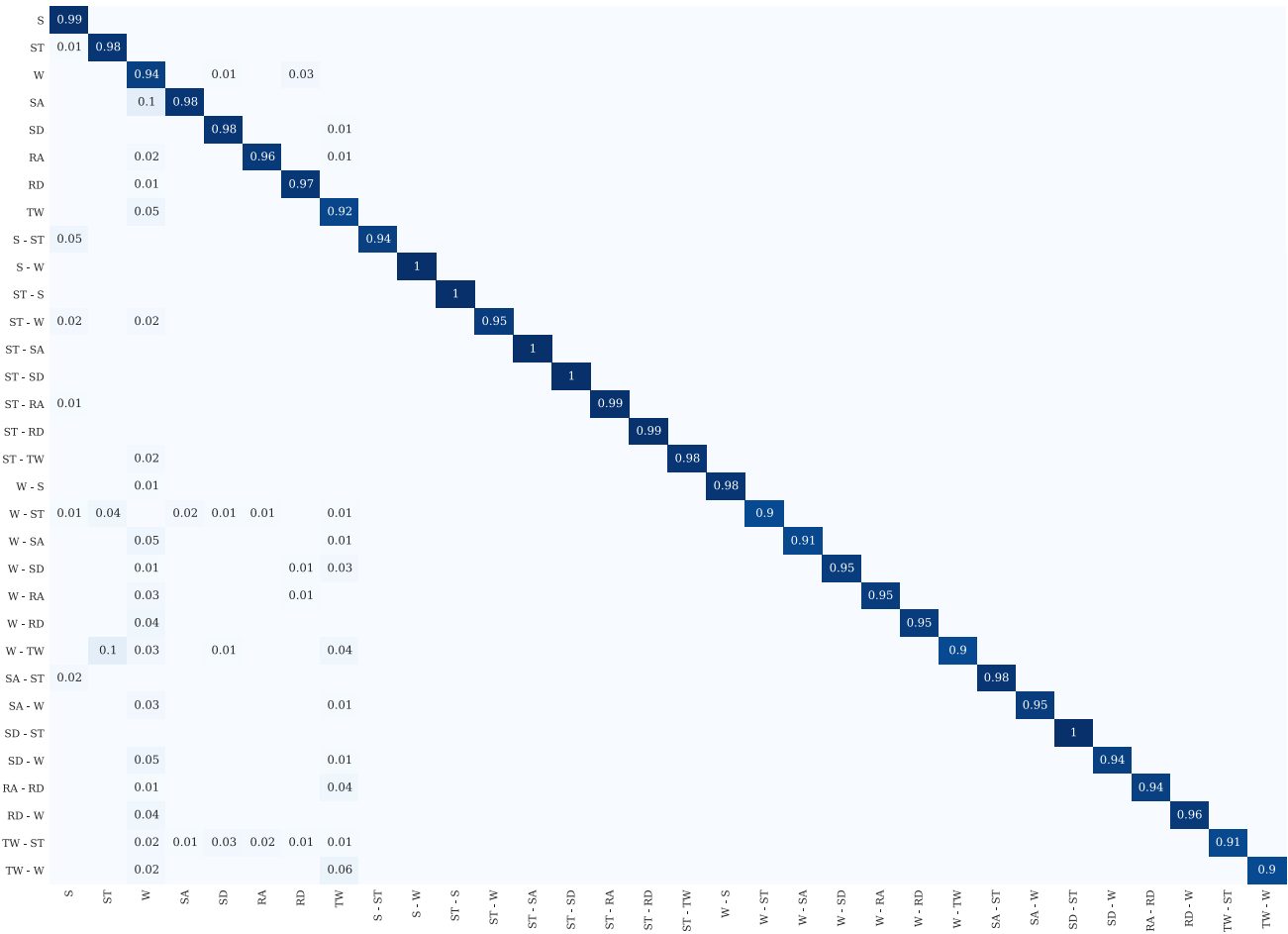


Fig. 8. Confusion matrix of the GRU neural network for the best experimental scenario (two IMUs above and below the prosthetic knee of the osseointegrated transfemoral amputee, including sitting).

F. Limitations and Future Outlook

1) *Proposed Methods*: In this study, nine different deep neural networks have been successfully trained, validated, and tested on the data collected on one transfemoral amputee. The potential of feature learning methods of being subject-independent has been shown in our previous work on the locomotor intention prediction of ten healthy subjects [16]. However, the generalization of these methods, and specifically of the GRU as the best performing one, to other transfemoral amputees (even with different K-levels) is left as future work.

2) *Real-Time Implementation*: To achieve an intuitive control, the intentions of the user should be predicted and converted to a proper control input to the prosthesis within 300 ms [2]. In this study, for each prediction of either locomotor or transition intentions, a neural network needs to acquire 10 ms of new raw IMU data (at 1000 Hz), to derive features from windows of 30 ms for a total of 70 ms of data (see Figure 1), and to process them for the intention prediction. Specifically, the best performing neural network (i.e., the GRU) needs 10 ms to acquire 10 ms of IMU raw data, 4.98 ± 0.038 ms (averaged on 70.000 data points) to preprocess 70 ms of data, and 17.79 ± 13.59 ms (averaged on 70.000 data points) for the intention prediction (see Table IV). This overall computation time is calculated on a desktop computer, whose computational power is comparable with

processors that can be placed on prosthetic leg prototypes, and it is far below the 300 ms requested for real-time control.

3) *Clinical Requirements*: Future research should focus on the implementation and evaluation of the proposed method on osseointegrated amputees in clinical trials, which could start from the amount of data already collected for this study since the F1 scores are always above 90% (see Figure 8). The F1 scores might be further improved by training the GRU neural network on more data. However, it is not surprising that the most challenging intentions predictions mainly concern transitions from/to walking on uneven terrain (which require a high level of adaptability of the prosthesis at both the knee and ankle joints), and the transitions from walking to stair ascent (which in general is performed by the amputee by stopping before starting ascending the stairs).

V. CONCLUSION

This study presented the design and comparison of nine deep neural networks for the prediction of eight locomotor (i.e., sitting, standing, level ground walking, stair ascent/descent, ramp ascent/descent, walking on uneven terrain) and twenty-four transition intentions of one osseointegrated transfemoral amputee. Inputs to these networks are derived from features in the time and the time-frequency domain, which are extracted from either one IMU

(above the prosthetic knee) or two IMUs (above and below the prosthetic knee). The features are (i) the means of the angular accelerations and angular velocities (i.e., the raw data obtained from the 3-axis accelerometers and the 3-axis gyroscopes of each IMU) computed within a time window; (ii) the quaternions calculated on the mean IMU data in the same time window; (iii) the time-localized frequency information of each IMU, calculated using the STFT within the same time window.

This study shows that a RNN, realized with four layers of gated recurrent unit networks, achieves (with a 5-fold cross-validation) a mean F1 score of 84.78% (SD of 1.33) and 86.50% (SD of 0.38) using one IMU, and 93.06% (SD of 1.21) and 89.99% (SD of 5.95) using two IMUs, with or without the sitting mode, respectively.

ACKNOWLEDGMENT

The authors would like to thank Prof. Dr. Hermie Hermens (Roessingh Research and Development, Enschede, The Netherlands) for providing the data of the osseointegrated amputee, which have been collected within the scope of the European Commission's Horizon 2020 MyLeg Project.

REFERENCES

- [1] H. Pernot, L. de Witte, E. Lindeman, and J. Cluitmans, "Daily functioning of the lower extremity amputee: An overview of the literature," *Clin. Rehabil.*, vol. 11, no. 2, pp. 93–106, May 1997.
- [2] B. Hudgins, P. Parker, and R. N. Scott, "A new strategy for multifunction myoelectric control," *IEEE Trans. Biomed. Eng.*, vol. 40, no. 1, pp. 82–94, Jan. 1993.
- [3] K. Zhang, C. W. de Silva, and C. Fu, "Sensor fusion for predictive control of Human-Prosthesis-Environment dynamics in assistive walking: A survey," 2019, *arXiv:1903.07674*. [Online]. Available: <http://arxiv.org/abs/1903.07674>
- [4] J. Wang, Y. Chen, S. Hao, X. Peng, and L. Hu, "Deep learning for sensor-based activity recognition: A survey," *Pattern Recognit. Lett.*, vol. 119, pp. 3–11, Mar. 2019.
- [5] J. Figueiredo, S. P. Carvalho, D. Goncalve, J. C. Moreno, and C. P. Santos, "Daily locomotion recognition and prediction: A kinematic data-based machine learning approach," *IEEE Access*, vol. 8, pp. 33250–33262, 2020.
- [6] R. Stolyarov, G. Burnett, and H. Herr, "Translational motion tracking of leg joints for enhanced prediction of walking tasks," *IEEE Trans. Biomed. Eng.*, vol. 65, no. 4, pp. 763–769, Apr. 2018.
- [7] B. Chen, E. Zheng, and Q. Wang, "A locomotion intent prediction system based on multi-sensor fusion," *Sensors*, vol. 14, no. 7, pp. 12349–12369, Jul. 2014.
- [8] A. J. Young, A. M. Simon, and L. J. Hargrove, "A training method for locomotion mode prediction using powered lower limb prostheses," *IEEE Trans. Neural Syst. Rehabil. Eng.*, vol. 22, no. 3, pp. 671–677, May 2014.
- [9] H. Li, S. Derronde, and W. Pieczynski, "An adaptive and on-line IMU-based locomotion activity classification method using a triplet Markov model," *Neurocomputing*, vol. 362, pp. 94–105, Oct. 2019.
- [10] V. Papapicco *et al.*, "A classification approach based on directed acyclic graph to predict locomotion activities with one inertial sensor on the thigh," *IEEE Trans. Med. Robot. Bionics*, vol. 3, no. 2, pp. 436–445, May 2021.
- [11] K. Bhakta, J. Camargo, L. Donovan, K. Herrin, and A. Young, "Machine learning model comparisons of user independent & dependent intent recognition systems for powered prostheses," *IEEE Robot. Autom. Lett.*, vol. 5, no. 4, pp. 5393–5400, Oct. 2020.
- [12] J. Camargo, W. Flanagan, N. Csomay-Shanklin, B. Kanwar, and A. Young, "A machine learning strategy for locomotion classification and parameter estimation using fusion of wearable sensors," *IEEE Trans. Biomed. Eng.*, vol. 68, no. 5, pp. 1569–1578, May 2021.
- [13] M. A. Alsheikh, A. Selim, D. Niyato, L. Doyle, S. Lin, and H.-P. Tan, "Deep activity recognition models with triaxial accelerometers," in *Proc. AAAI Conf. Artif. Intell.*, 2016, pp. 1–7.
- [14] W. Xu, Y. Pang, Y. Yang, and Y. Liu, "Human activity recognition based on convolutional neural network," in *Proc. 24th Int. Conf. Pattern Recognit. (ICPR)*, Aug. 2018, pp. 165–170.
- [15] W.-H. Chen *et al.*, "Determining motions with an IMU during level walking and slope and stair walking," *J. Sports Sci.*, vol. 38, no. 1, pp. 62–69, Jan. 2020.
- [16] H. Lu, L. R. B. Schomaker, and R. Carloni, "IMU-based deep neural networks for locomotor intention prediction," in *Proc. IEEE/RSJ Int. Conf. Intell. Robots Syst. (IROS)*, Oct. 2020, pp. 4134–4139.
- [17] B.-Y. Su, J. Wang, S.-Q. Liu, M. Sheng, J. Jiang, and K. Xiang, "A CNN-based method for intent recognition using inertial measurement units and intelligent lower limb prosthesis," *IEEE Trans. Neural Syst. Rehabil. Eng.*, vol. 27, no. 5, pp. 1032–1042, May 2019.
- [18] A. Bevilacqua, K. MacDonald, A. Rangarej, V. Widjaya, B. Caulfield, and T. Kechadi, "Human activity recognition with convolutional neural networks," in *Machine Learning and Knowledge Discovery in Databases (Lecture Notes in Computer Science)*, vol. 11053. Cham, Switzerland: Springer, 2019, pp. 541–552.
- [19] O. Dehzangi, M. Taherisadr, and R. ChagalVala, "IMU-based gait recognition using convolutional neural networks and multi-sensor fusion," *Sensors*, vol. 17, no. 12, p. 2735, Nov. 2017.
- [20] T. Zebin, P. J. Scully, and K. B. Ozanyan, "Human activity recognition with inertial sensors using a deep learning approach," in *Proc. IEEE Sensors*, Oct. 2016, pp. 1–3.
- [21] D. Ravi, C. Wong, B. Lo, and G.-Z. Yang, "Deep learning for human activity recognition: A resource efficient implementation on low-power devices," in *Proc. IEEE 13th Int. Conf. Wearable Implant. Body Sensor Netw. (BSN)*, Jun. 2016, pp. 71–76.
- [22] R. Rego Drumond, B. A. Dorta Marques, C. Nader Vasconcelos, and E. Clua, "PEEK—An LSTM recurrent network for motion classification from sparse data," in *Proc. 13th Int. Joint Conf. Comput. Vis., Imag. Comput. Graph. Theory Appl.*, 2018, pp. 215–222.
- [23] A. Fu and Y. Yu, "Real-time gesture pattern classification with IMU data," Stanford Univ., Stanford, CA, USA, Tech. Rep., 2017. [Online]. Available: http://stanford.edu/class/ee267/Spring2017/report_fu_yu.pdf
- [24] M. Kim, J. Cho, S. Lee, and Y. Jung, "IMU sensor-based hand gesture recognition for human-machine interfaces," *Sensors*, vol. 19, no. 18, p. 3827, Sep. 2019.
- [25] B. Hu, P. C. Dixon, J. V. Jacobs, J. T. Dennerlein, and J. M. Schiffman, "Machine learning algorithms based on signals from a single wearable inertial sensor can detect surface- and age-related differences in walking," *J. Biomech.*, vol. 71, pp. 37–42, Apr. 2018.
- [26] N. T. Pickle, S. M. Shearin, and N. P. Fey, "Dynamic neural network approach to targeted balance assessment of individuals with and without neurological disease during non-steady-state locomotion," *J. NeuroEng. Rehabil.*, vol. 16, no. 1, pp. 1–9, Dec. 2019.
- [27] B. Hu, E. Rouse, and L. Hargrove, "Benchmark datasets for bilateral lower-limb neuromechanical signals from wearable sensors during unassisted locomotion in able-bodied individuals," *Frontiers Robot. AI*, vol. 5, p. 14, Feb. 2018.
- [28] R. Mahony, T. Hamel, and J.-M. Pfimlin, "Nonlinear complementary filters on the special orthogonal group," *IEEE Trans. Autom. Control*, vol. 53, no. 5, pp. 1203–1218, Jun. 2008.
- [29] S. O. H. Madgwick, A. J. L. Harrison, and R. Vaidyanathan, "Estimation of IMU and MARG orientation using a gradient descent algorithm," in *Proc. IEEE Int. Conf. Rehabil. Robot.*, Jun. 2011, pp. 1–7.
- [30] A. van den Oord *et al.*, "WaveNet: A generative model for raw audio," 2016, *arXiv:1609.03499*. [Online]. Available: <http://arxiv.org/abs/1609.03499>
- [31] K. Cho *et al.*, "Learning phrase representations using RNN encoder-decoder for statistical machine translation," 2014, *arXiv:1406.1078*. [Online]. Available: <http://arxiv.org/abs/1406.1078>
- [32] P. Xia, J. Hu, and Y. Peng, "EMG-based estimation of limb movement using deep learning with recurrent convolutional neural networks," *Artif. Organs*, vol. 42, no. 5, pp. E67–E77, May 2018.
- [33] D. P. Kingma and J. Ba, "Adam: A method for stochastic optimization," 2014, *arXiv:1412.6980*. [Online]. Available: <http://arxiv.org/abs/1412.6980>
- [34] F. Pedregosa *et al.*, "Scikit-learn: Machine learning in Python," *J. Mach. Learn. Res.*, vol. 12, pp. 2825–2830, Oct. 2011.