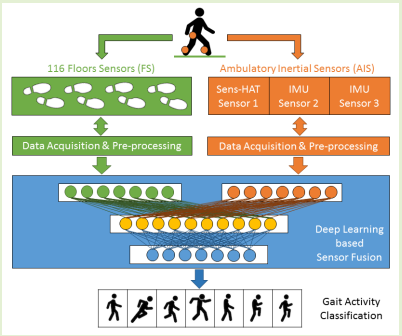
Gait Activity Classification Using Multi-Modality Sensor Fusion: A Deep Learning Approach

Syed U. Yunas and Krikor B. Ozanyan, *Senior Member, IEEE*

***Abstract*—Floor Sensors (FS) are used to capture information from the force induced on the contact surface by feet during gait. On the other hand, the Ambulatory Inertial Sensors (AIS) are used to capture the velocity, acceleration and orientation of the body during different activities. In this paper, fusion of the stated modalities is performed to overcome the challenge of gait classification from wearable sensors on the lower portion of human body not in contact with ground as in FS. Deep learning models are utilized for the automatic feature extraction of the ground reaction force obtained from a set of 116 FS and body movements from AIS attached at 3 different locations of lower body, which is novel. Spatio-temporal information of disproportionate inputs obtained from the two modalities is balanced and fused within deep learning network layers whilst reserving the categorical content for each gait activity. Our approach of fusion compensates the degradation in spatio-temporal accuracies in individual modalities and makes the overall classification outcomes more accurate. Further assessment of multi-modality based results show significant improvements in f-scores using different deep learning models i.e., LSTM (99.90%), 2D-CNN (88.73%), 1D-CNN (94.97%) and ANN (89.33%) respectively.**



***Index Terms*—Floor sensors, ambulatory inertial sensors, inertial measurement unit, deep learning, artificial neural networks, convolutional neural networks, long short-term memory.**

I. Introduction

GAIT defines unique walking pattern in humans which gets influenced by mutually independent factors such as height, weight, gender and age etc. Gait patterns get affected by many factors such as illness [1], fatigue [2], emotions [3], cognitive and motor tasks [4]. In addition, gait is also prone to influence from external factors such as clothing, wearing shoes or carrying load [5].

Gait analysis is on the way to maturity with applications in many research areas. In the medical field, the study of human gait is used to monitor and examine certain neurological diseases such as Alzheimer's and Parkinson's Disease (PD) [6]. Moreover, gait analysis has applications to assess the ability of sportsmen after injuries occurred during sport activities [7].

Ambulatory inertial sensors (AIS) are an inexpensive, convenient and efficient means to acquire particulars of human gait. Inertial measurement unit (IMU) is a type of ambulatory sensor that has been widely used to acquire gait information due to its small size, weight and cost [8]. A basic IMU comprises an accelerometer, gyroscope and sometimes a magnetometer, which allows a comprehensive report about the orientation, velocity and acceleration of the human body. An important factor to consider is that although ambulatory sensors are non-invasive, they require the individual to cooperate in wearing them on different body parts such as head, waist, chest, thigh, shank and foot to record gait signals [9]. The benefits of gait assessment and monitoring in patients can also be realized with smart phone based IMUs [10]. The smart phone now days is capable of performing all necessary tasks such as making decisions and contacting the health providers in case of emergency situations.

While walking, the interaction of human body with the environment is defined by the point of contact with the walking surface, which cannot be modified at will. Floor sensors (FS) are normally used to describe such interaction. FS can be unobtrusive and mainly based on resistive plates, capacitive plates, piezoelectric sensors or fiber optic cables [11]. These systems are typically installed indoors, in controlled environments such as offices and buildings. Most FS have been employed to record physiologically defined features, such as center of pressure, step length and cadence etc. [12], [13] rather than for collecting raw data over longer periods of time [14]. FS require minimal attention by the user

Manuscript received April 15, 2021; accepted April 28, 2021. Date of publication May 5, 2021; date of current version July 30, 2021. This article was presented in part at the Proceedings of the 2020 IEEE Sensors Applications Symposium (SAS), Kuala Lumpur, Malaysia. This work was supported by the Engineering and Physical Sciences Research Council (EPSRC) for the Award Sensor Fusion and Data Processing under Award 1925970. The associate editor coordinating the review of this article and approving it for publication was Dr. You Li. *(Corresponding author: Syed U. Yunas.)*

The authors are with the School of Electrical and Electronic Engineering, The University of Manchester, Manchester M13 9PL, U.K. (e-mail: syed.yunas@manchester.ac.uk; k.ozanyan@manchester.ac.uk).

Digital Object Identifier 10.1109/JSEN.2021.3077698

TABLE I
GAIT CLASSIFICATION USING MULTI-MODALITY SENSOR FUSION

References	Features	Models	Modalities	Test Measures
Vera-Rodriguez et al., 2013 [57]	Foot step and gait recognition	SVM	Video & floor sensors	EER = 4.8%
Mazumder at al., 2017 [58]	Stride time	RB-FNN	EMG, inertial & foot pressure sensors	Fused error rate = 0
Ding et al., 2018 [59]	Gait phase	LSTM	Inertial & foot switch sensors	Accuracy = 91.8%
Mun et al. 2018 [60]	Gait phase	DNN	Inertial,force sensors & depth camera	Accuracy = 95 %
Vu et al., 2018 [61]	Gait phase detection	ED-FNN	Inertial & force sensitive resistors	MAE = 2.1%±0.1
Beil et al., 2018[62]	Motion classification	HMM	Inertial & force sensors	Accuracy = 92.8%
Chalvatzaki et al., 2019 [63]	Gait stability prediction	LSTM	Depth camera & laser range finder	F-score = 86.79%
Kumar et al., 2019 [64]	Gait activity recognition	3D-CNN & LSTM	Video, inertial & pressure sensors	Accuracy = 91.3%
Ivanov et al., 2020 [65]	Gait recognition	CNN	Inertial & force sensors	Accuracy = 93.3%
Rahman et al., 2020 [66]	Gait classification	SVM	Inertial, ECG & respiratory sensors	F-score = 0.92
Syed et al., 2020 [67]	Gait activity recognition	K-SVM	Inertial & POF based floor sensors	Accuracy = 94%

and are suitable for continuous data capture. However, long term data acquisition motivates advances in sensor technology and data processing to extract unique features from the gait information.

In the past few years, the exponential rise in the efficiency and capabilities of sensor systems has enabled the extraction of more valuable information from sensing modalities [11]. Different sensing modalities have developed distinct set of features based on bio-mechanical measures related to physical body dimensions, body part masses and time varying forces generated by muscles during the gait cycle. Advances in gait sensing instruments have resulted in the evaluation of many human locomotion characteristics obtained from high quality information. However, the feasibility of a simple and widely used modality to adequately map the complex gait features is still unclear. Therefore, to capture the complex nature of gait information, a multisource and multi-modality sensor fusion approach is required. In this context, multi-modality sensor fusion uses information from multiple sources and provides a more comprehensive description of individual's gait [15]. Multi-sensor data fusion can be seen as combining data captured from multiple information sources however the resulting information pool produces a new representation, distinct from those captured by individual sensors. Still accuracy and performance of these systems is highly debated and there is significant amount of work for improving the quality of data from the gait sensors. A survey of results for gait classification using multi-modality sensor fusion is presented in table I.

Deep Learning (DL) has become the state-of-the-art in many pattern classification techniques such as iris [16], face [17], finger-print [18], palm vein [19], ECG [20], human action [21] and gait [22] etc. DL models require minimal pre-processing on complex data and are capable of achieving robust and improved accuracies when dealing with larger volumes and ranges of datasets. DL is called upon to maximize the use of data variance and remove the dependencies on hand-crafted features from individual whilst exploring the effectiveness of the combined information from a discriminant angle [59]–[62], [64]–[67].

Herewith, we present a unique DL approach to extract and fuse gait information from two different modalities i.e., FS and AIS. DL models, such as Feed-Forward Neural Networks (FFNN), Convolutional Neural Networks (CNN) and Long Short-Term Memory (LSTM) Networks are used to automatically extract the data representations, fusing rich features of gait patterns obtained from the two modalities and deliver high statistical confidence. Significant improvements in f-scores achieved using LSTM are discussed in section III.

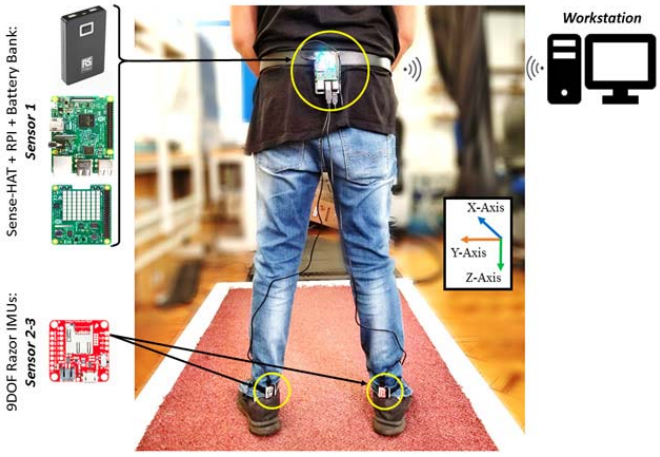


Fig. 1. AIS placed on the user, comprising the Sense-HAT board attached to the R-PI powered by a portable battery bank (sensor 1) and 9DOF Razor IMUs (sensors 2&3) connected through USB cables to RPI.

Our motivation in this work, is twofold: First, to analyse the change in gait occurring due to cognitive activities in lower portions of human body using two modalities i.e., FS and AIS; second, to fuse the spatio-temporal information with the aim to detect changes in human gait for health care scenarios i.e., age related factors [14], [23] and cognitive tasks [24], [25]. The effect of a dual tasks on gait at a certain age has a direct relation with the cognitive difficulty of the task and the type of gait performed [26]. Indeed, the importance of cognition is supported by the fact that gait changes are more common in people having cognitive impairment [27]. Sensors are contributing to the iterative process of engineering and development of new means to characterise gait. The comparison of results is complicated by the absence, to date, of a standard approach to experimental methodology to evaluate changes in gait from data acquired from the lower limbs by multiple modalities. In our previous work [28], feature extraction methods like PCA and CCA have been used with statistical methods to select the best optimal gait features for the fusion task. Feature domains containing many features increase the chances of redundancy and irrelevancy. In this paper using DL models, the automatic extraction of features from data leads to substantially more robust and accurate results as compared to the previous machine learning techniques. The proposed method is able to achieve reliable F-scores from a limited dataset.

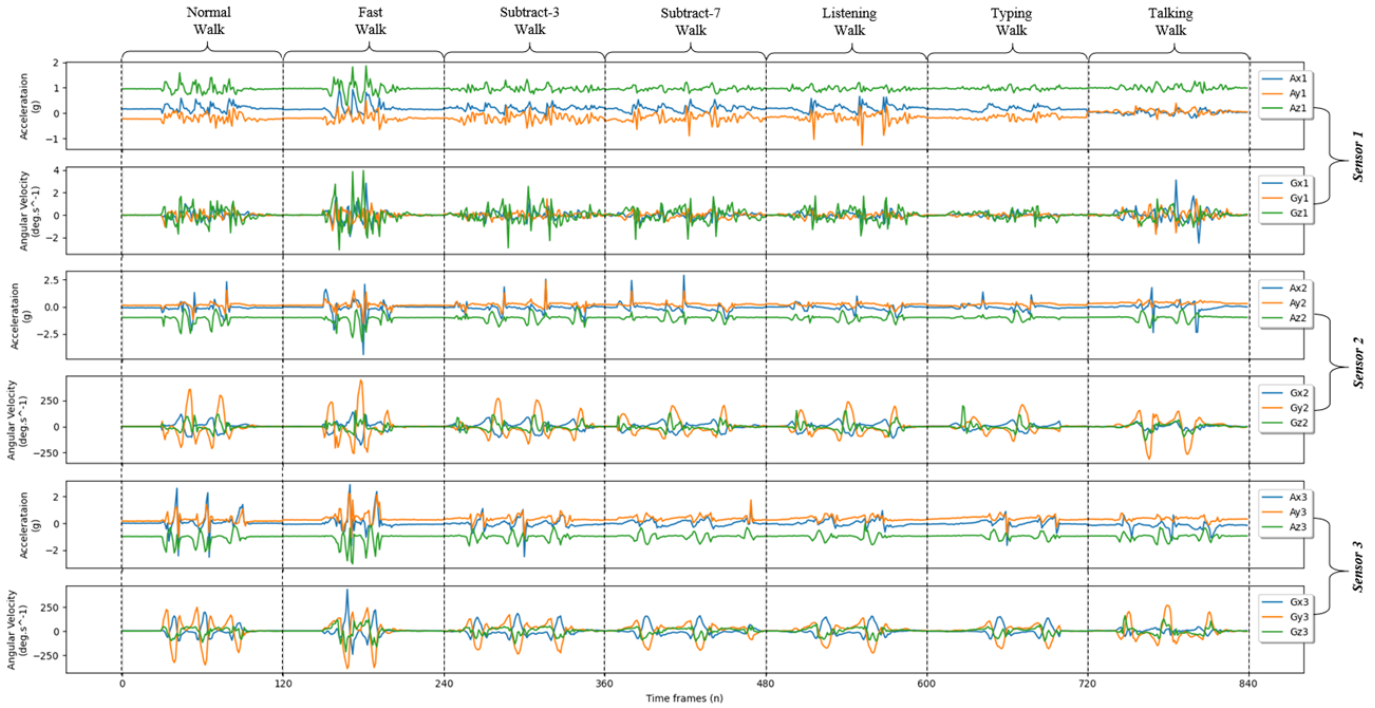


Fig. 2. Acceleration and angular velocity values from AIS for 7 gait activities i.e., normal walk, fast walk, dual tasks: subtracting number 3, number 7, listening, typing on mobile and talking whilst walking.

This paper is structured as follows: Section II describes the AIS and its functioning, together with the role of FS. Our approach of DL based sensor fusion is demonstrated further in that section. Section III presents results followed by discussions. In the end, section VI concludes paper.

II. Methods and Materials

A. Ambulatory Inertial Sensors (AIS)

A portable AIS system has been developed and deployed to study the effect of gait on the movements in the lower half of human body. The AIS system comprises: (i) a Raspberry-PI (R-PI) III Model b+ with a quad core 1.4GHz processor, 1GB RAM, Bluetooth and built-in Wi-Fi; (ii) Sense-HAT board [29] with a built-in 3D accelerometer (+/−16g) and a gyroscope (+/−2000dps), (iii) two 9DoF Razor IMUs [30] with Atmel SAMD21cortex-M0+ microprocessor (32 bit), a 3D accelerometer (+/−16g) and a gyroscope (+/−2000dps). The R-PI with the Sense-HAT board (attached at the top, see figure 1.) with a 2000-mAh portable battery bank (attached at the bottom) is called 'Sensor 1' which is connected through USB cables to both 9DoF Razor IMUs called 'Sensors 2 & 3'. Further, AIS is connected to a workstation for data transfer and control through a Wi-Fi connection as shown in figure 1.

Different number of sensors have been reported to capture gait activities in literature. However, deploying the minimum number of sensors may result in performance bottle necks whilst recording the complex gait activities [9]. The sensor positioning and number of sensors attached to the human body are also important factors whilst judging the quality of extracted data. Panebianco *et al.* [32], reported accuracies using 17 algorithms on 5 IMUs placed on back (1 IMU), shanks (2 IMUs) and feet (2 IMUs). To estimate the stance time, results obtained from the acceleration values of shank and foot performed better than the lower trunk. However, angular velocity estimation performed better in the detection of toe off and heel-strike events, with noticeable dependencies on sensor position.

From AIS, raw data on acceleration and angular velocity values is obtained from sensors 1-3 as shown in figure 2. The default sampling frequency of sensor 1 is 30Hz while sensor 2 and sensor 3 are sampled at 100Hz. After filtering and re-sampling, the spatio-temporal information from all three sensors of AIS is synchronized at 20Hz. Raw acceleration values are in two's compliment format, therefore they are converted into values between +16g and −16g. To calculate the angle (θ) from raw angular velocity (ω) values, the following formula is used:

$$\theta = \omega.\Delta t + \theta \tag{1}$$

where Δt is the time step. The nature of experiments requires subjects to start walking from one end of the FS to the other in forward direction whilst wearing the AIS. Therefore, all IMUs are aligned so that the highest acceleration (in forward direction) is represented by X-axis; weaker acceleration (vertical, in up/down direction between heel strike and toe off events of each foot) is represented by Z-axis; the weakest acceleration (lateral, in left/right direction) is represented by Y axis. For different gait activities, the results obtained from the three AIS sensors are shown in figure 2.

B. Floor Sensors (FS)

In this work, an original FS system (size: 2 m × 1 m approx.) is used to acquire the spatio-temporal dynamics of the ground reaction force during the chosen gait activities. This system comprises 116 plastic optical fiber (POF) sensor elements, each terminated with an LED as a light source and

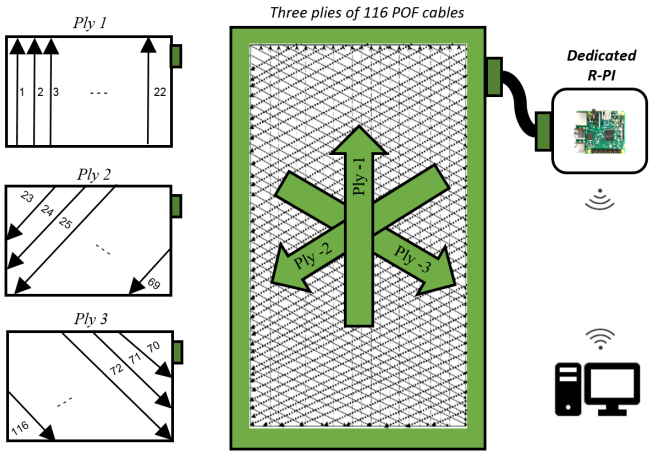


Fig. 3. Left: 3 plies of FS, Right: Overall connection of 116 POF sensors to the outside workstation through a dedicated RPI.

a photodiode as detector. The three-ply arrangement of POF based cables with circuit boards and wires are enclosed around the periphery of the FS and connecting with an umbilical cord to a R-PI in a shielded box as shown in figure 3. The set of POF sensors provides efficient sampling of the spatial-temporal distribution of the integrated transmission losses resulting from the applied pressure on the contact surface. The R-PI is used to transfer information to external work station using a Wi-Fi connection.

Data obtained from FS is a string of values output from 12bit ADC converter at every timestamp. These strings of information are processed and converted into transmitted light percentages. FS is synchronized at the same frequency of 20Hz used with AIS. Figure 4 shows the spatial average of the spatio-temporal information obtained from FS from gait activities,

$$SA\,[t] = \frac{1}{n}\sum_{k=1}^{n}(\mathrm{F_k}[t]) \tag{2}$$

where n is the total number of sensors and $\mathrm{F_k}$ is the amplitude of the individual sensor response. The full gait cycle can be represented (see figure 1 in [15]) as 5 events in the stance phase. Some of the gait events are possible to identify by visual inspection of data obtained separately from both modalities. For illustration purposes, in figure 5 only HS is indicated on time patterns from the FS signal (as the mean of all sensor values) and the AIS signal (as the root sum of squared maximum accelerations in all 3 directions, given by

$$A_{max} = \sqrt[2]{A_x^2 + A_y^2 + A_z^2} \tag{3}$$

C. Gait Activities

Human gait is no longer considered as an automated activity that utilises minimal higher-level cognitive input. In fact, the multi-faceted neuropsychological effects on gait and the interconnection between the mobility control and related factors incorporate new research pathways. Woollacott *et al.* reviewed the effect of dual task paradigms to observe the effect of age related changes in balance control and reductions in stability whilst performing an additional activity in healthier and elderly adults [33]. O'Sheas *et al.* [34] observed the performance of simultaneous motor or cognitive tasks such as walking at a certain speed (single task), transferring a coin (motor task) and performing number subtraction (cognitive task) on 15 PD patients. They concluded that gait changes whilst performing a cognitive and motor demanding task, however an additional secondary task does not necessarily determine the severity of disease. Costilla-Reyes *et al.* reported the capability of POF based FS (the "intelligent carpet" [35]) to detect changes in gait patterns using 10 manners of walking [36]. Zebin *et al.* [37] reported 6 daily life activities with 92% average recognition accuracy using only accelerometer and gyroscope data as inputs.

In this research, we have conducted 7 gait patterns with different activities i.e., normal walk, fast walk, subtracting 3 from a random number, subtracting 7, listening to the story, typing on mobile and talking to the operator. Data is collected from 11 healthy volunteers (gender: male/female; age[year]: 30.18 $\pm$ 7.7; weight[kg]: 71.18 $\pm$ 11.1; height[cm]: 173 $\pm$ 7.8), wearing AIS and walking on the FS at the same time. Each gait pattern is repeated 10 time for every activity. It takes the average user approximately 35 to 40 minutes to complete the 70 gait experiments including settling time between the experiments. A single gait activity starts when the user starts walking from one end of the FS and finishes when user steps off the FS on the other end. Therefore, one activity means recording of 2-3 step patterns and not a complete gait cycle. Data obtained from both modalities was stored in CSV format files on their dedicated RPIs and used on a workstation for further processing and fusion as presented further. For the proposed study, Manchester University Research Ethics Committee (MUREC) has granted the ethical approval to conduct experiments on healthy volunteers using FS and AIS. Written consent from each volunteer's was obtained prior to all experiments and research was conducted in accordance with the general guidelines of ethics board.

D. Deep Learning Based Multi-Modality Sensor Fusion

Multi-modality sensor fusion results in producing new data representations which are unique to the collection of individual sensors and modalities. Several modalities have demonstrated their capabilities to capture gait attributes and anomalies; however, most of these methods rely on handcrafted features. In such approaches, feature engineering might lose the salient features involved in problems. In our work, DL achieves the learning and extracting of highly statistically significant features from the gait activity data recorded from two different modalities. DL models implemented and used to extract gait features from both modalities, are discussed as follows:

1) Feed-Forward Neural Network (FFNN) Model: The neural network in which output from one layer is fed to the next layer in forward direction without any loops in the network is called a feed-forward neural network [38]. The basic architecture of a FFNN model consists of an input layer, few hidden layers and an output layer of neurons. In FFNN, the neurons in one layer are fully connected to the next layer through synapses or assigned weights to learn the complex representations of data.

In our work, for AIS, the training set is a 2D vector (73920 $\times$ 18) in which each row represents the spatial data at a single time instance. 18 Input values are passed to the fully connected (FC) layers of sizes 16, 12, 10, 8 and an

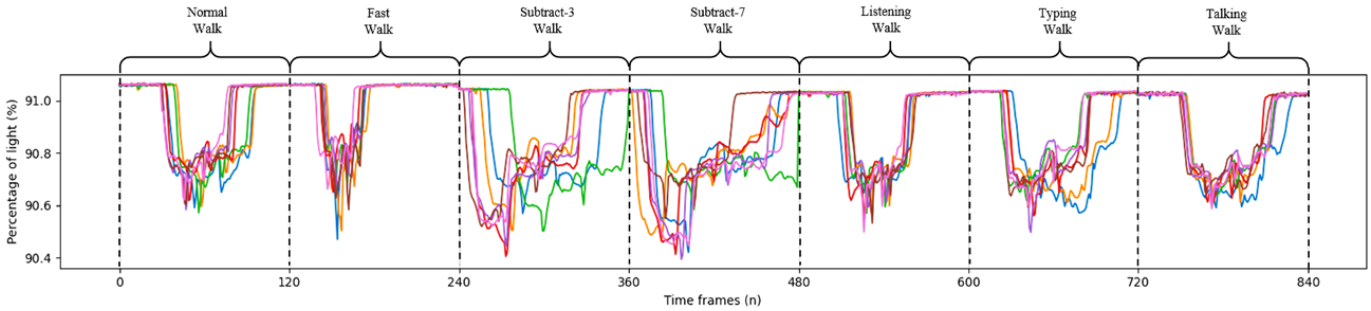


Fig. 4. FS results of a user performing 10 gait cycles for 7 gait activities i.e., normal walk, fast walk, dual tasks: subtracting number 3, number 7, listening, typing on mobile and talking whilst walking.

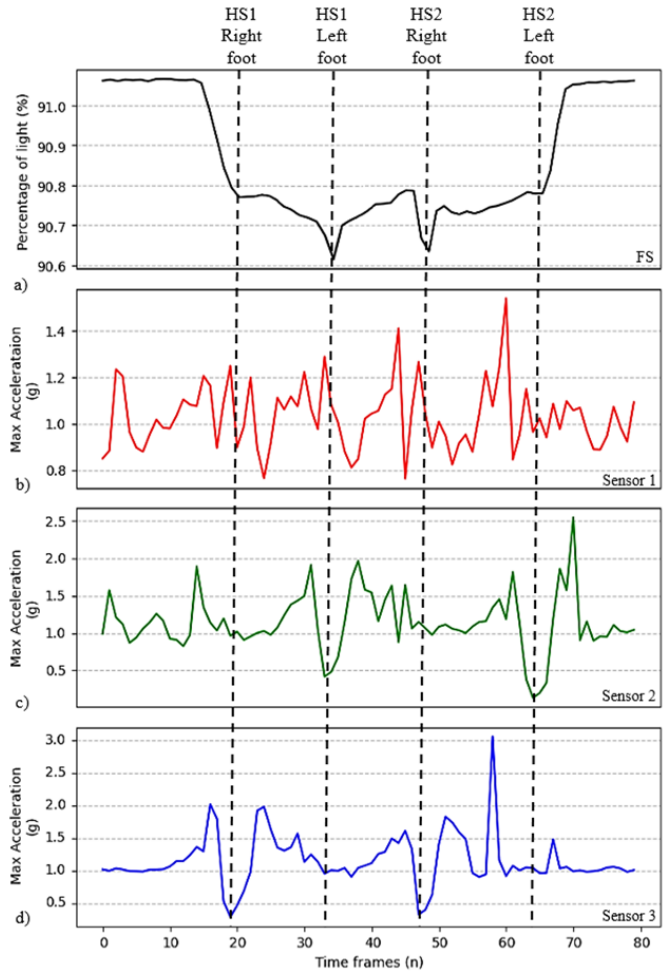


Fig. 5. (a) FS mean of sensor values vs time frames obtanied from FS; (b), (c), (d) Root sum of maximum accelerations vs time frames from AIS sensor-1,sensor-2,sensor-3 for a normal walk gait pattern. The notable heel strike dip at the dashed lines is alternating between the two legs: HS1-Right foot and HS2-Right foot from sensor 3 and HS1-Left foot and HS2-Left foot from sensor 2. Sensors 1 is not sensitive to HS1 and HS2, as expected, because of its position close to centre of mass and not on the limbs.

output layer of size 7 (representing 7 gait activities). The first layer size (16 being a multiple of 2 and closer to the average of input (18), output (1)) is selected, however, any number between number of input and output can be selected. Also higher accuracy is observed using first layer of size 16 than 8. The effect of every weight at FFNN layers is determined by the activation function which allows the model to achieve a desired output. To address the non-linearity of the spatio-temporal gait patterns in our dataset a Rectified Linear Unit (ReLU) activation function [39] is implemented at all the hidden layers. The weight of every neuron is multiplied by the input and passed through the activation function. Propagation continues until a prediction is achieved. At the output layer of size 7, a linear classifier *Softmax* is used to transform results into probabilities [40].

For FS, the training set is also a 2D vector (73920 $\times$ 116). 116 input values are passed to the FC layers of sizes 64, 32, 10, 8 and an output layer of size 7. For the multi-modality case, in order to create a balance between the number of features, the FC layers of size 10 from each modality are merged as shown in figure 6. The outputs from each layer are passed in forward direction to the next layer. For multi-modality case, forward propagation takes places over the merge layer.

Likewise forward propagation, the FC layers are responsible for the propagation of error in back ward direction. The predicted results are compared with the actual results and the error is quantified with the help of a cost function [41], [42]. We have used cross-entropy, based on a logarithmic function to handle very small errors. The error is back propagated in the form of updated weights send to the neurons layer-wise in backward direction. Among the gradient based algorithms such as stochastic gradient descent [43], conjugate gradient [44] and Adam [45], which are the commonly used methods for error optimization, the latter is used to determine the learning rate of new weights and biases in our research.

The above procedure is repeated and weights are updated after each batch of observations from the training set for each modality. Batch size of 120 observations is selected to update the weights which is equal to one activity. One epoch is completed when one whole training set passes through the FFNN. We have trained all experiments through 100 epochs for all cases. Results are further discussed in section III.

2) Convolutional Neural Networks (CNNs): CNN is a typical DL model which uses different levels of abstraction to learn the hierarchical representations of patterns existing in the dataset. CNN have been extensively utilised to classify and recognize humans based on various gait parameters i.e., foot-steps [46], gender and age [47], gait energy images [48], gestures [49] and freezing of gait [50] etc.

a) 1D-Approach: A basic CNN consists of an input layer, convolution layers, down-sampling or pooling layers, flattening layers, FC layers and an output layer [51]. In this work, the implementation of 1D-CNN for single and

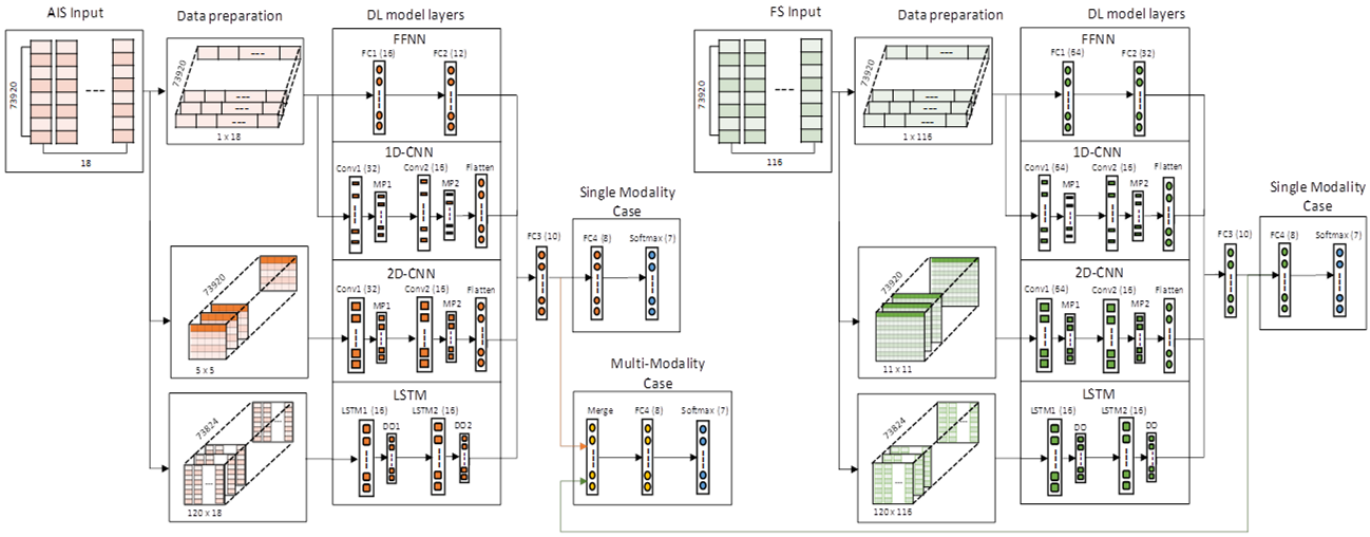


Fig. 6. Block diagram of single modality and multi-modality cases.

multi-modality cases can be seen in figure 6. For AIS, the training set 73920 × 18 is converted into 73920 arrays of size 1 × 18, where a single array determines the spatio information at a single time instance. Each array is passed to a 1D-Convolution layer (Conv1: 32 filters, kernel size 3, stride 1) to automatically extract the unique variability features from the training dataset. Max-pooling layer (MP1: kernel size 2) is used to down-sample the large volume of data after convolution. Results obtained from MP1 are feed to another 1D-Convolution layer (Conv2: 16 filters, kernel size 3, stride 1) and a Max-pooling layer (MP2: kernel size 2). Extracted features from max-pooling layers are in 2D format and therefore required to get aligned in a 1D feature vector of inputs for FC layers using the flattening function. FC layer of size 16 is used to connect flattening output. ReLU is the activation function used to handle the non-linearity at the convolution layers and the FC layers. Softmax function is used at the output layer (size 7) as discussed earlier.

For FS, the training set 73920 × 116 is converted into 73920 arrays of size 1 × 116, where a single array contains the spatial information at a single time instance. Each array is passed to a 1D-Convolution layer (Conv1: 64 filters, kernel size 3), Max-pooling layer (MP1: kernel size 2), another 1D-Convolution layer (Conv2: 16 filters, kernel size 3), Max-pooling layer (MP2: kernel size 2), flattening layer and a FC layer of size 16 followed by output layer of size 7. For the multi-modality case, in order to create a balanced number feature set, the FC layers of size 10 from each modality are merged. Results obtained for all cases are discussed in section III.

b) 2D-Approach: The implementation of 2D-CNN for single and multi-modality cases can be seen in figure 6. The same number of layers and filters at each layer are used for the 1D and 2D approach. However the dimensions of inputs and size of convolutional and max-pooling layer are different. Since CNN are most commonly applied to analyze visual images, therefore we have utilized their ability by transforming the 18 inputs of AIS into a 5 × 5 image and 116 inputs of FS into a 7 × 7 image with zero padded columns each. The filters kernel size is 3 × 3 for each convolutional layer and 2 × 2 and the max-pooling layer. Results obtained for all cases are discussed in section III.

3) Long Short Term Memory (LSTM): The basic structure of Recurrent Neural Networks (RNNs) is similar to FFNNs, where connections exist among hidden layer units based on time delays. These connections retain the information from previous inputs and help to find out the temporal correlations between events which are spread out in the dataset. However, the network output while cycling around recurrent connections gets affected from exponentially vanishing or exploding gradients [52]. Therefore, the efficient gradient-based technique, Long Short-Term Memory (LSTM), is introduced to cover the time lag between the time steps by enforcing constant error flow within special cells [53].

LSTM models work on time-processed data and are capable of learning time dependencies in sequence prediction problems. Since timestamps are equal in number for both modalities, the first layer of operation has been implemented with 16 blocks for both cases. Stacked layer LSTM models have been used to deeply exploit the dependencies between time-stamps [54]. The two stacked layered LSTMs, reported in many cases have been adopted in our approach to implement the individual [37] and multi-modality cases [55].

For AIS, the training data set is a 2D vector (73920 timestamps × 18 inputs) which is converted into a 3D vector (73824 time-stamps × 120 window samples × 18 inputs), with 120 window samples out of 119 serve as memory for the associated time-stamp. Training data is fed to the successive LSTM model in the form of batches of size 120 for different epoch values. The LSTM stack of two layers is implemented with 16 LSTM units on each layer, followed by a dropout layer (DO) utilizing 20% probability of data to prevent any overfitting. In case of FS, we have training data as a 3D vector (73824 time-stamps × 120 window samples × 116 inputs) following a similar LSTM model like AIS. After two layers a similar layered approach has been utilized as in case of FFNN, 1D-CNN, 2D-CNN for single and multi-modality cases as shown in figure 6.

4) Fused Approach: The fusion approach in this work is DL based fusion of lower human body joint angle trajectories

(obtained from an AIS modality) and ground reaction forces generated by feet (obtained using POF based FS modality). The implementation of AIS and FS that each modality records their datasets on their own RPI. The two RPIs are programmed to synchronize and record readings at 20Hz. Both modalities are checked and synchronisation is verified by test programs before starting experiments. From the point of view of the fusion task, the data is collected from synchronized RPIs separately and used for further analysis purposes.

Python environment libraries, including TensorFlow and Keras, are utilized to implement and run DL models. DL model layers are capable of processing the body orientation, positioning and forces in space and time using AIS. Likewise, these layers are equally useful to process the effect of forces resulted in foot on ground contact captured in FS data. In this research, we have used the first two layers from each DL models (as shown in figure 6) to automatically extract unique gait activity features from both modalities that mostly contribute towards gait classification whilst dropping the less significant values across the complex network layers. Fusion of such unique information helps to retain most of the gait activity dynamics from individual modalities. Keras functional API defines advanced network topologies and help to design complex problems unlike sequential APIs. We have used Keras functional API to build arbitrary graph of layers and handle shared layers to fit our fusion approach. There are many types of merge layers supported by Keras Functional API [56], some of the implemented merge layers used in our approach are as follows:

1. *Add:* Adds two same-sized input vectors (element-wise) into a single vector of the same size as input.
2. *Multiply:* Multiply two inputs vectors (similar to add).
3. *Average:* Computes the average of two input vectors.
4. *Maximum:* Computes the maximum of the two input vectors (element-wise) into a single vector of same size.
5. *Minimum:* Computes the minimum of the two input vectors.
6. *Concatenate:* Combines two inputs vectors into a single long vector, so that that the second input comes after first.

The listed layers perform arithmetic operations on their input layers and require them to be the same shape for fusion. However, concatenate layer can work with different shape inputs. Results obtained using these layers are discussed in section III.

III. Results and Discussions

The captured movements in the lower parts of the human body, by AIS and of foot falls by FS, in the general case are not independent from each other. The perceived coordination and complementarity of both data sources justify the need of fusion. It is also expected that fusion would partially accommodate possible inaccuracies in spatio-temporal data in certain situations resulting in improved robustness as compared to a single modality. In our work, comparing the results achieved using single and multi-modality systems are used to explore the benefits of complementary modalities in comparison with the cost and acceptability by the user. While retaining our focus on multi-modality fusion, significant differences in the performance of multiple DL models are summarized below:

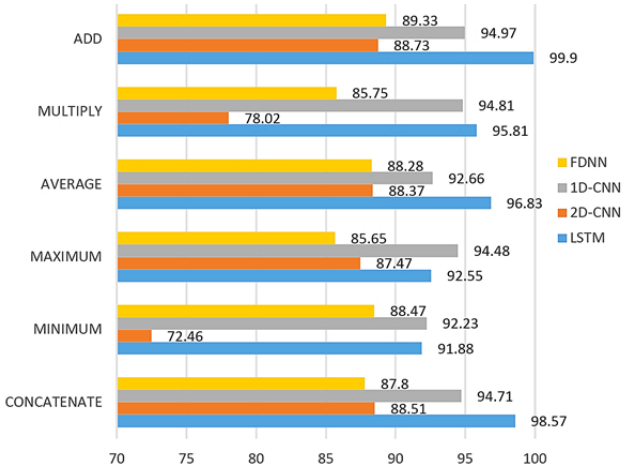


Fig. 7. F-scores for overall gait classification using merge layers (Epochs: 100, Batch size: 120).

A. Multi-Modality Fusion

For the classification of gait activities, 8,400 samples are obtained from one person during 7 gait activities using one modality. Therefore, total 92,400 samples are collected using one modality and 184800 using multi-modality case. Each case of single and multi-modality is split into 80% training (73,920 samples shown in figure 6), 10% validation and 10% test sets before feeding to the DL model. Training data from two modalities ($92{,}400 \times 2 = 184{,}800$ samples) is further tested and verified for epochs 1-100 and batch size 120 across the test datasets. All data processing and computational tasks were conducted on Lenovo ThinkPad with Intel® Core™ i7-8560U CPU @ 1.9GHz 2.11GHz and 8GB physical memory.

It is expected to achieve higher f-scores for FS (116 inputs) as compared to AIS (18 inputs) which is corroborated by Table II. It is also understandable that the execution time to generate classifications from FS is much higher than AIS. In our work, we have proposed a fusion strategy to balance the disproportional number of inputs between the two modalities, without substantial degradation of the information content. The classification features obtained using the fused multi-modality data yielded better f-scores as compared to individual modalities using all DL models (see Table II, columns 3, 5 & 7). However, this is achieved at higher execution time than single modalities. The effectiveness of this DL based multi-modality fusion has been further tested and verified using different fusion techniques as discussed in section II.D.4. The 'add' method appears to deliver the most accurate fused result among all. However worst f-scores are obtained using 'minimum' and 'multiply' methods in case of 2D-CNN, as shown in figure 7.

B. DL Models

The accuracies for DL models: FFNN, 1D-CNN, 2D-CNN and LSTM are higher in multi-modality cases compared to individual modalities. In case of FS (see Table II, column 3), 1D-CNN shows higher accuracies for all epochs when compared with FFNN and 2D-CNN. Comparison of 1D-CNN with LSTM shows mixed results with higher accuracies for 50 and 100 epochs in the latter case. LSTM models work on time processed data and are capable of learning time dependencies

TABLE II
F-SCORES FOR SINGLE MODALITY AND MULTI-MODALITY FUSION USING DL MODELS

DL Models	Epochs	FS	Execution Time (hh:mm:ss)	AIS	Execution Time (hh:mm:ss)	FS & AIS Multi-modality Fusion	Execution Time (hh:mm:ss)
FFNN	1	50.09	00:00:01	18.69	00:00:01	54.64	00:00:02
	5	72.56	00:00:07	30.11	00:00:06	73.84	00:00:09
	10	75.48	00:00:13	31.14	00:00:12	78.27	00:00:18
	50	81.19	00:01:06	36.69	00:01:04	87.35	00:01:27
	100	82.49	00:02:13	38.44	00:02:11	89.33	00:03:06
1D-CNN	1	50.34	00:00:12	27.19	00:00:04	51.25	00:00:12
	5	73.75	00:00:56	32.40	00:00:21	80.90	00:01:05
	10	79.72	00:02:01	34.73	00:00:43	84.75	00:02:06
	50	89.12	00:10:23	41.20	00:01:44	93.52	00:10:43
	100	94.71	00:20:01	42.47	00:03:30	94.97	00:21:14
2D-CNN	1	40.15	00:00:18	27.42	00:00:05	49.73	00:00:13
	5	64.70	00:00:59	34.63	00:00:26	73.31	00:01:05
	10	69.10	00:02:01	38.16	00:00:51	78.30	00:02:10
	50	76.89	00:10:14	44.29	00:03:34	85.70	00:10:49
	100	80.40	00:20:25	45.75	00:07:03	88.73	00:21:40
LSTM	1	34.56	00:02:32	33.20	00:01:08	41.43	00:03:17
	5	54.79	00:11:09	44.05	00:05:46	69.83	00:16:12
	10	72.49	00:23:51	68.26	00:11:15	90.93	01:06:37
	50	99.50	05:22:55	91.23	00:56:15	99.77	12:14:32
	100	99.78	22:51:17	95.19	01:52:31	99.90	23:23:45

in sequence prediction problems. Since the time-stamps are equal in number for both modalities, the first two layers of operation have been implemented with 16 units for both cases. A higher number i.e., 32 or 64, is reportedly beyond the capabilities of the computer system used. LSTM shows higher accuracies for all epochs when compared with FFNN, 1D-CNN and 2D-CNN in case of AIS (see Table II, column 5). For the multi-modality fusion case, LSTM has the highest accuracies for 10, 50 and 100 epochs of all DL models.

In case of 1D-CNN and 2D-CNN, the first two layers of operation have the same number of filters for single and multi-modality cases. FS data, considering a 5-fold larger number of inputs compared to AIS, have shown the maximum accuracy with 64 filters, as compared to AIS with 32 filters. AIS has been checked with 16 filters too manifesting reduced accuracies. The scope of this research is to report the most suitable approach for the fusion task. 1D-CNN proves itself as a second choice when compared with 2D-CNN and a single exception at 1 epoch with FFNN.

Columns 4 & 6 show that for FS, the execution time to train 1D-CNN, 2D-CNN and LSTM models is significantly higher than AIS for all epochs. Only FFNN has comparatively closer execution times using FS and AIS. In case of multi-modality fusion, FFNN takes much lesser time compared to LSTM which manifests the highest execution time for all epochs (see Table II, column 8). Therefore, the execution time is in a trade-off with the overall performance of the system. Best accuracy could be achieved using LSTM-based DL model when speed of execution is not of concern and data processing system with higher specifications is utilized.

Model-wise multi-modality fusion f-scores for all gait activities in figure 8 demonstrate that LSTM yielded f-scores superior to all other DL models. The 'typing' and 'talking' gait show worst f-score results among all activities: 64.09% (lowest) and 80.01% in case of FFNN; 75.87% and 70.09% in case of 2D-CNN. 1D-CNN model appears as the second choice due to its second highest f-scores for all gait activities, with some exceptions in 'subtracting-3'and 'listening', as well as 'subtracting-7' gait, showing lesser f-scores than 2D-CNN and FFNN respectively. It is noticeable that 1D-CNN shows worst f-score for 'listening', which is as high as 89.03% compared to FFNN (64.09% for 'typing' gait) and 2D-CNN (70.09% for 'talking' gait). Standard deviation in model- wise f-scores of multi-modality fusion for all classes is calculated as: LSTM (0.09%), 1D-CNN (3.05%), 2D-CNN (10.18%) and FFNN (11.81%).

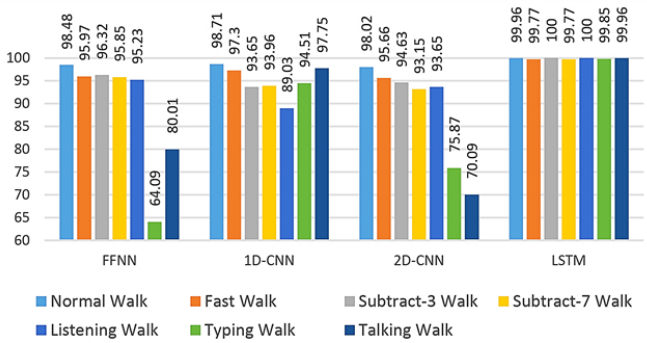


Fig. 8. Model-wise f-scores of multi-modality fusion for all classes (Epochs: 100, Batch Size: 120).

Furthermore, FFNN shows an overall f-score of 89.33% with minimum execution time (03min:06sec) and LSTM shows a highest f-score of 99.9% with maximum execution time (23hr:23min:45sec). However, 1D-CNN appears as the best DL model for overall performance for the proposed multi-modality fusion due to its performance f-score (94.97%) and a reasonable execution time (21min:14sec) to train the model (see table II, columns:7 & 8).

IV. CONCLUSION

Multi-modality sensor fusion based on DL is new and reports of such fusion are few, which should be interpreted in the light of scarcity of suitable datasets. We demonstrated multi-modality sensor fusion for gait activity classification using deep learning. FFNN, 1D-CNN, 2D-CNN and LSTM models were implemented and used to fuse spatio-temporal

gait activity data using two sensing modalities: ambulatory inertial sensors and floor sensors. Overall performance was studied in detail and revealed best f-score of 99.9% in case of LSTM and fastest execution time 3 min 06 sec in the case of 2D-CNN.

The classification obtained from multi-modality sensor fusion would undoubtedly be superior compared to that from a single modality. However, the choice of optimal fusion algorithms should also involve the assessment of practicality, design, built and maintenance characteristics of such complex systems.

REFERENCES

[1] M. Yoneyama, Y. Kurihara, K. Watanabe, and H. Mitoma, "Accelerometry-based gait analysis and its application to Parkinson's disease assessment—Part 2: A new measure for quantifying walking behavior," *IEEE Trans. Neural Syst. Rehabil. Eng.*, vol. 21, no. 6, pp. 999–1005, Nov. 2013, doi: 10.1109/TNSRE.2013.2268251.

[2] C. Strohrmann, H. Harms, C. Kappeler-Setz, and G. Troster, "Monitoring kinematic changes with fatigue in running using body-worn sensors," *IEEE Trans. Inf. Technol. Biomed.*, vol. 16, no. 5, pp. 983–990, Sep. 2012, doi: 10.1109/TITB.2012.2201950.

[3] M. Destephe, T. Maruyama, M. Zecca, K. Hashimoto, and A. Takanishi, "The influences of emotional intensity for happiness and sadness on walking," in *Proc. 35th Annu. Int. Conf. IEEE Eng. Med. Biol. Soc. (EMBC)*, Jul. 2013, pp. 7452–7455, doi: 10.1109/EMBC.2013.6611281.

[4] Y.-C. Liu, Y.-R. Yang, Y.-A. Tsai, R.-Y. Wang, and C.-F. Lu, "Brain activation and gait alteration during cognitive and motor dual task walking in stroke—A functional near-infrared spectroscopy study," *IEEE Trans. Neural Syst. Rehabil. Eng.*, vol. 26, no. 12, pp. 2416–2423, Dec. 2018, doi: 10.1109/TNSRE.2018.2878045.

[5] N. Samudin, W. N. M. Isa, T. H. Maul, and W. K. Lai, "Analysis of gait features between loaded and normal gait," in *Proc. 5th Int. Conf. Signal Image Technol. Internet Based Syst.*, Nov. 2009, pp. 172–179, doi: 10.1109/SITIS.2009.37.

[6] N. Margiotta, G. Avitabile, and G. Coviello, "A wearable wireless system for gait analysis for early diagnosis of alzheimer and parkinson disease," in *Proc. 5th Int. Conf. Electron. Devices, Syst. Appl. (ICEDSA)*, Dec. 2016, pp. 1–4, doi: 10.1109/ICEDSA.2016.7818553.

[7] H. Lee, S. J. Sullivan, and A. G. Schneiders, "The use of the dual-task paradigm in detecting gait performance deficits following a sports-related concussion: A systematic review and meta-analysis," *J. Sci. Med. Sport*, vol. 16, no. 1, pp. 2–7, Jan. 2013, doi: 10.1016/j.jsams.2012.03.013.

[8] W. Tao, T. Liu, R. Zheng, and H. Feng, "Gait analysis using wearable sensors," *Sensors*, vol. 12, no. 2, p. 2255, 2012, doi: 10.3390/S120202255.

[9] W. Kong, S. Sessa, M. Zecca, and A. Takanishi, "Anatomical calibration through post-processing of standard motion tests data," *Sensors*, vol. 16, no. 12, p. 2011, Nov. 2016, doi: 10.3390/s16122011.

[10] H. Chan, H. Zheng, H. Wang, and D. Newell, "Assessment of gait patterns of chronic low back pain patients: A smart mobile phone based approach," in *Proc. IEEE Int. Conf. Bioinf. Biomed. (BIBM)*, Nov. 2015, pp. 1016–1023, doi: 10.1109/BIBM.2015.7359823.

[11] P. Connor and A. Ross, "Biometric recognition by gait: A survey of modalities and features," *Comput. Vis. Image Understand.*, vol. 167, pp. 1–27, Feb. 2018, doi: 10.1016/j.cviu.2018.01.007.

[12] G. Qian, J. Zhang, and A. Kidane, "People identification using floor pressure sensing and analysis," *IEEE Sensors J.*, vol. 10, no. 9, pp. 1447–1460, Sep. 2010, doi: 10.1109/JSEN.2010.2045158.

[13] A. Muro-de-la-Herran, B. Garcia-Zapirain, and A. Mendez-Zorrilla, "Gait analysis methods: An overview of wearable and non-wearable systems, highlighting clinical applications," *Sensors*, vol. 14, no. 2, pp. 3362–3394, Feb. 2014, doi: 10.3390/s140203362.

[14] O. Costilla-Reyes, P. Scully, and K. B. Ozanyan, "Age-sensitive differences in single and dual walking tasks from footprint floor sensor data," in *Proc. IEEE Sensors*, Oct. 2017, pp. 1–3, doi: 10.1109/ICSENS.2017.8234299.

[15] A. S. Alharthi, S. U. Yunas, and K. B. Ozanyan, "Deep learning for monitoring of human gait: A review," *IEEE Sensors J.*, vol. 19, no. 21, pp. 9575–9591, Nov. 2019, doi: 10.1109/jsen.2019.2928777.

[16] E. Ribeiro, A. Uhl, and F. Alonso-Fernandez, "Iris super-resolution using CNNs: Is photo-realism important to iris recognition?" *IET Biometrics*, vol. 8, no. 1, pp. 69–78, Jan. 2019, doi: 10.1049/iet-bmt.2018.5146.

[17] R. Raghavendra, K. B. Raja, S. Venkatesh, and C. Busch, "Transferable deep-CNN features for detecting digital and print-scanned morphed face images," in *Proc. IEEE Conf. Comput. Vis. Pattern Recognit. Workshops (CVPRW)*, Jul. 2017, pp. 1822–1830, doi: 10.1109/CVPRW.2017.228.

[18] F. Wu, J. Zhu, and X. Guo, "Fingerprint pattern identification and classification approach based on convolutional neural networks," *Neural Comput. Appl.*, vol. 32, no. 10, pp. 5725–5734, May 2020, doi: 10.1007/s00521-019-04499-w.

[19] S. F. Chevtchenko, R. F. Vale, and V. Macario, "Multi-objective optimization for hand posture recognition," *Expert Syst. Appl.*, vol. 92, pp. 170–181, Feb. 2018, doi: 10.1016/j.eswa.2017.09.046.

[20] R. Donida Labati, E. Muñoz, V. Piuri, R. Sassi, and F. Scotti, "Deep-ECG: Convolutional neural networks for ECG biometric recognition," *Pattern Recognit. Lett.*, vol. 126, pp. 78–85, Sep. 2019, doi: 10.1016/j.patrec.2018.03.028.

[21] E. P. Ijjina and K. M. Chalavadi, "Human action recognition in RGB-D videos using motion sequence information and deep learning," *Pattern Recognit.*, vol. 72, pp. 504–516, Dec. 2017, doi: 10.1016/j.patcog.2017.07.013.

[22] Z. Wu, Y. Huang, L. Wang, X. Wang, and T. Tan, "A comprehensive study on cross-view gait based human identification with deep CNNs," *IEEE Trans. Pattern Anal. Mach. Intell.*, vol. 39, no. 2, pp. 209–226, Feb. 2017, doi: 10.1109/TPAMI.2016.2545669.

[23] A. W. Priest, K. B. Salamon, and J. H. Hollman, "Age-related differences in dual task walking: A cross sectional study," *J. NeuroEng. Rehabil.*, vol. 5, no. 1, p. 29, 2008, doi: 10.1186/1743-0003-5-29.

[24] J. M. Hausdorff, A. Schweiger, T. Herman, G. Yogev-Seligmann, and N. Giladi, "Dual-task decrements in gait: Contributing factors among healthy older adults," *J. Gerontol. A, Biol. Sci. Med. Sci.*, vol. 63, no. 12, pp. 1335–1343, Dec. 2008, doi: 10.1093/gerona/63.12.1335.

[25] G. Yogev-Seligmann, J. M. Hausdorff, and N. Giladi, "The role of executive function and attention in gait," *Movement Disorders*, vol. 23, no. 3, pp. 329–342, Feb. 2008, doi: 10.1002/mds.21720.

[26] R. Beurskens and O. Bock, "Age-related deficits of dual-task walking: A review," *Neural Plasticity*, vol. 2012, Oct. 2012, Art. no. 131608, doi: 10.1155/2012/131608.

[27] M. Montero-Odasso, S. W. Muir, and M. Speechley, "Dual-task complexity affects gait in people with mild cognitive impairment: The interplay between gait variability, dual tasking, and risk of falls," *Arch. Phys. Med. Rehabil.*, vol. 93, no. 2, pp. 293–299, Feb. 2012, doi: 10.1016/j.apmr.2011.08.026.

[28] S. U. Yunas and K. B. Ozanyan, "Gait activity classification from feature-level sensor fusion of multi-modality systems," *IEEE Sensors J.*, vol. 21, no. 4, pp. 4801–4810, Feb. 2021, doi: 10.1109/JSEN.2020.3028697.

[29] *Sense HAT—Raspberry Pi Documentation*. Accessed: Sep. 28, 2019.[Online]. Available: https://www.raspberrypi.org/documentation/hardware/sense-hat/

[30] *9DoF Razor IMU M0 Hookup Guide—Learn*. Accessed: Sep. 28, 2019. [Online]. Available: https://tinyurl.com/9DoFRazorIMU

[31] N. A. Capela, E. D. Lemaire, N. Baddour, M. Rudolf, N. Goljar, and H. Burger, "Evaluation of a smartphone human activity recognition application with able-bodied and stroke participants," *J. Neuroeng. Rehabil.*, vol. 13, no. 1, p. 5, Jan. 2016, doi: 10.1186/s12984-016-0114-0.

[32] G. Pacini Panebianco, M. C. Bisi, R. Stagni, and S. Fantozzi, "Analysis of the performance of 17 algorithms from a systematic review: Influence of sensor position, analysed variable and computational approach in gait timing estimation from IMU measurements," *Gait Posture*, vol. 66, pp. 76–82, Oct. 2018, doi: 10.1016/J.GAITPOST.2018.08.025.

[33] M. Woollacott and A. Shumway-Cook, "Attention and the control of posture and gait: A review of an emerging area of research," *Gait Posture*, vol. 16, no. 1, pp. 1–14, Aug. 2002, doi: 10.1016/S0966-6362(01)00156-4.

[34] S. O'Shea, M. E. Morris, and R. Iansek, "Dual task interference during gait in people with parkinson disease: Effects of motor versus cognitive secondary tasks," *Phys. Therapy*, vol. 82, no. 9, pp. 888–897, Sep. 2002, doi: 10.1093/ptj/82.9.888.

[35] J. A. Cantoral-Ceballos *et al.*, "Intelligent carpet system, based on photonic guided-path tomography, for gait and balance monitoring in home environments," *IEEE Sensors J.*, vol. 15, no. 1, pp. 279–289, Jan. 2015, doi: 10.1109/JSEN.2014.2341455.

[36] O. Costilla-Reyes, P. Scully, and K. B. Ozanyan, "Temporal pattern recognition in gait activities recorded with a footprint imaging sensor system," *IEEE Sensors J.*, vol. 16, no. 24, pp. 8815–8822, Dec. 2016, doi: 10.1109/JSEN.2016.2583260.

[37] T. Zebin, M. Sperrin, N. Peek, and A. J. Casson, "Human activity recognition from inertial sensor time-series using batch normalized deep LSTM recurrent networks," in *Proc. Annu. Int. Conf. IEEE Eng. Med. Biol. Soc.*, Oct. 2018, pp. 1–4, doi: 10.1109/EMBC.2018.8513115.

[38] M. Nielsen. *Neural Networks and Deep Learning*. Accessed: Jun. 3, 2019. [Online]. Available: http://neuralnetworksanddeeplearning.com

[39] *Rectified Linear Nits Improve Restricted Boltzmann Machines Semantic Scholar*. Accessed: Apr. 8, 2021. [Online]. Available: https://www.semanticscholar.org/paper/Rectified-Linear-Units-Improve-Restricted-Boltzmann-Nair-Hinton/a538b05ebb01a40323997629e171c91aa28b8e2f

[40] Y. A. LeCun, L. Bottou, G. B. Orr, and K.-R. Müller, "Efficient BackProp," in *Neural Networks: Tricks of the Trade* (Lecture Notes in Computer Science), vol. 7700. Cham, Switzerland: Springer, 2012, pp. 9–48.

[41] Z. Wang and A. C. Bovik, "Mean squared error: Love it or leave it? A new look at signal fidelity measures," *IEEE Signal Process. Mag.*, vol. 26, no. 1, pp. 98–117, Jan. 2009, doi: 10.1109/MSP.2008.930649.

[42] J. Shore and R. Johnson, "Axiomatic derivation of the principle of maximum entropy and the principle of minimum cross-entropy," *IEEE Trans. Inf. Theory*, vol. IT-26, no. 1, pp. 26–37, Jan. 1980, doi: 10.1109/TIT.1980.1056144.

[43] L. Bottou, "Large-scale machine learning with stochastic gradient descent," in *Proc. 19th Int. Conf. Comput. Statist.*, 2010, pp. 177–186, doi: 10.1007/978-3-7908-2604-3_16.

[44] M. F. Møller, "A scaled conjugate gradient algorithm for fast supervised learning," *Neural Netw.*, vol. 6, no. 4, pp. 525–533, Nov. 1993, doi: 10.1016/S0893-6080(05)80056-5.

[45] D. P. Kingma and J. Ba, "Adam: A method for stochastic optimization," 2014, *arXiv:1412.6980*. [Online]. Available: http://arxiv.org/abs/1412.6980

[46] O. Costilla-Reyes, R. Vera-Rodriguez, P. Scully, and K. B. Ozanyan, "Spatial footstep recognition by convolutional neural networks for biometric applications," in *Proc. IEEE SENSORS*, Jan. 2017, pp. 1–7, doi: 10.1109/ICSENS.2016.7808890.

[47] Y. Sun, F. P.-W. Lo, and B. Lo, "A deep learning approach on gender and age recognition using a single inertial sensor," in *Proc. IEEE 16th Int. Conf. Wearable Implant. Body Sensor Netw. (BSN)*, May 2019, doi: 10.1109/BSN.2019.8771075.

[48] P. Nithyakani, A. Shanthini, and G. Ponsam, "Human gait recognition using deep convolutional neural network," in *Proc. 3rd Int. Conf. Comput. Commun. Technol. (ICCCT)*, Feb. 2019, pp. 208–211, doi: 10.1109/ICCCT2.2019.8824836.

[49] P. Wang, Q. Zhang, L. Li, F. Ru, D. Li, and Y. Jin, "Deep learning-based gesture recognition for control of mobile body-weight support platform," in *Proc. 13th IEEE Conf. Ind. Electron. Appl. (ICIEA)*, May 2018, pp. 1803–1808, doi: 10.1109/ICIEA.2018.8398001.

[50] Y. Zhang and D. Gu, "A deep convolutional-recurrent neural network for freezing of gait detection in patients with Parkinson's disease," in *Proc. 12th Int. Congr. Image Signal Process., Biomed. Eng. Informat. (CISP-BMEI)*, Oct. 2019, doi: 10.1109/CISP-BMEI48845.2019.8965723.

[51] Y. LeCun, L. Bottou, Y. Bengio, and P. Haffner, "Gradient-based learning applied to document recognition," *Proc. IEEE*, vol. 86, no. 11, pp. 2278–2324, Nov. 1998, doi: 10.1109/5.726791.

[52] R. Pascanu, T. Mikolov, and Y. Bengio, "On the difficulty of training recurrent neural networks," in *Proc. 30th Int. Conf. Mach. Learn.*, vol. 3, Nov. 2012, pp. 2347–2355. Accessed: Jul. 1, 2020. [Online]. Available: http://arxiv.org/abs/1211.5063

[53] S. Hochreiter and J. Schmidhuber, "Long short-term memory," *Neural Comput.*, vol. 9, no. 8, pp. 1735–1780, Nov. 1997, doi: 10.1162/neco.1997.9.8.1735.

[54] I. Yeo and K. Balachandran, "Sentiment analysis on time-series data using weight priority method on deep learning," in *Proc. Int. Conf. Data Sci. Commun.*, Mar. 2019, pp. 1–6, doi: 10.1109/IconDSC.2019.8816985.

[55] W. Feng, N. Guan, Y. Li, X. Zhang, and Z. Luo, "Audio visual speech recognition with multimodal recurrent neural networks," in *Proc. Int. Joint Conf. Neural Netw. (IJCNN)*, May 2017, pp. 681–688, doi: 10.1109/IJCNN.2017.7965918.

[56] *Merge Layers—Keras Documentation*. Accessed: Sep. 28, 2019. [Online]. Available: https://keras.io/layers/merge/

[57] R. Vera-Rodriguez, J. Fierrez, J. S. D. Mason, and J. Ortega-Garcia, "A novel approach of gait recognition through fusion with footstep information," in *Proc. Int. Conf. Biometrics (ICB)*, Jun. 2013, pp. 1–6, doi: 10.1109/ICB.2013.6613014.

[58] O. Mazumder, A. S. Kundu, P. K. Lenka, and S. Bhaumik, "Multi-channel fusion based adaptive gait trajectory generation using wearable sensors," *J. Intell. Robot. Syst.*, vol. 86, nos. 3–4, pp. 335–351, Jun. 2017, doi: 10.1007/s10846-016-0436-y.

[59] Z. Ding *et al.*, "The real time gait phase detection based on long short-term memory," in *Proc. IEEE 3rd Int. Conf. Data Sci. Cyberspace (DSC)*, Jun. 2018, pp. 33–38, doi: 10.1109/DSC.2018.00014.

[60] K.-R. Mun, G. Song, S. Chun, and J. Kim, "Gait estimation from anatomical foot parameters measured by a foot feature measurement system using a deep neural network model," *Sci. Rep.*, vol. 8, no. 1, Dec. 2018, Art. no. 9879, doi: 10.1038/s41598-018-28222-2.

[61] H. Vu, F. Gomez, P. Cherelle, D. Lefeber, A. Nowé, and B. Vanderborght, "ED-FNN: A new deep learning algorithm to detect percentage of the gait cycle for powered prostheses," *Sensors*, vol. 18, no. 7, p. 2389, Jul. 2018, doi: 10.3390/s18072389.

[62] J. Beil, I. Ehrenberger, C. Scherer, C. Mandery, and T. Asfour, "Human motion classification based on multi-modal sensor data for lower limb exoskeletons," in *Proc. IEEE Int. Conf. Intell. Robots Syst.*, Dec. 2018, pp. 5431–5436, doi: 10.1109/IROS.2018.8594110.

[63] G. Chalvatzaki, P. Koutras, J. Hadfield, X. S. Papageorgiou, C. S. Tzafestas, and P. Maragos, "LSTM-based network for human gait stability prediction in an intelligent robotic rollator," in *Proc. IEEE Int. Conf. Robot. Autom.*, May, pp. 4225–4232, doi: 10.1109/ICRA.2019.8793899.

[64] P. Kumar, S. Mukherjee, R. Saini, P. Kaushik, P. P. Roy, and D. P. Dogra, "Multimodal gait recognition with inertial sensor data and video using evolutionary algorithm," *IEEE Trans. Fuzzy Syst.*, vol. 27, no. 5, pp. 956–965, May 2019, doi: 10.1109/TFUZZ.2018.2870590.

[65] K. Ivanov *et al.*, "Identity recognition by walking outdoors using multimodal sensor insoles," *IEEE Access*, vol. 8, pp. 150797–150807, 2020, doi: 10.1109/ACCESS.2020.3016970.

[66] M. J. Rahman, E. Nemati, M. Rahman, K. Vatanparvar, V. Nathan, and J. Kuang, "Toward early severity assessment of obstructive lung disease using multi-modal wearable sensor data fusion during walking," in *Proc. Annu. Int. Conf. IEEE Eng. Med. Biol. Soc.*, Jul. 2010, pp. 5935–5938, doi: 10.1109/EMBC44109.2020.9176559.

[67] S. U. Yunas, A. Alharthi, and K. B. Ozanyan, "Multi-modality fusion of floor and ambulatory sensors for gait classification," in *Proc. IEEE 28th Int. Symp. Ind. Electron. (ISIE)*, Jun. 2019, pp. 1467–1472, doi: 10.1109/ISIE.2019.8781127.

Syed U. Yunas received the M.Sc. degree in digital image and signal processing from The University of Manchester, Manchester, U.K., in 2011, where he is currently pursuing the Ph.D. degree in electrical and electronic engineering. He is currently working with Sensors and Sensing Systems Group. His research interests include sensor fusion and processing of data acquired from different gait modalities, such as floor and inertial sensors. For experimenting and research, he has designed and implemented his own wearable inertial sensor system to analyse human gait from the lower part of human body. His expertise involves designing of efficient and robust algorithms for sensor fusion of multi-modality systems.

Krikor B. Ozanyan (Senior Member, IEEE) received the M.Sc. degree in engineering physics (semiconductors) and the Ph.D. degree in solid-state physics, in 1980 and 1989, respectively. He is currently the Director of Research of EEE, The University of Manchester, U.K. He is a Fellow of the Institute of Engineering and Technology, U.K., and the Institute of Physics, U.K. He was a Distinguished Lecturer of the IEEE Sensors Council, in 2009 and 2010. He has more than 300 publications in the areas of photonic materials, devices, and systems for sensing and imaging. He was the General Co-Chair of the IEEE SENSORS 2017 Conference. He currently serves as the Vice-President for publications of the IEEE Sensors Council. He was a Guest Editor of the 10th Anniversary Issue of IEEE SENSORS JOURNAL in 2010, the Special Issues on Sensors for Industrial Process Tomography in 2005, and THz Sensing: Materials, Devices, and Systems in 2012. He was the Editor-in-Chief of the IEEE SENSORS JOURNAL from 2011 to 2018.