



Published in final edited form as:

*IEEE J Biomed Health Inform.* 2021 May ; 25(5): 1824–1831. doi:10.1109/JBHI.2020.3025049.

## Wearables and Deep Learning Classify Fall Risk from Gait in Multiple Sclerosis

**Brett M. Meyer [Student Member, IEEE], Lindsey J. Tulipani, Reed D. Gurchiek [Student Member, IEEE], Dakota A. Allen, Lukas Adamowicz, Dale Larie, Andrew J. Solomon, Nick Cheney, Ryan S. McGinnis [Member, IEEE]**

University of Vermont, Burlington, VT 05405 USA

### Abstract

Falls are a significant problem for persons with multiple sclerosis (PwMS). Yet fall prevention interventions are not often prescribed until after a fall has been reported to a healthcare provider. While still nascent, objective fall risk assessments could help in prescribing preventative interventions. To this end, retrospective fall status classification commonly serves as an intermediate step in developing prospective fall risk assessments. Previous research has identified measures of gait biomechanics that differ between PwMS who have fallen and those who have not, but these biomechanical indices have not yet been leveraged to detect PwMS who have fallen. Moreover, they require the use of laboratory-based measurement technologies, which prevent clinical deployment. Here we demonstrate that a bidirectional long short-term (BiLSTM) memory deep neural network was able to identify PwMS who have recently fallen with good performance (AUC of 0.88) based on accelerometer data recorded from two wearable sensors during a one-minute walking task. These results provide substantial improvements over machine learning models trained on spatiotemporal gait parameters (21% improvement in AUC), statistical features from the wearable sensor data (16%), and patient-reported (19%) and neurologist-administered (24%) measures in this sample. The success and simplicity (two wearable sensors, only one-minute of walking) of this approach indicates the promise of inexpensive wearable sensors for capturing fall risk in PwMS.

### Index Terms—

Wearables; Fall Risk; Digital Health; Machine Learning

## I. INTRODUCTION

More than half of the 2.3 million persons with multiple sclerosis (PwMS) experience a fall in any given three-month period [1]. Up to half of these falls result in an injury requiring medical attention, which increases fear of falling and decreases quality of life [2]. MS is a chronic condition [3], suggesting that these injurious falls pose a significant long-term

---

Personal use is permitted, but republication/redistribution requires IEEE permission.

R. S. McGinnis is the corresponding author for this work. ryan.mcginis@uvm.edu.

burden to the healthcare system. The time required to recover from fall-related injuries can worsen mobility problems and reduce independence in this population.

The progressive demyelination and axonal damage throughout the central nervous system [4], [5] caused by MS results in symptoms including debilitating fatigue and impaired coordination, muscle strength, and sensation [4]. Collectively, these symptoms lead to problems with balance and postural control, especially during dynamic activities [6]. Falls most often occur during balance-challenging daily activities such as position transfers, changes of direction, and walking [7]. Compared to healthy adults, PwMS fall more often and are more likely to be injured by a fall [8].

Fall history is one of the most important predictors of fall risk in PwMS [9], but only 51% of falls are self-reported [10]. In clinical practice, fall prevention interventions are not often prescribed until after a fall has been reported to a healthcare provider. An objective method of characterizing fall risk may enhance our ability to prescribe preventative interventions. However, quantitative measures of fall risk in PwMS remain nascent. While patient report measures [11], [12] and in-clinic functional assessments have been associated with risk of falling [13], [14], these approaches have poor performance when identifying fall risk [15]. Retrospective fall status classification commonly serves as an intermediate step in the development of prospective fall risk predictions [16]–[21] as it helps to identify biomarkers that have sufficient sensitivity to discriminate people who have fallen from those who have not. However, previous studies have not considered PwMS [16], [19]–[21], require expensive lab equipment [17], [18], and often leave room for improvement in classification performance [16].

Recent work has focused on capturing the differences in gait biomechanics between PwMS who have experienced a fall (*fallers*) and those who have not (*non-fallers*), and between PwMS and healthy controls [13], [22]–[24]. Fallers have less predictable trunk accelerations [22], slower gait speed [25], limited lower extremity joint angle excursion and range of motion [26], increased variability in spatiotemporal gait characteristics [13], [22], [24], and reduced margin of stability [22] compared to non-fallers with MS and/or healthy controls. These variables could provide objective indicators, or *biomarkers*, of fall risk that could be used to inform interventions. However, existing studies rely on data captured in constrained settings by technologies (e.g. optical motion capture, electronic walkway) that do not translate to clinical environments, limiting their utility.

Wearable inertial measurement units may provide an opportunity to quickly and unobtrusively capture the fall-related biomechanics of PwMS, and in clinical environments [27], [28]. To this end, several recent studies have demonstrated the ability of body worn sensors to capture biomechanical measures associated with fall risk in PwMS (e.g., walking speed [25], stride time variability [23]), and demonstrate correlations with clinical assessments [17]. However, these previous approaches fall short of developing a way to identify PwMS who are at high risk of falls, motivating the need for an objective risk measurement based on their biomechanics [29]. While machine learning techniques have been used to classify fall risk based on fall history using data from wearable sensors during gait (e.g., elderly [30], dementia [31]), these studies have not considered PwMS.

Importantly, gait biomechanics in PwMS differ significantly from the populations considered in previous work [22], [32], [33]. This indicates that statistical models for classifying fall risk may not generalize to PwMS.

Recent advances in machine learning, and particularly in artificial neural networks, have enabled learning of high-level outcomes directly from low-level raw sensory data streams. These *deep learning* models may be particularly well suited for developing statistical models for classifying fall risk as they automatically extract the most informative intermediate *latent* features for solving a given task (e.g. classification) and thus do not require the manual engineering of features from the wearable sensor data [34], [35]. Deep learning methods have been shown to provide superior results for time series classification tasks such as gait event detection from accelerometer data when compared to more classical machine learning techniques that rely on predetermined and manually-constructed features [36]. In fact, they have been used for several fall related tasks in other populations, including identifying fall status [20], [37], predicting fall risk [38], detecting falls [39]–[43], and predicting falls [44]. However, these approaches have not yet been applied to PwMS.

The primary objective of this work was to determine if a machine learning based analysis of wearable sensor data recorded during gait can be used to identify PwMS who have recently fallen. In contrast to fall event detection, this work aims to identify PwMS with elevated fall risk in hopes of prescribing preventative interventions. Future studies will be required to validate these methods prospectively. In addressing this objective, we explore the impact of different wearable sensor locations, model types, and feature sets on classification performance. Finally, we also examine the ability of clinically-accepted patient-reported measures and neurologist-administered assessments for detecting fallers using machine-learning based techniques.

## II. Methodology

### A. Subjects and Protocol

A sample of PwMS were recruited from the Multiple Sclerosis Center at University of Vermont Medical Center (inclusion: no major health conditions other than MS, no acute exacerbations within the previous three-months, ambulatory without the use of assistive devices). PwMS who have fallen within the previous six-months are identified as fallers, those who have not are considered non-fallers. Herein we consider data from 37 PwMS (18:19 fallers:non-fallers; 11:26 Male:Female, mean  $\pm$  standard deviation age  $51 \pm 12$  y/o; mean  $\pm$  standard deviation Expanded Disability Status Scale (EDSS)  $2.99 \pm 1.47$ ).

On the day of testing, subjects provided written informed consent to participate in the study. A neurologist with subspecialty expertise in MS completed the EDSS for each subject. Subjects were asked to complete the Fall Trips and Slips 6-month Survey, Activities-specific Balance Confidence Scale (ABC) [46], Modified Fatigue Impact Scale (MFIS) [47], Neurological Sleep Index (NSI) [48], and Twelve Item MS Walking Scale (MSWS) [49].

Subjects were instrumented with MC10 BioStamp and APDM Opal wearable sensors at various body locations as part of a larger study. In this study, we consider tri-axial

accelerometer data (31.25 Hz,  $\pm 16G$ ) sampled by BioStamp sensors secured to the sternum just below the clavicle and the anterior aspect of the right thigh as well as accelerometer and angular rate gyro data from Opal sensors secured to the lower sternum, lower back at the belt line, and anterior right and left shanks. Data from these sensors were recorded during a variety of simulated daily activities and several standard functional assessments. Here we consider data from a one-minute trial during which the subject walked at their self-selected pace in a hallway. All subjects completed these activities in the same order and controlled environment. The fall status of each patient was determined based on the Falls Trips and Slips 6-month survey. This protocol was approved by the University of Vermont's Institutional Review Board (CHRMS 18-0285).

## B. Gait Classification Pipeline

Data from this protocol were used to develop statistical classification models for identifying PwMS who have recently fallen. A high-level summary of our deep learning approach is provided in Fig. 1. Beginning with wearable sensor data from the chest and thigh, gait bouts were identified using statistical classification models and partitioned into individual strides using a simple state machine. These methods have been described previously [50], [51] and were leveraged herein to identify and analyze the 1-minute walking test completed by each subject. This method of walking extracting was employed to allow this method to be applied to un-supervised at-home monitoring in future work. Stride data from identified walking bouts were used to train and establish the performance of statistical models for classifying fall status using supervised machine learning techniques. We explore different modeling approaches for performing fall status classification including deep learning and traditional machine learning methods.

## C. Deep Learning Approaches

Deep learning methods, as applied here via a Recurrent Neural Network (RNN), do not require features to be manually calculated from sensor data prior to performing a statistical classification task, unlike traditional machine learning methods. Instead, models can be trained to take raw data as input, extract features and perform the classification. This approach has yielded performance improvements in a number of sensor-based classification tasks [36], [38], [52]. Long Short Term Memory (LSTM) networks, which are a particular type of RNN, explicitly model which aspects of a retained memory to remember and forget at each timestep by modeling dedicated gating functions for each; this behavior lends to the model's excellent ability to forecast and classify timeseries [35]. One version of this model employs two LSTM layers stacked sequentially (LSTM LSTM) such that the first extracts features and the second makes the classification [53]. An alternative to the LSTM LSTM arrangement, the Bi-Directional LSTM (BiLSTM) has two hidden states that allow the model to consider information from the past and the future, of each input, which has been shown to improve performance for some classification tasks [54].

Accelerometer data from each sensor were first calibrated to yield time series approximately aligned with the cranial-caudal, medio-lateral, and anterior-posterior directions (see [50], [51]). Herein, we train LSTM LSTM and BiLSTM models using calibrated and z-score normalized accelerometer data from the chest and thigh sampled from between one and five

sequential strides (hereafter referred to as Raw features). This processing framework is depicted in Figure 1, in which the deep learning models are represented in the ‘Determine Fall Risk’ segment. A total of 1422 stride combination observations were extracted from the 37 participants. All deep learning models were trained using the Adam optimizer, a minibatch size of 200 and with a maximum of 100 epochs. The LSTM LSTM models used the following architecture listed in order of layers: sequence input, LSTM layer with 215 hidden units, 30% dropout layer, LSTM layer with 125 hidden units, 40% dropout layer, fully connected layer with output size of 2, softmax layer, and classification layer. The BiLSTM models used the following architecture listed in order of layers: sequence input, BiLSTM layer with 215 hidden units, 40% dropout layer, fully connected layer of output size 2, softmax layer, and classification layer. All deep learning models were developed using MATLAB 2019a Deep Network Designer.

To explore the effect of sensor placement, models were also trained and evaluated using data from the chest and thigh sensors, individually.

#### D. Traditional Machine Learning Approaches

In order to compare to the deep learning methods, traditional machine learning methods were employed with several feature sets outlined below. Unlike deep learning methods, where raw data can serve as the input, traditional machine learning methods utilize metrics extracted from signals and other values to inform the statistical classification models. The remainder of this section will outline the stride, spatiotemporal, and patient-reported metric feature sets. Details on model training and feature selection are provided in Section II E.

Wearable sensor data from each stride identified for each subject were used to compute a single vector of time and frequency domain features (hereafter referred to as Stride features). This method varies from the framework in Figure 1 only after strides were extracted from the identified walking bouts. Accelerometer data from each sensor were first calibrated and following time and frequency domain features were computed from each of the resulting six time series: mean, standard deviation (std), root-mean-square (rms), maximum, minimum, distance between peaks, rms between peaks, skewness, kurtosis, sample entropy, power between 0.25 and 0.5 Hz, power between 0.5 and 1 Hz, power spectral densities, cross correlation of thigh and chest, ratio of thigh and chest rms, ratio of thigh and chest mean raw acceleration, and sample entropy. Additionally, the time derivative of each accelerometer time series was taken yielding a measure of jerk used to inform calculation of the following time and frequency domain features: mean, maximum, minimum, std, rms, power between 0.25 and 0.5 Hz, power between 0.5 and 1 Hz, and distance between peaks. Finally, stride time and duty cycle for the given stride were also included yielding a 403-element feature vector for each stride and 1422 observations across the 37 participants.

We also explored models trained on features developed by extracting spatiotemporal measures of gait, as these have been associated with fall risk [55]. These measures are reported by the Opal system, which has previously been validated in PwMS [56], during the 1-minute walking test completed by each subject (hereafter referred to as Spatiotemporal features). Notably, these features do not follow the processing methodology outlined in Fig. 1. The Spatiotemporal features were the mean, maximum, minimum, and standard deviation

of gait speed, cadence, double support time, elevation at mid-swing, gait cycle duration, lateral step deviation, lateral step variability, lateral swing max acceleration, maximum pitch, pitch at initial contact, pitch at mid-swing, pitch at toe off, single limb support time, stance time, step duration, stride length, swing time, mean toe out angle, maximum toe out angle, and minimum toe out angle. This yields one 81-element feature vector for each of the 37 participants.

To provide context for results from these sensing approaches, we also explored statistical models trained on a standard neurologist-administered assessment and patient reported measures (PRMs) in this sample. These models provide a baseline for the best results that could reasonably be expected from today's standard of care. Specifically, models were developed based on neurologist-administered EDSS (four EDSS features: EDSS Total, EDSS Sensory, EDSS Pyramidal, EDSS Cerebellar) and patient reported MFIS (4 features: MFIS Physical, MFIS Cognitive, MFIS Psychosocial, MFIS Total), ABC (one feature ABC Total), MSWS (one feature: MSWS Total), and NSI (one feature: NSI Total) (hereafter referred to as PRM features). The preceding PRMs were selected due to their relation to fall risk [9], [14], [57]. We explored the following combinations of possible model inputs: EDSS only, PRM only, and PRM with EDSS. These features were observed from each of the 37 participants.

## E. Model Performance Assessment and Analysis

Classifier performance was established using leave-one-subject-out cross validation (LOSO-CV). In this approach, data from all but one participant ( $N = 36$ ) were partitioned into a training dataset while data from the remaining subject was used for testing. This process was repeated until data from each subject had been included in the test set. The LOSO-CV approach ensures the model was tested on subjects it has not previously seen, which provides a realistic estimate of how the model would perform during real-world use. The normalized posterior probability assigned to each observation were combined to calculate an overall model performance by considering the area under the receiver operating characteristic curve (AUC). Additionally, accuracy, sensitivity, specificity, and F1-score were calculated. Where appropriate, models were assessed based on their individual input performance, referring to 1–5 strides, as well as their aggregate performance after taking the median decision score of the individual inputs across the entire one-minute walking test for each participant.

Models were trained on each of the feature sets described above (Stride, Spatiotemporal, Raw, EDSS, PRM). Features were normalized using z-scores then reduced using Davies-Bouldin index [58] based feature selection follow by principal components analysis (PCA) within each iteration of the LOSO-CV. The Stride features were reduced on average from 403 to 17 features using a DB threshold of 3.5. A DB threshold of 6 was used to reduce the Spatiotemporal features from 80 to an average of 12.67 features per evaluation. A DB threshold of 1.5 was used for both EDSS only and PRMS with EDSS, reducing on average from 4 to 2 features and from 12 to 3.5 features, respectively. The PRMs only used a DB threshold of 2.5 to reduce from 8 features to 4 on average. The principal components that explained 95% of the variance of these reduced feature sets were then extracted and used to



train Logistic Regression (LR) [59], Support Vector Machine (SVM) [60], Decision Tree [61], K-Nearest Neighbors (KNN) [62], and Ensemble (ENS) [61] binary statistical classification models to discriminate fallers from non-fallers. Model hyperparameters were optimized for each input feature set to provide the highest classification performance, this information has been provided in Table IV in the appendix. Deep learning models, the LSTM LSTM and BiLSTM, were trained to discriminate fallers from non-fallers based on data from the Raw feature set.

### III. Results

The BiLSTM deep learning model trained on four sequential strides with thigh and chest acceleration provided the best performance in discriminating PwMS who had previously fallen from those who had not, achieving an AUC of 0.80, accuracy of 73%, specificity of 76%, sensitivity of 69%, and F1-score of 0.70 (see Table I). The discriminative ability of this model improved considerably when considering the median decision score reported across the entire one-minute walking test for each subject, to an AUC of 0.88, accuracy of 86%, specificity of 83%, sensitivity of 88%, and F1-score of 0.86 (see Fig. 2). The 4-stride BiLSTM using aggregation provided the best AUC, accuracy, sensitivity, and F1-score.

Deep learning BiLSTM models trained on data from the chest and thigh (AUC of 0.88) outperformed models trained on data from the thigh (AUC of 0.55) or chest (AUC of 0.71) alone (see Fig. 3). For models trained on data from the chest and thigh, performance increased with each additional stride considered up until four strides (AUC 0.65 to 0.88) at which point it began to decrease (AUC 0.88 to 0.79, see Table I). This may indicate that additional data are needed to accommodate the variance added by considering five strides. Interestingly, the LSTM LSTM models performed better when considering one and two strides as input while the BiLSTM performed better for between three and five strides (see Table I). This difference may be related to the LSTM LSTM model providing the necessary power to learn the variance for a fewer number of strides, however a lack of data likely prevented this method from effectively learning the variance seen in a greater number of consecutive strides. The performance of the models generally increases with the number of strides used as input. We suspect that this is because considering data from more strides provides a better measure of gait variability which has been shown to be indicative of fall risk in PwMS [23]. In all cases, classification performance improved when the median decision score from the entire one-minute walking test was used for the classification (AUC increase 4–10%, Table I).

These results compare favorably to models trained on the Stride, Spatiotemporal, EDSS, and PRM feature sets (see Table II). The best performing feature-based models are listed in Table II, see the Appendix for details on the performance of all models tested. Models trained on the Stride feature set showed the lowest performance, when considering AUC, with an AUC of 0.69. However, this feature set demonstrated significant improvement when considering the median decision score across the entire one-minute walking test with an AUC of 0.76. Models trained on the EDSS feature set showed the next lowest performance when considering AUC and lowest F1-score, achieving an AUC of 0.71 and F1-score of 0.46. This reflects previous work demonstrating the poor sensitivity of the EDSS for

capturing fall risk [63]. The model trained on the Spatiotemporal feature set achieved the highest specificity of 89%, however, this model was less impressive in other metrics. Models trained on the PRM feature set provided reasonable performance, however they did not stand out in any performance metric. The best feature-based performance, when considering AUC, was achieved by models trained on a combination of the EDSS and PRM feature sets with an AUC of 0.79. As demonstrated by the results in Table II, it is clear that models trained on the Stride, Spatiotemporal, EDSS, and PRM features sets have lower performance across all metrics when compared to the deep learning-based approach (Table I).

Analysis of the features selected as part of the feature reduction process can provide an indication of the measures most important for identifying PwMS at elevated fall risk. For example the reduced PRM and EDSS feature set typically contained EDSS Pyramidal, EDSS Cerebellar, ABC, and MSWS, removing the MFIS and NSI PRMs. Similarly, for the EDSS feature set, the features used for classification were EDSS Cerebellar and EDSS Pyramidal. For the PRM only feature set, all PRM features were selected except for NSI. For the spatiotemporal feature set, a large number of features were selected in each LOSO fold prior to PCA (44 on average), however there were a consistent top five features observed. These were minimum elevation at mid swing, maximum elevation at mid swing, mean toe-out angle, minimum toe-out angle standard deviation, and minimum toe out angle mean.

#### IV. Discussion

In this paper, we demonstrated the identification of PwMS who have recently fallen based on only one minute of walking through deep neural network-analysis of accelerometer data. We show that this approach improves classification performance relative to more traditional machine learning approaches that consider wearable sensor data (both time and frequency domain features as well as spatiotemporal gait parameters), neurologist-administered assessments, and/or patient reported measures. We further discuss these results, identify potential reasons why these particular deep learning architectures may be improving performance, place the results in the context of the literature, and discuss their implications for preventing falls in PwMS.

The deep learning approaches used in this study have shown promise in other biomedical sensor-based classification and prediction tasks [20], [37]–[44], [52]. In line with a recent review [64], we chose these approaches here so that the memory of the network and time series input could capture gait variability, which has been shown to indicate fall risk in PwMS [23]. This approach yielded a substantial improvement in model performance (increase in F1-score of more than 28%), relative to the best performing traditional wearable sensor and machine learning based approach (F1-score of 0.86 vs. 0.67, see Tables I and II).

Sensitivity is a key parameter to consider in developing risk screening technologies as it is a direct measure of how many people with elevated risk the instrument is able to correctly detect. Our proposed deep learning approach improves the sensitivity for detecting PwMS with elevated fall risk by 44% relative to neurologist and patient report measures (88% for 4-stride, BiLSTM, median vs. 61% for PRM + EDSS, see Tables I and II).



The BiLSTM methods used in this study provide higher levels of performance than other studies classifying fall risk in PwMS [18] as well as many studies performed on healthy older adults [21], [65], [66], albeit with a smaller sample size in some cases. A recent study in neurological disorders (other than MS) was able to achieve a slightly higher AUC of 0.94 (vs. 0.88 here) for a similar task of classifying fall risk from a larger dataset ( $n=76$ ) using similar LSTM-based methods [20]. Studies conducted in healthy older adults have achieved AUC for models trained on walking data of 0.73–0.79 [30]. While it is difficult to compare results across studies, the relative performance of the methods described herein indicate the promise of this approach for identifying PwMS who have recently fallen and encourage future work to explore the use of this method for identifying fallers prospectively in this population in both lab and home environments.

The proposed deep learning approach for detecting PwMS who have fallen provides results faster than traditional PRMs or clinical assessments. For comparison, the one-minute walking test and associated BiLSTM model used for making a fall risk classification could be deployed in under 90 seconds, while the PRMs needed for the PRM + EDSS trained model take about 25 minutes to complete, and the EDSS requires 15 minutes of a certified neurologist's time to administer. Given the typical 30-minute neurological visit, these times suggest that this wearable sensor approach may provide a more feasible solution for objectively quantifying fall risk in PwMS in the clinical environment than existing clinical assessments or PRMs.

As seen in Fig. 3, when studying the effect of sensor location, the thigh and chest together provided greater performance than data from either sensor individually. This suggests that the relationship between the sensors is being leveraged to better classify fallers. Interestingly, the relative motion of the torso and legs directly impacts dynamic margin of stability which has previously been related to fall risk in PwMS [22]. These results point to the need for further refining the sensor locations which provide the best data for identifying PwMS who have fallen, but these results suggest that locations on both the upper and lower body may be critical.

Due to the minimal sensor configuration, level of performance achieved, and short period of time needed to generate a fall risk assessment, the 4-stride BiLSTM shows promise to serve as a powerful tool for classifying the fall risk of PwMS in a clinical setting. Further studies are needed to assess the ability of this approach to identify fallers prospectively [28], and the performance of this model in the home environment to test the viability of this approach outside of the clinic for prolonged fall risk monitoring.

## V. Conclusion

In this study, we analyzed a variety of machine learning models and feature sets for classifying the fall status of PwMS. Specifically, we assessed the impact of sensor location, number of strides, and aggregation of decision scores on the performance of LSTM LSTM, and BiLSTM deep learning models. For comparison, we assessed the performance of traditional machine learning models using feature sets manually calculated from wearable accelerometer data, derived from spatiotemporal gait parameters, and extracted from patient

reported measures and clinical assessments. The BiLSTM deep learning model was found to provide the highest performance with AUC of 0.88. These results support the use of deep learning with gait acceleration data for classifying the fall status of PwMS. While future studies are required to assess the performance of these methods for classifying fall risk in the home environment and predicting risk for future falls in this population, these promising results indicate that this deep learning approach is worthy of future consideration.

## Acknowledgments

This work was supported in part by NIH under Grant R21EB027852.

## Appendix

**TABLE III**

Machine learning model performance for various feature sets of all Models

Feature Set	Model	ACC	SPE	SEN	AUC	F1
Spatiotemporal	ENS	0.54	0.53	0.56	0.47	0.54
	LR	0.59	0.74	0.44	0.65	0.52
	KNN	0.57	0.53	<b>0.61</b>	0.54	0.58
	SVM	<b>0.73</b>	<b>0.89</b>	0.55	<b>0.73</b>	<b>0.67</b>
	Tree	0.38	0.32	0.44	0.44	0.41
Stride	ENS	0.61	0.74	0.44	0.65	<b>0.59</b>
	LR	<b>0.65</b>	0.73	<b>0.56</b>	<b>0.69</b>	<b>0.59</b>
	KNN	0.57	0.69	0.42	0.60	0.47
	SVM	0.62	<b>0.84</b>	0.36	<b>0.69</b>	0.46
	Tree	0.57	0.66	0.45	0.60	0.48
Aggregated Stride	ENS	0.57	0.78	0.35	0.75	0.44
	LR	<b>0.66</b>	0.78	<b>0.53</b>	<b>0.76</b>	<b>0.60</b>
	KNN	0.63	0.83	0.41	0.71	0.52
	SVM	0.62	<b>0.84</b>	0.36	0.69	0.46
	Tree	0.63	0.72	0.53	0.65	0.58
EDSS	ENS	0.65	0.74	0.56	0.54	0.61
	LR	<b>0.70</b>	0.74	<b>0.67</b>	0.63	<b>0.69</b>
	KNN	0.65	0.74	0.56	0.64	0.61
	SVM	0.62	<b>0.89</b>	0.33	<b>0.71</b>	0.46
	Tree	0.54	0.53	0.56	0.47	0.54
PRM	ENS	0.65	0.63	0.67	0.64	0.65
	LR	0.70	0.79	0.61	0.65	0.67
	KNN	0.73	0.68	0.78	0.72	0.74
	SVM	0.70	<b>0.84</b>	0.56	<b>0.74</b>	0.65
	Tree	<b>0.78</b>	0.68	<b>0.89</b>	0.60	<b>0.80</b>
PRM + EDSS	ENS	<b>0.73</b>	<b>0.84</b>	0.61	0.74	0.69
	LR	0.70	0.79	0.61	<b>0.79</b>	0.67
	KNN	0.68	0.58	<b>0.78</b>	0.69	<b>0.70</b>

Feature Set	Model	ACC	SPE	SEN	AUC	F1
	SVM	0.65	<b>0.84</b>	0.44	0.77	0.55
	Tree	0.68	0.58	<b>0.78</b>	0.69	<b>0.70</b>

ENS: Ensemble; LR: Logistic Regression; KNN: K-Nearest Neighbors; SVM: Support Vector Machine; Tree: Decision Tree; AUC: Area Under the Receiver Operating Characteristic Curve; ACC: Accuracy; SPE: Specificity; SEN: Sensitivity; F1: F1-Score.

**TABLE IV**

Optimization parameters for various feature sets of all Models

Feature Set	Model	Optimization Parameters
Spatiotemporal	ENS	Method: LogitBoost
	LR	Learner: Logistic, Lambda: 0.01169
	KNN	Num Neighbors: 3
	SVM	Solver: SMO, Kernel Function: Linear, Box Constraint: 41.478, Kernel Scale: 21.404
	Tree	Min Leaf Size: 19
Stride	ENS	Method: Bag
	LR	Learner: Logistic, Lambda: 0.01
	KNN	Distance: cityblock, Num Neighbors: 7
	SVM	Solver: SMO, Kernel Function: Linear, Box Constraint: 22.287, Kernel Scale: 23.763
	Tree	Min Leaf Size: 27
EDSS	ENS	Method: Bag
	LR	Learner: Logistic, Lambda: 0.0015
	KNN	Distance: chebychev, Num Neighbors: 20
	SVM	Solver: SMO, Kernel Function: Linear, Box Constraint: 575.38, Kernel Scale: 76.8350
	Tree	Min Leaf Size: 10
PRM	ENS	Method: Bag
	LR	Learner: Logistic, Lambda: 0.010281
	KNN	Distance: chebychev, Num Neighbors: 20
	SVM	Solver: SMO, Kernel Function: Linear, Box Constraint: 5.737, Kernel Scale: 10.624
	Tree	Min Leaf Size: 5
PRM + EDSS	ENS	Method: Bag
	LR	Learner: Logistic, Lambda: 0.0103
	KNN	Distance: chebychev, Num Neighbors: 20
	SVM	Solver: SMO, Kernel Function: Linear, Box Constraint: 5.737, Kernel Scale: 10.624
	Tree	Min Leaf Size: 5

ENS: Ensemble; LR: Logistic Regression; KNN: K-Nearest Neighbors; SVM: Support Vector Machine; Tree: Decision Tree; SMO: Sequential Minimal Optimization

## References

[1]. Coote S, Sosnoff JJ, and Gunn H, "Fall Incidence as the Primary Outcome in Multiple Sclerosis Falls-Prevention Trials," *Int J MS Care*, vol. 16, no. 4, pp. 178–184, 2014, doi: 10.7224/1537-2073.2014-059. [PubMed: 25694776]

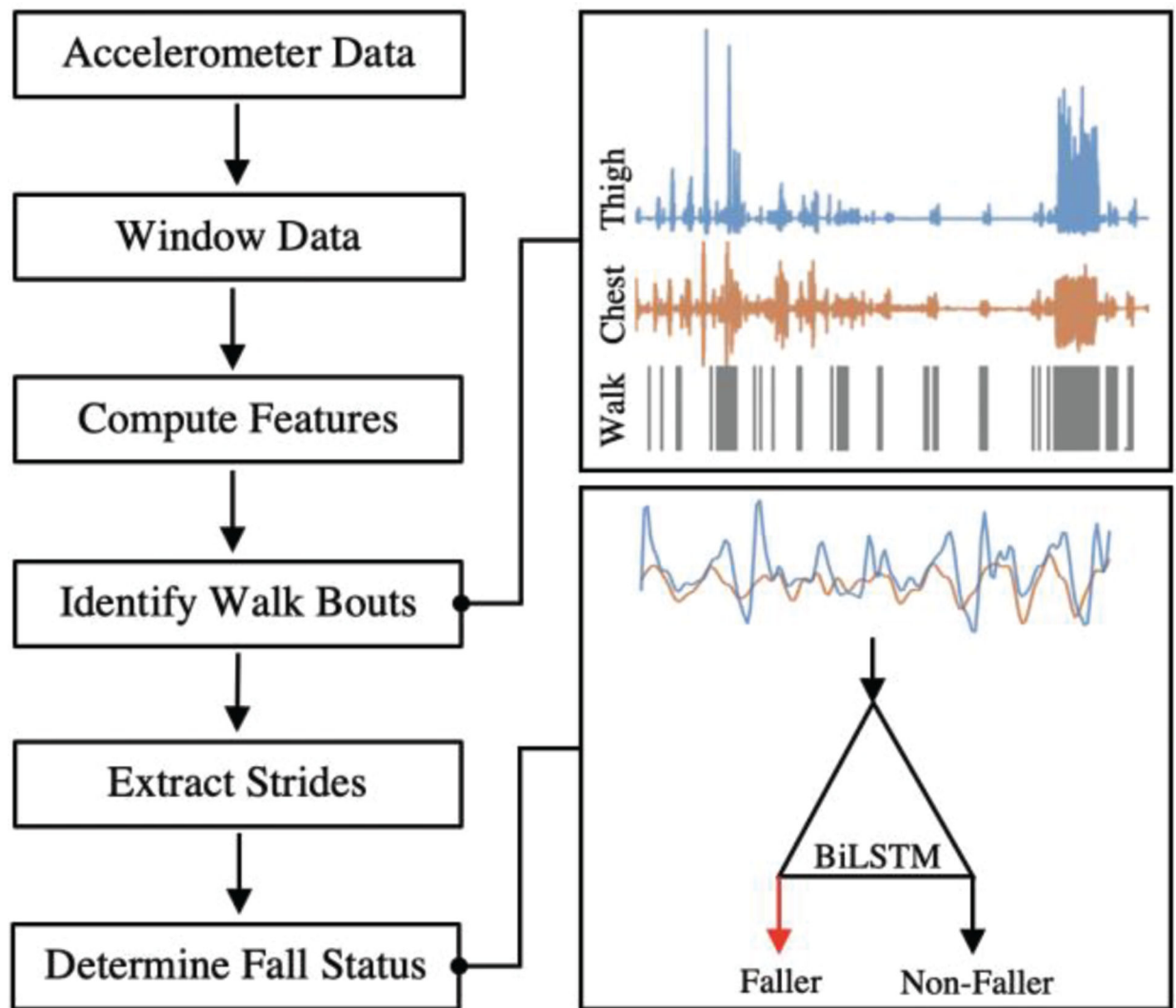
- [2]. Peterson EW, Cho CC, von Koch L, and Finlayson ML, “Injurious Falls Among Middle Aged and Older Adults With Multiple Sclerosis,” *Archives of Physical Medicine and Rehabilitation*, vol. 89, no. 6, pp. 1031–1037, 6. 2008, doi: 10.1016/j.apmr.2007.10.043. [PubMed: 18503796]
- [3]. Berg K, Wood-Dauphine S, Williams JI, and Gayton D, “Measuring balance in the elderly: preliminary development of an instrument,” *Physiotherapy Canada*, 4. 2009, doi: 10.3138/ptc.41.6.304.
- [4]. Podsiadlo D and Richardson S, “The timed ‘Up & Go’: a test of basic functional mobility for frail elderly persons,” *J Am Geriatr Soc*, vol. 39, no. 2, pp. 142–148, 2. 1991. [PubMed: 1991946]
- [5]. Nilsagård Y, Lundholm Cecilia, Denison E, and Gunnarsson L-G, “Predicting accidental falls in people with multiple sclerosis — a longitudinal study,” *Clin Rehabil*, vol. 23, no. 3, pp. 259–269, 3. 2009, doi: 10.1177/0269215508095087. [PubMed: 19218300]
- [6]. Kasser SL, Jacobs JV, Ford M, and Tourville TW, “Effects of balance-specific exercises on balance, physical activity and quality of life in adults with multiple sclerosis: a pilot investigation,” *Disability and Rehabilitation*, vol. 37, no. 24, pp. 2238–2249, 11. 2015, doi: 10.3109/09638288.2015.1019008. [PubMed: 25738911]
- [7]. Cattaneo D, De Nuzzo C, Fascia T, Macalli M, Pisoni I, and Cardini R, “Risks of falls in subjects with multiple sclerosis,” *Archives of Physical Medicine and Rehabilitation*, vol. 83, no. 6, pp. 864–867, 6. 2002, doi: 10.1053/apmr.2002.32825. [PubMed: 12048669]
- [8]. Mazumder R, Murchison C, Bourdette D, and Cameron M, “Falls in People with Multiple Sclerosis Compared with Falls in Healthy Controls,” *PLoS One*, vol. 9, no. 9, 9. 2014, doi: 10.1371/journal.pone.0107620.
- [9]. Cameron MH, Thielman E, Mazumder R, and Bourdette D, “Predicting falls in people with multiple sclerosis: fall history is as accurate as more complex measures,” *Multiple Sclerosis International*, 2013, Accessed: Jan. 02, 2020. [Online]. Available: [https://link.gale.com/apps/doc/A377777201/HWRC?u=vol\\_b92b&sid=HWRC&xid=552f58cd](https://link.gale.com/apps/doc/A377777201/HWRC?u=vol_b92b&sid=HWRC&xid=552f58cd).
- [10]. Matsuda PN, Shumway-Cook A, Bamer AM, Johnson SL, Amtmann D, and Kraft GH, “Falls in multiple sclerosis,” *PM R*, vol. 3, no. 7, pp. 624–632; quiz 632, 7. 2011, doi: 10.1016/j.pmrj.2011.04.015. [PubMed: 21777861]
- [11]. Kasser SL, Goldstein A, Wood PK, and Sibold J, “Symptom variability, affect and physical activity in ambulatory persons with multiple sclerosis: Understanding patterns and time-bound relationships,” *Disability and Health Journal*, vol. 10, no. 2, pp. 207–213, 4. 2017, doi: 10.1016/j.dhjo.2016.10.006. [PubMed: 27814947]
- [12]. Powell DJH, Liossi C, Schlotz W, and Moss-Morris R, “Tracking daily fatigue fluctuations in multiple sclerosis: ecological momentary assessment provides unique insights,” *J Behav Med*, vol. 40, no. 5, pp. 772–783, 10. 2017, doi: 10.1007/s10865-017-9840-4. [PubMed: 28281106]
- [13]. Socie MJ, Sandroff BM, Pula JH, Hsiao-Weckler ET, Motl RW, and Sosnoff JJ, “Footfall Placement Variability and Falls in Multiple Sclerosis,” *Ann Biomed Eng*, vol. 41, no. 8, pp. 1740–1747, 8. 2013, doi: 10.1007/s10439-012-0685-2. [PubMed: 23132152]
- [14]. Tajali S et al., “Predicting falls among patients with multiple sclerosis: Comparison of patient-reported outcomes and performance-based measures of lower extremity functions,” *Multiple Sclerosis and Related Disorders*, vol. 17, pp. 69–74, 10. 2017, doi: 10.1016/j.msard.2017.06.014. [PubMed: 29055478]
- [15]. Quinn G, Comber L, Galvin R, and Coote S, “The ability of clinical balance measures to identify falls risk in multiple sclerosis: a systematic review and meta-analysis,” *Clin Rehabil*, vol. 32, no. 5, pp. 571–582, 5 2018, doi: 10.1177/0269215517748714. [PubMed: 29260583]
- [16]. Howcroft J, Kofman J, and Lemaire ED, “Review of fall risk assessment in geriatric populations using inertial sensors,” *Journal of NeuroEngineering and Rehabilitation*, vol. 10, no. 1, p. 91, 8. 2013, doi: 10.1186/1743-0003-10-91. [PubMed: 23927446]
- [17]. Sebastião E, Learmonth YC, and Motl RW, “Mobility measures differentiate falls risk status in persons with multiple sclerosis: An exploratory study,” *NeuroRehabilitation*, vol. 40, no. 1, pp. 153–161, 2017, doi: 10.3233/NRE-161401. [PubMed: 27935560]
- [18]. Mañago MM, Cameron M, and Schenkman M, “Association of the Dynamic Gait Index to fall history and muscle function in people with multiple sclerosis,” *Disabil Rehabil*, pp. 1–6, 5 2019, doi: 10.1080/09638288.2019.1607912.

- [19]. Qiu H, Rehman RZU, Yu X, and Xiong S, “Application of Wearable Inertial Sensors and A New Test Battery for Distinguishing Retrospective Fallers from Non-fallers among Community-dwelling Older People,” *Scientific Reports*, vol. 8, no. 1, Art. no. 1, 11. 2018, doi: 10.1038/s41598-018-34671-6.
- [20]. Tunca C, Salur G, and Ersoy C, “Deep Learning for Fall Risk Assessment With Inertial Sensors: Utilizing Domain Knowledge in Spatio-Temporal Gait Parameters,” *IEEE Journal of Biomedical and Health Informatics*, vol. 24, no. 7, pp. 1994–2005, 7. 2020, doi: 10.1109/JBHI.2019.2958879. [PubMed: 31831454]
- [21]. Howcroft J, Lemaire ED, and Kofman J, “Wearable-Sensor-Based Classification Models of Faller Status in Older Adults,” *PLOS ONE*, vol. 11, no. 4, p. e0153240, 4. 2016, doi: 10.1371/journal.pone.0153240. [PubMed: 27054878]
- [22]. Peebles AT, Bruetsch AP, Lynch SG, and Huisinga JM, “Dynamic balance in persons with multiple sclerosis who have a falls history is altered compared to non-fallers and to healthy controls,” *Journal of Biomechanics*, vol. 0, no. 0, 8. 2017, doi: 10.1016/j.jbiomech.2017.08.023.
- [23]. Moon Y, Wajda DA, Motl RW, and Sosnoff JJ, “Stride-Time Variability and Fall Risk in Persons with Multiple Sclerosis,” *Mult Scler Int*, vol. 2015, 2015, doi: 10.1155/2015/964790.
- [24]. Sosnoff JJ et al., “Mobility, Balance and Falls in Persons with Multiple Sclerosis,” *PLOS ONE*, vol. 6, no. 11, p. e28021, 11. 2011, doi: 10.1371/journal.pone.0028021. [PubMed: 22132196]
- [25]. McGinnis RS et al., “A machine learning approach for gait speed estimation using skin-mounted wearable sensors: From healthy controls to individuals with multiple sclerosis,” *PLoS One*, vol. 12, no. 6, 6. 2017, doi: 10.1371/journal.pone.0178366.
- [26]. Filli L et al., “Profiling walking dysfunction in multiple sclerosis: characterisation, classification and progression over time,” *Scientific Reports*, vol. 8, no. 1, p. 4984, 3. 2018, doi: 10.1038/s41598-018-22676-0. [PubMed: 29563533]
- [27]. Moon Y et al., “Monitoring gait in multiple sclerosis with novel wearable motion sensors,” *PLOS ONE*, vol. 12, no. 2, p. e0171346, 2. 2017, doi: 10.1371/journal.pone.0171346. [PubMed: 28178288]
- [28]. Frechette ML, Meyer BM, Tulipani LJ, Gurchiek RD, McGinnis RS, and Sosnoff JJ, “Next Steps in Wearable Technology and Community Ambulation in Multiple Sclerosis,” *Curr Neurol Neurosci Rep*, vol. 19, no. 10, p. 80, 9. 2019, doi: 10.1007/s11910-019-0997-9. [PubMed: 31485896]
- [29]. Sparaco M, Lavorgna L, Conforti R, Tedeschi G, and Bonavita S, “The Role of Wearable Devices in Multiple Sclerosis,” *Multiple Sclerosis International*, vol. 2018, pp. 1–7, 10. 2018, doi: 10.1155/2018/7627643.
- [30]. Bet P, Castro PC, and Ponti MA, “Fall detection and fall risk assessment in older person using wearable sensors: A systematic review,” *International Journal of Medical Informatics*, vol. 130, p. 103946, 10. 2019, doi: 10.1016/j.ijmedinf.2019.08.006. [PubMed: 31450081]
- [31]. Gietzelt M, Feldwieser F, Gövercin M, Steinhagen-Thiessen E, and Marschollek M, “A prospective field study for sensor-based identification of fall risk in older people with dementia,” *Informatics for Health and Social Care*, vol. 39, no. 3–4, pp. 249–261, 9. 2014, doi: 10.3109/17538157.2014.931851. [PubMed: 25148560]
- [32]. Huisinga JM, Mancini M, George R. J. St., and Horak FB, “Accelerometry Reveals Differences in Gait Variability Between Patients with Multiple Sclerosis and Healthy Controls,” *Ann Biomed Eng*, vol. 41, no. 8, pp. 1670–1679, 8. 2013, doi: 10.1007/s10439-012-0697-y. [PubMed: 23161166]
- [33]. Huisinga JM, Schmid KK, Filipi ML, and Stergiou N, “Gait mechanics are different between healthy controls and patients with multiple sclerosis,” *J Appl Biomech*, vol. 29, no. 3, pp. 303–311, 6. 2013, doi: 10.1123/jab.29.3.303. [PubMed: 22923390]
- [34]. Yu D and Deng L, “Deep Learning and Its Applications to Signal and Information Processing [Exploratory DSP,” *IEEE Signal Process. Mag*, vol. 28, no. 1, pp. 145–154, 1. 2011, doi: 10.1109/MSP.2010.939038.
- [35]. Hochreiter S and Schmidhuber J, “Long short-term memory,” *Neural Computation*, vol. 9, no. 8, p. 1735, 11. 1997. [PubMed: 9377276]

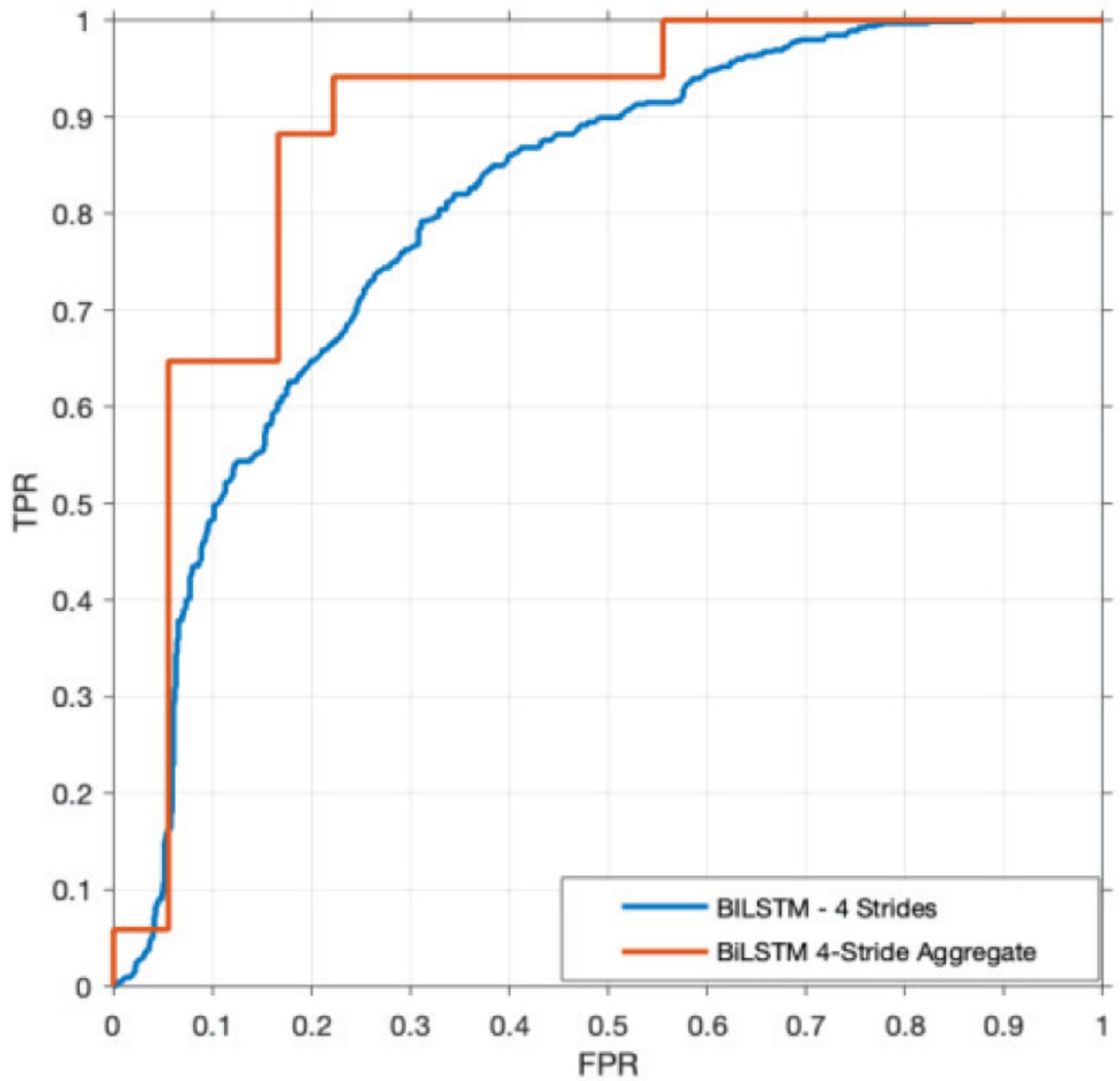
- [36]. Tan HX, Aung NN, Tian J, Chua MCH, and Yang YO, “Time series classification using a modified LSTM approach from accelerometer-based data: A comparative study for gait cycle detection,” *Gait & Posture*, vol. 74, pp. 128–134, 10. 2019, doi: 10.1016/j.gaitpost.2019.09.007. [PubMed: 31518859]
- [37]. Giansanti D, Macellari V, and Maccioni G, “New neural network classifier of fall-risk based on the Mahalanobis distance and kinematic parameters assessed by a wearable device,” *Physiological Measurement*, vol. 29, no. 3, pp. N11–N19, 3. 2008, doi: 10.1088/0967-3334/29/3/N01. [PubMed: 18367804]
- [38]. Nait Aicha A, Englebienne G, Van Schooten KS, Pijnappels M, and Kröse B, “Deep Learning to Predict Falls in Older Adults Based on Daily-Life Trunk Accelerometry,” *Sensors*, vol. 18, no. 5, p. 1654, 5 2018, doi: 10.3390/s18051654.
- [39]. Torti E et al., “Embedded Real-Time Fall Detection with Deep Learning on Wearable Devices,” in 2018 21st Euromicro Conference on Digital System Design (DSD), 8. 2018, pp. 405–412, doi: 10.1109/DSD.2018.00075.
- [40]. Wayan Wiprayoga Wisesa I and Mahardika G, “Fall detection algorithm based on accelerometer and gyroscope sensor data using Recurrent Neural Networks,” *IOP Conference Series: Earth and Environmental Science*, vol. 258, p. 012035, 5 2019, doi: 10.1088/1755-1315/258/1/012035.
- [41]. Musci M, Martini DD, Blago N, Facchinetti T, and Piastra M, “Fall Detection using Recurrent Neural Networks,” p. 7.
- [42]. Luna-Perejón F, Domínguez-Morales MJ, and Civit-Balcells A, “Wearable Fall Detector Using Recurrent Neural Networks,” *Sensors*, vol. 19, no. 22, Art. no. 22, 1. 2019, doi: 10.3390/s19224885.
- [43]. Luna-Perejón F et al., “An Automated Fall Detection System Using Recurrent Neural Networks,” in *Artificial Intelligence in Medicine*, Cham, 2019, pp. 36–41, doi: 10.1007/978-3-030-21642-9\_6.
- [44]. Yu X, Qiu H, and Xiong S, “A Novel Hybrid Deep Neural Network to Predict Pre-impact Fall for Older People Based on Wearable Inertial Sensors,” *Front Bioeng Biotechnol*, vol. 8, 2. 2020, doi: 10.3389/fbioe.2020.00063.
- [45]. Sun R, Hsieh KL, and Sosnoff JJ, “Fall Risk Prediction in Multiple Sclerosis Using Postural Sway Measures: A Machine Learning Approach,” *Sci Rep*, vol. 9, no. 1, p. 16154, 12. 2019, doi: 10.1038/s41598-019-52697-2. [PubMed: 31695127]
- [46]. Powell LE and Myers AM, “The Activities-specific Balance Confidence (ABC) Scale,” *J Gerontol A Biol Sci Med Sci*, vol. 50A, no. 1, pp. M28–M34, 1. 1995, doi: 10.1093/gerona/50A.1.M28. [PubMed: 7814786]
- [47]. “Modified Fatigue Impact Scale,” Shirley Ryan AbilityLab. <https://www.sralab.org/rehabilitation-measures/modified-fatigue-impact-scale> (accessed Jun. 16, 2020).
- [48]. Mills R, Tennant A, and Young C, “The Neurological Sleep Index: A suite of new sleep scales for multiple sclerosis,” *Mult Scler J Exp Transl Clin*, vol. 2, 4. 2016, doi: 10.1177/2055217316642263.
- [49]. Hobart JC, Riazi A, Lamping DL, Fitzpatrick R, and Thompson AJ, “Measuring the impact of MS on walking ability: The 12-Item MS Walking Scale (MSWS-12),” *Neurology*, vol. 60, no. 1, pp. 31–36, 1. 2003, doi: 10.1212/WNL.60.1.31. [PubMed: 12525714]
- [50]. Gurchiek RD et al., “Remote Gait Analysis Using Wearable Sensors Detects Asymmetric Gait Patterns in Patients Recovering from ACL Reconstruction,” presented at the 2019 IEEE International Conference on Body Sensor Networks (BSN), Chicago, IL, 5 2019.
- [51]. Gurchiek RD et al., “Open-Source Remote Gait Analysis: A Post-Surgery Patient Monitoring Application,” *Sci Rep*, vol. 9, no. 1, pp. 1–10, 11. 2019, doi: 10.1038/s41598-019-54399-1. [PubMed: 30626917]
- [52]. Pourbabae B, Roshtkhari MJ, and Khorasani K, “Deep Convolutional Neural Networks and Learning ECG Features for Screening Paroxysmal Atrial Fibrillation Patients,” *IEEE Transactions on Systems, Man, and Cybernetics: Systems*, vol. 48, no. 12, pp. 2095–2104, 12. 2018, doi: 10.1109/TSMC.2017.2705582.
- [53]. Sak H, Senior A, and Beaufays F, “Long Short-Term Memory Recurrent Neural Network Architectures for Large Scale Acoustic Modeling,” p. 5.



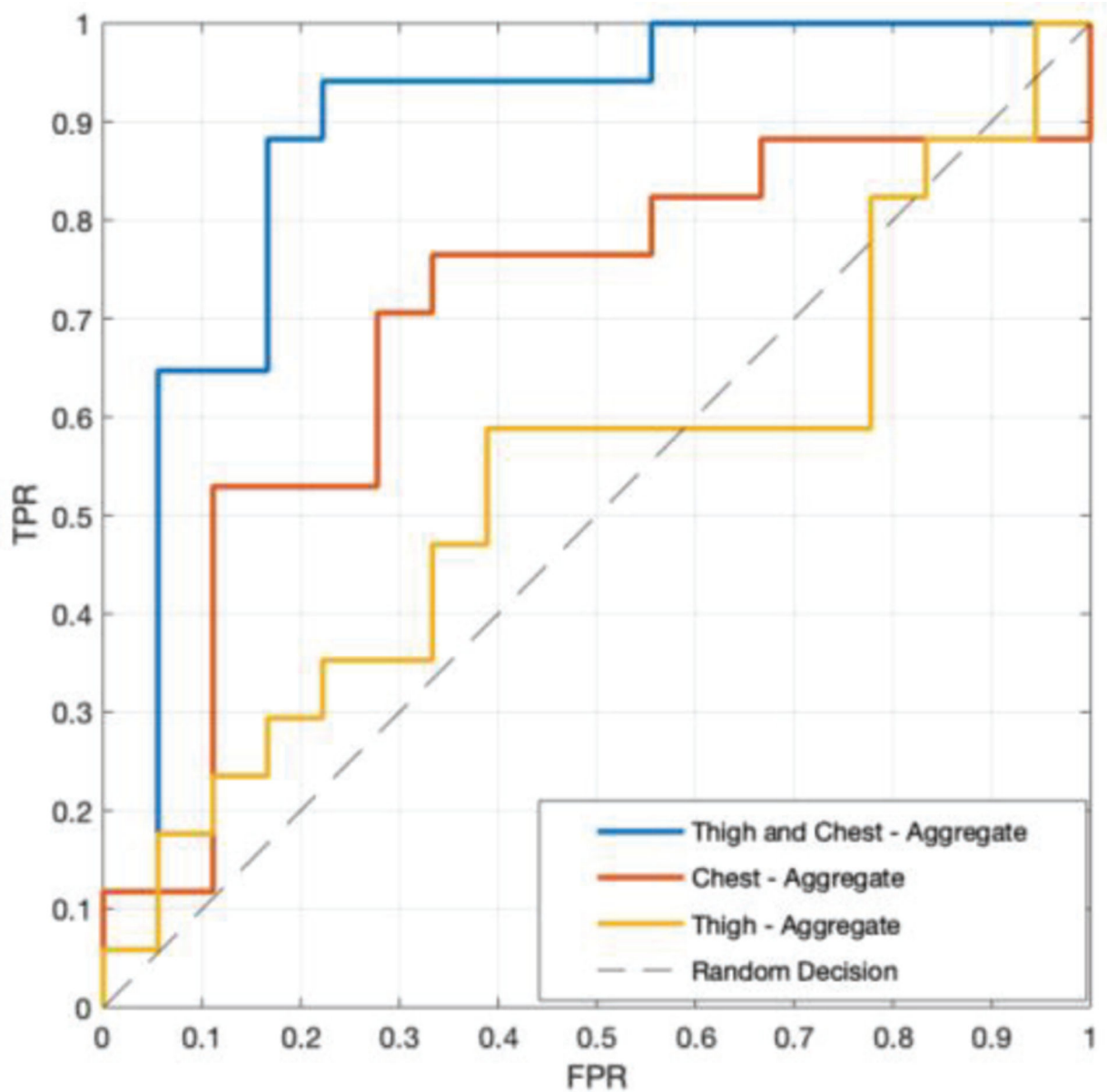
- [54]. Schuster M and Paliwal K, "Bidirectional recurrent neural networks," *Signal Processing, IEEE Transactions on*, vol. 45, pp. 2673–2681, 12. 1997, doi: 10.1109/78.650093.
- [55]. Mortaza N, Osman NAA, and Mehdikhani N, "Are the spatio-temporal parameters of gait capable of distinguishing a faller from a non-faller elderly?," *EUROPEAN JOURNAL OF PHYSICAL AND REHABILITATION MEDICINE*, vol. 50, no. 6, p. 15, 2014.
- [56]. Spain RI et al., "Body-worn motion sensors detect balance and gait deficits in people with multiple sclerosis who have normal walking speed," *Gait & Posture*, vol. 35, no. 4, pp. 573–578, 4. 2012, doi: 10.1016/j.gaitpost.2011.11.026. [PubMed: 22277368]
- [57]. Stone KL et al., "Actigraphy-Measured Sleep Characteristics and Risk of Falls in Older Women," *Archives of Internal Medicine*, vol. 168, no. 16, pp. 1768–1775, 9. 2008, doi: 10.1001/archinte.168.16.1768. [PubMed: 18779464]
- [58]. Davies DL and Bouldin DW, "A Cluster Separation Measure," *IEEE Transactions on Pattern Analysis and Machine Intelligence*, vol. PAMI-1, no. 2, pp. 224–227, 4. 1979, doi: 10.1109/TPAMI.1979.4766909.
- [59]. Hosmer DW, *Applied logistic regression*, Third edition / Hosmer David W. Jr., Lemeshow Stanley, Sturdivant Rodney X.. Hoboken, New Jersey : Wiley, 2013.
- [60]. Géron Aurélien, *Understanding support vector machines*. O'Reilly Media, Inc, 2017.
- [61]. Nagy Zsolt, *Artificial Intelligence and Machine Learning Fundamentals*. Packt Publishing, 2018.
- [62]. Lee W, *Python® Machine Learning*. Indianapolis, Indiana: Indianapolis, Indiana: John Wiley & Sons, Inc.
- [63]. Allali G et al., "Gait variability in multiple sclerosis: a better falls predictor than EDSS in patients with low disability," *Journal of Neural Transmission*, vol. 123, no. 4, pp. 447–450, 4. 2016, doi: 10.1007/s00702-016-1511-z. [PubMed: 26846418]
- [64]. Gurchiek RD, Cheney N, and McGinnis RS, "Estimating Biomechanical Time-Series with Wearable Sensors: A Systematic Review of Machine Learning Techniques," *Sensors*, vol. 19, no. 23, p. 5227, 1. 2019, doi: 10.3390/s19235227.
- [65]. Howcroft J, Kofman J, and Lemaire ED, "Feature selection for elderly faller classification based on wearable sensors," *J Neuroeng Rehabil*, vol. 14, no. 1, p. 47, 30 2017, doi: 10.1186/s12984-017-0255-9. [PubMed: 28558724]
- [66]. Yang Y, Hirdes JP, Dubin JA, and Lee J, "Fall Risk Classification in Community-Dwelling Older Adults Using a Smart Wrist-Worn Device and the Resident Assessment Instrument-Home Care: Prospective Observational Study," *JMIR Aging*, vol. 2, no. 1, p. e12153, 6. 2019, doi: 10.2196/12153. [PubMed: 31518278]



**Fig. 1.** Raw z-acceleration from chest and thigh with walking bouts and 4 steps extracted to provide input for BiLSTM classifier determining fall classification.



**Fig. 2.** ROC curve of 4-stride BiLSTM using the median decision scores over the 1-minute walking trial compared to evaluating each input.



**Fig. 3.** ROC curve (true positive rate - TPR vs. false positive rate - FPR) of best performing model for each sensor configuration; thigh and chest, thigh only, and chest only. All results were found using the 4-stride BiLSTM and taking the median decision score of all strides in the 1-minute walking trial.

**TABLE I**

PERFORMANCE OF DEEP LEARNING MODELS BY NUMBER OF STRIDES AND AMOUNT OF DATA CONSIDERED

Strides	Model	AGG	ACC	SPE	SEN	AUC	F1
5	BiLSTM	None	0.74	0.86	0.58	0.76	0.67
		Median	0.77	<b>0.89</b>	0.65	0.79	0.73
4	BiLSTM	None	0.73	0.76	0.69	0.80	0.70
		Median	<b>0.86</b>	0.83	<b>0.88</b>	<b>0.88</b>	<b>0.86</b>
3	BiLSTM	None	0.70	0.73	0.67	0.77	0.67
		Median	0.80	0.83	0.76	0.83	0.79
2	LSTM	None	0.72	0.77	0.66	0.74	0.68
		Median	0.80	0.83	0.76	0.78	0.79
1	LSTM	None	0.58	0.64	0.52	0.61	0.52
		Median	0.60	0.67	0.53	0.65	0.56

LSTM: Long-Short Term Memory Neural Network; BiLSTM: Bidirectional Long-Short Term Memory Neural Network; AGG: Aggregation technique (none or median of 1 minute of stride observations); AUC: Area Under the Receiver Operating Characteristic Curve; ACC: Accuracy; SPE: Specificity; SEN: Sensitivity; F1: F1-Score.

**TABLE II**

Machine learning model performance for various feature sets

Feature Set	Model	ACC	SPE	SEN	AUC	F1
Spatiotemporal	SVM	<b>0.73</b>	<b>0.89</b>	0.55	0.73	<b>0.67</b>
Stride	LR	0.65	0.73	0.56	0.69	0.59
Aggregated Stride	LR	0.66	0.78	0.53	0.76	0.60
EDSS	SVM	0.62	<b>0.89</b>	0.33	0.71	0.46
PRM	SVM	0.70	0.84	0.56	0.74	0.65
PRM + EDSS	LR	0.70	0.79	<b>0.61</b>	<b>0.79</b>	<b>0.67</b>

SVM: Support Vector Machine; LR: Logistic Regression; AUC: Area Under the Receiver Operating Characteristic Curve; ACC: Accuracy; SPE: Specificity; SEN: Sensitivity; F1: F1-Score.

Author Manuscript

Author Manuscript

Author Manuscript

Author Manuscript