Article

Tracking Foot Drop Recovery Following Lumbar-Spine Surgery, Applying Multiclass Gait Classification Using Machine Learning Techniques

Shiva Sharif Bidabadi [1,*] **, Tele Tan** [1] **, Iain Murray** [2] **and Gabriel Lee** [3]

[1] School of Civil and Mechanical Engineering, Curtin University of Technology, Perth 6102, Australia; T.Tan@curtin.edu.au

[2] School of Electrical Engineering, Computing and Mathematical Sciences, Curtin University of Technology, Perth 6102, Australia; I.Murray@curtin.edu.au

[3] St John of God Subiaco Hospital Perth 6008, Australia and School of Surgery of University of Western Australia, Perth 6009, Australia; glee.neurosurgery@bigpond.com

* Correspondence: Shiva.Sharif@curtin.edu.au; Tel.: +61-892663475

check for updates

Abstract: The ability to accurately perform human gait evaluation is critical for orthopedic foot and ankle surgeons in tracking the recovery process of their patients. The assessment of gait in an objective and accurate manner can lead to improvement in diagnoses, treatments, and recovery. Currently, visual inspection is the most common clinical method for evaluating the gait, but this method can be subjective and inaccurate. The aim of this study is to evaluate the foot drop condition in an accurate and clinically applicable manner. The gait data were collected from 56 patients suffering from foot drop with L5 origin gathered via a system based on inertial measurement unit sensors at different stages of surgical treatment. Various machine learning (ML) algorithms were applied to categorize the data into specific groups associated with the recovery stages. The results revealed that the random forest algorithm performed best out of the selected ML algorithms, with an overall 84.89% classification accuracy and 0.3785 mean absolute error for regression.

Keywords: foot drop; gait classification; machine learning; inertial measurement unit

---

1. Introduction

Walking ability is a key physical behavior that can strongly influence the individual's personal independence, and therefore, the successful execution of their daily activities. Thus, assessment of the gait is frequently required in the clinical setting. It becomes increasingly necessary and important to assess the gait during the treatment of a wide range of gait disorders [1].

Regardless of the significance of gait disorders, there is no widely accepted method for assessing the quality of walking. The most common methods of gait evaluation are the Berg Balance Scale (BBS) [2], dynamic gait index [3], 10-Meter Walk Test [4], 6-Min Walk Test [5], and the Functional Ambulation Categories (FACs) [6]. All these methods evaluate walking ability using different ranges and through the application of different tasks and specified ranges [7]. During these tests, the medical practitioner visually observes the walking ability of the patient and ranks this ability. Therefore, the outcome of these assessments is subjective and may be inaccurate [8].

To overcome this limitation, different methods and devices have been investigated and introduced in practice. As a general measure, walking velocity has been used as an indication of gait health [9,10]. Manual muscle testers are also used to measure muscle strength, which is indirectly related to walking gait [11]. Recently, several studies have been conducted in an attempt to utilize the technology of various sensors in gait analysis. For example, the GAITRite System is a waking platform that uses

a set of pressure sensors and a software system to track gait events [12]. The prosthetic activity monitor (PAM) is also used to assess physical activities based on acceleration measurements [1]. In addition, there are various complex optical motion capture systems, such as the Vicon system, that accurately measure human movements [13]. However, optical motion capture systems are expensive and require software expertise to operate; therefore, they are not practical for conducting daily clinical assessments [14]. Among the many different movement measurement methods, inertial measurement unit (IMU) sensors have been widely implemented for gait analysis due to their particular advantages, such as long-term monitoring and portable recording of ambulatory measurements [15–17].

The gait data collected from IMU sensors are generally large, noisy, complex, and interconnected. Automated methods (e.g., machine learning (ML), which can extract high-level information from raw data) are the preferred solution for managing this data characteristic [18,19]. These methods are currently being used in various medical applications [20]. For example, the walking-gait pattern of patients with Parkinson's disease has been assessed for identification of gait freeze and for distinguishing the characteristics of Parkinson's gait [21,22]. In addition, by utilizing ML algorithms, real-time monitoring of elderly patients' fall down has become possible [23].

In other research, IMU sensors used together with ML analysis have been reported to assist in the identification of different gait phases and human identification via gait patterns [24,25]. These methods have also been implemented to detect gait-related symptoms, such as fall detection or lower extremity muscular fatigue [26,27]. The integration of ML algorithms for the analysis of IMU gait data has been found to be a feasible solution for helping in the diagnosis of neurological disorders involving gait [28].

Foot drop is a common gait disorder in adults, which may be due to varied aetiologies [29]. In simplistic terms, a weakness of the muscles which dorsiflex the foot at the ankle leads to a "dropping" effect on the front of the foot when an induvial walks. This can cause tripping and recurrent falls, with potentially disastrous consequences [30,31]. In the longer term, the abnormal gait pattern leads to compensatory mechanisms, which can also have an impact on other joints. It is important to recognise that the term, foot drop, is an all-encompassing end point phenomenon, which does not relate to a precise aetiology or pathophysiological mechanism. In this particular study, the authors have attempted to recruit patients who have developed a foot drop specifically due to compressive L5 (the fifth lumbar spine vertebrae) radiculopathy. An L5 lumbar nerve root lesion results in paresis of the tibialis anterior, extensor hallucis longus, extensor digitorum brevis, and gluteal muscles of the lower limb on clinical examination, and ankle dorsiflexion, ankle eversion, toe extension, and hip abduction weakness is typically documented [32]. Consequently, a characteristic foot drop and a pathological gait develops. This uniform criteria overcomes an obvious criticism of current studies, which purport to include patients with foot drop, but which, in reality, are attributable to heterogeneous pathologies. The present research aims to provide an objective methodology for tracking the recovery process in foot drop disorder, specifically in patients with L5 radiculopathy following lumbar-spine surgery. Also introducing a gait quality index based on regression technique to assist medical practitioners in the assessment of foot drop severity and the recovery state of patients. For this purpose, the gait characteristics were captured using IMU sensors and multiple ML algorithms were applied and results compared.

2. Materials and Methods

A series of walking trials were recorded using a custom designed system based on three IMU sensors [33]. The data were captured while walking from participants' foot (barefoot), shank, and thigh in the form of different angle measurements over time (pitch (x), roll (y), and yaw (z)). This information was then transmitted and stored via wireless communication. This IMU system has been proven to be feasible for gait assessments in a previous study by the authors. That study compared the accuracy of the IMU system with the Vicon motion capture system (with an 18 camera setup) [31]. A strong correlation was observed (more than 96.9%) between the IMU system and the Vicon motion capture system [34].

2.1. Test Protocol

IMU sensors were attached to participants' lower limbs using straps and double-sided tape. Participants were asked to perform three to seven walking trials in a straight line in their usual walking style. They were also asked to pause and wait for two seconds before walking. The first 500 ms of data were used to offset the sensor readings during the post-processing stage.

2.2. Participants

The gait data were gathered from two groups of participants having specified inclusion and exclusion criteria: (1) A group of 30 participants with healthy gait styles and without any reported gait-related problems (normal group); (2) a group of 56 patients recruited from a neurosurgery practice, who presented with ankle-dorsiflexion weakness with L5 radiculopathy origins (foot drop group). Degenerative lumbar spine disorders, such as radiculopathy at L5, can cause foot drop. The mechanism is through a compression of the nerve fibers that constitute the peroneal nerve [32]. The compressive pathology of the L5 nerve root was confirmed using magnetic resonance imaging (MRI) of the lumbar spine region.

The data from the second group were captured from the affected side of the lower limb and in three different stages: First, before the lumbar spine surgery (pre); second, during the first two days following surgery (post 1); and third, two to three weeks after the surgery while recovering (post 2).

The walking capabilities of different subjects at different stages may have varied due to wound pain, patient fatigue, or other related problems. Therefore, the speed, distance, and the number of steps in different trials were not equal. To overcome this limitation, a resampling process was applied, which will be further explained in Section 2.3.

Given that the study involves human participants, the relevant ethical approvals were obtained from both the Curtin University of Technology (Human Research Ethics Office): HR 12/2016 and St John of God Hospital (Human Research Ethics Committee): 823.

2.3. Data Preprocessing

As stated, the gait data were collected from different participant groups over different periods of time. Therefore, the number of samples captured from the pre, post 1, post 2, and normal groups was 203, 199, 136, and 178 respectively. To resolve this data size problem, some of the long walking trials were subdivided into two walking sample sets, each set including at least three walking steps. For example, to match the number of samples in the post 2 data set, 67 long walking trials were selected and each was split into two trials. This increased the original number of post 2 samples by 67. After applying the mentioned resampling method to each data set, the number of walking samples for all groups was normalized to 203 trials.

The captured dataset consisted of three angle measurement in the format of a time series signal for pitch, roll, and yaw movement. Fast Fourier transform (FFT) was implemented on these time series signals to extract the signals' features, such as the fundamental harmonic, amplitude, and phase shifts. Previous studies have confirmed the capability of FFT in analyzing gait and IMU data [27,35]. The signals were modelled using FFT as follows:

$$F(t) = \sum_{i=0} P_i \sin(2\pi f_0 i + \phi_i),\tag{1}$$

where f_0, P_i, and ϕ_i are the fundamental harmonic, amplitude, and phase shift of the ith harmonic, respectively. The FFT was applied to each angle from the walking samples, meaning a 15-feature model in the frequency domain was obtained [36]. The foot sensor (S1), shank sensor (S2), and thigh sensor (S3) recorded the movements, with each sensor representing the movement in the sagittal, coronal, and transverse planes as the pitch, roll, and yaw, respectively. In addition to the 15-feature model in the

frequency domain, also, each angle in the time domain constituted an extra nine features. The final model consisted of 144 features as below:

$$144 \; Features \left(\underbrace{\overbrace{5}^{\text{Frequency domain}} + \underbrace{5}_{P_{1-5}} + \underbrace{5}_{\phi_{1-5}} + \overbrace{\underbrace{1}_{Angle}}^{\text{Time domain}} \right) \times \overbrace{\underbrace{3}_{\text{Pitch, Roll, Yaw}}}^{\text{Movments}} \times \overbrace{\underbrace{3}_{\text{Thigh, Shank, Foot}}}^{\text{Sensors}}. \qquad (2)$$

2.4. Feature Extraction, Classification, and Regression

The Waikato Environment for Knowledge Analysis (WEKA) software version 3.8 was used as the workbench for evaluation of the 11 ML algorithms used to classify gait pattern based on the model with 144 features [37]. The following are the 11 classification algorithms that were analyzed: Deep learning, multilayer perceptron, K-nearest neighbors (IBK), logistic regression, Bayes net, naive Bayes, C4.5 decision tree (J48), random forest (unlimited depth with 100 iterations), random tree (unlimited depth with 100 iterations), support vector machine (SVM) (radial basis function kernel), and OneR (1R). A 10-fold cross-validation methodology was applied for each classification.

Four measures were used to compare the performance of the 11 classification algorithms. First, classification accuracy was defined as the number of correct predictions over the total number of instances in that dataset. Second, the confusion matrix, which provides information about correct and incorrect predictions, was created for each classifier [38]. The confusion matrix is a square matrix in which $C_{i,j}$ indicates the number of instances predicted as class *i*, where they were from class *j* originally. The best classification will have only zero values outside the main diagonal. In addition, the precision and F-score were calculated:

$$Precision = \frac{TP}{(TP + FP)}, \qquad (3)$$

$$F \; score = \frac{2 \times TP}{2 \times TP + FP + FN}, \qquad (4)$$

where *TP*, *FP*, and *FN* are the true positive, false positive, and false negative, respectively. True positives are items correctly labeled as belonging to their class. False positives are items incorrectly labeled as belonging to the class. False negatives are items which were not labeled as belonging to the class, but should have been. Among all 144 features describing gait in this model, some may have a higher effect in describing the severity level of foot drop. To find features with the most significant effect, the wrapper feature selection technique was implemented. The wrapper technique-based method was implemented alongside the classification algorithm to review the subset of the input features that maximizes a predefined objective function. In this case, the objective was to maximize the classification accuracy and to minimize the false alarm rate. A vector of scores for all features indicates the significance of the features. In this study, the wrapper feature selection technique was conducted using all 11 classification algorithms and the results are presented in the next section [39]. In this procedure, the data set was shuffled randomly and split into 10 groups. Each group was taken as the hold out set (or test data set) once and the remaining groups as the training data set.

Classification was done on the training set and evaluated on the test set retaining an evaluation score. The 10 fold cross-validation procedure was applied 11 times using each classification algorithm and the whole dataset was evaluated each time.

In addition to the aim of classifying gait, a further aim of this study was to find an objective index to indicate the severity of observed foot drop symptoms. To achieve this index, eight regression ML algorithms were investigated using the WEKA framework. The following are the eight regression ML algorithms that were analyzed: Deep learning, multilayer perceptron, IBK, random forest, random tree, linear regression, simple linear regression, and SVM regression. Some of the classification algorithms provide an index indicating the likelihood of their prediction, therefore they can be used as the regression algorithem. To be able to perform the regression analysis on the dataset, the state variable,

which indicates in which state the sample was captured (e.g., pre, post 1, post 2), was changed to a numerical value from zero to four, which refers to the pre, post 1, post 2, and normal states, respectively.

Different measures were used to evaluate the performance of the regression algorithms. The first measure was the error between the predictions and the actual value of the class. For example, a regression algorithm may have predicted a post 2 (i.e., 2) walking sample as 2.8, and therefore the error for this sample is 0.8. Additionally, the correlation coefficient, mean absolute error, root mean square error (RMSE), relative absolute error, and root relative squared error were used as measures for evaluating the regression performance as shown below:

$$Correlation\ coefficient = \frac{N \sum y\acute{y} - (\sum y)(\sum \acute{y})}{\sqrt{\left[N \sum y^2 - (\sum y)^2\right]\left[N \sum \acute{y}^2 - (\sum \acute{y})^2\right]}}, \tag{5}$$

$$Mean\ absolute\ error = \frac{1}{N} \sum_{i=1}^{N} |y - \acute{y}|, \tag{6}$$

$$Root\ mean\ absolute\ error = \sqrt{\frac{1}{N} \sum_{i=1}^{N} (y - \acute{y})^2}, \tag{7}$$

$$Relative\ absolute\ error = \frac{\sum_{i=1}^{N} |y_i - \acute{y}_i|}{\sum_{i=1}^{N} |\overline{y} - \acute{y}_i|} \times 100, \tag{8}$$

$$Root\ relative\ squared\ error = \sqrt{\frac{\sum_{i=1}^{N} (\acute{y}_i - y_i)^2}{\sum_{i=1}^{N} (\acute{y}_i - \overline{y})^2}} \times 100, \tag{9}$$

where y and $\acute{y}$ are the actual and prediction values and N is the number of samples.

3. Results

This section compares the performance of the ML algorithms on the collected data.

3.1. Analysis of the Four Classes

First, the gait patterns were classified into four classes using the entire available dataset from the patients at different stages of treatment and the participants in the normal group. These four classes were the pre, post 1, post 2, and normal class.

Table 1 presents the accuracy of the ML classification algorithms for the four classes.

Table 1. Accuracy from 11 ML algorithms classifying data into four classes.

Algorithm	Accuracy (%)
Bayes net	55.0493
Deep learning	52.3399
IBK	50.6158
J48	57.1429
Logistic regression	54.1872
Multilayer perceptron	54.6798
Naïve Bayes	51.4778
OneR	47.9064
Random forest	67.3645
Random tree	55.665
SVM	62.3153
Average	55.3400

Table 1 demonstrates that random forest and OneR have the maximum and minimum accuracy, respectively. The average overall accuracy of the algorithms is 55.34%, which is low. To investigate the cause of the low accuracy of the algorithms, the confusion matrix was generated and investigated. Figure 1 presents the confusion matrix, precision, and F-score observed from all classification algorithms. The figure is color coded so that as the value of the cell increases, the cell is colored with a darker red.

Algorithm	Confusion matrix				Prediction	Precision	F-score
	Pre	Post 1	Post 2	Normal			
One R	81	85	5	32	Pre	0.40	0.41
	90	70	14	29	Post 1	0.34	0.35
	4	7	139	53	Post 2	0.68	0.67
	18	32	54	99	Normal	0.49	0.48
Bayes Net	65	101	8	29	Pre	0.32	0.38
	56	118	11	18	Post 1	0.58	0.53
	4	5	147	47	Post 2	0.72	0.70
	18	16	52	117	Normal	0.58	0.57
Naïve Bayes	62	125	7	9	Pre	0.31	0.34
	60	128	9	6	Post 1	0.63	0.51
	12	17	116	58	Post 2	0.57	0.63
	29	27	35	112	Normal	0.55	0.58
SVM	104	24	74	1	Pre	0.51	0.64
	16	117	69	1	Post 1	0.58	0.68
	0	1	200	2	Post 2	0.99	0.60
	1	1	116	85	Normal	0.42	0.58
IBK	82	64	22	35	Pre	0.40	0.41
	72	86	20	25	Post 1	0.42	0.45
	17	12	125	49	Post 2	0.62	0.61
	26	19	40	118	Normal	0.58	0.55
Logistic	88	72	22	21	Pre	0.43	0.44
	81	88	22	12	Post 1	0.43	0.45
	19	14	123	47	Post 2	0.61	0.60
	11	13	38	141	Normal	0.69	0.67

Algorithm	Confusion matrix				Prediction	Precision	F-score
	Pre	Post 1	Post 2	Normal			
J48	95	84	11	13	Pre	0.47	0.46
	76	102	13	12	Post 1	0.50	0.49
	17	15	132	39	Post 2	0.65	0.67
	18	15	35	135	Normal	0.67	0.67
Random Tree	84	87	17	15	Pre	0.41	0.44
	60	110	16	17	Post 1	0.54	0.51
	16	12	145	30	Post 2	0.71	0.67
	20	21	49	113	Normal	0.56	0.60
Random Forest	77	98	13	15	Pre	0.38	0.47
	38	141	13	11	Post 1	0.69	0.61
	3	11	168	21	Post 2	0.83	0.79
	4	8	30	161	Normal	0.79	0.78
Multilayer perceptron	76	91	11	25	Pre	0.37	0.39
	86	89	13	15	Post 1	0.44	0.43
	13	14	138	38	Post 2	0.68	0.70
	15	15	32	141	Normal	0.69	0.67
Deep learning	64	97	20	22	Pre	0.32	0.35
	80	90	15	18	Post 1	0.44	0.44
	13	13	132	45	Post 2	0.65	0.64
	11	10	43	139	Normal	0.68	0.65

Figure 1. Confusion matrix, precision, and F-score from classifying data into four classes.

In addition to the classification algorithms, the regression method was used to compare and evaluate each class data on a numerical basis. Here, the pre category was given a base value of 1, post 1 a value of 2, post 2 a value of 3, and normal a value of 4. Table 2 presents the error measures of the eight regression algorithms for the four classes (i.e., pre, post 1, post 2, and normal).

Table 2. Error measures from eight regression algorithms classifying data into four classes.

Algorithm	Correlation Coefficient	Mean Absolute Error	Root Mean Squared Error	Relative Absolute Error	Root Relative Squared Error
Deep learning	−0.0856	1.3104	1.5825	130.8814	141.4032
IBk	0.4181	0.7808	1.2217	77.9844	109.1655
Linear regression	0.6430	0.7003	0.8910	69.9453	79.6130
Multilayer perceptron	0.0026	1.2089	9.3257	120.7437	833.2870
Random forest	0.7807	0.5426	0.7119	54.1923	63.6136
Random tree	0.5748	0.6252	1.0010	62.4422	89.4466
Simple linear regression	0.4454	0.7936	1.0010	79.2604	89.4430
SVM	0.6244	0.7062	0.9067	70.5317	81.0133

Figure 2 presents the error of the predictions and the error bar plot of the regression algorithms.

In Figure 2, the mean value of the prediction of each algorithm is represented by a black dot, while the standard deviation is represented by a rectangle. The minimum and maximum values are presented using thin lines in each bar. Figure 2 represents the eight regression algorithms for the four classes.

It is notable in Figure 1 that in all 11 classifications, the predictions for the pre and post 1 categories are often confused (i.e., the data from the pre class were classified as post 1 and vice versa). This error in prediction is the principal cause of the low accuracy level of the algorithms presented in Table 1.

Therefore, the following section presents the results from the classification and regression algorithms on the dataset without post 1 data.
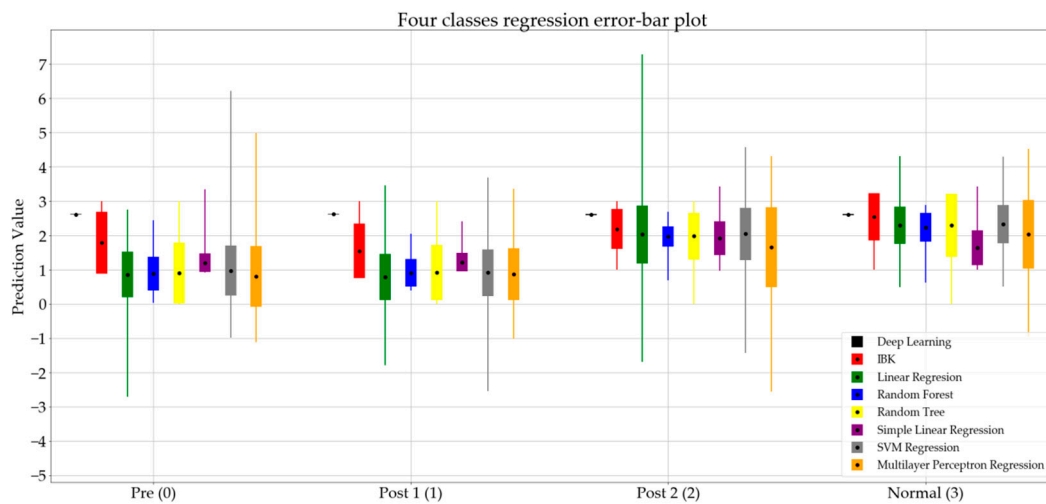

Figure 2. Error bar plot for regression algorithms classifying data into four classes.

3.2. Analysis of Three Classes

This section presents the results of the classification tests when the post 1 class was removed from the analysis. Table 3 presents the accuracy of ML algorithms when the dataset was classified into three classes of pre, post 2, and normal. The second column shows the classification results when all 144 features were used. In general, Table 3 compares the accuracy observed before and after applying the wrapper feature selection technique.

Table 3. Classification accuracy from 11 ML algorithms classifying data into three classes before and after feature selection.

Algorithm	All Features (n = 144)	Selected Features	
	Accuracy (%)	Accuracy (%)	Number of Selected Features
Bayes net	70.94	79.97	20
Deep learning	70.61	76.52	21
IBk	62.89	75.37	46
J48	73.56	76.03	22
Logistic regression	68.47	79.15	22
Multilayer perceptron	75.21	76.52	20
Naïve Bayes	68.64	76.52	44
OneR	65.85	67.98	1
Random forest	83.25	84.89	33
Random tree	69.62	75.04	21
SVM	67.82	77.83	30
Average	70.62	76.89	NA

Figures 3 and 4 present the confusion matrix, precision, and F-score of the classification algorithms before and after the feature selection, respectively.

Algorithm	Confusion matrix			Prediction	Precision	F-score
	Pre	Post 2	Normal			
One R	147	9	47	Pre	0.72	0.75
	11	135	57	Post 2	0.67	0.68
	33	51	119	Normal	0.59	0.56
Bayes Net	168	7	28	Pre	0.83	0.81
	9	148	46	Post 2	0.73	0.72
	34	53	116	Normal	0.57	0.59
Naïve Bayes	185	8	10	Pre	0.91	0.79
	29	116	58	Post 2	0.57	0.64
	53	33	117	Normal	0.58	0.60
SVM	127	75	1	Pre	0.63	0.76
	2	199	2	Post 2	0.98	0.67
	1	115	87	Normal	0.43	0.59
IBK	129	25	49	Pre	0.64	0.67
	24	123	56	Post 2	0.61	0.63
	31	41	131	Normal	0.65	0.60
Logistic	152	26	25	Pre	0.75	0.73
	34	128	41	Post 2	0.63	0.64
	25	41	137	Normal	0.67	0.67

Algorithm	Confusion matrix			Prediction	Precision	F-score
	Pre	Post 2	Normal			
J48	159	21	23	Pre	0.78	0.78
	26	145	32	Post 2	0.71	0.71
	20	39	144	Normal	0.71	0.72
Random Tree	150	25	28	Pre	0.74	0.77
	17	148	38	Post 2	0.73	0.69
	21	56	126	Normal	0.62	0.64
Random Forest	172	15	16	Pre	0.85	0.87
	12	174	17	Post 2	0.86	0.82
	10	32	161	Normal	0.79	0.81
Multilayer perceptron	163	16	24	Pre	0.80	0.81
	17	145	41	Post 2	0.71	0.73
	20	33	150	Normal	0.74	0.72
Deep learning	146	27	30	Pre	0.72	0.74
	24	137	42	Post 2	0.67	0.68
	21	35	147	Normal	0.72	0.70

Figure 3. Confusion matrix, precision, and F-score from classification algorithms classifying data into three classes before feature selection.

Algorithm	Confusion matrix			Prediction	Precision	F-score
	Pre	Post 2	Normal			
One R	147	9	47	Pre	0.72	0.75
	12	141	50	Post 2	0.69	0.70
	29	48	126	Normal	0.62	0.59
Bayes Net	175	18	10	Pre	0.86	0.86
	17	163	23	Post 2	0.80	0.77
	14	40	149	Normal	0.73	0.77
Naïve Bayes	167	9	27	Pre	0.82	0.80
	28	144	31	Post 2	0.71	0.75
	20	28	155	Normal	0.76	0.75
SVM	165	16	22	Pre	0.81	0.83
	14	161	28	Post 2	0.79	0.77
	16	39	148	Normal	0.73	0.74
IBK	167	14	22	Pre	0.82	0.84
	13	136	54	Post 2	0.67	0.71
	16	31	156	Normal	0.77	0.72
Logistic	176	9	18	Pre	0.87	0.85
	23	144	36	Post 2	0.71	0.75
	12	29	162	Normal	0.80	0.77

Algorithm	Confusion matrix			Prediction	Precision	F-score
	Pre	Post 2	Normal			
J48	165	16	22	Pre	0.81	0.80
	20	151	32	Post 2	0.74	0.75
	23	33	147	Normal	0.72	0.73
Random Tree	164	16	23	Pre	0.81	0.82
	20	149	34	Post 2	0.73	0.72
	14	45	144	Normal	0.71	0.71
Random Forest	177	17	9	Pre	0.87	0.88
	15	171	17	Post 2	0.84	0.82
	7	27	169	Normal	0.83	0.85
Multilayer perceptron	173	14	16	Pre	0.85	0.84
	18	148	37	Post 2	0.73	0.73
	17	41	145	Normal	0.71	0.72
Deep learning	175	12	16	Pre	0.86	0.84
	24	149	30	Post 2	0.73	0.73
	17	44	142	Normal	0.70	0.73

Figure 4. Confusion matrix, precision, and F-score from classification algorithms classifying data into three classes after feature selection.

As seen in Table 3, the classification performance improved after applying feature selection. The best performing classifier was random forest before and after feature selection. Also, the random forest as the best classifier had 33 selected features (Table 3) using the wrapper technique. Table 4 indicates the type and body part of the selected features.

Table 4. The features selected by wrapper techniques using the random forest algorithm.

Feature Type	Foot	Shank	Thigh	Total
Pitch	5	5	6	16
Roll	6	4	1	11
Yaw	4	1	1	6
Total	15	10	8	33

For the next step of the analysis, the regression algorithms were applied to the three classes (i.e., pre, post 2, and normal), and the regression error measures calculated are presented in Table 5.

Table 5. Error measures from eight regression algorithms classifying data into four classes.

Algorithm	Correlation Coefficient	Mean Absolute Error	Root Mean Squared Error	Relative Absolute Error (%)	Root Relative Squared Error (%)
Deep learning	−0.0598	0.9396	1.1555	140.4993	141.4422
IBk	0.4966	0.4631	0.8245	69.2432	100.9249
Linear regression	0.6109	0.5319	0.5319	79.5453	88.4125
Random forest	0.7931	0.3785	0.5099	56.6055	62.4162
Random tree	0.5861	0.3992	0.7267	59.6992	88.9477
Simple linear regression	0.2998	0.6580	0.7963	98.3908	97.4729
SVM	0.5930	0.5470	0.7221	81.8015	88.3848
Multilayer perceptron	0.5469	0.5649	0.8279	84.4696	101.3445

Figure 5 presents the error bar plot of the regression algorithms. In this figure, the pre, post 2, and normal states are represented by the numbers 0, 1, and 2, respectively.
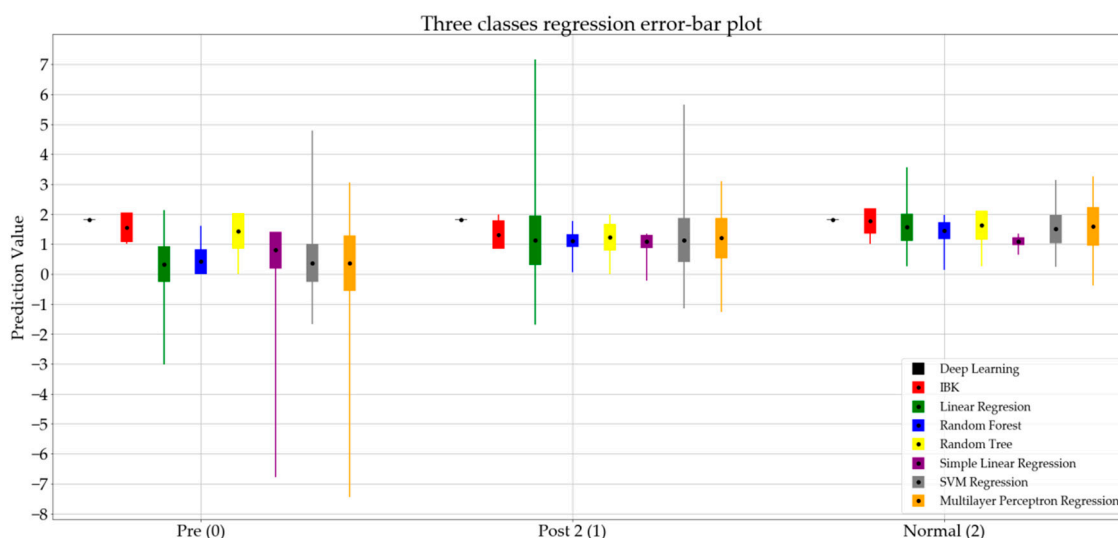
Figure 5. Error bar plot for eight regression algorithms classifying data into three classes.

4. Discussion

This research demonstrated a systematic and objective methodology for the evaluation of foot drop with L5 lumber radiculopathy origins.

As presented in Table 5, the random forest regression shows the lowest mean absolute error. After investigating the performance of the random forest regression more closely in Figure 5, and comparing the results for the pre and post 2 states, a jump in the mean value of the predictions is noticeable. This indicates that the prediction values for the random forest algorithm can be used as an index to determine the severity level of foot drop in the walking gait pattern. Therefore, the method presented in this research shows promise as a potential measurement tool for tracking the recovery process of foot drop with L5 origins in adults. However, outliers in the random forest algorithm require further investigation.

This study found that the random forest algorithm provides the best classification, with an 84.89% accuracy. The average accuracy of the classifiers improved significantly after removing the post 1 class and applying the classification to three instead of four classes. Additionally, the use of wrapper feature selection proved to be effective in improving the classification performance of the algorithms in the three class analysis (Table 3). The improvements in accuracy when decreasing the number of features

indicates that the current IMU system can be simplified by reducing the number of sensors, which will lower the computation expenses. Also, Table 4 indicates that the wrapper technique, which was applied to the random forest classifier, selected features mainly from the foot and shank regions, demonstrating the correlation between the IMU sensor location and the ability to classify foot drop conditions. In addition, 48.5% of the all selected features were from pitch (flexion) movement that is known to be affected by foot drop.

As noted in Section 3 and presented in Figure 1, the confusion matrix of all the classification algorithms revealed similarity in the gait patterns of the pre and post 1 groups, which led to confusion between these two classes of movement. This raises an important question about the timeline of monitoring foot drop patients after lumbar spine surgery. According to Section 3, the improvement can be fully observed at least two to three weeks after the surgery [40]. In addition, the confusion matrixes before and after feature selection (Figures 3 and 4) showed that the highest level of confusion occurs between the post 2 and normal stages, which indicates that two to three weeks after surgery, the walking patterns of the patients are similar to the walking patterns of people who are not suffering from foot drop (i.e., the normal group). Also, referring to the F1 scores in Figures 1 and 3, it is noticeable that the false positive or false negative rates are highly reduced while remaining cases can be addressed in the clinical environment by using simultaneous assessments.

In this study, it was demonstrated that the ML algorithms are capable of classifying patients with foot drop from normal patients, without any knowledge of specific gait events (i.e., swing phase, heel contact, toe-off, etc.). This is beneficial since no extra steps are required to identify gait events before the application of ML algorithms.

Figure 2 summarizes the prediction values of the regression algorithms. Comparing the standard deviation for the normal set among the four groups, it is notable that most of the algorithms show a smaller standard deviation for the normal group, which indicates the diversity of gait patterns at different stages of spinal surgery treatment.

In addition to the work in this study, the proposed system has the potential to be used in the clinical environment for an objective evaluation and assessment of gait in the case of any gait-related disorder. While this study only investigated foot drop patients with L5 origins, the study's approach can be applied to any other group of patients with gait-related disorders, such as children with cerebral palsy [41,42].

The system presented in this study has the potential to be used for long-term patient monitoring at home, which not only helps to enable continuous tracking of patient recovery, but also provides more gait data that can help to improve the accuracy of the analysis.

In addition, the system could be used to track the walking pattern of both legs of one patient to compare the gait patterns of both the left and right leg of a patient.

5. Conclusions

This study tested the feasibility of an IMU system with ML analysis to assess the level of severity in foot drop patients by utilizing an applicable method in clinical environments. An application of this approach is to evaluate gait conditions and track the recovery of gait disorders, such as foot drop. The study applied two ML approaches to achieve its aim: Classification and regression. In each approach, multiple ML algorithms were evaluated and compared over the datasets of walking gait from a sample of healthy participants with normal gait styles and a group of patients with foot drop in different stages of lumbar spine surgery. The evaluation was based on the accuracy, confusion matrix, and mean absolute error of the algorithms after classification of the different characteristics of the gaits of participants. The random forest classifier initially resulted in the best accuracy (83.25%). The application of the wrapper feature selection technique to the random forest algorithm improved the accuracy to 84.89%.

Conflicts of Interest: The authors declare no conflict of interest.

References

1. Bussmann, J.B.J.; Culhane, K.M.; Horemans, H.L.D.; Lyons, G.M.; Stam, H.J. Validity of the prosthetic activity monitor to assess the duration and spatio-temporal characteristics of prosthetic walking. *IEEE Trans. Neural Syst. Rehabil. Eng.* **2004**, *12*, 379–386. [CrossRef] [PubMed]

2. Giné-Garriga, M.; Guerra, M.; Manini, T.M.; Marí-Dell'Olmo, M.; Pagès, E.; Unnithan, V.B. Measuring balance, lower extremity strength and gait in the elderly: Construct validation of an instrument. *Arch. Gerontol. Geriatr.* **2010**, *51*, 199–204. [CrossRef] [PubMed]

3. Whitney, S.; Wrisley, D.; Furman, J. Concurrent validity of the Berg Balance Scale and the Dynamic Gait Index in people with vestibular dysfunction. *Physiother. Res. Int.* **2003**, *8*, 178–186. [CrossRef] [PubMed]

4. Van Hedel, H.J.A. Gait dpeed in relation to categories of functional ambulation after spinal cord injury. *Neurorehabilit. Neural Repair* **2009**, *23*, 343–350. [CrossRef] [PubMed]

5. Uszko-Lencer, N.H.; Mesquita, R.; Janssen, E.; Werter, C.; Brunner-La Rocca, H.P.; Pitta, F.; Wouters, E.F.; Spruit, M.A. Reliability, construct validity and determinants of 6-minute walk test performance in patients with chronic heart failure. *Int. Cardiol. J.* **2017**, *240*, 285–290. [CrossRef] [PubMed]

6. Kesikburun, S.; Yavuz, F.; Güzelküçük, Ü.; Yaşar, E.; Balaban, B. Effect of ankle foot orthosis on gait parameters and functional ambulation in patients with stroke. *Turk. J. Phys. Med. Rehab.* **2017**, *63*, 143–149. [CrossRef]

7. Debono, B.; Bousquet, P.; Sabatier, P.; Plas, J.Y.; Lescure, J.P.; Hamel, O. Postoperative monitoring with a mobile application after ambulatory lumbar discectomy: An effective tool for spine surgeons. *Eur. Spine J.* **2016**, *25*, 3536–3542. [CrossRef] [PubMed]

8. Cloete, T.; Scheffer, C. Repeatability of an off-the-shelf, full body inertial motion capture system during clinical gait analysis. In Proceedings of the 2010 Annual International Conference of the IEEE Engineering in Medicine and Biology Society, Buenos Aire, Argentina, 30 August–4 September 2010; pp. 5125–5128.

9. Collen, F.M.; Wade, D.T.; Bradshaw, C.M. Mobility after stroke: Reliability of measures of impairment and disability. *Int. Disabil. Stud.* **1990**, *12*, 6–9. [CrossRef]

10. Viosca, E.; Martínez, J.L.; Almagro, P.L.; Gracia, A.; González, C. Proposal and validation of a new functional ambulation classification scale for clinical use. *Arch. Phys. Med. Rehabil.* **2005**, *86*, 1234–1238. [CrossRef]

11. Marino, M.; Nicholas, J.A.; Gleim, G.W.; Rosenthal, P.; Nicholas, S.J. The efficacy of manual assessment of muscle strength using a new device. *Am. J. Sports Med.* **1982**, *10*, 360–364. [CrossRef]

12. Roche, B.; Simon, A.L.; Guilmin-Crepon, S.; Boizeau, P.; Andriss, B.; Alberti, C.; Presedo, A.; Ilharreborde, B.; Husson, I. Test-retest reliability of an instrumented electronic walkway system (GAITRite) for the measurement of spatio-temporal gait parameters in young patients with Friedreich's ataxia. *Gait Posture* **2018**, *66*, 45–50. [CrossRef] [PubMed]

13. Windolf, M.; Götzen, N.; Morlock, M. Systematic accuracy and precision analysis of video motion capturing systems-exemplified on the Vicon-460 system. *J. Biomech.* **2008**, *41*, 2776–2780. [CrossRef] [PubMed]

14. Higginson, B.K. Methods of running gait analysis. *Curr. Sports Med. Rep.* **2009**, *8*, 136–141. [CrossRef] [PubMed]

15. Tao, W.; Liu, T.; Zheng, R.; Feng, H. Gait Analysis Using Wearable Sensors. *Sensors* **2012**, *12*, 2255–2283. [CrossRef] [PubMed]

16. Pappas, I.P.I.; Keller, T.; Mangold, S.; Popovic, M.R.; Dietz, V.; Morari, M. A Reliable Gyroscope-Based Gait-Phase Detection Sensor Embedded in a Shoe Insole. *IEEE Sens. J.* **2004**, *4*, 268–274. [CrossRef]

17. Li, Q.; Young, M.; Naing, V.; Donelan, J.M. Walking speed and slope estimation using shank-mounted inertial measurement units. In Proceedings of the 2009 IEEE International Conference on Rehabilitation Robotics, Kyoto, Japan, 23–26 June 2009; pp. 839–844.

18. Foster, K.R.; Koprowski, R.; Skufca, J.D. Machine learning, medical diagnosis, and biomedical engineering research-commentary. *Biomed. Eng. Online* **2014**, *13*, 94. [CrossRef]

19. Begg, R.K.; Palaniswami, M.; Owen, B. Support vector machines for automated gait classification. *IEEE Trans. Biomed. Eng.* **2005**, *52*, 828–838. [CrossRef]

20. Hu, B.; Dixon, P.C.; Jacobs, J.V.; Dennerlein, J.T.; Schiffman, J.M. Machine learning algorithms based on signals from a single wearable inertial sensor can detect surface- and age-related differences in walking. *J. Biomech.* **2018**, *71*, 37–42. [CrossRef]

21. Mazilu, S.; Hardegger, M.; Zhu, Z.; Roggen, D.; Tröster, G.; Plotnik, M.; Hausdorff, J.M. Online Detection of Freezing of Gait with Smartphones and Machine Learning Techniques. In Proceedings of the 6th International Conference on Pervasive Computing Technologies for Healthcare (PervasiveHealth) and Workshops, San Diego, CA, USA, 21–24 May 2012; pp. 123–130.

22. Tahir, N.M.; Manap, H.H. Parkinson Disease Gait Classification based on Machine Learning Approach. *J. Appl. Sci.* **2012**, *12*, 180–185. [CrossRef]

23. Pogorelc, B.; Bosnić, Z.; Gams, M. Automatic recognition of gait-related health problems in the elderly using machine learning. *Multimed. Tools Appl.* **2012**, *58*, 333–354. [CrossRef]

24. Williamson, R.; Andrews, B.J. Gait event detection for FES using accelerometers and supervised machine learning. *IEEE Trans. Rehabil. Eng.* **2000**, *8*, 312–319. [CrossRef] [PubMed]

25. Nowlan, M.F. Human Identification via Gait Recognition Using Accelerometer Gyro Forces. Available online: https://pdfs.semanticscholar.org/a63e/04fefd2be621488646ae11bfe66c98d9649e.pdf (accessed on 3 May 2017).

26. Özdemir, A.; Barshan, B. Detecting Falls with Wearable Sensors Using Machine Learning Techniques. *Sensors* **2014**, *14*, 10691–10708. [CrossRef]

27. Zhang, J.; Lockhart, T.E.; Soangra, R. Classifying lower extremity muscle fatigue during walking using machine learning and inertial sensors. *Ann. Biomed. Eng.* **2014**, *42*, 600–612. [CrossRef] [PubMed]

28. LeMoyne, R.; Heerinckx, F.; Aranca, T.; De Jager, R.; Zesiewicz, T.; Saal, H.J. Wearable body and wireless inertial sensors for machine learning classification of gait for people with Friedreich's ataxia. In Proceedings of the 13th International Conference on Wearable and Implantable Body Sensor Networks (BSN), San Francisco, CA, USA, 14–17 June 2016; pp. 147–151.

29. Westhout, F.D.; Paré, L.S.; Linskey, M.E. Central causes of foot drop: Rare and underappreciated differential diagnoses. *J. Spinal Cord Med.* **2007**, *30*, 62–66. [CrossRef] [PubMed]

30. Sackley, C.; Disler, P.B.; Turner-Stokes, L.; Wade, D.T.; Brittle, N.; Hoppitt, T. Rehabilitation interventions for foot drop in neuromuscular disease. In *Cochrane Database of Systematic Reviews*; Sackley, C., Ed.; John Wiley & Sons, Ltd.: Chichester, UK, 2009.

31. Stewart, J.D. Foot drop: Where, why and what to do? *Pract. Neurol.* **2008**, *8*, 158–169. [CrossRef] [PubMed]

32. Young, A.; Getty, J.; Jackson, A.; Kirwan, E.; Sullivan, M.; Parry, C.W. Variations in the pattern of muscle innervation by the L5 and S1 nerve roots. *Spine* **1983**, *8*, 616–624. [CrossRef] [PubMed]

33. Sharif Bidabadi, S.; Murray, I.; Lee, G. The application of Inertial Measurements Unit for the clinical evaluation and assessment of gait events. *Med. J. Eng. Technol.* **2017**, *41*, 612–622. [CrossRef] [PubMed]

34. Sharif Bidabadi, S.; Murray, I.; Lee, G.Y.F. Validation of foot pitch angle estimation using inertial measurement unit against marker-based optical 3D motion capture system. *Biomed. Eng. Lett.* **2018**, *8*, 283–290. [CrossRef] [PubMed]

35. Yoo, J.H.; Nixon, M.S.; Harris, C.J. Model-driven statistical analysis of human gait motion. *Time* **2002**, *1*, 285–288.

36. Wong, A.M.; Simon, S.R.; Olshen, R.A. Statistical analysis of gait patterns of persons with cerebral palsy. *Stat. Med.* **1983**, *2*, 345–354. [CrossRef]

37. Frank, E.; Hall, M.; Holmes, G.; Kirkby, R.; Pfahringer, B.; Witten, I.H.; Trigg, L. Weka-A Machine Learning Workbench for Data Mining. In *Data Mining and Knowledge Discovery Handbook*; Springer: Boston, MA, USA, 2009; pp. 1269–1277.

38. Guan, C.; Qin, S.; Ling, W.; Ding, G. Apparel recommendation system evolution: An empirical review. *Int. J. Cloth. Sci. Technol.* **2016**, *28*, 854–879. [CrossRef]

39. El Aboudi, N.; Benhlima, L. Review on wrapper feature selection approaches. In Proceedings of the 2016 International Conference on Engineering and MIS, ICEMIS 2016, Agadir, Morocco, 22–24 September 2016; pp. 1–5.

40. MacKi, M.; Syeda, S.; Kerezoudis, P.; Gokaslan, Z.L.; Bydon, A.; Bydon, M. Preoperative motor strength and time to surgery are the most important predictors of improvement in foot drop due to degenerative lumbar disease. *J. Neurol. Sci.* **2016**, *361*, 133–136. [CrossRef] [PubMed]

41. Patane, F.; Rossi, S.; Del Sette, F.; Taborri, J.; Cappa, P. WAKE-Up Exoskeleton to Assist Children With Cerebral Palsy: Design and Preliminary Evaluation in Level Walking. *IEEE Trans. Neural Syst. Rehabil. Eng.* **2017**, *25*, 906–916. [CrossRef] [PubMed]

42. Mileti, I.; Taborri, J.; Rossi, S.; Petrarca, M.; Patane, F.; Cappa, P. Evaluation of the effects on stride-to-stride variability and gait asymmetry in children with Cerebral Palsy wearing the WAKE-up ankle module. In Proceedings of the 2016 IEEE International Symposium on Medical Measurements and Applications (MeMeA), Benevento, Italy, 15–18 May 2016; pp. 1–6.