

1 Machine learning algorithms in spatiotemporal gait analysis can 2 identify patients with Parkinson's disease.

3 4 Machine learning, gait analysis and identifying Parkinson's gait.

5
6 P. Vinuja R. Fernando^{1,2,3}, Marcus Pannu^{1,2,3}, Pragadesh Natarajan^{1,2,3}, R. Dineth Fonseka^{1,2,3},
7 Naman Singh^{2,3}, Shivanthika Jayalath^{2,3}, Monish M. Maharaj^{1,2,3,4,5}, Ralph J. Mobbs^{1,2,3,4,5},

8
9 ¹Faculty of Medicine, University of New South Wales, Sydney, Australia;

10 ²NeuroSpine Surgery Research Group (NSURG), Sydney, Australia;

11 ³Wearables and Gait Analysis Research Group (WAGAR), Sydney, Australia;

12 ⁴Department of Neurosurgery, Prince of Wales Hospital, Sydney, Australia;

13 ⁵NeuroSpine Clinic, Prince of Wales Private Hospital, Sydney, Australia

14

15 Corresponding author: vfernandoresearch@gmail.com P. Vinuja R. Fernando

16

17

18 Abstract

19 Changes to spatiotemporal gait metrics in gait-altering conditions are characteristic of the
20 pathology. This data can be interpreted by machine learning (ML) models which have recently
21 emerged as an adjunct to clinical medicine. However, the literature is undecided regarding its
22 utility in diagnosing pathological gait and is heterogeneous in its approach to applying ML
23 techniques. This study aims to address these gaps in knowledge. This was a prospective
24 observational study involving 32 patients with Parkinson's disease and 88 'normative' subjects.
25 Spatiotemporal gait metrics were gathered from all subjects using the MetaMotionC inertial
26 measurement unit and data obtained were used to train and evaluate the performance of 10
27 machine learning models. Principal component analysis and Genetic Algorithm were amongst
28 the feature selection techniques used. Classification models included Logistic Regression,
29 Support Vector Machine, Naïve – Bayes, Random Forest, and Artificial Neural Networks. ML
30 algorithms can accurately distinguish pathological gait in Parkinson's disease from that of
31 normative controls. Two models which used the Random Forest classifier with Principal
32 Component analysis and Genetic Algorithm feature selection techniques separately, were 100%
33 accurate in its predictions and had an F_1 score of 1. A third model using principal component
34 analysis and Artificial neural networks was equally as successful (100% accuracy, $F_1 = 1$). We
35 conclude that ML algorithms can accurately distinguish pathological gait from normative
36 controls in Parkinson's Disease. Random Forest classifiers, with Genetic Algorithm feature
37 selection are the preferred ML techniques for this purpose as they produce the highest
38 performing model.

39 NOTE: This preprint reports new research that has not been certified by peer review and should not be used to guide clinical practice.

40

41 **Author summary**

42

43 The way humans walk, are emblematic of their overall health status. These walking patterns,
44 otherwise, can be captured as gait metrics from small and portable wearable sensors. Data
45 gathered from these sensors can be interpreted by machine learning algorithms which can then
46 be used to accurately distinguish healthy and non-healthy patients based on their gait or walking
47 pattern. The applications of this technology are many and varied. Firstly, it can be used to
48 simply aid in diagnosis as explored in this paper. In future, researchers may use their
49 understanding of normal and pathological gait, and their differences to quantify how severely
50 one's gait is affected in a disease state. This data can be used to track, and quantify,
51 improvements or further deteriorations post treatment, whether these be medication-based or
52 interventions like surgery. Retrospective analyses on data such as this can be used to judge the
53 value of an intervention in reducing a patient's disability, and advise health related expenditure.

54

55

56

57

58 **1. Background**

59

60 **1.1. Introduction to Gait analysis**

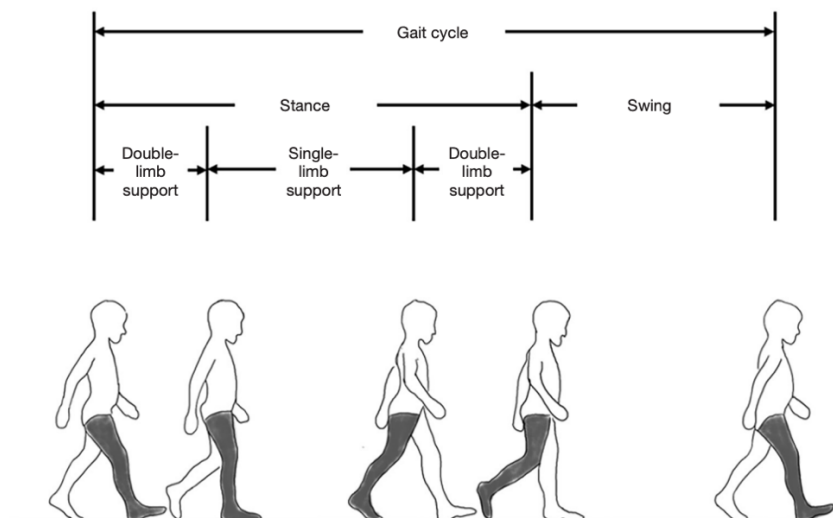
61 Gait refers to the way a person or animal walks or runs and is a simple yet informative measure
62 of overall health. A meta-analysis by Studenski et al(1) showed that with each increment of 0.1
63 m/s in walking speed there was a 12% increase in 10-year survival rate in older adults (HR 0.88,
64 95% CI, 0.87- 0.90; $P < 0.001$)(1). Walking speed as a health metric is not restricted to the
65 context of ageing but can also be predictive of neurological, cardiovascular, orthopaedic, and
66 psychiatric conditions(2-6).

67 Gait, however, is remarkably complex and is not restricted to the metric of walking speed alone.
68 Gait analysis can be subdivided into qualitative and quantitative methods. Qualitative
69 observational methods utilised by clinicians day-to-day are convenient, yet highly subjective
70 and correlate poorly with validated computerised sensors (mean $r = 0.55$)(7). Kinetic data
71 investigates forces involved in locomotion such as ground reaction force. These measures
72 present limited clinical utility(8) and are more suited to the realm of high-performance sports
73 where the focus of gait analysis is not to identify disease states but rather to maximise the
74 efficiency of locomotion(9).

75 In contrast kinematic analyses have shown clinically significant differences in pathological and
76 healthy gait patterns. Table 1 summarises findings from several studies where spatiotemporal
77 parameters in a range of conditions are compared to healthy age-matched controls. Table 1 is
78 merely a snapshot of the unique gait ‘signatures’ of various pathologies which illuminates the
79 diagnostic potential of spatiotemporal gait metrics. For example, appreciable differences can
80 be noted between Parkinson’s disease(10-17) and Lumbar disc herniation(18) in terms of
81 cadence (-6% vs -66%) and double support time (+24% vs +53%) whilst those with Lumbar
82 spinal stenosis(19-23) present with a more modest decrease in cadence (10-14%). Furthermore,
83 statistical models created by Verghese et al. and Lord et al. using spatiotemporal data alone,
84 were able to explain up to 90% of gait variance between healthy and pathological gait using
85 only five factors: pace, rhythm, variability, asymmetry, and postural control(24, 25).

86 A normal gait cycle for each leg involves a stance and a swing phase. Stance (also known as
87 support) phase describes the entire period during which a foot is on the ground, and swing
88 describes the time this same foot is in the air as the limb advances in space. When one limb is

89 in stance, the contralateral limb is in swing, except for an overlapping period where both feet
 90 are on the ground, known as the double support time, as seen in Figure 1.



91

92 **Figure 1:** Gait cycle for right leg (shaded). The figure shows that the gait cycle for any one leg
 93 is comprised of a stance and a swing phase. The right-leg is shaded and used as an example.

94 Figure taken from Natarajan et al.(26)

95

96 The single support time is the period during which only one limb is on the ground.

97 Several other spatiotemporal gait metrics exist and are described in Table 2.

Gait variable	Definition	Units	Type
Step Length	Average distance between two consecutive contacts of any foot with the ground	Metres (m)	Spatial
Stride Length	Average distance between two consecutive contacts of the same foot with the ground.	Metres (m)	Spatial
	Average time between two	Seconds (s)	Temporal

Step time	consecutive contacts of any foot with the ground		
Stride time	Average time between two consecutive contacts of the same foot with the ground	Seconds (s)	Temporal
Walking speed (or gait velocity)	Average distance travelled per second	Metres/second (m/s)	Spatiotemporal
Cadence	Average rate (or frequency) of steps	Steps/minute	Spatiotemporal
Step time variability	Step-to-step variability of step time	Standard deviation (SD) coefficient of variance (cov = SD/mean)	Gait variability
Step length variability	Step-to-step variability of step length	Standard deviation (SD) coefficient of variance (cov = SD/mean)	Gait variability
Walking speed (or gait velocity) variability	Step-to-step variability of walking speed	Coefficient of variance (cov = SD/mean)	Gait variability
Step time asymmetry	Average difference in time taken for	Seconds (s)	Gait asymmetry

	successive steps on left and right foot		
Step length asymmetry	Average difference in length for successive steps on left and right foot	Metres (m)	Gait asymmetry

98

99 **Table 2:** Common spatiotemporal gait metrics. The figure above summarises the most
100 common spatiotemporal metrics. Spatial parameters such as step and stride length can be
101 considered alongside temporal metrics of step and stride time to calculate spatiotemporal data
102 pertaining to gait velocity and cadence. Furthermore, more complex ‘derived’ metrics such as
103 variability and asymmetry in step time, step length and gait velocity can also be calculated.
104 Table adapted from Natarajan et al.(26)

105

106 1.2. Measuring Gait

107 1.2.1. Laboratory Techniques

108 When it comes to gait assessments, optoelectronic stereophotogrammetry is a highly precise
109 laboratory technique and is the gold standard for clinical spatiotemporal gait analysis(27).
110 Infrared cameras capture three-dimensional trajectories of reflective markers placed on points
111 of interest on the subject’s body. However, these require expensive equipment, skilled
112 technicians and are ultimately not feasible for the fast-paced everyday clinical environment(19).
113 Furthermore, these methods are susceptible to the psychological Hawthorne and “white-coat”
114 effects as individuals are more likely to be conscious of their gait when closely observed by a
115 clinician. Hence laboratory techniques fail to capture ‘free-living gait’ which refers to the way
116 people walk in everyday life(19). One study by Brodie et al. highlights this well, finding that
117 lab-based technologies tend to overestimate parameters such as cadence (8.91%, $p < 0.001$)
118 whilst underestimating the variability in gait (81.55%, $p < 0.001$)(28). These drawbacks may
119 limit the validity of the study and decrease the generalisability of the findings.

120 1.2.2. Inertial Measurement Units

121

122 In contrast, inertial measurement units (IMU's) are wearable single-point devices with an
123 accelerometer, magnetometer, and a gyroscope. Measurements made with IMU's have shown
124 to be largely consistent with that of the laboratory analysis techniques ($r > 0.83$). These are very
125 promising as they can capture free-living gait in community and home environments as they
126 are small, inexpensive, and unobtrusive to the activities of daily living(29-31).

127 After measuring gait, scientists are concerned with distinguishing healthy and pathological gait
128 patterns. This has proven to be challenging and the literature shows that mathematical(32, 33)
129 and statistical techniques(34, 35) are popular due to their simplicity. However, purely
130 mathematical transforms provide limited insight as they rely solely on univariate signals and
131 data processed from wavelets, whilst statistical techniques assume normal distributions which
132 tend to oversimplify the complex non-linear relationships in gait data(36, 37). In contrast, recent
133 applications of machine learning (ML), a special subset of artificial intelligence (AI), have
134 shown their ability to model non-linear multidimensional data whilst being versatile in
135 incorporating new data to improve accuracy of predictions(38, 39).

136 **1.3. Machine Learning in Gait Analysis**

137 The workflow in classifying healthy and pathological gait has four key stages.

138 **1.3.1. Feature Selection**

139 Feature selection techniques aim to optimise the model's performance by selecting only the
140 features with maximal separation between classes to ensure the model is both time and cost-
141 efficient(40, 41). Methodologies fall under three categories: filter, wrapper, and embedded
142 methods.

143 Filter methods are the least computationally intensive as they evaluate the dataset without
144 evaluating the performance of the model(40). Wrapper methods are the most computationally
145 intensive as they select features tailored to the performance of the ML model(40). Embedded
146 methods consider both the dataset and the performance of the model with the advantage of
147 being much less computationally intensive than wrapper methods(40).

148 The most common feature selection methods used in gait analysis are Principal Component
149 Analysis (PCA) a filter method, Genetic Algorithm (GA) a wrapper method, and Hill-climbing
150 (HC) an embedded method(42-44). PCA aims to find the minimum number of features or
151 variables required to explain the majority of variance in the data(45). The GA is a different
152 technique which uses the Darwinian theory of natural selection to determine the 'fittest'

153 features. i.e., those that are most discriminative and contribute meaningfully to the performance
154 of the model. Successive iterations of the genetic algorithm are termed ‘generations’ and see
155 the ‘natural selection’ of fitter features and allow for the ‘breeding’ of fit features to form newer
156 and fitter composite features(46) to enhance performance. In contrast, HC is a heuristic search
157 for a solution which maximises the separation between classes but as it is an embedded method,
158 HC may miss the global optimal maximum and instead settle on local maxima. Hence its
159 heuristic nature may provide a sufficient solution in a reasonable amount of time, but this may
160 not be the optimal solution to the classification problem(47).

161 PCA is the simplest technique computationally, and produces the most reliable results(48)
162 (model accuracy >95%) (Table 3). Theoretically speaking, HC is expected to be quite promising
163 as an embedded method and has been highly successful (>96% accuracy) in heart monitors(49).
164 However, it still provides relatively low classification accuracy (75.5-83.3%)(50) with
165 spatiotemporal gait data, showing that its use has not yet been optimised to gait analysis. Further
166 research is recommended to realise its potential in gait analysis.

167 **1.3.2. Classification**

168 Support vector machine (SVM), Naïve-Bayes (NB), and Artificial Neural networks (ANN)
169 were the most common ML models used for classification purposes in the literature.

170 SVM utilises supervised learning methods to compute a hyperplane with greatest separability
171 between the analysed classes(50) whilst NB utilises the Bayes theorem and assumes that all
172 features are independent to create a probabilistic model(51). Finally, ANN’s feature a feed-
173 forward networks where multiple nodes ‘synapse’ upon each other in a layered system, and rely
174 on a ‘transfer-function’ for forward propagation and classification.(52)

175 SVM has shown the greatest success with model accuracies as high as 100%(48) (Table 3). It
176 is also the most used ML model(48, 53-55). NB has been featured sparingly in the literature,
177 and more papers featuring this model are required before its utility can be determined.

178 **1.3.3. Cross Validation**

179 Cross-validation (CV) is used to evaluate the generalisability and external validity of a model
180 by training the algorithm on a training set and evaluating its performance on a validation set(44,
181 50, 56). The most common CV techniques are the k-fold and leave-one-out (LOO) methods as
182 seen in Table 3. K-fold techniques randomly partition data into k subsets where k is an integer.
183 K-1 subsets are used as training subsets, whilst the remaining subset is used to validate the

184 model(50). This is done k times where a different subset is chosen as the validation set on each
185 iteration of the process. LOO methodology uses the same concept except that it is not random
186 as each subset belongs to an individual subject. Consequently, LOO trains the model more
187 rigorously compared to k -fold and introduces levels of complexity which may overfit the model
188 and reduce its external validity. Hence, LOO should be reserved for smaller datasets(48, 53).
189 However, the literature does not indicate an appropriate size for a dataset using LOO and this
190 is likely since it is not only the number of subjects that determines the ‘size’ of the dataset, but
191 also the amount of information associated with each subject. Hence, the size of spatiotemporal
192 gait datasets must be evaluated with both CV techniques before a recommendation can be made.

193 1.3.4. Evaluation of model performance

194 A confusion matrix (Figure 2a) is used to represent the results of the classification model.
195 Metrics such as accuracy, recall, precision, specificity and F1 score can be calculated from the
196 matrix(57).

197 Example Confusion Matrix

198

	Has condition	Does not have condition
Positive test result	<i>TP</i>	<i>FP</i>
Negative test result	<i>FN</i>	<i>TN</i>

199

200 **Figure 2a:** Confusion matrix for a binary classification test. *TN* = true negative, *TP*=true positive,
201 *FN*=false negative, *FP*=false positive.

202

203 Furthermore, the generalisability of the model can be evaluated by the Mean Squared Error
204 (MSE) which is a reflection of the degree of underfitting or overfitting(58). All metrics are
205 summarised in Figure 2b. Whilst accuracy is used as a metric in almost all papers (see Table
206 3), the literature is quite heterogeneous in its use of other metrics. Further research in this field
207 is required in order recommend a more consistent and holistic approach to evaluating model
208 performance as it pertains to clinical use cases.

209 1.4. Research Questions

210 This study will use multiple ML models to distinguish normative subjects from those with
211 Parkinson’s disease using spatiotemporal gait data gathered by wearable IMU’s.

212 **1.4.1. Primary Research Question**

213 Can ML algorithms accurately distinguish patients with Parkinson disease from normative
214 controls?

215 **1.4.2. Secondary Research Question**

216 Which combination of feature selection and classification techniques are most suited to an AI
217 model tasked with gait analysis?

218 **1.4.3. Study Rationale**

219 Spatiotemporal gait data are discriminative of pathologies and IMUs are valid and convenient
220 methods of gathering spatiotemporal data. ML has emerged as a promising adjunct to clinical
221 medicine but has not been optimised for clinical gait analysis. The study aims to determine
222 whether a ML model can accurately distinguish patients suffering from Parkinson disease from
223 normative controls, and the combination of feature selection and classification techniques
224 which are best suited to this purpose.

225 **1.4.4. Study Significance**

226 Such a model would allow for significantly earlier diagnosis of gait-altering pathologies such
227 as Parkinson's disease, compared to current means which depend on clinicians' observational
228 analysis. This will facilitate early intervention, improve long-term outcomes and patient quality
229 of life.

230

231

232

233

234

235

236

237

238 **2. Results**

239 **2.1 Study Population**

240 After cleaning our data prior to applying ML techniques, we excluded data pertaining to 68
241 normative subjects due to missing demographic values, 4 normative subjects due to an
242 IMUGaitPy bug, 8 normative subjects with excessive noise evidenced by their clearly incorrect
243 spatiotemporal parameters. After exclusion of these records, the study population consisted of
244 32 subjects with Parkinson's disease and 88 normative subjects.

245 **2.2 Demographic characteristics**

246 A summary of the demographic characteristics can be found in Table 4b. There were no
247 statistically significant differences in height, weight, BMI, and sex. However significant
248 differences were noted in age, daily step count, smoking, diabetes, cholesterol, and 12-month
249 falls status as well as problems with balance.

250 **2.3 Model Performance**

251 Confusion matrices for the classification for each of the models outlined in Figure 4 are
252 available in Appendix 4. The performance of the model according to metrics outlined in Figure
253 2b are available in Table 5a.

254

255 Models 4, 5 and 9 were the most accurate (100%), sensitive (100%), and had the highest F1
256 score (1.000).

257

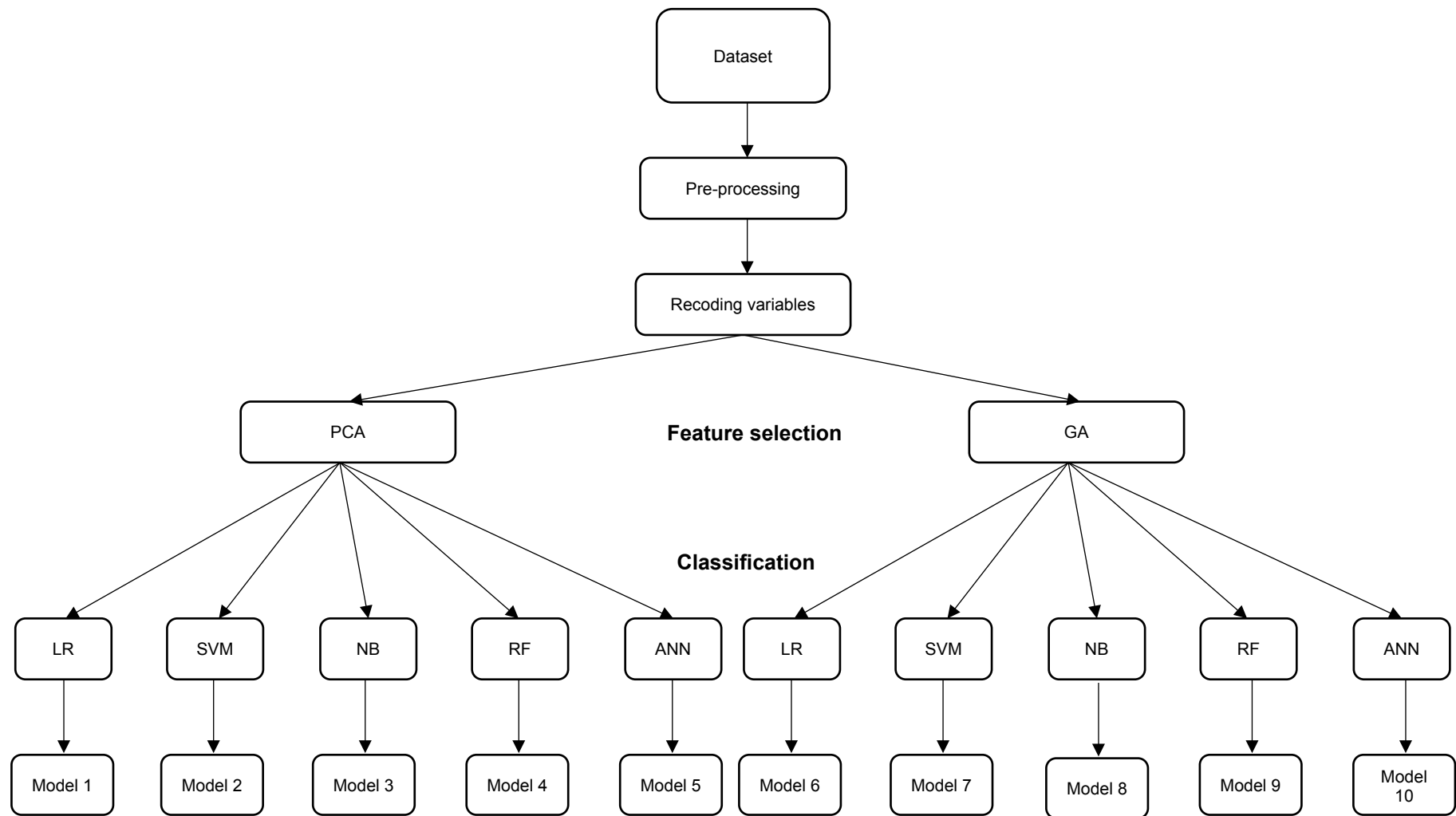


Figure 4: Descriptions of all iterations of machine learning models used for this project. PCA = Principal Component Analysis, GA = Genetic Algorithm, LR = logistic regression, SVM = Support Vector Model, NB = Naïve Bayes, RF = Random Forest, ANN = Artificial Neural Network, LOO = Leave One out. As shown in the figure, the dataset was pre-processed, categorical variables and classification outcome recoded to 0 vs 1 values before feature selection with PCA and GA separately. Each of classification models (five in total) were applied independently to the feature sets selected by PCA and GA to create 10 separate machine learning models.

Model	Feature selection	Classifier	Accuracy (%)	MSE with LOO	MSE with k-fold	Recall	Specificity	Precision	F1 Score
1	PCA	LR	98.33	0.0500	0.0500	0.9375	1.0000	1.0000	0.9677
2		SVM	97.50	0.0833	0.0833	0.9375	0.9886	0.9677	0.9524
3		NB	92.50	0.1417	0.1583	0.7813	0.9886	0.9615	0.8621
4		RF	100.0	0.1083	0.1000	1.0000	1.0000	1.0000	1.0000
5		ANN	100.0	0.0916	0.0666	1.0000	1.0000	1.0000	1.0000
6	GA	LR	94.17	0.0666	0.0917	0.8750	0.9659	0.9032	0.8584
7		SVM	94.17	0.0666	0.0833	0.8750	0.9659	0.9032	0.8889
8		NB	89.17	0.1250	0.1667	0.8125	0.9205	0.7879	0.0801
9		RF	100.0	0.1250	0.1667	1.0000	1.0000	1.0000	1.0000
10		ANN	95.83	0.1333	0.1583	0.9375	0.9886	0.9677	0.9524

Table 5a. Performance metrics of all Machine Learning models. PCA = Principal Component Analysis, GA = Genetic Algorithm, LR = Logistic Regression, SVM = Support Vector Model, NB = Naïve Bayes, RF = Random Forest, ANN = Artificial Neural Network, MSE = Mean squared error, LOO = Leave one out. Recall is otherwise known as sensitivity and Precision is otherwise known as the positive predictive value. The k in k fold has a value k=5 for all models.

See Table 5b for rankings of models according to the aforementioned metrics.

Rank	Accuracy	F1	Sensitivity
1	4 = 5 = 9	4 = 5 = 9	4 = 5 = 9
2			
3			
4	1	1	1
5	2	2 = 10	2
6	10		10
7	6 = 7	7	6 = 7
8		3	
9	3	6	8
10	8	8	3

Table 5b. Models ranked as per the metrics Accuracy, F1 and Sensitivity. Cells are merged and '=' used where models rank equally according to a specific metric. Refer to table 5a to see relevant values.

3. Discussion

Spatiotemporal gait patterns detailed in Table 2 have proven to be sufficiently discriminative of gait altering pathologies such as lumbar spinal stenosis, multiple sclerosis, and Parkinson's disease. Mathematical and statistical techniques have shown their ability to distinguish between healthy and pathological gait(32-35) but are limited in their inability to model the complex non-linear relationships that are inherent to human gait metrics(38, 39).

Recently ML has emerged as a promising new technique which can model both linear and non-linear relationships and is versatile in its ability to incorporate new information to improve the performance of the model. However, this field is still largely in its infancy, especially where it pertains to medicine. The current literature is largely heterogeneous and undecided on the best approach to applying ML techniques to spatiotemporal gait features.

The present study applies a wide range of ML techniques to spatiotemporal gait metrics gathered (using MetaMotionC) from participants with Parkinson's disease and normative subjects. The feature selection techniques will be applied separately to each classification technique as illustrated in Figure 4 to determine the combination of techniques which produces the highest performing model. The aim of the study is to determine the utility of ML in diagnosing pathological gait and finding the combination of ML techniques which produces the highest performing model.

3.1 Justification of study design

3.1.1 Data collection protocols

The present study was inspired by research done by Fonseka et al(59) and Natarajan et al(60) who profiled a variety pathological gait signatures of lumbar spinal stenosis, chronic mechanical lower back pain as well as rheumatological hip and knee conditions. These authors found >92% of agreement between measurements taken from the MetaMotionC and a reference standard (single-camera videography) with an intraclass coefficient >0.86 ($p<0.001$) and was hence deemed valid. Despite other gait analysis studies placing IMUs at the lower back(22, 61-64), wrist(65), ankle(63, 66) or thigh(64), the sternal angle was chosen as the flat surface of the sternum provides a simple and highly repeatable sensor attachment even for unskilled users(59). Accordingly, several studies(29-31) validate chest-based sensor

placements for spatiotemporal metrics by demonstrating high correlation ($r > 0.83$) with optoelectronic stereophotogrammetry which is the current gold standard in gait analysis(27).

3.1.2 Machine learning techniques

The present study utilises PCA and GA feature selection techniques, but omitted HC in its investigation as it is computationally intensive(67) yet performs inconsistently with model accuracies ranging from 75%(54) to 100%(53). Hair et al's(45) recommendation that features chosen by PCA should explain at least 60% of variance in a dataset is widely cited in the literature. Hence the present study uses this notion to choose the top 15 variables which explains 86% of variance in the dataset. In GA, although the ideal population size is specific to the application(46), the literature recommends a larger population size (up to $n=300$ (68)) to allow GA to converge on a robust solution. Since our total population was $n=120$, we applied GA to our entire dataset with $n=50$ generations, to obtain a solution with 11 hybrid or 'mutated' features.

The LR and ANN classifier were fitted using the Limited-memory Broyden-Fletcher-Goldfarb-Shanno algorithm (lbfgs) which is derived from Broyden-Fletcher-Goldfarb-Shanno algorithm (bfgs). Both are mathematical techniques applied to non-linear optimisation problems(69), with lbfgs having the added advantage of reduced runtime and memory usage. The lbfgs has been validated by biomechanics papers which found a limited increase in model performance for a considerably larger investment of computational power(70, 71) with the original bfgs. The SVM was fitted using a linear kernel which is known for its shorter runtime and is preferred for datasets with many features(72). For RF models, the literature recommends using 64-128 trees as a tradeoff between high ROC AUC values and processing time(73). Hence, we utilised 100 decision trees which were merged to increase the accuracy of predictions. Whilst the number of hidden layers possible for an ANN are unlimited, a higher number of layers incurs greater computational costs. The author decided on a moderate number of hidden layers ($n=6$) to increase the accuracy of predictions with a reasonable computational cost. This is based off previous applications in biochemistry and genetics(74-76) which used a similar number of features to the present study ($n=10-16$).

3.2 Evaluating Model performance

3.2.1 Metrics used to evaluate model performance

The literature is heterogeneous in the metrics used to report the performance of ML models. Different combinations of the metrics in Figure 2b have been used(48, 50, 53-55). For example, Eskofier et al(48) and Pogorelc et al(55) report only accuracy, whilst Begg et al(54) and Khandoker et al(53) report accuracy, recall (sensitivity) and precision (positive predictive value). Reporting accuracy alone is problematic with an imbalanced dataset when the condition has a low prevalence(77) and can lead to misleading conclusions. Hence recall is useful as it quantifies the true-positive rate whilst precision reflects the false positive rate ($FPR = 1 - \text{precision}$). Ideally, a good test has a high sensitivity, so as not to miss subjects suffering from a condition, but also has a high precision (low FPR) so as not to incur additional costs to the healthcare system by necessitating clinic visits for healthy individuals(78).

However, models which have high recall, do not necessarily have high precision. As seen in Table 5b, models can be ranked differently by different metrics. For example, Model 3 in this present study is more precise than model 8 (0.9615 vs 0.7879) but has a lower recall (0.7813 vs 0.9310). Here the F_β score is useful (Figure 2b) as it is a composite metric of both recall and precision. The β -parameter controls the tradeoff of importance between recall and precision. $\beta < 1$ focuses on precision, $\beta > 1$ focuses on recall and $\beta = 1$ assigns equal importance to both. The use of the F_1 score ($\beta = 1$) has not yet been used in the literature concerning gait analysis but has proven to be insightful in studies related to COVID-19(79) as well as the wider statistical literature(80-83). The F_1 score is suitable to the present study where the maximisation of true positives and minimisation of false positives are of equal importance.

$$Accuracy = \frac{TP + TN}{TP + TN + FP + FN} \times 100\%$$

$$Recall = \frac{TP}{TP + FN} \times 100\%$$

$$Specificity = \frac{TN}{TP + FP} \times 100\%$$

$$Precision = \frac{TP}{TP + FP} \times 100\%$$

$$F_{\beta} = (1 + \beta^2) \times \frac{(\text{precision} \times \text{recall})}{(\beta^2 \cdot \text{precision}) + \text{recall}} \times 100\%$$

$$F_1 = 2 \times \frac{(\text{precision} \times \text{recall})}{(\text{precision} + \text{recall})} \times 100\%$$

$$MSE = (y_{true}, y_{predicted}) = \frac{1}{n_{samples}} \sum (y_{true} - y_{predicted})^2$$

Figure 2b: Overview of metrics used to analyse the performance of machine learning models. TN = true negative, TP=true positive, FN=false negative, FP=false positive. Recall is also known as Sensitivity and Precision is otherwise known as the Positive predictive value. MSE = Mean Squared Error. y_{true} = true value, $y_{predicted}$ = predicted value. MSE is a measure of the average of the squared error calculated when comparing the true and predicted values. Note that the F_1 score is fundamentally the F_{β} score when β is assigned a value of 1.

Furthermore, the Mean-squared Error (MSE) is obtained (Figure 2b) after performing cross-validation techniques such as k-fold and LOO. Whilst ML studies (see Table 3) all perform cross validation; none report the error in any form, whether it be MSE, Mean absolute error (MAE) or others. The MSE is a representation of the degree of bias in a model(84). A highly complex and overfitted model tends to be less biased towards its training data, have a lower MSE, but in turn these models show greater variance with external data, are less generalisable and have poor external validity. The opposite is true for models with higher error. If they retain

a high classification accuracy and F_1 score, a higher error value is desirable as it means that the model is more biased towards its training data, less overfitted, less likely to show variance with external data, generalisable and clinically useful(85).

Hence, the author recommends the combined use of accuracy and F_1 score to judge the performance of ML models considering its MSE after cross-validation which is an indication of the bias-variance tradeoff(85).

3.2.2 Performance of models

The present study finds that there is a high classification accuracy amongst all models (>89%). Models 4,5 and 9 (Table 5a) are the highest performers with 100% accuracy, and F_1 score of 1, which is the highest possible score. Out of these Model 9 performs best as it has the highest MSE (0.125) after cross-validation and is likely to have higher bias in favour of lower variance, thus greater generalisability and greater external validity. Following this, are models 1,2 and 10 which are ranked highest to lowest in terms of accuracy *and* F_1 score. The remaining models cannot be ranked as they perform inconsistently based on accuracy and F_1 score.

The success of Models 4 and 9 which use RF is consistent with Arora et al(86) who used the tri-axial accelerometry data from smartphones to distinguish participants with Parkinson's disease from normative controls. The models had an average sensitivity of 98.5% and specificity of 97.5%. The likely reason for the slightly lower performance of their model is due to the lack of features used in their analysis as well as the lack of a feature selection technique. The MetaMotionC has not only an accelerometer but a magnetometer and gyroscope and inevitably, the present study works with more features. In summary, our findings are consistent with the literature and suggest that the RF classifier is promising in gait analysis. The author recommends the use of a feature selection technique, namely GA (Model 9) in combination with RF to increase the performance of the model.

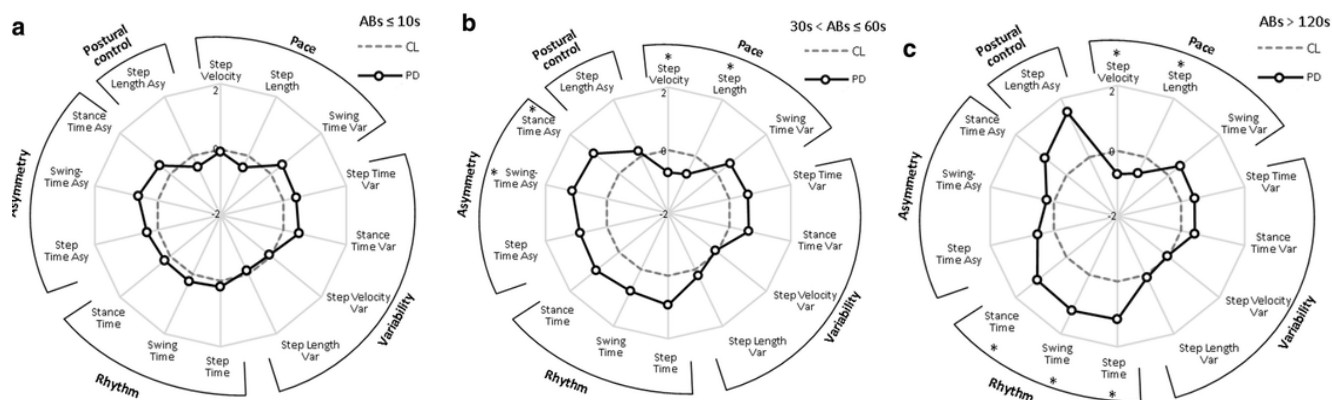


Figure 5: Radar plot illustrating 14 spatiotemporal gait metrics for patients with Parkinson's Disease (PD) and controls (CL) as evaluated ambulatory bouts (ABs) in free-living contexts. Central dotted line represents CL data and bolded line represents PD data measured in standard deviations from CL values (range $\pm 2SD$). (a) represents $Abs < 10s$, (b) represents $30s < Abs < 60s$ and (c) represents $Abs > 120s$. Figure taken from Del Din et al.(87)

In comparison, Model 5 featuring ANN, greatly outperforms a recent study by Iosa et al(88) who used a very similarly capable IMU and hence had access to a very similar feature set. It is understood that this team did not apply any feature selection techniques and that their participants only walked for 10m in data collection. A 2016 study by Del Din et al.(87) found that longer ambulatory bouts were more discriminative of pathological gait (Figure 5). The present study utilises a minimum walking distance of 50m and is a likely reason for our increased classification performance. Hence, we find that ANN is a good classifier in spatiotemporal gait analysis but should be used with a feature selection technique and longer ambulatory bouts for more valid predictions.

Both models 3 and 8 performed poorly (accuracies 92.5% and 89.1%) in the present study and performs similarly poorly in other models using sensor-based data(89, 90). However, a study by Pogorelc et al(55) which used video-analysis as opposed to sensor-based techniques in gait analysis achieved a 97.2% classification accuracy. No feature selection techniques were used, indicating that classification accuracies could be further increased. Early interpretations may suggest that NB classifiers are more suited to visual data compared to sensor-based data, but further research is necessary to make a firm conclusion.

3.3 Significance of findings

The findings confirm that ML algorithms can accurately distinguish pathological from healthy gait. This field is still largely in its infancy and extant literature is heterogeneous in its approach to the use of ML techniques. Through this study, we have contributed to the existing knowledge by showing that feature selection improves performance and should hence be used routinely hereon. Furthermore, we have shown that RF classifiers in conjunction with GA outperform other spatiotemporal gait analysis models which combine other techniques. In addition, through our analysis we add to the extant literature by recommending the routine use of accuracy and F_1 score to evaluate model performance.

3.4 Strengths and limitations

The main strength of this study is in the wide scope of techniques investigated. By combining two feature selection techniques iteratively with five different classifiers, we were able to form 10 different models to make a comprehensive recommendation on the combination of methods best suited to distinguishing pathological from healthy gait using spatiotemporal gait data.

The main limitation is in the statistically significant age difference between the Parkinson's and normative groups. This makes age a confounding variable which may obscure the 'true' impact of the pathology(91) and limit the internal validity of the study. This arose largely due to difficulty obtaining older subjects who satisfied the inclusion criteria for the normative group.

In addition, the lack of an external validation dataset precludes determination of the generalisability and external validity of the model. Cross-validation techniques and MSE values calculated are only a prediction of the likely generalisability. CV techniques are common in the literature because models are often 'bootstrapped' for data. Increasing the size of the dataset, would allow researchers to have a separate validation dataset that is not used at all in training the model(92).

4. Future directions and Conclusion

Firstly, the number of older participants should be increased in a follow-up study. Participants should be stratified by age to reduce the confounding influence. Secondly, we aim to introduce a second pathological group such as patients with lumbar spinal stenosis to evaluate the

performance of the model in a three-way classification problem similar to that conducted by Mannini et al who achieved 90.5% accuracy in classifying elderly subjects from Post-stroke and Huntington's disease patients using Support Vector Machines (SVM)(93).

In addition, the utility of models described in this paper must be determined by examining the degree of disease progression by assessing the severity of gait deterioration. Similarly, the team aim to investigate whether ML models can quantify patients' response to therapy. For example, whether a model distinguish a Parkinson's patient before and after they take medication.

In conclusion, this study found that ML algorithms in combination with feature selection techniques could accurately distinguish pathological from healthy gait. In relation to Parkinson's disease the findings suggest that a RF classifier paired with the GA feature selection is the best performing model with 100% accuracy and F_1 score.

These findings are invaluable considering that such a tool can allow early diagnosis of conditions such as Parkinson's disease, facilitate early intervention and increase patient outcomes and quality of life.

Future research should have larger datasets stratified with age and construct a model that is not only able to distinguish Parkinson's patients from healthy ones but also from patients suffering from other gait-altering pathologies (e.g., post-stroke, lumbar spinal stenosis).

5. Materials and Methods

5.1 Objectives

The present study is an observational case-control study of participants with Parkinson's disease who were compared to healthy controls. Spatiotemporal gait metrics summarised in Table 2 were collected from both groups using an IMU and several ML models were used to classify the study population based on whether they suffer from Parkinson's disease.

5.2 Ethics

Approval was obtained from the South-Eastern Sydney Local Health District, New South Wales, Australia (HREC 17/184). All participants provided written informed consent.

5.3 Study Population

A total of 168 normative subjects and 32 participants with Parkinson's Disease were recruited for the study. Details regarding the locations from which participants were recruited as well as age ranges can be found in Table 4a.

Inclusion criteria for normative subjects included being older than 18 years of age and inclusion criteria for the group with Parkinson's disease included being older than 18 years of age and a clinical diagnosis of Parkinson's disease.

Exclusion criteria for both groups included a BMI greater than 25, inability to walk at least 50m independently, women who are pregnant and any concurrent gait altering pathologies including but not restricted to stroke, lumbar spinal stenosis, multiple sclerosis, rheumatological conditions of hip, knee and spine and cauda equina syndrome.

5.4 Data collection

Participants provided informed written consent after which they were interviewed to obtain demographic data summarised in Table 4b. The wearable IMU used was the MetaMotionC developed by Mbientlab Inc. and contains a 16bit triaxial accelerometer (100Hz), gyroscope (100Hz), and $0.3\mu T$ magnetometer (25Hz). Participants were fitted with this sensor at the sternal angle (Figure 3) and following a short pause to orient the device, instructed to walk 50m, unobserved, along a flat concrete pathway, at their natural walking pace. Data was downloaded via Bluetooth™ to an Android™ smartphone running the IMUGait Recorder application which was developed for this study. IMUGaitPY, a modified version of the open-source GaitPY Python(94) package by Czech and Patel was used to extract spatiotemporal gait metrics (Table 2) from the raw data. Appendix 1 elaborates on this process. Setup instructions for IMUGaitPY as well as details regarding configuration files and mathematical derivations can be found in Appendix B.

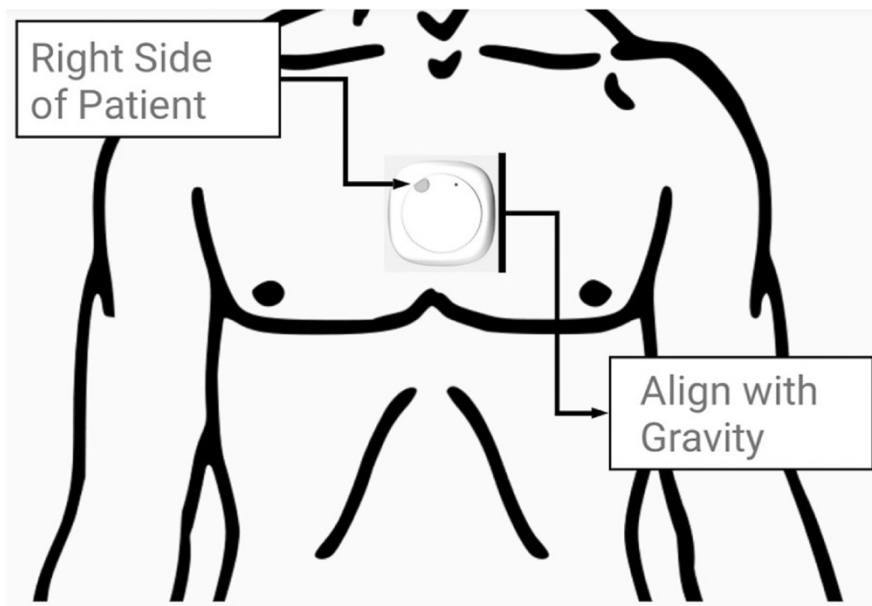


Figure 3: The MetaMotion the MetaMotionC® (MMC) inertial measurement unit (IMU) developed by Mbientlab Inc. pictured as it will be fitted on the sternal angle of patients. Figure taken from Natarajan et al(26).

5.5 Data Analysis

5.5.1 Demographic variables

Demographic data were assessed for normality using the Shapiro-Wilk test and visual inspection of histograms. Continuous variables such as age, height, weight, BMI were compared between groups using the independent sample t-test for normal data and the Mann-Whitney U test for non-normal data. Categorical variables such as sex, smoking, diabetes, hypertension, cholesterol, and 12-month falls status were compared using the Chi-square test of independence. The level of statistical significance was set to $p=0.05$ and analysis was performed using IBM SPSS Statistics Version 26.0 (IBM, New York, United States).

5.5.2 Machine learning models

Pre-processing

The dataset was cleaned by removing duplicate records and records with missing values. Structural errors such as spelling mistakes were corrected as they have the potential to return error codes. Following this, the data was standardised to remove outliers.

Recoding variables and outcomes

In preparation for binary classification, normative (healthy) subjects were assigned a value of 0 whilst those with Parkinson's disease were assigned a value of 1. Similarly, categorical demographic variables such as smoking, diabetes, hypertension, cholesterol, and 12-month falls status which were previously answered as yes or no, were recoded to be 1 and 0 respectively.

Feature selection

Principal Component Analysis (PCA) was used to reduce the features (variables). The sum of the first 15 features (from a total of 75) explained over 86% of variance and was deemed sufficient to represent the data.

Separately, the Genetic Algorithm (GA) reduced the dataset to the 11 most descriptive features.

Classification

Classification models used include Logistic regression, Support Vector Model (SVM), Naïve Bayes classifier (NB), Random Forest (RF) and Artificial Neural Network (ANN). Each of these models were applied separately to each of the reduced feature sets determined by PCA and GA to create 10 separate ML models in total. The process thus far is summarised in Figure 4.

The LR model was fitted using the Limited-memory Broyden-Fletcher-Goldfarb-Shanno (lbfgs) optimisation algorithm. The SVM was fitted using a linear kernel whilst the NB classifier was applied using the Gaussian Naïve Bayes method. The RF model utilized 100 decision trees which were merged to increase the accuracy of predictions. Finally, the connected multi-layer artificial neural network (ANN) multilayer perceptron was also fitted with the lbfgs optimisation algorithm and included six hidden layers (n=6).

Cross-Validation

All models were validated independently using both the k-fold (k value set to 5) and leave-one-out (LOO) techniques.

Evaluating performance

All metrics outlined in Figure 2b were used to evaluate the performance of the models.

The models above were coded using Jupyter Notebook, an open-source software (Project Jupyter, 2014). See Appendix 3 for the full code.

Conflict of interest

The authors declare no conflict of interest.

Acknowledgments/ Funding information

V.Fernando was responsible for data collection, curation and analysis as well as writing the manuscript. RJM and MMM were crucial in the conceptualisation of the study. RJM, MMM, RDF and PN all provided editorial input into the manuscript through their reviews. MP and NS aided also in data collection. SMJ provided all patients with Parkinson's disease for assessment through her clinics. This study was not funded.

REFERENCES

1. Studenski S, Perera S, Patel K, Rosano C, Faulkner K, Inzitari M, et al. Gait Speed and Survival in Older Adults. *JAMA*. 2011;305(1):50-8.
2. Brandler TC, Wang C, Oh-Park M, Holtzer R, Verghese J. Depressive symptoms and gait dysfunction in the elderly. *Am J Geriatr Psychiatry*. 2012;20(5):425-32.
3. Dumurgier J, Elbaz A, Ducimetière P, Tavernier B, Alperovitch A, Tzourio C. Slow walking speed and cardiovascular death in well functioning older adults: prospective cohort study. *Bmj*. 2009;339:b4460.
4. Perry J, Garrett M, Gronley JK, Mulroy SJ. Classification of walking handicap in the stroke population. *Stroke*. 1995;26(6):982-9.
5. Hollman JH, Beckman BA, Brandt RA, Merriwether EN, Williams RT, Nordrum JT. Minimum detectable change in gait velocity during acute rehabilitation following hip fracture. *J Geriatr Phys Ther*. 2008;31(2):53-6.
6. Motyl JM, Driban JB, McAdams E, Price LL, McAlindon TE. Test-retest reliability and sensitivity of the 20-meter walk test among patients with knee osteoarthritis. *BMC Musculoskelet Disord*. 2013;14:166.
7. McGinley JL, Goldie PA, Greenwood KM, Olney SJ. Accuracy and Reliability of Observational Gait Analysis Data: Judgments of Push-off in Gait After Stroke. *Physical Therapy*. 2003;83(2):146-60.
8. Dicharry J. Kinematics and kinetics of gait: from lab to clinic. *Clin Sports Med*. 2010;29(3):347-64.
9. Hollis CR, Koldenhoven RM, Resch JE, Hertel J. Running biomechanics as measured by wearable sensors: effects of speed and surface. *Sports Biomech*. 2021;20(5):521-31.
10. Godi M, Arcolin I, Giardini M, Corna S, Schieppati M. A pathophysiological model of gait captures the details of the impairment of pace/rhythm, variability and asymmetry in Parkinsonian patients at distinct stages of the disease. *Scientific Reports*. 2021;11(1).

11. Schlachetzki JCM, Barth J, Marxreiter F, Gossler J, Kohl Z, Reinfelder S, et al. Wearable sensors objectively measure gait parameters in Parkinson's disease. *PLoS one*. 2017;12(10):e0183989.
12. Morris R, Hickey A, Del Din S, Godfrey A, Lord S, Rochester L. A model of free-living gait: A factor analysis in Parkinson's disease. *Gait Posture*. 2017;52:68-71.
13. Espay AJ, Bonato P, Nahab FB, Maetzler W, Dean JM, Klucken J, et al. Technology in Parkinson's disease: Challenges and opportunities. *Mov Disord*. 2016;31(9):1272-82.
14. Muro-de-la-Herran A, Garcia-Zapirain B, Mendez-Zorrilla A. Gait analysis methods: an overview of wearable and non-wearable systems, highlighting clinical applications. *Sensors (Basel, Switzerland)*. 2014;14(2):3362-94.
15. Geroin C, Nonnekes J, de Vries NM, Strouwen C, Smania N, Tinazzi M, et al. Does dual-task training improve spatiotemporal gait parameters in Parkinson's disease? *Parkinsonism Relat Disord*. 2018;55:86-91.
16. Zhang S, Qian J, Zhang Z, Shen L, Wu X, Hu X. Age- and Parkinson's disease-related evaluation of gait by General Tau Theory. *Experimental Brain Research*. 2016;234(10):2829-40.
17. Maetzler W, Klucken J, Horne M. A clinical view on the development of technology-based tools in managing Parkinson's disease. *Mov Disord*. 2016;31(9):1263-71.
18. Bonab M, Colak TK, Toktas ZO, Konya D. Assessment of Spatiotemporal Gait Parameters in Patients with Lumbar Disc Herniation and Patients with Chronic Mechanical Low Back Pain. *Turkish neurosurgery*. 2020;30(2):277-84.
19. Perring J, Mobbs R, Betteridge C. Analysis of Patterns of Gait Deterioration in Patients with Lumbar Spinal Stenosis. *World Neurosurg*. 2020;141:e55-e9.
20. Loske S, Nüesch C, Byrnes KS, Fiebig O, Schären S, Mündermann A, et al. Decompression surgery improves gait quality in patients with symptomatic lumbar spinal stenosis. *Spine J*. 2018;18(12):2195-204.
21. Sun J, Liu Y-c, Yan S-h, Wang S-s, Lester DK, Zeng J-z, et al. Clinical Gait Evaluation of Patients with Lumbar Spine Stenosis. *Orthopaedic Surgery*. 2018;10(1):32-9.
22. Papadakis NC, Christakis DG, Tzagarakis GN, Chlouverakis GI, Kampanis NA, Stergiopoulos KN, et al. Gait variability measurements in lumbar spinal stenosis patients: part A. Comparison with healthy subjects. *Physiol Meas*. 2009;30(11):1171-86.
23. Odonkor C, Kuwabara A, Tomkins-Lane C, Zhang W, Muaremi A, Leutheuser H, et al. Gait features for discriminating between mobility-limiting musculoskeletal disorders: Lumbar spinal stenosis and knee osteoarthritis. *Gait Posture*. 2020;80:96-100.
24. Verghese J, Robbins M, Holtzer R, Zimmerman M, Wang C, Xue X, et al. Gait dysfunction in mild cognitive impairment syndromes. *J Am Geriatr Soc*. 2008;56(7):1244-51.
25. Lord S, Galna B, Verghese J, Coleman S, Burn D, Rochester L. Independent domains of gait in older adults and associated motor and nonmotor attributes: validation of a factor analysis approach. *J Gerontol A Biol Sci Med Sci*. 2013;68(7):820-7.
26. Natarajan P, Fonseka RD, Kim S, Betteridge C, Maharaj M, Mobbs RJ. Analysing gait patterns in degenerative lumbar spine diseases: a literature review. *Journal of Spine Surgery*. 2022.
27. Cappozzo A, Dellacroce U, Leardini A, Chiari L. Human movement analysis using stereophotogrammetry: Part 1: Theoretical background. *Gait & Posture - GAIT POSTURE*. 2005;21:186-96.

28. Brodie MA, Coppens MJ, Lord SR, Lovell NH, Gschwind YJ, Redmond SJ, et al. Wearable pendant device monitoring using new wavelet-based methods shows daily life and laboratory gaits are different. *Med Biol Eng Comput*. 2016;54(4):663-74.
29. Washabaugh EP, Kalyanaraman T, Adamczyk PG, Claflin ES, Krishnan C. Validity and repeatability of inertial measurement units for measuring gait parameters. *Gait & Posture*. 2017;55:87-93.
30. Kluge F, Gaßner H, Hannink J, Pasluosta C, Klucken J, Eskofier BM. Towards Mobile Gait Analysis: Concurrent Validity and Test-Retest Reliability of an Inertial Measurement System for the Assessment of Spatio-Temporal Gait Parameters. *Sensors (Basel, Switzerland)*. 2017;17(7):1522.
31. Rantalainen T, Pirkola H, Karavirta L, Rantanen T, Linnamo V. Reliability and concurrent validity of spatiotemporal stride characteristics measured with an ankle-worn sensor among older individuals. *Gait Posture*. 2019;74:33-9.
32. Ismail AR, Asfour SS. Discrete wavelet transform: a tool in smoothing kinematic data. *J Biomech*. 1999;32(3):317-21.
33. Mezghani N, Husse S, Boivin K, Turcot K, Aissaoui R, Hagemester N, et al. Automatic classification of asymptomatic and osteoarthritis knee gait patterns using kinematic data features and the nearest neighbor classifier. *IEEE Trans Biomed Eng*. 2008;55(3):1230-2.
34. Jones L, Beynon MJ, Holt CA, Roy S. An application of the Dempster-Shafer theory of evidence to the classification of knee function and detection of improvement due to total knee replacement surgery. *J Biomech*. 2006;39(13):2512-20.
35. Takahashi T, Ishida K, Hirose D, Nagano Y, Okumiya K, Nishinaga M, et al. Trunk deformity is associated with a reduction in outdoor activities of daily living and life satisfaction in community-dwelling older people. *Osteoporos Int*. 2005;16(3):273-9.
36. Chau T. A review of analytical techniques for gait data. Part 2: neural network and wavelet methods. *Gait Posture*. 2001;13(2):102-20.
37. Su FC, Wu WL. Design and testing of a genetic algorithm neural network in the assessment of gait patterns. *Med Eng Phys*. 2000;22(1):67-74.
38. Abujrida H, Agu E, Pahlavan K. Machine learning-based motor assessment of Parkinson's disease using postural sway, gait and lifestyle features on crowdsourced smartphone data. *Biomed Phys Eng Express*. 2020;6(3):035005.
39. Aich S, Pradhan PM, Chakraborty S, Kim HC, Kim HT, Lee HG, et al. Design of a Machine Learning-Assisted Wearable Accelerometer-Based Automated System for Studying the Effect of Dopaminergic Medicine on Gait Characteristics of Parkinson's Patients. *Journal of Healthcare Engineering*. 2020;2020.
40. Saeys Y, Inza I, Larrañaga P. A review of feature selection techniques in bioinformatics. *Bioinformatics*. 2007;23(19):2507-17.
41. Zhang J, Lockhart TE, Soangra R. Classifying Lower Extremity Muscle Fatigue During Walking Using Machine Learning and Inertial Sensors. *Annals of Biomedical Engineering*. 2014;42(3):600-12.
42. Lu Y, Boukharouba K, Boonært J, Fleury A, Lecœuche S. Application of an incremental SVM algorithm for on-line human recognition from video surveillance using texture and color features. *Neurocomputing*. 2014;126:132-40.
43. Martins M, Santos C, Costa L, Frizera A. Feature reduction with PCA/KPCA for gait classification with different assistive devices. *INTERNATIONAL JOURNAL OF INTELLIGENT COMPUTING AND CYBERNETICS*. 2015;8(4):363-82.

44. Martins M, Costa L, Frizera A, Ceres R, Santos C. Hybridization between multi-objective genetic algorithm and support vector machine for feature selection in walker-assisted gait. *Computer Methods and Programs in Biomedicine*. 2014;113(3):736-48.
45. Hair J, Sarstedt M, Pieper T, Ringle C. The Use of Partial Least Squares Structural Equation Modeling in Strategic Management Research: A Review of Past Practices and Recommendations for Future Applications. *Long Range Planning*. 2012;45:320-40.
46. Rajakumar BR, George A. APOGA: An Adaptive Population Pool Size based Genetic Algorithm. *AASRI Procedia*. 2013;4:288-96.
47. Pereda E, García-Torres M, Melián-Batista B, Mañas S, Méndez L, González JJ. The blessing of Dimensionality: Feature Selection outperforms functional connectivity-based feature transformation to classify ADHD subjects from EEG patterns of phase synchronisation. *PLOS ONE*. 2018;13(8):e0201660.
48. Eskofier BM, Federolf P, Kugler PF, Nigg BM. Marker-based classification of young–elderly gait pattern differences via direct PCA feature extraction and SVMs. *Computer Methods in Biomechanics and Biomedical Engineering*. 2013;16(4):435-42.
49. Ashfaq Z, Mumtaz R, Rafay A, Zaidi SM, Saleem H, Mumtaz S, et al. Embedded AI-Based Digi-Healthcare. *Applied Sciences*. 2022;12(1).
50. Begg R, Kamruzzaman J. A machine learning approach for automated recognition of movement patterns using basic, kinetic and kinematic gait data. *Journal of Biomechanics*. 2005;38(3):401-8.
51. Badesa FJ, Morales R, Garcia-Aracil N, Sabater JM, Casals A, Zollo L. Auto-adaptive robot-aided therapy using machine learning techniques. *Computer Methods and Programs in Biomedicine*. 2014;116(2):123-30.
52. Ardestani MM, Moazen M, Jin Z. Gait modification and optimization using neural network–genetic algorithm approach: Application to knee rehabilitation. *Expert Systems with Applications*. 2014;41(16):7466-77.
53. Khandoker AH, Lai DTH, Begg RK, Palaniswami M. Wavelet-Based Feature Extraction for Support Vector Machines for Screening Balance Impairments in the Elderly. *IEEE Transactions on Neural Systems and Rehabilitation Engineering*. 2007;15(4):587-97.
54. Begg R, Palaniswami M, Owen B. Support Vector Machines for Automated Gait Classification. *IEEE transactions on bio-medical engineering*. 2005;52:828-38.
55. Pogorelc B, Bosnić Z, Gams M. Automatic recognition of gait-related health problems in the elderly using machine learning. *Multimedia Tools and Applications*. 2012;58(2):333-54.
56. López V, Fernández A, Herrera F. On the importance of the validation technique for classification with imbalanced datasets: Addressing covariate shift when data is skewed. *Information Sciences*. 2014;257:1-13.
57. Shanbehzadeh M, Kazemi-Arpanahi H, Orooji A, Mobarak S, Jelvey S. Performance evaluation of selected machine learning algorithms for COVID-19 prediction using routine clinical data: With versus Without CT scan features. *J Educ Health Promot*. 2021;10:285.
58. Emil Eskildsen C, Næs T. Sample-Specific Prediction Error Measures in Spectroscopy. *Appl Spectrosc*. 2020;74(7):791-8.
59. Fonseka RD, Natarajan P, Maharaj M, Mobbs R. Tracking the disease progression of lumbar spinal stenosis using objective gait metrics: a case report. *Journal of Spine Surgery*. 2021;8.

60. Natarajan P, Fonseka RD, Sy L, Maharaj M, Mobbs R. Analysing Gait Patterns in Degenerative Lumbar Spine Disease Using Inertial Wearable Sensors: An Observational Study. *World Neurosurgery*. 2022;163.
61. Romijnders R, Warmerdam E, Hansen C, Welzel J, Schmidt G, Maetzler W. Validation of IMU-based gait event detection during curved walking and turning in older adults and Parkinson's Disease patients. *Journal of NeuroEngineering and Rehabilitation*. 2021;18(1):28.
62. Mancini M, Chiari L, Holmstrom L, Salarian A, Horak FB. Validity and reliability of an IMU-based method to detect APAs prior to gait initiation. *Gait Posture*. 2016;43:125-31.
63. Hansen C, Beckbauer M, Romijnders R, Warmerdam E, Welzel J, Geritz J, et al. Reliability of IMU-Derived Static Balance Parameters in Neurological Diseases. *Int J Environ Res Public Health*. 2021;18(7).
64. Hsu W-C, Sugiarto T, Lin Y-J, Yang F-C, Lin Z-Y, Sun C-T, et al. Multiple-Wearable-Sensor-Based Gait Classification and Analysis in Patients with Neurological Disorders. *Sensors*. 2018;18(10):3397.
65. Tripuraneni KR, Foran JRH, Munson NR, Racca NE, Carothers JT. A Smartwatch Paired With A Mobile Application Provides Postoperative Self-Directed Rehabilitation Without Compromising Total Knee Arthroplasty Outcomes: A Randomized Controlled Trial. *J Arthroplasty*. 2021;36(12):3888-93.
66. Baghdadi A, Cavuoto LA, Crassidis JL. Hip and Trunk Kinematics Estimation in Gait Through Kalman Filter Using IMU Data at the Ankle. *IEEE Sensors Journal*. 2018;18(10):4253-60.
67. Chira C, Horvath D, Dumitrescu D. Hill-Climbing search and diversification within an evolutionary approach to protein structure prediction. *BioData Min*. 2011;4:23.
68. Odetayo MO, editor Optimal population size for genetic algorithms: an investigation. *IEE Colloquium on Genetic Algorithms for Control Systems Engineering*; 1993 28-28 May 1993.
69. Guo J, Wan Z. Two Modified Single-Parameter Scaling Broyden–Fletcher–Goldfarb–Shanno Algorithms for Solving Nonlinear System of Symmetric Equations. *Symmetry [Internet]*. 2021; 13(6).
70. Liu X, Belcher AH, Grelewicz Z, Wiersma RD. Robotic real-time translational and rotational head motion correction during frameless stereotactic radiosurgery. *Med Phys*. 2015;42(6):2757-63.
71. Liu X, Wiersma RD. Optimization based trajectory planning for real-time 6DoF robotic patient motion compensation systems. *PLOS ONE*. 2019;14(1):e0210385.
72. Ben-Hur A, Ong CS, Sonnenburg S, Schölkopf B, Rätsch G. Support vector machines and kernels for computational biology. *PLoS Comput Biol*. 2008;4(10):e1000173.
73. Denisko D, Hoffman MM. Classification and interaction in random forests. *Proc Natl Acad Sci U S A*. 2018;115(8):1690-2.
74. Dias R, Torkamani A. Artificial intelligence in clinical and genomic diagnostics. *Genome Medicine*. 2019;11(1):70.
75. Vaz JM, Balaji S. Convolutional neural networks (CNNs): concepts and applications in pharmacogenomics. *Mol Divers*. 2021;25(3):1569-84.
76. Cao C, Liu F, Tan H, Song D, Shu W, Li W, et al. Deep Learning and Its Applications in Biomedicine. *Genomics Proteomics Bioinformatics*. 2018;16(1):17-32.

77. Jeni LA, Cohn JF, De La Torre F. Facing Imbalanced Data Recommendations for the Use of Performance Metrics. *Int Conf Affect Comput Intell Interact Workshops*. 2013;2013:245-51.
78. Trevethan R. Sensitivity, Specificity, and Predictive Values: Foundations, Plabilities, and Pitfalls in Research and Practice. *Front Public Health*. 2017;5:307.
79. Alakus TB, Turkoglu I. Comparison of deep learning approaches to predict COVID-19 infection. *Chaos Solitons Fractals*. 2020;140:110120.
80. Seo S, Kim Y, Han HJ, Son WC, Hong ZY, Sohn I, et al. Predicting Successes and Failures of Clinical Trials With Outer Product-Based Convolutional Neural Network. *Front Pharmacol*. 2021;12:670670.
81. Hicks SA, Strümke I, Thambawita V, Hammou M, Riegler MA, Halvorsen P, et al. On evaluation metrics for medical applications of artificial intelligence. *medRxiv*. 2021:2021.04.07.21254975.
82. Chicco D, Jurman G. The advantages of the Matthews correlation coefficient (MCC) over F1 score and accuracy in binary classification evaluation. *BMC Genomics*. 2020;21(1):6.
83. Hermine O, Mariette X, Tharaux P-L, Resche-Rigon M, Porcher R, Ravaud P, et al. Effect of Tocilizumab vs Usual Care in Adults Hospitalized With COVID-19 and Moderate or Severe Pneumonia: A Randomized Clinical Trial. *JAMA Internal Medicine*. 2021;181(1):32-40.
84. Batah F, Gore S, Verma MR. Effect of jackknifing on various ridge type estimators. *Model Assisted Statistics and Applications*. 2008;3.
85. Doroudi S. The Bias-Variance Tradeoff: How Data Science Can Inform Educational Debates. *AERA Open*. 2020;6(4):2332858420977208.
86. Arora S, Venkataraman V, Donohue S, Biglan KM, Dorsey ER, Little MA, editors. High accuracy discrimination of Parkinson's disease participants from healthy controls using smartphones. 2014 IEEE International Conference on Acoustics, Speech and Signal Processing (ICASSP); 2014 4-9 May 2014.
87. Del Din S, Godfrey A, Galna B, Lord S, Rochester L. Free-living gait characteristics in ageing and Parkinson's disease: impact of environment and ambulatory bout length. *Journal of NeuroEngineering and Rehabilitation*. 2016;13(1):46.
88. Iosa M, Capodaglio E, Pelà S, Persechino B, Morone G, Antonucci G, et al. Artificial Neural Network Analyzing Wearable Device Gait Data for Identifying Patients With Stroke Unable to Return to Work. *Frontiers in Neurology*. 2021;12.
89. Joshi D, Mishra A, Anand S. A naïve Gaussian Bayes classifier for detection of mental activity in gait signature. *Comput Methods Biomech Biomed Engin*. 2012;15(4):411-6.
90. De Laet T, Papageorgiou E, Nieuwenhuys A, Desloovere K. Does expert knowledge improve automatic probabilistic classification of gait joint motion patterns in children with cerebral palsy? *PLoS ONE*. 2017;12.
91. Jager KJ, Zoccali C, Macleod A, Dekker FW. Confounding: what it is and how to deal with it. *Kidney Int*. 2008;73(3):256-60.
92. Eertink JJ, Heymans MW, Zwezerijnen GJC, Zijlstra JM, de Vet HCW, Boellaard R. External validation: a simulation study to compare cross-validation versus holdout or external testing to assess the performance of clinical prediction models using PET data from DLBCL patients. *EJNMMI Research*. 2022;12(1):58.
93. Mannini A, Trojaniello D, Cereatti A, Sabatini AM. A Machine Learning Framework for Gait Classification Using Inertial Sensors: Application to Elderly, Post-Stroke and Huntington's Disease Patients. *Sensors (Basel)*. 2016;16(1).

94. Czech M, Patel S. GaitPy: An Open-Source Python Package for Gait Analysis Using an Accelerometer on the Lower Back. *Journal of Open Source Software*. 2019;4:1778.