A Gait Analysis Tool Based on Machine Learning to Support the Rehabilitation Strategy of Post-stroke Patients

Nicoletta Balletti[1,2][a], Gennaro Laudato[1][b] and Rocco Oliveto[1,3][c]

[1]*STAKE Lab, University of Molise, Pesche (IS), Italy*

[2]*Defense Veterans Center, Ministry of Defense, Rome, Italy*

[3]*Datasound srl, Pesche (IS), Italy*

Keywords: Gait Analysis, Motion Tracking, Post Stroke Rehabilitation, Machine Learning.

Abstract: Stroke is a serious medical condition that can result in permanent brain damage and other pathological issues. Conditions suffered by survivors ranged in severity from full recovery to significant movement disability. Even though some may recover quickly, many stroke survivors require long-term support to help them achieve as much independence as they can. Thanks to a proper rehabilitation, patients who have experienced a stroke can work to regain skills that are suddenly lost when a section of their brain is injured. Due to the breakdown of neuronal networks in the motor cortex, abnormal gait patterns are a typical disability after a stroke. Therefore, gait analysis can be a powerful tool to support stroke patients during rehabilitation. In this work we propose GIULYO, a Machine Learning based tool that offers support in the assessment of video gait trials in stroke patients by providing an automatic analysis on the muscle activity of the assisted subject. GIULYO is a device-agnostic tool because it accepts motion tracking data in terms of 3d trajectories regardless of the type of instrumentation. GIULYO has been validated on the ARRA Stroke dataset and the results showed an overall accuracy of 0.74 while on a subset a patients—with common clinical assessment of mobility impairments—the accuracy increased to 0.92, therefore demonstrating the feasibility of involving a ML-based approach for the rehabilitation support of post stroke patients.

1 INTRODUCTION

Stroke is a clinically defined condition of quickly evolving symptoms or indicators of localized loss of brain function (Warlow et al., 1997). Survivors experienced conditions that can range in severity from complete recovery to severe movement impairment (Warlow, 1998). As a result, stroke patients continue to have serious locomotor deficits. Additionally, patients and rehabilitation professionals continue to face daily challenges in achieving a good gait recovery following a stroke (Nadeau et al., 2013).

The idea behind this work is based on the consideration that biomechanical components of steady-state walking in healthy individuals have been demonstrated to be produced by separated, coexcited muscle groups (Allen and Neptune, 2012; Neptune et al., 2009). These specific muscle groups can be detected thanks to the Surface Electromyography (S-EMG) (Routson et al., 2014), by evaluating the number of channels that registered EMG activity. This value is referred as modules (Allen and Neptune, 2012; Neptune et al., 2009).

Nevertheless, those who have had a stroke have poor intermuscular coordination, which is defined by the merging of modules that are ordinarily independent in healthy people (Clark et al., 2010). Following a stroke, having more independent modules has been linked to better performance in a variety of clinical and biomechanical walking assessments, including faster walking, a better Dynamic Gait Index (DGI) (Jonsdottir and Cattaneo, 2007), and better step length and propulsion symmetry (Bowden et al., 2010; Clark et al., 2010).

Many efforts were provided by the scientific community to support the rehabilitation strategy of post-stroke patients by proposing video gait analysis tools (Liu et al., 2021; Mirelman et al., 2010; Swank et al., 2020). Liu et al. (2021) developed a method based on the Microsoft Kinect camera that could detect the ro-

[a] https://orcid.org/0000-0002-6617-7074
[b] https://orcid.org/0000-0002-5241-1608
[c] https://orcid.org/0000-0002-7995-8582

Balletti, N., Laudato, G. and Oliveto, R.
A Gait Analysis Tool Based on Machine Learning to Support the Rehabilitation Strategy of Post-stroke Patients.
DOI: 10.5220/0011697800003414
In *Proceedings of the 16th International Joint Conference on Biomedical Engineering Systems and Technologies (BIOSTEC 2023) - Volume 5: HEALTHINF*, pages 400-407
ISBN: 978-989-758-631-6; ISSN: 2184-4305

tation and movement of the Center of Mass in many planes. The CoM is regarded as a crucial indicator for evaluating the impact of therapy and recovery. Swank et al. (2020) used video gait analysis to demonstrate the validity of Robotic Exoskeletons (EKSO) session in rehabilitation therapies of post-stroke individuals.

However, we believe that there is still room to contribute this research field. Indeed, in this paper, we present GIULYO (video-based **G**a**I**t tool for the a**U**tomatic ana**L**ysis of rehabilitation qualit**Y** in post str**O**ke survivors), a tool designed to provide an automatic quantitative assessment to the gait trials of a rehabilitation session. Indeed, GIULYO is an approach capable of analyzing a video gait trial and automatically classifying it according to the number of modules detected. Furthermore, GIULYO is a device-agnostic tool, *i.e.,* it does not depend on any specific motion tracking instrumentation because its workflow begins with the processing of 3d trajectories. Also, GIULYO allows the clinical evaluation of a gait trial without the need of installing S-EMG electrodes on the interested limb. The proposed approach has been validated on the ARRA post-stroke database (Routson et al., 2014). GIULYO was submitted to an extensive ML experimentation and validated using the most fitting validation scheme when involving data from different subjects, *i.e.,* the Leave 1 Subject Out (L1SO) cross validation. Results showed that GIULYO is capable of predicting the number of moduls of a gait trial with an overall accuracy of 0.74. However, on a specific subset—more than a third—of subjects presenting the common characteristic of poor motor skills, GIULYO achieves an overall accuracy of 0.92.

The rest of the paper is structured as follows: Section 2 provides details on related works focused on gait analysis for post-stroke rehabilitation. Section 3 presents GIULYO, our novel approach for the assessment of a gait trial in terms of muscle activity. Section 4 reports the design and the results of the empirical study we conducted to evaluate GIULYO. Section 5 reports the results achieved by GIULYO and a qualitative analysis in terms of features importance and performances at subject level. Finally, Section 7 concludes the paper and provides suggestions for possible future research directions.

2 RELATED WORKS

In this section a review of the state-of-the-art related to the main contribution in the research field of gait analysis in post-stroke individuals is proposed.

The study conducted by Nadeau et al. (2013) provided an overview of the gait analysis procedure as well as the most significant gait parameters and deviations in stroke survivors, with a focus on the impacts of gait speed and the significance of ground response forces (GRFs). Reduced walking speed, an unbalanced gait pattern, and a drop in peak moments and powers on the hemiparetic side were all considered as characteristics of the hemiparetic gait following stroke. The findings showed that While some traits are shared by all patients, hemiparetic individuals may have markedly different gait patterns, even if their walking speeds are equivalent. GRFs are useful in evaluating the gait pattern abnormalities of stroke patients as well as other neurologic groups, just as the other gait characteristics.

The objective of the observational study conducted by Ferrarin et al. (2015) was to evaluate the influence of gait analysis on therapeutic decision-making (both surgical and non-surgical) for adult patients with chronic walking difficulties due to stroke. The idea was that clinical recommendations based only on clinical examination and visual gait observation are considerably different from those based additionally on gait analysis data. The present study's findings were consistent with the notion that gait analysis has a considerable impact on treatment planning for chronic post-stroke patients with locomotor dysfunction, both surgically and non-surgically, and supports decision-making.

Through comparisons with traditional gait measurements, the study proposed by Guzik and Drużbicki (2020) examined the concurrent validity of the Gait Deviation Index (GDI) as an outcome measure of gait deficits at a chronic stage of recovery following a stroke. A group of 65 people with stroke and 65 healthy people without gait abnormalities were enrolled. An analytic system for movement was used to measure the kinematic gait characteristics. The results supported the contemporaneous validity of the GDI in post-stroke patients, but only for the afflicted limb and mGDI. The authors concluded that GDI for the unaffected limb, however, may be helpful in locating any compensatory mechanisms appearing in post-stroke gait patterns.

In the work presented by Li et al. (2019), the symmetry, regularity, and stability of post-stroke hemiparetic gaits were obtained as features using the dynamic temporal warping (DTW) technique, sample entropy approach, and empirical mode decomposition-based stability index. A cluster of 15 stroke survivors and 15 healthy control persons participated in the studies. The findings achieved by the authors suggested that hypothesized characteristics were considerably able to distinguish post-stroke hemiparetic patients from healthy control participants.

Khera and Kumar (2020) proposed a review to provide researchers with critical recommendations for applying ML approaches for gait analysis and gait rehabilitation. This review article demonstrated the effectiveness of ML approaches in identifying illnesses, forecasting the period of rehabilitation, and controlling rehabilitation devices, making them appropriate for clinical diagnosis.

Much effort was dedicated from the scientific community to contribute the research field of gait analysis for the support in the rehabilitation of post-stroke patients. To the best of our knowledge, no effort was provided in the automatic classification of muscle activity from video gait analysis, in terms of S-EMG channels detected during the walking activities to support the decision-making strategy in the rehabilitation of post-stroke individuals.

3 USING ML TO PREDICT THE QUALITY OF WALKING

In this section we present GIULYO, a novel approach for the automatic assessment of gait quality designed to support post-stroke patients during specific rehabilitation therapy.

3.1 The Workflow of GIULYO

A motion capture system is needed to measure subject kinematics data and an electromyograph to measure the muscle activity during the walking tasks. These two instruments provide the two sources of information necessary for GIULYO's analysis. Once the 3D trajectories are acquired, a features calculator module is activated which is in charge to evaluate three sets of aggregate features: (i) the first contains stride measures and timing info for all subjects, (ii) the second is about the measures derived from the walking cycle, such as the single and double support and (iii) the third contains aggregated leg angle and foot height data for Paretic (P) and Non-Paretic (N) legs/feet, acquired within the laboratory reference frame. These three sets of features—together with demographic and clinical descriptors—compose the final feature vector for the classification module. This latter is the component aimed at providing the final assessment of the walking activity.

Two studies were conducted within the experimentation proposed in this paper. The first one aimed at automatically assessing the raw number of independent muscle co-excited in the set 2,3,4 while the second is focused on a binary classification in *Low* and *High* activation of the muscle. Details about this choice of clustering are offered in Section 4.

3.2 Gait Features

As already described, the features can be conceptually divided into 3 distinct sets:

- **Set A (Stride Measures and Timing Info):** treadmill speed, Paretic Step Ratio (PSR) averaged over all steps, Paretic Step Length/Stride Length, PSR Standard deviation, Paretic Stride length (distance between same foot), Paretic stride length standard deviation, non-paretic stride length, non-paretic stride length standard deviation, etc.
- **Set B (Measures Derived from the Walking Cycle):** paretic affected leg side, left single support percentage over all cycle, right single support percentage over all cycle, left single support time, right single support time, step length, step time, stride length and stride time standard deviation for left and right sides, single and double support time standard deviation for left and right sides.
- **Set C (Aggregated Leg Angle and Foot Height Data):** paretic leg angle from pelvis to foot, paretic frontal angle from pelvis to foot, non-paretic leg angle from pelvis to foot, non-paretic frotal angle from pelvis to foot, paretic Distance from pelvis to foot, non-Paretic Distance from pelvis to foot, paretic (leg length/pelvis height), non-paretic (leg length/pelvis height), paretic vertical leg distance, non-Paretic vertical leg distance, etc.

Details about how these features were registered and obtained are available in Section 4 and in the paper proposed by Routson et al. (2014).

3.3 Demographic and Clinical Features

A set of demographic and clinical features were integrated to the features vector in order to increase the knowledge of the model embedded in GIULYO. Demographic descriptors were gender, age and the affected side (left or right) while clinical features included individual and overall scores based on the DGI (Jonsdottir and Cattaneo, 2007), the 6 Minutes Walk Test (6MWT) (Enright, 2003), the Berg Balance Scale (BBS) (Berg, 1992), and the Fugl-Meyer (FM) Score (FUGL et al., 1975).

3.4 Putting All Together

GIULYO combines all the features we previously described. After the training phase, GIULYO is able—given a walking activity—to classify it in terms of muscle activation to describe the quality of the rehabilitation therapy and the progresses made. The final features vector is composed by the following features: anonymized subject ID, Trial Number, Gender, Age, Affected-Side, Speed, [DGI-1,...,DGI-8,DGI-TOT], 6MWT-Distance, [BERG-1,...,BERG-14,BERG-TOT], [FM-1,...,FM-17,FM-TOT,FM-Sinergy], Set-A, Set-B, Set-C, number of Modules.

4 EMPIRICAL EVALUATION

The *goal* of this study is to evaluate the accuracy of GIULYO in classifying walking trials in terms of muscle activity in rehabilitation patients. The *perspective* is of a researcher who wants to understand if combining several walking features is useful for the automatic assessment of the muscle activity. Thus, the study is steered by the following:

Research Question:
Can Machine Learning be used to predict the quality of walking in post-stroke patients?

The above RQ is then divided into two sub-RQ:

RQ_A: To what extent can GIULYO *predict the number of muscles activated during walking?*

RQ_B: To what extent can GIULYO *predict the level of muscle activation during walking?*

With RQ_A we aimed at verifying if ML can predict the single number of muscle activated during a walking activity as an indicator of the quality of the gait while with RQ_B our purpose was to evaluate the capability of the ML in discriminating a walking activity in low and high muscle activation.

4.1 Context Selection

The context of this study is represented by ARRA Post-Stroke Database (Routson et al., 2014). The data were obtained from 27 post-stroke participants and 17 healthy control participants. However, only 38 subjects presented the data with the complete set of descriptors. Five conditions were tested while walking on a treadmill at intervals of 30 seconds. Examples of such conditions include (i) self-Selected (SS) walking pace, which the participant determined to be their typical walking speed, High Step (HS) conditions, where subjects were asked to take as high a step as they could while walking at their SS pace. Kinematics, kinetics (from split belt treadmill force plates), and electromyography data were gathered for each condition. The following tools were employed to gather the data:

- to evaluate subject kinematics, a 12-camera motion capture system (PhaseSpace, Inc., San Leandro, CA) with two linear detectors in each camera was used. In order to establish the parameters of segment size, the system additionally makes use of active markers that produce infrared light and are positioned on anatomical landmarks of a patient.
- split-belt treadmill (FIT, Bertec, Inc.) with an inclination for measuring ground reaction forces and moments in three dimensions
- electromyograph MA400, 16 channel EMG system (Motion Lab Systems, Baton Rouge, LA).

For each subject, more than one gait trial is available. This is because in the experimental protocol proposed by Routson *et al.* (Routson et al., 2014), multiple trials were provided for each condition in order to select the best recording for data analysis. So, the dataset is composed of different information and measures related to walking, captured through the marker-based optoelectronic instrument. In addition, each subject—who underwent the experimental protocol of this study—had to install a set of electrodes for EMG signal acquisition. This was done bilaterally from the tibialis anterior, soleus, medial gastrocnemius, vastus medialis, rectus femoris, medial hamstrings, lateral hamstrings, and gluteus medius (Routson et al., 2014).

So, each walking trial can be classified in terms of co-excited muscles or modules, thanks to the information from EMG channels. To recap, each walking trial was assigned a class in the set [2,3,4] to indicate the number of independent co-excited muscles. Class 2, 3, and 4 indicate that two, three, and four modules were respectively detected by the instrumentation. These numerical classes can be considered as an index of quality of gait, as supported by several studies which found that a higher number of independent poststroke modules is associated with better gait performance (Bowden et al., 2010; Clark et al., 2010).

4.2 Features Selection and Classification

In the context of our study, we experimented two techniques of features engineering: (i) first a correlation analysis was applied to the features vector in order to discard the features with a correlation index greater than 0.95 (ii) then, an automatic features selection algorithm was triggered to evaluate the best descriptor to be used as input to the classification model.

A large set of algorithms for the features selection and for the training of the classification model of GIULYO was involved. Especially, we experimented: Random Forest (RF) (Barandiaran, 1998), MultiLayer Perceptron (MLP) (Pal and Mitra, 1992), Logistic Regression (LR) (Cramer, 2002), K Nearest Neighbour (KNN) (Dasarathy, 1991), Gaussian Naive Bayes (GNB) (Zhang, 2004), Stochastic Gradient Descent (SGD) (Ruder, 2016), Decision Tree (DT) (Wu et al., 2008), Bagging Classifier (BC) (Breiman, 1996), Gradient Boosting Classifier (GBC) (Friedman, 2001), AdaBoost (AB) (Freund and Schapire, 1997), Passive Aggressive Classifier (PAC) (Crammer et al., 2006), Extra Trees Classifier (ETC) (Geurts et al., 2006), Support Vector Machine (SVM) (Cortes and Vapnik, 1995).

For this experiment, we used the python library Scikit-learn (Pedregosa et al., 2011).

4.3 Experimental Procedure

A typical Leave-1-Person Out (L1PO) cross-validation was involved to assess the accuracy of GIULYO. To do this, we divided the data into *n* folds, one for each patient, and used—one at a time—these folds as test set. This indicates that a patient's data were embedded *n-1* times in the training dataset and *1* time in the test dataset. This method enables the creation of a classifier that is not tested and trained on the same patient's data. We did this in order to test the approach under the most difficult conditions possible because individual patients' gait data might vary consistently.

To answer this RQ_A, a specific study with the raw number of independent co-excited muscles was conducted. This meant that the classification component of GIULYO was in charge of discriminating a gait features vector in the classes 2, 3 or 4.

The dataset is composed by 50, 219, 208 instances for the class 2,3, and 4 respectively.

To face RQ_B, a binary ML experiment was conducted. Two classes were created. The first one *Low* aimed at representing the gait records in terms of low or poor muscle co-excitement; indeed, this class groups the records with 2 or 3 number of modules. The second class, namely *High* aimed at assess the gait records with 4 modules.

The dataset is composed by 269, and 208 instances for the class *Low* and *High* respectively.

For both studies, due to the class imbalance (especially in RQ_A), the SMOTE technique was experimented (Chawla et al., 2002).

Table 1: Classification performances achieved by the top 3 ML configurations in Study 1.

SMOTE	Corr.	Feat. Sel.	Class.	Overall Acc.
No	No	PAC	RF	0.67
No	Yes	RF	RF	0.63
No	Yes	Extra	RF	0.63

Table 2: Detailed classification performances achieved by the best configuration of GIULYO in Study 1.

Class	Classification Metrics		
	Precision	Recall	F1-score
2	0.53	0.18	0.27
3	0.64	0.73	0.68
4	0.71	0.72	0.71
Overall Accuracy			**0.67**

The results from the two studies conducted to answer the above RQs are presented according to the following class-level metrics: Precision, Recall, F1-score.

5 ANALYSIS OF THE RESULTS

The analysis of the results is described according to the specific RQ in the next subsections.

5.1 RQ_A: To What Extent Can GIULYO Predict the Number of Muscles Activated During Walking?

The top 3 configurations of ML settings are depicted in Table 1. The one with the best results is composed by a PAC model for features selection and a Random Forest for classification. Detailed results—achieved by this latter—during the experiment to answer RQ_A are shown in Table 2. It is evident that the best performances —at class level—are achieved by GIULYO for the subjects who had a S-EMG with 4 modules detected. Indeed, in this case, all the metrics are above 0.7. For what concerns class *3*, the performance show a decrease except for the recall value, which is 0.73. Class 2 is the one with the poorest classification performances. However, the overall accuracy is 0.67.

Table 3: Classification performances achieved by the top 3 ML configurations in Study 2.

SMOTE	Corr.	Feat. Sel.	Class.	Overall Acc.
Yes	Yes	DT	RF	0.74
Yes	No	PAC	RF	0.73
Yes	Yes	Extra	RF	0.73

Table 4: Classification performances achieved by GIULYO in Study 1.

Class	Classification Metrics		
	Precision	Recall	F1-score
Low	0.79	0.73	0.76
High	0.68	0.75	0.71
Overall Accuracy			**0.74**

5.2 RQ_B: To What Extent Can GIULYO Predict the Level of Muscle Activation During Walking?

The top 3 configurations of ML settings are depicted in Table 3. The one with the best results is composed by an operation of SMOTE, a correlation analysis, a Decision Tree model for features selection and a Random Forest algorithm for classification. Detailed results—achieved by this latter configuration—during the experiment to answer RQ_B are shown in Table 4. In this binary classification experiment, the *high* class shows classification performances around 0.7 with a peak of 0.75 for the Recall. On the other hand, the *low* class shows definitely better performances, with a F1-score of 0.76 and a Precision of 0.79. The overall accuracy achieved by GIULYO is 0.74.

The results obtained show that a classification model based on the Random Forest algorithm achieves the best classification performance.

5.3 Qualitative Analysis

For sake of space limitation, the qualitative analysis of the results achieved by GIULYO in this study is focused only on the version of the approach with the highest accuracy, *i.e.,* GIULYO used to discriminate a gait trial in *Low* or *High* class of modules triggered (Study 2, RQ_B).

5.3.1 Features Importance

From the L1SO validation scheme, a set of most informative instances were selected for each test set, *i.e.,* for each set composed of trial data related to a single subject. The top three features that were highly taken into consideration by the model embedded in GIULYO are: (i) the standard deviation of the leg length measures on the paretic side evaluated during the gait trial, (ii) the cadence, the number of steps over the minute, (iii) the maximum measure of the leg length and the minimum measure of the sagittal angle obtained during the gait trial.

It is not surprising that the leg length measurement, on the paretic side, was highly taken into account by the feature selection models as this is a determining factor in the analysis of the balance of post-stroke patients (Gardas and Shah, 2020).

Other examples of selected features include additional measures on the paretic side together with some descriptor on the normal side.

5.3.2 GIULYO's Performances on a Subset of Subjects

Qualitative analysis included the observation of the results at subject level. Within this study, we found that the performances of GIULYO are better for a subset of 14 subjects (out of the 38 that compose the whole dataset), *i.e.,* with an accuracy consistently greater than the 0.74 overall accuracy.

On this specific cluster of subject, 162 out of 176 gait trials were correctly identified as *low* (2 or 3 modules detected) or *high* (4 modules detected). The overall accuracy on this group of subjects is therefore approximately **0.92**. To foster this analysis, we tried to find the peculiarities that describe this group of subjects. During this study, we found that they have a low value of DGI-2 and an overall value of DGI smaller than 18. DGI is a gait index proposed by Shumway-Cook and Woollacott (Shumway-Cook and Woollacott, 1995) that showed high reliability in persons With chronic stroke and also evidence of concurrent validity with other balance and mobility scales (Jonsdottir and Cattaneo, 2007). The DGI includes many tasks that are evaluated in terms of numeric indexes, and their sum compose the overall DGI. The DGI-2 is known as ”Change in gait speed” and it is a test where the assisted patient has to start walking normally (for 5 minutes), and then walking quickly, for 5 minutes with five slow steps at the end (Jonsdottir and Cattaneo, 2007).

The DGI overall score of 19 or below indicates a risk of falling in older adults and those with vestibular impairment (Whitney et al., 2000, 1998).

6 LIMITATIONS OF THE STUDY

The followings could represent threats to validity of this study:

- The reduction to a binary state classification could seem a choice far from practical use in rehabilitation. However, we believe that this choice may provide a rapid and more accurate screening of muscle activity, with respect to the raw number of modules. This should represent an initial indication for the medical staff.
- The dataset used in this study may be too small for the purposes of a medical study. However, it was one of the public datasets—found in the relevant literature—with EMG, motion tracking and clinical assessment data for a substantial number of patients. It is our opinion that the results of this study should be considered as preliminary.
- There is no comparison between GIULYO and a scientific baseline. To the best of our knowledge, there is not a study to compare with.

7 CONCLUSIONS

This paper proposed GIULYO, a tool designed for the support of the rehabilitation strategy of post stroke patients. The tool is device-agnostic, meaning that no specific motion tracking instrumentation is needed to make it work and it is designed to provide an assessment of the gait in terms of a novel indication, the number of muscle group co-excited. This information could be raw (*i.e.,* the exact number) or clustered (*i.e.,* low or high muscle activation). GIULYO was validated on the ARRA post stroke dataset. Results showed an overall accuracy of 0.74 in the binary case, and of 0.92 on a specific subset of patients, therefore demonstrating higher performances when subjects have poor locomotor skills, according to the DGI. GIULYO aims at be embedded in modern Decision Support Systems (DSS) for the support to the medical equipe thanks to a rapid screening of the assisted patients.

Future works will be devoted to validate GIULYO (i) on a larger set of patients and (ii) on data acquired from another motion tracking instrument to verify how the accuracy of the 3d trajectories may affect the precision of the automatic classification.

ACKNOWLEDGMENT

The authors have been supported by the project "*EDAM*": A Diagnosis Recommender System based on Explainable Artificial Intelligence and the Combination of Motion Analysis and Others Clinical Biomarkers funded by the Italian Ministry of Defense.

Special thank to A. Ciccotelli for the support in the generation of the features vector.

REFERENCES

Allen, J. L. and Neptune, R. R. (2012). Three-dimensional modular control of human walking. *Journal of biomechanics*, 45(12):2157–2163.

Barandiaran, I. (1998). The random subspace method for constructing decision forests. *IEEE Trans. Pattern Anal. Mach. Intell*, 20(8):1–22.

Berg, K. (1992). Measuring balance in the elderly: Development and validation of an instrument.

Bowden, M. G., Clark, D. J., and Kautz, S. A. (2010). Evaluation of abnormal synergy patterns poststroke: relationship of the fugl-meyer assessment to hemiparetic locomotion. *Neurorehabilitation and neural repair*, 24(4):328–337.

Breiman, L. (1996). Bagging predictors. *Machine learning*, 24(2):123–140.

Chawla, N. V., Bowyer, K. W., Hall, L. O., and Kegelmeyer, W. P. (2002). Smote: synthetic minority oversampling technique. *Journal of artificial intelligence research*, 16:321–357.

Clark, D. J., Ting, L. H., Zajac, F. E., Neptune, R. R., and Kautz, S. A. (2010). Merging of healthy motor modules predicts reduced locomotor performance and muscle coordination complexity post-stroke. *Journal of neurophysiology*, 103(2):844–857.

Cortes, C. and Vapnik, V. (1995). Support-vector networks. *Machine learning*, 20(3):273–297.

Cramer, J. S. (2002). The origins of logistic regression (technical report). In *Tinbergen Institute*.

Crammer, K., Dekel, O., Keshet, J., Shalev-Shwartz, S., and Singer, Y. (2006). Online passive aggressive algorithms.

Dasarathy, B. V. (1991). Nearest neighbor (nn) norms: Nn pattern classification techniques. *IEEE Computer Society Tutorial*.

Enright, P. L. (2003). The six-minute walk test. *Respiratory care*, 48(8):783–785.

Ferrarin, M., Rabuffetti, M., Bacchini, M., Casiraghi, A., Castagna, A., Pizzi, A., and Montesano, A. (2015). Does gait analysis change clinical decision-making in poststroke patients? results from a pragmatic prospective observational study. *Eur J Phys Rehabil Med*, 51(2):171–84.

Freund, Y. and Schapire, R. E. (1997). A decision-theoretic generalization of on-line learning and an application to boosting. *Journal of computer and system sciences*, 55(1):119–139.

Friedman, J. H. (2001). Greedy function approximation: a gradient boosting machine. *Annals of statistics*, pages 1189–1232.

FUGL, M. et al. (1975). The post-stroke hemiplegic patient. i. a method for evaluation of physical performance.

Gardas, S. and Shah, H. (2020). Influence of leg length discrepancy on balance and gait in post-stroke patients: a correlational study. *Bulletin of Faculty of Physical Therapy*, 25(1):1–9.

Geurts, P., Ernst, D., and Wehenkel, L. (2006). Extremely randomized trees. *Machine learning*, 63(1):3–42.

Guzik, A. and Drużbicki, M. (2020). Application of the gait deviation index in the analysis of post-stroke hemiparetic gait. *Journal of Biomechanics*, 99:109575.

Jonsdottir, J. and Cattaneo, D. (2007). Reliability and validity of the dynamic gait index in persons with chronic stroke. *Archives of physical medicine and rehabilitation*, 88(11):1410–1415.

Khera, P. and Kumar, N. (2020). Role of machine learning in gait analysis: a review. *Journal of Medical Engineering & Technology*, 44(8):441–467.

Li, M., Tian, S., Sun, L., and Chen, X. (2019). Gait analysis for post-stroke hemiparetic patient by multi-features fusion method. *Sensors*, 19(7):1737.

Liu, Y., Liu, B., Zhou, Z., Cai, S., and Xie, L. (2021). A novel center of mass (com) perception approach for lower-limbs stroke rehabilitation. In *International Conference on Social Robotics*, pages 606–615. Springer.

Mirelman, A., Patritti, B. L., Bonato, P., and Deutsch, J. E. (2010). Effects of virtual reality training on gait biomechanics of individuals post-stroke. *Gait & posture*, 31(4):433–437.

Nadeau, S., Betschart, M., and Bethoux, F. (2013). Gait analysis for poststroke rehabilitation: the relevance of biomechanical analysis and the impact of gait speed. *Physical Medicine and Rehabilitation Clinics*, 24(2):265–276.

Neptune, R. R., Clark, D. J., and Kautz, S. A. (2009). Modular control of human walking: a simulation study. *Journal of biomechanics*, 42(9):1282–1287.

Pal, S. and Mitra, S. (1992). Multilayer perceptron, fuzzy sets, and classification. *IEEE Transactions on Neural Networks*, 3(5):683–697.

Pedregosa, F., Varoquaux, G., Gramfort, A., Michel, V., Thirion, B., Grisel, O., Blondel, M., Prettenhofer, P., Weiss, R., Dubourg, V., Vanderplas, J., Passos, A., Cournapeau, D., Brucher, M., Perrot, M., and Duchesnay, E. (2011). Scikit-learn: Machine learning in Python. *Journal of Machine Learning Research*, 12:2825–2830.

Routson, R. L., Kautz, S. A., and Neptune, R. R. (2014). Modular organization across changing task demands in healthy and poststroke gait. *Physiological reports*, 2(6):e12055.

Ruder, S. (2016). An overview of gradient descent optimization algorithms. *arXiv preprint arXiv:1609.04747*.

Shumway-Cook, A. and Woollacott, M. H. (1995). Theory and practical applications. *Motor Control*.

Swank, C., Almutairi, S., Wang-Price, S., and Gao, F. (2020). Immediate kinematic and muscle activity changes after a single robotic exoskeleton walking session post-stroke. *Topics in Stroke Rehabilitation*, 27(7):503–515.

Warlow, C. (1998). Epidemiology of stroke. *The Lancet*, 352:S1–S4.

Warlow, C. P., Dennis, M., Gijn, J. v., Hankey, G., Sandercock, P., Bamford, J., Wardlaw, J., and Brown, M. M. (1997). Stroke: a practical guide to management. *BMJ-British Medical Journal-International Edition*, 314(7097):1840.

Whitney, S., Hudak, M., and Marchetti, G. (2000). The dynamic gait index relates to self-reported fall history in individuals with vestibular dysfunction. *Journal of Vestibular Research*, 10(2):99–105.

Whitney, S. L., Poole, J. L., and Cass, S. P. (1998). A review of balance instruments for older adults. *The American Journal of Occupational Therapy*, 52(8):666–671.

Wu, X., Kumar, V., Ross Quinlan, J., Ghosh, J., Yang, Q., Motoda, H., McLachlan, G. J., Ng, A., Liu, B., Yu, P. S., et al. (2008). Top 10 algorithms in data mining. *Knowledge and information systems*, 14(1):1–37.

Zhang, H. (2004). The optimality of naive bayes. *Aa*, 1(2):3.