Machine Learning-Based Qualitative Analysis of Human Gait Through Video Features

Nicoletta Balletti[1,2][a], Roberto Zinni[3], Marco Russodivito[1][b], Gennaro Laudato[1][c],
Simone Scalabrino[1,4][d] and Rocco Oliveto[1,4][e]

[1]*STAKE Lab, University of Molise, Pesche (IS), Italy*
[2]*Defense Veterans Center, Ministry of Defense, Rome, Italy*
[3]*Word Power SRL, Italy*
[4]*Datasound srl, Pesche (IS), Italy*

Keywords: Gait Analysis, Motion Tracking, DGI, Machine Learning.

Abstract: Strokes constitute a major cause of both mortality and disability, carrying significant economic implications for healthcare systems. Evaluating the quality of gait in post-stroke patients during rehabilitation is essential for providing effective care. The Dynamic Gait Index (DGI) is a valuable metric for evaluating gait quality. However, the assessment of such an index typically requires invasive tests or specialized sensors. In this paper, we introduce a machine learning-based approach for estimating DGI exclusively from video recordings. Our research encompasses a comprehensive set of experiments, including data preprocessing, feature selection, and the application of various machine learning algorithms. To ensure the robustness of our findings, we employ the Leave 1 Subject Out (L1SO) cross-validation method. Our results underscore the challenge of accurately estimating DGI using solely video data. We achieved an R-squared (R^2) value of only 0.19 and a mean absolute error (MAE) of 2.2. Notably, we observed that our approach yielded notably poorer results for a specific subset of three patients. Upon excluding this subset, the R^2 increased to 0.30, and the MAE improved to 1.9. This observation suggests that incorporating patient-specific features into the model may hold the key to enhancing its overall accuracy.

1 INTRODUCTION

Strokes provide a significant contribution to both mortality and disability. From a clinical perspective, a stroke occurs when there is a temporary interruption in the blood supply to a part of the brain.[1] In the US, the economic burden associated with strokes amounted to approximately $57B between 2018 and 2019.[1] Most of such costs are directed toward addressing the long-term disabilities resulting from strokes, which affect over half of stroke survivors aged 65 and older.[1] Gait abnormalities are frequently encountered in patients who have suffered a stroke.[1]

These individuals must undergo extended rehabilitation before they can regain the ability to walk. Prior research has introduced various techniques to support the rehabilitation of post-stroke patients by assessing the quality of their gait. For instance, Swank et al. (2020) employed video gait analysis to establish the effectiveness of Robotic Exoskeletons (EKSO) in post-stroke rehabilitation therapy. Liu et al. (2021) used the Microsoft Kinect camera to monitor the rotation and movement of the Center of Mass in multiple planes for gait analysis.

Recently, Balletti et al. (2023) introduced GIU-LYO, an approach designed to automatically predict the number of independent co-excited muscle groups by analyzing patients' video recordings of their gait. Indeed, specific muscle groups can be identified through Surface Electromyography (S-EMG) by analyzing the number of channels displaying EMG activity (Routson et al., 2014). GIULYO provides this information without the need for installing sensors on

---

[a] https://orcid.org/0000-0002-6617-7074
[b] https://orcid.org/0009-0004-8860-1739
[c] https://orcid.org/0000-0002-5241-1608
[d] https://orcid.org/0000-0003-1764-9685
[e] https://orcid.org/0000-0002-7995-8582
[1]https://www.cdc.gov/stroke/facts.htm

the human body, and it was validated using the ARRA post-stroke database (Routson et al., 2014). Although GIULYO's predictions support the evaluation of rehabilitation sessions, they are limited to distinguishing only three classes (2, 3, or 4 co-excited muscles).

Ideally, healthcare practitioners would benefit from a more granular measurement of gait quality. The literature presents a metric known as the Dynamic Gait Index (DGI) (Shumway-Cook and Woollacott, 1995), which aims to achieve this goal. DGI assesses gait quality by providing an index ranging from 0 to 24, with a lower index indicating lower gait quality and a higher risk of falls (Herdman, 2000). Currently, DGI can only be measured through invasive tests, requiring patients to perform specific exercises or wear specialized sensors (Herdman, 2000; Shumway-Cook and Woollacott, 1995).

This work aims to automatically estimate a rehabilitation session's DGI using a video recording, providing a fine-grained assessment of gait quality. To accomplish this, we introduce a machine learning (ML)-based approach that, based on features extracted from the gait video, predicts the DGI value. Similar to GIULYO, our approach is device-agnostic and does not rely on specific motion tracking instruments.

To validate our approach, we conducted extensive experiments involving various data preprocessing techniques, feature selection strategies, and ML algorithms to develop the best DGI prediction model. We employed a Leave 1 Subject Out (L1SO) cross-validation approach, ensuring that no subject's data used for testing was included in the training set. Our results reveal the considerable challenge of accurately measuring DGI exclusively from video data. The most successful combination of preprocessing techniques, feature selection, and ML algorithms achieved an R^2 value of only 0.19 for the regression problem, with a mean absolute error of 3.1. However, we observed that the model's errors are most pronounced in patients with slower walking speeds. By excluding such patients from the dataset upfront, we achieved more favorable results ($R^2 = 0.24$) with a mean absolute error of 2.7. This figure falls below the minimum detectable change for DGI (Romero et al., 2011), which is 2.9.

2 BACKGROUND AND RELATED WORK

We first introduce the Dynamic Gait Index (DGI). Then, we review related studies concerning DGI estimation and the assessment of gait quality.

2.1 Dynamic Gait Index

The Dynamic Gait Index (DGI) is a clinical assessment tool originally proposed by Shumway-Cook and Woollacott in 1995 (Shumway-Cook and Woollacott, 1995). It is widely employed by healthcare professionals, particularly physical therapists and rehabilitation specialists, to evaluate an individual's dynamic balance and walking ability. DGI is often utilized to assess those at risk of falling, such as the elderly population, post-stroke patients, individuals with vestibular disorders or brain injuries, and those experiencing balance issues resulting from non-vestibular causes (Jonsdottir and Cattaneo, 2007; Wrisley et al., 2003).

DGI comprises a sequence of eight distinct walking tasks or conditions designed to evaluate various facets of gait and dynamic balance. These tasks are typically administered in a *controlled* setting, with the individual being assessed receiving a performance score for each task.

The equipment necessary for the assessment includes *a box, two cones, stairs, and a 6-meter-long, 40-centimeter-wide walkway*. The specific activities involved in the DGI assessment are as follows (Herdman, 2000; Shumway-Cook and Woollacott, 1995):

1. **Gait Level Surface.** Walk a distance of approximately twenty feet (or six meters) at your normal pace.

2. **Change in Gait Speed.** Begin walking at your usual speed, then, on the assessor's cue, walk as fast as possible. On the cue to slow down, walk as slowly as you can.

3. **Gait with Horizontal Head Turns.** Walk at your usual pace. When instructed to "look right", continue walking straight but turn your head to the right. Keep your gaze to the right until instructed to "look left", at which point you should turn your head left. When told to "look straight", return your gaze to the center.

4. **Gait with Vertical Head Turns.** Tip your head up and maintain your upward gaze until instructed to "look down". When prompted to look down, continue walking straight while lowering your head. Upon receiving the cue to "look straight", return your head to a neutral position while continuing to walk.

5. **Gait and Pivot Turn.** Begin walking at your usual pace. Upon the cue to "turn and stop", swiftly pivot to face the opposite direction and come to a stop.

6. **Step Over Obstacle.** Walk at your normal speed. When you encounter a shoebox, step over it di-

rectly, rather than circumventing it, and continue walking.

7. **Step Around Obstacles.** Begin walking at your normal pace. When you reach the first cone, approximately two meters away, circumnavigate it on the right side. As you approach the second cone, located approximately six meters beyond the first, navigate around it on the left side.

8. **Steps.** Ascend the stairs as you would at home, using the railing if necessary. Upon reaching the top, pivot and descend the stairs.

Each task is assessed by a specialist on a scale ranging from a minimum score of *0*, indicating severe impairment, to a maximum of *3*, signifying normal function. Consequently, the overall DGI score can range from *0* to *24* (Shumway-Cook and Woollacott, 1995).

The DGI score serves as an indicator of a patient's propensity for falls. An index exceeding 22 characterizes individuals as safe ambulators, whereas an index below 20 denotes an increased risk of falls (Herdman, 2000; Whitney et al., 2000).

2.2 Related Work

We direct our focus towards gait analysis in individuals who have undergone a stroke, with a brief reference to works concerning DGI reliability.

Nadeau et al. (2013) conducted a comprehensive study that delved into the gait analysis process, shedding light on key gait parameters and deviations observed in stroke survivors. Their investigation revolved around the effects of gait speed and ground response forces (GRFs). Notable findings included reduced walking speed, an unsteady gait pattern, and diminished peak moments and powers on the affected side as common traits in post-stroke hemiparetic gait. Remarkably, even when two hemiparetic individuals exhibited similar walking speeds, their gait patterns could exhibit significant disparities. This underscores the critical role of GRFs in assessing gait abnormalities among stroke patients, among other gait characteristics.

Ferrarin et al. (2015) conducted an observational study aimed at evaluating how gait analysis influences therapeutic decision-making, be it surgical or non-surgical, for adult patients with chronic walking difficulties resulting from a stroke. Their research unveiled substantial differences in recommendations based solely on clinical examination and visual gait observation compared to those supplemented with gait analysis data. This study underscored the profound impact of gait analysis on treatment planning for chronic post-stroke patients with locomotor

dysfunction, endorsing both surgical and non-surgical decision-making processes.

Li et al. (2019) employed dynamic time warping (DTW), sample entropy, and empirical mode decomposition-based stability index to extract features associated with symmetry, regularity, and stability in post-stroke hemiparetic gaits. Their study encompassed 15 stroke survivors and 15 healthy control subjects, with findings strongly supporting the efficacy of the identified features in distinguishing post-stroke hemiparetic patients from their healthy counterparts.

Eichler et al. (2022) proposed an approach to automate the BBS fall risk assessment test. Subjects are required to execute the BBS tasks while a computer vision system captures poses and motion. Subsequently, a multi-level machine learning model predicts the overall score. Results confirm the feasibility of predicting gait assessment indexes through non-invasive video features.

Liuzzi et al. (2023) introduced machine learning techniques for predicting mDGI, suggesting an insightful comparison with its MDC (Minimally Detectable Change). Given that this research targets the unmodified DGI, Romero et al. (2011) is used as a reference, as it established an MDC$_{95\%}$ of 2.9 points for DGI.

Balletti et al. (2023) presented GIULYO, an approach designed to automatically predict the number of co-excited muscles by visually analyzing patients' gait. GIULYO offers an estimation of the number of co-excited muscles during walking, categorizing gait into three classes (2, 3, or 4 co-excited muscles). As the authors note, the predictions furnished by this approach have the potential to support the evaluation of rehabilitation sessions.

While the latter work lays a solid foundation for assisting practitioners in the automatic evaluation of patients' gait, it remains a coarse-grained approach. In this study, we strive to bridge this gap by the automatic prediction of the DGI, a more refined measure of gait quality.

3 ESTIMATING THE QUALITY OF WALKING

Our primary design objective is to develop a system capable of assessing the quality of a patient's gait without the need for specialized staff in the least invasive manner. To achieve this, we employ a computer vision markerless device with the capability to detect the positions of subject joints and bone rotations in real-time. Consequently, the input to our approach is

Table 1: The features used to estimate the quality of walking in our ML-based approach.

	Aspect	Aggregations
Stride	Walking speed	Mean, SD
	Number of strides	Mean, SD
	Paretic step ratio over all steps	Mean, SD
	Paretic propulsion	Mean, SD
	Paretic stride length	Mean, SD
	Non-paretic stride length	Mean, SD
	Paretic step length	Mean, SD
	Non-paretic step length	Mean, SD
	Norm. foot height for paretic side	Mean, SD
	Norm. foot height for non-paretic side	Mean, SD
Walking Cycle	Left single support percentage	//
	Right single support percentage	//
	Left step length	Mean
	Right step length	Mean
	Left stride length	Mean
	Right stride length	Mean
Legs+Feet	Paretic leg sagittal-plane angle	Sum, Min, Max, Mean, SD
	Non-paretic leg sagittal-plane angle	Sum, Min, Max, Mean, SD
	Paretic leg frontal-plane angle	Sum, Min, Max, Mean, SD
	Non-paretic leg frontal-plane angle	Sum, Min, Max, Mean, SD
	Paretic leg length	Sum, Min, Max, Mean, SD
	Non-paretic leg length	Sum, Min, Max, Mean, SD
	Normalized paretic leg length	Sum, Min, Max, Mean, SD
	Normalized non-paretic leg length	Sum, Min, Max, Mean, SD
	Paretic leg vertical-length	Sum, Min, Max, Mean, SD
	Non-paretic leg vertical-length	Sum, Min, Max, Mean, SD
	Normalized paretic leg vertical-length	Sum, Min, Max, Mean, SD
	Normalized non-paretic leg vertical-length	Sum, Min, Max, Mean, SD
	Normalized paretic foot height	Sum, Min, Max, Mean, SD
	Normalized non-paretic foot height	Sum, Min, Max, Mean, SD
	Non-paretic leg angle	Sum, Min, Max, Mean, SD
	Paretic foot height	Sum, Min, Max, Mean, SD
	Non-paretic foot height	Sum, Min, Max, Mean, SD
Demogr.	Age	//
	Gender	//
	Affected side	//
	Trial condition	//

a video recording of the patient's gait, while the output is a numerical estimation of the DGI.

Irrespective of the selected acquisition device, our system comprises two key modules: a *recorder* module responsible for translating gait data into a standardized format, and a *calculator* module designed to extract features related to the acquired video. Once acquired the gait data, we perfom a video analysis procedure aiming to obtain a feature vector that represents a single gait session for a given patient.

The features extracted capture four distinct aspects of the gait: *stride-related*, *walking cycle-related*, *leg and feet-related*, and *demographics*. In the following we provide a detailed description of each of these features, which are also succinctly summarized in Table 1.

Stride-Related Aspects. These aspects focuses on stride measurements and timing data. They offer insight into the impact of pathology on the subject's stride, which can be instrumental in assessing overall gait quality. For patients with a paretic side and a non-paretic side, we consider the following aspects: (i) walking speed, (ii) number of strides, where a stride represents a complete gait cycle consisting of two steps, starting with one foot making contact with

the ground and ending when the same foot repeats this contact, (iii) paretic step ratio calculated as the average of *(paretic step length) / (stride length)* over all strides in the trial, (iv) paretic propulsion, (v) paretic and non-paretic stride length, (vi) paretic and non-paretic step length, and (vii) normalized foot height for both paretic and non-paretic sides. Mean and standard deviation are computed for these aspects to extract relevant features.

Walking Cycle-Related Aspects. This aspects comprise metrics derived from the walking cycle, which can help in assessing disparities between the paretic and non-paretic sides, contributing to the estimation of DGI. These aspects include: (i) single support percentages, representing the *(total time in which the subject is supported by a single leg on the chosen side) / (total trial time)* for both sides, and (ii) step and stride length averages across the entire trial.

Legs and Feet-Related Aspects. The third set of aspects encompasses measurements related to the legs and feet, both paretic and non-paretic, providing insights into how this gait differs from a normal one in terms of movement. Specific aspects include: (i) sagittal plane leg angle, measured from pelvic Center Of Mass (COM) to foot COM, (ii) frontal plane leg angle, measured from pelvic COM to foot COM, (iii) leg length, defined as the distance between pelvis COM and foot COM, (iv) normalized leg length, expressed as *(leg length) / (height of pelvis COM from the floor)*, (v) vertical-only leg length, representing the difference between the vertical components of pelvis COM and foot COM, (vi) normalized vertical length, defined as *(vertical-only length) / (height of pelvis COM)*, and (vii) normalized foot height, calculated as *(foot COM height) / (pelvis COM height)*. For each of these seven aspects, measured for both sides, we compute the sum, maximum, minimum, mean, and standard deviation.

Demographic and Clinical Aspects. To enhance the effectiveness of the evaluation system, we include demographic information known to the patient. This data, collected alongside gait analysis, consists of the patient's age, gender, the affected side expressed as 1/2 for left/right, and the trial condition, represented by an integer from 0 to 4, addressing one of the five trial conditions described in Section 4.

The above features are then used to estimate the DGI. Given the numerical nature of DGI, estimating it naturally falls under the category of regression problems. Thus, we employ machine learning techniques to train a regression model capable of estimating DGI. Further details regarding the algorithms used are presented in our description of the experimental design (Section 4).

4 EMPIRICAL STUDY DESIGN

The *goal* of our empirical study is to understand to what extent our approach can be used to automatically assess DGI. Our study is guided by the following research question: *Can DGI be estimated through machine learning approaches trained only on video recordings of human gait?*

The answer to this research question is crucial in determining whether video data alone is sufficient for DGI estimation.

4.1 Context Selection

For our study, we utilized the ARRA Post-Stroke Database (Kautz, 2018; Routson et al., 2014). This database originates from a study focused on elucidating the cause-and-effect relationship between neural output and the level of walking functionality in post-stroke patients. It encompasses a wide range of data, including kinematics, kinetics (captured using split belt treadmill force plates), and electromyography (EMG) data collected from 27 post-stroke individuals (who were at least 6 months post-stroke) and 17 healthy control participants.

The dataset entails multiple gait trials performed by each subject under five distinct conditions:

- *Self-Selected (SS) walking pace*: The participant's usual walking speed.

- *Fastest Comfortable (FC)*: The participant's maximum comfortable walking speed.

- *High Step (HS)*: Participants were instructed to take the highest possible steps while maintaining their self-selected pace.

- *Quick step (QS)*: Subjects walked at their self-selected speed, taking the quickest possible steps.

- *Long step (LS)*: While retaining their self-selected speed, participants were instructed to take the longest possible steps.

Data collection involved several measurement techniques and equipment, including a 12-camera motion capture system from PhaseSpace, Inc., which employed active infrared markers placed on anatomical landmarks to measure body movements. Ground reaction forces and moments in three dimensions were measured using a split-belt treadmill from FIT, Bertec, Inc. Electromyography (EMG) data was acquired using the MA400, a 16-channel EMG system from Motion Lab Systems.

Each participant's gait data consisted of multiple trials, ensuring data diversity and representativeness. The dataset includes raw measurements required to derive the features utilized in our approach. Additionally, it contains EMG modules, a feature indicating the number of independent co-excited muscle groups measured during the trial, which we do not use due to its requirement for invasive sensors. Similarly, clinical assessments, such as the 6 Minutes Walk Test (6MWT), the Berg Balance Scale (BBS), and the Fugl-Meyer (FM) Score, were not used, as they necessitate specialized medical staff and would render the estimation of DGI redundant.

Furthermore, the dataset provides the previously assessed Dynamic Gait Index (DGI) for each participant. DGI scores in the patient population ranged from 8 to 22, though, as explained earlier, the theoretical DGI range is 0 to 24.

4.2 Experimental Procedure

The dataset was pre-processed to create a single row for each trial instance. This involved collecting demographic data (gender, age, affected side, trial condition, and treadmill speed, which serves as a proxy for *walking speed*) directly from the dataset. Subsequently, stride measurements, timing information, and *stride-related* metrics were included for both the paretic and non-paretic sides. For each bone type, the mean and standard deviation were calculated from the angles measured across the acquisition frames to quantify these metrics.

The resulting dataset, derived from the ARRA Post-Stroke Database (Kautz, 2018; Routson et al., 2014), comprises 322 instances and 130 features.

We employed several machine learning algorithms to train a regression model for DGI estimation, including RandomForestRegressor (Liu et al., 2012), MLPRegressorg, LogisticRegression, LinearSVR (Boser et al., 1992), SVR (Gunn et al., 1998), KNeighborsRegressor (Kramer, 2011), SGDRegressor (Ketkar and Ketkar, 2017), DecisionTreeRegressor (Loh, 2011), BaggingRegressor (Breiman, 1996), GradientBoostingRegressor (Friedman, 2001), AdaBoostRegressor (Freund and Schapire, 1997), PassiveAggressiveRegressor (Crammer et al., 2006), ExtraTreesRegressor (Geurts et al., 2006).

In terms of pre-processing, we utilized two techniques: *automatic feature selection* to exclude unnecessary features and make the problem more manageable for the machine learning algorithm, and *correlation-based feature selection* to eliminate highly correlated features that might not significantly contribute to the estimation. We tested various combinations of these pre-processing techniques to optimize performance.

For automatic feature selection, we employed

a wrapper approach based on different regression techniques, including LogisticRegression, Random-ForestRegressor (Liu et al., 2012), AdaBoostRegressor (Freund and Schapire, 1997), SGDRegressor (Ketkar and Ketkar, 2017), PassiveAggressiveRegressor (Crammer et al., 2006), ExtraTreesRegressor (Geurts et al., 2006), DecisionTreeRegressor (Loh, 2011), GradientBoostingRegressor (Friedman, 2001). Instead, in correlation-based feature selection, features with a correlation coefficient exceeding 0.90 were discarded. We considered and combined all possible pre-processing options for the two steps (automatic feature selection and correlation-based feature selection) with the machine learning techniques chosen for testing. This resulted in testing 208 configurations (8 for automatic feature selection $\times$ 2 for correlation-based feature selection $\times$ 13 machine learning techniques). To prevent overfitting and ensure that the model did not learn from data already observed for the specific patient on which it was tested, we employed a Leave-1-Person Out (L1PO) cross-validation. This approach partitioned the data into folds, with one fold assigned to each patient. Subsequently, we used these folds one by one as the test set, while the union of the remaining folds served as the training set. This method ensured that an individual patient's data contributed to the training dataset *n-1* times and to the test dataset once.

To answer our research question, we initiated our analysis by evaluating the goodness of fit for our models. To accomplish this, we employed the R^2 metric, which falls within the range of 0 to 1. A higher R^2 value indicates a better fit of the model to the data. Furthermore, we utilized two error metrics for assessing accuracy, namely MAE (Mean Absolute Error) and MSE (Mean Squared Error). MSE penalizes larger errors, offering a more precise estimate of the *cost* of errors, while MAE provides a realistic assessment of the error range in our approach for DGI assessment. Lower values in both metrics correspond to better performance. Lastly, we employed the explained variance metric, which indicates the extent to which the variance in the response variable can be accounted for by the features in our model. A higher explained variance value reflects a greater ability to explain data variation and implies a higher level of predictive quality.

5 EMPIRICAL STUDY RESULTS

We report the results achieved by the top four configurations in terms of the goodness of fit (R^2) in Table 2. Subsequently, we detail their respective performance
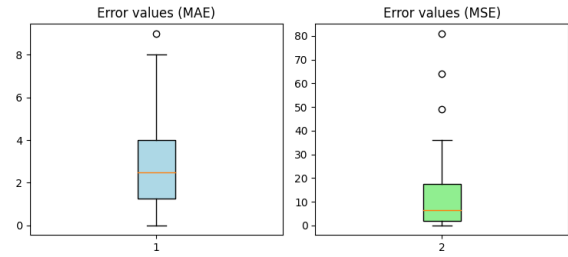

Figure 1: MAE and MSE values for the first configuration.

Table 2: The four best machine learning configurations.

Conf.	Algo.	Auto. Sel.	Corr. Sel.
#1	tree	gbc	False
#2	tree	tree	False
#3	log	gbc	False
#4	tree	adaboost	False

outcomes in Table 3.

The DecisionTreeRegressor model exhibited the highest R^2 and explained variance values. Notably, the top four configurations employed various automatic selection algorithms, while correlation-based selection was not utilized in any of these cases.

However, it is evident that the results are somewhat underwhelming in terms of R^2. The best model achieved only a R^2 of 0.19, signifying that even the most effective model cannot adequately fit the data. This model produced a *Mean Absolute Error (MAE)* of 3.11, indicating that the model's estimates are expected to deviate by an average of 3 DGI points in either direction.

Upon analyzing the selected features in the top four models, we observed that all walking cycle features and a subset of demographic features were consistently included. Additionally, all stride-related features were retained. In contrast, most features related to legs and feet were discarded during the feature selection process. To validate this observation and further explore feature importance, we tested the best configuration with each of the four feature categories isolated. A model based on walking cycle features achieved the highest R^2 (0.19), followed by demographic features (0.11), stride-related features (0.03), and legs-and-feet features (0.0). While raw measurements from legs and feet features failed to guide ma-

Table 3: Regression results achieved by the best four machine learning configurations accoding to R^2.

Conf.	R^2	MSE	MAE	Explained Variance
#1	0.19	16.04	3.11	0.96
#2	0.15	21.90	3.73	1.00
#3	0.15	17.84	3.33	0.92
#4	0.12	21.47	3.92	0.96

chine learning in establishing a correlation between gait and DGI, domain-specific and calculated aggregations, such as walking-related features, provided a better description of this relationship.

5.1 Discussion

We observed that the overall results of the best model we trained are slightly underwhelming: we achieve an R^2 of 0.19 and a mean absolute error of 3.11. In practice, this error might be low enough for some DGI values (*e.g.,* very high or very low), while it might be detrimental for in-between values. We compared the MAE we achieved with the variations observed when different human experts measure DGI through the standard state-of-the-art procedure. Such a variation is also called Minimum Detectable Change (MDC). Romero et al. (2011) computed the MDC$_{95\%}$ for DGI and reported that it is 2.9 (*i.e.,* in 95% of the cases, the error in the estimation of DGI is ±2.9). The MAE of our approach is above such a value. Thus, on average, the approach makes larger errors than the minimum detectable change.

We report in Figure 1 the MAE and MSE registered during the training of the best configuration for different patients. It can be noticed that the model is particularly inaccurate on three subjects. We tried to identify features common to those patients so that we can characterize them. To do this, we used agglomerative clustering (Murtagh and Legendre, 2014) on the dataset, considering all the features selected to train the first configuration. Specifically, we configured the algorithm to identify five clusters using the *Euclidean linkage metric* with *ward linkage criterion*, which minimizes the variance of the clusters being merged. We found that the smallest cluster containing outlier patients is composed of 9 subjects. To understand which features are most responsible for this cluster aggregation, we used the silhouette score(Shahapure and Nicholas, 2020), a metric that provides a measure of how similar a data entry is to its own cluster compared to other clusters. Specifically, we first calculated the average silouette score for the outliers cluster. Then, for each feature, we created a copy of the dataset replacing the selected feature with its mean. In this way, we nullify the feature variance and its clustering power. For each dataset copy we computed the cluster average silouette score again and compared each score with the original one. The worse the new score, the more important the feature is for clustering.

Through this procedure, we discovered that the most characterizing feature for this cluster is *Right single support percentage*. Besides, we found that this cluster has a single-support-percentage difference, defined as the absolute value of the difference between left and right single support percentages, higher than the other data points (*11.56* percentage points versus *5.3* percentage points on average for all the other instances). This analysis led us to conclude that this cluster contains trials with more impaired gaits which are assigned with better-ranged DGI values. In particular, we compared single-support-percentage differences with the original DGI value assigned to the trials and noticed higher ratios for cluster data (*0.22*) than for the remaining part of the dataset (*0.08*), which suggests how the outliers cluster is characterized by a less inverse-proportionality between gait impairment and DGI values. Then, we performed an experiment in which we excluded the cluster that contains the outlier patients to understand what theoretical improvements could be achieved. Overall, we obtained better results (R^2 = 0.24, explained variance = 1.0). The error also significantly decreased (MSE = 13.18, MAE = 2.70). It is also worth noting that this new experiment allowed the model to achieve a MAE score below the MDC$_{95\%}$ value.

6 CONCLUSION

We developed a machine learning-based approach for automatically estimating the Dynamic Gait Index (DGI) for post-stroke patients during rehabilitation sessions using video recordings. The results of our study showed that our approach performed poorly, particularly on patients who walked very slowly. When we removed these subjects from the analysis, the overall R^2 improved to 0.24, and the MAE reduced to 2.7, which falls below the minimum detectable change. In summary, our study demonstrates the feasibility of using video data for DGI estimation in post-stroke patients during rehabilitation. While challenges in achieving high precision persist, further research should focus on finding features that allow for more accurate DGI measurement, especially for patients who walk at slower speeds.

ACKNOWLEDGMENT

REFERENCES

Balletti, N., Laudato, G., and Oliveto, R. (2023). A gait analysis tool based on machine learning to support the rehabilitation strategy of post-stroke patients. In *HEALTHINF*, pages 400–407.

Boser, B. E., Guyon, I. M., and Vapnik, V. N. (1992). A training algorithm for optimal margin classifiers. In *WCLT*, pages 144–152.

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

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

Eichler, N., Raz, S., Toledano-Shubi, A., Livne, D., Shimshoni, I., and Hel-Or, H. (2022). Automatic and efficient fall risk assessment based on machine learning. *Sensors*, 22(4):1557.

Ferrarin, M., Rabuffetti, M., Bacchini, M., Casiraghi, A., Castagna, A., Pizzi, A., Montesano, A., and Palsy, C. (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):171–184.

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.

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

Gunn, S. R. et al. (1998). Support vector machines for classification and regression. *ISIS technical report*, 14(1):5–16.

Herdman, S. (2000). Physical therapy diagnosis for vestibular disorders. *Vestibular Rehabilitation. 2nd ed. Philadelphia, PA: FA Davis Company*.

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.

Kautz, S. A. (2018). *Medical University of South Carolina Stroke Data (ARRA)*.

Ketkar, N. and Ketkar, N. (2017). Stochastic gradient descent. *Deep learning with Python: A hands-on introduction*, pages 113–132.

Kramer, O. (2011). Unsupervised k-nearest neighbor regression. *arXiv preprint arXiv:1107.3600*.

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 *ICSR*, pages 606–615. Springer.

Liu, Y., Wang, Y., and Zhang, J. (2012). New machine learning algorithm: Random forest. In *ICICA*, pages 246–252. Springer.

Liuzzi, P., Carpinella, I., Anastasi, D., Gervasoni, E., Lencioni, T., Bertoni, R., Carrozza, M. C., Cattaneo, D., Ferrarin, M., and Mannini, A. (2023). Machine learning based estimation of dynamic balance and gait adaptability in persons with neurological diseases using inertial sensors. *Scientific Reports*, 13(1):8640.

Loh, W.-Y. (2011). Classification and regression trees. *Wiley interdisciplinary reviews: data mining and knowledge discovery*, 1(1):14–23.

Murtagh, F. and Legendre, P. (2014). Ward's hierarchical agglomerative clustering method: which algorithms implement ward's criterion? *Journal of classification*, 31:274–295.

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.

Romero, S., Bishop, M. D., Velozo, C. A., and Light, K. (2011). Minimum detectable change of the berg balance scale and dynamic gait index in older persons at risk for falling. *Journal of geriatric physical therapy*, 34(3):131–137.

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.

Shahapure, K. R. and Nicholas, C. (2020). Cluster quality analysis using silhouette score. In *DSAA*, pages 747–748. IEEE.

Shumway-Cook, A. and Woollacott, M. H. (1995). Theory and practical applications. *Motor control*, pages 89–90.

Swank, C., Sikka, S., Driver, S., Bennett, M., and Callender, L. (2020). Feasibility of integrating robotic exoskeleton gait training in inpatient rehabilitation. *Disability and rehabilitation. Assistive technology*, 15(4):409—-417.

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.

Wrisley, D. M., Walker, M. L., Echternach, J. L., and Strasnick, B. (2003). Reliability of the dynamic gait index in people with vestibular disorders. *Archives of physical medicine and rehabilitation*, 84(10):1528–1533.