

Article

Not peer-reviewed version

Sensor-Based Classification of Post-Stroke Motor Impairment Using Fugl-Meyer Lower Extremity Scores

[Cristiana Pinheiro](#)*, [Luís Abreu](#), [Joana Figueiredo](#), [Cristina Cruz](#), [João Cerqueira](#), [Cristina P. Santos](#)

Posted Date: 15 May 2026

doi: 10.20944/preprints202605.1064.v1

Keywords: biomarker; clinical scale; EMG; gait; machine learning; spatiotemporal



Preprints.org is a free multidisciplinary platform providing preprint service that is dedicated to making early versions of research outputs permanently available and citable. Preprints posted at Preprints.org appear in Web of Science, Crossref, Google Scholar, Scilit, Europe PMC, OpenAlex.

Copyright: This open access article is published under a [Creative Commons CC BY 4.0 license](#), which permit the free download, distribution, and reuse, provided that the author and preprint are cited in any reuse.

Disclaimer/Publisher's Note: The statements, opinions, and data contained in all publications are solely those of the individual author(s) and contributor(s) and not of MDPI and/or the editor(s). MDPI and/or the editor(s) disclaim responsibility for any injury to people or property resulting from any ideas, methods, instructions, or products referred to in the content.

Article

Sensor-Based Classification of Post-Stroke Motor Impairment Using Fugl-Meyer Lower Extremity Scores

Cristiana Pinheiro ^{1,*}, Luís Abreu ¹, Joana Figueiredo ^{1,2}, Cristina Cruz ³, João Cerqueira ^{4,5} and Cristina P. Santos ^{1,2}

¹ Center for MicroElectroMechanical Systems (CMEMS), University of Minho, 4800-058 Guimarães, Portugal

² LABBELS Associative Laboratory, University of Minho, Braga/Guimarães, Portugal

³ Department of Physical Medicine and Rehabilitation, Hospital of Braga, 4710-243 Braga, Portugal

⁴ Life and Health Sciences Research Institute (ICVS), University of Minho, 4710-057 Braga, Portugal

⁵ Clinical Academic Center (2CA-Braga), Hospital of Braga, 4710-243 Braga, Portugal

* Correspondence: cristianapinheiro6@hotmail.com

Abstract

This study aims to evaluate multiple feature sets composed of sensor-based biomarkers acquired during walking for the automated estimation of post-stroke motor impairment levels using Fugl-Meyer Lower Extremity Assessment (FMA-LE) derived classes. Sensor-based walking data from the open-source ARRA dataset were combined with data collected at the Hospital of Braga. Data from 32 post-stroke individuals (FMA-LE:24±3) were included. A decision tree classifier was evaluated using stratified 6-fold cross-validation across different feature configurations, including: correlated versus full feature sets; spatiotemporal versus electromyographic (EMG) features; inclusion of demographic variables; and the use of data augmentation. The best performance was achieved using correlated EMG features combined with age, paretic side, and body mass, along with noise-based data augmentation, yielding a validation MCC of 0.85±0.16 and a test MCC of 0.70. EMG features provided improved classification performance compared to spatiotemporal features, and comparable results were obtained using a reduced subset of muscles. These results demonstrate the feasibility of using EMG-based features acquired during walking to classify post-stroke motor impairment levels. Feature reduction and inclusion of demographic variables may support efficient model design, while data augmentation may enhance generalization. Further validation in larger and more diverse datasets is required to assess robustness and clinical applicability.

Keywords: biomarker; clinical scale; EMG; gait; machine learning; spatiotemporal

1. Introduction

Fugl-Meyer Assessment (FMA) clinical scale is the primary outcome to evaluate motor impairment in the post-stroke population [1]. FMA characterizes motor impairment in the domains of movement, coordination, speed, and reflex action of upper and lower extremities [2–4]. The patients are visually examined by healthcare professionals while performing multiple pre-defined structured physical tasks, resulting in an overall score of motor impairment [2–4]. Evaluating these scores is standard in clinical practice and meaningful for diagnostic and therapeutic purposes once they facilitate comparisons of patients and treatments for, consequently, guiding critical treatment choices for rehabilitation [5].

Although FMA presents excellent inter and intra-rater reliability, its interpretation may include some subjectivity and dependency on exposure of clinicians to a wide range of post-stroke impairments [6]. This drawback can be suppressed by exploring objective sensor-based biomarkers correlated with FMA scores once sensor data acquisition provides objective and real-time

measurements of motor impairment [7]. Sensor-based biomarkers can be combined with machine learning (ML) methods for automated estimation of clinical scores. After being trained by experienced clinicians, the model may be capable of interpreting sensor-based biomarkers to automatically deliver the clinical score, requiring just technical training to place sensors and configure equipment that may be achieved with a few supervised sessions as wearable sensor technologies develop.

Current literature has investigated the development of ML methods for automatically estimating FMA scores based on sensor-based biomarkers. Both biomechanical and physiological metrics were investigated [8–12]. Tozlu et al. [11] combined demographic, clinical, neurophysiological, and Magnetic Resonance Imaging metrics as features of a random forest achieving a median r-square of 0.88. Song et al. [10] introduced cellphone movement data while executing FMA-related tasks into a decision tree regression model, succeeding with an average r-square of 0.97. These works only focus on the upper extremity of the FMA scale (FMA-UE) [8–13] and, as far as the authors know, there is no study focusing on the FMA-LE segment of the clinical scale. However, lower extremity motor function impacts activities of daily living such as walking. In this manner, objective measurement of FMA-LE is also essential for monitoring post-stroke recovery progress, as appointed by Rech et. al [14].

This work aims to evaluate multiple feature sets composed of sensor-based biomarkers for automated estimation of post-stroke motor impairment levels during walking. The findings will contribute to maximizing the use of FMA-LE in clinical practice to support clinical decisions toward prompt rehabilitation for stroke patients. This work uses both collected (for testing) and open-source datasets (for training, validating, and testing). For this purpose, we studied the performance of a state-of-the-art Decision Tree classifier and multiple feature sets (EMG, spatiotemporal, and demographic) in estimating FMA-LE-derived motor impairment levels.

2. Materials and Methods

2.1. Participants

Data from open-source ARRA dataset [15] were merged with data acquired at the Hospital of Braga, producing a larger dataset with 32 post-stroke subjects (12 female, 53±8 years, 80±8 kg, 24±3 FMA-LE score, 16 paretic left side). All eligible subjects gave their informed consent for inclusion before they participated in the studies [15]. Data acquired at the Hospital of Braga followed the Declaration of Helsinki and the protocol approval by the Ethics Committee CEHB 157_2021 (“Comissão Ética Hospital Braga 157_2021”). We collected data from 5 post-stroke subjects (3 female, 45±18 years, 69±15 kg, 24±5 FMA-LE score, 2 paretic left side, 12±7 months after the stroke), using the following subject inclusion criteria: 1) history of single unilateral stroke; 2) lower limb muscle spasticity medically controlled; 3) able to complete the 10-meter walk test. The subjects were excluded according to the following exclusion criteria: 1) significant cognitive impairment limiting their active participation in the study; 2) neurological, orthopedic, cardiac, or respiratory disease affecting locomotion; 3) aphasia. These data were acquired to be used as a separate test set to evaluate model generalization.

2.2. Experimental Protocol and Data Collection

The participants were instructed to perform three walking trials at their self-selected speed. EMG data were acquired from a maximum of 8 muscles in the paretic limb while walking (tibialis anterior, soleus, gastrocnemius, vastus medialis, rectus femoris, medial hamstring, lateral hamstring, and gluteus medius). The following widely used frequency domain features were extracted from the EMG data overcoming possible noise in the signal by extracting its stable characteristics: mean (MNF), median (MDF), and peak (PKF) power frequencies [16]. Kinematic and ground reaction forces (GRF) data were also recorded during walking. ARRA dataset provided spatiotemporal parameters derived from these data, including non-paretic and paretic step length (m), stride length (m), stride

time (s), and cadence (steps/min). Demographic features including gender (male/female), age (years), body mass (kg), and paretic side (right/left) were also considered. More detailed information regarding data collection and processing can be found in Supplementary Material (Sections S1.2 and S1.3).

Spearman's correlation coefficients, r , were calculated between FMA-LE-derived classes (mid and low motor impairment levels) and EMG or spatiotemporal features. The correlation strength was interpreted as moderate if $r = 0.40 - 0.69$ and strong if $r \geq 0.7$ [17]. The correlated features were identified by presenting a moderate to strong correlation ($r \geq 0.4$).

2.3. FMA and Data Labelling

The Fugl-Meyer Assessment was accomplished by an expert (biomedical background) to determine the FMA-LE clinical score. Participants were classified with a low, mid, or high motor impairment level. High motor impairment was automatically considered for patients unable to walk (FMA-LE < 9 according to Smith et al. [18]). Smith et al. [18] reported the FMA-LE score for a sample of 93 post-stroke patients able to walk independently ranging from 9. Mid (label 0) and low (label 1) motor impairment levels were defined for $9 \leq \text{FMA-LE} < 21$ and $\text{FMA-LE} \geq 21$, respectively, according to Kwong et. al [19]. According to the inclusion criteria of both experimental protocols, patients with high motor impairment are not included. Ten participants were labelled with mid impairment (1 from the Hospital of Braga and 9 from the ARRA dataset) and 22 participants with low impairment (4 from the Hospital of Braga and 18 from ARRA dataset).

2.4. Data Splitting

Data from the ARRA dataset (27 subjects) was split at subject level into training and test sets using an 80–20% random stratified split according to motor impairment class (mid/low), ensuring class balance in both partitions.

The training set was used exclusively for model development and was further evaluated using StratifiedGroupKFold cross-validation with 6 folds. This approach ensured that all gait cycles from a given subject were assigned to the same fold, preventing subject-level data leakage between training and validation partitions. Stratification was applied at the subject level to preserve the distribution of motor impairment classes (mid/low) across folds, and each fold contained a disjoint set of subjects.

Within the training set, this configuration resulted in 19 samples (33%) classified as mid impairment and 38 samples (67%) classified as low impairment, with each sample corresponding to one average gait cycle.

We used two different datasets for testing the model, including a) only ARRA dataset subjects, and b) subjects from ARRA merged with the ones from the Hospital of Braga—the hybrid test set.

2.5. Training Dataset Preparation

Training dataset preparation methods were applied, namely adding noise or SMOTE methods. White or pink noise was added to all features from the training set as a data augmentation technique to foster model generalization ability [20] and to balance the classes. This resulted in three training sets: a) without noisy samples, b) with white noisy samples, and c) with pink noisy samples. Also, the SMOTE method was compared to balance the dataset. Thus, a fourth training set was created by applying SMOTE to the raw data. You can find detailed information on Supplementary Material (Section S1.5).

2.6. ML Model

We evaluated the performance of Decision Tree [8,10,11] state-of-the-art classifier. As a preliminary study, the decision tree classifier was considered a suitable choice due to its interpretability which allows healthcare professionals to make informed decisions based on the model's estimation, ability to handle both continuous (e.g., sensor-based features) and categorical

(e.g., demographic features) data, and capacity to capture non-linear relationships between features. [21] Moreover, it is robust to outliers, scalable for larger datasets, and adaptable to imbalanced data. [21] Regressor performance was not evaluated due to the limited open-source data from post-stroke patients covering the full-range FMA-LE. All algorithms were imported from the Scikit-Learn Python library.

Decision Tree classifier was trained with multiple feature sets to explore which are the most suitable for classifying the post-stroke motor impairment levels (Section S1.5 from Supplementary Material). Figure 1 resumes the methodological workflow of this work.

Model performance was evaluated with the MCC, F1-score, recall, and confusion matrix metrics, averaged over the 6 folds of cross-validation (S1.6 and S1.7 Section of Supplementary Material). The maximum score (value of 1) represents a perfect classification.

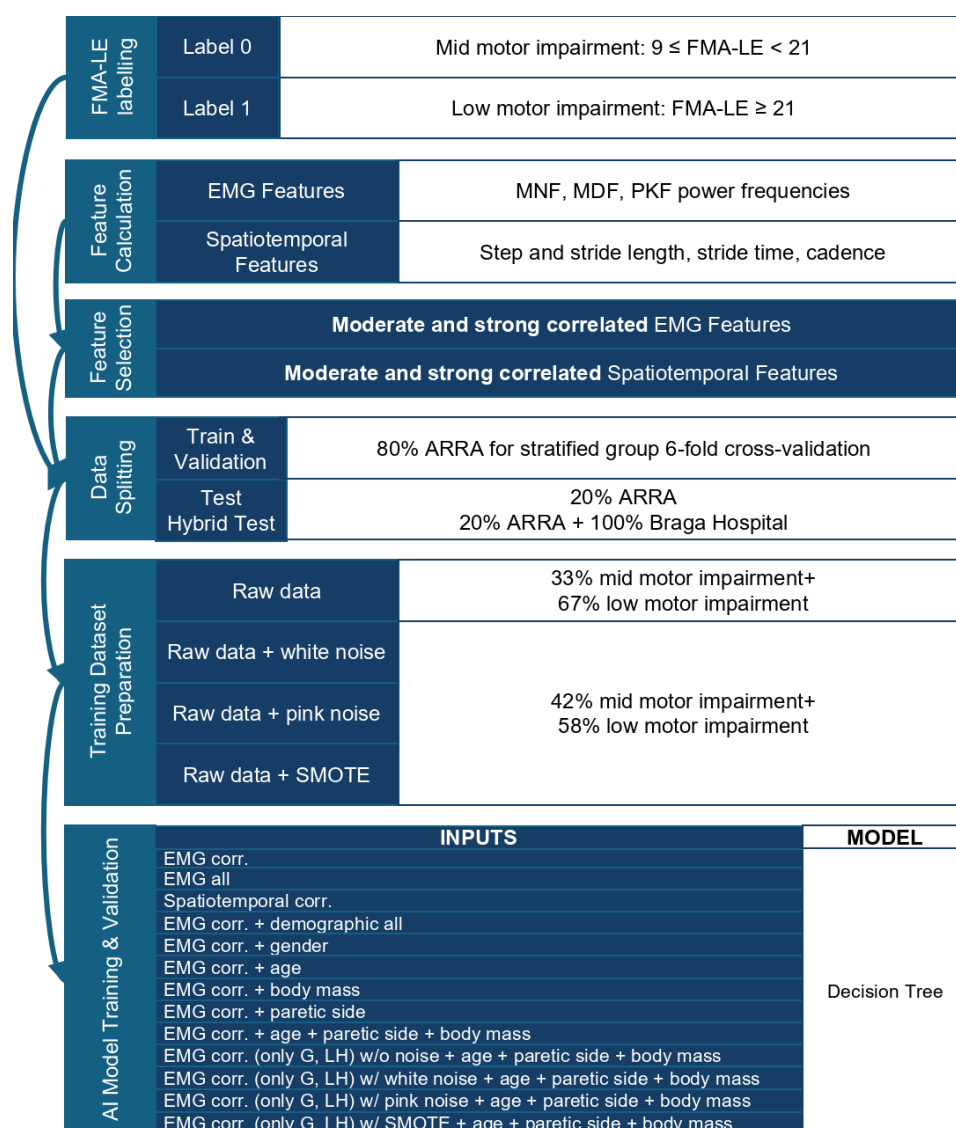


Figure 1. Diagram of the methodological workflow (“corr.”: moderate and strong correlated, “G”: gastrocnemius, “LH”: lateral hamstring).

2.7. Statistical Analysis

Non-parametric Wilcoxon Signed Rank Test was used with the SciPy Python library to evaluate the models’ performance using a confidence level of 0.05 and considering model performance metrics in each fold. The following null hypotheses were analyzed: there are no significant differences regarding the model performance metric between a) correlated vs all EMG features; b) correlated

EMG vs spatiotemporal features; c) correlated EMG features vs combination with demographic features; d) correlated EMG features from gastrocnemius, vastus, gluteus, and lateral hamstring muscles vs correlated EMG features from only gastrocnemius and lateral hamstring muscles (matching the correlated muscles from both ARRA and Hospital of Braga); e) noise-free vs noisy augmented or SMOTE features. The null hypotheses were iteratively defined based on the feature set that achieved the best performance in the previous comparison, which was then used as the reference condition for subsequent tests.

3. Results

The resulting moderate to strong correlated EMG features (Figure S1 in Supplementary Material) comprise all EMG features from paretic gluteus medius (GM, $0.43 \leq r \leq 0.49$) and median and peak frequencies from medial gastrocnemius (MG, $0.46 \leq r \leq 0.53$), vastus medialis (VM, $0.45 \leq r \leq 0.50$), and lateral hamstring (LH, $0.46 \leq r \leq 0.48$). Regarding spatiotemporal features (Figure S2 in Supplementary Material), the moderate to strong correlated ones include paretic and non-paretic stride length ($r = 0.44$), non-paretic step length ($r = 0.57$), paretic stride time ($r = -0.51$), and cadence ($r = 0.51$).

Figure 2 exhibits mean and standard deviation F1, recall, and MCC validation scores considering multiple features set: all EMG features, correlated EMG features, correlated spatiotemporal features, or correlated EMG features combined with demographic features. Correlated EMG features achieved higher training (Table S3 from Supplementary Material) and validation MCC scores than correlated spatiotemporal features, being the training scores significantly different (p -value = 0.03). Furthermore, the use of correlated EMG features revealed higher validation scores than using all features. However, training and validation scores were not considered significantly different (p -value ≥ 0.28). A higher MCC validation score was observed for the combination of EMG features with age, body mass, and paretic side. Individually, although not significantly different, age, body mass, and paretic side increased the average and decreased the standard deviation of the MCC validation score (p -value ≥ 0.14). In contrast, gender decreased the average and increased the standard deviation MCC validation score.

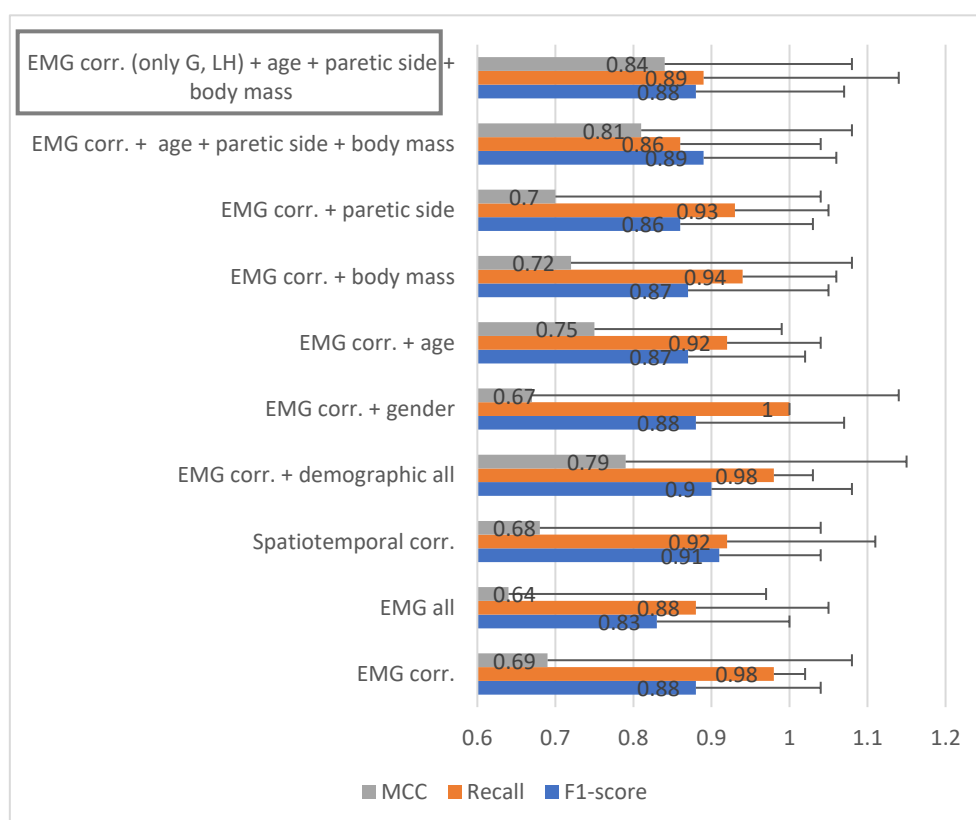


Figure 2. Mean and standard deviation values of F1, recall, and MCC validation scores using different features: all EMG features, strongly correlated (corr.) EMG features, strongly correlated spatiotemporal features, and strongly correlated EMG features combined with demographic features (age, gender, body mass, paretic side). A decision tree classifier was used. The result with the maximum MCC validation score is highlighted. G and LH define gastrocnemius and lateral hamstring muscles, respectively.

Despite the significant decrease in the training scores (p -value = 0.03), the MCC validation score was not significantly affected by using EMG features from only the gastrocnemius and lateral hamstring muscles (p -value ≥ 0.32). Furthermore, the combination of correlated EMG features from only gastrocnemius and lateral hamstring muscles with age, body mass, and paretic side demographic features allowed the MCC validation score of 0.84 ± 0.24 . This combination of features was used in testing the model with the hybrid test set (Figure S5 in Supplementary Material).

Figure 3 shows the mean and standard deviation values of F1, recall, and MCC validation and test scores considering correlated EMG features with the addition of noise or SMOTE samples. The EMG features (only from gastrocnemius and lateral hamstring) were combined with age, paretic side, and body mass demographic features. By analyzing Figure 3, a higher MCC validation score appeared for the non-noisy features. White noise presented a slightly higher MCC validation score (MCC 0.77 ± 0.24) compared with pink noise (MCC 0.72 ± 0.37) and similar with SMOTE features (MCC 0.77 ± 0.35) but is not significantly different compared to non-noisy features (p -value ≥ 0.46). However, the highest test score appeared using both white and pink noisy samples in training data preparation, being considered the best model found (Table S4 and Figures S3–S5 in Supplementary Material). The hybrid dataset presented a lower test score than ARRA.



Figure 3. Mean and standard deviation values of F1, recall, and MCC validation and test scores (ARRA test set or hybrid test set from ARRA and Braga Hospital) using features after different data preparation methods: strongly correlated EMG features (only gastrocnemius and lateral hamstring muscles) combined with demographic features (age, body mass, paretic side) and the addition of noisy (white or pink noise) or SMOTE samples. A decision tree classifier was used. The result with the maximum MCC test score is highlighted. The best hyperparameters found were the gini function, random split strategy, 4 maximum depths of the tree, a minimum of 2 samples to split an internal node and 4 samples required to be at a leaf node, square of features for the number of features to consider for the best split, 8 maximum leaf nodes, balanced weights, and complexity parameter equal to 0.

4. Discussion

Sensors provide real-time data, such as muscle activity, that can objectively capture motor impairment, which can be difficult to assess through visual physical examination, specially to

distinguish close impairment levels. [7] In this manner, this objective assessment has the potential to enhance accuracy in monitoring the post-stroke recovery process. [10] Combining sensor data with ML methods can allow automate support for clinical evaluation. ML algorithms may detect complex patterns in sensor data that might not be immediately apparent, enabling more informed evaluations of motor impairment. [21] After being trained by clinicians who experienced years of exposure to a wide range of post-stroke impairments, the model may have the potential to automatically interpret sensor-based biomarkers to deliver the clinical score, improving accuracy and reliability of clinical decisions. [10] This work evaluates the performance of a decision tree classifier with multiple sensor-based features acquired during walking to estimate post-stroke motor impairment by FMA-LE-derived classes.

Although spatiotemporal parameters revealed an individual higher correlation with FMA-LE-derived classes ($0.44 \leq |r| \leq 0.57$), EMG group features ($0.43 \leq |r| \leq 0.53$) were more suitable to achieve better model performance. In line with the literature, physiological features demonstrated often higher performance scores than biomechanical features for estimating FMA-UE [8,9,11].

Correlated EMG features were shown to be sufficient to achieve higher model performance scores when compared to the use of all EMG features, revealing training and validation scores statistically similar between them. In this context, feature selection based on correlation analysis allowed a reduction in input dimensionality without a significant loss in performance, consistent with prior work using correlation analysis to identify informative and non-redundant features [22].

Considering the correlated EMG features, the results allow a preliminary exploration of sensor placement requirements within this experimental context. The findings indicate that EMG signals from the medial gastrocnemius, vastus medialis, lateral hamstring, and gluteus medius showed stronger association with FMA-LE classes, whereas signals from the tibialis anterior, soleus, rectus femoris, and medial hamstring exhibited weaker correlations. Furthermore, model performance was not significantly affected when only EMG features from the medial gastrocnemius and lateral hamstring were used. These results suggest that, within this dataset, FMA-LE classification during walking may be achievable using EMG data from a reduced set of muscles. This reduction could potentially decrease subject preparation time for sensor placement; however, this observation is specific to the present model and dataset and should be validated in broader populations and experimental conditions.

Although a reduced number of sensors may simplify the acquisition protocol, initial technical training for healthcare professionals, including supervised sessions on proper sensor placement and equipment configuration, is still likely required to ensure reliable data acquisition and accurate model input [23]. Advances in wearable sensor technologies may further reduce the complexity of these procedures over time. Future research should evaluate the extent of training required to achieve consistent and reliable measurements, particularly in comparison with the expertise needed for traditional clinical scale assessments [4].

The addition of demographic features can improve the model performance, consistent with previous findings by Tozlu et al. [11]. Specifically, age, body mass, and paretic side were associated with higher classification scores in this work. These observations are consistent with prior literature indicating that long-term functional recovery after stroke may vary with age, body mass index, and side of impairment, with older age associated with slower or reduced recovery [24], higher body mass index linked to less favorable rehabilitation outcomes [25], and differences in recovery trajectories observed between hemiparetic sides [26]. However, in the present work, these variables should be interpreted as contributing to model discrimination rather than as direct indicators of recovery mechanisms.

Although previous studies report that female patients may experience greater challenges in post-stroke recovery [27], gender did not show a positive contribution to FMA-LE classification performance in this dataset, suggesting that its predictive value may be limited or context-dependent.

The data augmentation did not significantly improve validation scores. However, the addition of white and pink noise data during training was associated with higher test performance, increasing

MCC scores from 0.40 to 0.70. This suggests that noise-based augmentation may have enhanced model robustness to variability in unseen data. Similar findings have been reported in wearable sensor-based machine learning studies, where data augmentation techniques were shown to improve generalization in time-series signals collected from movement disorders [20].

Overall, the best model achieved validation F1-score, recall, and MCC of 0.85 ± 0.16 , 0.84 ± 0.23 , and 0.77 ± 0.24 , respectively. Training and validation scores were similar, indicating that no marked overfitting was observed (Figure S3 in Supplementary Material). Previous studies, such as Song et al. [10] and Riahi et al. [12], have reported strong predictive performance for FMA-UE estimation using alternative data modalities, including smartphone-based movement data and resting-state EEG. However, these studies evaluated regression performance (e.g., R^2), which is not directly comparable to the classification framework adopted in the present work.

The model also achieved acceptable performance on both the ARRA (MCC = 0.70) and hybrid datasets (MCC = 0.60). The lower performance observed in the hybrid test set may reflect reduced generalization capability under dataset shift conditions. This could be related to differences in acquisition protocols, including overground versus treadmill walking, variability in examiner experience, and differences in demographic distributions such as age and body mass (Braga dataset: 45 ± 18 years, 69 ± 15 kg; ARRA dataset: 60 ± 12 years, 92 ± 19 kg). However, these factors cannot be isolated in the present study and therefore their individual contribution cannot be determined. Although data collection constraints such as equipment availability, patient recruitment, and clinical scheduling are common in applied biomedical studies, future work should aim to better control or explicitly model these sources of variability to improve generalizability.

The performance of the decision tree classifier used in this study is likely constrained by the limited sample size and variability of the dataset, which may affect model stability and generalizability. In addition, the results highlight the need for larger and more diverse datasets of post-stroke patients using wearable sensors [10,28]. The availability of such datasets, particularly in open-access formats, would support more robust model training and enable evaluation across heterogeneous populations, thereby improving generalizability and potential clinical applicability.

Despite these limitations, this work demonstrates the feasibility of using EMG-based features acquired through wearable sensors during walking to estimate FMA-LE-derived classes with a decision tree model. Future research should build on these findings by testing whether reduced EMG sensor configurations, targeting a minimal subset of muscles, can maintain classification performance across larger and more diverse post-stroke populations. The contribution of demographic variables should also be explored in greater depth, particularly to determine whether they capture inter-subject variability not represented in sensor-derived features. Furthermore, the reduced performance observed under hybrid testing conditions highlights the need to systematically evaluate the impact of dataset shift, including differences in acquisition protocols such as treadmill versus overground walking. Finally, future studies should assess the trade-off between reduced sensor configurations and operator-dependent factors, such as sensor placement variability and required training, to determine the practical feasibility of deploying such systems in clinical settings.

5. Conclusions

This study demonstrates the feasibility of using EMG-based features acquired through wearable sensors during walking to classify post-stroke motor impairment through FMA-LE-derived classes. The results show that comparable model performance can be achieved using a reduced subset of muscles and that the inclusion of demographic features may enhance classification performance. In addition, noise-based data augmentation was associated with improved test performance, suggesting a potential role in enhancing model generalization. However, model performance remained sensitive to dataset variability and acquisition conditions, highlighting the impact of dataset shift on generalization. These findings support the potential of wearable sensor-based approaches for assisting motor impairment assessment, while emphasizing the need for validation in larger and more diverse populations and under standardized acquisition protocols.

Supplementary Materials: The following supporting information can be downloaded at the website of this paper posted on Preprints.org, Figure S1: boxplot of the correlated EMG features per class; Figure S2: boxplot of spatiotemporal features per class; Table S1: intervals for optimizing hyperparameters; Table S2: best hyperparameters found for each input; Figure S3: training and validation score curves per model complexity; Figure S4: confusion matrix; Figure S5: tree diagram.

Author Contributions: Conceptualization, C.P., J.F. and C.S.; methodology, C.P., J.F. and C.S.; software, L.A. and C.P.; validation, C.P., L.A. and J.F.; formal analysis, C.P., L.A. and J.F.; investigation, C.P., L.A., J.F. and C.S.; resources, C.C., J.C. and C.S.; data curation, C.P., L.A., J.F. and C.S.; writing—original draft preparation, C.P.; writing—review and editing, J.F. and C.S.; visualization, C.P.; supervision, C.S., J.F. and J.C.; project administration, C.S.; funding acquisition, C.S., C.P. and J.F. All authors have read and agreed to the published version of the manuscript.

Funding: This research was funded by the Fundação para a Ciência e Tecnologia under the scholarship reference 2020.05709.BD and under the Stimulus of Scientific Employment with the grant 2020.03393.CEECIND, and by the FEDER Funds through the COMPETE 2020—Programa Operacional Competitividade e Internacionalização (POCI) and P2020 with the Reference Project SmartOs Grant POCI-01-0247-FEDER-039868, under research grant reference POCI-01-0247-FEDER-039868_BI_07_2022_CMEMS.

Institutional Review Board Statement: The study was conducted in accordance with the Declaration of Helsinki, and approved by the Ethics Committee of Hospital of Braga (protocol code CEHB 157_2021 and date of approval 2021).

Informed Consent Statement: Informed consent was obtained from all subjects involved in the study. Written informed consent has been obtained from the patient(s) to publish this paper.

Data Availability Statement: The ARRA dataset is publicly available as described in the original publication. The clinical dataset collected at the Hospital of Braga is not publicly available due to the limited size of the cohort and to avoid potential re-identification risks. Data supporting the findings of this study are available from the corresponding author upon reasonable request.

Acknowledgments: The results published here are partially based on data obtained from Steven A. Kautz and Richard R. Neptune's Dataset15: Medical University of South Carolina Stroke Data (ARRA) (ICPSR 37122).

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

Abbreviations

The following abbreviations are used in this manuscript:

EEG	electroencephalographic
EMG	electromyographic
FMA	Fugl-Meyer Assessment
FMA-LE	Fugl-Meyer Lower Extremity Assessment
FMA-UE	Fugl-Meyer Upper Extremity Assessment
GRF	ground reaction forces
MCC	Mathew Correlation Coefficient
MDF	median power frequency
ML	Machine Learning
MNF	mean power frequency
PKF	peak power frequency
SMOTE	Synthetic Minority Over-sampling Technique
TUG	Timed Up and Go

References

1. Bushnell, C.; Bettger, J.P.; Cockroft, K.M.; Cramer, S.C.; Edelen, M.O.; Hanley, D.; Katzan, I.L.; Mattke, S.; Nilsen, D.M.; Piquado, T.; et al. Chronic Stroke Outcome Measures for Motor Function Intervention Trials. *Circ. Cardiovasc. Qual. Outcomes* **2015**, *8*, S163–S169, doi:10.1161/CIRCOUTCOMES.115.002098.

2. Duncan, P.W.; Propst, M.; Nelson, S.G. Reliability of the Fugl-Meyer Assessment of Sensorimotor Recovery Following Cerebrovascular Accident. *Phys. Ther.* **1983**, *63*, 1606–1610, doi:10.1093/ptj/63.10.1606.
3. Sanford, J.; Moreland, J.; Swanson, L.R.; Stratford, P.W.; Gowland, C. Reliability of the Fugl-Meyer Assessment for Testing Motor Performance in Patients Following Stroke. *Phys. Ther.* **1993**, *73*, 447–454, doi:10.1093/ptj/73.7.447.
4. Sullivan, K.J.; Tilson, J.K.; Cen, S.Y.; Rose, D.K.; Hershberg, J.; Correa, A.; Gallichio, J.; McLeod, M.; Moore, C.; Wu, S.S.; et al. Fugl-Meyer Assessment of Sensorimotor Function After Stroke. *Stroke* **2011**, *42*, 427–432, doi:10.1161/STROKEAHA.110.592766.
5. Quinn, T.; Harrison; McArthur Assessment Scales in Stroke: Clinimetric and Clinical Considerations. *Clin. Interv. Aging* **2013**, *201*, doi:10.2147/CIA.S32405.
6. Gladstone, D.J.; Danells, C.J.; Black, S.E. The Fugl-Meyer Assessment of Motor Recovery after Stroke: A Critical Review of Its Measurement Properties. *Neurorehabil. Neural Repair* **2002**, *16*, 232–240, doi:10.1177/154596802401105171.
7. Routson, R.L.; Kautz, S.A.; Neptune, R.R. Modular Organization across Changing Task Demands in Healthy and Poststroke Gait. *Physiol. Rep.* **2014**, *2*, e12055, doi:10.14814/phy2.12055.
8. Julianjatsono, R.; Ferdiana, R.; Hartanto, R. High-Resolution Automated Fugl-Meyer Assessment Using Sensor Data and Regression Model. In Proceedings of the 2017 3rd International Conference on Science and Technology - Computer (ICST); IEEE, July 2017; pp. 28–32.
9. Gebruers, N.; Truijien, S.; Engelborghs, S.; De Deyn, P.P. Prediction of Upper Limb Recovery, General Disability, and Rehabilitation Status by Activity Measurements Assessed by Accelerometers or the Fugl-Meyer Score in Acute Stroke. *Am. J. Phys. Med. Rehabil.* **2014**, *93*, 245–252, doi:10.1097/PHM.0000000000000045.
10. Song, X.; Chen, S.; Jia, J.; Shull, P.B. Cellphone-Based Automated Fugl-Meyer Assessment to Evaluate Upper Extremity Motor Function After Stroke. *IEEE Trans. Neural Syst. Rehabil. Eng.* **2019**, *27*, 2186–2195, doi:10.1109/TNSRE.2019.2939587.
11. Tozlu, C.; Edwards, D.; Boes, A.; Labar, D.; Tsagaris, K.Z.; Silverstein, J.; Pepper Lane, H.; Sabuncu, M.R.; Liu, C.; Kuceyeski, A. Machine Learning Methods Predict Individual Upper-Limb Motor Impairment Following Therapy in Chronic Stroke. *Neurorehabil. Neural Repair* **2020**, *34*, 428–439, doi:10.1177/1545968320909796.
12. Riahi, N.; Vakorin, V.A.; Menon, C. Estimating Fugl-Meyer Upper Extremity Motor Score From Functional-Connectivity Measures. *IEEE Trans. Neural Syst. Rehabil. Eng.* **2020**, *28*, 860–868, doi:10.1109/TNSRE.2020.2978381.
13. Chen, S.; Lin, X.; Fu, J.; Qian, Y.; Chen, Z.; Huang, Z.; Liu, Q.; Lu, X.; Jia, J. Prediction of the Hand Function Part of the Fugl-Meyer Scale after Stroke Using an Automatic Quantitative Assessment System. *Brain-X* **2023**, *1*, doi:10.1002/brx.2.26.
14. Rech, K.D.; Salazar, A.P.; Marchese, R.R.; Schifino, G.; Cimolin, V.; Pagnussat, A.S. Fugl-Meyer Assessment Scores Are Related With Kinematic Measures in People with Chronic Hemiparesis after Stroke. *J. Stroke Cerebrovasc. Dis.* **2020**, *29*, 104463, doi:10.1016/j.jstrokecerebrovasdis.2019.104463.
15. Kautz, S.A.; Neptune, R.R. Medical University of South Carolina Stroke Data (ARRA) Available online: <https://www.icpsr.umich.edu/web/ICPSR/studies/37122>.
16. Oskoei, M.A.; Huosheng Hu Support Vector Machine-Based Classification Scheme for Myoelectric Control Applied to Upper Limb. *IEEE Trans. Biomed. Eng.* **2008**, *55*, 1956–1965, doi:10.1109/TBME.2008.919734.
17. Dancey, C.P.; Reidy, J. *Statistics without Maths for Psychology*; Pearson education, 2007;
18. Smith, M.-C.; Barber, A.P.; Scrivener, B.J.; Stinear, C.M. The TWIST Tool Predicts When Patients Will Recover Independent Walking After Stroke: An Observational Study. *Neurorehabil. Neural Repair* **2022**, *36*, 461–471, doi:10.1177/15459683221085287.
19. Kwong, P.W.H.; Ng, S.S.M. Cutoff Score of the Lower-Extremity Motor Subscale of Fugl-Meyer Assessment in Chronic Stroke Survivors: A Cross-Sectional Study. *Arch. Phys. Med. Rehabil.* **2019**, *100*, 1782–1787, doi:10.1016/j.apmr.2019.01.027.
20. Um, T.T.; Pfister, F.M.J.; Pichler, D.; Endo, S.; Lang, M.; Hirche, S.; Fietzek, U.; Kulić, D. Data Augmentation of Wearable Sensor Data for Parkinson's Disease Monitoring Using Convolutional Neural Networks. In

- Proceedings of the Proceedings of the 19th ACM International Conference on Multimodal Interaction; ACM: New York, NY, USA, November 3 2017; pp. 216–220.
21. Breiman, L. *Classification And Regression Trees*; TAYLOR & FRANCIS LTD, Ed.; 1984; ISBN 9780412048418.
 22. Kumar, S.; Chong, I. Correlation Analysis to Identify the Effective Data in Machine Learning: Prediction of Depressive Disorder and Emotion States. *Int. J. Environ. Res. Public Health* **2018**, *15*, 2907, doi:10.3390/ijerph15122907.
 23. Hilty, D.M.; Armstrong, C.M.; Edwards-Stewart, A.; Gentry, M.T.; Luxton, D.D.; Krupinski, E.A. Sensor, Wearable, and Remote Patient Monitoring Competencies for Clinical Care and Training: Scoping Review. *J. Technol. Behav. Sci.* **2021**, *6*, 252–277, doi:10.1007/s41347-020-00190-3.
 24. Yoo, J.; Hong, B.; Jo, L.; Kim, J.-S.; Park, J.; Shin, B.; Lim, S. Effects of Age on Long-Term Functional Recovery in Patients with Stroke. *Medicina (B. Aires)*. **2020**, *56*, 451, doi:10.3390/medicina56090451.
 25. Leszczak, J.; Czenczek-Lewandowska, E.; Przysada, G.; Baran, J.; Weres, A.; Wyszynska, J.; Mazur, A.; Kwolek, A. Association Between Body Mass Index and Results of Rehabilitation in Patients After Stroke: A 3-Month Observational Follow-Up Study. *Med. Sci. Monit.* **2019**, *25*, 4869–4876, doi:10.12659/MSM.915586.
 26. Bindawas, S.M.; Mawajdeh, H.M.; Vennu, V.S.; Alhaidary, H.M. Functional Recovery Differences after Stroke Rehabilitation in Patients with Uni- or Bilateral Hemiparesis. *Neurosciences* **2017**, *22*, 186–191, doi:10.17712/nsj.2017.3.20170010.
 27. Kim, J.-S.; Lee, K.-B.; Roh, H.; Ahn, M.-Y.; Hwang, H.-W. Gender Differences in the Functional Recovery after Acute Stroke. *J. Clin. Neurol.* **2010**, *6*, 183, doi:10.3988/jcn.2010.6.4.183.
 28. Sanchez, N. Stroke Initiative for Gait Data Evaluation (STRIDE), United States, 2012-2020 Available online: <https://www.icpsr.umich.edu/web/ICPSR/studies/38002#>.

Disclaimer/Publisher’s Note: The statements, opinions and data contained in all publications are solely those of the individual author(s) and contributor(s) and not of MDPI and/or the editor(s). MDPI and/or the editor(s) disclaim responsibility for any injury to people or property resulting from any ideas, methods, instructions or products referred to in the content.