Journal of NeuroEngineering and Rehabilitation

RESEARCH

Open Access

A multimodal machine learning approach to forecast upper limb motor recovery after stroke using kinematic and electromyographic data – A pilot-study

Luigi Privitera[1,2*], Michael Lassi[1], Stefania Dalise[3], Valentina Azzollini[4], Luca Maggiani[4], Adrian Guggisberg[5,6], Alberto Mazzoni[1], Carmelo Chisari[3,4], Silvestro Micera[1,7,8] and Andrea Bandini[1,9]

Abstract

Background Forecasting post-stroke rehabilitation outcome is essential to personalize therapeutic strategies. Traditional approaches rely on clinical assessment scales, which, while essential, may benefit from complementary objective measures. In this direction, robot-assisted assessment offers the unprecedented possibility of precisely collecting patients' physiological signals, enabling a data-driven approach to recovery assessments. Leveraging the use of multimodal features collected during assessment sessions performed before and after one month of robotic rehabilitation, we developed a machine learning approach to forecast the upper limb (UL) motor recovery in stroke survivors after rehabilitation.

Methods This study evaluated a 4-week rehabilitation program, using both standard physical therapy and the ALEx robot to promote UL motor recovery in 11 subacute stroke survivors, compared to 6 healthy individuals. Kinematic measures and surface electromyography (sEMG) were collected during a 3D reaching task involving six target points. From these tasks, 76 sEMG features, 18 kinematics features and 1 multimodal feature were extracted. To forecast the UL motor recovery post intervention, a two-step machine learning approach was devised: a machine learning regression model was developed and validated to predict the Fugl-Meyer Assessment for UL (FMA-UL) post rehabilitation, whereas an anomaly detection algorithm identified patients who exhibited limited or no motor recovery post intervention. The anomaly detection approach used a fully-connected autoencoder that identified patients with reduced recovery likelihood. The regression models, optimized via a nested Leave-One-Subject-Out approach, guided feature selection and refined hyperparameters to predict FMA-UL scores post intervention.

Results The optimized regression model achieved an RMSE in predicting the FMA-UL post intervention of 5.45. The autoencoder effectively identified patients with reduced recovery potential, showing a higher distribution of reconstruction errors for these individuals.

*Correspondence:
Luigi Privitera
luigi.privitera@santannapisa.it

Full list of author information is available at the end of the article

Conclusions The findings confirm that combining kinematic and sEMG data improves motor recovery assessment. The proposed machine learning approach holds potential for aiding clinicians and therapists in identifying patients who are more likely to recover UL motor functions before rehabilitation begins. By accurately predicting recovery outcomes, this method can help guide the development of personalized therapeutic strategies, optimizing treatment planning in advance.

Keywords Stroke recovery, Machine learning, Rehabilitation robots, Upper limb recovery, Surface EMG, Kinematics

Introduction

Stroke is the second leading cause of global disability, significantly impairing the quality of life of the affected individuals, with only 20% of them managing to resume their previous professional and personal activities [1, 2]. Restoring upper limb (UL) motor functions is crucial for stroke survivors, as even minor improvements can substantially enhance their independence [3, 4].

Intensive training, particularly during the acute phase, is essential for promoting functional recovery and preventing complications associated with inactivity [5]. The use of exoskeletons and rehabilitation robots has shown promising results in supporting the recovery of stroke survivors [6, 7]. Although a range of neurorehabilitation technologies for promoting UL recovery are available [8], predicting which patients will benefit from these tools remains a challenge [9]. This prediction largely depends on clinical assessment scales, which, despite their limitations, remain the standard for evaluating post-stroke motor impairment [10]. For instance, the Fugl-Meyer Assessment (FMA) scale – one of the most used tools for evaluating post-stroke motor impairment [11] – exhibits a certain degree of inter-rater and intra-rater variability [12, 13], and is influenced by floor and ceiling effects [14, 15]. Nevertheless, it continues to serve as a key benchmark for both clinical assessments and machine learning models.

Therefore, there is a significant need to develop predictive models to forecast the effectiveness of specific rehabilitation interventions. One such model is the proportional recovery rule (PRR), which suggests that most stroke survivors typically recover around 70% of their initial FMA score [16, 17]. However, this model has primarily been used to predict spontaneous recovery in the acute post-stroke phase and, to the best of our knowledge, has not been adapted to predict responses to rehabilitative interventions.

Over the past 20 years, the evolution of upper limb rehabilitation methodologies has seen an increase in the adoption of robotic technologies, particularly exoskeletons [18, 19]. These devices not only facilitate patients in rehabilitation activities but also offer valuable support during assessment sessions [20–22], providing precise kinematic measures, particularly during functional tasks such as "reaching and grasping," frequently used in post-stroke upper limb rehabilitation [19, 23, 24]. For this reason, in this study we adopted a multi-modal approach, integrating surface sEMG and kinematic data, extracted during clinical assessment using a robotic exoskeleton, with the specific goal of predicting upper limb motor recovery outcomes.

Studies in the literature that focus on predicting upper limb functional recovery, after robotic therapy, are very limited. Most rely solely on clinical or kinematic data [25–27], and while some have explored EEG data [28, 29], the use of sEMG remains particularly scarce. Importantly, the few available studies show that sEMG can effectively track motor recovery post-stroke [30, 31], which motivated us to investigate whether sEMG may also hold predictive value in this context with a data-driven approach.

The current literature underexplores the joint potential of kinematic and electrophysiological information captured during robotic-assisted assessments, particularly when combined with machine learning approaches, to enhance the understanding of patient progress and improve outcome prediction. Our hypothesis is that, kinematic data obtained from robotic rehabilitation devices, when combined with sEMG data, can form the basis for multimodal predictive models aimed at identifying which patients are most likely to benefit from rehabilitative interventions. In fact, sEMG and kinematic data provide complementary perspectives on motor function: sEMG captures the neural drive and muscular activation patterns that precede movement, while kinematics describes the resulting motion [32, 33].

Hence, in this study, we introduce a novel multimodal machine learning approach to predict the degree of UL motor recovery after standard and robotic therapy with the Arm Light Exoskeleton Rehab Station (ALEx RS) [34, 35]. Our method integrates sEMG and kinematic data in a data-driven framework to achieve two objectives: (1) to predict changes in FMA-UL scores after four weeks of intervention, and (2) to identify patients unlikely to benefit from the treatment using anomaly detection techniques. Notably, our models use only data collected during the initial visit to forecast outcomes at the end of the treatment period.

The focus of this work is on the subacute phase of recovery, with the aim of developing a practical tool to support physiotherapists in making timely, personalized decisions about the most appropriate therapy for each patient.

Materials and methods

Participants

Eleven individuals recovering from stroke were recruited (6 females, average age: 61 ± 14 years old; Table 1). All participants had an ischemic stroke within the last 2 to 6 weeks before the experiments, showing right or left-side hemiplegia and at least 10° of residual movement in their shoulder and elbow joints (see Table 1 for details). Additionally, six neurologically intact individuals (4 females, age 58 ± 16 years old) were also included as the control group, to gather comparable data for evaluating the therapy's outcome. The research was carried out at the Neurorehabilitation Unit of the University Hospital of Pisa (Cisanello hospital), Italy, and the University Hospital of Geneva (HUG), Switzerland. Ethical approval was obtained from the Commission Cantonale d'Ethique de la Recherche (CCER) de Genève in Switzerland and the Comitato Etico Area Vasta Nord Ovest (CEAVNO) in Pisa. All participants signed informed consent according to the requirements of the Declaration of Helsinki.

Experimental set-up

The experimental setup and protocol follow those outlined in the study by Pierella et al. [31]. All stroke survivors participated in a four-week training program, which included either three 30-minute sessions per week with the ALEx robot (10 subjects IDs [01,02,04,05,06,07,08 ,09,10,11]) or additional standard physical therapy (ID 03) (Tab.1). In addition to robot-assisted training, participants received standard therapy consisting of two 30-minute sessions per day, five days a week. The difference in rehabilitation protocol arises because the participants in this study were drawn from a larger research project that included multiple rehabilitation groups. All participants, regardless of their assigned rehabilitation protocol, underwent the same robotic assessments used in the current study.

Pre- and post-training assessments were conducted in all participants using the ALEx robot (denoted as A_{pre} and A_{post}, respectively) to evaluate the effect of the rehabilitation protocol on UL motor functions after one month. Control participants underwent the same assessment session with ALEx robot, but only once. In each assessment session, a physiotherapist administered the FMA to the stroke survivors. The upper limb section (FMA-UL, ranging from 0 to 66) was used as the outcome measure to be predicted with the proposed machine learning approach.

Both the robotic assessments and treatment sessions with the ALEx robot involved a 3D point-to-point reaching task. In this task, participants were required to reach 18 targets on the surface of a virtual sphere using their impaired arm (see Fig. 1).

They were instructed to reach each target and return to the center at a comfortable speed, with visual feedback displayed on a monitor in front of them. The spherical workspace was aligned with the acromion of the right arm, ensuring maximum exploration while accommodating different body sizes.

During the assessment sessions, participants were asked to reach all 18 targets within 30 min, whereas the training sessions focused on 8 targets selected by the therapist. If participants had difficulty reaching a target, the ALEx robot provided assistance using a minimum speed profile [35, 36].

For this study, we focused on 6 targets, collected during the assessment sessions, which represent the three primary movement directions (highlighted in blue in Fig. 1): upward and downward, rightward and leftward, and toward and away from the body. Only movements in which participants were able to actively reach the target without assistance from the robot were included in the analysis. In each assessment session, participants performed 1 to 5 repetitions per target. For each target,

Table 1 List of participants' data: age, sex, time since the stroke, FMA-UL scores during the assessment sessions (before the therapy, A_{pre}, and after therapy, A_{post}), type of therapy

IDs	Age (years)	Sex	Time since stroke	FMA-UL A_{pre}	FMA-UL A_{post}	Therapy
01	69	F	4 weeks	49	56	ST + RT
02	34	F	4 weeks	34	46	ST + RT
03	79	F	3 weeks	27	29	ST
04	73	F	6 weeks	59	63	ST + RT
05	64	M	6 weeks	61	62	ST + RT
06	82	F	3 weeks	48	55	ST + RT
07	55	M	3 weeks	18	26	ST + RT
08	65	M	5 weeks	27	45	ST + RT
09	56	M	2 weeks	50	63	ST + RT
10	49	F	2 weeks	60	63	ST + RT
11	43	M	3 weeks	54	62	ST + RT

(standard therapy – ST; standard therapy + robotic therapy – ST+RT)

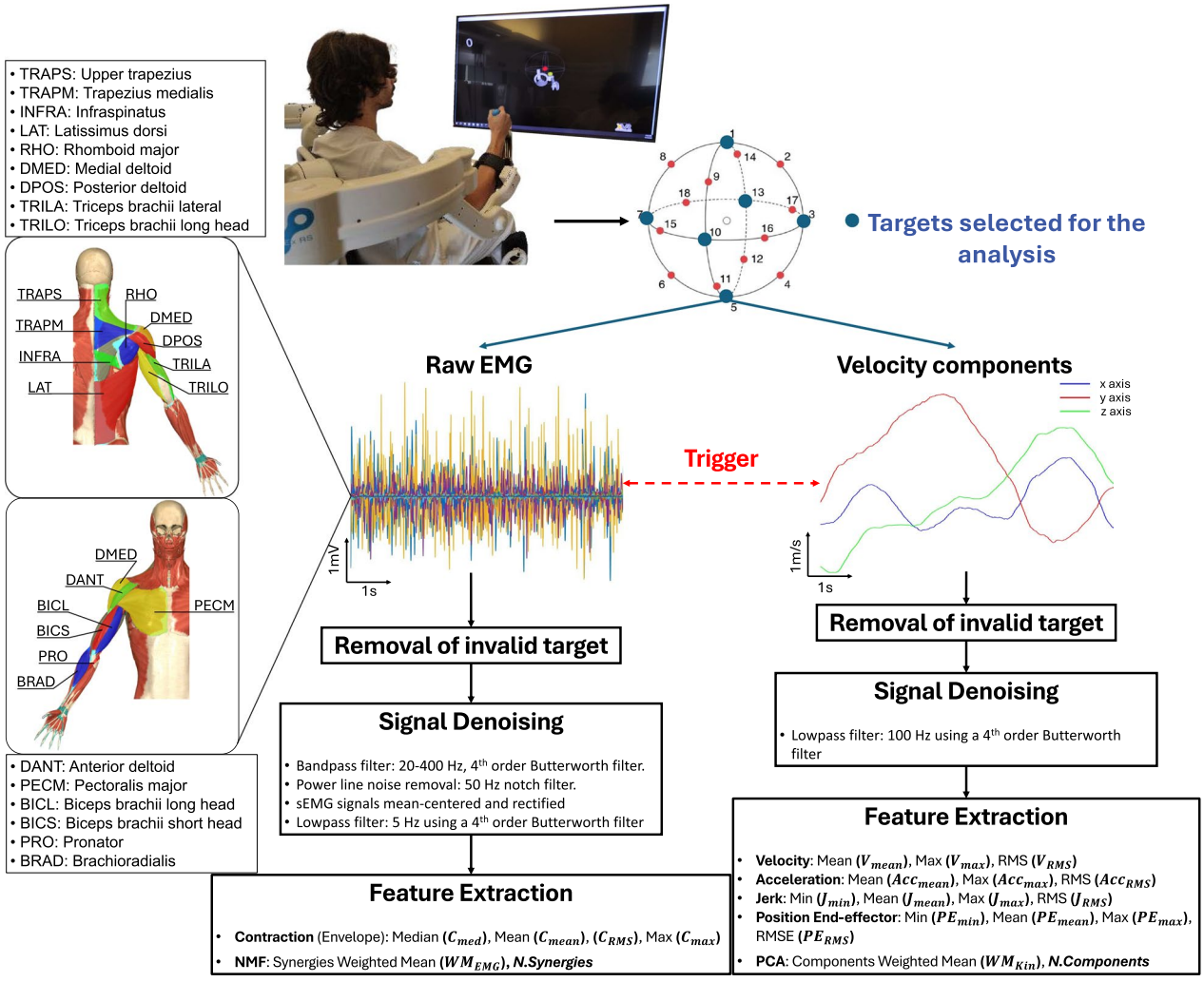

Fig. 1 Overview of the acquisition and pre-processing pipeline: Participants were asked to reach the target shown on the screen in front of them while sEMG and kinematic data were collected during the movements. Data were synchronized by a trigger and once collected they were filtered and used for feature extraction. The sphere shows all possible reaching targets: highlighted in blue are the 6 targets selected in this study for the analyses, and the type of data extracted for each of movement. For each movement, sEMG signals from 15 muscles were collected, filtered, and rectified to extract features such as envelopes and synergies. Velocity components were also collected, filtered, and used to derive kinematic features, including velocity, acceleration, jerk, end-effector position, and PCA components. Data corresponding to the same movement type were concatenated, resulting in 12 unique movements (2 distinct movements for each of the 6 targets)

participants performed two types of movements: one involved reaching the target, and the other involved returning to the resting position. Multiple repetitions of each movement type were concatenated, creating two combined sets of movements per target. Hence, across the six targets, this resulted in a total of 12 distinct movement types for each participant.

During these assessment sessions, we recorded the sEMG signals from specific upper body muscles and the kinematics of the end-effector, from which we extracted features to train the predictive machine learning algorithms.

sEMG recordings

We recorded the activity of the 15 muscles that most of all are involved in arm lifting, reaching, and grasping movements [37, 38], as illustrated in Fig. 1: upper trapezius (TRAPS), trapezius medialis (TRAPM), anterior deltoid (DANT), medial deltoid (DMED), posterior deltoid (DPOS), pectoralis major (PECM), latissimus dorsi (LAT), infraspinatus (INFRA), rhomboid major (RHO), biceps brachii long head (BICL), biceps brachii short head (BICS), brachioradialis (BRAD), triceps brachii lateral (TRILA), triceps brachii long head (TRILO), and pronator (PRO). The sEMG was recorded with a Noraxon Desktop DTS wireless system at 1.5 kHz, using superficial Ag-AgCl electrodes (Kendall H124SG, ECG electrodes

30 x 24 mm) after proper skin preparation. Electrodes were placed according to the guidelines from the European Community project Surface Electromyography for Non-Invasive Assessment of Muscles (SENIAM) [39] and anatomical guidelines [40].

Kinematic recordings

To obtain precise and consistent measurements of upper limb kinematics during assessment sessions, we used the end-effector of the ALEx robot. Robotic exoskeletons like ALEx provide high-resolution kinematic data across repeated movements, which is essential for quantitatively assessing motor performance and recovery in post-stroke rehabilitation. ALEx is a CE-certified medical device whose measurement capabilities have been validated in prior clinical studies [41, 42], supporting its reliability for research and clinical applications.

Kinematic data were obtained from the positions of the handle, serving as the end-effector of the exoskeleton, and were recorded using ALEx during each reaching task. The measurements included the point-to-point position and the speed of the handle, sampled at 1 kHz. The start and end of each movement were determined around the target's appearance by ALEx's software on a trial-by-trial basis, identifying instances where the velocity profile of the exoskeleton's end-effector exceeded or dropped below 2% of the local maximum value.

Feature extraction

For each movement, 76 sEMG features, 18 kinematic features, 1 multimodal feature and 1 clinical feature were extracted:

1. sEMG features: 15 median contractions (C_{med}; where "contraction" refers to the envelope of the sEMG signal of each muscle), 15 mean contractions (C_{mean}), 15 root mean square values contraction (C_{RMS}), 15 maximum contractions (C_{max}), 15 weighted mean values of muscle synergies' activation WM_{EMG}, 1 number of muscle synergies (*N.Synergies*).
2. Kinematic features: 1 velocity mean value (V_{mean}), 1 velocity maximum value (V_{max}), 1 velocity root mean square value (V_{RMS}), 1 acceleration mean value (Acc_{mean}), 1 acceleration maximum value (Acc_{max}), 1 acceleration root mean square value (Acc_{RMS}), 1 jerk minimum value (J_{min}), 1 jerk mean value (J_{mean}), 1 jerk maximum value (J_{max}), 1 jerk root mean square value (J_{RMS}), 1 position end-effector minimum value (PE_{min}), 1 position end-effector mean value (PE_{mean}), 1 position end-effector maximum value (PE_{max}), 1 position end-effector root mean square value (PE_{RMS}), 3 weighted means of principal component weights (WM_{Kin}), and 1 number of principal components ($N_{Components}$).
3. Multimodal feature: 1 Gesture Similarity Index *GSI*.
4. Clinical feature: FMA-UL score assessed at A_{pre}.

An in-depth description of the above features is reported in the following sections.

sEMG: pre-processing and feature extraction

The raw sEMG signals were detrended and bandpass filtered between 20–400 Hz using a 4th order Butterworth filter. Power line noise was eliminated with a 50 Hz notch filter. Subsequently, the sEMG signals were mean-centered, rectified, and then low-pass filtered with a cutoff frequency of 5 Hz using a 4th order Butterworth filter to obtain envelopes [43, 44]. To mitigate variations in sEMG amplitudes caused by electrode placement and to prevent bias against muscles with lower amplitudes during synergy extraction, the sEMG signal for each muscle during an entire movement, was normalized to its median value, calculated individually for each muscle and each participant across each session. Median normalization, as opposed to using the maximum value, offers better robustness to outliers [43]. All trials directed towards the same target were concatenated into movements, and the following parameters were extracted for each muscle: C_{med}, C_{mean}, C_{RMS}, C_{max}, and muscle synergies. Muscle synergies of each movement, for each participant, were identified using the non-negative matrix factorization (NNMF) algorithm [45]. This algorithm decomposes the sEMG envelope into positive components or synergies, organized into two matrices: the H_{EMG} matrix, which contains the time course of the synergies, and the W_{EMG} matrix, representing the weight of muscle activations in each synergy. The product $H_{EMG} \times W_{EMG}$ reconstructs the original sEMG envelope matrix with minimal error [45]. To address the issue of local minima, 50 extraction repetitions were performed with different initializations, selecting the iteration that explained the highest sEMG variance. The minimum number of synergies required was determined by employing the variance accounted for (VAF) method, setting the threshold at the number of synergies where VAF exceeded 95% [43]. The varying number of synergies for each movement led to the creation of W_{EMG} matrices with dimensions $s \times M$, where s represents the number of synergies (which varies for each movement), and M is the number of muscle channels, consistent across all movements. To enable the comparison of W_{EMG} matrices with different numbers of synergies, a method was devised to harmonize their sizes:

1. For each synergy of the W_{EMG} matrix, a ranking of muscle activation coefficients is performed.

2. Each activation coefficient is multiplied by its position in the synergy ranking (e.g., the lowest coefficient is multiplied by 1, the second by 2, and so on).
3. The average of the new coefficients, obtained in the previous step, is calculated for synergies related to the same movement.

This process results in a matrix, WM_{EMG}, of size $1\times M$ for each movement, suitable for comparison. In this new matrix each element is the result of:

$$WM_m = \frac{\sum_{s=1}^{S}(A_{s,m}\cdot p_{s,m})}{\sum_{s=1}^{S}(p_{s,m})}, m = 1\dots 15 \quad (1)$$

where WM_m are the weighted mean values of muscle *m* along all its synergies *S*, $A_{s,m}$ is the muscle activation coefficient of muscle *m* and synergy *s*, $p_{s,m}$ is the ranking of muscle *m* inside the synergy *s*.

This matrix represents the weighted average of the contributions of each muscle to the synergies of the movement it describes. While a typical average would flatten the contribution values, using the rankings as weights for the average allowed us to preserve the distinctiveness of the muscle contributions within individual synergies.

Kinematics: pre-processing and feature extraction

The raw speed data of the end-effector was lowpass filtered at 100 Hz, using a 4th order Butterworth filter, and then used to derive all kinematic features. The velocity magnitude was differentiated twice to calculate the acceleration and jerk. Various kinematic features were computed, including the tangential velocity, tangential acceleration, and tangential jerk of the handle. For each kinematic feature, the mean value (V_{mean}, Acc_{mean}, and J_{mean}, respectively), the maximum value (V_{max}, Acc_{max}, and J_{max}, respectively), the root mean square (V_{RMS}, Acc_{RMS}, and J_{RMS}, respectively), and the minimum value of the jerk (J_{min}) were determined. Furthermore, measures regarding the interaction with the target were considered, such as the distance from the target, expressed as "position end-effector" (*PE*). From this, the minimum value (PE_{min}), the mean value (PE_{mean}), the maximum value (PE_{max}) and the root mean square (PE_{RMS}) were calculated. The acceleration signal provides rich and dynamic information about movement patterns, making it particularly well-suited for capturing meaningful variations in motor control [46]. Acceleration is also highly sensitive to changes such as corrections, hesitations, or jerky transitions, which are often indicative of neurological impairments or adaptations [47, 48]. In contrast, position and velocity can mask these small but critical dynamic changes due to smoothing effects, as position accumulates over time and averages out fluctuations, and velocity is less sensitive to abrupt variations in movement. For these reasons, we conducted a more in-depth analysis of the acceleration signal by means of Principal Component Analysis (PCA). The number of components was chosen to achieve 80% of the R^2 value. The resulting matrices, W_{Kin} and H_{Kin}, were then used similarly to the NNMF process described earlier, computing the matrix WM_{Kin} with dimensions 1x3.

Gesture similarity index (GSI)

By combining WM_{EMG} and WM_{Kin}, we computed the GSI. This metric was originally developed for telerehabilitation purposes for stroke survivors, where an algorithm evaluates the similarity between movements performed by patients during rehabilitation sessions with their therapists and those performed independently at home during exercise sessions [49]. The result is a score indicating how similar the participants' gestures are to those of a normative reference group, with greater similarity resulting in a higher score. In this work, we chose to repurpose the GSI to compare the movements of stroke survivors with those of the neurotypical individuals. The metrics used for GSI are described below:

1. First, reference values (*Ref*) were extracted by using the WM_{EMG} and WM_{Kin} matrices from the healthy individuals:

$$\begin{aligned} Ref_{EMG} &= \frac{WM_{EMG_{Healthy}}}{N.Synergies_{Healthy}} \\ Ref_{Kin} &= \frac{WM_{Kin_{Healthy}}}{N.Components_{Healthy}} \cdot J_{mean_{Healhty}} \end{aligned} \quad (2)$$

   In these equations, $WM_{EMG_{Healthy}}$ and $WM_{Kin_{Healthy}}$ represent the WM_{EMG} and WM_{Kin} matrices calculated from the NNMF and PCA applied to the sEMG and kinematic data collected from healthy control participants. *N.Synergies* is the number of synergies extracted through NNMF for the specific movement, and *N.Components* is the number of principal components extracted through the PCA for the specific movement. $J_{mean_{Healthy}}$ represents the mean jerk of the reference group and is used to scale the reference value in proportion to the smoothness of the specific movement.
2. Similarly, movement metrics (*Mov*) were extracted from the participants:

$$Mov_{EMG} = \frac{WM_{EMG_{Patients}}}{N.Synergies_{Patients}}$$
$$Mov_{Kin} = \frac{WM_{Kin_{Patients}}}{N.Components_{Patients}} \cdot J_{mean_{Patients}} \quad (3)$$

By including J_{mean} in both Ref_{Kin} and Mov_{Kin}, the reference and movement metrics are scaled to reflect the smoothness of movements, a key indicator of motor control. The mean jerk emphasizes the role of movement smoothness in the kinematic reference metric, as smooth movements are a distinctive sign of healthy motor control and a clinically relevant indicator of recovery in patients [48]. Individuals with motor impairments, such as stroke survivors, often demonstrate jerky or irregular movements due to neuromuscular deficits [50]. By including jerk characteristics, this approach captures differences in movement quality between patients and healthy individuals, ensuring the similarity comparison accounts for spatial, temporal, and smoothness characteristics, providing a comprehensive evaluation of movement.

3. The similarity score for both the sEMG and the kinematics characteristics was then calculated:

$$GSI_{EMG} = 1 - \frac{|Ref_{EMG} - Mov_{EMG}|}{Ref_{EMG}}$$
$$GSI_{Kin} = 1 - \frac{|Ref_{Kin} - Mov_{Kin}|}{Ref_{Kin}} \quad (4)$$

4. Finally, the cumulative GSI (GSI_{TOT}) was calculated as the average value between GSI_{EMG} and GSI_{Kin}.

The score ranges from 0 to 1, where higher values indicate greater similarity between the participant's movements and the reference movements derived from the normative group. A score of 1 reflects perfect similarity, whereas the score decreases towards 0 with the increasing of discrepancies in the assessed metrics. This score captures not only how closely the movements align with the normative standard, but also which movements deviate most significantly, offering valuable insights for designing and optimizing therapeutic interventions.

Machine learning algorithms for recovery prediction

We developed and validated a two-step approach to predict UL recovery using robotic assessment with ALEx. In the first step, we predicted the post-treatment FMA-UL score for all patients using the extracted features at A_{pre}. We also analyzed the model to improve its explainability, identifying the most predictive features. In the second step, we automatically identified patients who exhibited limited or no recovery after the treatment, therefore flagging cases of non-recovery. To address the class imbalance between non-recoverers ($NO-REC$), who are typically fewer [51, 52], and recoverers (*REC*), for the second step, we employed an anomaly detection approach. An outline of the proposed approach is reported in Fig. 2.

We trained and validated multimodal predictive models that combined sEMG, kinematic, and multimodal features, both with and without including the FMA-UL score from the initial assessment as an input feature. To provide a meaningful comparison - and given that, to the best of our knowledge, no existing predictive approaches can be directly applied to our dataset - we also developed two single-modality models: one using only sEMG features and another using only kinematic features. This enabled a comparative analysis of multimodal versus single-modality approaches for predicting post-rehabilitation outcomes.

FMA-UL regression

To predict the FMA-UL score following rehabilitation, we trained machine learning regression models using features extracted from the A_{pre} session to estimate the FMA-UL score at A_{post}. Given the limited sample size, we employed a nested leave-one-subject-out cross-validation (LOSO-CV) approach. This framework consists of two nested loops: 1) outer loop (for test set selection) – one participant at a time (and all their corresponding instances) was excluded as the test subject, while the remaining participants formed the training set; 2) inner loop (for feature selection and model optimization) – within the training set, another LOSO-CV was conducted to search the most predictive features and optimize the model hyperparameters using a grid search.

Feature selection To reduce the feature dimensionality, we used two widely adopted feature ranking algorithms: minimum redundancy maximum relevance (MRMR) and ReliefF [53, 54]. These methods ranked features based on their relevance to the outcome variable (FMA-UL at A_{post}) while minimizing redundancy between features. For each iteration of the inner LOSO-CV, features were ranked using data from the training participants. However, the effectiveness of the MRMR algorithm decreases as the set of features increases, and it also neglects the synergistic effects between variables, occasionally leading to suboptimal classification results [53]. For this reason, we chose to split the selection of the best features based on the type of data (sEMG, kinematic, multimodal), to make the most of the dataset at our disposal. At the end of the inner loop, the frequency with which each feature was selected across the iterations was obtained. The two most frequently selected sEMG features and the two most frequently selected kinematic measures were combined

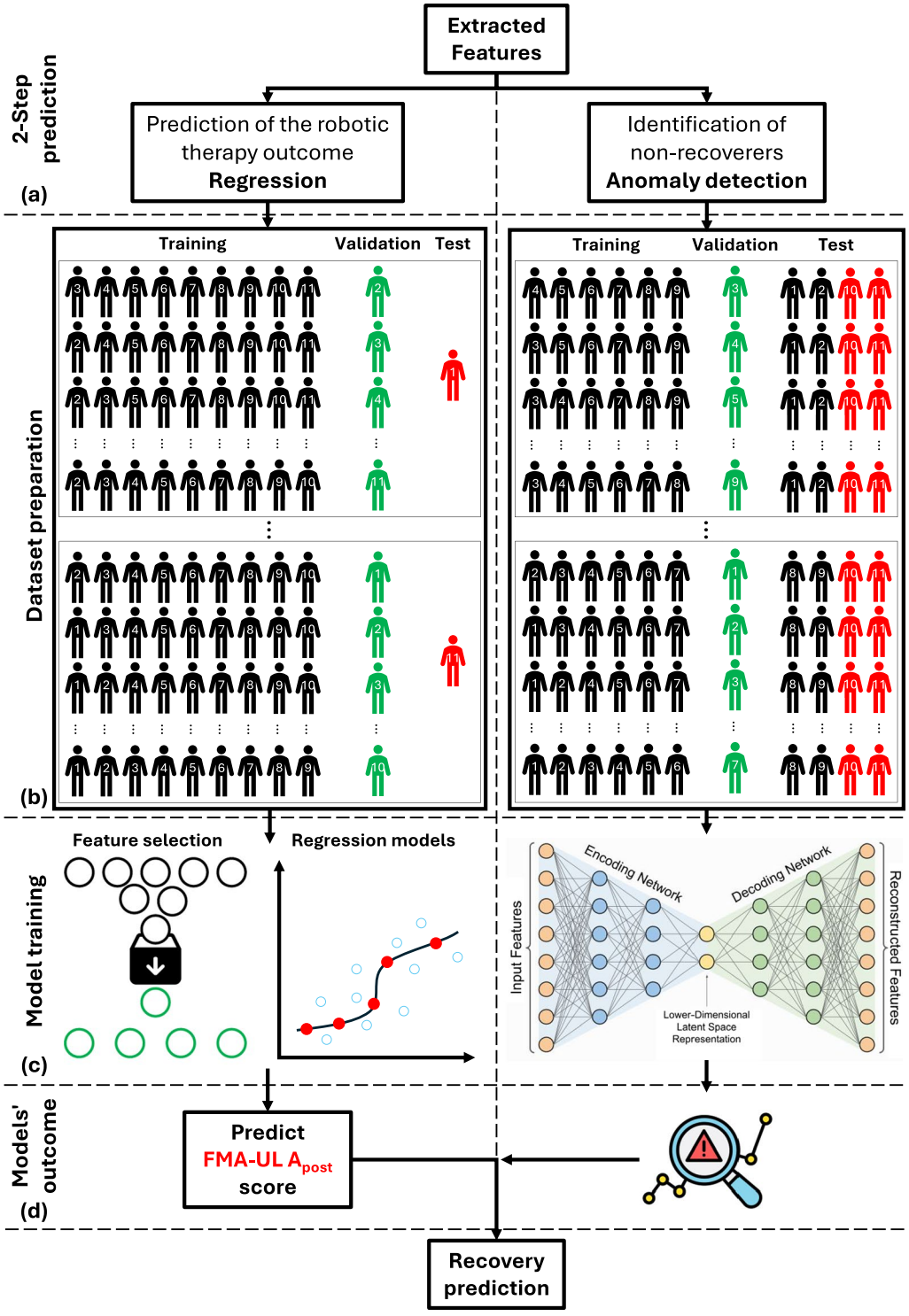

Fig. 2 Two-step machine learning approach to predict UL recovery. **a** Overall structure of the two steps: FMA-UL regression and anomaly detection. **b** Dataset preparation and cross-validation strategies adopted for both approaches: a nested LOSO-CV for the regression and a Leave-Two-Subject-Out CV for the anomaly detection. **c** Representation of the models used during training, validation and test. **d** The outcome of both approaches: the prediction of the final FMA-UL score at the end of the therapy and the identification of NO-REC participants for the anomaly detection

with the GSI and FMA-UL at A_{pre}, resulting in a total of six features per participant. To identify the optimal feature set, all possible combinations of subsets of these six features (63 combinations in total) were generated and evaluated during the inner LOSO-CV.

Model training and validation (Inner LOSO-CV) Using the selected features, we trained and validated three regression algorithms during the inner LOSO loop: support vector regression (SVR), multilayer perceptron (MLP), and random forest for regression (RF). For each algorithm, a grid search was conducted to identify the best combination of hyperparameters, specifically:

- For SVR: kernel types (linear, radial basis function – RBF); Regularization parameter – C (0.001, 0.01, 0.1, 1, 10, 100, 1000); kernel coefficient – γ (0.001, 0.01, 0.1, 1, 10, 100, 1000); Epsilon – ϵ ((iqr(Target)/1.349) [0.001, 0.01, 0.1, 1, 10, 100])
- For the MLP: number of hidden layers (1, 2, 3); number of neurons per layer (5, 10, 15, 20, 25).
- For the RF: number of trees (1, 5, 10, 20, 50, 100, 1000).

For each combination of features and hyperparameters, the model predicted the FMA-UL score at A_{post} for the validation participant left out in the inner loop. Additionally, two feature set configurations were evaluated: (1) Models including the FMA-UL A_{pre} feature, and (2) Models excluding the FMA-UL A_{pre} feature. This dual evaluation ensured that the model's predictive performance could be assessed both with and without the inclusion of the initial FMA-UL score, providing insight into the influence of prior functional status on predictions. The two models were further compared against two other predictive models trained exclusively with sEMG data and kinematic data, respectively. These comparisons aimed to provide a further evaluation of the multimodal approach against the single modalities. The root mean square error (RMSE) between the predicted and actual A_{post} FMA-UL scores was calculated for each movement in the training set to evaluate the model's performance.

Final evaluation (Outer LOSO-CV) The model configuration (algorithm, hyperparameters, and feature set) yielding the lowest RMSE during the inner LOSO-CV was selected as the optimal model for predicting the left-out participant's FMA-UL score in the outer loop. Predictions were made for all movements performed by the test participant, and the average predicted FMA-UL score was compared to the actual A_{post} score, using the RMSE. This approach was repeated for all participants, ensuring that each participant served as the test subject exactly once.

Anomaly detection

The use of anomaly detection approaches is based on two key hypotheses: (1) there is typically a class imbalance, with NO-REC individuals being fewer than REC ones [51, 52]; and (2) the data distribution of NO-REC participants differs significantly from that of REC participants.

Considering the significant class imbalance of our dataset, where NO-REC participants are much fewer than REC ones, a one-class classification strategy is more appropriate than binary classification. Such a framework consists in training the model exclusively on the majority class (REC participants), effectively learning the characteristics of this group, and then treating deviations from this learned pattern as anomalies (NO-REC participants).

To investigate these hypotheses, we developed a fully connected autoencoder to reconstruct the feature set obtained from the movements performed by participants during the A_{pre} session. Exploiting the error between the input feature set and the reconstructed one, we sought to detect NO-REC participants, who should exhibit higher reconstruction errors.

Dataset preparation To define the two classes in the anomaly detection framework, we plotted each participant on a plane (see Fig. 3), with the horizontal axis representing the remaining potential for improvement (66 - FMA-UL A_{pre}) and the vertical axis representing the Delta FMA-UL (difference between FMA-UL A_{post} and FMA-UL A_{pre}).

We based the distinction between REC and NO-REC participants on the proportional recovery rule (PRR), taking inspiration from it and therefore considering recovery relative to each participant's remaining potential rather than baseline severity alone. Although the PRR has known limitations and has been critically discussed in the literature [55, 56], it remains an empirically supported framework for describing recovery patterns, particularly in patients with moderate to severe baseline impairment [16, 57].

Clustering was performed using the k-means algorithm, which grouped participants based on the difference between their initial and post-intervention FMA scores. The algorithm iteratively assigned participants to 2 clusters and adjusted the cluster centers until convergence. To ensure stability and reduce the risk of converging to local minima, the k-means clustering was repeated 100 times with different random initial centroids, and the best clustering solution (minimizing within-cluster variance) was selected. This process identified two participants (ID03 and ID07) as NO-REC, characterized by limited recovery, and a larger group of participants (REC) exhibiting more improvements.

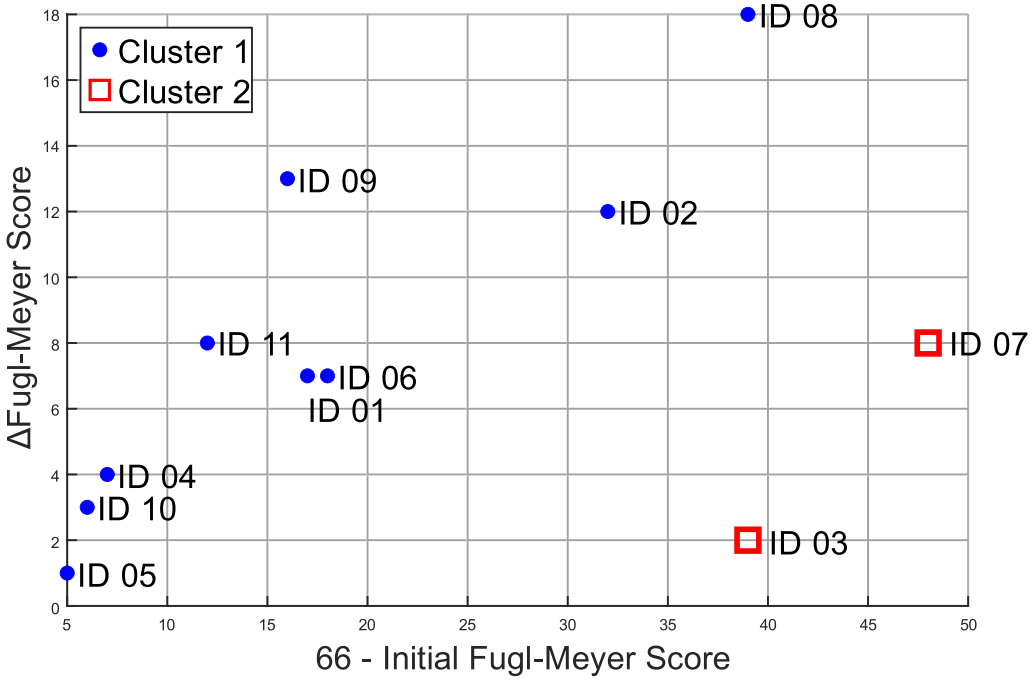

Fig. 3 Visualization of the clustering of the participants according to their FMA-UL score with the horizontal axis representing the remaining potential for improvement (66 - FMA-UL A_{pre}) and the vertical axis representing the Delta FMA-UL (difference between FMA-UL A_{post} and FMA-UL A_{pre})

Autoencoder architecture We developed a fully connected autoencoder with two main components: an encoder, which compresses the input data into a latent-space representation, and a decoder, which reconstructs the input data from this compressed representation [58]. Two versions of the model were implemented: one including the FMA-UL score assessed at A_{pre}, and one excluding it. The input data consisted of the full feature set: sEMG, kinematic, and multimodal features, with or without the FMA-UL A_{pre} score. This resulted in an input layer of dimensionality 96 for the model that included FMA-UL A_{pre}, and dimensionality 95 for the model that excluded it. During the validation phase, we optimized the two autoencoders by testing various combinations of hyperparameters, such as the number of hidden layers, the number of nodes per layer, the activation function and the optimizer. The tested hidden layer configurations were:

- Encoder: [32, 16]; Decoder: [16, 32]
- Encoder: [64, 32, 16]; Decoder: [16, 32, 64]
- Encoder: [64, 32, 16, 8]; Decoder: [8, 16, 32, 64]

with the number of units in the hidden layers specified in square brackets. We evaluated two activation functions, ReLU and tanh, and used both Adam and RMSprop optimizers, each configured with a learning rate of 0.001. The training process was carried out for 50 epochs with a batch size of 16, and the Mean Squared Error (MSE) was used as the loss function to assess the reconstruction quality. Since the initial clustering was based on FMA-UL A_{pre} and FMA-UL A_{post}, and FMA-UL A_{pre} serves as a predictive feature for the autoencoder, we built models with the same architecture and features, excluding FMA-UL A_{pre}, to assess potential bias.

Training-validation strategy We implemented a nested cross-validation approach to optimize and evaluate the performance of the fully connected autoencoder. This process comprised two levels: (1) Outer Loop (test set selection) – the test set consisted of the two NO-REC participants and two randomly selected REC participants, whereas the remaining seven REC participants were used for training and validation; (2) Inner Loop (hyperparameter tuning) – within the training and validation set, a LOSO-CV was performed on the seven REC participants to identify the optimal hyperparameters of the autoencoder. The autoencoder's performance during validation was evaluated using the MSE between the input feature array and the reconstructed array. This metric guided the selection of the best combination of hyperparameters.

Final model evaluation The best-performing hyperparameters from the inner LOSO-CV were used to train the autoencoder on all 7 REC participants. The model was then tested on the four held-out participants (2 REC and 2 NO-REC). To assess the model's robustness, this process was repeated by randomly selecting different pairs of REC participants for the test set, ensuring all possible combinations were covered. This resulted in 36 unique

test sets, allowing for a comprehensive evaluation of the model under various conditions.

To classify REC and NO-REC participants, we used the root mean squared error (RMSE) of the reconstruction for the movements of the four test participants. A threshold was set at the 50th percentile of the sorted RMSE values: movements with RMSE in the upper half of the distribution were classified as NO-REC (indicating a higher error and a greater deviation from the REC distribution), while those with RMSE in the lower half were classified as REC. A test participant was flagged as NO-REC if at least half of their movement repetitions were classified as NO-REC.

Statistical analysis

The regression RMSEs of the predicted FMA-UL A_{post} scores on the test set were evaluated under four conditions: (1) multimodal approach including FMA-UL A_{pre} as an input feature (Multimodal with FMA), (2) multimodal approach excluding FMA-UL A_{pre} (Multimodal without FMA), (3) single-modality approach using only sEMG features (sEMG-only), and (4) single-modality approach using only kinematic features (Kinematic-only). For each condition, the Shapiro–Wilk test was applied to assess the normality of RMSE distributions. As some distributions did not meet the normality assumption, we used a non-parametric Friedman test to examine overall differences across the four groups. When the Friedman test indicated significance ($p < .05$), post-hoc pairwise comparisons were performed using Nemenyi's Critical Difference (CD) test to identify which model(s) achieved the lowest prediction errors.

Results

FMA-UL regression

We first developed a cross-validated machine learning regression model able to predict the FMA-UL A_{post} based on kinematic, sEMG and clinical data. At the conclusion of the inner validation loop, the features, models, and hyperparameters that minimized the RMSE in predicting the FMA-UL A_{post} were recorded for each participant.

Best features used

The features used by the best-performing models are shown in Fig. 4a, which compares the features selected by multimodal models that included FMA-UL A_{pre} with those that excluded it. In multimodal models that included FMA-UL A_{pre}, the most frequently used features were C_{med} BRAD (5/11), *GSI* (8/11), J_{mean} (9/11), and FMA-UL A_{pre} itself (10/11). In contrast, formultimodal models that excluded FMA-UL A_{pre}, the most frequently used features were J_{mean} (5/11), C_{med} BRAD (7/11), and *GSI* (8/11).

Figure 4b compares the features selected by the sEMG-only and kinematic-only models. Among the sEMG-only models, the most frequently selected feature was C_{med} BRAD (5/11). For the kinematic-only models, the most commonly selected feature was J_{mean} (8/11).

Best models used

For multimodal models that included FMA-UL A_{pre} in the feature set, participants ID 04, ID 05, ID 09, ID 10, and ID 11 achieved the best results using MLP models. In contrast, participants ID 01, ID 02, ID 03, ID 06, ID 07, and ID 08 performed best with SVR models. For multimodal models trained without including FMA-UL A_{pre} in the feature set, participants ID 03 and ID 04 achieved the best results with SVR models. All other participants achieved their best performance with MLP models. The list of hyperparameters used for each model is shown in Table 2.

For the sEMG-only models, participants ID 01, ID 03, and ID 06 achieved the best results using SVR models, while participants ID 02, ID 04, ID 05, ID 07, ID 08, ID 09, ID 10, and ID 11 performed best with MLP models. Regarding the kinematic-only models, participants ID 02, ID 06, ID 07, ID 08, ID 09, and ID 11 obtained their best results using SVR models, whereas participants ID 01, ID 03, ID 04, ID 05, and ID 10 performed best with MLP models. The specific hyperparameters used for each model are detailed in Table 3.

Statistical results

For each test participant in the outer loop of the LOSO-CV, the optimal hyperparameters identified during the inner loop were used to train the best regression model. This trained model was then applied to predict the FMA-UL A_{post} scores based on the movements of the left-out test participant. The predicted scores were subsequently compared to the actual FMA-UL A_{post} scores.

Figure 5a presents and compares the predicted scores for each model with the actual FMA-UL scores for all participants. The median RMSE was 4.0 (interquartile range-$IQR = 3.5$) for multimodal model that included FMA-UL at A_{pre}, 7.0 ($IQR = 12.5$) for multimodal model trained excluding FMA-UL at A_{pre}, 4.0 ($IQR = 4.0$) for sEMG-only model, 8.0 ($IQR = 9.5$) for kinematic-only model.

Due to the violation of normality distribution of the RMSEs in at least one group, a non-parametric Friedman test was used to assess overall differences in RMSE distributions across the four models. The test did not reveal statistically significant differences ($\chi^2 = 1.69$, $p = 0.64$), as a result, no post-hoc pairwise comparisons were conducted.

As no significant differences were found among the four methods, we applied a tiebreaker rule by computing

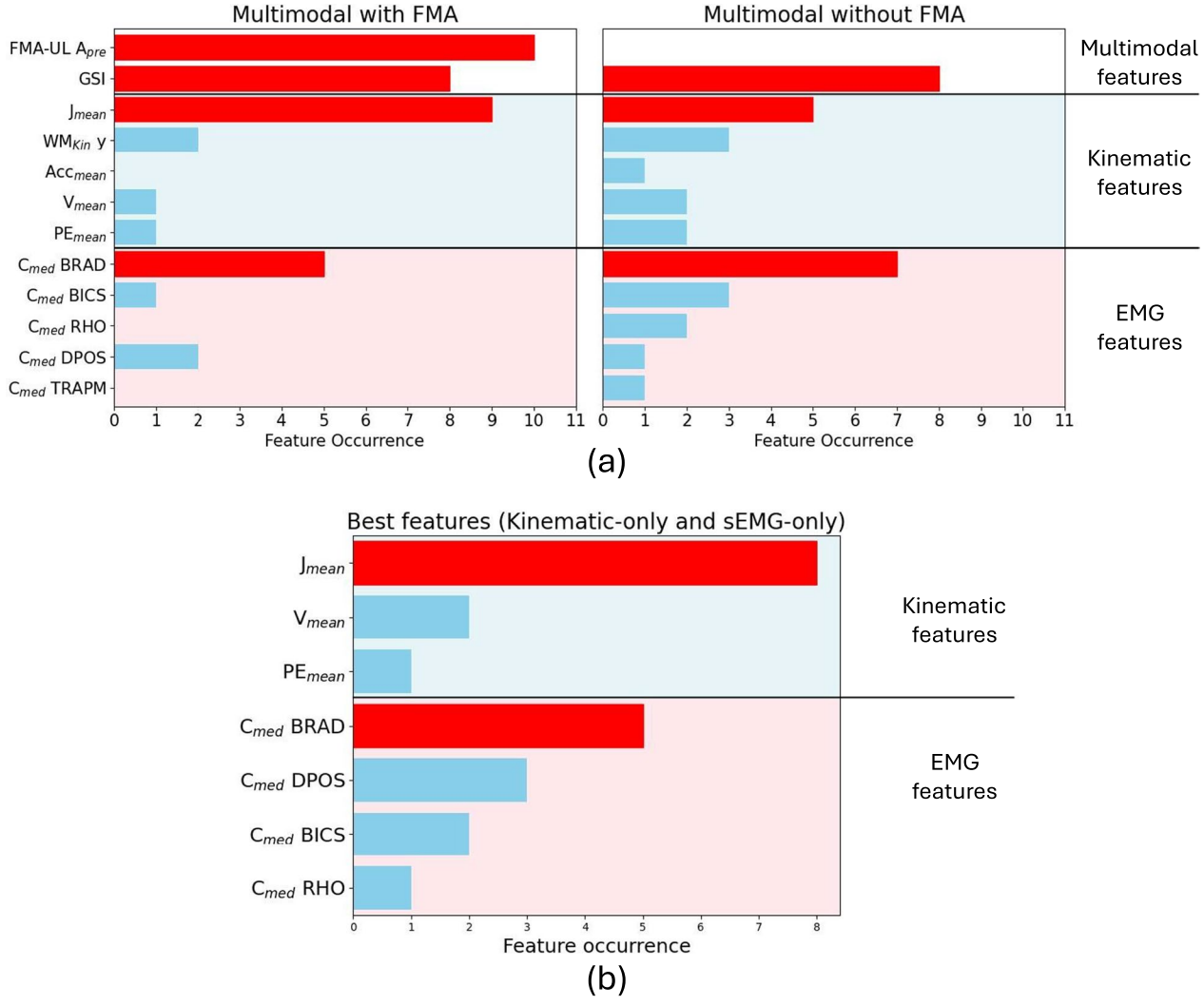

Fig. 4 **a** List of features used in the best performing multimodal models, both when FMA-UL A_{pre} was included and when it was excluded, grouped according to the type (sEMG, kinematic or multimodal). **b** List of features used in the best performing models when using exclusively sEMG data and kinematic data, grouped according to the type. The bar graphs show the occurrence of each feature for each model, the most used ones are highlighted in red

the percentage of participants for whom the RMSE was less than or equal to the minimal clinically important difference for the Fugl-Meyer upper limb assessment (MCID = 12.4) [59]. We found that the percentage of participants with an error below this value was 90.91% for multimodal model including FMA-UL A_{pre}, 63.64% for multimodal model excluding FMA-UL A_{pre}, 81.82% for sEMG-only model, and 72.73% for kinematic-only model.

Anomaly detection

We cross-validated a fully connected autoencoder, with the intention of flagging non-recoverer stroke survivors, therefore determining which predictions from the regression model could be considered as unreliable. At the conclusion of the validation loop, 36 autoencoder configurations were obtained, each one corresponding to an independent test set comprising 2 NO-REC and 2 REC participants.

The hyperparameters optimized during the validation phase were used to retrain each model, enabling anomaly detection on the test sets. Each test participant performed 12 distinct movements, resulting in a total of 48 reconstructed movements per test set.

The reconstruction RMSEs for the 48 movements were ranked in descending order. A threshold was set at the median RMSE value for each test set: movements with RMSEs above the median were labeled as NO-REC (indicating higher reconstruction error, therefore an anomaly), while those below the median were labeled as REC (indicating lower reconstruction error, therefore closer to the training sample). A test participant was classified as NO-REC if at least half of their movements were labeled as NO-REC based on this threshold.

Table 2 List of models and hyperparameters used by each participant for both cases: with and without FMA-UL APRE as a training feature

Multimodal with FMA			Multimodal without FMA	
ID	**Model**	**Model's Hyperparameters**	**Model**	**Model's Hyperparameters**
ID1	SVR	`rbf`, $C = 100$, $\gamma = 1, \epsilon = 0.013343$	MLP	[15, 10]
ID2	SVR	`rbf`, $C = 100$, $\gamma = 1, \epsilon = 0.13343$	MLP	[5, 25, 20]
ID3	SVR	`linear`, C=1000, $\gamma = 1, \epsilon = 1.2602$	SVR	`linear`, C=1000, γ=0.1, $\epsilon = 0.12602$
ID4	MLP	[20, 15, 25]	SVR	`linear`, $C = 100, \gamma = 0.1$, $\epsilon = 0.012602$
ID5	MLP	[25, 10, 25]	MLP	[10, 25, 20]
ID6	SVR	`linear`, $C = 100$, $\gamma = 1, \epsilon = 0.013343$	MLP	[5, 5, 15]
ID7	SVR	`linear`, $C = 100$, $\gamma = 1, \epsilon = 0.12602$	MLP	[20, 10]
ID8	SVR	`linear`, $C = 100$, $\gamma = 1, \epsilon = 0.12602$	MLP	[25, 10, 25]
ID9	MLP	[5, 25, 25]	MLP	[5, 25, 25]
ID10	MLP	[15, 5, 20]	MLP	[25, 5, 25]
ID11	MLP	[20, 5, 20]	MLP	[5, 10, 15]

Table 3 List of models and corresponding hyperparameters used for each participant in the best-performing sEMG-only and kinematic-only models

sEMG-only			Kinematic-only	
ID	**Model**	**Model's Hyperparameters**	**Model**	**Model's Hyperparameters**
ID1	SVR	`rbf`, $C = 100, \gamma = 1$, $\epsilon = 0.013343$	MLP	[5, 10, 10]
ID2	MLP	[5, 25]	SVR	`rbf`, $C = 100$, $\gamma = 1, \epsilon = 0.13343$
ID3	SVR	`linear`, $C = 1000, \gamma = 1$, $\epsilon = 1.2602$	MLP	[25, 25, 25]
ID4	MLP	[20, 15, 25]	MLP	[25, 20, 10]
ID5	MLP	[25, 10, 25]	SVR	`rbf`, $C = 100$, $\gamma = 1, \epsilon = 1.3343$
ID6	SVR	`linear`, $C = 100, \gamma = 1$, $\epsilon = 0.013343$	SVR	`linear`, $C = 100$, $\gamma = 1, \epsilon = 0.13343$
ID7	MLP	[5, 20, 15]	SVR	`linear`, $C = 100$, $\gamma = 1, \epsilon = 0.12602$
ID8	MLP	[25, 25, 5]	SVR	`linear`, $C = 100$, $\gamma = 1, \epsilon = 0.12602$
ID9	MLP	[5, 25, 25]	SVR	`linear`, $C = 100$, $\gamma = 1, \epsilon = 1.2602$
ID10	MLP	[5, 10, 15]	MLP	[15, 5, 20]
ID11	MLP	[20, 5, 20]	SVR	`rbf`, $C = 100$, $\gamma = 1, \epsilon = 0.13343$

To evaluate the performance of this classification approach, we calculated the number of true positives, true negatives, false positives, and false negatives for each test set. These results were aggregated into a cumulative confusion matrix across all 36 test sets, as shown in Fig. 5b, more specifically: on the left is shown the confusion matrix of the model trained using FMA-UL A_{pre}, on the right the model trained without using FMA-UL A_{pre}. Using this approach, we achieved a global F1-score of 0.985 for the model that included FMA-UL A_{pre} while the model rained excluding this feature obtained a global F1-score of 0.94. In both cases, the method successfully classified nearly all test participants, with REC participants misclassified only 2 out of 72 instances as NO-REC in the first case, and 8 out of 72 misclassified REC instances in the second. Additionally, the average RMSE for each test participant (calculated across their 12 movements) was plotted for all 36 test sets (Fig. 5c), with the model trained with FMA-UL A_{pre} among their features on the left and those that didn't include it on the right. The plot of our multimodal models demonstrates a clear separation between the RMSEs of REC and NO-REC participants for both families of models. This result highlights the feasibility of using anomaly detection to identify stroke survivors who are less likely to benefit from rehabilitative interventions.

Discussion

This study highlights the potential of combining regression models with anomaly detection techniques to predict UL motor recover post-stroke using robot-assisted assessment. A key strength of this study is its novel integration of a two-step prediction, combining machine-learning models for regression with fully connected autoencoders for anomaly detection. This combination of methodologies allowed us to estimate the post-rehabilitation outcome, identify patients who are not going to experience improvements after therapy, therefore flagging cases for which predictions can be considered unreliable. This a priori information may enable the development of more personalized rehabilitation plans, tailored to the specific needs of individual patients. There are few studies in the literature that focus on predicting upper limb functional recovery following robotic therapy. Many rely solely on clinical or kinematic data [26], while others have opted for EEG data [29]. Studies incorporating sEMG are particularly scarce and are typically used for evaluation rather than prediction [30, 31]. Our work aimed to integrate data from multiple sources in a multimodal approach, providing a more comprehensive analysis of participants and enabling more accurate predictions.

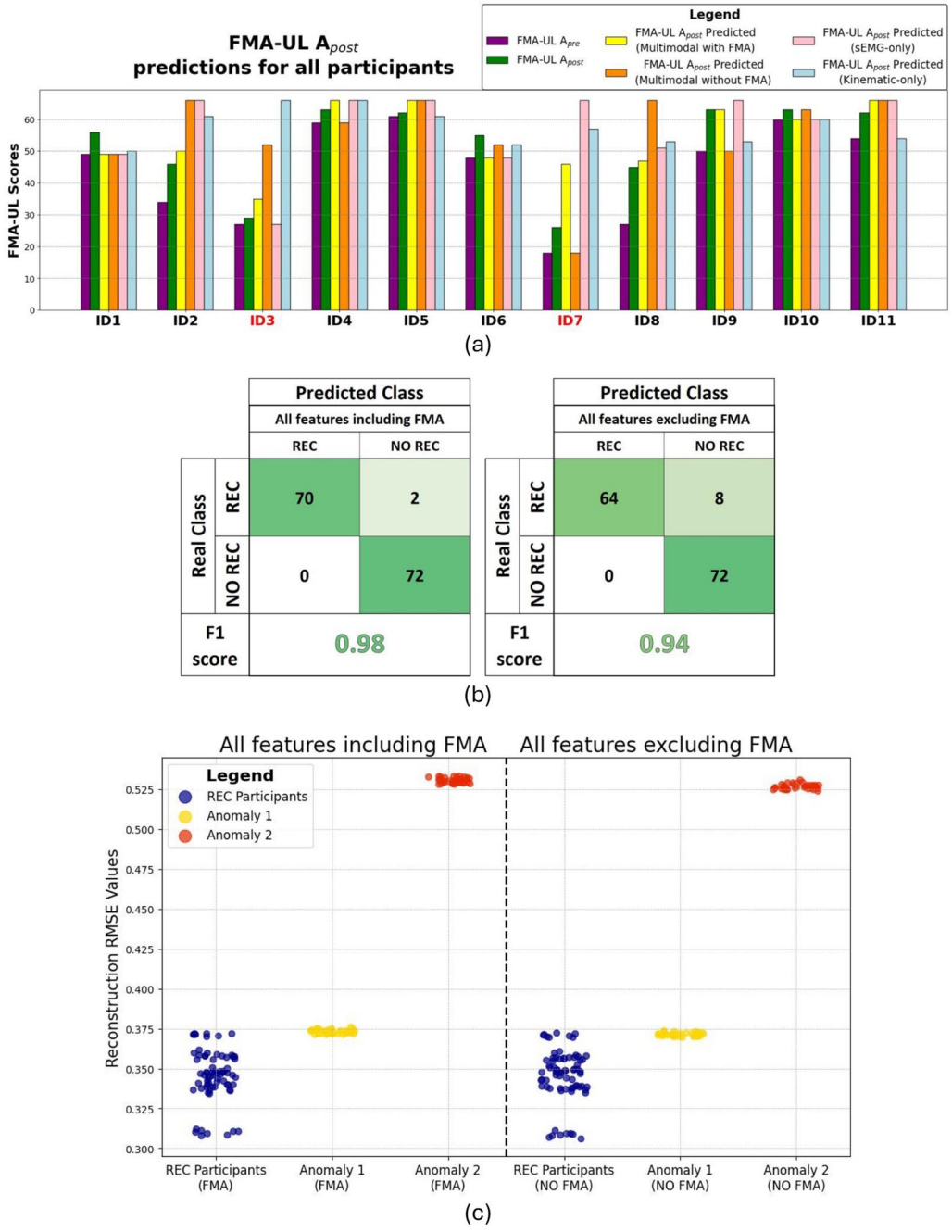

Fig. 5 **a** Comparison of FMA-UL scores across all participants. For each participant, six bars are shown: the FMA-UL scores at Apre and Apost, the scores predicted by the multimodal models using the best combination of features with and without the Apre score, and the scores predicted by the sEMG-only and kinematic-only models. Participants with anomalous recovery are highlighted in red. **b** Confusion matrices, for anomaly detection classification, of the models trained with and without the inclusion of the FMA-UL Apre score. **c** Distribution of reconstruction RMSEs from the best-performing autoencoders, tested with and without the FMA-UL Apre score

Models' performance

The multimodal model for FMA-UL A_{post} regression that included FMA-UL A_{pre} as a feature achieved a median RMSE of 4.00 (IQR = 3.5), whereas the multimodal predictive model that excluded this feature had a higher median RMSE of 7.00 (IQR = 12.5). Examining the performance for individual patients, in the first case, the prediction error remained well below the MCID for all participants except ID 07. The RMSE values were [7, 4, 6, 3, 4, 7, 20, 2, 0, 3, 4], with ID 07 showing the highest error at 20. This shows how the model performed highly effectively and made accurate predictions of therapy outcomes.

In contrast, when FMA-UL A_{pre} was not included, the prediction error was generally higher. The RMSE values in this case were [7, 20, 23, 4, 4, 3, 8, 21, 13, 0, 4], with ID 02, ID 03, ID 08, and ID 09 exceeding the clinically important difference threshold. The largest error was observed for ID 03, with an RMSE of 23. While the model still demonstrated predictive efficacy, its accuracy was lower, demonstrating how the combined use of kinematic and sEMG data, used in a data-driven approach, adds predictive value to the FMA for recovery after robotic rehabilitation.

The sEMG-only model achieved a median RMSE of 4.00 (IQR = 4.00), which is comparable to the multimodal model including FMA-UL A_{pre}. The RMSE values were [7, 20, 2, 3, 4, 7, 40, 6, 3, 3, 4], with IDs 02 and 07 exceeding the MCID threshold. While errors were generally comparable to those of the multimodal model, IDs 02 and 07 showed significantly higher errors than when FMA-UL A_{pre} was included. This suggests that although sEMG captures relevant motor activity, it lacks sufficient predictive power when used on its own.

The kinematic-only model showed the weakest performance among all trained models, with median RMSE values of 8.00 (IQR = 9.50). RMSE values were [6, 15, 37, 3, 1, 3, 31, 8, 10, 3, 8], with IDs 02, 03, and 07 consistently above the MCID threshold. These results indicate that kinematic model struggle to achieve reliable predictive performance on their own.

Although some models exhibited similar median RMSE, substantial differences were observed in mean performance and participant-level consistency; however, these differences did not reach statistical significance. At the individual level, the FMA-inclusive multimodal model and the sEMG-only model both achieved the lowest median RMSE (4.00), but their mean RMSE values differed considerably (5.45 vs. 9.00, respectively). To further differentiate model performance, we calculated the percentage of participants with RMSE $\leq$ MCID: 90.91% for multimodal model including FMA-UL A_{pre}, 63.64% for multimodal model excluding FMA-UL A_{pre}, 81.82% for sEMG-only model, and 72.73% for kinematic-only model. These results suggest that a multimodal approach, integrating sEMG and kinematic data, provides the most accurate and reliable predictions and that these two modalities complement and enrich the Fugl-Meyer score in forecasting UL recovery post-rehabilitation.

The anomaly detection models, a novel feature of this approach, demonstrated outstanding performance, achieving an overall F1 score of 0.985. This method effectively distinguished between patients who were more likely to benefit from therapy (REC) and those with a lower response (NO-REC). Setting the threshold at the 50th percentile, determined empirically without the need for fine-tuning, allowed for a clear distinction between the two groups. This approach allowed the identification of participants with limited or uncertain recovery predictions, such as IDs 03 and 07, who had the highest prediction errors in the regression phase and were flagged as potential outliers. These findings support our hypothesis that a two-step predictive approach enhances the overall assessment of stroke patients, helping clinicians identify individuals who may not respond as expected to the selected therapy.

This result not only confirms the feasibility of integrating autoencoders with sEMG and kinematic data but also establishes reconstruction errors as a robust indicator of patient response to therapy. By offering a more comprehensive assessment of therapy outcomes, this approach strengthens confidence in the predicted recovery of REC patients while highlighting the need for caution when interpreting predictions for NO-REC patients.

These findings highlight the potential of the proposed framework to support clinical decision-making in rehabilitation. One of the persistent challenges in stroke care is selecting the most suitable therapeutic approach for each patient. A data-driven tool capable of predicting recovery outcomes before the intervention could assist clinicians in tailoring therapies more effectively. By analyzing a patient's initial sEMG and kinematic data, the system could provide early insights into which available therapies are more or less likely to yield meaningful improvements, potentially including estimated gains on clinical scales. For instance, participant ID03, classified as a non-recoverer (NO-REC) following a non-robotic therapy, might have achieved better recovery with an alternative intervention. This example suggests that recovery potential is not absolute but may vary depending on the therapeutic pathway chosen. To realize this vision, dedicated predictive models would need to be trained for each type of therapy, enabling a personalized rehabilitation strategy. While the development of such therapy-specific models is complex, it represents a valuable direction for future research aimed at optimizing outcomes and resource allocation in stroke rehabilitation.

Key predictors

These results highlight the critical role that FMA-UL A_{pre} still plays in improving predictive accuracy for motor recovery but also reveal new key variables such as median brachioradialis contraction (C_{med} BRAD), mean jerk (J_{mean}), and GSI, all of which consistently emerged as critical predictors in the best-performing models. These features, partly electromyographic and partly kinematic, highlight not only their critical relevance in predicting upper limb motor recovery in stroke patients, but also the need for a multimodal approach. By analyzing upper limb movement from multiple perspectives, this

approach allows for more comprehensive assessments and more precise predictions.

Interestingly, the sEMG-only model, which predominantly relied on brachioradialis activity (frequently selected as a key feature), achieved a median RMSE of 4.00, comparable to the performance of the multimodal model that included FMA-UL A_{pre}. This underscores the strong predictive power of sEMG features and highlights the brachioradialis muscle as a critical predictor of stroke recovery. The brachioradialis is a well-known and easily analyzed muscle that plays a critical role in forearm function and is involved in various upper limb activities, including elbow movements, supination and pronation, and elbow stabilization during rapid flexion and extension [60–62]. The results confirm its value as a key predictor variable, aligning with previous research and suggesting that recovery outcomes can be effectively predicted using fewer sensors than initially anticipated.

Another significant insight from this study is the critical role of kinematic smoothness as a predictor of stroke recovery. The J_{mean} feature was among the most frequently selected kinematic variables, confirming its clinical relevance. However, kinematic-only model, showed the worst regression performance, achieving a median RMSE of 8.00 (IQR = 9.5) and several participants exceeding the MCID threshold. This highlights the limitations of using either kinematic or clinical features alone and underscores the need for multimodal integration to improve predictive accuracy.

Equally noteworthy is the importance of GSI, introduced here as a novel metric for predicting recovery. GSI captures key aspects of muscle contraction and movement fluidity, providing a more multidimensional and comprehensive assessment of a patient's progress. Originally conceived as a simple measure of movement quality during home exercises, the results of this study demonstrate that the GSI has exceptional potential to predict the effectiveness of therapeutic interventions, highlighting the value of a multimodal approach to stroke patient assessments. Its utility could be further expanded by identifying the optimal combination of muscles, including the brachioradialis, to predict upper limb recovery with even greater accuracy. For example, our findings indicate that, among the fifteen muscles originally analyzed, only four, namely the brachioradialis (BRA), dorsal posterior (DPOS), biceps (BICS), and rhomboid (RHO), were selected across all sEMG-only models, with the BRA being the most frequently used. This would lead to a reduction in the number of sensors used, which would improve the practicality and cost-effectiveness of rehabilitation monitoring in clinical settings.

These results indicate that, although sEMG and kinematic data each provide valuable insights into movement quality, neither modality alone offers sufficient predictive power. In contrast, integrating both feature types within multimodal models achieved the most accurate and reliable predictions.

Limitations and future work

We recognize the limitations in our work. The small dataset is a major constraint, potentially impacting on the generalizability of the findings. While the use of nested leave-one-subject-out cross-validation helped estimate robust predictions, a larger and more diverse dataset would allow for more generalized validation and improve model reliability across diverse patient populations. Unexpectedly, no significant statistical difference was observed between regression multimodal models with and without the inclusion of FMA-UL A_{pre}, despite the superior performance of the model incorporating this feature. This suggests that with a larger dataset, multimodal models excluding FMA-UL A_{pre} could achieve similar levels of predictive accuracy. This finding highlights the potential for alternative predictors that do not rely on clinical scales but complement them to provide a more nuanced understanding of the optimal rehabilitation pathway for each patient.

Regarding anomaly detection, the model can distinguish and classify the two groups of participants with excellent accuracy, but it is also fair to point out that many of the values obtained during the testing phase are very close to the threshold found, both by the RECs and the NO RECs. This could also be due to an overfitting problem with the model, although with a different effect than seen in regression. For anomaly detection to work effectively, the model needs to be overfitted to a single category, allowing it to better recognize deviations as anomalies. Although the percentage of anomalies $(20-30\%)$ is in line with the results of other studies [51, 52], the use of only seven REC participants as a training base, despite the precautions taken by using all possible combinations for training and testing, may still not be enough for obtaining an optimal performance. We tried to mitigate this issue by applying a majority vote on the classification of individual movments: if at least six of the twelve movements were classified as NO REC, the subject was labeled as NO REC; otherwise, they were classified as REC. This approach refined the classification process, reducing overlap between the result distributions of the two groups and more clearly separating them. However, a larger dataset with more REC participants would allow a better fitting of the model to this group and could further clarify the observed distinctions, even with the current limited data. It should also be noted that the initial grouping of the two models, REC and NO REC, was based on clustering using FMA-UL A_{pre} and FMA-UL A_{post} values. While this approach could inadvertently lead to bias, we attempted to assess this issue by

training two families of models with the same architecture but trained using different features, specifically, with and without FMA-UL A_{pre}. The results showed that both model families performed similarly, though those trained without FMA-UL A_{pre} had slightly worse performance and similar RMSE distributions. While this disparity suggests a difference in model performance, it does not directly confirm bias. A larger dataset could offer more conclusive insights into potential bias.

Another limitation of our study is the lack of EEG data for analysis, as found in other studies [63], which could have provided more in-depth information on the recovery of participants during the 4 weeks of therapy. Future work should incorporate these data to assess how their inclusion might impact the performance of the multimodal models.

Another important consideration is that difficulty in reaching targets may stem not only from motor impairments but also from unfamiliarity with the ALEx robotic system. Although all participants received standardized instructions, a learning effect cannot be ruled out. The current dataset does not allow us to assess this impact. Future studies should include dedicated familiarization sessions, as suggested by [64], to better isolate motor deficits from adaptation effects.

Future studies could also explore the development of multimodal models that rely on a reduced set of sensors, incorporating sEMG data from the four most important muscles we found (brachioradialis (BRA), dorsal posterior (DPOS), biceps (BICS), and rhomboid (RHO)), jerk as the sole kinematic feature, the GSI calculated from this specific set of muscles and jerk, and FMA-UL A_{pre}. Comparing the performance of such sensor-efficient models with the current results could help assess the feasibility of using fewer sensors, potentially reducing patient stress and minimizing setup time for both patients and physiotherapists.

Conclusions

In conclusion, this study demonstrated that integrating machine learning approaches for regression and anomaly detection with sEMG and kinematic data significantly help the prediction of UL recovery post rehabilitation using robotic-assisted assessment in stroke survivors. The combined use of data-driven predictive models with anomaly detection can be a valuable aid to clinicians, enabling them to personalize the individual recovery journey and maximize the therapeutic outcomes by targeting the most appropriate interventions for each patient.

Despite its promising results, the study highlighted the need to evaluate the predictive model with larger and more diverse patient populations. The anomaly detection method has the potential to be applied to many different types of therapies, including non-robotic rehabilitation approaches, offering broader utility in clinical practice. Furthermore, the role of GSI and its relationship with other established metrics require further exploration to fully understand its potential as a primary indicator of recovery. Future research should focus on addressing these gaps by expanding the dataset, incorporating patients from different therapeutic settings, and validating the model in broader scenarios. These efforts will help assess its generalizability and robustness, paving the way for a versatile tool that supports personalized rehabilitation strategies in various settings.

List of abbreviations

Technical terms for clinical assessment and data collection:

sEMG	Surface electromyography
Kin	Kinematic
UL	Upper limb
FMA	Fugl-Meyer Assessment
FMA-UL	Fugl-Meyer Assessment for UL
MCID	Minimal clinically important difference for the FMA-UL
PRR	Proportional recovery rule
ALEx RS	Arm light exoskeleton rehab station
A_{pre}	Assessment sessions before the therapy
A_{post}	Assessment sessions after the therapy
ST	Standard therapy
ST+RT	Standard therapy + robotic therapy
$NO-REC$	Non-recoverers (less likely to recover) participants
REC	Recoverers (more likely to recover) participants

Technical terms for muscles and sEMG data:

TRAPS	Upper trapezius
TRAPM	Trapezius medialis
DANT	Anterior deltoid
DMED	Medial deltoid
DPOS	Posterior deltoid
PECM	Pectoralis major
LAT	Latissimus dorsi
INFRA	Infraspinatus
RHO	Rhomboid major
BICL	Biceps brachii long head
BICS	Biceps brachii short head
BRAD	Brachioradialis
TRILA	Triceps brachii lateral
TRILO	Triceps brachii long head
PRO	Pronator
SENIAM	Surface Electromyography for Non-Invasive Assessment of Muscles
C_{med}	Median contractions
C_{RMS}	Root mean square values contraction
C_{max}	Maximum contractions
NNMF	Non-Negative Matrix Factorization
H_{EMG}	Matrix containing the time course of the synergies
W_{EMG}	Matrix representing the weight of muscle activations in each synergy (activation coefficient)
VAF	Variance Accounted For
s	Number of synergies for each movement
M	Number of muscle channels
WM_m	Weighted Mean values of muscle *m* along all its synergies *S*
$A_{s,m}$	Muscle activation coefficient of muscle *m* and synergy *s*
$p_{s,m}$	Ranking of muscle *m* inside the synergy *s*

Technical terms for kinematic data:

V_{mean}	Tangential velocity mean value
V_{max}	Tangential velocity maximum value
V_{RMS}	Tangential velocity root mean square
Acc_{mean}	Tangential acceleration mean value

Acc_{max}	Tangential acceleration maximum value
Acc_{RMS}	Tangential acceleration root mean square
J_{min}	Tangential jerk minimum value
J_{mean}	Tangential jerk mean value
J_{max}	Tangential jerk maximum value
J_{RMS}	Tangential jerk root mean square
PE_{min}	Position end-effector maximum value
PE_{mean}	Position end-effector maximum value
PE_{max}	Position end-effector maximum value
PE_{RMS}	Position end-effector maximum value
PCA	Principal Component Analysis

Technical terms for the GSI:

GSI	Gesture Similarities Index
Ref	Reference metrics from healthy participants
Mov	Movement metrics from stroke participants
GSI_{EMG}	SEMG component of the GSI
GSI_{Kin}	Kinematic component of the GSI
GSI_{TOT}	Cumulative GSI

Technical terms for machine learning and statistics:

LOSO-CV	Leave-One-Subject-Out cross-validation
MRMR	Minimum Redundancy Maximum Relevance
SVR	Support Vector Regression
MLP	MultiLayer Perceptron
RF	Random Forest for regression
RMSE	Root Mean Square Error
MSE	Mean Squared Error
IQR	Inter-Quartile Range
W	Shapiro-Wilk Test Statistic
χ^2	Friedman Test Statistic
p	P-value

Acknowledgements
Special thanks to Dr. Fulvio Vercillo for carefully following the participants during training and evaluation. Work partly supported by: #NEXTGENERATIONEU (NGEU) and funded by the Ministry of University and Research (MUR), National Recovery and Resilience Plan (NRRP), projects MNESYS (PE0000006) – A Multiscale integrated approach to the study of the nervous system in health and disease (DN. 1553 11.10.2022) and THE (IECS00000017) - Tuscany Health Ecosystem (DN. 1553 11.10.2022), project PERSONA "PERSonalized rObotic NeurorehAbilitation for stroke survivors" Regione Toscana – Bando FAS Salute 2018, Bertarelli Foundation.

Author contributions
Luigi Privitera: Conceptualization (equal); Formal Analysis (lead); Software (lead); Investigation (equal); Methodology (equal); Validation (lead); Visualization (lead); Writing - Original Draft Preparation (lead); Writing - Review & Editing (equal). Michael Lassi: Conceptualization (equal); Methodology (equal); Data Curation (lead); Writing - Review & Editing (equal). Stefania Dalise: Resources (equal); Writing - Review & Editing (equal). Valentina Azzollini: Resources (equal); Writing - Review & Editing (equal). Luca Maggiani: Resources (equal); Writing - Review & Editing (equal). Adrian Guggisberg: Resources (equal); Funding Acquisition (equal); Writing - Review & Editing (equal). Alberto Mazzoni: Writing - Review & Editing (equal). Carmelo Chisari: Supervision (equal); Funding Acquisition (equal); Writing - Review & Editing (equal). Silvestro Micera: Supervision (equal); Funding Acquisition (equal); Writing - Review & Editing (equal). Andrea Bandini: Supervision (lead); Project Administration (equal); Conceptualization (lead); Methodology (equal); Writing - Review & Editing (equal).

Funding
#NEXTGENERATIONEU (NGEU) and funded by the Ministry of University and Research (MUR), National Recovery and Resilience Plan (NRRP), projects MNESYS (PE0000006) – A Multiscale integrated approach to the study of the nervous system in health and disease (DN. 1553 11.10.2022) and THE (IECS00000017) - Tuscany Health Ecosystem (DN. 1553 11.10.2022). Project PERSONA "PERSonalized rObotic NeurorehAbilitation for stroke survivors" Regione Toscana – Bando FAS Salute 2018. Bertarelli Foundation.

Data availability
The datasets used and/or analysed during the current study are available from the corresponding author on reasonable request.

Declarations

Ethics approval and consent to participate
The research was carried out at the Neurorehabilitation Unit of the University Hospital of Pisa (Cisanello hospital), Italy, and the University Hospital of Geneva (HUG), Switzerland. Ethical approval was obtained from the Commission Canton-ale d'Ethique de la Recherche (CCER) de Genève in Switzerland and the Comitato Etico Area Vasta Nord Ovest (CEAVNO) in Pisa. All participants signed informed consent according to the requirements of the Declaration of Helsinki.

Consent for publication
Not applicable.

Competing interests
The authors declare no competing interests.

Author details
[1]The BioRobotics Institute and the Department of Excellence in Robotics and AI, Scuola Superiore Sant'Anna, Pisa, Italy
[2]School of Advanced Studies, Università di Camerino, Camerino, Italy
[3]Neurorehabilitation Unit, Department of Neuroscience, University Hospital of Pisa, Pisa, Italy
[4]Department of Translational Research and New Technologies in Medicine and Surgery, University of Pisa, Pisa, Italy
[5]Division of Neurorehabilitation, Department of Clinical Neurosciences, University Hospital Geneva, Geneva, Switzerland
[6]Laboratory of Cognitive Neurorehabilitation, Department of Clinical Neurosciences, Medical School, University of Geneva, Geneva, Switzerland
[7]Modular Implantable Neuroprostheses (MINE) Laboratory, Università Vita-Salute San Raffaele, Milan, Italy
[8]Translational Neural Engineering Laboratory, Neuro-X Institute, École Polytechnique Federale de Lausanne (EPFL), Lausanne, Switzerland
[9]Interdisciplinary Research Center "Health Science", Scuola Superiore Sant'Anna, Pisa, Italy

Received: 9 May 2025 / Accepted: 29 October 2025
Published online: 05 January 2026

References

1. Lee KB, Lim SH, Kim KH, Kim KJ, Kim YR, Chang WN, et al. Six-month functional recovery of stroke patients: a multi-time-point study. Int J Rehabil Res. 2015;38(2):173–80.
2. Organization WH. World health statistics 2008. World Health Organization; 2008.
3. Bertani R, Melegari C, De Cola MC, Bramanti A, Bramanti P, Calabrò RS. Effects of robot-assisted upper limb rehabilitation in stroke patients: a systematic review with meta-analysis. Neurol Sci. 2017;38:1561–9.
4. Hatem SM, Saussez G, Della Faille M, Prist V, Zhang X, Dispa D, et al. Rehabilitation of motor function after stroke: a multiple systematic review focused on techniques to stimulate upper extremity recovery. Front Hum Neurosci. 2016;10:442.
5. Kwakkel G, van Peppen R, Wagenaar RC, Wood Dauphinee S, Richards C, Ashburn A, et al. 2004. Effects of augmented exercise therapy time after stroke: a meta-analysis. stroke. 35(11):2529
6. Laut J, Porfiri M, Raghavan P. The present and future of robotic technology in rehabilitation. Curr Phys Med Rehabil Rep. 2016;4:312–9.
7. Volpe BT, Ferraro M, Krebs HI, Hogan N. Robotics in the rehabilitation treatment of patients with stroke. Curr Atheroscler Rep. 2002;4:270–6.
8. Everard G, Declerck L, Detrembleur C, Leonard S, Bower G, Dehem S, et al. New technologies promoting active upper limb rehabilitation after stroke: an overview and network meta-analysis. Eur J Phys Rehabil Med. 2022;58(4):530.
9. Boyd LA, Hayward KS, Ward NS, Stinear CM, Rosso C, Fisher RJ, et al. Biomarkers of stroke recovery: consensus-based core recommendations from the stroke recovery and rehabilitation roundtable. Int J Stroke. 2017;12(5):480–93.

10. Bosecker C, Dipietro L, Volpe B, Igo KH. Kinematic robot-based evaluation scales and clinical counterparts to measure upper limb motor performance in patients with chronic stroke. Neurorehabil Neural Repair. 2010;24(1):62–9.
11. Fugl-Meyer AR, Jääskö L, Leyman I, Olsson S, Steglind S. A method for evaluation of physical performance. Scand J Rehabil Med. 1975;7(1):13–31.
12. Harrison JK, McArthur KS, Quinn TJ. Assessment scales in stroke: clinimetric and clinical considerations. Clinical interventions in aging. 2013;p. 201–211.
13. Krebs HI, Krams M, Agrafiotis DK, DiBernardo A, Chavez JC, Littman GS, et al. Robotic measurement of arm movements after stroke establishes biomarkers of motor recovery. Stroke. 2014;45(1):200–4.
14. Alt Murphy M, Resteghini C, Feys P, Lamers I. An overview of systematic reviews on upper extremity outcome measures after stroke. BMC Neurol. 2015;15:1–15.
15. Lamers I, Kelchtermans S, Baert I, Feys P. Upper limb assessment in multiple sclerosis: a systematic review of outcome measures and their psychometric properties. Arch Phys Med Rehabil. 2014;95(6):1184–200.
16. Kundert R, Goldsmith J, Veerbeek JM, Krakauer JW, Luft AR. What the proportional recovery rule is (and is not): methodological and statistical considerations. Neurorehabil Neural Repair. 2019;33(11):876–87.
17. Winters C, Van Wegen EE, Daffertshofer A, Kwakkel G. Generalizability of the proportional recovery model for the upper extremity after an ischemic stroke. Neurorehabil Neural Repair. 2015;29(7):614–22.
18. Volpe BT, Ferraro M, Krebs HI, Hogan N. Robotics in the rehabilitation treatment of patients with stroke. Curr Atheroscler Rep. 2002;4:270–6.
19. Zeiaee A, Soltani-Zarrin R, Langari R, Tafreshi R. Design and kinematic analysis of a novel upper limb exoskeleton for rehabilitation of stroke patients. In: 2017 international conference on rehabilitation robotics (ICORR) IEEE; 2017. p. 759–764.
20. Bosecker C, Dipietro L, Volpe B, Igo KH. Kinematic robot-based evaluation scales and clinical counterparts to measure upper limb motor performance in patients with chronic stroke. Neurorehabil Neural Repair. 2010;24(1):62–9.
21. Krebs HI, Hogan N, Aisen ML, Volpe BT. Robot-aided neurorehabilitation. IEEE Trans Rehabil Eng. 1998;6(1):75–87.
22. Volpe BT, Krebs HI, Hogan N. Is robot-aided sensorimotor training in stroke rehabilitation a realistic option? Curr Opin Neurol. 2001;14(6):745–52.
23. Lassi M, Dalise S, Bandini A, Spina V, Azzollini V, Vissani M, et al. Neurophysiological underpinnings of an intensive protocol for upper limb motor recovery in subacute and chronic stroke patients. Eur J Phys Rehabil Med. 2023;60(1):13.
24. Laut J, Porfiri M, Raghavan P. The present and future of robotic technology in rehabilitation. Curr Phys Med Rehabil Rep. 2016;4:312–9.
25. Harari Y, O'Brien MK, Lieber RL, Jayaraman A. Inpatient stroke rehabilitation: prediction of clinical outcomes using a machine-learning approach. J Neuroeng Rehabil. 2020;17:1–10.
26. Quattrocelli S, Russo EF, Gatta MT, Filoni S, Pellegrino R, Cangelmi L, et al. Integrating machine learning with robotic rehabilitation may support prediction of recovery of the upper limb motor function in stroke survivors. Brain Sci. 2024;14(8):759.
27. Saes M, Mohamed Refai M, van Beijnum BJF, Bussmann J, Jansma EP, Veltink PH, et al. Quantifying quality of reaching movements longitudinally post-stroke: a systematic review. Neurorehabil Neural Repair. 2022;36(3):183–207.
28. Lin PJ, Jia T, Li C, Li T, Qian C, Li Z, et al. CNN-based prognosis of BCI rehabilitation using EEG from first session BCI training. IEEE Trans Neural Syst Rehabil Eng. 2021;29:1936–43.
29. Saes M, Meskers CG, Daffertshofer A, Van Wegen EE, Kwakkel G. Are early measured resting-state EEG parameters predictive for upper limb motor impairment six months poststroke? Clin Neurophysiol. 2020;132(1):56–62. https://doi.org/10.1016/j.clinph.2020.09.031.
30. Hussain I, Park SJ. Prediction of Myoelectric Biomarkers in Post-Stroke Gait. Sensors. 2021;21(16). https://www.mdpi.com/1424-8220/21/16/5334, https://doi.org/10.3390/s21165334.
31. Pierella C, Pirondini E, Kinany N, Coscia M, Giang C, Miehlbradt J, et al. A multimodal approach to capture post-stroke temporal dynamics of recovery. J Neural Eng. 2020;17(4):045002.
32. Beuter A, Carrière L, Boucher JP. Relationships between electromyography and kinematics in human stepping strategies. Neurosci Lett. 1987;77(1):119–23.
33. Mustard B, Lee R. Relationship between EMG patterns and kinematic properties for flexion movements at the human wrist. Exp Brain Res. 1987;66(2):247–56.
34. Lenzo B, Bergamasco M, Salsedo F. Actuating method and device for human interaction multi-joint mechanisms, WO Patent Application; 2015.
35. Pirondini E, Coscia M, Marcheschi S, Roas G, Salsedo F, Frisoli A, et al. Evaluation of the effects of the Arm Light Exoskeleton on movement execution and muscle activities: a pilot study on healthy subjects. J Neuroeng Rehabil. 2016;13:1–21.
36. Sadaka-Stephan A, Pirondini E, Coscia M, Micera S. Influence of trajectory and speed profile on muscle organization during robot-aided training. In: 2015 IEEE International Conference on Rehabilitation Robotics (ICORR) IEEE; 2015. p. 241–246.
37. American Society for Surgery of the Hand.: Muscles of the Hand; 2025. Accessed July 19, 2025. URL: https://www.assh.org/handcare/safety/muscles. ASSH Hand Care & Safety.
38. Dipietro L, Sabatini AM, Dario P. Artificial neural network model of the mapping between electromyographic activation and trajectory patterns in free-arm movements. Med Biol Eng Compu. 2003;41(2):124–32.
39. Hermens HJ, Freriks B, Disselhorst-Klug C, Rau G. Development of recommendations for SEMG sensors and sensor placement procedures. J Electromyogr Kinesiol. 2000;10(5):361–74.
40. Perotto AO. Anatomical guide for the electromyographer: the limbs and trunk. Charles C Thomas Publisher; 2011.
41. Lencioni L. Bimanual Robotic Exoskeletal Platform for the Treatment of the Upper Limb in Patients With Stroke: A Feasibility Study (Birehab); 2022. ClinicalTrials.gov Identifier: NCT05176600. Sponsor: Wearable Robotics srl. Responsible Party: Lucia Lencioni (Wearable Robotics srl.). https://clinicaltrials.gov/study/NCT05176600.
42. Wearable Robotics srl . Effect of an Automatic Personalized Robot-assisted Rehabilitation on Cortical Organization and Clinical Recovery After Stroke; 2016. ClinicalTrials.gov Identifier: NCT02770300. https://clinicaltrials.gov/study/NCT02770300.
43. Cheung VC, Piron L, Agostini M, Silvoni S, Turolla A, Bizzi E. Stability of muscle synergies for voluntary actions after cortical stroke in humans. Proc Natl Acad Sci. 2009;106(46):19563–8.
44. Kieliba P, Tropea P, Pirondini E, Coscia M, Micera S, Artoni F. How are muscle synergies affected by electromyography pre-processing? IEEE Trans Neural Syst Rehabil Eng. 2018;26(4):882–93.
45. Lee D, Seung HS. Algorithms for non-negative matrix factorization. Advances in neural information processing systems. 2000;13.
46. Van Der Pas SC, Verbunt JA, Breukelaar DE, Van Woerden R, Seelen HA. Assessment of arm activity using triaxial accelerometry in patients with a stroke. Arch Phys Med Rehabil. 2011;92(9):1437–42.
47. Flash T, Hogan N. The coordination of arm movements: an experimentally confirmed mathematical model. J Neurosci. 1985;5(7):1688–703.
48. Rohrer B, Fasoli S, Krebs HI, Hughes R, Volpe B, Frontera WR, et al. Movement smoothness changes during stroke recovery. J Neurosci. 2002;22(18):8297–304.
49. Ariano P, Rosa GN, Giordano S, Privitera L, SRL MT.: A method of detecting and identifying muscle activations made by a user, corresponding system and computer program product; 2023. WO2024157145A1. https://patents.google.com/patent/WO2024157145A1/en?q=(paolo)&inventor=ariano &oq=paolo+ariano.
50. Feng CJ, Mak AF. Three-dimensional motion analysis of the voluntary elbow movement in subjects with spasticity. IEEE Trans Rehabil Eng. 1997;5(3):253–62.
51. Aprile I, Germanotta M, Cruciani A, Loreti S, Pecchioli C, Cecchi F, et al. Upper limb robotic rehabilitation after stroke: a multicenter, randomized clinical trial. J Neurol Phys Ther. 2020;44(1):3–14.
52. Dehem S, Gilliaux M, Stoquart G, Detrembleur C, Jacquemin G, Palumbo S, et al. Effectiveness of upper-limb robotic-assisted therapy in the early rehabilitation phase after stroke: A single-blind, randomised, controlled trial. Ann Phys Rehabil Med. 2019;62(5):313–20.
53. Ding C, Peng H. Minimum redundancy feature selection from microarray gene expression data. J Bioinform Comput Biol. 2005;3(02):185–205.
54. Kononenko I, Šimec E, Robnik-Šikonja M. Overcoming the myopia of inductive learning algorithms with RELIEFF. Appl Intell. 1997;7:39–55.
55. Hawe RL, Scott SH, Dukelow SP. Taking proportional out of stroke recovery. Stroke. 2019;50(1):204–11.
56. Hope TM, Friston K, Price CJ, Leff AP, Rotshtein P, Bowman H. Recovery after stroke: not so proportional after all? Oxford University Press; 2019.
57. Goldsmith J, Kitago T, de la Garza AG, Kundert R, Luft A, Stinear C, et al. Arguments for the biological and predictive relevance of the proportional recovery rule. Elife. 2022;11:e80458.
58. Li P, Pei Y, Li J. A comprehensive survey on design and application of autoencoder in deep learning. Appl Soft Comput. 2023;138:110176.

59. Hiragami S, Inoue Y, Harada K. Minimal clinically important difference for the fugl-meyer assessment of the upper extremity in convalescent stroke patients with moderate to severe hemiparesis. J Phys Ther Sci. 2019;31(11):917–21. https://doi.org/10.1589/jpts.31.917.
60. Galvão S, De Oliveira LF, De Lima R, Xerez D, Menegaldo LL. Shear wave elastography of the brachioradialis spastic muscle and its correlations with biceps brachialis and clinical scales. Clin Biomech. 2022;97:105687. https://doi.org/10.1016/j.clinbiomech.2022.105687.
61. Seo G, Houston M, Portilla M, Fang F, Park J, Lee H, et al. Expanding the repertoire of intermuscular coordination patterns and modulating intermuscular connectivity in stroke-affected upper extremity through electromyogram-guided training: a pilot study. In: 2022 44th Annual International Conference of the IEEE Engineering in Medicine & Biology Society (EMBC); 2023. p. 1–https://doi.org/10.1109/EMBC40787.2023.10341085.
62. Yu B, Zhang X, Cheng Y, Liu L, YanJiang N, Wang J, et al. The effects of the biceps brachii and brachioradialis on elbow flexor muscle strength and spasticity in stroke patients. Neural Plast. 2022;2022:1–15. https://doi.org/10.1155/2022/1295908.
63. Lassi M, Bandini A, Spina V, Azzollini V, Dalise S, Mazzoni A, et al. Classification of upper limb impairment in acute stroke patients using resting-state EEG markers and machine learning. In: 2023 11th International IEEE/EMBS Conference on Neural Engineering (NER) IEEE; 2023. p. 1–4.
64. Giang C, Pirondini E, Kinany N, Pierella C, Panarese A, Coscia M, et al. Motor improvement estimation and task adaptation for personalized robot-aided therapy: a feasibility study. Biomed Eng Online. 2020;19(1):33.

Publisher's Note