Research Square

Preprints are preliminary reports that have not undergone peer review.
They should not be considered conclusive, used to inform clinical practice,
or referenced by the media as validated information.

Towards At-Home Stroke Rehabilitation: Fully Automated Identification of Upper Extremity Functional Movements in Daily-Life Activities of Stroke Patients

Sahel Akbari

s.akbari@erasmusmc.nl

Erasmus MC

Herwin L. D. Horemans

Erasmus MC

Johannes B.J. Bussmann

Erasmus MC

Marc Evers

Rijndam Revalidatiecentrum

Arkady Zgonnikov

Delft University of Technology

Research Article

Keywords: At-Home Stroke Rehabilitation, Upper Extremity Functional Movements, Upper Extremity Movement Identification, Movement Temporal Segmentation, Inertial Measurement Units (IMUs)

Posted Date: June 22nd, 2026

DOI: https://doi.org/10.21203/rs.3.rs-9901929/v1

License: This work is licensed under a Creative Commons Attribution 4.0 International License.
Read Full License

Additional Declarations: No competing interests reported.

Towards At-Home Stroke Rehabilitation: Fully Automated Identification of Upper Extremity Functional Movements in Daily-Life Activities of Stroke Patients

Sahel Akbari[a,b,*], Herwin L. D. Horemans[a,c], Johannes B.J. Bussmann[a], Marc Evers[c], Arkady Zgonnikov[b]

[a] *Dept. Rehabilitation Medicine, Erasmus MC University Medical Center, Rotterdam, the Netherlands*
[b] *Dept. Cognitive Robotics, Faculty of Mechanical Engineering, Delft University of Technology, Delft, the Netherlands*
[c] *Rijndam Rehabilitation, Rotterdam, the Netherlands*

Abstract

Upper extremity (UE) recovery remains a major challenge after stroke, requiring high-intensity practice, increasingly delivered through home-based rehabilitation. Monitoring UE functional movements during daily life activities is essential for assessing and providing feedback on patient performance. Inertial Measurement Units (IMU) combined with machine learning approaches have advanced UE monitoring. However, existing methods often lack validation in home-like environments, show limited generalization across patients, do not account for clinically irrelevant movements, and do not explicitly perform movement temporal segmentation. To address these limitations, we developed a method to identify Reaching, Lifting, and Supination/Pronation functional movements from IMU sensor data collected from stroke patients. A CNN+BiLSTM+Attention deep-learning classifier was evaluated using free-living measurements under two scenarios: a subject-specific (baseline) scenario and a group-cohort scenario to assess generalization. Lastly, we extended the classifier into a fully automated segment identification framework by introducing an "Unknown" class to account for irrelevant movements and integrating the classifier with a Hidden Markov

*s.akbari@erasmusmc.nl

Model (HMM) sequence-decoder for temporal segmentation. Results showed that the group-cohort classification model outperformed the subject-specific baseline (F1-score=0.90 vs. 0.69), demonstrating improved generalization. The automated framework achieved reliable segment identification, with the highest performance for Lifting movements (F1-score=0.84). Most correctly identified segments had Intersection-over-Union (IoU) values > 0.50, indicating good temporal localization, and movement duration was well preserved for Lifting and Supination/Pronation. Reaching was the most difficult movement to identify (F1-score=0.63), and to preserve the duration. These findings support the potential of the proposed framework for continuous monitoring of UE functional movements in at-home rehabilitation.

Keywords: At-Home Stroke Rehabilitation, Upper Extremity Functional Movements, Upper Extremity Movement Identification, Movement Temporal Segmentation, Inertial Measurement Units (IMUs)

1. Introduction

Upper extremity (UE) recovery remains a major clinical challenge after stroke, as UE impairments are among the most persistent and disabling consequences of stroke [1, 2, 3]. These impairments severely impact daily life upper extremity activities, reducing independence and quality of life [4, 5]. Achieving meaningful UE recovery requires high-intensity, repetitive practice [6, 7]. However, current clinical pathways often provide insufficient therapy time and repetitions, creating a gap between recommended and delivered rehabilitation intensity [8, 9, 10]. This limitation has accelerated the adoption of home-based rehabilitation, which increases training frequency and demonstrates comparable effectiveness to conventional in-clinic therapy for post-stroke motor recovery [11, 12].

Daily-life upper-extremity activity is a central component of patient functioning within the performance domain of the International Classification of Functioning (ICF), which emphasizes how patients act in real-life rather than controlled settings [13, 14]. Knowledge of daily life upper extremity activity is therefore essential for assessing patient status and guiding treatment goals [15, 16]. In stroke rehabilitation, increased arm use is a therapeutic goal in itself, as evidenced by the effectiveness of constraint-induced movement therapy (CIMT) [17, 18]. Based on the CIMT principles, daily life upper extremity activity feedback systems have been developed and evaluated [19].

However, such systems typically provide feedback on general arm use rather than specific UE functional movements, although understanding how the arm is functionally used in daily life is equally important [20]. In this context, Human Activity Recognition (HAR) can provide objective and precise information on both the general arm use and the quality and type of UE functional movements performed during daily life [21, 22, 23]. While optical motion capture provides detailed measurement, it remains laboratory-based [24], whereas body-worn Inertial Measurement Units (IMUs) enable continuous at-home monitoring of UE functional movements [21, 25, 26].

Considerable research has focused on monitoring UE functional movements using wearable sensors combined with numerous machine learning (ML) approaches. These include conventional ML models requiring manual feature engineering [27, 28, 29, 30], as well as neural network–based methods that eliminate the need for handcrafted features [31, 32, 33, 34, 35]. More recently, deep learning approaches, particularly convolutional neural networks (CNNs) and recurrent neural networks (RNNs), have gained attention due to simpler feature processing and superior performance compared to traditional ML techniques [36, 37, 38].

However, despite this progress, existing algorithms are not yet ready for practical deployment in at-home stroke rehabilitation settings. This gap stems from several unresolved limitations. First, existing approaches have not been validated for continuous identification of UE functional movement types in stroke patients performing *unconstrained activities in home-like environments*. In prior studies [27, 29, 30, 32, 33, 35], patients typically performed *predefined activities* or exercises under explicit instructions, limiting the ecological validity of the resulting models to variable, context-dependent movement patterns that may occur in home-like environments. Second, existing approaches show *limited generalization across the stroke population*. Clinical deployment requires models that generalize across subjects to minimize *subject-specific training*, as extensive per-patient training reduces usability and scalability in real-world settings. However, studies evaluating unconstrained scenarios rarely address between-subject generalization [31], while broader HAR literature reports reduced performance on unseen subjects [39, 40, 41]. Consequently, deployment across patients still often requires burdensome individualized data collection and model adaptation. Third, existing approaches rarely account for *unknown or clinically irrelevant movements* inherent in naturalistic recordings or describe how they are *distinguished from intended UE functional movements* [28, 31]. As a result, non-functional or

incidental movements may be misclassified as meaningful activity, leading to overestimation of UE functional use and potentially misleading clinical interpretation and feedback. Fourth, most existing deep learning approaches rely on *fixed-size inputs* (commonly obtained via sliding windows over continuous sensor data), whereas arm movements in daily life *do not follow predefined durations or sequences*. Although some algorithms support variable-length sequences (e.g., dense labeling), explicit *temporal segmentation remains insufficiently addressed* [42, 43]. Window-based approaches depend on arbitrary lengths and often fail to recover coherent movement segments due to substantial jitter between window-level predictions, thereby limiting accurate identification of movement onset, offset, and duration. This reflects a recognized gap in arm movement activity identification during at-home use: the lack of integrated classification and temporal segmentation strategies [28].

Overall, although each limitation has been partially addressed, *no existing method overcomes all four*, limiting applicability for practical deployment in at-home rehabilitation settings. Therefore, the higher-level aim of this study is to develop a method for the identification (including both classification and segmentation) of *three UE functional movements* – Reaching, Lifting (e.g., drinking/eating), and Supination/Pronation – in stroke patients, and to evaluate it in a *setting representative of at-home rehabilitation*.

More specifically, this study addresses the identified limitations through three key components. First, to address the lack of validation in unconstrained environments, we incorporate semi-structured and, in particular, free-living measurements to better reflect real-life conditions. These were recorded in a rehabilitation kitchen within a clinical setting, alongside structured measurements for baseline evaluation. Second, we assess a group-trained deep-learning model based on a previously proposed classification model [44], now enhanced to a CNN+BiLSTM architecture and a self- attention mechanism [45], to examine generalization across patients. The model is trained on cohort data and tested on unseen free-living measurements. This design minimizes subject-specific training, avoids prior in-clinic calibration, and enables immediate home-to-home transfer. To provide a baseline reference, we evaluate a subject-specific model trained and tested on data from the same participant across different measurement conditions (structured and free-living). Lastly, to account for clinically irrelevant movements and the lack of explicit temporal segmentation, we extend the proposed method into a fully automated movement segment identification framework. The CNN+BiLSTM+Attention deep-learning classification model is augmented

with an additional "Unknown" movement class, representing movements outside the three target functional categories. A Hidden Markov Model (HMM)-based sequence decoder is then integrated to smooth window-level predictions and recover coherent movement segments. This fully automated identification framework represents a solution that has been validated in patients performing complex daily life activities and is now ready for future deployment in real-home stroke rehabilitation settings. In addition, the recorded dataset will be made publicly available to support future research and benchmarking.

2. Methodology

2.1. Participants

We included fifteen individuals with subacute stroke, ranging from mild to moderate upper limb impairments, who were recruited from inpatient rehabilitation at Rijndam Rehabilitation Center (mean age 61.1 $\pm$ 8.8 years; 3 female). Participants were assessed at 39.1 $\pm$ 17.5 days post-stroke. Most participants had ischemic stroke (13/15), while two had hemorrhagic stroke. The majority had right-sided dominance (13/15) and right-sided paresis (10/15). Clinical scores indicated moderate impairments overall (FMA-UE: 54.6 $\pm$ 8.6; ARAT: 46.4 $\pm$ 12.6).

Inclusion criteria were: Fugl–Meyer Assessment for the upper extremity (FMA-UE) $\geqslant$ 32/66, sufficient grasping ability to handle objects, and sufficient walking ability to navigate a kitchen environment, as judged by the treating physiotherapist; age $\geqslant$ 18 years; and proficiency in Dutch or English. Exclusion criteria included severe cognitive impairments, severe upper extremity impairments, motor planning deficits, or aphasia affecting communication. Ethical approval was obtained from the Ethics Committee of Erasmus Medical Center (MEC-2024-0640, Towards@HomeRehab). All participants provided written informed consent prior to their participation in accordance with the Declaration of Helsinki.

2.2. Experimental Protocol

For each participant, data were collected using structured, semi-structured, and free-living protocols to capture UE functional movements under progressively less constrained conditions. Structured measurements were used for baseline evaluation, while semi-structured and free-living measurements better reflect movement behavior during everyday activities. Experiments

were conducted in a rehabilitation kitchen environment at the Rijndam Rehabilitation Center. All measurements were video recorded for subsequent annotation.

Structured Protocol. The structured protocol followed the procedure described in [44] and involved three isolated UE movements:

1. **Reaching**: forward extension of the arm involving shoulder flexion and elbow extension;
2. **Lifting**: elevation of the hand toward the mouth (e.g., drinking);
3. **Supination/Pronation**: rotation of the forearm around its longitudinal axis, resulting in the palm facing downward or upward.

Participants were seated in front of a table and performed three measurements of ten repetitions per movement under explicit instructions, with five-second pauses between repetitions.

Semi-Structured Protocol. The semi-structured protocol also followed [44], with reducing the activities from fourteen to ten to improve task flow and reduce patient burden, including the removal of butter spreading with a knife (activities 9, 10) and the cup/bread replacement tasks. It included a list of consecutive kitchen-related activities that combined variations of the same three UE movements within everyday actions (e.g., reaching, drinking, opening/closing a water tap, washing the face, and page turning). The activities were carried out without explicit instructions regarding task execution. Object use and placement were predefined across activities. No constraints were imposed on body position or movement speed. Participants completed the activity list sequentially once, with five-second pauses between activities.

Free-Living Protocol. In addition to the previous protocols, a free-living protocol was included in which patients performed three higher-level activities of daily living, each composed of multiple subordinate activities. The higher-level activities were: (1) prepare and eat a slice of bread with peanut butter; (2) make tea and have it with a cookie; and (3) wash a cup and set the table. These activity titles were assigned solely to elicit goal-directed behavior involving at least some of the three predefined UE functional movements. Patients were provided only with the activity title, while kitchen objects were placed randomly, and no further instructions were given regarding task execution, task order, or object use and placement. The measurement was

intentionally non-standardized, allowing recording duration and the number of subordinate activities to vary across patients. Measurements were included even if the higher-level activity was not fully completed, provided that instances of the three UE functional movements could be identified.

2.3. Motion Data Acquisition

We employed four QSense 9DOF motion tracking sensors (2M Engineering, [46]), each integrating a 3-axis accelerometer, gyroscope, and magnetometer. However, only data from accelerometers and gyroscopes were used, as environmental factors, such as magnetic field disturbances, could significantly influence magnetometer data. These IMUs were placed on the dorsal side of the affected arm, arranged from distal to proximal at the hand (along the third metacarpal), distal forearm (proximal to the wrist), distal upper arm (proximal to the lateral epicondyle), and over the acromion of the shoulder. Data were sampled at 50 Hz and transmitted via Bluetooth to a computer running the QSense Motion Application.

2.4. Proposed Method

The proposed method is initially based on a deep-learning classification model to identify movement types in free-living measurements, without explicitly recovering movement segments. It is then extended to a fully automated movement segment identification framework that integrates the classifier with a sequence decoding approach for temporal segmentation. The proposed method is illustrated in Figure 1.

2.4.1. Data Pre-Processing

General Pre-Processing. IMU data were preprocessed by constructing 16 channels per unit, including three-dimensional acceleration and angular velocity, their Euclidean norms, acceleration jerk signals and angular acceleration signals (first temporal derivatives) [47, 48], and their corresponding norms. To compensate for inter-sensor synchronization discrepancies, measurements were truncated to a common length, and missing samples were linearly interpolated [49]. The IMU data were subsequently band-pass filtered using a third-order Butterworth filter with cutoff frequencies of 0.1–12 Hz [28, 29].

Fixed-Size Windowing. To be able to use our existing deep-learning classification model [44], the IMU data were processed into fixed-sized windows using a 1-second sliding-window approach with 50% overlap. Two windowing strategies were employed: (i) windowing constrained within manually annotated ground-truth segment boundaries, obtained via synchronization with video recordings, and (ii) windowing applied continuously across the entire data stream, independent of segment boundaries. The first strategy included three movement classes (Reaching, Lifting, and Supination/Pronation) and was used for movement type classification (Section 2.4.3), whereas the second additionally incorporated an "Unknown" class and enabled fully automated movement segment identification by integrating classification with sequence decoding (Section 2.4.4) (Figure1). For labeling, if the midpoint of a sliding window fell within one of the three ground-truth segments, it was assigned the corresponding label; otherwise, it was labeled as "Unknown".

2.4.2. Deep-Learning Classification Model

The deep-learning classification model consists of four blocks, and is illustrated in Figure 2.

- **CNN block:** The CNN block acts as a feature extractor, which slides 1D-filters across the temporal dimension, and looks for local temporal motion patterns of the data. Two Conv1D layers are used with kernel size 3 and stride 1: the first has 32 filters and the second has 64 filters, both using ReLU activation and L_2 weight regularization. Each convolution is followed by batch normalization to normalize the input across channels, thereby accelerating convergence. Max-pooling with pool size 2 is applied after each convolutional stage to reduce the temporal resolution (from 50 (fs= 50 Hz) to 25 and then to 12 time steps).

- **BiLSTM block:** The resulting feature sequence from the CNN block is fed to a bidirectional LSTM with 64 units in each direction. This allows the model to capture longer-range temporal dependencies among the patterns the CNN identified by encoding both past and future context at each time step, yielding a 128-dimensional representation per step.

- **Attention Mechanism block:** A temporal attention mechanism is applied to the BiLSTM outputs to emphasize which part of the data

is more informative. A dense layer with one neuron computes a scalar attention score for each time step. Then a softmax converts scores into weights that sum to one. These weights are used to scale the BiLSTM timestep's 128-dimensional vector by its attention weight. Finally, a weighted sum over time is performed to produce a single 128-dimensional window-level representation that emphasizes the most relevant temporal moments for classification.

- **Output block:** The output block comprised two dense layers with 256 units, followed by a final classification layer of 3 or 4 neurons, depending on whether windowing was restricted to ground-truth movement segment boundaries or applied continuously over the full data stream, including the "Unknown" movement class.

The proposed model used the Adam optimizer [50] with a learning rate of 0.001 and a batch size of 32, training for 20 epochs, and employed categorical cross-entropy as the loss function.

2.4.3. Movement Type Classification Within Ground-Truth Segment Boundaries

Movement type classification within ground-truth segment boundaries refers to identifying one of three upper-extremity movement types using a hybrid CNN+BiLSTM+Attention model. Ground-truth annotations were manually obtained by synchronizing IMU data with video recordings, and classification was performed using fixed-size sliding windows within the annotated movement segment boundaries. Each window was classified independently (window-level classification), meaning predictions were constrained to manually labeled movement segment boundaries and were not generated for the full continuous signal.

2.4.4. Fully Automated Movement Segment Identification Framework

Window-level IMU classification, constrained by ground-truth segment boundaries, limits the classifier to (i) identifying only three movement types and (ii) not explicitly modeling their temporal segments. However, in real applications of prolonged home monitoring, continuous recordings include a substantial proportion of "Unknown" movements (movements other than the intended types), which may lead the deep neural classifier to over-predict this class. In addition, window-level predictions may jitter among the three

UE movements because deep classifiers do not enforce long-range temporal consistency. Therefore, the observed sequence of predicted labels may lack precision to recover the movement segment, as it can be influenced by the predominance of "Unknown" movements as well as rapid state transitions among the other three movement types. This corresponds to the standard HMM decoding task, which can be restated as follows: given a sequence of observed window-level predictions from the neural classifier, how can we infer the most probable underlying sequence of movement states? To address this, we developed a fully automated movement segment identification framework that combines the deep learning classification model (trained to include the fourth "Unknown" class) with an HMM sequence-decoder as a post-processing block to smooth window-level predictions and recover the temporal movement segments (Figure 1). Motivated by the work of [51] and [52], the purpose of this post-processor is to aggregate the sequence of fixed-size rolling windows, in order to smooth the window-level prediction and ultimately to determine the timing and full duration of an activity, without relying on the ground-truth movement segment boundaries. The decoder receives (i) a transition probability matrix estimated from training data, (ii) emission probabilities given by the posterior output of the hybrid CNN+BiLSTM+Attention classifier, and (iii) initial state probabilities. The decoding problem is solved using the Viterbi algorithm [53]; therefore, the Baum–Welch algorithm is not used to optimize the HMM parameters, as the transition and emission matrices are assumed to be derived from the CNN+BiLSTM+Attention classifier. The Viterbi decoder operates according to the following principle:

- Let x_t denote the observed window at time step t, and let $S_t \in \{1, \ldots, K\}$ denote the hidden movement state at time t, where K is the number of movement classes. The sequence of hidden states is denoted by $S = (S_1, \ldots, S_N)$, where N is the total number of windows in the sequence.

- The transition probabilities a_{ij}, form the state transition matrix, expressing the probability of moving from state i to state j between two consecutive windows. These probabilities are estimated from ground-truth training label sequences and are defined as:

$$a_{ij} = P(S_t = j \mid S_{t-1} = i) \tag{1}$$

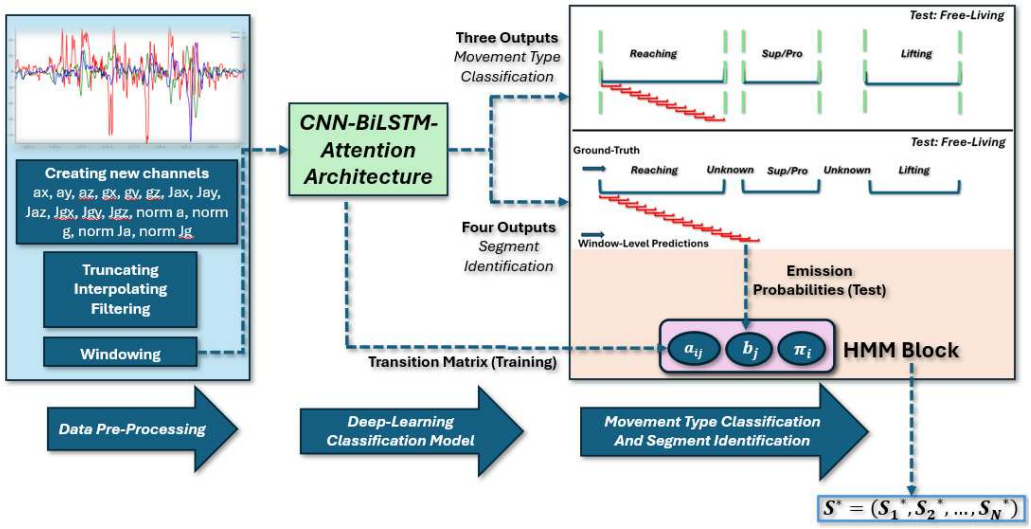


Figure 1: Illustration of the proposed method. Window-level outputs, constrained by ground-truth segment boundaries, are used for movement type classification, while windowing across the full data stream (including "Unknown" segments), integrates the classifier's window-level outputs with an HMM to enable fully automated segment identification and outputs the optimal hidden-state path. The 3D raw gyroscope data are depicted, and the CNN+BiLSTM+Attention model is presented in detail in Figure 2.

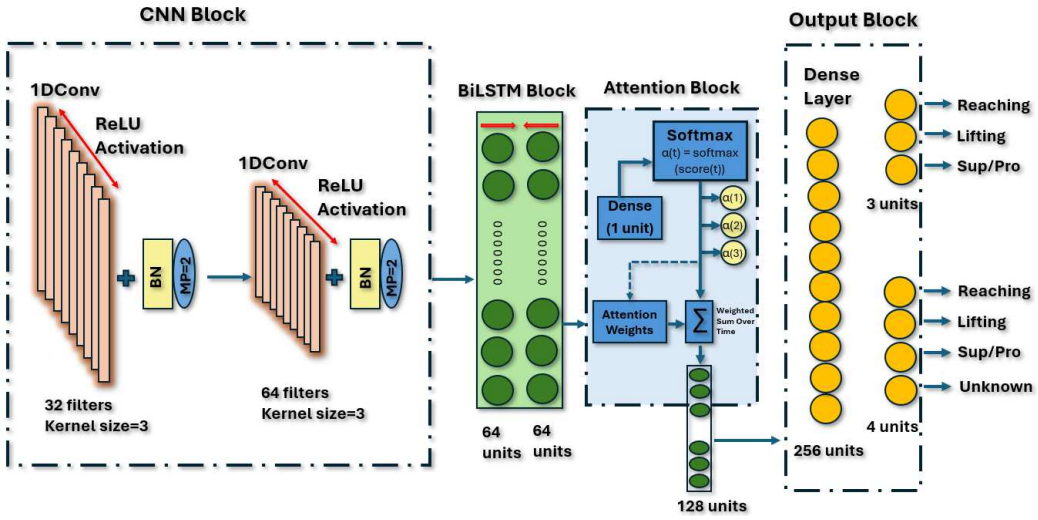


Figure 2: Illustration of the deep-learning classification model. This model is an extension of a previously proposed classification model [44].

- The emission probabilities are defined as the posterior probabilities produced by the proposed classifier, representing the likelihood that window x_t belongs to movement state j (e.g., Reaching, Lifting, Supination/Pronation, or Unknown) at time step t:

$$b_j(t) = P(S_t = j \mid x_t) \tag{2}$$

- The Initial state probabilities π_i, define the likelihood that a recording starts in state i:

$$\pi_i = P(S_1 = i) \tag{3}$$

- In the final step, Viterbi decoding infers the optimal hidden-state path $S^* = (S_1^*, \ldots, S_N^*)$, which is a smoother prediction series than S and best explains the continuous sequence of observations.

$$S^* = \arg\max_S \left[\log \pi_{s_1} + \sum_{t=2}^{N} \log a_{s_{t-1}, s_t} + \sum_{t=1}^{N} \log b_{s_t}(t) \right] \tag{4}$$

2.5. Evaluation Scenarios

Our methodology is evaluated following two experimental scenarios:

Subject-Specific (Clinic-Home Scenario). The clinic-home scenario serves as a baseline evaluation, and it refers to a situation in which movement predictions are personalized to each subject's performance. This scenario assumes that the patient begins using the system at the clinic and is then transferred to use it at home. Hence, the model is trained on all "structured" measurements from a single subject and then tested on the "free-living" measurement from the same subject. During training, one of the three structured measurements was used iteratively as a validation set, yielding a 3-fold Leave-One-Trial-Out (LOTO) cross-validation. In this baseline scenario, our goal was to evaluate whether the model could identify arm movement across different measurement conditions using only the subject-specific data. Therefore, we limited the problem to a movement type classification task, constrained by ground-truth segment boundaries, as explained in subsection 2.4.3; movement segmentation was not considered as the primary aim was to compare evaluation scenarios.

Group-Cohort (Home-Home Scenario). The home-home scenario represents the main evaluation, and it refers to a situation in which the patient begins using the arm movement identification system immediately at home, without prior training in clinics. To ensure that the model captures the substantial variability in daily life performance, rather than relying solely on a personalized free-living measurement, the model was trained on all semi-structured and free-living measurements from all subjects (referred to as the free-style training set), except for one subject, and was tested on the free-living measurement of the left-out subject. This process was repeated iteratively in a Leave-One-Subject-Out (LOSO) manner across all participants. Hence, we employed a group-cohort design to assess the generalizability of identifying arm movements during daily life for an unseen subject. The group-cohort design was then followed in two ways: (i) assessed through movement type classification within ground-truth segment boundaries (subsection 2.4.3), and (ii) assessed through the fully automated movement segment identification framework (subsection 2.4.4). The latter allows the system to be used directly and immediately in daily life by patients, without the need for manual labeling or individual pre-training.

2.5.1. Evaluation Metrics

The CNN+BiLSTM+Attention classifier performance was evaluated using accuracy (overall correctness), precision, recall, and F1-score (harmonic balance between precision and recall) [29]. This evaluation was performed at the window-level, where classification was constrained by the ground-truth segment boundaries and therefore considered only the three movement types. Due to class imbalance, weighted averages were reported for all metrics. Results were averaged across all participants as mean ± standard deviation. In this context, strong performance is indicated by high F1-scores across movement classes rather than accuracy alone. Although no validated thresholds exist, top HAR systems often report F1-scores > 0.75 [54, 55]. Therefore, we considered values < 0.50 low, 0.50–0.75 moderate, 0.75–0.90 high, and > 0.90 excellent.

In contrast to the window-level evaluation of the classifier, the performance of the automated segment identification framework was assessed at the movement segment level. Segment identification performance was evaluated using precision (reflecting false positive segments), recall (reflecting false negative/missed segments), and F1-score as the main metric, capturing the trade-off between over-identifying segments and missed segments. Accu-

racy was not considered for this stage, as the inclusion of the fourth class ("Unknown" movements) results in a large number of overlapping segments, making the assessment of overall correctness less informative.

To evaluate temporal localization (how well each predicted segment overlapped in time with the corresponding ground-truth segment), the mean Intersection-over-Union (IoU) [56] between predicted and ground-truth segments was computed for all correctly identified segments. Further, the mean onset and offset errors were used, defined as the difference between predicted and ground-truth segment boundaries. IoU is interpreted as follows: values < 0.50 are low, > 0.50 are generally considered good, and > 0.75 are high; although no standard exists for movement segment identification, these thresholds follow those used in object detection and image segmentation [57].

Finally, to assess total duration preservation, the total duration of all predicted segments (correctly or incorrectly identified, including false positives and false negatives segments) was compared with the corresponding ground-truth duration for each movement type, and the ratio between predicted and ground-truth durations was reported. Good performance for segment identification is reflected by a high F1-score, high IoU values, small onset and offset errors, and duration ratios close to one. A schematic representation (derived from a snapshot of real data) of the evaluation metrics for the automated movement segment identification framework is shown in Figure 3. Metrics are defined as follows:

$$\text{Accuracy} = \frac{TP + TN}{TP + TN + FP + FN} \tag{5}$$

$$\text{Precision} = \frac{TP}{TP + FP}, \quad \text{Recall} = \frac{TP}{TP + FN} \tag{6}$$

$$\text{F1-score} = \frac{2 \cdot TP}{2 \cdot TP + FP + FN} \tag{7}$$

$$\text{IoU}_k = \frac{|\hat{A}_k \cap A_k|}{|\hat{A}_k \cup A_k|}, \quad (A_k, \hat{A}_k) \in \mathcal{M} \tag{8}$$

$$\text{Onset Error}_k = \hat{t}_k^{\text{start}} - t_k^{\text{start}}, \qquad \text{Offset Error}_k = \hat{t}_k^{\text{end}} - t_k^{\text{end}}, \qquad (A_k, \hat{A}_k) \in \mathcal{M} \tag{9}$$

$$\text{Duration Ratio} = \frac{\sum_{\hat{A}_k \in \hat{\mathcal{A}}} \left(\hat{t}_k^{\text{end}} - \hat{t}_k^{\text{start}}\right)}{\sum_{A_k \in \mathcal{A}} \left(t_k^{\text{end}} - t_k^{\text{start}}\right)} \tag{10}$$

where: TP = True Positive, TN = True Negative, FP = False Positive, FN = False Negative, $\mathcal{A} = \{A_k\}$ and $\hat{\mathcal{A}} = \{\hat{A}_k\}$ denote the sets of ground-truth and predicted segments, respectively, with k indexing individual segments. $\mathcal{M} = \{(A_k, \hat{A}_k)\}$ denotes the set of correctly identified segment pairs used for IoU and onset/offset error computation. $\cap$ and $\cup$ represent the intersection and union, respectively. $t_k^{\text{start}}, t_k^{\text{end}}$ and $\hat{t}_k^{\text{start}}, \hat{t}_k^{\text{end}}$ denote the start and end times of the k-th ground-truth and predicted segments.

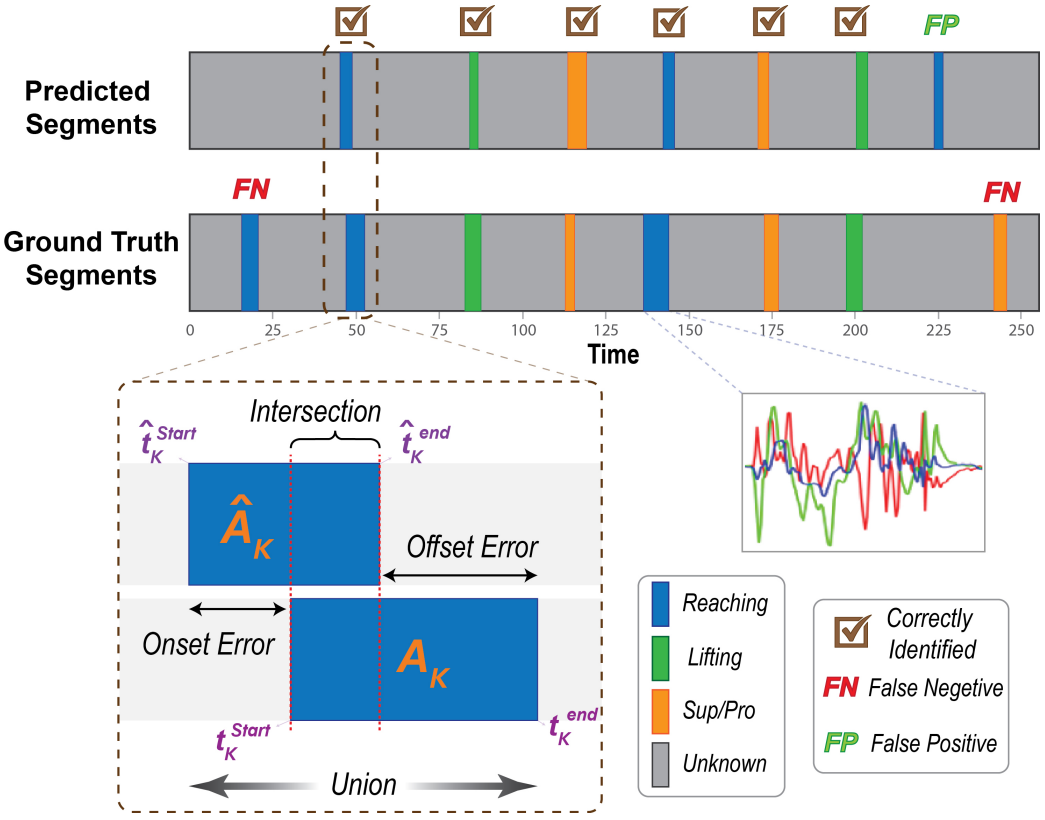


Figure 3: A schematic representation (derived from a snapshot of real data) illustrating evaluation metrics for movement segment identification and temporal localization. The data of a Reaching segment is shown to illustrate that each colored interval represents a temporal instance of a specific segment.

3. Results

Results are reported for each evaluation scenario (subject-specific and group-cohort), including the performance of the classifier when constrained by ground-truth segment boundaries, followed by their comparison, and finally, the performance of the fully automated movement segment identification framework.

3.1. Subject-Specific (Clinic-Home Scenario): Classification Within Ground-Truth Segment Boundaries

According to Table 1, the subject-specific (baseline) models achieved excellent classification performance under structured measurements, with a weighted F1-score of 0.96 ± 0.07. When evaluated on free-living measurement, performance decreased to 0.69 ± 0.18, indicating moderate baseline personalized performance under naturalistic conditions.

	Structured Validation				Corresponding Free-living Test			
Metric	Accuracy	Precision	Recall	F1-score	Accuracy	Precision	Recall	F1-score
Mean (SD)	0.95 (0.07)	0.97 (0.04)	0.95 (0.07)	0.96 (0.07)	0.69 (0.16)	0.79 (0.11)	0.69 (0.16)	0.69 (0.18)

Table 1: Performance of the subject-specific (baseline) classification model within ground-truth segment boundaries in the clinic-home scenario. Values are averaged across all participants. Structured validation corresponds to structured 3-fold cross-validation during training, while the corresponding free-living test reflects the average performance on the corresponding subject's free-living measurement.

3.2. Group-Cohort (Home-Home Scenario)

The group-cohort scenario served as the main evaluation and was assessed in two ways to examine generalization to an unseen subject.

3.2.1. Movement Type Classification Within Ground-Truth Segment Boundaries

When evaluated within ground-truth segment boundaries (Table 2), the group model achieved a weighted F1-score of 0.89 ± 0.02 on the cohort validation, reflecting model performance during training. On an unseen free-living test data, it achieved an F1-score of 0.90 ± 0.08, indicating high generalization performance.

	Cohort Validation				Unseen Free-living Test			
Metric	Accuracy	Precision	Recall	F1-score	Accuracy	Precision	Recall	F1-score
Mean (SD)	0.89 (0.02)	0.89 (0.02)	0.89 (0.02)	0.89 (0.02)	0.90 (0.08)	0.92 (0.05)	0.90 (0.08)	0.90 (0.08)

Table 2: Performance of the group-cohort classification model within ground-truth segment boundaries in the home-home scenario. Results are averaged across leave-one-subject-out iterations. Cohort validation corresponds to training on the free-style dataset, while the unseen free-living test result reflects the performance for an unseen subject's free-living measurement.

Figure 4 compares subject-specific (baseline) and group-cohort weighted F1-scores for free-living test measurements across participants. The group-cohort model achieved significantly higher weighted F1-scores and lower variability compared with the subject-specific baseline (0.90 ± 0.08 vs. 0.69 ± 0.18; Wilcoxon signed-rank test, $W = 120$, $p < 0.001$). Overall, the group-cohort model improved window-level classification performance in free-living measurements compared with the subject-specific baseline.

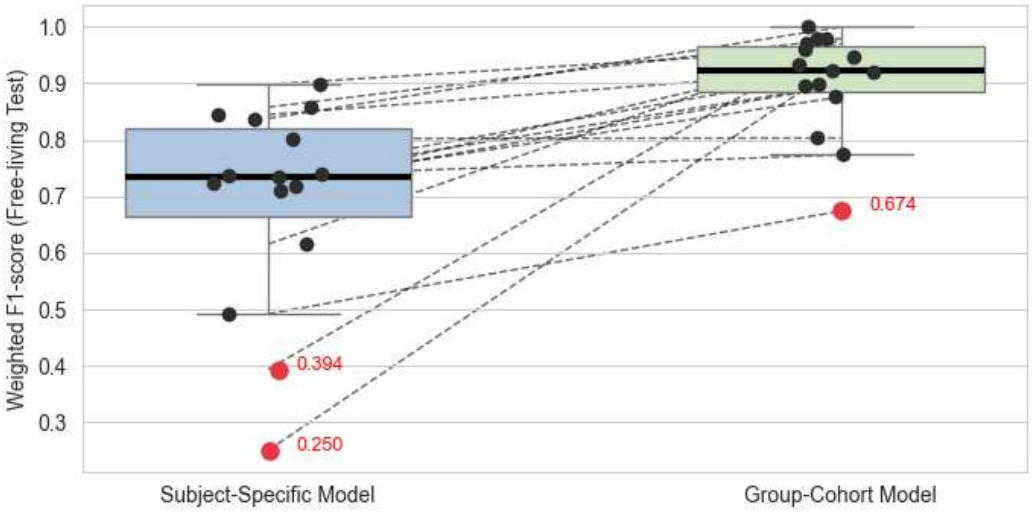


Figure 4: Performance of the classification model within ground-truth segment boundaries: Subject-specific (baseline) and group-cohort weighted F1-scores for free-living test measurements across participants. Each point represents an individual subject, and dashed lines connect points belonging to the same subject.

3.2.2. Fully Automated Movement Segment Identification Framework

Segment Identification Performance. Table 3 indicates that the automated framework achieved high identification performance for Lifting (F1-score = 0.84) and Supination/Pronation (F1-score = 0.80) segments, whereas Reaching segments exhibited lower identification performance (F1-score = 0.63). Reaching segments also showed lower recall (0.57) compared to Lifting (0.88) and Supination/Pronation (0.79), indicating that reaching movements were more frequently missed by the framework. The Unknown class is not included in the interpretation of the evaluation metrics because it was predominant across all measurements, making it difficult to treat it as an identifiable movement segment.

GT \ Pred	R	L	S/P	U	Precision (SD)	Recall (SD)	F1-score (SD)
R	59	0	0	44	0.72 (0.22)	0.57 (0.26)	0.63 (0.24)
L	0	62	0	8	0.82 (0.20)	0.88 (0.18)	0.84 (0.17)
S/P	0	0	71	18	0.81 (0.14)	0.79 (0.24)	0.80 (0.17)
U	22	13	16	—	—	—	—

Table 3: Performance of the fully automated movement segment identification framework: The confusion matrix and segment identification performance metrics for segment types "Reaching" (**R**), "Lifting" (**L**), "Supination/Pronation" (**S**/**P**), and "Unknown" (**U**), across all subjects' recordings. Precision, recall, and F1-score are reported with standard deviation for each segment type, except for "Unknown" segments.

Temporal Localization Performance. According to Table 4, the framework demonstrated overall good temporal localization performance for correctly identified segments, as reflected by mean IoU values > 0.5 and relatively small onset and offset errors. Lifting segments demonstrated the highest overlap between predicted and ground-truth segments (IoU = 0.81), with the lowest combined absolute onset and offset errors ($0.04 + 0.17 = 0.21$), indicating the most accurate temporal localization (coverage). In contrast, Reaching and Supination/Pronation showed lower overlap and relatively larger combined

absolute onset and offset errors. The framework also exhibited high temporal localization for Unknown segments (IoU = 0.93).

Metric	IoU (SD)	Onset Error (SD)	Offset Error (SD)	Correctly Identified Segments
R	0.68 (0.22)	0.32 (0.55)	-0.07 (0.49)	59
L	0.81 (0.18)	0.17 (0.39)	-0.04 (0.47)	62
S/P	0.66 (0.25)	0.05 (1.13)	0.35 (1.26)	71
U	0.93 (0.02)	— —	— —	—

Table 4: Performance of the fully automated movement segment identification framework: Mean (standard deviation) of Intersection-over-Union (IoU) and onset/offset errors for correctly identified segment types "Reaching" (**R**), "Lifting" (**L**), "Supination/Pronation" (**S/P**), and "Unknown" (**U**), across all subjects' recordings. IoU and onset/offset error analyses were based on the correctly identified segments (**R**:59, **L**:62, **S/P**:71). Onset/offset errors are expressed in seconds; positive values indicate that the predicted segment occurs later than the ground truth, whereas negative values indicate earlier predictions.

To further characterize temporal localization performance beyond the mean values reported in Table 4, Figure 5 illustrates the distribution of IoU values across all correctly identified segments. Most IoU values for Lifting segments exceeded 0.75, confirming their high overlap with the ground truth. Across all movement types, the majority of correctly identified segments achieved IoU values > 0.5, indicating generally good overlap and suggesting that the proposed classifier + HMM framework reliably captured the temporal extent of movement episodes and approximated their boundaries in free-living measurements.

Duration Analysis. Table 5 shows the ratio between the total duration of all predicted segments (correctly or incorrectly identified, including false positives and false negatives segments) and the corresponding ground-truth duration. Lifting segments showed a predicted duration close to the ground-truth duration (ratio = 0.90), Supination/Pronation segments slightly overestimated it (ratio = 1.09), whereas Reaching segments captured a smaller

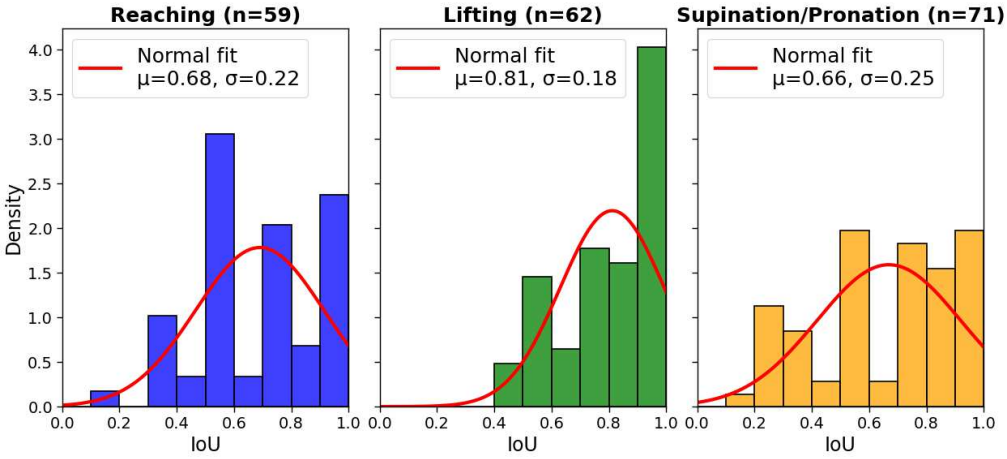


Figure 5: Distribution of IoU values for correctly identified segment types "Reaching" (**R**), "Lifting" (**L**) and "Supination/Pronation" (**S**/**P**), across all subjects' recordings. Each histogram illustrates the IoU values for the corresponding segment type, with an overlaid normal fitting curve.

portion of the ground-truth duration (ratio = 0.60), reflecting more missed segment identifications and lower temporal localization.

Metric	Total GT Segments Duration (Seconds)	Total Predicted Segments Duration (Seconds)	Duration Ratio
R	165.2	99.5	0.60
L	133.5	120.0	0.90
S/P	192.5	210.0	1.09

Table 5: Total ground-truth segments duration, total duration of all predicted segments, and the ratio between predicted and ground-truth durations across all subjects' recordings. "Reaching" is denoted by (**R**), "Lifting" by (**L**), and "Supination/Pronation" by (**S**/**P**). Ratios greater than 1 indicate that the predicted segments cover a longer total duration than the duration of the corresponding ground-truth segment.

4. Discussion

This study aimed to develop a method to identify three upper-extremity (UE) functional movements, – Reaching, Lifting, and Supination/Pronation

–, from IMU data recorded in stroke patients, and to evaluate it in a setting representative of at-home rehabilitation. The proposed deep-learning classification model (CNN+BiLSTM+Attention) demonstrated higher performance in the group-cohort evaluation scenario compared to the subject-specific (baseline) scenario, with strong generalization to unseen participants during free-living measurements. Moreover, integrating the classifier with an HMM–Viterbi decoder for movement segmentation yielded a fully automated framework that accurately identifies movement segments, reliably localizes them in time, and preserves their total duration.

The superior performance of the group-cohort model in free-living measurements indicates that models trained on diverse, multi-subject data provide a more realistic assessment of generalization and are better suited for real-world deployment. In contrast, subject-specific models showed reduced robustness when trained on structured, clinic-like recordings and applied to free-living measurements. This drop is likely due to two factors: (i) changes in environmental context and the resulting variability in movement execution, and (ii) the limited generalization capacity of models trained on a single subject. Previous work suggests that subject-specific models may improve robustness to environmental variability when trained on larger and more diverse datasets from the same individual [58]. However, the limited amount of labeled data typically available from individual patients makes such approaches difficult to implement in practice. Consequently, group-cohort training provides a more feasible alternative by improving robustness to inter-subject and contextual variability, even with a relatively small patient cohort. Although some studies report higher performance for subject-dependent validation (e.g., k-fold), where data from the same individual are used for both training and testing, these approaches are known to overestimate performance compared to subject-independent methods such as Leave-One-Subject-Out (LOSO) [40, 41, 59]. The differences observed in our results are likely related to the distinct training and testing settings used in the subject-specific models.

The performance of the fully automated movement segment identification framework varies across functional movement types, reflecting differences in their inherent variability and detectability. The framework demonstrated high performance in identifying Lifting and Supination/Pronation segments, suggesting that their temporal patterns are relatively consistent and well captured by the classifier combined with HMM-based decoding. In contrast, Reaching segments showed lower performance, primarily due

to reduced recall, indicating that a number of reaching movements were missed. This likely reflects greater variability in reaching movement during free-living measurements, where differences in execution and compensatory strategies (e.g., trunk displacement) can alter expected kinematic patterns [60]. Such variability in post-stroke reaching has been reported in previous studies, where patients with greater impairments exhibit reduced working area, shorter movement distances, and lower vertical hand elevation during reaching tasks in simulated home environments [61]. This highlights the challenges of identifying reaching movements during unsupervised daily activities. Future improvements may involve incorporating contextual information (e.g., object interaction) and additional kinematic features, such as elbow joint angle and trunk displacement, to better capture reaching variability. Additionally, the confusion matrix indicated that most errors arose from missed segments (false negatives) and false positives (i.e., misclassifying unknown movements as functional segments), rather than confusion between movement types. This suggests that the framework generally distinguishes between movement categories but occasionally fails to detect segment onset or misclassifies unknown movements as one of the three functional movements.

With respect to temporal localization, the fully automated framework generally achieved good performance (IoU > 0.5) in approximating movement segment boundaries in free-living measurements, as reflected by the IoU distribution (Figure 5). Temporal alignment was highest for Lifting, while Reaching and Supination/Pronation showed lower overlap and greater variability in onset/offset estimation. The observed onset/offset errors were comparable to those reported in previous ADL segmentation studies using adaptive peak-threshold detection methods, despite differences in activities and populations. These studies included lower-limb activities such as walking, sitting, and standing in Parkinson's disease patients, and reported temporal deviations of approximately 350–500 ms from manual annotations [62, 63]. Duration estimation was more consistent across movements, particularly for Lifting and Supination/Pronation. From an application perspective, accurate duration estimation may be sufficient for many rehabilitation use cases, even when boundary precision is limited.

Our study has a concrete clinical application, namely, identifying stroke patients' UE functional movements in free-living measurements with minimal per-subject training and more realistic generalization on unseen subjects. More specifically, the proposed fully automated framework enables objective

tracking of both the frequency (segment identification) and duration of three functional arm movements performed in daily life, which are key indicators of post-stroke recovery [14, 64, 65]. Therefore, this method shows potential for remotely monitoring arm use in settings representative of at-home rehabilitation. Such capabilities are particularly relevant given the increasing emphasis on home-based rehabilitation and the need to support growing patient populations with limited clinical resources. Future work should assess its feasibility in fully unsupervised, real-home environments, extend the approach to a wider range of functional or transitional UE movements, and support the deployment of such systems as clinically usable tools for long-term patient monitoring.

5. Limitations

Some limitations of this study should be acknowledged. First, although the proposed fully automated movement segment identification framework is designed for real-time deployment in home environments, the present evaluation was conducted in a rehabilitation kitchen within a clinical setting, and model training was performed offline. As such, the findings may not fully reflect performance in real-home environments, where greater contextual and environmental variability is expected. Second, the relatively small sample size and the inclusion of predominantly moderately impaired stroke patients, as reflected by their FMA-UE and ARAT scores, may limit the generalizability of the findings across the full spectrum of post-stroke upper-limb impairment levels. Third, the evaluation of temporal localization (IoU) relied on thresholds adopted from image segmentation due to the lack of established standards for movement segment identification, which may affect the interpretation of what constitutes good or poor performance in this context. Future work should therefore aim to establish movement-specific evaluation criteria for the temporal localization of upper-extremity movements. Fourth, the study focused on a limited set of predefined functional arm movements, which may not fully capture the diversity and complexity of upper-limb use during daily activities.

6. Conclusions

In this study, we proposed a deep learning–based framework for the identification of three UE functional movements in free-living measurements of

stroke patients. By combining a CNN+BiLSTM+Attention classifier with HMM-based sequence decoding, the framework enables both accurate movement type classification and segmentation. The results demonstrate that group-cohort training substantially improves generalization to unseen subjects compared to subject-specific evaluation, while the fully automated framework achieves reliable segment identification, temporal localization, and duration estimation in a setting representative of at-home rehabilitation. These findings highlight the potential of the proposed method for objective, continuous monitoring of upper-limb function, and support its further evaluation in real-home rehabilitation settings.

Declaration of Competing Interest

The authors declare that they have no known competing financial interests or personal relationships that could have appeared to influence the work reported in this paper.

Funding

The study was funded by Erasmus MC TKI-LSH (project number: EMCLSHM22015), with financial support from Rijndam Rehabilitation.

Availability of Data and Materials

The dataset will be made publicly available upon acceptance/publication of the manuscript.

Ethical Approval and Informed Consent

These experiments received approval from the Ethical Committee of Erasmus Medical Center, Rotterdam, The Netherlands (application number: MEC-2024-0640, Towards@HomeRehab) for inclusion of patients. All participants provided written informed consent before enrollment in the respective studies, in accordance with the Declaration of Helsinki.

Acknowledgment

We would like to thank all participants for generously taking part in these experiments and the students who assisted in conducting them.

References

[1] Pollock A, Farmer S E, Brady M C, Langhorne P, Mead G E, Mehrholz J, F. van Wijck, Interventions for improving upper limb function after stroke, Cochrane Database of Systematic Reviews (11) (2014). doi:10.1002/14651858.CD010820.pub2.
URL https://doi.org//10.1002/14651858.CD010820.pub2

[2] P. Langhorne, J. Bernhardt, G. Kwakkel, Stroke rehabilitation, The Lancet 377 (9778) (2011) 1693–1702. doi:10.1016/S0140-6736(11)60325-5.
URL https://doi.org/10.1016/S0140-6736(11)60325-5

[3] M. Kolmos, M. Munoz-Novoa, K. Sunnerhagen, M. Alt Murphy, C. Kruuse, Upper-extremity motor recovery after stroke: A systematic review and meta-analysis of usual care in trials and observational studies, Journal of the Neurological Sciences 468 (2025) 123341. doi:https://doi.org/10.1016/j.jns.2024.123341.
URL https://www.sciencedirect.com/science/article/pii/S0022510X24004775

[4] C. J. Winstein, J. Stein, R. Arena, B. Bates, L. R. Cherney, S. C. Cramer, F. Deruyter, J. J. Eng, B. Fisher, R. L. Harvey, C. E. Lang, M. MacKay-Lyons, K. J. Ottenbacher, S. Pugh, M. J. Reeves, L. G. Richards, W. Stiers, R. D. Zorowitz, Guidelines for Adult Stroke Rehabilitation and Recovery, Stroke 47 (6) (2016) e98–e169. doi:10.1161/STR.0000000000000098.
URL https://www.ahajournals.org/doi/abs/10.1161/STR.0000000000000098

[5] K. Kim, Y. M. Kim, E. K. Kim, Correlation between the Activities of Daily Living of Stroke Patients in a Community Setting and Their Quality of Life, Journal of physical therapy science 26 (3) (2014) 417—419. doi:10.1589/jpts.26.417.
URL https://europepmc.org/articles/PMC3976015

[6] M. Maier, B. R. Ballester, P. F. M. J. Verschure, Principles of Neurorehabilitation After Stroke Based on Motor Learning and Brain Plasticity Mechanisms, Frontiers in Systems Neuroscience Volume 13 - 2019 (2019). doi:10.3389/fnsys.2019.00074.

URL https://www.frontiersin.org/journals/systems-neuroscience/articles/10.3389/fnsys.2019.00074

[7] Veerbeek Janne Marieke, Erwin van Wegen, Roland van Peppen, van der Wees Philip Jan, Rietberg Marc, Hendriks Erik, Kwakkel Gert, What Is the Evidence for Physical Therapy Poststroke? A Systematic Review and Meta-Analysis, PLOS ONE 9 (2) (2014) 1. doi:10.1371/journal.pone.0087987.
URL https://doi.org/10.1371/journal.pone.0087987

[8] C. E. Lang, J. R. MacDonald, C. Gnip, Counting Repetitions: An Observational Study of Outpatient Therapy for People with Hemiparesis Post-Stroke, Journal of Neurologic Physical Therapy 31 (1) (2007).
URL https://journals.lww.com/jnpt/fulltext/2007/03000/counting_repetitions__an_observational_study_of.4.aspx

[9] C. E. Lang, J. R. MacDonald, D. S. Reisman, L. Boyd, T. Jacobson Kimberley, S. M. Schindler-Ivens, T. G. Hornby, S. A. Ross, P. L. Scheets, Observation of Amounts of Movement Practice Provided During Stroke Rehabilitation, Archives of Physical Medicine and Rehabilitation 90 (10) (2009) 1692–1698. doi:https://doi.org/10.1016/j.apmr.2009.04.005.
URL https://www.sciencedirect.com/science/article/pii/S0003999309003530

[10] G. Kwakkel Senior Researcher, Impact of intensity of practice after stroke: Issues for consideration, Disability and Rehabilitation 28 (13-14) (2006) 823–830. doi:10.1080/09638280500534861.
URL https://doi.org/10.1080/09638280500534861

[11] A. Y. Gelaw, B. Janakiraman, B. F. Gebremeskel, H. Ravichandran, Effectiveness of Home-based rehabilitation in improving physical function of persons with Stroke and other physical disability: A systematic review of randomized controlled trials, Journal of Stroke and Cerebrovascular Diseases 29 (6) (2020) 104800. doi:https://doi.org/10.1016/j.jstrokecerebrovasdis.2020.104800.
URL https://www.sciencedirect.com/science/article/pii/S1052305720301841

[12] F. S. Sarfo, U. Ulasavets, O. K. Opare-Sem, B. Ovbiagele, Tele-Rehabilitation after Stroke: An Updated Systematic Review of the Literature, Journal of Stroke and Cerebrovascular Diseases 27 (9) (2018) 2306–2318. doi:https://doi.org/10.1016/j.jstrokecerebrovasdis.2018.05.013.
URL https://www.sciencedirect.com/science/article/pii/S1052305718302313

[13] World Health Organization, How to Use the ICF: A Practical Manual for Using the International Classification of Functioning, Disability and Health (ICF), World Health Organization, Geneva, Switzerland, 2013.

[14] T. Proietti, A. Bandini, Wearable Technologies for Monitoring Upper Extremity Functions During Daily Life in Neurologically Impaired Individuals, IEEE Transactions on Neural Systems and Rehabilitation Engineering 32 (2024) 2737–2748. doi:10.1109/TNSRE.2024.3435042.

[15] J.-H. Kim, S. B. Lee, Evaluation of Activities of Daily Living: Current Insights and Future Horizons, Ann Geriatr Med Res 29 (2) (2025) 143–158. doi:10.4235/agmr.24.0172.
URL https://doi.org/10.4235/agmr.24.0172http://www.e-agmr.org/journal/view.php?number=1185

[16] C. E. Lang, J. M. Wagner, D. F. Edwards, A. W. Dromerick, Upper Extremity Use in People with Hemiparesis in the First Few Weeks After Stroke, Journal of Neurologic Physical Therapy 31 (2) (2007).
URL https://journals.lww.com/jnpt/fulltext/2007/06000/upper_extremity_use_in_people_with_hemiparesis_in.4.aspx

[17] G. Kwakkel, J. M. Veerbeek, E. E. H. van Wegen, S. L. Wolf, Constraint-induced movement therapy after stroke, The Lancet Neurology 14 (2) (2015) 224–234. doi:10.1016/S1474-4422(14)70160-7.
URL https://doi.org/10.1016/S1474-4422(14)70160-7

[18] S. L. Wolf, C. J. Winstein, J. P. Miller, E. Taub, G. Uswatte, D. Morris, C. Giuliani, K. E. Light, D. Nichols-Larsen, f. t. EXCITE Investigators, Effect of Constraint-Induced Movement Therapy on Upper Extremity Function 3 to 9 Months After StrokeThe EXCITE Randomized Clinical Trial, JAMA 296 (17) (2006) 2095–2104. doi:10.1001/jama.296.17.

2095.
URL https://doi.org/10.1001/jama.296.17.2095

[19] A. Langerak, G. Regterschot, M. Evers, B.-J. BEIJNUM, C. Meskers, R. Selles, G. Ribbers, J. Bussmann, A Sensor-Based Feedback Device Stimulating Daily Life Upper Extremity Activity in Stroke Patients: A Feasibility Study, Sensors 23 (5 2023). doi:10.3390/s23135868.

[20] N. Balestra, G. Sharma, L. M. Riek, A. Busza, Automatic Identification of Upper Extremity Rehabilitation Exercise Type and Dose Using Body-Worn Sensors and Machine Learning: A Pilot Study, Digital Biomarkers 5 (2) (2021) 158–166. doi:10.1159/000516619.
URL https://doi.org/10.1159/000516619

[21] Q. Wang, P. Markopoulos, B. Yu, W. Chen, A. Timmermans, Interactive wearable systems for upper body rehabilitation: a systematic review, Journal of NeuroEngineering and Rehabilitation 14 (1) (2017) 20. doi:10.1186/s12984-017-0229-y.
URL https://doi.org/10.1186/s12984-017-0229-y

[22] N. Gebruers, C. Vanroy, S. Truijen, S. Engelborghs, P. P. De Deyn, Monitoring of Physical Activity After Stroke: A Systematic Review of Accelerometry-Based Measures, Archives of Physical Medicine and Rehabilitation 91 (2) (2010) 288–297. doi:10.1016/j.apmr.2009.10.025.
URL https://doi.org/10.1016/j.apmr.2009.10.025

[23] R. R. Bailey, J. W. Klaesner, C. E. Lang, Quantifying Real-World Upper-Limb Activity in Nondisabled Adults and Adults With Chronic Stroke, Neurorehabilitation and Neural Repair 29 (10) (2015) 969–978. doi:10.1177/1545968315583720.
URL https://doi.org/10.1177/1545968315583720

[24] A. M. Valevicius, P. Y. Jun, J. S. Hebert, A. H. Vette, Use of optical motion capture for the analysis of normative upper body kinematics during functional upper limb tasks: A systematic review, Journal of Electromyography and Kinesiology 40 (2018) 1–15. doi:https://doi.org/10.1016/j.jelekin.2018.02.011.
URL https://www.sciencedirect.com/science/article/pii/S1050641117303516

[25] S. P. Marco Iosa Pietro Picerno, G. Morone, Wearable inertial sensors for human movement analysis, Expert Review of Medical Devices 13 (7) (2016) 641–659. doi:10.1080/17434440.2016.1198694.
URL https://doi.org/10.1080/17434440.2016.1198694

[26] S. E. Grace J. Kim Avinash Parnandi, H. Schambra, The use of wearable sensors to assess and treat the upper extremity after stroke: a scoping review, Disability and Rehabilitation 44 (20) (2022) 6119–6138. doi:10.1080/09638288.2021.1957027.
URL https://doi.org/10.1080/09638288.2021.1957027

[27] A. Parnandi, J. Uddin, D. M. Nilsen, H. M. Schambra, The Pragmatic Classification of Upper Extremity Motion in Neurological Patients: A Primer, Frontiers in Neurology 10 (2019). doi:10.3389/fneur.2019.00996.
URL https://www.frontiersin.org/journals/neurology/articles/10.3389/fneur.2019.00996

[28] D. Biswas, A. Cranny, N. Gupta, K. Maharatna, J. Achner, J. Klemke, M. Jöbges, S. Ortmann, Recognizing upper limb movements with wrist worn inertial sensors using k-means clustering classification, Human Movement Science 40 (2015) 59–76. doi:https://doi.org/10.1016/j.humov.2014.11.013.
URL https://www.sciencedirect.com/science/article/pii/S0167945714002115

[29] J. P. Gomez-Arrunategui, J. J. Eng, A. J. Hodgson, Monitoring Arm Movements Post-Stroke for Applications in Rehabilitation and Home Settings, IEEE Transactions on Neural Systems and Rehabilitation Engineering 30 (2022) 2312–2321. doi:10.1109/TNSRE.2022.3197993.

[30] Y. Benachour, M. Rehman, F. Flitti, Towards improved human arm movement analysis: Advanced feature engineering and model optimization, Engineering Applications of Artificial Intelligence 156 (2025) 111194. doi:https://doi.org/10.1016/j.engappai.2025.111194.
URL https://www.sciencedirect.com/science/article/pii/S0952197625011959

[31] M. Panwar, D. Biswas, H. Bajaj, M. Jöbges, R. Turk, K. Maharatna, A. Acharyya, Rehab-Net: Deep Learning Framework for Arm Movement

Classification Using Wearable Sensors for Stroke Rehabilitation, IEEE Transactions on Biomedical Engineering 66 (11) (2019) 3026–3037. doi:10.1109/TBME.2019.2899927.

[32] S. H. Chae, Y. Kim, K.-S. Lee, H.-S. Park, Development and Clinical Evaluation of a Web-Based Upper Limb Home Rehabilitation System Using a Smartwatch and Machine Learning Model for Chronic Stroke Survivors: Prospective Comparative Study, JMIR Mhealth Uhealth 8 (7) (2020) e17216. doi:10.2196/17216.
URL http://www.ncbi.nlm.nih.gov/pubmed/32480361

[33] M. Panwar, S. R. Dyuthi, K. C. Prakash, D. Biswas, A. Acharyya, K. Maharatna, A. Gautam, G. R. Naik, CNN based approach for activity recognition using a wrist-worn accelerometer, in: 2017 39th Annual International Conference of the IEEE Engineering in Medicine and Biology Society (EMBC), 2017, pp. 2438–2441. doi:10.1109/EMBC.2017.8037349.

[34] Y. Chen, Y. Xue, A Deep Learning Approach to Human Activity Recognition Based on Single Accelerometer, in: 2015 IEEE International Conference on Systems, Man, and Cybernetics, 2015, pp. 1488–1492. doi:10.1109/SMC.2015.263.

[35] S. Wang, J. Liu, S. Chen, S. Wang, Y. Peng, C. Liao, L. Liu, Recognizing wearable upper-limb rehabilitation gestures by a hybrid multi-feature neural network, Engineering Applications of Artificial Intelligence 127 (2024) 107424. doi:https://doi.org/10.1016/j.engappai.2023.107424.
URL https://www.sciencedirect.com/science/article/pii/S0952197623016081

[36] F. Jin, M. Zou, X. Peng, H. Lei, Y. Ren, Deep Learning-Enhanced Internet of Things for Activity Recognition in Post-Stroke Rehabilitation, IEEE Journal of Biomedical and Health Informatics 28 (7) (2024) 3851–3859. doi:10.1109/JBHI.2023.3332735.

[37] A. Kaku, A. Parnandi, A. Venkatesan, N. Pandit, H. Schambra, C. Fernandez-Granda, Towards data-driven stroke rehabilitation via wearable sensors and deep learning, in: Machine Learning for Healthcare Conference, PMLR, 2020, pp. 143–171.

[38] A. Murad, J.-Y. Pyun, Deep Recurrent Neural Networks for Human Activity Recognition, Sensors 17 (11) (2017). doi:10.3390/s17112556.
URL https://www.mdpi.com/1424-8220/17/11/2556

[39] P. Sarcevic, Z. Kincses, S. Pletl, Online human movement classification using wrist-worn wireless sensors, Journal of Ambient Intelligence and Humanized Computing 10 (1) (2019) 89–106. doi:10.1007/s12652-017-0606-1.
URL https://doi.org/10.1007/s12652-017-0606-1

[40] S. U. Rehman, A. Ali, A. M. Khan, C. Okpala, Human Activity Recognition: A Comparative Study of Validation Methods and Impact of Feature Extraction in Wearable Sensors, Algorithms 17 (12) (2024). doi:10.3390/a17120556.
URL https://www.mdpi.com/1999-4893/17/12/556

[41] D. Gholamiangonabadi, N. Kiselov, K. Grolinger, Deep Neural Networks for Human Activity Recognition With Wearable Sensors: Leave-One-Subject-Out Cross-Validation for Model Selection, IEEE Access 8 (2020) 133982–133994. doi:10.1109/ACCESS.2020.3010715.

[42] R. Yao, G. Lin, Q. Shi, D. Ranasinghe, Efficient Dense Labeling of Human Activity Sequences from Wearables using Fully Convolutional Networks (2017).
URL https://arxiv.org/abs/1702.06212

[43] Y. Zhang, Y. Zhang, Z. Zhang, J. Bao, Y. Song, Human activity recognition based on time series analysis using U-Net (2018).
URL https://arxiv.org/abs/1809.08113

[44] S. Akbari, H. L. D. Horemans, J. B. J. Bussmann, A. Zgonnikov, Evaluating generalization of arm movement identification using machine learning: From structured to semi-structured environments, Computers in Biology and Medicine 198 (2025) 111167. doi:https://doi.org/10.1016/j.compbiomed.2025.111167.
URL https://www.sciencedirect.com/science/article/pii/S0010482525015203

[45] M. A. Khatun, M. A. Yousuf, S. Ahmed, M. Z. Uddin, S. A. Alyami, S. Al-Ashhab, H. F. Akhdar, A. Khan, A. Azad, M. A. Moni, Deep

CNN-LSTM With Self-Attention Model for Human Activity Recognition Using Wearable Sensor, IEEE Journal of Translational Engineering in Health and Medicine 10 (2022) 1–16. doi:10.1109/JTEHM.2022.3177710.

[46] 2M Engineering, 9DOF Motion Tracking Sensors (2024).
URL https://www.2mel.nl/wearable-product-development/motion-tracking-sensors

[47] Y.-C. P. Pan, B. Goodwin, E. Sabelhaus, K. M. Peters, K. F. Bjornson, K. L. D. Pham, W. Walker, K. M. Steele, Feasibility of using acceleration-derived jerk to quantify bimanual arm use, Journal of NeuroEngineering and Rehabilitation 17 (1) (2020) 44. doi:10.1186/s12984-020-0653-2.
URL https://doi.org/10.1186/s12984-020-0653-2

[48] A. Favata, R. Gallart-Agut, R. Pàmies-Vilà, C. Torras, J. M. Font-Llagunes, IMU-Based Systems for Upper-Limb Kinematic Analysis in Clinical Applications: A Systematic Review, IEEE Sensors Journal 24 (18) (2024) 28576–28594. doi:10.1109/JSEN.2024.3436532.

[49] M. Lepot, J.-B. Aubin, F. H. L. R. Clemens, Interpolation in Time Series: An Introductive Overview of Existing Methods, Their Performance Criteria and Uncertainty Assessment, Water 9 (10) (2017). doi:10.3390/w9100796.
URL https://www.mdpi.com/2073-4441/9/10/796

[50] D. P. Kingma, Adam: A method for stochastic optimization, arXiv preprint arXiv:1412.6980 (2014).

[51] M. Ramadan, A. El-Jaroudi, Action detection and classification in kitchen activities videos using graph decoding, The Visual Computer 39 (3) (2023) 799–812. doi:10.1007/s00371-021-02346-5.
URL https://doi.org/10.1007/s00371-021-02346-5

[52] L. Trinh, B. Ha, An incorporation of deep temporal convolutional networks with hidden markov models post-processing for sensor-based human activity recognition, in: Proceedings of the 11th International Symposium on Information and Communication Technology, SoICT '22, Association for Computing Machinery, New York, NY, USA, 2022, pp.

96–102. doi:10.1145/3568562.3568610.
URL https://doi.org/10.1145/3568562.3568610

[53] L. R. Rabiner, A tutorial on hidden Markov models and selected applications in speech recognition, Proceedings of the IEEE 77 (2) (1989) 257–286. doi:10.1109/5.18626.

[54] M. Alanazi, R. S. Aldahr, M. Ilyas, Human Activity Recognition through Smartphone Inertial Sensors with ML Approach, Engineering, Technology & Applied Science Research (2024).
URL https://api.semanticscholar.org/CorpusID:267642997

[55] T. Stampfler, M. Elgendi, R. R. Fletcher, C. Menon, The use of deep learning for smartphone-based human activity recognition, Frontiers in Public Health Volume 11 - 2023 (2023). doi:10.3389/fpubh.2023.1086671.
URL https://www.frontiersin.org/journals/public-health/articles/10.3389/fpubh.2023.1086671

[56] M. A. Hossen, P. E. Abas, Machine Learning for Human Activity Recognition: State-of-the-Art Techniques and Emerging Trends, Journal of Imaging 11 (3) (2025). doi:10.3390/jimaging11030091.
URL https://www.mdpi.com/2313-433X/11/3/91

[57] A. N. Gajjar, J. Jethva, Intersection over Union based analysis of Image detection/segmentation using CNN model, in: 2022 Second International Conference on Power, Control and Computing Technologies (ICPC2T), 2022, pp. 1–6. doi:10.1109/ICPC2T53885.2022.9776896.

[58] A. Bulling, U. Blanke, B. Schiele, A tutorial on human activity recognition using body-worn inertial sensors, ACM Comput. Surv. 46 (3) (1 2014). doi:10.1145/2499621.
URL https://doi.org/10.1145/2499621

[59] S. Mekruksavanich, N. Hnoohom, W. Phaphan, A. Jitpattanakul, A Comparative Study of Validation Methods for Sensor-Based Human Activity Recognition Using Deep Learning Models, in: 2025 Joint International Conference on Digital Arts, Media and Technology with ECTI Northern Section Conference on Electrical, Electronics, Computer and

Telecommunications Engineering (ECTI DAMT & NCON), 2025, pp. 686–691. doi:10.1109/ECTIDAMTNCON64748.2025.10962074.

[60] S. K. Subramanian, J. Yamanaka, G. Chilingaryan, M. F. Levin, Validity of Movement Pattern Kinematics as Measures of Arm Motor Impairment Poststroke, Stroke 41 (10) (2010) 2303–2308. doi:10.1161/STROKEAHA.110.593368.
URL https://www.ahajournals.org/doi/abs/10.1161/STROKEAHA.110.593368

[61] F. B. van Meulen, J. Reenalda, J. H. Buurke, P. H. Veltink, Assessment of Daily-Life Reaching Performance After Stroke, Annals of Biomedical Engineering 43 (2) (2015) 478–486. doi:10.1007/s10439-014-1198-y.
URL https://doi.org/10.1007/s10439-014-1198-y

[62] H. Nguyen, K. Lebel, S. Bogard, E. Goubault, P. Boissy, C. Duval, Using Inertial Sensors to Automatically Detect and Segment Activities of Daily Living in People With Parkinson's Disease, IEEE Transactions on Neural Systems and Rehabilitation Engineering 26 (1) (2018) 197–204. doi:10.1109/TNSRE.2017.2745418.

[63] H. Nguyen, K. Lebel, P. Boissy, S. Bogard, E. Goubault, C. Duval, Auto detection and segmentation of daily living activities during a Timed Up and Go task in people with Parkinson's disease using multiple inertial sensors, Journal of NeuroEngineering and Rehabilitation 14 (1) (2017) 26. doi:10.1186/s12984-017-0241-2.
URL https://doi.org/10.1186/s12984-017-0241-2

[64] G. R. H. Regterschot, J. B. J. Bussmann, M. H. J. Fanchamps, C. G. M. Meskers, G. M. Ribbers, R. W. Selles, Objectively measured arm use in daily life improves during the first 6 months poststroke: a longitudinal observational cohort study, Journal of NeuroEngineering and Rehabilitation 18 (1) (2021) 51. doi:10.1186/s12984-021-00847-x.
URL https://doi.org/10.1186/s12984-021-00847-x

[65] J. Pohl, G. Verheyden, J. P. O. Held, A. R. Luft, C. Easthope Awai, J. M. Veerbeek, Construct validity and responsiveness of clinical upper limb measures and sensor-based arm use within the first year after stroke: a longitudinal cohort study, Journal of NeuroEngineering and

Rehabilitation 22 (1) (2025) 14. doi:10.1186/s12984-024-01512-9.
URL https://doi.org/10.1186/s12984-024-01512-9