

Empowering stroke recovery with upper limb rehabilitation monitoring using TinyML based heterogeneous classifiers

Article

Creative Commons: Attribution-Noncommercial-No Derivative Works 4.0

Open Access

Xie, J., Wu, Q., Dey, N., Shi, F., Sherratt, R. S. ORCID: <https://orcid.org/0000-0001-7899-4445> and Kuang, Y. (2025) Empowering stroke recovery with upper limb rehabilitation monitoring using TinyML based heterogeneous classifiers. *Scientific Reports*, 15. 18090. ISSN 2045-2322 doi: 10.1038/s41598-025-01710-y Available at <https://centaur.reading.ac.uk/122851/>

It is advisable to refer to the publisher's version if you intend to cite from the work. See [Guidance on citing](#).

To link to this article DOI: <http://dx.doi.org/10.1038/s41598-025-01710-y>

Publisher: Nature Publishing Group

All outputs in CentAUR are protected by Intellectual Property Rights law, including copyright law. Copyright and IPR is retained by the creators or other copyright holders. Terms and conditions for use of this material are defined in the [End User Agreement](#).

www.reading.ac.uk/centaur

CentAUR

Central Archive at the University of Reading

Reading's research outputs online



OPEN

Empowering stroke recovery with upper limb rehabilitation monitoring using TinyML based heterogeneous classifiers

Jiayu Xie¹, Qun Wu¹, Nilanjan Dey², Fuqian Shi³, R. Simon Sherratt⁴ & Yuxiang Kuang⁵✉

Stroke is one of the leading causes of disability worldwide, with approximately 70% of survivors experiencing motor impairments in the upper limbs, significantly affecting their quality of life. Home-based rehabilitation offers a cost-effective approach to improving motor function, but it faces challenges, including inaccurate movement reporting, lack of real-time feedback, and the high cost of rehabilitation equipment. Therefore, there is a need for affordable, lightweight home-based rehabilitation monitoring systems. This paper presents an intelligent wearable sensor system that utilizes TinyML AI technology to classify eight upper limb rehabilitation movements with minimal sensors. The system is designed for patients with upper limb impairments who retain antigravity voluntary movement, enabling them to monitor rehabilitation progress at home. The study recruited 10 healthy volunteers to perform rehabilitation movements, creating a standardized dataset for model training. Data normalization, preprocessing, model training, and deployment were carried out using the Edge Impulse platform. A hybrid classifier, combining multilayer perceptron and k-means clustering, achieved 96.1% training accuracy, 95.09% testing accuracy, and 88.01% deployment accuracy. The proposed TinyML-based system shows promising potential for home-based rehabilitation of stroke patients.

Keywords TinyML, Wearable devices, Stroke, Home rehabilitation, Inertial motion capture

Cerebral apoplexy, commonly referred to as stroke, is one of the principal aetiologies of enduring disabilities worldwide¹. Statistical analyses indicated that approximately 70% of individuals who recovered from stroke experienced motor dysfunctions after discharge from medical facilities, with upper limb hemiparesis being the most prevalent complication. This impairment significantly hampers the ability of stroke survivors to perform activities of daily living (ADL)^{2,3}, thereby imposing profound psychological and financial distress upon patients and their kin.

Empirical evidence suggests that prolonged and consistent rehabilitation training can significantly ameliorate motor functional impairments resulting from stroke^{4,5}. However, the vast majority of stroke survivors lack the resources necessary for sustained clinical rehabilitation, and the rehabilitation process is hampered by factors such as time, distance, costs, the scarcity of specialists, and limited clinical facilities^{6–9}. Studies have indicated that home-based rehabilitation represents a more sustainable model of recovery, facilitating more effective promotion of long-term rehabilitation training for patients and transitioning care services towards a decentralized model^{10–12}, which is divergent from traditional clinical therapeutic approaches^{13,14}. Compared with hospital-based rehabilitation, home rehabilitation offers greater flexibility and autonomy, affording patients substantial amounts of time for self-directed training, and home-based recovery has been shown to reduce rehabilitation healthcare costs by up to 15%¹⁵. Nonetheless, for patients receiving rehabilitation therapy within the home environment, accurately assessing the quality and quantity of rehabilitation movements poses a significant challenge. Research has documented patients' difficulties in precisely reporting the quantity and quality of exercises performed^{16–18}, as well as their struggles to adhere to rehabilitation training over extended

¹Zhejiang Sci-Tech University, Hangzhou 310029, Zhejiang, China. ²Department of Computer Science and Engineering, Techno International New Town, Kolkata 700156, India. ³Laura and Isaac Perlmutter Cancer Center, New York University Langone Health, New York, NY 10003, USA. ⁴Department of Biomedical Engineering, the University of Reading, Berkshire RG6 6AY, UK. ⁵Jiangxi University of Finance and Economics, Nanchang 330013, China. ✉email: kyxiang@jxufe.edu.cn

periods¹⁶. Consequently, the classification of rehabilitation movements has emerged as a pressing issue that needs addressing in the context of home-based rehabilitation.

Rapid advancements in artificial intelligence (AI) technologies are pivotal in supporting decentralized care models. Specifically, the development of unobtrusive motion capture technologies to establish simple, safe, effective, and objective systems for movement classification has become a strategic application in supporting remote home-based rehabilitation models^{4,19}. Traditional systems for movement classification include (1) camera-based systems^{20–22}, among which depth cameras, exemplified by Kinect²³, have achieved significant milestones in the domain of rehabilitation motion recognition^{24–27}. However, camera-based solutions necessitate ample, unobstructed observational space to monitor participants, a requirement that is challenging to meet outside laboratory environments^{28,29}. Moreover, the deployment of visual recognition technologies within domestic settings may trigger privacy invasion concerns^{30,31}. (2) Wearable-based inertial sensor systems, which employ either single or multiple inertial measurement units (IMUs), facilitate the determination of each limb's position by resolving the corresponding kinematic equations^{32–34}.

Inertial motion analysis has attracted increasing interest in recent decades because of its advantages over classical optical systems³⁵. Owing to their integrative capabilities, low cost, and ease of implementation, IMUs are extensively employed for monitoring rehabilitation exercises within home settings^{36–39}. Table 1 provides a detailed presentation of the relevant works associated with this study, including achievements in movement classification using either a single or multiple sensors. For example, Zinnen et al.⁴⁰ reported a classification accuracy of 93% for 20 movements via a five-device system and an accuracy of 86% with two wrist devices, covering actions such as manipulating various car doors and a writing task. Zhang et al.⁴¹ introduced a two-device framework that achieved 97.2% accuracy in classifying four arm movements. Lui and Menon³³, Wang et al.⁴², Alessandri et al.⁴³, and Li et al.⁴⁴ all rely on multiple sensors for classification. Similarly, studies by Basterretxea et al.⁴⁵, Xu and Yuan⁴⁶, and Choudhury et al.⁴⁷ also demonstrate high accuracy, with substantial diversity in the classified movements. However, as shown in the relevant studies listed in Table 1, few studies have achieved high-precision motion classification using a single sensor. Multi-device systems are generally more complex than single-device setups are, particularly when devices are affixed to multiple limbs or when interdevice calibration is needed. Whenever possible, a single-device system is preferred due to its ease of wearability and configuration³³. The rapid development of machine learning algorithms has made it possible to achieve rehabilitation movement classification using a single inertial sensor. For example, Zhang et al.⁴⁸

Author	Sensors	Power consumption	Activities	Performance metrics
Xie et al. (The present study)	An accelerometer is utilized	The peak power consumption is 54.05 mW	Eight types of upper limb rehabilitation movements are analysed	The testing accuracy achieved is 95.09%, with a final deployment accuracy of 88.01%
Lui and C. Menon. ³³	One wrist-worn inertial device (MPU9250) and an Omron D6T thermal sensor	-	Twenty-four types of upper limb movements	The classification accuracy for the 24 upper limb movements reaches 93.55%
Giordano, et al. ³²	One accelerometer	Peak power consumption is approximately 8 mA. 8 mA represents the operating current of the device during operation (the common voltage of NRF52832 is 3 V, and the corresponding power consumption is 24mW)	four different usage classes (Tool transportation, no-load, metal, and wood drilling)	achieving an accuracy of 90.6% with a model size of approximately 30 kB
Wang et al. ⁴²	The upper limb wearable rehabilitation training data collection device is composed of three inertial sensor units, MPU6050	-	There are eight types of upper limb rehabilitation exercises	The CNN-LSTM model demonstrates the best performance with an identification accuracy of 99.67%, followed by the multilayer LSTM model with an identification accuracy of 97.00%
Basterretxea et al. ⁴⁵	The device utilizes accelerometers and stretch sensors to capture data	1.13 mW sensing and 11.24 mW computation power consumption	It can distinguish between various activities such as walking, sitting, standing, driving, lying down, jumping, and the transitions between these states	With a power consumption of less than 12.5 mW, the system achieves an accuracy of 97.7% in recognizing six activities and their transitions
Alessandri et al. ⁴³	photoplethysmogram (PPG) sensors and accelerometers	-	the device can differentiate between walking, ascending stairs, descending stairs, sitting, standing, and lying down	Average test time is 150.1 at the highest test accuracy of 95.54% on the microcontroller unit (MCU)
Xu and Yuan ⁴⁶	The accelerometers are worn on the mid-waist and front of the ankle	-	The device can distinguish between nine types of activities: walking, running, ascending stairs, descending stairs, resting, standing, sitting-to-standing, standing-to-squatting, and squatting-to-standing	The average recognition rate for these nine activities is 97.26%, with an identification rate of 98.56% for the four transition movements
Choudhury et al. ⁴⁷	The device employs accelerometers, microphones, and light sensors	Power consumption is approximately 43 mW (with all sensors running continuously)	It can differentiate between activities such as walking, running, cycling, using an elliptical machine, and using a stair machine	Highest Accuracy is 93.8% (Inference v2.0 with temporal information included and supervised learning)
Li et al. ⁴⁴	Two IMU modules (MPU9250), each comprising a tri-axis accelerometer, a tri-axis gyroscope, and a tri-axis magnetometer	-	Five types of upper limb rehabilitation exercises: forearm pronation & supination, lumbar touch, shoulder touch, shoulder anteflexion, and shoulder extension	The overall recognition accuracy for five motions achieves 99.34%

Table 1. Comparative analysis of related work.

achieved an accuracy of 96.1% for nine activities among healthy participants, including actions such as walking, running, sitting, and standing. In another study, Khan et al.⁴⁹ achieved a precision of 97.9% using chest-mounted devices for 15 movements across various activities (e.g., lying, sitting, standing, and transitions among these postures). The rapid advancement of machine learning algorithms has facilitated the implementation of rehabilitation movement classification via a single inertial sensor. Tseng et al.⁵⁰ reported a single-device system that achieved a classification accuracy of 93.3% for three movements associated with door opening in a study involving five healthy participants. Zhang et al.⁵¹ presented a system employing a solitary IMU for discerning six-arm manoeuvres among fourteen stroke subjects, with classification accuracies of 99.4% with an innovative algorithm and 98.8% when an SVM framework was used. These manoeuvres, which encapsulate four of the human arms' seven degrees of freedom, include flexion and extension at the shoulder joint, signifying the arm's forward and backwards motion; abduction and adduction at the same joint, indicating lateral movement away from and towards the torso; pronation and supination of the forearm, involving rotational movement to orient the palm downwards and upwards, respectively; and flexion and extension at the elbow joint, denoting the arm's bending and straightening. This delineation underscores the integral relationship of movements with specific joint actions, which are crucial for comprehensive upper limb rehabilitation in stroke recovery.

However, the aforementioned wearable sensor-based rehabilitation systems are still often categorized as traditional cloud computing solutions due to the limited computational capabilities of wearable devices, which are often inadequate for extracting useful information from collected data^{31,52}. The data gathered by the sensors are typically streamed by a node and transmitted to an external device via a wireless network for subsequent processing in the cloud. These cloud-based processing methodologies inevitably increase the energy requirements for wearable device communication, development costs, and system latency and introduce security and privacy concerns. Consequently, most of the current research on wearable sensor-based rehabilitation systems is conducted within a clinical context and primarily serves to aid clinicians in rehabilitation centers.

TinyML, a burgeoning and relatively nascent technology within the AI domain^{53,54}, is attracting considerable interest across academic and industrial landscapes. This technology represents a confluence of two distinct paradigms, machine learning (ML) and the Internet of Things (IoT), promulgating the deployment of optimized ML models in devices constrained by computational resources, notably, microcontroller units (MCUs) powered by small batteries^{53,55}. Notably, Prado et al.⁵⁶ employed TinyML techniques to enhance the autonomous driving capabilities of microvehicles. Lahade et al.⁵⁷ introduced a TinyML methodology to quantify the outputs of alcohol sensors on MCUs and to calibrate the sensor responses locally. Piątkowski and Walkowiak⁵⁸ devised a strategy employing TinyML techniques to ascertain the correct wearing of masks, achieving remarkably low power consumption and high accuracy. Avellaneda et al.²² presented a TinyML-based deep learning strategy for indoor asset tracking, with the system attaining a classification accuracy of 88%. In essence, TinyML can address the challenges associated with cloud-based solutions, such as potential privacy infringement risks, response latency, additional network load, and significant power consumption⁵⁴, and synergizes with algorithms capable of executing intricate tasks. This facilitates local inference on wearable devices with severely limited computational ability, bypassing the elaborate cloud-based process and heralding the commercialization of wearable devices in domestic rehabilitation contexts.

This paper describes the design, execution, and appraisal of a domiciliary upper limb rehabilitation framework leveraging TinyML technology aimed at furnishing stroke survivors with an inconspicuous, economically viable home-based upper limb rehabilitation system equipped with instantaneous feedback mechanisms. The architecture of the system is bifurcated into hardware and software segments. Notably, the hardware infrastructure is orchestrated around a microchip equipped with Bluetooth low-energy (BLE) and an integrated IMU system, rendering it highly suitable for the procurement and scrutiny of inertial datasets. The microchip can conduct sophisticated movement classifications locally in an energy-efficient manner while dispensing real-time feedback to the user via an LED light ring and a vibrating motor module. The software is engineered via the MIT APP inventor framework, establishing a wireless conduit with hardware through BLE, thereby facilitating efficacious directional and feedback communication to the user. To our knowledge, no research papers have been published, nor have any devices developed that have utilized TinyML, as used in this paper, for the task of locally analysing upper limb rehabilitation movements.

The main contributions of this paper are summarized as follows:

- (1) A comprehensive architecture is proposed, implemented, and evaluated for the recognition of upper limb rehabilitation actions, leveraging acceleration data to facilitate this process.
- (2) A multilayer perceptron (MLP) neural network model is presented to ensure the efficacious and precise identification of activities pertinent to upper limb rehabilitation.
- (3) Substantial optimization of the model size was achieved, subsequently enabling its integration into an MCU to facilitate ultralow power operation.
- (4) A flexible lightweight wearable device is presented to discern rehabilitation actions within the upper limb.

The rest of the paper is structured as follows. Section "Materials and methods" describes the methodology used to build the home upper limb rehabilitation system. Section "Results" describes the experiments conducted to evaluate the system and the results. Section "Discussion" discusses the results obtained. Finally, Section "Conclusion" concludes the paper and suggests future work.

Materials and methods

This section delineates the proposed architecture of a home-based upper limb rehabilitation system leveraging TinyML, along with the corresponding design and implementation process. Figure 1 presents a representation of the proposed system.

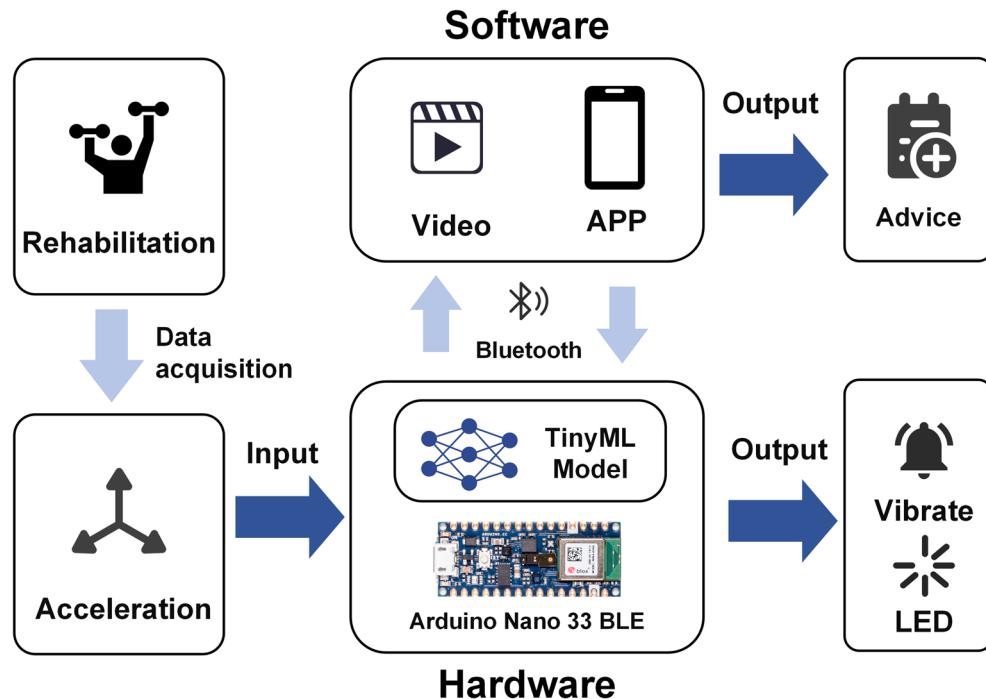


Fig. 1. Schematic representation of the system.

Microcontroller	Price	CPU, Speed	Power Consumption	Memory	Size
SparkFun Edge (Apollo3 Blue)	\$16.50	ARM Cortex-M4F, 48 MHz	6uA/MHz	384 KB SRAM, 1 MB FLASH	40.6 mm*40.6 mm
Arduino Nano 33 BLE Sense Rev2	\$41.90	ARM Cortex-M4F, 64 MHz	14.0625uA/MHz	256 KB SRAM, 1 MB FLASH	18 mm*45 mm
STMicroelectronics STM32F746G	\$55.13	ARM Cortex M7, 216 MHz	0.463 mA/MHz	340 KB SRAM, 1 MB FLASH	80 mm*130 mm

Table 2. Comparison of selected CPUs.

Hardware design

Selecting the appropriate deployment device constitutes a pivotal factor influencing the overall design of the TinyML system. The principal characteristics to consider include cost, speed, power consumption, memory size, and product dimensions. Warden and Situnayake⁵³ recommended three representative microcontrollers that support the TinyML framework, as illustrated in Table 2, and a detailed examination of these three microcontrollers was conducted. Among the three MCUs, the STMicroelectronics STM32F746G exhibits a marked superiority in performance; however, it presents critical drawbacks in terms of price, size, and power consumption. Although the SparkFun Edge (Apollo3 Blue) possesses certain advantages over the Arduino Nano 33 BLE Sense Rev2 in terms of price and power efficiency, the dimensions are of paramount importance considering the aim of this study to provide an unobtrusive home rehabilitation system. Hence, the Arduino Nano 33 BLE Sense Rev2 was selected as the microcontroller for this research. It employs the Nordic Semiconductor nRF52840 microcontroller as its primary processing unit, which is sufficiently rapid for the computational tasks of this study and relatively low in power consumption. The Arduino Nano 33 BLE Sense Rev2 also features an onboard 9-axis IMU system, comprising a 6-axis BMI270 and a 3-axis BMM150, offering adequate precision for measuring acceleration, rotation, and magnetic fields in 3D space. Importantly, within this study, the detection of upper limb rehabilitation movements relies solely on accelerometer data, which measure the vector of the gravitational field and the linear acceleration produced by forces acting on the sensor across each of its three mutually orthogonal sensor axes. (<https://x-io.co.uk/open-source-imu-and-ahrs-algorithms>) The rationale for this selection is that accelerometers consume significantly less power than gyroscopes do (microwatts versus milliwatts)⁵⁹, making them more suitable for prolonged continuous monitoring. Previous research has corroborated the feasibility of using accelerometer data to assess upper limb movements accurately^{60–62}. In this study, we used the Arduino Nano 33 BLE Sense Rev2 chip paired with the Edge Impulse platform. This platform offers ready-made, open-source firmware designed for this specific development board, available on GitHub at [edge impulse/firmware-arduino-nano33-ble-sense](https://github.com/edge-impulse/edge-impulse-firmware-arduino-nano33-ble-sense). By using this firmware, we streamlined our development process, avoiding the need to build firmware from scratch. This approach not only saved substantial development time but also allowed us to focus more on the fundamental aspects of our research. Figure 2 presents the details of the hardware system, where Fig. 2a represents the schematic diagram of the device wearing scenario. Figure 2b

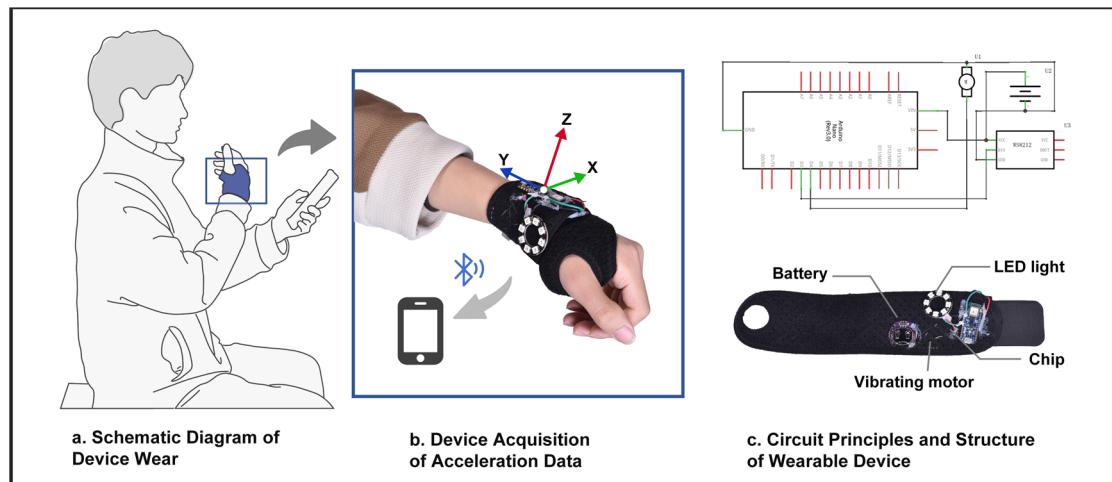


Fig. 2. Details of the Hardware System: (a) Wearable Device Wearing Scenario. (b) Device Acceleration Data Collection. (c) Wearable Device Circuit Principles and Structural Layout.

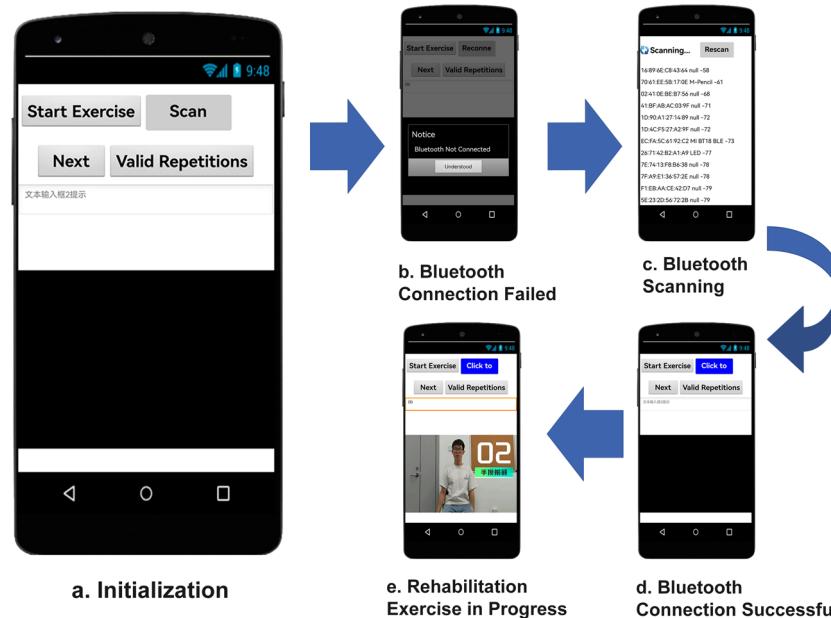


Fig. 3. Illustrates the interface interactions of the mobile application: (a) The initialization interface. (b) Alerting the user that the app has not yet established a BLE connection with the hardware device. (c) Scanning for nearby BLE devices. (d) Successful BLE connection with the hardware device. (e) The interface after initiating the rehabilitation exercises.

indicates the acceleration data collection of the wearable device. Figure 2c illustrates the circuit principles and structural layout of the wearable device.

Software design

A mobile application (app) compatible with the hardware device was developed on the MIT App Inventor platform⁶³ and was designed to facilitate more intuitive rehabilitation training for patients. In Fig. 3, a) illustrates the application in its initialization state, which prompts the user to establish a BLE connection with the corresponding hardware device; In Fig. 3, b) presents how the user is alerted when the app has not yet been connected to the hardware device; In Fig. 3, c) displays the BLE scanning status page, where users select the desired device for connection on the basis of the hardware device's name; In Fig. 3, d) indicates successful device connection, allowing users to initiate rehabilitation training by pressing the "Start Exercise" button on the interface; and In Fig. 3,e) the app guides users through rehabilitation exercises via video. By tapping the button labelled "Effective Repetitions", users can view the number of exercise repetitions marked as effective by the

app. The app provides instructional videos for eight upper limb rehabilitation exercises (these exercises will be described in Section "Materials and methods").

The interactive logic between the app and the hardware device is depicted in Fig. 4, where the app functions as the central device (server) and the hardware device operates as the peripheral device (client). Upon the activation of the physical switch on the hardware device, the periphery initiates BLE initialization to ascertain the successful activation of the BLE module. Following successful activation, the periphery repetitively broadcasts BLE advertisements, and upon acquiring the advertisement data, the central device attempts to establish a connection with the node. After connection establishment, the central device interacts with the accessible information held by the node, a process predicated on services, each of which is further delineated into characteristics. In this study, two characteristics were instantiated for the BLE service of the hardware device: one facilitating read and write operations pertinent to action characteristics and another designated for notifications related to score characteristics. After the establishment of a BLE connection between the server and the client, information is included in the action characteristic, which conveys the current rehabilitation action ID to the client. Following the acquisition of acceleration data, the client executes local inference through the TinyML model. If the classification result aligns with the rehabilitation action ID inscribed by the server, it denotes the correctness of the patient's rehabilitation action, prompting the operation of the LED ring and vibration motor module on the hardware device to furnish feedback to the patient. Moreover, the score characteristic designated for notifications apprises the central device of the count of effective actions.

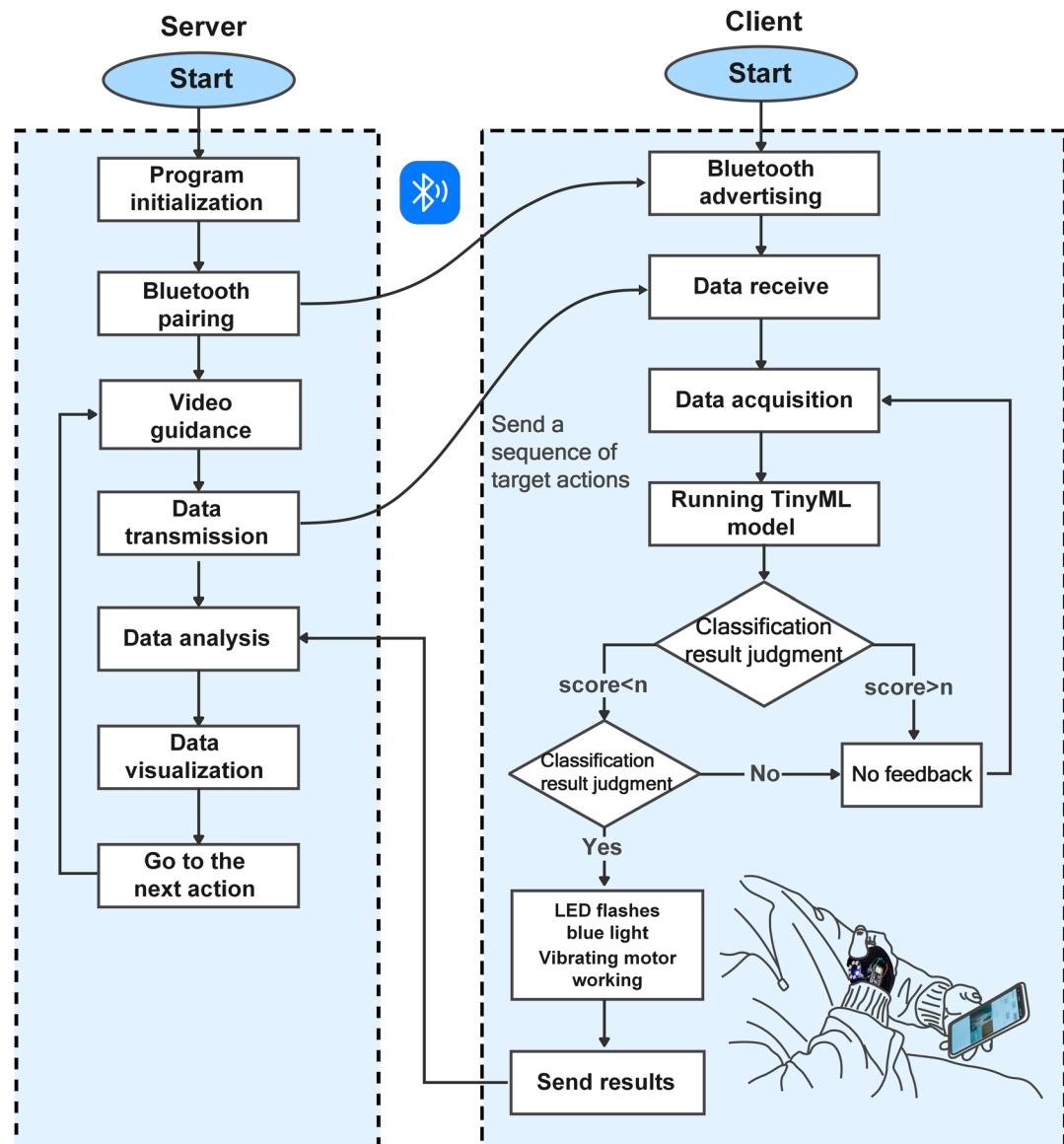


Fig. 4. System interaction logic diagram. The "score" represents the distance between a new data point and its nearest centroid, where "n" denotes the threshold.

Rehabilitation movements

The Brunnstrom approach constitutes a comprehensive therapeutic methodology tailored for the motor recovery of stroke patients⁶⁴. The Brunnstrom theory divides the restoration of limb functionality into six distinct phases, within which patients acquire the ability to perform active exercises against gravity during phases 4 to 6, facilitating advanced rehabilitative training within a home environment. In accordance with the Brunnstrom approach to rehabilitation, eight upper limb rehabilitation exercises were identified for this study, as illustrated in Fig. 5. The selection of these exercises is attributed to their prevalence in hospital settings, their familiarity among stroke patients, and the comprehensive nature of the training content they offer, encompassing various aspects of motor impairments, including multijoint flexibility, muscle strength, and spasticity. Notably, exercises 1–5 are categorized as upper limb flexion movement training, whereas exercises 6–8 focus on upper limb extension movement training.

Data acquisition and analysis

This study focused on upper limb movements in poststroke rehabilitation training and included a total of ten physically healthy volunteers rather than actual stroke patients. The cohort comprised six males and four females, with an average weight of 59.2 ± 7.56 kg and height of 170 ± 7.32 cm; the mean age was 23.9 years (ranging from 22 to 26 years). Participants reporting any prior musculoskeletal disorders, pain, or discomfort were excluded. Additionally, anthropometric data (height, weight, sex, and age) were collected from the participants, all of whom signed a written informed consent.

The device utilized for data acquisition affixed to the participant's dominant hand at the wrist via Velcro is illustrated in Fig. 6. Data collection and labelling were conducted via the data acquisition tools provided by the Edge Impulse platform⁶⁵. Previous studies reported that a signal data frequency of 10 Hz is sufficient for detecting arm movements⁶⁶. The default sampling rate of the Arduino Nano 33 BLE Rev2 chip is 119 Hz; however, we adjusted the sampling rate to 50 Hz (resulting in an actual effective sampling rate of 48.8 Hz) through downsampling, which meets the sampling precision required for this study.

As depicted in Fig. 6, participants are seated on a stool without armrests, maintaining an upright posture. The participants completed the eight rehabilitation exercises per the instructions displayed on a computer. Upon completion of each exercise, the arm is allowed to fall naturally to a vertical resting position before the exercise is repeated. This methodology is utilized to collect data on the fundamental movement units of the eight rehabilitation exercises during the upper limb motion process. Each posture was repeated 20 times for data collection, with each collection lasting 5 s, culminating in a total of 141 min and 5 s of raw accelerometer data, which were uploaded to the Edge Impulse platform in JSON format. The participants were afforded appropriate rest periods after the collection of each action. During the data acquisition process, participants were instructed to avoid contact with surrounding objects to prevent excessive acceleration and to refrain from making other significant body movements that could affect data collection. Notably, the data collection was conducted via a cable connected to a computer.

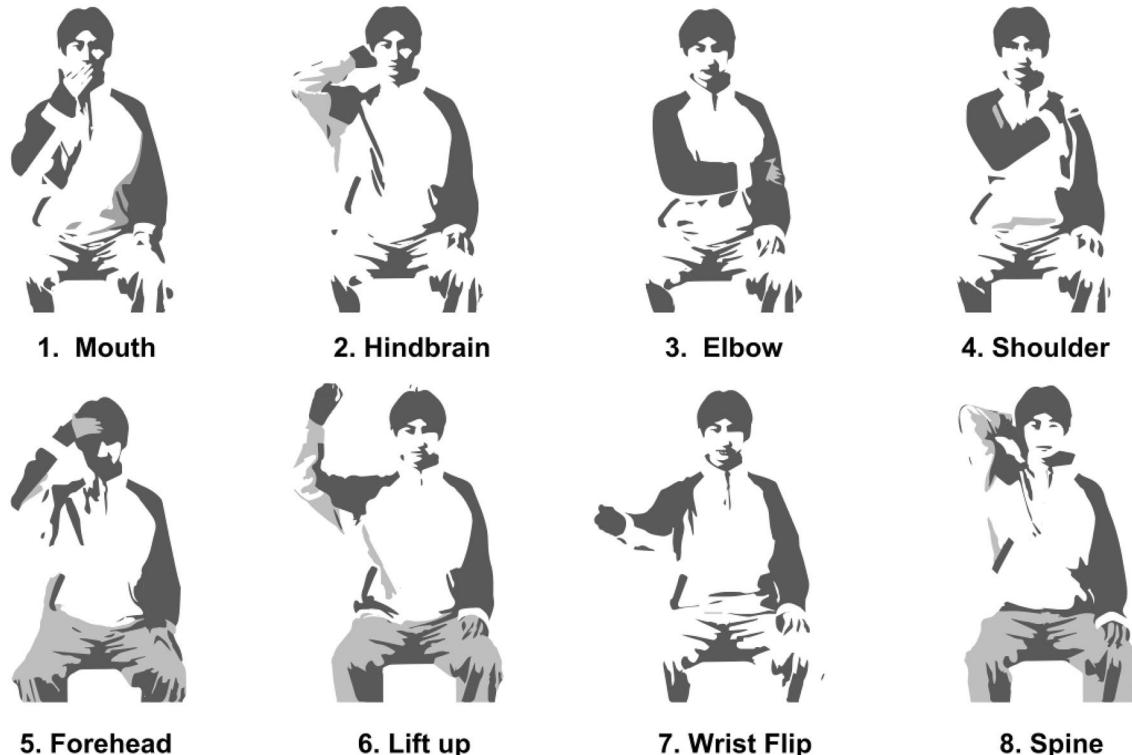


Fig. 5. Schematic depiction of eight rehabilitation assessment exercises based on the Brunnstrom approach.

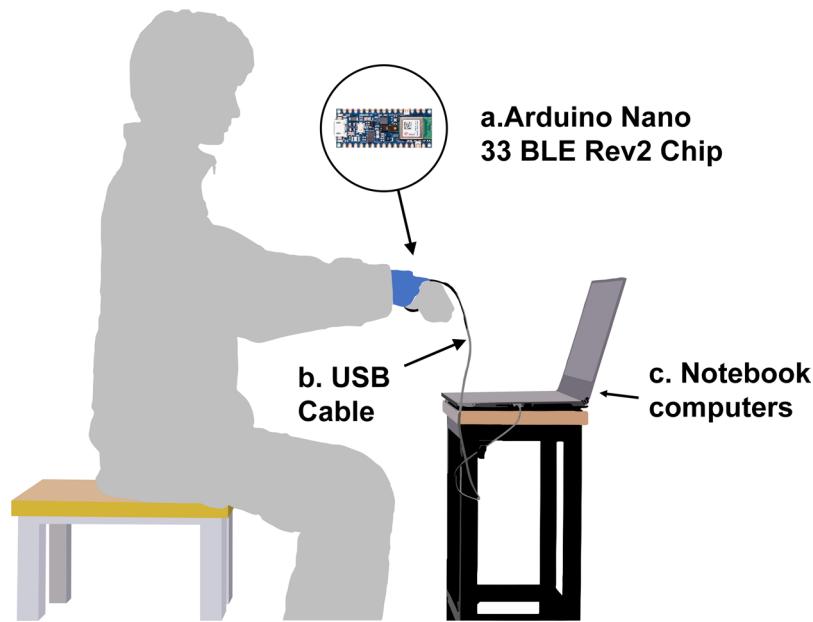


Fig. 6. Schematic of the data collection process: (a) The chip utilized for data acquisition. (b) The USB cable employed for data transmission. (c) Notebook computer.

Window size (ms)	Stride (ms)	Stride: WindowSize Ratio (%)	Accuracy (%)	Loss
4000	1000	25	92.2	0.45
3600	900	25	87.0	0.58
3200	800	25	84.4	0.47
2800	700	25	85.1	0.69
4000	2000	50	90.4	0.26
3600	1800	50	93.0	0.2
3200	1600	50	83.9	0.51
2800	1400	50	84.6	0.48

Table 3. Validation set accuracy for different window lengths and overlapping, best result displayed in bold.

Each action constitutes a time series composed of acceleration signals along three orthogonal axes, XYZ. However, extracting motion patterns directly from raw acceleration data is exceedingly challenging because of high-frequency oscillations and noise. Therefore, prior to modelling, the raw data undergo preprocessing for feature extraction and selection. Window size and stride are critical hyperparameters in neural networks and other machine learning algorithms. To identify the optimal combination for our specific network, we conducted a series of tests varying these two parameters. Given that the discrete wavelet transform was employed for data preprocessing, the minimum window size required was 2560. Therefore, in our experiments, the smallest window size was set to 2800, considering the maximum data length of 5000, with the largest window size capped at 4000. This upper limit was chosen because increasing the window size further would not sustain an adequate number of data points to support the creation of multiple windows given the stride, thus impacting data utilization and the comprehensiveness of the analysis. As demonstrated in Table 3, the results indicate that the optimal validation accuracy was achieved with a window size of 3600 and a step size of 1800. Through overlapping and repetitive segmentation, each piece of raw acceleration data was divided into two windows, each 3600 in size with a stride of 1800, resulting in a total of 3064 training windows. Within the edge impulse platform, a robust digital signal processing (DSP) module is subsequently provided. Given the data typology of this study, we employed an input processing module based on spectral analysis for feature extraction. The spectral feature block encompasses two practical tools: the fast Fourier transform (FFT) and the discrete wavelet transform (DWT). We opted for DWT because it allows for signal decomposition on the basis of local changes in time and frequency, making it more suitable for analysing complex signals with transient or irregular waveforms, such as those emanating from accelerometer-based motion or vibration.

The raw acceleration data are decomposed into multiple approximations and detail levels through the application of the discrete wavelet transform (DWT), facilitating a nuanced analysis of the signal. Specifically, the features encompass acceleration axes (AccX, AccY, AccZ), levels (L0, L1, L2), and 14 statistical measures or transformations, yielding a total of 126 features. Following signal decomposition, relevant features are computed

at each level, including time-domain features (zero cross, mean cross, median, mean, standard deviation, variance, and root mean square), frequency-domain features (entropy), and statistical domain features (5th percentile, 25th percentile, 75th percentile, 95th percentile, skewness, and kurtosis).

Implementation of the TinyML model

In this section, the neural network architecture and related work are described. To identify eight distinct upper limb rehabilitation actions exclusively through accelerometer data, we selected Edge Impulse, a web-based machine learning development platform, which furnished us with a comprehensive deep learning solution.

To identify the optimal training parameters for the model, this study conducted a series of experiments aimed at constructing an effective data structure for deep learning model inputs, ensuring the model's ability to accurately learn and predict. Table 4 presents the optimal combination of hyperparameters determined through these experiments. The dataset used for training the model was randomly divided into two subsets: a training subset comprising 80% of the data and a testing subset consisting of the remaining 20%. The model was trained on the training subset and subsequently tested on the testing subset.

The Edge Impulse Studio offers a straightforward and effective module for classification learning. A multilayer perceptron (MLP) was constructed and trained utilizing the TensorFlow and Keras libraries for this classification task. Specifically, model training employs a sequential model architecture, providing a linear stack for the hidden layers. The deep neural network is engineered with an input layer that captures 126 distinct features and is meticulously designed to identify the unique attributes inherent to each category. To enhance the model's robustness and prevent overfitting, it includes two hidden dense layers equipped with 40 and 20 neurons. A dropout layer, introduced with a dropout rate of 0.25, is strategically positioned between these dense layers to diminish the likelihood of overfitting, thereby ensuring a more generalized model performance.

For the choice of the activation function, the hidden layer uses the rectified linear unit (ReLU) activation function. Because ReLU can introduce nonlinearity while maintaining computational simplicity, it helps the model learn complex data representations and improves the model training speed, generalizability and robustness. Its function expression is shown in Eq. (1):

$$\text{ReLU}(x) = \max(0, x) \quad (1)$$

where x is the input to the neuron, \max represents taking the larger of the two numbers, and 0 is the cut-off point for ReLU, which outputs x if x is positive and 0 if x is negative.

For the output layer, a Softmax activation function was used. Since this study involves the classification of eight different upper limb rehabilitation actions, the output layer is configured with eight neurons, and the output of each neuron represents the probability of belonging to the corresponding class. The Softmax function can convert the output into a probability distribution, and the Softmax output of the j^{th} neuron is defined as Eq. (2) For a problem with K classes:

$$\text{Softmax}(x)_j = \frac{e^{x_j}}{\sum_k^K e^{x_k}} \quad (2)$$

where x_j is the input of the j^{th} output neuron, e^{x_j} is the exponent of x_j and the denominator is the sum of the exponent values of all the output neurons. The Softmax function ensures that the sum of all the output values is 1 and that each of the output values ranges from 0 to 1, which represents the probability of belonging to the corresponding category.

During training, the model is compiled via the Adam optimizer with a loss function of categorical_crossentropy, as shown in Eq. (3):

$$\text{Cross - Entropy Loss} = - \sum_{c=1}^M y_{o,c} \log(p_{o,c}) \quad (3)$$

where M is the number of categories, and for each category c , y is a binary indicator (0 or 1) of whether category c is correctly categorized, and $p_{o,c}$ is the probability that the model predicts that observation o belongs to category c . The function $\log(p_{o,c})$ takes the natural logarithm of this predicted probability. Equation (3) The cross-entropy loss function, an extensively employed metric in classification endeavours to gauge the disparity

Considerations	Optimal value
Training set	80%
Test set	20%
Input layer	126 features
Dense layer	40 neurons
Dense layer	20 neurons
Dropout	rate 0.25
Number of training cycles	60
Learning rate	0.0005

Table 4. The optimal combination of hyperparameters.

between the forecasted probabilities and the empirical distribution of outcomes, is delineated. The term 'Cross' is indicative of this function's deployment over both predicted and factual outcomes, whereas 'entropy' encapsulates a metric of indeterminacy or disarray prevalent within the prognostications. Overall, the 'cross-entropy loss' advantageously encapsulates the divergence of the model's forecasts from the authentic labels, furnishing a formidable metric for optimization during the neural network model training regimen.

Model training was performed via the `model.fit` method with inputs from training and validation datasets, and the number of rounds, batch size, and callback functions were set. Figure 7 shows the final structure of the neural network after training. Among them, x in the input layer represents a 126 dimensional feature vector formed by extracting features from three—axis acceleration data. The first hidden layer $a^{[1]}$ contains 40 neurons, and the second hidden layer $a^{[2]}$ has 20 neurons. Through further feature refinement, the final output layer consists of $-p_1 p_8$, corresponding to the 8 action classifications in the task of this paper. Each p represents the predicted probability of the corresponding category.

K-Means-based Anomaly Detection

Neural networks exhibit remarkable learning capabilities within the confines of their trained sample space, yet they present distinct limitations when encountering novel data beyond this predefined scope, such as new upper limb movements. This limitation stems from the inherent capacity of neural networks to recognize and comprehend only the data patterns they have been exposed to during their training regimens. Given that data significantly diverge from the training sample space, neural networks tend to erroneously categorize these novel inputs into one of the few known categories. To address this issue, we employed the K-means anomaly detection block, which is designed to discern data points that might otherwise be misconstrued within the conventional action classification framework.

K-means, as an unsupervised learning technique, furnishes an efficacious supplementary mechanism. It delves into the intrinsic structure of data, autonomously discerning latent clusters, thereby revealing unlabelled patterns within the dataset. The edge impulse platform has two adjustable parameters: the cluster count and axes. The cluster count denotes the segmentation of akin data points into a predetermined number of K clusters; the axes pertain to features generated by the preprocessing blocks, with the selected axes utilizing these features as input data for training.

Throughout the inference process, after extensive training, the optimal number of clusters was determined to be 32. The digital signal processing (DSP) module of the Edge Impulse Studio platform is capable of automatically extracting features from acceleration data and offers an option for computing feature importance. This process ensures the identification and utilization of the most critical features to enhance the model's performance. Consequently, guided by the feature importance list generated by the Edge Impulse Studio platform, this study ultimately selected AccX L0-var, AccY L0-n95, AccY L0-rms, and AccZ L0-rms as the most significant features. For each cluster C_k , the center of mass μ_k is calculated as shown in Eq. (4):

$$\mu_k = \frac{1}{|C_k|} \sum_{x \in C_k} x \quad (4)$$

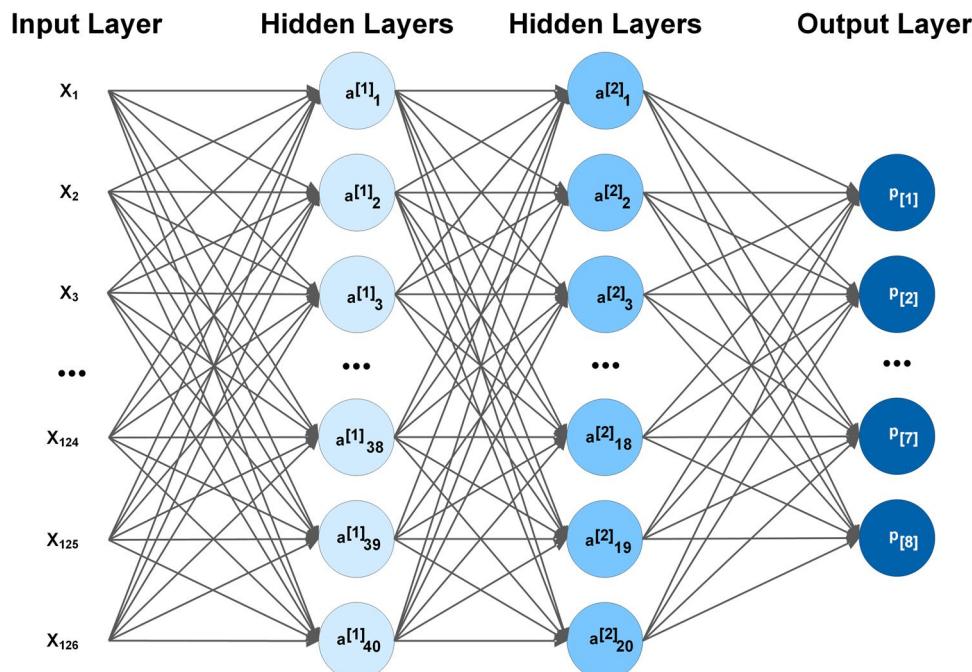


Fig. 7. Final structure of the implemented and deployed neural network.

where $|C_k|$ is the number of data points in cluster C_k and where C_k is the number of data points in the cluster. The data points x_i are assigned to the nearest center of mass μ_k to form the clusters, and this process can be expressed as:

$$\text{assign}(x_i) = \text{argmin}_k \|x_i - \mu_k\|^2 \quad (5)$$

We calculate the distance between a new data point and its nearest center of mass and detect anomalies by setting a threshold as:

$$d(x_i, \mu_k) = \|x_i - \mu_k\| \quad (6)$$

If $d(x_i, \mu_k)$ is greater than the threshold, then x_i is an anomaly. The threshold in this study is set to 0. An anomaly score below 0 indicates that the data are within a cluster (no anomaly). Using the k-means anomaly detection module, this study computed the anomaly scores for new data points, which were then evaluated within the software. This program concurrently provides the classification outcomes from the neural network alongside the corresponding anomaly scores. The classification results are deemed valid when the anomaly score falls below a predetermined threshold; conversely, if the anomaly score exceeds this threshold, this indicates that the user's performed action does not align with any of the eight predefined movements, thereby rendering the classification result invalid. This mechanism significantly enhances the robustness and accuracy of the overall classification system.

Model compression

Given the deployment of neural networks on MCUs with severely constrained computational resources, quantization has emerged as one of the most efficacious optimization strategies. Conventionally, weights and biases within the model are stored as 32-bit floating-point numbers to facilitate high-precision computations during the training phase; however, this incurs excessive resource consumption when executed on MCUs. The technique of quantization reduces the precision of these weights and biases, enabling their storage as 8-bit integers. Nonetheless, transitioning models from 32-bit floating-point representations to 8-bit integer representations has historically been challenging, necessitating profound knowledge of model architecture and the ranges of activation layers. Fortunately, the edge impulse platform has introduced the edge optimized neural compiler (EON), which, compared with TensorFlow Lite for microcontrollers, enhances neural network execution speeds by approximately 25–55% and reduces flash storage by up to 35%, all while maintaining comparable accuracy levels. This innovation has significantly reduced engineering time and the costs associated with repeated iterative experiments for this study.

Results

Input processing module

Acceleration data encapsulate a wealth of information pertaining to the dynamic characteristics of an object, information that may fluctuate over time and potentially reside within the frequency components of the data. The selection of an appropriate data processing methodology is paramount for the extraction of pertinent features and patterns when dealing with acceleration data. The input processing modules under consideration include the following:

- (1) **Raw data:** An input block that eschews data preprocessing.
- (2) **Spectral Analysis:** This input module is particularly adept at analysing repetitive motions by extracting features related to the variation in frequency and power over time and is applicable to data from accelerometers or audio signals.
- (3) **Flatten:** This block flattens axes into singular values, which is beneficial for slowly varying means, such as temperature data when combined with other blocks.

Table 5 lists the precision, loss, inference time, peak RAM usage, and peak flash consumption associated with three input processing modules: flatten, spectral analysis, and raw data. Overall, spectral analysis demonstrated a significant advantage in handling acceleration data compared with flattening and raw data. By converting time-domain signals into frequency-domain signals, spectral analysis can reveal the periodic characteristics and frequency distributions of the signal, which might be obscured in the raw time-domain signal. Raw data retain all original information but may also encompass a considerable amount of noise and irrelevant information, complicating analysis and pattern recognition and demanding more resources. Flattening transforms multidimensional arrays into one-dimensional arrays, simplifying data handling but potentially losing temporal and structural information in the data, leading to suboptimal training outcomes. Hence, this study opts for spectral analysis as the input processing block to generate input features for neural network classification.

	Accuracy (%)	Loss	Inferencing time (ms)	Peak ram usage (KB)	Flash usage (kB)
Flatten	82.4	0.48	1	1.4	16.5
Spectral Analysis	91.9	0.25	2s	1.5	20.6
Raw data	90.8	0.36	11	2.1	45.0

Table 5. Performance metrics of the flattening, spectral analysis, and raw data input processing modules.

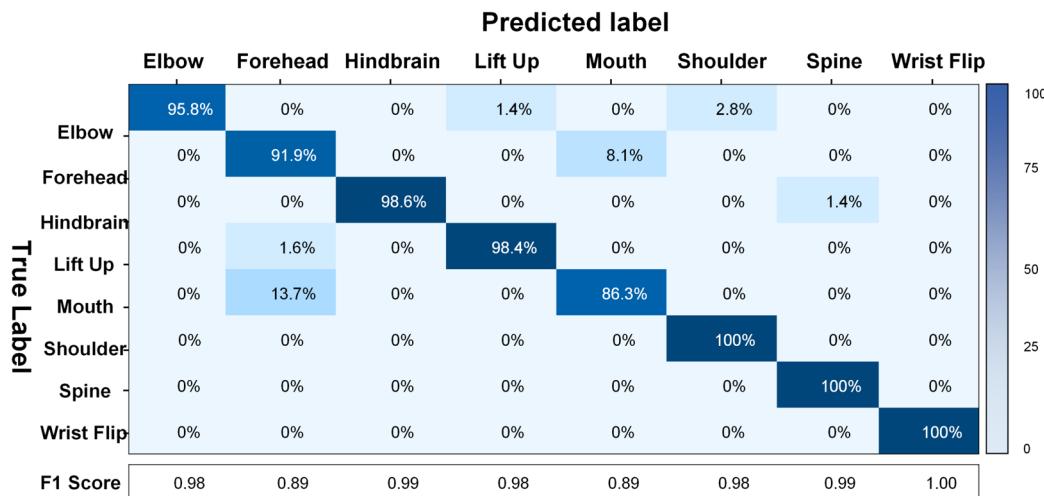


Fig. 8. Confusion matrix for full-precision neural networks.

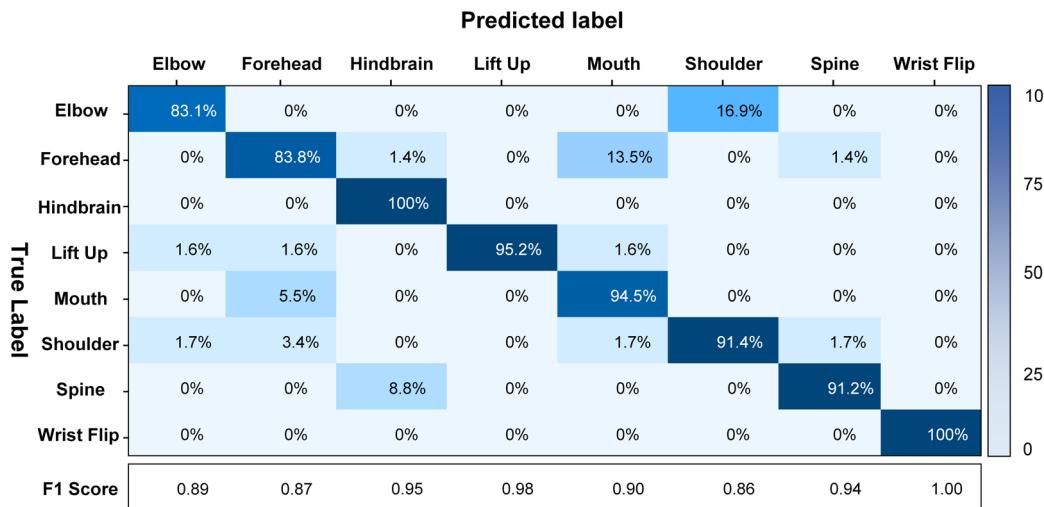


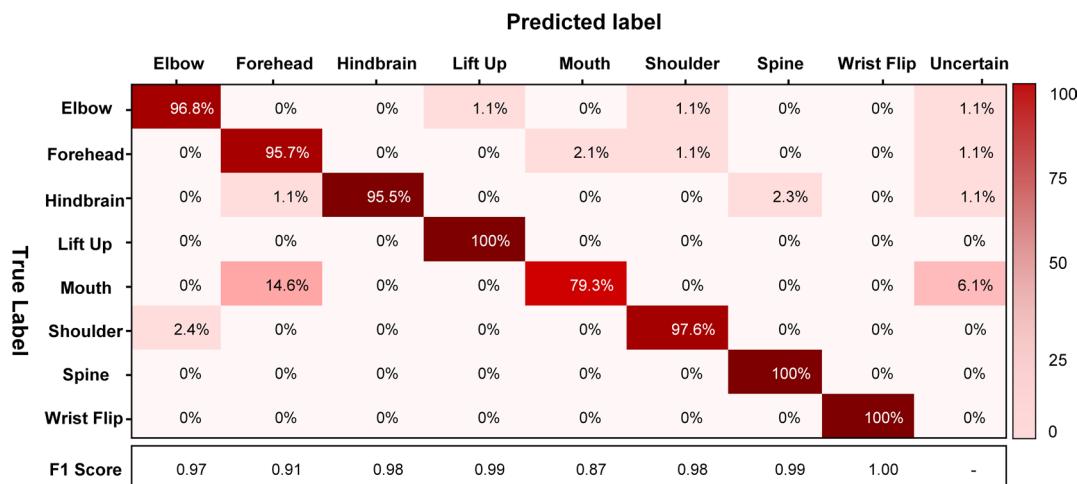
Fig. 9. Confusion matrix of the quantized neural network.

Performance modelling

As illustrated in Fig. 8, the full-precision neural network achieved an accuracy of 96.1% on the training set, with a loss of 0.14 and an F1 score of 0.9625. The confusion matrix of the 8-bit quantized neural network, depicted in Fig. 9, indicates only a marginal decline in model performance, as the quantized neural network still yielded an accuracy of 92.2% on the same training set, with a loss of 0.45 and an F1 score of 0.924. Figure 10 shows the performance of the final neural network model on the test set, where it achieved an accuracy of 95.09% and an F1 score of 0.961. Table 6 displays the detailed scoring for the eight categorized actions.

Model performance

Table 7 delineates the model parameters with float32 precision weights, whereas Table 8 presents the parameters of the model postquantification to int8 precision. A comparative analysis revealed a decrease in model size from a full-precision neural network of 52.2 kB to a quantized dimension of 25.3 kB, resulting in a 51.53% reduction in model size at a modest precision cost of 7.08%. Notably, the impact of quantization on neural network optimization is particularly pronounced. The inference time of the classifier was augmented from 54 to 4, indicating a 13.5-fold acceleration. However, the enhancement attributable to quantization was not significantly evident for the spectral analysis input processing module. During the model training process, accuracy and loss demonstrated effective convergence as the number of training epochs increased, reaching saturation around the 60th epoch. Both training and validation losses converged towards the same lower value. This indicates that the model has successfully assimilated the characteristics of the data and displayed commendable generalization capabilities.

**Fig. 10.** Confusion matrix of the model in the test set.

Category	Precision (%)	Recall (%)	F1-Score (%)	Support
Elbow	100.0	95.75	97.83	47
Forehead	86.96	85.11	86.02	47
Hindbrain	100.0	95.45	97.67	44
Lift up	97.62	100.0	98.80	41
Mouth	83.33	85.37	84.34	41
Shoulder	97.67	100.0	98.82	42
Spine	95.65	100.0	97.78	44
Wrist Flip	100.0	100.0	100.0	40
Overall/Average	95.14	95.09	95.06	346

Table 6. Detailed performance metrics for eight action classifications.

	Spectral Analysis	Classifier	Total
Latency	13	54	67
RAM	4.6 kB	1.9 KB	4.6 KB
FLASH	-	34.8 KB	-
Accuracy	-	-	87.43%

Table 7. Unoptimized (float32) model metrics.

	Spectral Analysis	Classifier	Total
Latency	13	2	15
RAM	4.6 kB	1.5 KB	4.6 KB
FLASH	-	20.6 KB	-
Accuracy	-	-	84.88%

Table 8. Quantized model indicators (int8).

Power consumption

The greatest advantage of embedded systems over computer and mobile systems is their very low power consumption. The energy consumption can be calculated via:

$$E = U * I * t \quad (7)$$

Mode	Power (mW)	Current (mA)
Idle	24.75	7.50
Modelled reasoning	29.9	9.065
BLE advertising model	53.96	16.35
BLE connection mode	54.05	16.38

Table 9. System core power consumption parameters.

where E is the energy, U is the voltage, I is the current and t is time. We tested the hardware system in both stationary and operational modes to evaluate its power consumption.

In the present investigation, a DT830B digital multimeter was utilized, with its measurement electrodes serially integrated into the circuitry of the wearable apparatus, facilitating an exhaustive appraisal and elucidation of the principal energy consumption associated with the wearable system, specifically during the TinyML model inference process. This approach encompasses a detailed examination of power consumption metrics across various operational states, including the model inference phase, BLE advertisement phase, BLE connectivity phase, and device energy expenditure while in standby mode. Table 9 shows the details of the core power consumption of the system. According to our measurements, in idle mode, the system consumes 24.75 mW of power and 7.50 mA of current. This reflects the baseline energy consumption of the system when there is no significant computational or communication activity. The system communicates at 53.96 mW when the BLE is advertised. The power consumption of the system is very close at 54.05 mW in the BLE connectivity mode. The power consumption of the acceleration sensor is also low, with a power of 1 mW, which has a relatively small effect on the overall power consumption of the system.

Discussion

Even with a solitary inertial sensor, the model achieved a commendable accuracy of 96.1% on the training dataset and 95.09% on the test dataset. Although this model does not surpass those of some of the previous studies in terms of accuracy, the novelty of this study lies in its ability to recognize eight distinct upper limb rehabilitation movements with a streamlined sensor setup. In this study, the eight types of upper-limb rehabilitation movements display high kinematic similarity, exhibiting subtle yet crucial differences in aspects such as motion trajectory, duration, and limb velocity. Therein lies a significant challenge in achieving precise classification solely through accelerometer data. To address this issue, the current research employs a novel strategy that combines a multilayer perceptron (MLP) with an anomaly recognition K-means algorithm for precise classification. This approach demonstrates the heterogeneity of the model by combining the strengths of deep learning and traditional machine learning algorithms. The multilayer perceptron, as a powerful nonlinear model, is capable of extracting deep features from complex data, whereas the K-means algorithm, known for its high computational efficiency and simplicity, performs efficient anomaly detection on the basis of these extracted features. This heterogeneous combination not only enhances the model's ability to process diverse data but also optimizes the use of computational resources. Model quantization was facilitated by TinyML technology, resulting in a final deployed model size of only 20.6 KB. Notably, TinyML technology also offers substantial benefits in terms of data security and privacy; since only the data processed by the model are transmitted, the raw data captured by the sensors remain confined to the microcontroller node, thereby negating the risk of data interception.

From the power consumption assessment of this system, it is evident that the power consumption of the deployed model is comparatively low, with an inference time per model of approximately 15, thus resulting in lower energy usage. Wireless communication has emerged as the primary energy consumer within this wearable setup. Despite the implementation of low-power BLE technology aligning with the expectations for low-power communication standards, the energy expended on wireless communication still constitutes the majority of the system's total power consumption⁶⁷. This can be attributed primarily to the several energy-intensive tasks involved in the wireless communication protocol, the CPU being active, and the current for the RF transmission stage. The transmission of wireless signals requires a relatively substantial amount of energy to power the RF transmitter, ensuring that the signal's strength is adequate for reception, particularly over extended distances or in the presence of interference. Moreover, wireless modules are required to sustain high-speed data handling and frequency alterations during both the transmission and reception phases, which further increases energy consumption.

In conclusion, the home upper limb rehabilitation system presented in this study, which uses TinyML technology on microcontroller units, achieves proficient motion classification, shows benefits such as reduced power consumption, minimal latency, and a lightweight framework. Nonetheless, this system is encumbered by certain constraints. First, the research is predicated on data derived from healthy subjects and has not been comprehensively validated within real stroke patient cohorts, which may limit its immediate applicability in clinical environments. The healthy population in this study does not directly represent real stroke patients but serves merely as a data source for model training, collecting limb movement data. The ultimate rehabilitation goal for the upper limb dysfunction group was to achieve such healthy standards through subsequent rehabilitation training. Furthermore, TinyML's drive towards miniaturization and energy efficiency imposes architectural limitations on the model owing to the constraints of hardware resources, rendering the deployment of intricate machine learning models for tasks necessitating high accuracy and elaborate feature extraction a formidable endeavor. As a result, the ability of the proposed upper limb rehabilitation motion monitoring apparatus to furnish motion classification insights without improving the quality of movements is limited. There is also a

discernible need for enhancement in the differentiation of akin movements, particularly those with motion amplitudes closely resembling one another. The efficacy of hardware devices during data acquisition may be compromised by a myriad of factors, including device placement, the nature of user activities, and individual variances, thereby infusing additional noise into the dataset and influencing the model's analytical precision. The ability of the system to identify eight fundamental upper limb rehabilitation movements does not encapsulate the entire spectrum of movement types that may manifest throughout the rehabilitation trajectory of stroke survivors. Finally, model training is executed on the Edge Impulse web platform, resulting in a static neural network model; any requisite modifications to the detection objectives mandate retraining and redeployment processes on the platform. While this limitation might be inconsequential for diminutive systems, it represents a substantial impediment for expansive detection frameworks encompassing thousands of MCUs, where the task of batch updating models emerges as an onerous challenge. In future research, we will explore the potential of online learning and dynamic deployment within TinyML. Specifically, by integrating online learning and incremental learning techniques, we aim to enhance the model's capability for dynamic updates. Online learning allows for real-time model adjustments directly on the device without the need to transfer sensitive data, thus improving performance while maintaining data privacy.

Furthermore, we will continue to incorporate research on model deployment into our long-term plans, consistently monitoring and exploring innovative technologies in this field to drive further advancements in related research.

The focus of this study was to explore the feasibility of using TinyML devices in the recognition of upper limb rehabilitation movements. Our methodology is based on a minimal sensor setup, employing an embedded microcontroller unit to execute the TinyML model for precise motion classification. We have adopted a single accelerometer approach, which, despite its limitations in detecting rotational movements, offers significant advantages due to its lightweight and low-cost attributes. Additionally, its low power consumption provides a crucial benefit for home rehabilitation devices, especially under the demands of long-term continuous monitoring. The low power design of the accelerometer allows the device to operate for extended periods on battery power, which is vital for practical applications in home rehabilitation settings. Through meticulous preprocessing and feature extraction of the accelerometer data, our system demonstrates robust performance in classifying rehabilitation movements that primarily occur in the sagittal and coronal planes. This characteristic enables effective recognition of common rehabilitation movements without reliance on complex sensors. Nonetheless, given the current hardware configuration, we recognize that there is room for improvement in the recognition capabilities for more complex movements, particularly those involving rotational motions. Especially when compared to high-end robotic neurorehabilitation equipment used in clinical environments, our device may have limitations in functionality and precision. However, we acknowledge that compared with high-end robotic neural rehabilitation devices used in clinical settings, our device may have limitations in terms of functionality and precision. This study aimed to complement traditional clinical rehabilitation methods by providing patients with more home rehabilitation options rather than completely replacing professional clinical rehabilitation services. Robotic rehabilitation devices can offer fine motion assistance and feedback, but their cost and operational complexity limit their widespread application in home rehabilitation. We emphasize that the research and the application recommendations for the methodology and devices described are primarily suitable for the later stages of patient rehabilitation. At this stage, patients have transitioned from initial clinical rehabilitation to home-based rehabilitation (with a Fugl-Meyer assessment score ≥ 20), possessing sufficient autonomous movement ability to engage in active exercise training. Their range of movement and complexity are relatively limited. Therefore, the system can provide adequate precision for handling these patients' common rehabilitation movements. At this time, our study offers an affordable, lightweight home rehabilitation movement monitoring system. However, for patients with severe motor disabilities (Fugl-Meyer assessment score < 10), their movement patterns may include more compensatory rotational actions, which would require more precise sensor support. In such cases, activating the full 9-axis IMU functionality becomes necessary to ensure accurate monitoring, enabling the system to meet the rehabilitation needs of patients at different stages of recovery.

Certainly, the primary objective of this study remains the feasibility validation of TinyML technology. The above discussion primarily explores the future potential of this technology in home rehabilitation scenarios. Consequently, to increase the rigor and practicality of future research, we plan to conduct an in-depth investigation into the perspectives of actual stroke patients regarding our designed wearable system and to evaluate the system's usability. In future work, we will focus on how methodological innovations can improve the service experience for the target population. Specifically, we intend to explore the deep integration of TinyML technology with smart product service systems, aiming to develop a new ecosystem for wearable device services and extend the application of our research findings to real home environments. In response to the limitations previously mentioned, we will continue to work on improving computational efficiency, optimizing model architecture, enhancing model adaptability and precision, and facilitating the effective application and dissemination of the technology.

Conclusion

Drawing from the training and outcomes illustrated within this paper, the TinyML-based home upper limb rehabilitation system proposed herein is capable of effectively classifying eight distinct upper limb rehabilitation movements with a minimal sensor setup on a device constrained by significantly limited resources. In contrast to extant research findings, this methodology circumvents the laborious cloud computing workflow, thereby furnishing a more economical, compact, and energy-efficient solution for domiciliary rehabilitation, concurrently mitigating the risk of privacy intrusions within domestic settings. The principal conduit for data acquisition, preprocessing, model training, evaluation, and deployment is facilitated through the edge impulse platform.

In this investigation, we crafted and realized the comprehensive architecture of the IoT solution, wherein the concept of an embedded system is employed for data gathering, action classification, and rendering both visual and haptic feedback to the user, which collectively form the physical layer of the system. Concurrently, a corresponding application was developed to facilitate information interchange with hardware units leveraging the BLE wireless communication protocol, aiming to provide users with a visualization of the imparted and logged rehabilitation training. This solution integrates an innovative approach that combines an MLP and the anomaly detection K-means algorithm for precise classification, enhancing the ability to differentiate similar motions and exhibiting significant advantages in terms of accuracy and robustness. Moreover, the model delineated in this study is deployed on devices with exceedingly constrained computational resources via TinyML technology, catering to the requisites of wearable apparatuses characterized by minimal power consumption, affordability, and portability.

In conclusion, the domiciliary upper limb rehabilitation system delineated in this investigation provides an efficacious novel methodology for the precise classification of upper limb rehabilitation movements and has promising application prospects for wearable technologies in the forthcoming era.

Data availability

The dataset from the current analysis is not public but is available from the corresponding author upon reasonable request.

Received: 17 October 2024; Accepted: 7 May 2025

Published online: 24 May 2025

References

1. Feigin, V. L. et al. World Stroke Organization (WSO): Global stroke fact sheet 2022. *Int. J. Stroke* **17**, 18–29 (2022).
2. Veerbeek, J. M., Kwakkel, G., Van Wegen, E. E. H., Ket, J. C. F. & Heymans, M. W. Early prediction of outcome of activities of daily living after stroke. *Stroke* **42**, 1482–1488 (2011).
3. Benjamin, E. J. et al. Heart disease and stroke statistics—2018 update: A report from the American Heart Association. *Circulation* <https://doi.org/10.1161/CIR.0000000000000573> (2018).
4. Mennella, C., Maniscalco, U., De Pietro, G. & Esposito, M. The role of artificial intelligence in future rehabilitation services: A systematic literature review. *IEEE Access* **11**, 11024–11043 (2023).
5. Taub, E., Uswatte, G., Mark, V. W. & Morris, D. M. M. The learned nonuse phenomenon: implications for rehabilitation. *PubMed* **42**, 241–256 (2006).
6. Banks, G., Bernhardt, J., Churilov, L. & Cumming, T. B. Exercise Preferences Are Different after Stroke. *Stroke Res. Treat.* **2012**, 1–9 (2011).
7. Stinear, C. M., Lang, C. E., Zeiler, S. & Byblow, W. D. Advances and challenges in stroke rehabilitation. *Lancet Neurol.* **19**, 348–360 (2020).
8. Young, J. A. & Tolentino, M. Stroke evaluation and treatment. *Top. Stroke Rehabil.* **16**, 389–410 (2009).
9. European Commission, Council of the European Union, Directorate-General for Economic and Financial Affairs, Economic Policy Committee, “The 2015 ageing report – Economic and budgetary projections for the 28 EU Member States (2013–2060),” Publications Office, 2015.
10. Mayo, N. E. et al. There’s no place like home. *Stroke* **31**, 1016–1023 (2000).
11. Hillier, S. & Inglis-Jassiem, G. Rehabilitation for community-dwelling people with stroke: Home or centre based? A systematic review. *Int. J. Stroke* **5**, 178–186 (2010).
12. Mayo, N. E. Stroke rehabilitation at home. *Stroke* **47**, 1685–1691 (2016).
13. Patel, S., Park, H., Bonato, P., Chan, L. & Rodgers, M. A review of wearable sensors and systems with application in rehabilitation. *J. NeuroEng. Rehabil.* <https://doi.org/10.1186/1743-0003-9-21> (2012).
14. Rimmer, J. H. Barriers associated with exercise and community access for individuals with stroke. *J. Rehabil. Res. Dev.* **45**, 315–322 (2008).
15. Anderson, C., Mhurchu, C. N., Brown, P. M. & Carter, K. Stroke rehabilitation services to accelerate hospital discharge and provide Home-Based care. *Pharmacoeconomics* **20**, 537–552 (2002).
16. Chen, C.-Y., Neufeld, P. S., Feely, C. A. & Skinner, C. S. Factors influencing compliance with home exercise programs among patients with Upper-Extremity Impairment. *Am. J. Occup. Ther.* **53**, 171–180 (1999).
17. Daly, J. et al. Barriers to Participation in and adherence to cardiac Rehabilitation Programs: A Critical Literature review. *Prog. Cardiovasc. Nurs.* **17**, 8–17 (2002).
18. Komatireddy, R. et al. Quality and quantity of rehabilitation exercises delivered by a 3-D motion controlled camera: A pilot study. *Int. J. Phys. Med. Rehabil.* <https://doi.org/10.4172/2329-9096.1000214> (2014).
19. Majumder, S., Mondal, T. & Deen, M. Wearable sensors for remote health monitoring. *Sensors* **17**, 130 (2017).
20. Ahmed, E., Yaqoob, I., Gani, A., Imran, M. & Guizani, M. Internet-of-things-based smart environments: state of the art, taxonomy, and open research challenges. *IEEE Wirel. Commun.* **23**, 10–16 (2016).
21. Čolaković, A. & Hadžić, M. Internet of Things (IoT): A review of enabling technologies, challenges, and open research issues. *Comput. Netw.* **144**, 17–39 (2018).
22. Avellaneda, D., Mendez, D. & Fortino, G. A TinyML deep learning approach for indoor tracking of assets. *Sensors* **23**, 1542 (2023).
23. Antico, M. et al. Postural control assessment via Microsoft Azure Kinect DK: An evaluation study. *Comput. Methods Programs Biomed.* **209**, 106324 (2021).
24. Faity, G., Motte, D. & Froger, J. Validity and reliability of Kinect v2 for quantifying upper body kinematics during seated reaching. *bioRxiv (Cold Spring Harbor Laboratory)* (2022) <https://doi.org/10.1101/2022.01.18.476737>.
25. Otten, P., Kim, J. & Son, S. A framework to automate assessment of upper-limb motor function impairment: A feasibility study. *Sensors* **15**, 20097–20114 (2015).
26. Lee, S., Lee, Y.-S. & Kim, J. Automated evaluation of upper-limb motor function impairment using Fugl-meyer assessment. *IEEE Trans. Neural Syst. Rehabil. Eng.* **26**, 125–134 (2017).
27. Sheng, B. et al. A novel scoring approach for the Wolf Motor Function Test in stroke survivors using motion-sensing technology and machine learning: A preliminary study. *Comput. Methods Programs Biomed.* **243**, 107887 (2023).
28. Zhou, H. & Hu, H. Human motion tracking for rehabilitation—A survey. *Biomed. Signal Process. Control* **3**, 1–18 (2007).
29. Hadjidi, A., Souil, M., Bouabdallah, A., Challa, Y. & Owen, H. Wireless sensor networks for rehabilitation applications: Challenges and opportunities. *J. Netw. Comput. Appl.* **36**, 1–15 (2012).

30. Henriksen, E., Burkow, T. M., Johnsen, E. & Vognild, L. K. Privacy and information security risks in a technology platform for home-based chronic disease rehabilitation and education. *BMC Med. Inf. Decision Making* <https://doi.org/10.1186/1472-6947-13-8> 5 (2013).
31. Nahavandi, D., Alizadehsani, R., Khosravi, A. & Acharya, U. R. Application of artificial intelligence in wearable devices: Opportunities and challenges. *Comput. Methods Programs Biomed.* **213**, 106541 (2021).
32. Villeneuve, E., Harwin, W., Holderbaum, W., Janko, B. & Sherratt, R. S. Reconstruction of angular kinematics from Wrist-Worn inertial sensor data for smart home healthcare. *IEEE Access* **5**, 2351–2363 (2017).
33. Lui, J. & Menon, C. Would a thermal sensor improve arm motion classification accuracy of a single wrist-mounted inertial device?. *BioMed. Eng. OnLine* <https://doi.org/10.1186/s12938-019-0677-7> (2019).
34. Sardini, E., Serpelloni, M. & Pasqui, V. Wireless wearable T-Shirt for posture monitoring during rehabilitation exercises. *IEEE Trans. Instrum. Meas.* **64**, 439–448 (2014).
35. García-De-Villa, S., Casillas-Pérez, D., Jiménez-Martín, A. & García-Domínguez, J. J. Inertial sensors for human motion analysis: A comprehensive review. *IEEE Trans. Instrum. Meas.* **72**, 1–39 (2023).
36. Connell, L. A., McMahon, N. E., Simpson, L. A., Watkins, C. L. & Eng, J. J. Investigating measures of intensity during a structured upper limb exercise program in stroke Rehabilitation: an exploratory study. *Arch. Phys. Med. Rehabil.* **95**, 2410–2419 (2014).
37. Lang, C. E., MacDonald, J. R. & Gnip, C. Counting repetitions: An observational study of outpatient therapy for people with hemiparesis post-stroke. *J. Neurol. Phys. Ther.* **31**, 3–10 (2007).
38. Yurur, O., Liu, C.-H. & Moreno, W. Unsupervised posture detection by smartphone accelerometer. *Electron. Lett.* **49**, 562–564 (2013).
39. Guo, H., Chen, L., Chen, G. & Lv, M. Smartphone-based activity recognition independent of device orientation and placement. *Int. J. Commun. Syst.* **29**, 2403–2415 (2015).
40. Zinnen, A., Blanke, U. & Schiele, B. An analysis of sensor-oriented vs. model-based activity recognition. *Int. Symp. Wear. Comput.* <https://doi.org/10.1109/iswc2009.32> (2009).
41. Zhang, N. Z., Fang, N. Q. & Ferry, F. Upper limb motion capturing and classification for unsupervised stroke rehabilitation. *IECON 2020 the 46th Annual Conference of the IEEE Industrial Electronics Society* **26**, 3832–3836 (2011).
42. Wang, S., Liao, J., Yong, Z., Li, X. & Liu, L. Inertial sensor-based upper limb rehabilitation auxiliary equipment and upper limb functional rehabilitation evaluation. *Commun. Comput. Inf. Sci.* https://doi.org/10.1007/978-981-19-4546-5_40 (2022).
43. Alessandrini, M., Biagiotti, G., Crippa, P., Falaschetti, L. & Turchetti, C. Recurrent neural network for human activity recognition in embedded systems using PPG and accelerometer data. *Electronics* **10**, 1715 (2021).
44. Li, Q. et al. A motion recognition model for upper-limb rehabilitation exercises. *J. Ambient. Intell. Humaniz. Comput.* **14**, 16795–16805 (2023).
45. Koldo Basterretxea, Echanobe, J. & Ines del Campo. A wearable human activity recognition system on a chip. *Conference on Design and Architectures for Signal and Image Processing* (2014).
46. Xu, J. & Yuan, K. Wearable muscle movement information measuring device based on acceleration sensor. *Measurement* **167**, 108274 (2020).
47. Choudhury, T. et al. The mobile sensing platform: An embedded activity recognition system. *IEEE Pervasive Comput.* **7**, 32–41 (2008).
48. Zhang, M. & Sawchuk, A. A. Human daily activity recognition with sparse representation using wearable sensors. *IEEE J. Biomed. Health Inform.* **17**, 553–560 (2013).
49. Khan, A. M., Lee, N.Y.-K., Lee, S. Y. & Kim, N.T.-S. A triaxial Accelerometer-Based Physical-Activity recognition via Augmented-Signal features and a hierarchical recognizer. *IEEE Trans. Inf. Technol. Biomed.* **14**, 1166–1172 (2010).
50. Tseng, M.-C., Liu, K.-C., Hsieh, C.-Y., Hsu, S. J. & Chan, C.-T. Gesture spotting algorithm for door opening using single wearable sensor. In: *2018 IEEE International Conference on Applied System Invention (ICASI)* 854–856 (2018) <https://doi.org/10.1109/icasi.2018.8394398>.
51. Zhang, Z., Liparulo, L., Panella, M., Gu, X. & Fang, Q. A fuzzy kernel motion classifier for autonomous stroke rehabilitation. *IEEE J. Biomed. Health Inf.* **20**, 893–901 (2015).
52. Giordano, M. et al. Design and performance evaluation of an Ultralow-Power smart IoT device with embedded TinyML for asset activity monitoring. *IEEE Trans. Instrum. Meas.* **71**, 1–11 (2022).
53. Warden, P. & Situnayake, D. *TinyML: Machine Learning with TensorFlow Lite on Arduino and Ultra-Low-Power Microcontrollers*. (2019).
54. Sanchez-Iborra, R. & Skarmeta, A. F. TinyML-Enabled Frugal Smart Objects: Challenges and opportunities. *IEEE Circuits Syst. Mag.* **20**, 4–18 (2020).
55. Banbury, C. R. et al. MicroNets: Neural network architectures for deploying TinyML applications on commodity microcontrollers. *Proceedings of Machine Learning and Systems* **3**, 517–532 (2020).
56. De Prado, M. et al. Robustifying the deployment of TinyML models for autonomous Mini-Vehicles. *Sensors* **21**, 1339 (2021).
57. Lahade, S. V., Namuduri, S., Upadhyay, H. & Bhansali, S. Alcohol sensor calibration on the edge using tiny machine learning (Tiny-ML) hardware. *Meeting Abstracts/Meeting Abstracts (Electrochemical Society. CD-ROM) MA2020-01*, 1848 (2020).
58. Piątkowski, D. & Walkowiak, K. TinyML-Based concept system used to analyze whether the face mask is worn properly in Battery-Operated conditions. *Appl. Sci.* **12**, 484 (2022).
59. Lee, S. I. et al. Enabling stroke rehabilitation in home and community settings: A wearable sensor-based approach for upper-limb motor training. *IEEE J. Translat. Eng. Health Med.* **6**, 1–11 (2018).
60. De Niet, M., Bussmann, J. B., Ribbers, G. M. & Stam, H. J. The stroke upper-limb activity monitor: Its sensitivity to measure hemiplegic upper-limb activity during daily life. *Arch. Phys. Med. Rehabil.* **88**, 1121–1126 (2007).
61. Muscillo, R., Schmid, M., Conforto, S. & D'Alessio, T. Early recognition of upper limb motor tasks through accelerometers: real-time implementation of a DTW-based algorithm. *Comput. Biol. Med.* **41**, 164–172 (2011).
62. Vega-Gonzalez, A., Bain, B. J., Dall, P. M. & Granat, M. H. Continuous monitoring of upper-limb activity in a free-living environment: A validation study. *Med. Biol. Eng. Comput.* **45**, 947–956 (2007).
63. Patton, E. W., Tissenbaum, M. & Harunani, F. MIT App Inventor: Objectives, design, and development. in *Springer eBooks* 31–49 (2019). https://doi.org/10.1007/978-981-13-6528-7_3.
64. Brunnstrom, S. Motor testing procedures in hemiplegia: based on sequential recovery stages. *Phys. Ther.* **46**, 357–375 (1966).
65. Hymel, S. et al. Edge Impulse: an MLOPS platform for tiny machine learning. *arXiv (Cornell University)* (2022) <https://doi.org/10.48550/arxiv.2212.03322>.
66. Zhou, H., Stone, T., Hu, H. & Harris, N. Use of multiple wearable inertial sensors in upper limb motion tracking. *Med. Eng. Phys.* **30**, 123–133 (2007).
67. Nguyen, D., Caffrey, C. M., Silven, O. & Kogler, M. Physically flexible Ultralow-Power Wireless sensor. *IEEE Trans. Instrum. Meas.* **71**, 1–7 (2022).

Author contributions

Conceptualization, X.J. and W.Q.; methodology, X.J., K.Y. and W.Q.; validation, X.J.; formal analysis, N.D.; investigation, F.S.; data curation, X.J.; writing—original draft preparation, X.J.; writing—review and editing, W.Q., K.Y., N.D., F.S. and R.S.S.; visualization, X.J. and W.Q.; supervision, W.Q., K.Y., N.D., F.S. and R.S.S.; W.Q. All

authors have read and agreed to the published version of the manuscript.

Funding

This research received funding by Zhejiang Province philosophy and social science planning project, grant number 20NDJO084YB; Key Research & Development Program of Zhejiang Province, grant number 2023C01041; and Key Research & Development Program of Zhejiang Province, grant number 2021C02012.

Declarations

Competing interests

The authors declare no competing interests.

Ethical approval

The study was conducted in accordance with the Declaration of Helsinki, and approved by the Institutional Review Board of School of Art and Design of Zhejiang, Sci-Tech University (protocol code 1030/2023 and 1030.2023). The experiments for this study had obtained informed consent from all subjects.

Informed consent

Informed consent was obtained from all subjects involved in the study.

Additional information

Correspondence and requests for materials should be addressed to Y.K.

Reprints and permissions information is available at www.nature.com/reprints.

Publisher's note Springer Nature remains neutral with regard to jurisdictional claims in published maps and institutional affiliations.

Open Access This article is licensed under a Creative Commons Attribution-NonCommercial-NoDerivatives 4.0 International License, which permits any non-commercial use, sharing, distribution and reproduction in any medium or format, as long as you give appropriate credit to the original author(s) and the source, provide a link to the Creative Commons licence, and indicate if you modified the licensed material. You do not have permission under this licence to share adapted material derived from this article or parts of it. The images or other third party material in this article are included in the article's Creative Commons licence, unless indicated otherwise in a credit line to the material. If material is not included in the article's Creative Commons licence and your intended use is not permitted by statutory regulation or exceeds the permitted use, you will need to obtain permission directly from the copyright holder. To view a copy of this licence, visit <http://creativecommons.org/licenses/by-nc-nd/4.0/>.

© The Author(s) 2025