University Library



A gateway to Melbourne's research publications

Minerva Access is the Institutional Repository of The University of Melbourne

Author/s:
Sengupta, N;Rao, AS;Yan, B;Palaniswami, M

Title:
A Survey of Wearable Sensors and Machine Learning Algorithms for Automated Stroke Rehabilitation

Date:
2024

Citation:
Sengupta, N., Rao, A. S., Yan, B. & Palaniswami, M. (2024). A Survey of Wearable Sensors and Machine Learning Algorithms for Automated Stroke Rehabilitation. IEEE Access, 12, pp.36026-36054. https://doi.org/10.1109/ACCESS.2024.3373910.

Persistent Link:
https://hdl.handle.net/11343/345382

License:
cc-by

This article has been accepted for publication in IEEE Access. This is the author's version which has not been fully edited and content may change prior to final publication. Citation information: DOI 10.1109/ACCESS.2024.3373910

IEEE Access
Multidisciplinary | Rapid Review | Open Access Journal

Date of publication xxxx 00, 0000, date of current version xxxx 00, 0000.

Digital Object Identifier 10.1109/ACCESS.2024.DOI

A Survey of Wearable Sensors and Machine Learning Algorithms for Automated Stroke Rehabilitation

NANDINI SENGUPTA[1], ARAVINDA S. RAO[1], (Senior Member, IEEE), BERNARD YAN[2], AND MARIMUTHU PALANISWAMI.[1], (Fellow, IEEE)

[1]Department of Electrical and Electronic Engineering, The University of Melbourne, Parkville, Victoria—3010, Australia (e-mail: nsengupta@student.unimelb.edu.au, aravinda.rao@unimelb.edu.au, palani@unimelb.edu.au)
[2]The Royal Melbourne Hospital, Parkville, Victoria—3052, Australia (e-mail: bernard.yan@mh.org.au)

Corresponding author: Nandini Sengupta (e-mail: nsengupta@student.unimelb.edu.au).

This research was supported partially by the Australian Government through the Australian Research Council's Discovery Projects funding scheme (project DP190101248). The views expressed herein are those of the authors and are not necessarily those of the Australian Government or Australian Research Council.

ABSTRACT Stroke is one of the leading causes of disability among the elderly population and is a significant public health problem worldwide. The main impact of stroke is functional disabilities due to motor impairment after stroke. Advances in modern medicine and technology have significantly improved diagnosis and treatment; however, most post-stroke care is based on the effectiveness of rehabilitation. Stroke rehabilitation depends on two main components: (i) training (or therapy) to restore the patient to pre-stroke mobility and (ii) assessing motor functionality of affected patients performing activities to track motor recovery. This article highlights how combining wearable devices and machine learning (ML) produces new pathways for effective stroke rehabilitation. While wearable devices help capture patient movements at much finer time resolutions, ML allows us to build predictive models from wearable data to assist clinicians in diagnosis and treatments. Specifically, we expand on how wearable devices and ML can improve monitoring quality in training intervention, assessment, and remote monitoring. In addition, we provide our main findings from the literature, research challenges, and future directions in post-stroke therapies using wearable devices and ML.

INDEX TERMS Stroke rehabilitation, wearable devices, machine learning, interventions, remote monitoring.

I. INTRODUCTION

STROKE is the second leading cause of death, and 17 million people worldwide suffer from stroke each year [1]. Stroke can have devastating effects, including death or severe disability, which can cause social and family burdens. Even though the majority of the victims are older adults, the number of people 60 years of age or older is projected to increase from an estimated 488 million in 1990 to nearly 1,363 million in 2030 [2]. Since stroke is the leading cause of adult disability in the world and 70–85% of stroke patients have hemiplegia after the first stroke [3], motor recovery is one of the most crucial aspects for stroke victims. Wearable devices, when paired with machine learning (ML), enable us to continuously monitor subjects and ascertain the progressive course towards enhanced motor recovery. It is vital to monitor stroke patients' daily activities and exercise schedules in the absence of a physician to monitor progression regularly. Wearable devices will enable us to collect patient data continuously, which would otherwise lead to missed opportunities for diagnosis and treatment. Combining ML with wearable devices will improve the remote monitoring of stroke patients. Figure 1 provides an overview of stroke monitoring using wearable devices, smartphones, cloud computing, and the Internet of Things.

To improve and regain mobility following strokes, we require two primary components: (i) more practice and training (including physiotherapist instruction sessions and personal practice) and (ii) periodic evaluations (to gauge the impact of training). Therefore, monitoring the performance to assess the progress remotely is essential when the patient is at home

This work is licensed under a Creative Commons Attribution 4.0 License. For more information, see https://creativecommons.org/licenses/by/4.0/

This article has been accepted for publication in IEEE Access. This is the author's version which has not been fully edited and
content may change prior to final publication. Citation information: DOI 10.1109/ACCESS.2024.3373910

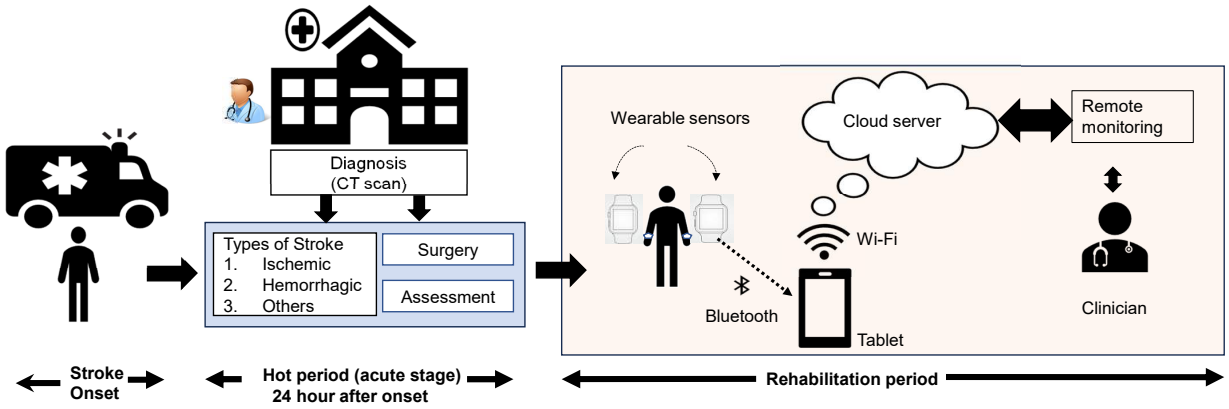


FIGURE 1: Depicts a visionary diagram of wearable sensor-based remote monitoring of stroke patients' activity during rehabilitation. Whenever a person suffers a stroke, they are rushed to the hospital, and a computed tomography (CT) scan is performed to determine the next course of action, including the type of stroke. The rehabilitation period commences 24 hours after the onset of stroke. Wearable sensor-based remote monitoring detects motor recovery when patients engage in exercises in a rehabilitation facility or at home without the assistance of a medical professional. Patients wear a wearable device (wrist-based), and sensors onboard the device record their activities. The data is then transferred to a phone or tablet via Bluetooth. The app will send data to a cloud server using Wi-Fi or internet-enabled devices. The cloud server allows clinicians to retrieve and analyze data to track patients remotely. An alert process can be incorporated if the clinician needs to provide some feedback to patients. The primary advantage of wearable devices for remote monitoring is that they can capture finer time resolutions for near-real-time monitoring and data visualization. Integrating this functionality, specifically fine-time resolution data, with ML enables clinicians to access real-time predictive analytics.

or practicing exercises alone. It is imperative to conduct objective monitoring without any biases or prejudices. No single measure is capable of predicting all dimensions of recovery and disability [4], despite the availability of several stroke assessment scales to measure functional outcomes at each level of the World Health Organization International Classification of Functioning, Disability, and Health (WHO-ICF) [5]. For further details about stroke assessment scales, readers are referred to Table 1 (with 17 scales for impairment), Table 2 (with 16 scales for activity), and Table 3 (with eight scales for participation restriction class) of the included supplementary material.

A comprehensive overview of the numerous techniques employed in stroke therapy for motor recovery is depicted in Figure 2. As shown in Figure 2, we note that training and monitoring (*i.e.*motor evaluation) are crucial components of motor recovery. In addition to conventional physical therapy (manual), six widely utilized rehabilitation training programs are presented in Figure 2. Either in-person or remote evaluations can be employed to evaluate motor recovery. The physical therapist conducts a manual evaluation; however, it is also feasible to perform it remotely using an inexpensive wearable sensor.

Mobile health (mHealth) technology [6] uses smartphone processing power and mobility to provide and support healthcare solutions. The utilization of current cell phone technology and applications enables patients to access their health status by interacting with healthcare providers at any time and location. The primary advantage of mHealth is its ability to provide remote healthcare services, such as medical advice and consultations, through smartphones when patients cannot visit a healthcare center. mHealth has become essential, especially in rural or regional areas with limited access to health facilities. With the rise of smartphone applications, it is now possible for people to download and use easily usable applications to record, monitor, and report different aspects of their health to relatives and health professionals. Smartphones and other health monitoring tools can help provide information on different health-related parameters, such as physical activity, body mass index, heart rate, and blood pressure. The apps can be used to promote physical activity and motivate individuals to engage in more physical activity [7]–[12].

Furthermore, the data from these apps can be shared with clinicians to monitor the condition for consultation and treatment. *Ambient Intelligence* (AmI) is another related technology that uses sensors and artificial intelligence to create an environment that can respond to needs, for example, by adjusting the air conditioning according to the temperature [13, 14]. However, it can be challenging to monitor stroke patients with smartphones because of health conditions. Hence, portable devices, such as the form factor of a wristwatch, are more appealing to observing stroke subjects.

Wearable devices coupled with ML will significantly aid in assessing patients' motor recovery status by automatically collecting continuous data objectively and inferring progress conditions. Patients can wear wearable devices to gather data relating to their health, and the sensors on such devices

This work is licensed under a Creative Commons Attribution 4.0 License. For more information, see https://creativecommons.org/licenses/by/4.0/

This article has been accepted for publication in IEEE Access. This is the author's version which has not been fully edited and
content may change prior to final publication. Citation information: DOI 10.1109/ACCESS.2024.3373910

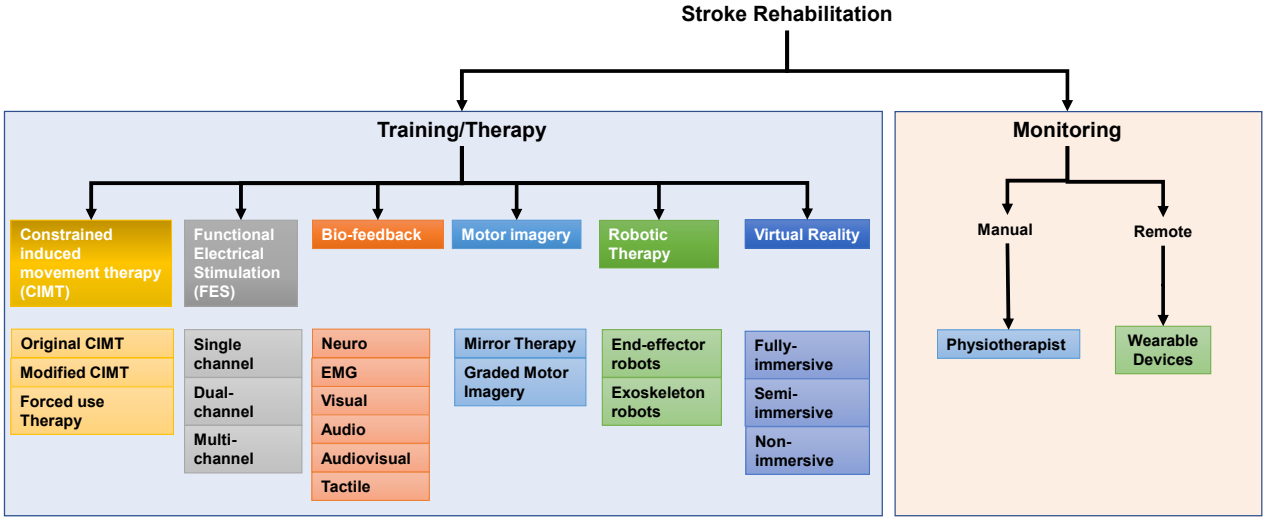


FIGURE 2: Shows the overview of different methods used in stroke rehabilitation for motor recovery. Two main components of stroke rehabilitation for motor recovery are training and motor assessment. In training, six common rehabilitation training schemes are shown in addition to conventional physical therapy. Manual or remote assessments can be performed to assess motor recovery. A physical therapist performs a manual evaluation, and remote evaluation can be performed using a low-cost wearable sensor. This article will focus on the use of wearable devices for stroke rehabilitation.

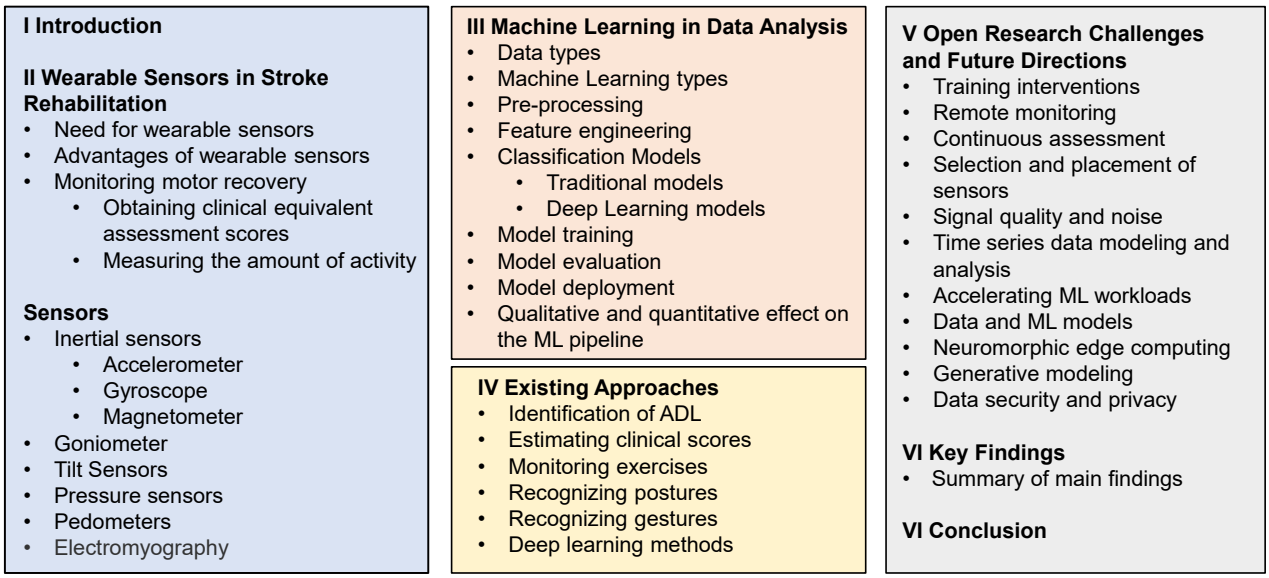


FIGURE 3: The figure shows this article's different sections and outlines.

will monitor various activities. While wearable devices can capture movements at finer time resolutions, signal processing helps us analyze wearable data, and ML helps us build models to help clinicians diagnose and treat diseases. Therefore, wearable devices and ML present a suitable means of remotely analyzing the wearable data of stroke patients and obtaining details on their motor recovery status. This approach enables the creation of a continuous record of movements and a diary for the resulting assessments.

In this article, we concentrate on various wearable devices and machine-learning techniques to monitor stroke patients' motor recovery during rehabilitation. Specifically, the methods by which they can enhance the quality of monitoring in training intervention, assessment, and remote monitoring (as outlined in Section V). Figure 3 provides a summary of the main sections of this article. Our review differs from a recent review paper [15] in examining the advantages and limitations of various wearable sensors extensively utilized in stroke rehabilitation. Furthermore, we provide open research challenges and prospective directions in diverse aspects of stroke rehabilitation research, which are absent in [15].

Therefore, this review aims to help:

This work is licensed under a Creative Commons Attribution 4.0 License. For more information, see https://creativecommons.org/licenses/by/4.0/

This article has been accepted for publication in IEEE Access. This is the author's version which has not been fully edited and
content may change prior to final publication. Citation information: DOI 10.1109/ACCESS.2024.3373910

- determine the type of remote monitoring that is most relevant and objective
- determine which sensors are most relevant and best suited to specific problems
- identify deficiencies in the current techniques and approaches
- compare the results of various ML models to pick the most effective ML algorithms
- suggest deep learning models suitable for real-time outcomes by considering cost, power, time, and model complexity requirements.

II. WEARABLE SENSORS IN STROKE REHABILITATION

This section outlines the significance of wearable sensors in stroke rehabilitation, their methodologies, the monitoring of motor recovery progress, and the diverse types of sensors utilized in wearable devices.

A. NEED FOR WEARABLE SENSORS

During rehabilitation, different therapeutic exercises are recommended for patients, who are asked to try some daily activities (motor tasks). In hospital rehabilitation, healthcare professionals evaluate patients' motor abilities and provide scores based on assessment scales. This evaluation procedure is conducted regularly at crucial intervals. Nonetheless, these assessment scales possess limited predictive capability [4]. Since this periodic monitoring is not continuous, we do not receive an accurate estimate of motor activity progress, leading to missed opportunities and early interventions. This is where wearable sensors play a crucial role in facilitating the continuous monitoring of individuals in hospitals, rehabilitation facilities, and homes.

B. ADVANTAGES OF WEARABLE SENSORS

In stroke rehabilitation, the use of wearable sensors presents two significant advantages. First, utilizing a wearable device equipped with appropriate sensors and algorithms coupled with an internet connection enables the real-time monitoring of subjects. Second, the small form factor of these wearable devices makes them comfortable for patients to wear without causing discomfort or inconvenience, encouraging them to participate in rehabilitation exercises and activities actively.

C. MONITORING MOTOR RECOVERY

Two approaches are used to monitor motor recovery through wearable sensors: 1) obtaining clinical equivalent assessment scores by detecting tasks, and 2) measuring the amount of activity performed by the affected part.

1) **Obtaining clinical equivalent assessment scores by detecting tasks**. Clinical scores are calculated remotely (without a clinician) using data recorded on wearable devices [16, 17]. Sensors acquire movement data from the affected body parts while executing prescribed tasks at a rehabilitation facility or home. Using this sensor data, patterns can be detected for different activities. It is possible to convert the data later into another index representing the patient's status, similar to the assessment score. The newly released index is transmitted to the healthcare professional via either an Internet web application or a smartphone application, enabling the healthcare professional to guide patients residing in remote locations. For instance, in a hospital setting, equivalent National Institutes of Health Stroke Scale (NIHSS) scores were calculated using accelerometer sensors to assess motor recovery during the initial 24 hours following a stroke [18].
2) **Measuring the amount of activity carried out by the affected part**. The total amount of exercise the affected body part performs is determined by wearable sensor data. Different movement metrics are calculated, such as the complete linear or angular movement, the distance covered, and the amount of daily use of the affected body part [19, 20]. For instance, a radial basis function (RBF) was employed to correlate differential magnetometer readings from a "manumeter" to wrist and finger joint angles [19]. These techniques are not only cost-effective and remotely accessible (*i.e.*, time-efficient) but also provide objective methods to measure body movements and calculate scores.

D. SENSORS

Table 1 depicts wearable sensors commonly used in healthcare applications, capabilities, and limitations. It presents the eight sensors widely used with wearable devices, including an accelerometer, gyroscope, magnetometer, goniometer, tilt sensor, pressure sensor, pedometer, and electromyography (EMG) sensor. Each sensor measures a particular physical phenomenon (second column of Table 1), which means it also has its limitations (third column of Table 1). For instance, accelerometers can measure acceleration, but they cannot measure angular velocity, which can be accomplished by utilizing gyroscope sensors. Depending on the problem, many studies use single sensors, while others use a combination of sensors that complement each other to capture the desired signals. For further details on the use and applications of wearable devices in healthcare applications, we encourage readers to refer to [25]. The following subsections provide a brief overview of each sensor and its primary application in stroke rehabilitation.

1) Inertial Measurement Units (IMUs)

Micro-electromechanical systems (MEMS) facilitate the miniaturization of electrical and mechanical components. These components are fabricated into integrated circuits (ICs) with a form factor ranging from a few micrometers to millimeters [26]. MEMS technology allows for a smaller form factor while reducing costs and requiring low power. These characteristics make MEMS-based sensors particularly attractive for wearable devices. The three commonly used IMUs are:

This work is licensed under a Creative Commons Attribution 4.0 License. For more information, see https://creativecommons.org/licenses/by/4.0/

This article has been accepted for publication in IEEE Access. This is the author's version which has not been fully edited and
content may change prior to final publication. Citation information: DOI 10.1109/ACCESS.2024.3373910

TABLE 1: List of eight wearable sensors commonly used to detect motor activities in stroke monitoring and rehabilitation, along with their measures and limitations.

Sensor	Measure	Limitations
Accelerometer	Accelerations or decelerations, velocity, and displacement	(i) Unable to differentiate between (i) active and passive movements, and (ii) loaded and unloaded activities [17].
Gyroscope	Angular velocity, rotation	Sensor values drift over time. Therefore, additional mechanisms in hardware or software must be needed to calibrate the sensors regularly. Drift correction algorithms may also be needed.
Magnetometer	Spatial orientation	Ferromagnetic materials can cause distortions in measurements. When a subject is in motion, the measurement may be affected when the subject is near a ferromagnetic material.
Goniometer	Range of motion	The presence of a fixed hinge imposes a fixed center of rotation when measuring the bending angle in some joints, leading to inaccurate results [21]
Pressure	Force and pressure distribution	The accuracy of measurements is dependent on hysteresis, linearity, temperature sensitivity, pressure range, sensor detection area, operating frequency, creep, and repeatability [22]
Pedometer	Number of footsteps	(i) Measures motion in only one plane, which may not be sufficient for measuring stroke recovery progress (ii) Inaccurate in measuring sedentary behavior and missing data [23]
Electromyography (EMG)	Muscle activity	Activation level of serratus anterior during isometric abduction and flexion, dynamic abduction and flexion, and bench press is not reliable [24]
Tilt	Angle of tilt	(i) The force-balanced sensor is expensive (ii) Micro-electromechanical systems (MEMS) based tilt sensors have very high thermal coefficients, which can damage the physical as the temperature increases. (iii) The vibration of sensors tends to create erratic or random signals. (iv Fluid-based sensors have limited response time, whereas mercury-based sensors are toxic.

- **Accelerometer** sensors are used in wearable devices to measure acceleration (the rate of velocity change). Accelerometer sensors typically measure three-dimensional accelerations (x, y, and z) [27] and are capable of measuring up to $\pm$ 16 g [28].
- **Gyroscope** sensors are utilized to measure angular velocity, commonly measured in degrees per second or revolutions per second [27]. Gyroscopes are advantageous instruments for determining orientations and evaluating rotations on three axes.
- **Magnetometer** is used to measure magnetic fields and fluxes in the Earth's magnetic field [27].

Accelerometers are limited to linear accelerations, whereas gyroscopes can measure rotations and orientation. Therefore, it is possible to combine accelerometers and gyroscopes to obtain a complete picture of object movement in 3D space. This combination is called an IMU and has several applications, including stroke monitoring [18] and epilepsy monitoring [29]. Adding a magnetometer allows us to estimate the direction, which is helpful in many applications, such as navigation systems [30] and sleep monitoring [31].

IMUs possess significant practicality in evaluating motor function in individuals who have undergone strokes. Numerous researchers have utilized accelerometers in Upper Extremity (UE) studies [16, 32]–[34] due to their ability to provide reliable and valid objective measurement and their potent psychometric attributes [35]. For instance, accelerometers were employed to assess arm usage [33], calculate Functional Ability Scale (FAS) scores from accelerometer data [16], acquire gait patterns [36], and document Activity of Daily Living (ADL) [34]. Further, IMUs were used to determine the body's specific force and angular rotation rate [37]. A wristwatch device and a magnetic ring worn on the index finger ("manumeter") are utilized to calculate the total angular distance traveled by the wrist and finger joints [19].

2) Goniometer

A goniometer is an instrument used to measure a joint's angle (range of motion) and is commonly used in orthopedics to measure the angles of the joints [38]. In rehabilitation, wearable goniometer-equipped systems aid in reconstructing human posture, providing support both in clinics and at home [39]. A sensing glove with three knitted piezoresistive fabric (KPF) goniometers was used to track the flexion and extension movement of the metacarpophalangeal joint of the thumb, index, and middle fingers [40].

3) Tilt sensors

Tilt sensors, also called inclinometers, measure the angle of gravity. Compared to goniometers, inclinometers possess a higher degree of reliability [41], and studies have demonstrated their superiority in detecting intricate movements [42]. Since inclinometers use gravity, they cannot measure angles in a horizontal plane.

4) Pressure Sensors

Pressure sensors measure the applied pressure in the sensing area, and force resistive sensors (FSRs) are commonly used to detect physical tension, squeezing, and contact. Examples include detecting gait patterns with FSR sensors embedded in shoes [43], gait abnormalities [44], and monitoring of daily life among stroke survivors [45]. An FSR can monitor multiple major muscle groups that govern the movements of the hand and wrist [46]. It has been demonstrated that a

This work is licensed under a Creative Commons Attribution 4.0 License. For more information, see https://creativecommons.org/licenses/by/4.0/

This article has been accepted for publication in IEEE Access. This is the author's version which has not been fully edited and
content may change prior to final publication. Citation information: DOI 10.1109/ACCESS.2024.3373910

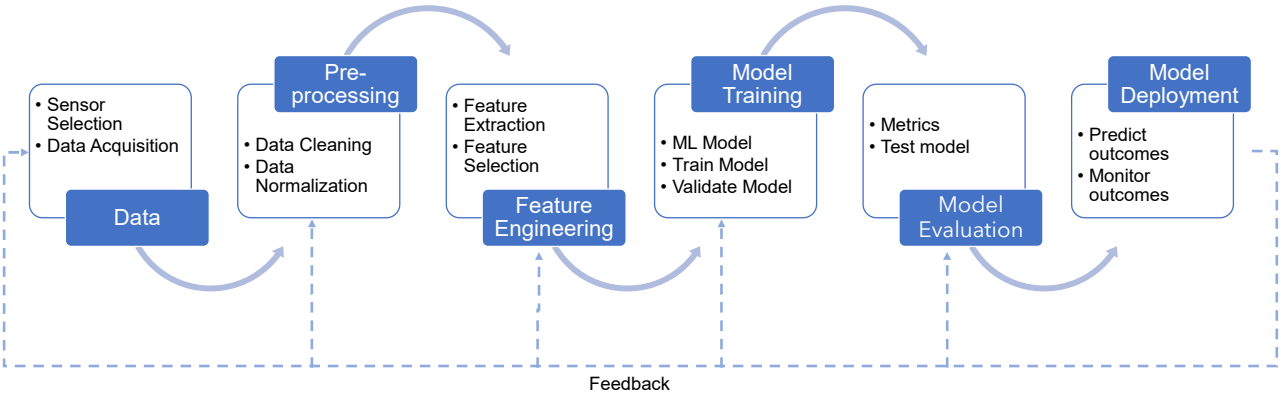


FIGURE 4: An overall ML pipeline comprises six steps, from data collection to building models to model deployment. Data is collected from sensors and equipment, and the first step is capturing electronic health records (EHR). In the next step, the data is preprocessed to remove unwanted noise and missing values, then normalized. The third step is feature engineering, which comprises feature extraction and selection. The fourth step is building the appropriate ML model, training the model, and validating the model. In the fifth step, the model's performance is measured through various metrics using a test dataset. In the final step, the model is deployed to an edge server, cloud server, or wearable device for real-time predictions. The model's performance is constantly measured, and the feedback is used to improve the processes and the model's performance.

combination of FSR and gyroscopes can accurately detect precise gait patterns [47]. Furthermore, a gait shoe consisting of an IMU, FSR, and electric field height sensors was used to quantitatively analyze gait parameters (heel strike and toe-off) between healthy people and patients with Parkinson's disease [48]. In stroke patients, shoe-based accelerometers and FSR sensors were used to detect temporal gait parameters to differentiate posture allocation and activities [49].

5) Pedometers

A pedometer is used to measure the number of steps an individual takes and to record the vertical acceleration [50]. Wearable pedometers often use a 3-axis MEMS-based accelerometer to estimate 3D motion patterns and deduce the number of steps [51]. The pedometer is a more valid method for measuring ambulatory activity in stroke patients [52]. It proved advantageous in objectively assessing the number of steps walked during walking and describing the pattern and intensity of an activity [53].

6) Electromyography (EMG)

EMG measures electrical activity in muscles and can detect activities such as moving fingers or clenching the fist. There are primarily two types: intramuscular EMG (iEMG) and surface EMG (sEMG) [54]. Intramuscular EMG is an invasive technique in which needles are inserted through the skin to measure electrical signals. On the other hand, sEMG is a non-invasive technique that measures electrical signals from the skin's surface. Due to the non-invasiveness of sEMG, it is commonly used in wearable systems [55]. The accelerometer and sEMG sensors were combined to monitor ADL [17] and provide feedback during ADL [32]. Additionally, they were used to assess the quality of motor performance in home rehabilitation so that appropriate feedback could be provided to promote high-quality exercises.

Until now, we have discussed commonly used sensors in wearable devices. The next section, provides an overview of ML steps in stroke rehabilitation, including data types, ML types, and ML models.

III. MACHINE LEARNING IN DATA ANALYSIS

In this section, we discuss various data types used in ML, types of ML, pre-processing steps, feature engineering, ML models, training of the models, evaluation of the models, deployment of the models, and qualitative and quantitative effects of various steps on the outcomes. The overview of these steps is shown in Figure 4.

A. DATA TYPES

Different methods are used to collect information from patients to provide the appropriate medical attention in medical establishments. Sensor data is primarily *quantitative* (numerical or a natural number). On the contrary, the data gathered through electronic health records (EHR) is predominantly *qualitative* (categorical) and typically comprises a digital patient history form. The EHR is a longitudinal data collected over time. It contains information about age, gender, demographics, problem history, immunizations, physician's observations, reports, and laboratory results. It is also known as electronic medical records (EMR) and computer-based patient records (CPR) [56, 57]. Figure 5 shows a graph depicting the data collection in the medical field is divided into four different categories: discrete, continuous, nominal, and ordinal. The *numerical* (quantitative) data is categorized into *discrete* (such as the number of hospital visits) and *continuous* (such as the weight of patients). On the contrary,

This work is licensed under a Creative Commons Attribution 4.0 License. For more information, see https://creativecommons.org/licenses/by/4.0/

This article has been accepted for publication in IEEE Access. This is the author's version which has not been fully edited and
content may change prior to final publication. Citation information: DOI 10.1109/ACCESS.2024.3373910

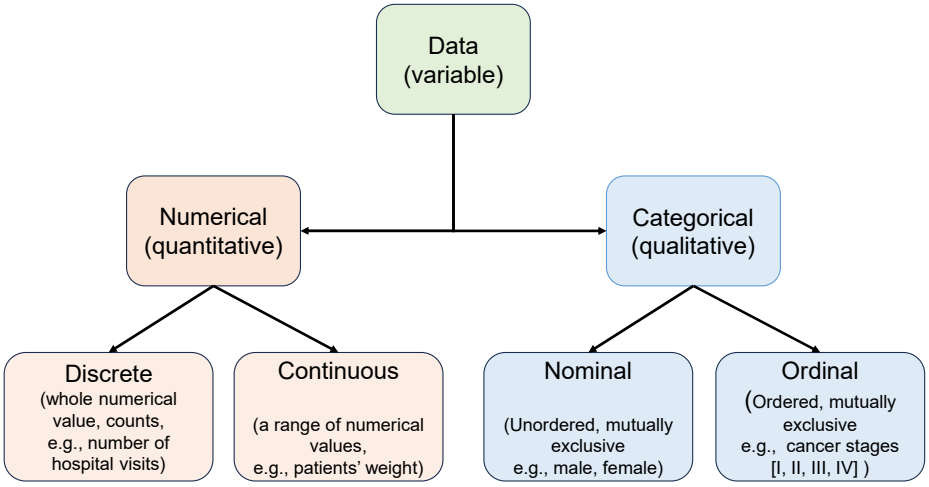


FIGURE 5: The figure illustrates how data collected in the medical domain can be classified into four types: discrete, continuous, nominal, and ordinal. The *numerical* (quantitative) data is categorized into discrete (such as the number of hospital visits) and continuous (such as the weight of patients). On the other hand, *categorical* (qualitative) data can be partitioned into nominal (unordered data, such as male or female) and ordinal (ordered data, such as the cancer stages [I–IV]).

categorical (qualitative) data can be classified into *nominal* (unordered data, such as male or female) and *ordinal* (ordered data, such as cancer stages [I–IV])). These data types are crucial in defining ML problems and creating appropriate models.

B. MACHINE LEARNING TYPES

ML is a field of artificial intelligence (AI) concerned with performing tasks by learning patterns from the dataset and generalizing without being specifically programmed. In other words, ML algorithms try to automate tasks by improving the learning process without human participation. There are four types of ML: (i) supervised learning, (ii) unsupervised learning, (iii) semi-supervised learning, and (iv) reinforcement learning. Figure 6 shows the four types of ML. Below, we briefly describe the four types of ML paradigms [58]:

1) In **supervised learning**, the ML algorithm is presented with a dataset whose input (feature vector) and the desired output (label) are explicitly provided. The algorithm learns to map the input to the output by learning the hidden patterns. Suppose we have N labeled examples: $\{(\mathbf{x}_i, y_i)\}_i^N$, where $\mathbf{x}_i$ is a vector input and y_i is a (scalar) output. The label y_i can be a real number (*regression problem*) or a finite set of classes $\{0, 1, 2, \ldots, C\}$ (*classification problem*). Then, the supervised learning algorithm uses the dataset to generate a model that takes x as input and delivers output labels (y). The desired output is also called *target variable*. Examples of supervised classification algorithms include Support Vector Machine (SVM) [59], k-nearest neighbors (kNN) [60], Naive Bayes [61], decision trees (DT) [61], random forest (RF) [61], and neural networks (NN) [62]. Examples of supervised regression algorithms [61] include linear, non-linear, Bayesian, polynomial regression, ensemble methods, and NN.
Supervised learning is beneficial for learning from data and making predictions. For example, data collected from wearable devices during stroke rehabilitation can be used to build a personalized ML model for individuals and subsequently use these models to track and predict the quality of exercises during remote monitoring.
2) In **unsupervised learning**, we do not have specific labels, *i.e.*, the dataset will be a collection of unlabeled examples $\{(\mathbf{x}_i)\}_i^N$. The goal here is to build a model that transforms the input data (or a feature vector) into another vector or value such that it helps to solve the problem at hand. For example, in *clustering* problems, unsupervised learning can be used to find clusters in the data so that it helps to make decisions. Another application of unsupervised learning is to find *associations* between variables in large databases. Examples of unsupervised algorithms include K-means [63], fuzzy c-means [64], hierarchical [65], Density-based spatial clustering of applications with noise (DBSCAN) [66], Gaussian Mixture Model (GMM) [67], Principal Component Analysis (PCA) [68], and NN.
Unsupervised learning is advantageous when we do not have the labels for the data and want to explore the data for hidden patterns. For example, data from wearable devices without labels can be extracted and clustered according to stroke severity level and exercises during rehabilitation, establishing connections between subjects and exercises.
3) In **semi-supervised learning** [69], the dataset contains both labeled and unlabeled examples. Ordinarily, the number of unlabeled examples will be much higher

This work is licensed under a Creative Commons Attribution 4.0 License. For more information, see https://creativecommons.org/licenses/by/4.0/

This article has been accepted for publication in IEEE Access. This is the author's version which has not been fully edited and
content may change prior to final publication. Citation information: DOI 10.1109/ACCESS.2024.3373910

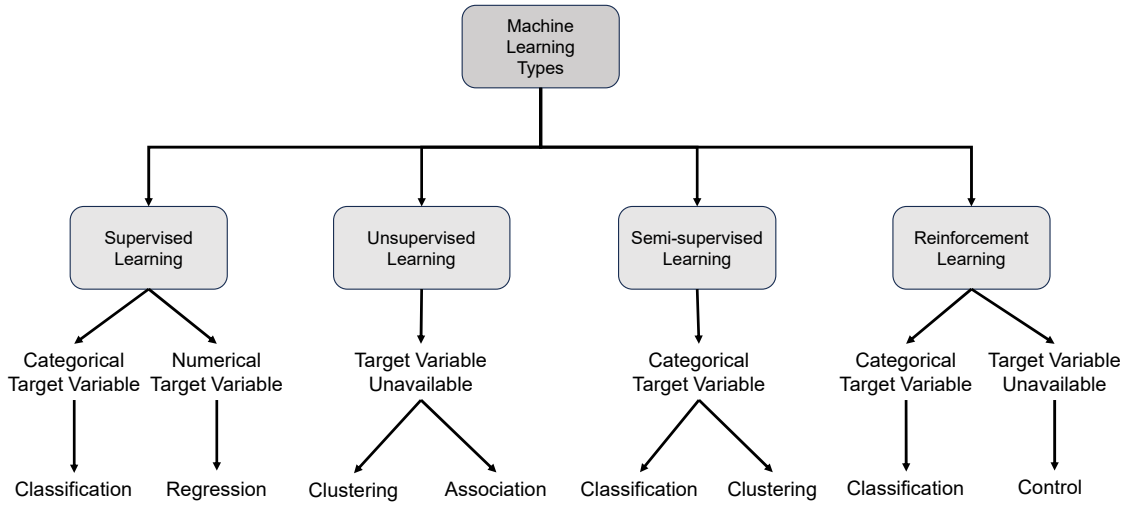


FIGURE 6: The figure illustrates the four types of ML: (i) supervised learning, (ii) unsupervised learning, (iii) semi-supervised learning, and (iv) reinforcement learning. Depending on the target variable, the supervised learning can be further categorized into classification (categorical) and regression (numerical, real number). In the case of unsupervised learning, the target variable is unavailable. However, we can perform clustering and association-type problems. In the case of semi-supervised learning, it primarily handles categorical target variables, which can be grouped into classification and clustering problems. Reinforcement learning deals with categorical target variables to manage classification problems. When the target variable is unavailable, the problems usually fall into the category of control (*e.g.*, self-driving cars)

.

than labeled examples. The objective here is similar to supervised learning, *i.e.*, to produce a model that learns not only from the labeled examples but also uses unlabeled examples. The idea here is that the unlabeled examples add more uncertainty, and this helps to generate better models. Examples of semi-supervised learning algorithms include Expectation-Maximization [70] and semi-supervised GMM [71].

Semi-supervised learning becomes extremely useful when we have limited labeled data. For example, time-series data from stroke rehabilitation exercise sessions may consist of data of a few minutes for each exercise for each subject. In such a scenario, semi-supervised learning can be used to train a supervised learning model from the unlabeled data by exploring temporal associations between exercises and segments of data.

4) In **reinforcement learning** [72], an *agent* (ML algorithm) interacts with an *environment* and is capable of comprehending the *state* of the environment. The agent can execute *actions* in all states. Different actions bring different rewards, and the agent can move to a different environment. A *policy* is a function that accepts a state's feature vector as input and produces the best course of action to take while in that state, much like the model in supervised learning. The goal of the RL is to learn a policy, which is optimal if a course of action maximizes the *expected average reward*.

Reinforcement learning is beneficial when the decision-making is sequential to solve the problem over the longer term. For example, robots can learn to interact with stroke victims to help improve motor recovery in the stroke rehabilitation process. Examples of RL algorithms include Trust Region Policy Optimization (TRPO) [73], Proximal Policy Optimization (PPO) [74], Q-learning [75], and Deep Q Neural Network (DQN) [76].

C. PRE-PROCESSING

This subsection provides a brief overview of the steps involved in the ML approach to data analysis. Once sensor data is collected, the next step is to separate activity types and provide a comparative index equivalent to a clinical score. As shown in Figure 4, pre-processing steps are performed depending on the acquired data (after collecting the data). The data may consist of noise and artifacts. Typical pre-processing efforts include centering and scaling (zero mean with unit standard deviation), removing outliers, and handling missing values [77]. The condition of the data determines whether these steps are required.

D. FEATURE ENGINEERING

A *feature* (also known as a predictor) compactly represents the sensor data and reveals its distinguishing characteristics. The fundamental steps in feature engineering consist of *feature extraction* and *feature selection*.

Feature extraction transforms the input data (feature vector) into a different feature space for better model performance. Feature extraction requires a solid understanding of

This work is licensed under a Creative Commons Attribution 4.0 License. For more information, see https://creativecommons.org/licenses/by/4.0/

This article has been accepted for publication in IEEE Access. This is the author's version which has not been fully edited and
content may change prior to final publication. Citation information: DOI 10.1109/ACCESS.2024.3373910

the problem, domain knowledge, and related aspects of the problem. Often, new features are created from existing ones to improve ML model performance. The features extracted from the sensor data can be primarily classified into five groups (or domains): (1) time domain, (2) frequency domain, (3) statistical features, (4) morphological features, and (5) data-specific features.

1) **Time domain features** encompass the characteristics extracted directly from the time series data (either raw or preprocessed), capturing the temporal patterns and dynamics without transforming them into alternative domains.
2) **Frequency domain features** are obtained by transforming time series data into the frequency domain, revealing information about the distribution of frequency components and their corresponding magnitudes.
3) **Statistical features** provide the quantitative measures calculated from valuable insights into data distribution, its central tendency, and variability.
4) **Morphological features** capture the shape and structural information of data, focusing on the time series' patterns, trends, or structural characteristics.
5) **Data-specific features** are patterns extracted from specific datasets, tailored to the characteristics of that particular dataset, rather than being generalized for all datasets. For example, a velocity sequence can be extracted as a data-specific feature in accelerometer sensor data.

On the other hand, *feature selection* deals with selecting the best k features to improve the model's performance and help with the models' interpretability [61]. *Feature selection* (also known as *subset selection*) helps reduce redundancy of features and time complexity [61]. There are three main approaches to feature selection [78]: (1) filter methods, (2) wrapper methods, and (3) embedded methods.

1) **Filter methods** use various statistical tests to determine their correlation with the classes and then rank these features. Some frequently used statistical tests include the $\mathcal{X}^2$ test, Fisher's exact test, the Euclidean distance, Pearson's correlation, information gain, and others.
2) **Wrapper methods** select features by eliminating features that do not contribute (unimportant or less significant) by using the performance (typically of classification algorithms). As a result, wrapper methods depend on classification algorithms to select the best features. Some algorithms include sequential forward and backward selection and genetic algorithms.
3) **Embedded methods** select features by dynamically adjusting the weights given to each part during training, and this mechanism is built into the classifier. Some techniques include tree-based classification (*e.g.*, decision trees, random forests) and regularization models (ridge regression, most minor absolute shrinkage, and selection operator).

Both feature extraction and selection are part of feature engineering. Furthermore, procedures such as dimension reduction may be applied for specific problems, which require projecting the data into a lower-dimensional space (using PCA) and reducing the correlation among features.

E. CLASSIFICATION MODELS

The literature shows that most machine-learning models developed using wearable sensors in stroke rehabilitation tend to address classification problems. In this article, we have identified five application areas concerning stroke rehabilitation. They are (i) identifying ADL, (ii) estimating clinical scores, (iii) monitoring exercises, (iv) recognizing postures, and (v) recognizing gestures. These approaches are reviewed in Section IV. These application areas use various data types, but most target categorical outputs. Therefore, we focus on classification models in this article.

Classification models are mathematical models designed to make decisions on a new observation (*e.g.*, sensor data) and categorize the outcome based on the learned features of the collected data. For example, classifiers can help clinicians make an informed decision on whether the patient is recovering or deteriorating after treatment by using wearable devices [18]. The selection of models depends mainly on the type of feature, the complexity of the task, and the sample size. After the model is trained using training data, it is evaluated, validated, and calibrated before deployment to predict new test data.

In this subsection, we provide a high-level description of the most commonly used *traditional classifier models* (k-nearest neighbors, SVM, RF, and k-means algorithms) and the most recent *deep learning models* (convolutional neural network, recurrent neural networks, transformers). The reader is encouraged to follow the relevant references for more detailed explanations.

a: Traditional Models:

Most models in this category require a set of selected features to be used as input to build classification models. The previous subsection III-D presented a brief overview of feature extraction, where features are hand-crafted. The hand-crafted features may or may not be good, as this is subjective. Therefore, the classification performance may suffer.

K-nearest neighbors (**K-NN**) is a supervised algorithm used for classification. Suppose that we have a set of data points $(x_1, y_1), (x_2, y_2), \ldots, (x_n, y_n)$, where x_i is the data value and y_i are the class labels. The $k-$NN algorithm classifies a new data point x_j into one of the y_i classes using a distance metric that minimizes the distance to one of the existing data points x_i. Distance metrics can include Euclidean, City Block, Manhattan, and others [61]. The parameter k determines the number of neighbors it looks for while deciding the label for the new data point. It uses majority voting to make the final class label for this new data point. For example, with $k = 5$, the algorithm looks for the five closest data points of x_j and assigns the labels based on most of the labels y_i of these 5 data points. It is one of the

This work is licensed under a Creative Commons Attribution 4.0 License. For more information, see https://creativecommons.org/licenses/by/4.0/

This article has been accepted for publication in IEEE Access. This is the author's version which has not been fully edited and
content may change prior to final publication. Citation information: DOI 10.1109/ACCESS.2024.3373910

simplest supervised learning algorithms. It performs faster than other models as it has no sophistication (*i.e.*, explicit training step) to classify a new data point.

Support Vector Machine (SVM) is a supervised learning algorithm for binary and multi-class classification. The idea behind the SVM is to find the hyperplane so the distance from the closest points (that belong to two classes in the binary) to the hyperplane is maximized and is given by [79]

$$\min_{\omega,b,\xi} \frac{1}{2}\|\omega\|^2 + \frac{C}{n}\sum_{i=1}^{n} l(\xi^k), \tag{1}$$

subject to

$$y_i(\omega \cdot \mathbf{x}_i + b) \geq 1 - \xi_i, \forall \quad i \in 1, \ldots, n, \tag{2}$$

where $\mathbf{x}_i$ is a feature vector, n is the number of feature vectors, b is the bias term, ω is the weight vector, C is a positive regularization constant and ξ is the slack term. The points close to the decision boundary are called *support vectors*. In cases where the data is not linearly separable using the hyperplane, the kernel techniques allow for class separation by projecting the data points into a high-dimensional space. SVM can be trained with kernels, including linear, polynomial, radial basis function (RBF), and Sigmoid [80]. The RBF kernel is often used, as it allows the points to be separated into different classes and is given by [79]

$$K(\mathbf{x}, \mathbf{x}') = e^{\left(-\frac{\left\|\mathbf{x}-\mathbf{x}'\right\|^2}{2\sigma^2}\right)}, \tag{3}$$

where $K(\cdot,\cdot)$ is the RBF kernel, $\mathbf{x}$ and $\mathbf{x}'$ are feature samples in the input space and a free parameter σ. The grid search optimization algorithm [81] provides the optimal positive regularization constant C and the kernel parameter σ. SVM can handle high-dimensional data and is very effective in cases where the number of predictors exceeds the observations. SVM is primarily used for classification problems, and the binary classification can be extended to multi-class classification problems. However, a variant of SVM called the *Support Vector Regression (SVR)* can be used for regression problems.

Random Forest (RF) is a supervised learning algorithm that combines a collection of tree-structured classifiers to classify a new data point [82]. Predictor trees are created by randomly sampling data points with replacements from the training dataset (also referred to as *Bootstrap Aggregation* or *Bagging*). The decisions of each tree are aggregated (majority vote) for the new data point, that is, the predicted class [61]. Bootstrap Aggregation helps to reduce variance by building uncorrelated trees by selecting random samples and then splitting them into binary trees. There are several hyperparameters, such as the depth of the tree, the number of leaf nodes, and many others. RF can be used both for classification and regression problems. We urge the readers to refer to [61, 83] for more details.

K-means [67] is a clustering algorithm that assigns one of the k labels to unlabeled data points in the dataset. Suppose we have a dataset consisting of N observations: $\{\mathbf{x}_1, \mathbf{x}_2, \mathbf{x}_3, \ldots, \mathbf{x}_N\}$, where $\mathbf{x}$ is a d-dimensional vector. Then, our goal is to partition the dataset into k clusters. This is primarily used in exploratory data analysis. We encourage the readers to refer to [67] for further details.

b: Neural Networks and Deep Learning:

In contrast to traditional classification models that employ manually designed features, modern deep learning models use artificial neural networks (ANN) to automatically learn effective representations (attributes) hidden deep in the data. In addition, as the dataset grows, it becomes increasingly challenging to identify good features from complex datasets using handcrafted features. Therefore, deep learning models tend to have a better generalization power (better predictions for unseen samples) than traditional classification models, leading to the widespread adoption of algorithms in the recent past. Neural networks, the fundamental building block of deep learning models, began in the 1950s with the creation of the perceptron [84]. In the 1980s and 1990s, many techniques were invented, such as the backpropagation algorithm [85], which is the powerhouse of modern deep learning algorithms; however, they became more prominent after the 2012 research work by Krizhevsky *et al.* [86], especially with the availability of large data sets and computing facilities.

ANNs use *artificial neurons* (modeled like biological neurons) to create an artificial brain-like structure with links between artificial neurons, representing synapses in the physical brain. Artificial neurons are often organized as layers with connections between them. ANNs are also called as neural networks (NN) in short. We provide our data (to be modeled) to the first layer (or the input layer) and the expected output to the last layer (the output layer). We then use the backpropagation algorithm and various mathematical optimization techniques to *learn the weights* (of the links between artificial neurons) so that the difference between the expected output and the predicted output by the NN is minimized. The crucial part of these deep learning models is learning the proper weights, and the optimization algorithms adjust these weights through various mathematical procedures. ANNs can be used both for classification and regression problems. We recommend the readers refer to [62, 87] for more information.

Convolution Neural Network (CNN) is a class of ANN, and the name "convolution" is due to a mathematical convolution operation. NNs implement convolution layers, which are a series of convolution filters. CNNs are widely used in image processing, natural language processing, and time series data [87]. CNNs are robust to noise, extract deep features, and are equivariant (*i.e.*, they provide the correct output corresponding to changing input in time-series data, independent of time). Although the CNNs were primarily made for non-sequential data, they sometimes perform well on specific applications. This makes CNN very useful for

This work is licensed under a Creative Commons Attribution 4.0 License. For more information, see https://creativecommons.org/licenses/by/4.0/

This article has been accepted for publication in IEEE Access. This is the author's version which has not been fully edited and
content may change prior to final publication. Citation information: DOI 10.1109/ACCESS.2024.3373910

time series classification. For example, CNNs have been used in activity recognition [88], detection of atrial fibrillation from electrocardiography (ECG) signals [89], stroke rehabilitation [90, 91], and many others.

Recurrent Neural Network (RNN) is another class of ANN where they are explicitly geared to handle sequential input (such as text, audio, and video). RNNs incorporate a cyclical process between neurons, allowing the display of dynamic temporal behavior. This makes RNNs highly useful for time series data and applications such as speech recognition, language translation, handwriting recognition, and many other applications [62, 87].

Long Short-Term Memory (LSTM) model is a type of RNN that uses "short-term memory" and long-term dependencies of RNNs and was introduced in 1997 [92]. They use efficient gradient-based learning techniques to reduce the longer training time (owing to slow changes in weight) of the default RNNs by repeating backpropagation [92]. Therefore, LSTMs avoid the problem of vanishing gradients, commonly seen in deep NNs [93], and are used in applications such as robot control, music composition, time series prediction, human action recognition, and others. We encourage readers to refer to [94] for a comprehensive treatment of LSTMs and their extensions.

Transformer is a sequence-to-sequence model that employs the *attention mechanism* [95], allowing the models to capture dependencies independent of the distance between input and output. Transformer models avoid the recurrent relations present in RNNs and LSTMs. Instead, they use attention mechanisms to calculate global dependencies between the information and work to provide the predictions. We urge the readers to go through [95] for further details on the attention mechanisms and how they are used in sequence modeling applications. Furthermore, the reader can refer to [96] for in-depth information and usage of the transformers.

F. MODEL TRAINING

In this step, we choose the appropriate ML classification model, which can be based on the expertise of the domain, the nature of the data, and the expected results. Once a classification model is chosen, the next step is to divide the data set into three sets: the train set, the validation set, and the test set. Often, the proportion of the data for the training, validation, and test sets is chosen to be 80:10:10 percentages (likewise, 70:20:10 and 60:20:20 are also utilized in some cases), and partitioning of the data set is performed randomly to eliminate the introduction of bias.

The train set is used as input to the chosen classification model. It is trained using optimization methods to minimize the error between the given labels and the model's predicted output. In the case of classification problems, we try to reduce class errors by leaning on the model classification parameters. Each classification model requires a type of optimization algorithm to minimize the mistakes. For example, deep learning models use Stochastic Gradient Descent, Adagrad, Adadelta, RMSprop, Adam, AdamW, and others to learn the weights and biases in deep learning models [97].

The training process may also include dividing the train set into batches so that extensive data is divided into batches to train the model. An *epoch* refers to the process in which the model has used all the training data, and an *iteration* refers to the number of batches in an epoch. The training process is completed when the error is minimized or the training process has reached a predefined number of epochs. Once the model is trained, we use the validation set to assess the model's performance. The *hyperparameters* of the model are tuned [98] to obtain the best model for the given metric. The metric is chosen based on the outcome needs and is tied to the domain.

Once the fine-tuning of the model is completed, the model is used in test environments. In training, we decided to partition the data sets into training, validation, and testing sets. This is also called the *holdout method*. However, this may often cause models to have a high variance or have limited data samples. In such cases, we often refer to model training using *k-fold cross-validation*, where the data is split into k subsets, and we train the model on $k-1$ subsets and validate the model using the remaining set. The process is repeated until all training is performed on all k subsets, and the performance is averaged. The k-fold cross-validation helps reduce the models' variance to learn the best parameters. Usually, the value of k is chosen to be 5 or 10. When the number of samples is highly limited in the data set, the *leave-one-out* cross-validation is used and is similar to k-fold, where one sample data point is used for validation and the rest for training. We urge the readers to refer to [99] to explore related methods and their variations.

Often, one has to train several classification models and use different validation strategies to select the best-performing classification model for the problem at hand.

G. MODEL EVALUATION

Once the best-performing model has been selected, we have several metrics to evaluate classification performance. The most commonly used classifier prediction metrics in new unseen data are precision, recall, F1-Score, sensitivity, specificity, accuracy, root-mean-square error (RMSE), and coefficient of determination (R^2). They are given by:

$$\text{Precision} = \frac{\text{TP}}{\text{TP} + \text{FP}}, \tag{4}$$

$$\text{Recall} = \frac{\text{TP}}{\text{TP} + \text{FN}}, \tag{5}$$

$$\text{F1-score} = 2 \times \frac{\text{Precision} \times \text{Recall}}{\text{Precision} + \text{Recall}}, \tag{6}$$

$$\text{Sensitivity} = \frac{\text{TP}}{\text{TP} + \text{FN}}, \tag{7}$$

$$\text{Specificity} = \frac{\text{TN}}{\text{TN} + \text{FP}}, \tag{8}$$

$$\text{Accuracy} = \frac{\text{TP} + \text{TN}}{\text{TP} + \text{TN} + \text{FP} + \text{FN}}, \tag{9}$$

This work is licensed under a Creative Commons Attribution 4.0 License. For more information, see https://creativecommons.org/licenses/by/4.0/

This article has been accepted for publication in IEEE Access. This is the author's version which has not been fully edited and
content may change prior to final publication. Citation information: DOI 10.1109/ACCESS.2024.3373910

$$\text{RMSE} = \sqrt{\frac{1}{n}\sum_{i=1}^{n}(y_i - \hat{y}_i)^2}, \tag{10}$$

$$R^2 = \frac{\sum_{i=1}^{n}(\hat{y}_i - \bar{y})^2}{\sum_{i=1}^{n}(y_i - \bar{y})^2}, \tag{11}$$

where TP represents the number of true positive predictions, TN the number of true negative predictions, FP the number of false positive predictions, and FN the number of false negative predictions; y represents the actual value $\hat{y}$ represents the predicted value, $\bar{y}$ represents the mean value, n is the number of data points. Each metric measures certain aspects of the model's performance and allows us to interpret the results specific to the application problems. For example, we may need a model that requires both precision and recall to be high for specific medical applications. In such cases, we have to choose a classification model that provides higher precision and high recall, not simply based on the model of choice.

H. MODEL DEPLOYMENT

The best-performing model against the set criteria is then used for real-world predictions (often called the *production model*). The model predicts the outcomes on newer data samples, and labels for these samples may or may not exist. The model's outcome is recorded, and the actual labels are registered (if they exist). The model performance against the desired metrics is continuously monitored, and feedback is generated to identify the cases where model performance needs improvement. This feedback may apply to any previous steps in the ML pipeline, as shown in Figure 4. Feedback is then analyzed, an appropriate corrective measure is taken, and the model is retained and deployed. The model's development, monitoring, feedback analysis, and fine-tuning are an ongoing process.

I. QUALITATIVE AND QUANTITATIVE EFFECT ON ML PIPELINE

Data: The impact of data on shaping ML models is significant, both in qualitative and quantitative aspects. Qualitatively, noise, data imbalance, outliers, and other factors pose a potential risk of overfitting. Quantitatively, the dimension, diversity, correlations, and overlap of classes in the data directly impact a model's efficacy in extracting features from the data or learning from the data. Consequently, acquiring representative and balanced data is pivotal in determining whether ML models can qualitatively develop meaningful patterns and quantitatively utilize their knowledge with new data.

Sensors: The first step is to choose a sensor, considering factors such as activity type, ease of use, and financial implications. There are several advantages and disadvantages to each sensor. For instance, a triaxial accelerometer solely records movement along three axes without providing physiological data. However, sEMG sensors only detect the activity of muscles. Therefore, employing numerous sensors to collect diverse data types enhances efficiency and presents diverse viewpoints.

Nonetheless, it is imperative to balance the quantity and placement of sensors while avoiding any unnecessary burden on patients. Furthermore, we must use only relevant sensors to acquire valuable information. The number of sensors and the quality of the data they generate will impact the qualitative aspects of the ML model's learning capabilities and quantitative outcomes.

Preprocessing: Preprocessing removes unwanted noise from sensor data and improves the data quality. This helps ML models find hidden patterns, which makes them very effective. For example, preprocessing accelerometer readings can have a quantitative effect on improving the detection of movements and prediction score accuracy on new data. Similarly, preprocessing steps can help improve the quantitative outcomes of ML models by incorporating data from additional sensors.

Feature Engineering: To detect activities from time-series data, windowed data is preferred over long data streams, providing equal information in smaller segments while reducing complexity and dimensionality. The features extracted are deemed representative of identifying valuable patterns, thereby facilitating data interpretation. Therefore, feature extraction helps improve the model's quantitative performance. Feature selection reduces model complexity by eliminating redundant and irrelevant features, improving quality. The reduction of data dimension enables the visualization and modeling of features by removing feature correlations and enhancing the qualitative aspects of the modeling.

Classification Models: When selecting a classifier model, it is essential to consider the data types, complexity, and sample size. For example, linear models, such as linear discriminant analysis (LDAs), are easier to use but lack the structure to handle complex data. They may underfit intricate data, resulting in suboptimal performance. However, LDAs are less likely to overfit with small sample sizes. On the contrary, ANNs can handle complex data; however, they may exhibit overfitting without regularization or with fewer samples.

Furthermore, the outcomes of the ANN models pose a significant challenge in their interpretation and rationalization. Therefore, the choice of classifier has a qualitative impact on the quality of the model. Quantitatively, classifiers impact performance metrics such as accuracy, F1-score, sensitivity, and specificity. The selection of the correct classification model is, therefore, crucial and will have a significant impact on the prediction's performance.

IV. EXISTING APPROACHES IN STROKE REHABILITATION

This section reviews the relevant research on stroke rehabilitation and provides an overview of each work from 2009 to 2023. Based on the primary objectives of this review, we have cataloged the literature into six subsections for easy understanding and comprehension. These include (i) iden-

This work is licensed under a Creative Commons Attribution 4.0 License. For more information, see https://creativecommons.org/licenses/by/4.0/

This article has been accepted for publication in IEEE Access. This is the author's version which has not been fully edited and
content may change prior to final publication. Citation information: DOI 10.1109/ACCESS.2024.3373910

tifying ADL, (ii) estimating clinical scores, (iii) monitoring exercises, (iv) recognizing postures, (v) recognizing gestures, and (vi) deep learning methods.

At the end of this section, Table 2 summarizes the research objectives, sensors, devices, number of subjects, research context, placement of sensors, ML models, results, and limitations from 2009 to 2023. In addition, Table 3 shows the high-level summary of features (extracted and processed) in stroke rehabilitation using wearable sensor devices.

A. IDENTIFYING ADLS

Identifying ADLs is primarily a classification problem. In this direction, Roy *et al*. [17] used an accelerometer and sEMG sensors to classify 11 ADLs from 10 subjects (five men and five women). Sensors were placed on the shoulders, thigh, wrist, and waist. They used the adaptive neuro-fuzzy ML model and ANN to identify ADL with a sensitivity of 95% and a specificity of 99.7%. One of the significant limitations is that they designed ML models for each subject individually, which limits the usage of models for a more generic population.

Arif *et al*. [100] used three triaxial accelerometers to monitor the physical activities of nine healthy subjects (eight men and one woman). Sensors were placed on the ankle, chest, and wrist. They used k-NN, RF, and NN to identify 12 types of physical activity. The Rotation Forest model achieved 98% classification accuracy. One of the significant drawbacks of this work is that it requires selecting optimal features, which may not always be generalizable and may need re-selection of features upon change in the data.

Sadarangani *et al*. [101] uses FSR sensors attached to the forearm to classify three tasks (reach, grasp, and move an object). They collected data from 16 subjects (eight strokes and eight healthy). They used SVM with RBF kernel and LDA to classify the three tasks. SVM with the RBF model provided 92.2% accuracy in subjects with stroke and 96% precision in healthy subjects. However, they did not investigate the effect of the object's weight on grasp detection accuracy and did not change the importance of the entities. Furthermore, they did not seek to study the impact of re-donning the FSR sensors and the tightness levels.

O'Brien *et al*. [102] used the Samsung Galaxy S4 smartphone to collect data from 45 subjects (30 stroke and 15 healthy). The smartphone was strapped to a belt pouch and tied to the waist. The data included an accelerometer, gyroscope, and barometer sensor readings. Subjects self-labeled the activities and were sent to a server via Wi-Fi and 4G communications. They created 270 features (131 from the accelerometer, 131 from the gyroscope, and eight from the barometer), and RF was used to classify the six activities. The findings suggested that the model performed better when trained on stroke data (75% recall) than healthy activity data, and models trained on data collected in the laboratory performed poorly (56% recall) when the same models were tested at home. Some limitations include (i) a lack of study of environmental effects on the model performance and (ii) a limited number of subjects and data samples outside the laboratory environment.

Tran *et al*. [103] uses a single wrist-worn accelerometer to classify functional and nonfunctional arm movements. The study included 20 subjects (10 healthy and ten strokes) and extended the previous research [104]. The study strongly suggests that RF performs better overall when compared with SVM-RBF, linear SVM, and k-means. However, the performance of RBF-SVM was better than RF in inter-subject for the standard group, and k-NN was slightly better than RF in inter-normal and intra-stroke group [103]. One of the significant limitations of this work is that it is limited to only 4 ADLs and requires additional sensors to measure angular velocity, which is crucial sensing information to consider in measuring ADLs.

Lee *et al*. [32] use a wrist-worn triaxial accelerometer and gyroscope (Shimmer Ireland) to (i) detect upper limb movements detected with goal-directed (GD) during ADL and (ii) evaluate motor performance quality during home rehabilitation of 30 subjects (20 strokes and ten healthy). Logistic regression was used to classify GD and non-GD movements, and the model detected GD movements with 87% Area Under the Curve (AUC). RF detects whether exercise requires feedback for home rehabilitation with an F1 score of 84.3%. Some limitations of this work include that they did not consider both movements that belong to GD and non-GD movements. In addition, the approach has not been tested in home settings. Furthermore, the process requires two algorithms to detect tasks and quality assessments.

Chen *et al*. [105] uses five IMUs (one on each wrist, one on each of the upper arms, and one on the hip) to detect 10 ADLs using four ML models. They collected data from 11 subjects after the stroke, and Apple Watches (Series 3) were used to manage the data. Decision Tree, RF, SVM, and Extreme Gradient Boosting (XGBoost) were used in the analysis, and they found that XGBoost achieved 82% accuracy on 10 ADLs and 90% accuracy on seven tasks. One of the main limitations of this work is the limited sample size ($n = 11$), which is limited in diversity (data variations) due to a limited number of samples. Furthermore, the data was collected in a structured script-based simulated environment, which is unnatural (as in a home).

B. ESTIMATING CLINICAL SCORES

Estimating clinical scores can be both a classification and regression problem. This depends on the desired clinical scores. If the scores are whole numbers (1, 2, 3, and so on), they fall under the classification category. On the other hand, if the clinical scores are a range of numerical values, they fall under the category of regression.

Patel *et al*. [16] used accelerometer sensors on the hand, forearm, upper arm, and trunk to estimate FAS scores [106][1]. The study included 24 hemiparesis subjects and used RF to

[1]The Wolf Motor Function Test (WMFT) quantifies the upper extremity (UE) motor ability through performance time (WMFT-TIME) and functional ability (WMFT-FAS)

This work is licensed under a Creative Commons Attribution 4.0 License. For more information, see https://creativecommons.org/licenses/by/4.0/

This article has been accepted for publication in IEEE Access. This is the author's version which has not been fully edited and
content may change prior to final publication. Citation information: DOI 10.1109/ACCESS.2024.3373910

estimate FAS scores on 15 motor tasks. The RF model and further evaluation of motor tasks revealed that they could calculate scores with 0.04 points and a standard deviation of 2.43 points (with a coefficient of determination $R^2 = 0.96$). We note that the tasks in the study were segmented and separated, which helps the RF model select the features and identify the tasks correctly. Furthermore, further work is required to evaluate the model's performance when the jobs are isolated (as in the real world) and when the subjects wear the sensors (self-calibration and position variation).

Yu *et al*. [107] used two accelerometers and seven flex sensors to estimate the Fugl-Meyer Assessment (FMA) scores of 24 stroke subjects (consisting of 16 men and eight women). They used ensemble regression using an Extreme Learning Machine (ELM) of 7 weak models and could predict the FMA scores of the seven exercises with the coefficient of determination (R^2 = 0.917). The FMA estimation framework was designed to work in both laboratories and homes. However, the system was not tested in an open, free-living environment, nor was the model deployed on a wearable device.

Oubre *et al*. [108] used two inertial sensors (attached to the wrist and sternum) to estimate the levels of upper extremity impairment using FMA scores. They collected data from 23 subjects (seven men and 16 women) who had experienced a stroke. As a first step, they used DBSCAN [66] to group time series segmented data to extract features and later support vector regression (SVR) to estimate FMA scores. Their regression approach calculated the FMA scores with a normalized root-mean-square error of 18.2% (R^2 = 0.70) and only required one minute of time series data to estimate. One of the limitations of this method is that the model is dependent on large continuous movements. This dependency, in turn, may cause exhaustion among subjects. The second limitation is that they do not evaluate the correctness of activities, which is necessary to determine the quality of movements.

Park *et al*. [109] developed an automatic grading system to measure subtle weaknesses in the upper and lower extremities. They collected 60 instances of kinematic characteristics of motor disorders in stroke patients with the NIHSS score (0 or 1) or Medical Research Council (MRC) grades (7, 8, or 9). This included 15 stroke subjects (10 men and five women). They also used a synthetic minority oversampling technique [110] to complement the imbalance. They achieved 0.912 AUC for NIHSS scores of 83.3% accuracy using an ensemble of SVM kernels and boost algorithms, while they achieved 0.86 AUC for MRC scores of 80% accuracy with SVM. One of the drawbacks of this approach is that it generates synthetic data to handle class imbalance, which is not ideal, as the generated data could have different statistical characteristics. Therefore, the models may not generalize well for out-of-distribution (OOD) samples.

Adans-Dester *et al*. [111] use wearable sensors (consisting of 6 accelerometers) strapped to the arm, chest, fingers, and wrist to estimate the severity and quality of the impairment. They collected data from 37 subjects (16 stroke, 21 traumatic brain injury) and developed a modified RF model. They estimated FMA scores (severity of impairment) with the coefficient of determination $R^2 = 0.86$. On the other hand, they estimated FAS (quality of movement) with $R^2 = 0.79$. The approach assumes that the data precisely segments activities, meaning the models could calculate the scores. However, more studies are needed to assess performance in home and community settings, where tasks may not be clearly understood.

C. MONITORING EXERCISES

Pan *et al*. [112] developed a home-based self-rehabilitation system to monitor and detect exercises of the shoulder joint using a smartphone and two wireless accelerometer sensors. The plan was tested on 10 healthy subjects (three men and seven women) and 14 stroke subjects (five men and seven women). In the study, the authors included five exercises: ear touch, hand raise, climbing wall, pendulum movement, and assisted active stretching. The SVM classifier classified these exercises with 96% accuracy. Some limitations of this system include: (i) the data was easily segmented into different exercises with the use of the Android app; (ii) the authors did not specify whether the SVM model was deployed on the smartphone to classify the exercises or the data was collected in a centralized location, and then the analysis was performed offline.

Yurtman and Barshan [113] developed an autonomous system to detect and evaluate physical therapy exercises using wearable sensors. They used five IMUs strapped to the arms and legs of five healthy subjects (three men and two women). They proposed a multi-template, multi-match dynamic time warping algorithm (MTMM-DTW) to detect multiple occurrences of more than one exercise type. The algorithm could see and classify eight types of exercise with 93.46% accuracy. Furthermore, it counted and evaluated the practices as correct or incorrect with 88.65% accuracy. However, this method has some limitations. First, the template of each exercise for each subject must be re-recorded. Second, it has not been used in clinical settings, which limits its applicability in healthcare institutions. Third, sensors must be attached to the exact locations (as in template sequence recordings) to correctly detect the exercises, which can be practically difficult at home and in other community settings.

Zhang *et al*. [37] developed a system to identify everyday rehabilitation exercise movements during a routine exercise session. The study included 14 stroke patients (10 men and four women), and IMUs were attached to the patient's wrist to collect data while performing six joint UE exercises (Bobath handshake, straight arm palm press, horizontal flexion and extension of the shoulder, reaching the forehead with the elbow, shoulder touching and wrist turn). They compared eight different ML models to classify the exercises, and the fuzzy kernel classifier (FKC) obtained a 0.56% error (with a standard deviation of 0.64). Some limitations include: (i) the system has not been tested in an open free-living environ-

This work is licensed under a Creative Commons Attribution 4.0 License. For more information, see https://creativecommons.org/licenses/by/4.0/

This article has been accepted for publication in IEEE Access. This is the author's version which has not been fully edited and
content may change prior to final publication. Citation information: DOI 10.1109/ACCESS.2024.3373910

ment, and (ii) needs expert fine-tuning of model parameters for new data points.

Yu *et al*. [114] proposed a wearable sensor network system to monitor and assess upper extremity motor function quantitatively. Compared to traditional approaches, they used compressed sensing technology to reduce data transfer to the computer (using a sparse representation). The study included 230 stroke subjects (13 men and ten women), and accelerometer sensors were strapped to the forearm and shoulder. The exercises had a Bobath handshake and shoulder touch, and the results indicated that accelerometer signals could be compressed by one-third of the raw signal length. They used the ELM algorithm to detect and compare the exercises, resulting in an accuracy of 89.5% (compared to the accuracy of 92.5% of the raw signal model). One of the main limitations is that the study has not considered additional sensing modalities and how they affect the performance of the compressed sensing model.

D. RECOGNIZING POSTURES

Sazonov *et al*. [115] proposed a shoe-based wearable sensor device to monitor postures and activities. They observed that heel acceleration and plantar pressure uniquely characterized postures and typical movements. The study included nine subjects (3 men and six women). Each shoe had five FSRs and an accelerometer sensor implanted at the critical contact points. SVM was used to classify six postures (sitting, standing, walking/jogging, ascending stairs, descending stairs, cycling) with an accuracy of 95% in the complete sensor set and more significant than 98% in the optimized sensor set.

Fulk *et al*. [116] also proposed a similar shoe-based wearable sensor to classify three postures (sitting, standing, and walking) with SVM. They included data from eight stroke subjects (six men and two women). Using SVM, they achieved a classification accuracy of 99% (recall and precision from 0.99 to 1) for individual models and an accuracy of over 76.9% (recall and precision from 0.82 to 0.99) for group models.

In another work, Fulk *et al*. [117] developed a smart shoe to identify three postures (sitting, standing, and walking) using ANN instead of SVM. The study included 12 stroke subjects (6 men and six women). They achieved an accuracy of more than 97.2% (95–99% recall and 95.4–98.7% precision). However, the three studies were limited to laboratory settings. Another drawback of these studies is that they do not monitor activities in the upper extremities. Furthermore, subjects without ambulatory abilities cannot use these systems. Moreover, they have not measured the power of the sensors to measure the transition from one activity to another.

Cheng *et al*. [118] used four sEMGs and two accelerometer sensors to identify seven body postures (including 50 categories of dynamic activities). The seven body postures include standing, sitting, squatting, lying on the right, lying on the left, lying face up, and lying face down. The sEMG and accelerometer sensors were strapped to the chest and right thigh. The hidden Markov model (HMM) [119, 120] was used to classify gestures by combining the sEMG and accelerometer data. The average accuracy of posture recognition was 98.3%, where the model sometimes misclassified sitting and squat postures. However, the classification model is complex, requires much time to process data to detect activities, and is unsuitable for practical use on wearable devices.

Xiao *et al*. [46] developed a prototype to monitor the FMG signals from the upper extremities using FSR straps attached to the forearms. The study includes only six healthy subjects (six men). They used non-kernel ELM to classify six postures related to drinking tasks by organizing the forearm FMG in real-time, with an accuracy of 92.33% and a standard deviation of 3.19%. However, the study is limited to the discrete classification of six postures of the upper extremities. Moreover, the system cannot automatically identify the signatures of different types of multijoint movements.

Masse *et al*. [121] developed the characterization of sit-to-stand and stand-to-sit (STS) using inertial sensors and barometric pressure (BP) sensors. Each subject was simultaneously paired with an inertial sensor (3D accelerometer and 3D gyroscope) sampling at 200 Hz and a BP sensor sampling at 25 Hz. The study included 12 stroke subjects (seven men and five women), 345 STS were recorded, and subjects were monitored in a semi-structured conditioned protocol. They used Logistic Regression to classify STS with a single device comprising inertial with 75.4% accuracy. In addition, they combined inertial and BP sensors to achieve 90.6% accuracy. Therefore, the classification algorithm is not easily generalizable and depends on the collected dataset. Furthermore, the BP sensor is vulnerable to external perturbations, so validating further in different settings is critical.

E. RECOGNIZING GESTURES

Yang *et al*. [122] introduced an IoT-enabled stroke rehabilitation system based on a smart wearable armband (SWA), ML algorithms, and a 3D printed dexterous robotic hand to recognize nine gestures (agree, close hand, open hand, pointer thumb and middle finger, thumb and little finger, flex hand, extend hand and relax). The study included three subjects (two men and one woman). The authors compared three ML models to identify the nine gestures: LDA, multilayer perceptron (MLP), and SVM. The MLP model identified the nine gestures with 96.02% classification accuracy. The system and the approach come with the following limitations: (i) the train and test data came from the same subjects and were not tested on data from unknown subjects; (ii) the dataset itself is limited to three topics and lacks diversity; (iii) the system is limited to only nine gestures; (iv) the accuracy of the system decreases as and when other gestures are introduced; and (v) the position of the armband changes the results, which is not such an attractive option if we were to use such a system in home settings.

This work is licensed under a Creative Commons Attribution 4.0 License. For more information, see https://creativecommons.org/licenses/by/4.0/

This article has been accepted for publication in IEEE Access. This is the author's version which has not been fully edited and
content may change prior to final publication. Citation information: DOI 10.1109/ACCESS.2024.3373910

F. DEEP LEARNING METHODS

In this subsection, we highlight the deep learning approaches. The (last) four references (on page 21) in Table 2, comprehensively summarize four deep learning approaches.

Panwar *et al*. [90] proposed a "Rehab-Net" framework to classify three upper limb movements from wrist-based accelerometer data. The *Rehab-Net* uses a CNN-based deep learning model. The framework was tested in two situations: (i) a semi-naturalistic environment (making tea) with four-stroke subjects and (ii) a natural environment (any desired arm movement for 120 min) with ten-stroke subjects. They achieved an accuracy of 97.89% on semi-naturalistic data and 88.87% on naturalistic data. The CNN-based model outperforms LDA, SVM, and k-means. The model has also been optimized (algorithmic level) for real-time hardware implementation. One of the drawbacks of this study is that the deep learning (DL) models were trained specifically for individuals and, therefore, lacked generalizability.

Kaku *et al*. [123] developed an LSTM-based approach to identify five functional primitives using nine internal measurement units. Data were collected from 48 stroke subjects (22 men and 26 women). The nine IMUs were attached to the cervical vertebra C7, the thoracic vertebra T12, the pelvis, arms, forearms, and hands. They achieved an accuracy of 78%. The study included only subjects with dominant rights. This hand dominance may have a differential influence on the preferential roles of the UEs and, therefore, needs further investigation. Furthermore, classification performance deteriorates significantly for severely impaired patients, lacking generalizability.

Chae *et al*. [124] developed a home rehabilitation system using a smartwatch and a smartphone. They attached a smartwatch (equipped with an accelerometer and a gyroscope sensor) to a wrist and collected data during exercises. Data were collected from 10 control subjects and 22 subjects with stroke. They used a CNN model and achieved 99.80% accuracy (when both accelerometer and gyroscope were combined) versus 98.13% (only accelerometer data) and 96.07% (gyroscope data). The study selected four exercises to suit chronic stroke patients, particularly those based on bilateral movement therapy. However, these exercises may not fit all stroke subjects. Therefore, we cannot apply the same approach to issues with stroke.

Recently, Nair *et al*. [91] proposed a system to identify the completion status of rehabilitation exercises. They used triaxial accelerometers attached to the hands and forearms to collect data. The study included approximately five subjects and wrist, forearm, and shoulder exercises. They used a CNN model to classify 12 activities (six wrists, four forearms, two shoulders) with 98.61% accuracy. The model outperforms the Decision Tree, SVM, Linear discriminant (LD), and Naive Bayes classifiers. One of the limitations of this approach is that it is not automated, *i.e*., we need to convert time series data into images for feeding to CNN, which may not be a practical approach, as LSTMs are robust models for processing time series data. It will be inconvenient to use such an approach for real-world processing.

V. OPEN RESEARCH CHALLENGES AND FUTURE DIRECTIONS

Wearable sensors and ML will play an important role in stroke rehabilitation. Wearable sensors are the most practical option, from different training interventions to automated objective assessment and creating predictive models using ML. Figure 7 shows the combined view of wearable devices and ML algorithms to improve stroke rehabilitation outcomes. In this section, we discuss potential open research challenges and future research directions to solve problems in stroke rehabilitation.

A. TRAINING INTERVENTIONS

The two most prevalent issues encountered in training interventions relate to the insufficient knowledge of new physiotherapists and the execution of the optimal dose of therapy required for a particular patient. Except for studies that use biofeedback, the mandatory use of wearable sensors during the application of all other interventions has not yet been implemented. Wearable sensors can establish a connection between training and its effects in real-time, which helps improve training and maximize potential results. It enhances patient performance and provides a valuable opportunity for a novice physiotherapist to gain insight into the patterns (such as movement patterns and muscle activity) of patients' motor ability and progression in real-time. Furthermore, by using wearable sensor data from healthy subjects, physiotherapists may better understand the patient's status. Thus, the appropriate form of therapy and individualized treatment dosage can be implemented in real-time.

B. REMOTE MONITORING

Physicians currently recommend that patients perform prescribed exercises with wearable devices when they gain voluntary movement without a physiotherapist, in addition to routine care. Nonetheless, whether the patients effectively complete the tasks as per the instructions remains uncertain.

In [32], the researchers examined data from the unaffected UE. However, assessing the health of the affected body parts, specifically the upper or lower extremities, is advisable. This is critical for making improvements. Furthermore, the optimal amount of self-directed exercise and research related to selecting more efficient activities for individual patients to help regain pre-stroke movement ability remain to be explored.

Although wearable devices hold promise for remotely capturing movement data, it is essential to note that movement data alone is not always sufficient to reveal the fundamental physiological conditions of patients. It is essential to include sensors that record physiological signals such as nerve conduction velocity, EMG, ECG, photoplethysmogram (PPG), and others to provide additional information about muscle and nerve activity. Furthermore, incorporating data on heart

This work is licensed under a Creative Commons Attribution 4.0 License. For more information, see https://creativecommons.org/licenses/by/4.0/

This article has been accepted for publication in IEEE Access. This is the author's version which has not been fully edited and content may change prior to final publication. Citation information: DOI 10.1109/ACCESS.2024.3373910

TABLE 2: A comprehensive review of existing stroke rehabilitation approaches using wearable devices.

Year	Aim	Sensors	Device	Device Type	Subjects	Context	Placement	ML Model	Results	Limitations
2009 [17]	ADL	ACC, sEMG	Analog Devices, DelSys	Wired	10	Lab	Shoulders, thigh, wrist, waist	ANN	Identify 11 daily living activities with a sensitivity of 95% and a specificity of 99.7%	Algorithms were trained separately; needs proper segmentation of tasks.
2010 [16]	Clinical scores	ACC	Viatport	Wired	24	Lab	Chest, fingers, forearm, upper arm	RF	Derive accurate estimates of FAS scores during the performance of 15 motor tasks using an accelerometer	Tested only in controlled environments. Needs evaluation on the placement of sensors.
2011 [125]	Postures	ACC, FSR	Custom (Interlink, ST Microelectronics)	Wireless	9	Lab	Feet	SVM	Classify the accuracy of six postures of 95% in the complete sensor set and over 98% in the optimized sensor set.	Sensor readings may differ in uncontrolled environments. Tasks of the upper limb are not monitored.
2011 [116]	Postures	ACC, FSR	Custom (Interlink)	Wireless	8	Lab	Feet	SVM	Classify three postures with 99% accuracy for individual models and 76.9% for group models	Limited to subjects with ambulatory abilities only. Unmeasured ability to detect transitions from sitting to/from standing.
2012 [117]	Postures	ACC, FSR	Custom (Interlink)	Wireless	12	Lab	Feet	ANN	Identify posture with an accuracy of 97.2%	Not tested in open free-living environments and during task transitions.
2013 [112]	Exercises	ACC	Custom (Smartphone)	Wireless	24	Remote	Shoulders	SVM	Classify 5 activities with 96% accuracy.	Tasks were segmented using the Android app; whether the ML model was deployed on the smartphone is unclear.
2013 [118]	Postures	ACC, sEMG	Delsys Myomonitor IV	Wired	10	Lab	Lower limb, chest	HMM	Fusion of sEMG and accelerometer achieved 98.3% classification accuracy	ML model is complex and requires significant time and resources. Not suitable for wearable devices.

ACC: accelerometer; BP: barometric pressure; FSR: force sensing resistor; GYR: gyroscope; IMU: inertial measurement unit; sEMG: surface electromyography. ANN: artificial neural network; DT: decision tree; ELM: extreme learning machine; FAS: functional ability scale; FMA: Fugl-Meyer assessment; HMM: hidden Markov model; LDA: linear discriminant analysis; LR: logistic regression; MLP: multi-layer perceptron; RBF: radial basis function; RF: random forest; RMSE: root-mean-square error; SVM: support vector machine; SVR: support vector regression; XGBoost extreme gradient boosting.

This work is licensed under a Creative Commons Attribution 4.0 License. For more information, see https://creativecommons.org/licenses/by/4.0/

TABLE 2 : (continued. . .) A comprehensive review of existing stroke rehabilitation approaches using wearable devices.

Year	Aim	Sensors	Device	Device Type	Subjects	Context	Placement	ML Model	Results	Limitations
2014 [46]	Postures	FSR	Custom (Interlink)	Wired	6	Lab	Forearm	ELM	Classify six postures related to the drinking task with 92.33% accuracy	Limited to discrete classification of different postures of UEs. It cannot identify the signature of multi-joint movements automatically.
2014 [113]	Exercises	IMU	XSens Technologies	Wired	5	Lab	Arms, legs	MTMM-DTW	Detect and classify the 8 exercise types (93.46% accuracy); Evaluate exercises as correct or incorrect (88.65%)	Needs sensor placements to be precise and re-recording of exercise templates for new exercises.
2015 [100]	ADL	IMU	Colibri	Wireless	9	Lab	Ankle, chest, wrist	k-NN, Rotation Forest, ANN	Identify 12 different types of physical activities using three acceleration sensors; Rotation Forest achieving 98% classification accuracy in individual sensors; k-NN achieving 97.9% accuracy on combined (3 sensors) data	Needs further evaluation on optimal placement of sensors.
2016 [37]	Exercises	ACC, GYR	XSens Technologies	Wired	14	Lab	Wrist	Fuzzy kernel	Achieves 0% error rate for the low impairment level patient group and 0.56% for all patients	Not tested on wearable devices and needs expert knowledge to tune the model.
2016 [121]	Postures	BP, IMU	Physilog	Wireless	12	Lab	Chest	LR	IMU alone achieves 75.4% accuracy whereas combined IMU and BP sensors achieve 90.6% accuracy	Model parameters are highly dependent on data and need re-training of ML model for new data samples

ACC: accelerometer; BP: barometric pressure; FSR: force sensing resistor; GYR: gyroscope; IMU: inertial measurement unit; sEMG: surface electromyography. ANN: artificial neural network; DT: decision tree; ELM: extreme learning machine; FAS: functional ability scale; FMA: Fugl-Meyer assessment; HMM: hidden Markov model; LDA: linear discriminant analysis; LR: logistic regression; MLP: multi-layer perceptron; RBF: radial basis function; RF: random forest; RMSE: root-mean-square error; SVM: support vector machine; SVR: support vector regression; XGBoost extreme gradient boosting.

This work is licensed under a Creative Commons Attribution 4.0 License. For more information, see https://creativecommons.org/licenses/by/4.0/

This article has been accepted for publication in IEEE Access. This is the author's version which has not been fully edited and content may change prior to final publication. Citation information: DOI 10.1109/ACCESS.2024.3373910

This article has been accepted for publication in IEEE Access. This is the author's version which has not been fully edited and content may change prior to final publication. Citation information: DOI 10.1109/ACCESS.2024.3373910

TABLE 2 : (continued. . .) A comprehensive review of existing stroke rehabilitation approaches using wearable devices.

Year	Aim	Sensors	Device	Device Type	Subjects	Context	Placement	ML Model	Results	Limitations
2016 [107]	Exercises	ACC, Flex	Custom	Wireless	24	Lab, Remote	Forearm, fingers, shoulder	ELM	Predict the FMA scores of the seven exercises with coefficient of determination (R2 = 0.917)	Not tested on wearable devices.
2016 [114]	Exercises	ACC	Custom	Wireless	23	Lab	Forearm, shoulder	ELM	Monitor and quantitatively assess upper limb motion function with 89.5% accuracy	Not tested on wearable devices.
2017 [101]	ADL	FSR	Interlink Electronics	Wired	18	Lab	Forearm	SVM, LDA	SVM-RBF achieved 92.2% accuracy (stroke) and 96% (healthy) in grasp detection	Assessed on a single grasp object of constant weight. It did not study the effect of removing and redonning sensors on the effect of an object's weight.
2017 [102]	ADL	ACC, GYR, BP	Samsung Galaxy S4	Wireless	45	Lab, Remote	Waist	RF	The model achieved better performance when trained in stroke data (75% recall) rather than training on healthy activity data	Limited home data. Performance dropped when tested in home environments (56% recall).
2018 [122]	Gesture	sEMG	Custom	Wireless	3	Lab	Forearm	LDA, MLP, SVM	Identify nine gestures with 96.20% classification accuracy with MLP	The train and test data came from the same subjects and have not been tested on unseen subjects' data. The data set itself is limited to 3 subjects and lacks diversity.

ACC: accelerometer; BP: barometric pressure; FSR: force sensing resistor; GYR: gyroscope; IMU: inertial measurement unit; sEMG: surface electromyography. ANN: artificial neural network; DT: decision tree; ELM: extreme learning machine; FAS: functional ability scale; FMA: Fugl-Meyer assessment; HMM: hidden Markov model; LDA: linear discriminant analysis; LR: logistic regression; MLP: multi-layer perceptron; RBF: radial basis function; RF: random forest; RMSE: root-mean-square error; SVM: support vector machine; SVR: support vector regression; XGBoost extreme gradient boosting.

This work is licensed under a Creative Commons Attribution 4.0 License. For more information, see https://creativecommons.org/licenses/by/4.0/

TABLE 2 : (continued. . .) A comprehensive review of existing stroke rehabilitation approaches using wearable devices.

Year	Aim	Sensors	Device	Device Type	Subjects	Context	Placement	ML Model	Results	Limitations
2018 [103]	ADL	IMU	Analog Devices	Wireless	20	Lab	Wrist	k-NN, RF, SVM	Intrasubject classification accuracy of 96% (on controls) and 94% (stroke) and intersubject accuracy of 90% (controls) and 83% (stroke) with SVM-RBF	Limited to only 4 ADLs and may need additional sensors to monitor other tasks.
2018 [32]	ADL	ACC, GYR	Shimmer	Wireless	30	Lab	Wrists	LR, RF	Detected goal-directed (GD) movements with 87% area under the curve (AUC) and also detects whether exercise requires feedback (F1 score: 84.3%)	Did not consider movements that were both GD and non-GD; requires different algorithms for task detection and quality assessment.
2020 [108]	Clinical scores	IMU	MTw Awinda, XSens Technologies	Wireless	23	Lab	Ankles, Wrists	SVR	Estimates the FMA score using random movements with the coefficient of determination (R2 = 0.7), RMSE of 18.2%	Requires large, continuous movements, which can tire patients and does not assess the correctness of the movements.
2020 [111]	Clinical Scores	ACC	Shimmer	Wireless	37	Lab	Arm, chest, fingers, wrist	RF	Estimates FMA (impairment severity) scores with R2 of 0.86 and FAS scores (movement quality) with R2 of 0.79	Requires expertise in deriving clinical scores from wearable sensor data.
2021 [105]	ADL	IMU	Apple Watch (series 3)	Wireless	11	Lab	Wrists, upper arms, hip	DT, RF, SVM, XGBoost	Accuracy 82% on 10 ADLs and 90% accuracy on 7 ADLs with XGBoost	It uses artificial data synthesis to handle data imbalance, which may not be appropriate in all cases. Restricted study to mild stroke patients.

ACC: accelerometer; BP: barometric pressure; FSR: force sensing resistor; GYR: gyroscope; IMU: inertial measurement unit; sEMG: surface electromyography. ANN: artificial neural network; DT: decision tree; ELM: extreme learning machine; FAS: functional ability scale; FMA: Fugl-Meyer assessment; HMM: hidden Markov model; LDA: linear discriminant analysis; LR: logistic regression; MLP: multi-layer perceptron; RBF: radial basis function; RF: random forest; RMSE: root-mean-square error; SVM: support vector machine; SVR: support vector regression; XGBoost extreme gradient boosting.

This work is licensed under a Creative Commons Attribution 4.0 License. For more information, see https://creativecommons.org/licenses/by/4.0/

This article has been accepted for publication in IEEE Access. This is the author's version which has not been fully edited and content may change prior to final publication. Citation information: DOI 10.1109/ACCESS.2024.3373910

This article has been accepted for publication in IEEE Access. This is the author's version which has not been fully edited and content may change prior to final publication. Citation information: DOI 10.1109/ACCESS.2024.3373910

TABLE 2 : (continued. . .) A comprehensive review of existing stroke rehabilitation approaches using wearable devices.

Year	Aim	Sensors	Device	Device Type	Subjects	Context	Placement	ML Model	Results	Limitations
2019 [90]	ADL	ACC	Shimmer, Physilog	Wireless	14	Lab, Remote	Arm, wrist, elbow, chest	CNN	Classifies movements with an accuracy of 97.89% on semi-naturalistic data and with 88.87% accuracy on naturalistic data; It is an optimized model for real-time hardware implementation	Trained on individual subjects, hence lacks generalizability.
2020 [123]	ADL	IMU	Noraxon	Wireless	48	Lab	Neck, Back, pelvis, arms, forearms, hands	LTSM with CNN	Classify five functional (FMA) primitives with 70% accuracy	The study included only subjects with dominant right and classification performance deteriorates significantly in severely impaired patients
2020 [124]	Exercises	ACC, GYR	LG Smartwatch	Wireless	32	Lab, Remote	Wrist	CNN	Classify five exercises with 99.80% accuracy (when both accelerometer and gyroscope included) versus 98.13% (with only accelerometer data) and 96.07% (gyroscope)	Some participant exercises were not captured. The actual precision of the training at home was not evaluated.
2023 [91]	Exercises	ACC	Custom (ST Microelectronics)	Wireless	5	Lab	Hand, forearm	CNN	Classify 12 exercises (6 wrists, four forearms, two shoulders) with 98.61% accuracy; outperforms DT, SVM, LDA and Naive Bayes models	Converting time series data into images for feeding to CNN is not efficient practically.

ACC: accelerometer; BP: barometric pressure; FSR: force sensing resistor; GYR: gyroscope; IMU: inertial measurement unit; sEMG: surface electromyography. ANN: artificial neural network; DT: decision tree; ELM: extreme learning machine; FAS: functional ability scale; FMA: Fugl-Meyer assessment; HMM: hidden Markov model; LDA: linear discriminant analysis; LR: logistic regression; MLP: multi-layer perceptron; RBF: radial basis function; RF: random forest; RMSE: root-mean-square error; SVM: support vector machine; SVR: support vector regression; XGBoost extreme gradient boosting.

This work is licensed under a Creative Commons Attribution 4.0 License. For more information, see https://creativecommons.org/licenses/by/4.0/

TABLE 3: A high-level summary of the features used in stroke rehabilitation using wearable sensor devices is provided. The first column of the table categorizes the literature into five categories: (1) time domain, (2) frequency domain, (3) statistical, (4) morphological, and (5) data specific. The second column identifies the types of features. The last two columns provide an overview of the advantages and disadvantages of five categories of feature groups.

Domain	Features	Advantages	Disadvantages
Time domain	**Amplitude:** Minimum and maximum values: [16, 32, 100, 109, 111, 112, 121]; Amplitude or sum of amplitude: [107, 114, 117]; Range [17, 32, 102, 121], Root Mean Square (RMS): [16, 17, 32, 100, 107, 111, 112, 122] ; Wilson amplitude [100] **Temporal dynamics:** Short time segment [46, 115, 116, 118, 121] **Complexity or structure:** Correlation coefficient [16, 100, 102, 105, 111]; Entropy [16, 32, 111, 112]; Approximate entropy [107, 121]; Linear prediction coefficients [118]; **Energy:** [37, 100, 103, 111]	• Capture temporal trends and dynamics within the data • Easy to interpret • Computationally less intensive • Suitable for real-time analysis using small segments	• Limited frequency information (does not consider frequency information of the underlying signal, specific frequency events) • Insufficient for complex signals • Less robust to noise and outliers
Frequency domain	**Frequency dominance:** Dominant frequency component [17, 37, 111] **Power and energy:** Power spectral density [37]; mean power [37, 102]; energy ratio (dominant frequency to signal energy) [16]	• Captures spectral information (fundamental frequency and harmonics) • Provides compact representation • Less vulnerable to noise (ways to separate signal and noise components)	• Need expertise to understand complexity in understanding the physical or meaningful implications • Needs careful windowing techniques to handle non-stationary signals • Computationally pricier than time domain
Statistical	**Central tendency:** Mean [16, 32, 37, 102, 103, 105, 107, 109, 111, 112, 122]; Harmonic mean [100]; offset from the mean [121] **Dispersion:** Skewness [32, 100, 102, 108, 111]; kurtosis [32, 100, 102, 108]; standard deviation or variance [32, 100, 102, 103, 105, 117, 122]; interquartile range [102]	• Interpretable (provide meaningful insights of signal behavior and patterns) • Quantify key characteristics of the signal (such as central tendency, spread, and distribution shape) • Comparatively robust (such as measures central tendency, less affected by random noise or temporary fluctuations)	• Representation may be inadequate (due to the absence of intricate details) • Sensitive to scaling of data (may alter data distribution, affects statistics) • Depends on the quality of data
Morphological	**Signal characteristics:** Mean crossing number [117], zero crossing number [32, 100, 122], slope sign change [100, 121] **Shape and peak:** Waveform length [122], number of peaks [108], position of maxima [108] **Task specific:** Task duration [111, 121]	• Captures shape information • Visually interpretable (specific to patterns and shapes)	• Large time complexity and computationally expensive • Sensitivity to data processing techniques
Data specific	**Posture and movement:** Axial angle of posture [126], co-activation interval between muscle pairs [17], cumulative length [100] **Angle:** Angular position [37] **Temporal dynamics:** Jerk [16, 32, 111, 112]; start and stop time duration [121], elevation change during the transition of movement [121], velocity data sequence [37] **Distribution:** Histogram bin counts [102] **Regression and trend:** Slope of linear regression [102] **Demography:** Different demographic features [109]	• Direct relevance to the problem • Contextually understandable (identification of relevant patterns to form specific decisions)	• Limited to specific data (may not apply to other available data in general) • Subjective factors or data bias may affect the outcomes • Data-specific processing (requires complex analysis and is time-intensive)

This work is licensed under a Creative Commons Attribution 4.0 License. For more information, see https://creativecommons.org/licenses/by/4.0/

This article has been accepted for publication in IEEE Access. This is the author's version which has not been fully edited and content may change prior to final publication. Citation information: DOI 10.1109/ACCESS.2024.3373910

This article has been accepted for publication in IEEE Access. This is the author's version which has not been fully edited and
content may change prior to final publication. Citation information: DOI 10.1109/ACCESS.2024.3373910

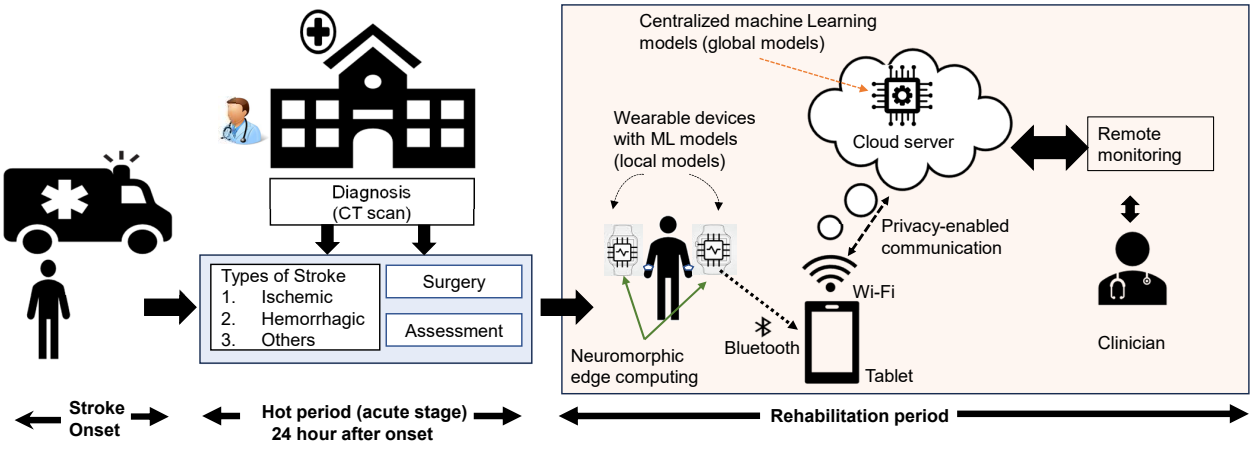


FIGURE 7: The future of wearable devices in stroke rehabilitation. The figure depicts the vision for wearable devices and ML algorithms for rehabilitation after a stroke. ML algorithms can be applied to wearable devices to make predictions (local model) and have a central database (global model) to predict outcomes and advise people, doctors, and health workers. The figure also shows privacy-enabled communication between data collection devices and third-party servers. It also highlights neuromorphic edge computing capabilities for providing efficient (near real-time) ML inferences on portable devices such as a smartwatch.

rate, respiration rate, body temperature, blood pressure information, and movement data may prove advantageous. The correlation, however, has not yet been established.

Another crucial aspect of remote monitoring relates to the ability of ML models to adapt to individual requirements. The current systems posit that offline machine-learning models can be utilized directly. Nonetheless, this is not the case, as the ML models have only been exposed to labeled data and cannot adapt to novel, unobserved data. This necessitates researchers to consider adaptable ML models and novel approaches to cater to personalized requirements.

C. CONTINUOUS ASSESSMENT

Visual assessment or solely utilizing clinical scales can result in subjective errors, leading to inappropriate physical therapy application. Nonetheless, wearable sensor-based continuous monitoring of subjects during clinical evaluation, regular hospital routines, and conventional physiotherapy can be utilized to compile a log of activities. This approach will assist in comparing the evaluated scores recorded by the physiotherapist to the objectively recorded scores by the wearable devices and inferred by the ML models.

Nonetheless, it is imperative to select the appropriate assessment scale. Assessment scales are subject to limitations, including but not limited to the ceiling effect, making it impossible to detect improvements in motor activities precisely. [107]. The use of wearable sensors that apply patient-specific movement information (such as angle of movement, time duration to complete a task, and total amount of activity) for evaluation could help remove this error. Furthermore, the non-linear analysis of wearable sensor data, such as multiscale entropy [127] or the maximum Lyapunov index [128, 129], can address this issue.

D. SELECTION AND PLACEMENT OF SENSORS

The placement of sensors on the body is essential for obtaining the desired signal correctly. As each sensor plays a crucial role in getting precise movement patterns, artifacts, or noise, incorrect placements or displacements of sensors can result in significant issues. Although doctors can prescribe several training interventions and exercises, it is always possible to misplace body sensors. We require methods to monitor correct placements and observe sensor placements automatically. ML algorithms are needed to detect misplacement, and they should also identify the artifacts caused by incorrect positions and compensate for incorrect information. To our knowledge, no such ML model or methodology is utilized in rehabilitation.

Due to wearability issues and the sensors' inefficiency, we cannot capture the patterns of all the activities captured by the sensors. Therefore, further studies are required to determine the optimal location of the wearable sensors.

E. SIGNAL QUALITY AND NOISE

The signal quality depends significantly on the sensitivity of the particular sensors and the environment. Sensors that capture external body movements exhibit greater sensitivity and are more vulnerable to noise. A minor error during data collection from healthy individuals can result in noisy data and a noisy model of healthy individuals. Therefore, a stroke patient can be attributed to being healthy, or vice versa; that is, an inaccurate movement from a healthy subject can be considered the movement of a stroke patient. Moreover, when the sensor is positioned on a body part with a significant degree of freedom, the signal may exhibit increased noise if the number and placement of sensors are not precisely calibrated.

Sometimes, simple movement sensors (for example, ac-

This work is licensed under a Creative Commons Attribution 4.0 License. For more information, see https://creativecommons.org/licenses/by/4.0/

This article has been accepted for publication in IEEE Access. This is the author's version which has not been fully edited and
content may change prior to final publication. Citation information: DOI 10.1109/ACCESS.2024.3373910

celerometers) cannot distinguish one movement from another due to a limited pattern dissimilarity. Therefore, combining more than one sensor (multi-sensor fusion) and sensors that capture the body's internal activity during a task is necessary to obtain more accurate information [130, 131].

Monitoring tasks would be more accurate if data were recorded for extended periods. As in the conventional approach, regular monitoring poses a challenge; however, long-term monitoring by wearable devices offers a solution. During specific exercise practice, patients are required to wear the device, and information is gained on the improvement or deterioration of motor function by tracking changes in patterns from time to time. A single exercise is generally of short duration, and efficient features can be extracted from these temporary signals.

F. TIME-SERIES DATA MODELING AND ANALYSIS

The analysis of sensor data signals is carried out primarily based on the characteristics of the time domain. Time-frequency analysis may prove advantageous due to the non-stationary nature of sensor data. Furthermore, the feature extraction process can be further enhanced by incorporating characteristics of the frequency domain, the distribution of sensor data, and the fractal nature of the sensor data. Recent advances in deep learning, such as LSTM models, have successfully extracted time-series features from wearable sensors [123, 132]. However, LSTM models are more complex and require much more data to learn than traditional models. Hence, implementing LSTM in stroke motor recovery necessitates further investigation, as data accessibility is frequently constrained.

Most studies use and analyze individual databases, making comparing different studies difficult. Furthermore, the limited collection of a large amount of data to determine the degree of variability between subjects from stroke patients is difficult, making the analysis process inefficient. Therefore, future studies should include large cohorts and, where possible, make these databases available to researchers outside organizations with due ethics considerations.

G. ACCELERATING ML WORKLOADS

Using IoT and deep learning to process wearable sensor data collected from stroke patients in the cloud requires significant power consumption. Graphics Processing Units (GPUs) are commonly utilized as the standard equipment for training deep learning networks, including but not limited to CNN, RNN, LSTM, and other large models. Images, speech, and sensor data can be introduced in milliseconds. However, the power consumption for training a deep learning model is not less than 100 W when using a high-performance GPU system [133]. According to [134], the power budget for devices such as smartphones, tablets, or watches is approximately 1.5 W, significantly lower than the power requirements for the cloud-based deep learning model. Therefore, it is recommended to use cloud-based services to train deep learning models and perform inference using wearable sensors.

Google has addressed this issue via the Tensor Processing Unit (TPU) [135]. TPUs were explicitly designed for deep learning networks, and they require a smaller number of resources and are much faster than GPUs. Google introduced an Edge TPU into a custom system-on-chip (SoC) named Google Tensor, and it was released in 2021 with the Pixel 6 line of smartphones. Its advantage is that it can achieve more excellent performance with lower power consumption. Suppose TPU is integrated into stroke watches or wearable sensors; in that case, the desired scores of stroke patients' progression can be obtained immediately, potentially opening up a new avenue for the future of monitoring stroke rehabilitation systems.

H. DATA AND ML MODELS

Most current rehabilitation methodologies are limited to controlled environments (laboratory settings). Furthermore, the rehabilitation tasks and exercises are predetermined and must be adhered to by the subjects. These restrictions limit user mobility and hinder the free movement of individuals. Therefore, newer approaches are needed, considering uncontrolled environments (laboratory and home settings) and protocols that are not scripted. Using such techniques to identify, classify, and categorize tasks would benefit stroke survivors.

Current approaches develop ML/DL models, considering limited data (refer to Table 2). Increasing the number of samples and diversifying the dataset with varying tasks and scenarios is necessary to make the models robust to different scenarios. Researchers need to share datasets and models to collaborate and develop strong models.

Although ML models are trained using diverse variables, they often cannot be transferred from one environment to another. For example, we cannot guarantee that the performance of models developed specifically for laboratory settings will be the same when used on data collected at home. Models created in laboratories are typically not tailored to the needs of individuals at home (local or personalized models). We currently do not possess models that can be tailored to individual requirements.

The literature indicates that most researchers use supervised or unsupervised learning to address their issues. Semi-supervised learning is also a promising direction to explore when there are many unlabeled samples. Furthermore, the study of sequential decision-making has not been conducted in stroke rehabilitation research. To this end, integrating deep learning techniques with reinforcement learning may provide promising solutions.

I. NEUROMORPHIC EDGE COMPUTING

Wearable computing units frequently necessitate additional energy for novel applications that rely on numerous sensors and possess high learning capacities. However, with current portable battery technology, the power supply to wearable devices is insufficient to provide long-term monitoring capabilities. This limitation must be addressed to improve the

This work is licensed under a Creative Commons Attribution 4.0 License. For more information, see https://creativecommons.org/licenses/by/4.0/

This article has been accepted for publication in IEEE Access. This is the author's version which has not been fully edited and
content may change prior to final publication. Citation information: DOI 10.1109/ACCESS.2024.3373910

longevity of wearable devices for continuous monitoring in stroke rehabilitation [136].

The intricacy of utilizing a wireless module, its substantial power consumption, considerable data volume, spatial limitations arising from wireless transmission range, and privacy concerns arising from signal broadcasting. Non-negligible time latency from communication channels renders this solution less than optimal [137]. These limitations in technology severely limit the applicability of wearable sensors.

In contrast to traditional methods that rely on a binary digital system, brain-inspired neuromorphic hardware is a promising solution. Still, it needs to be improved regarding data removal, storage, and transmission across various units [137]. From this angle, the front-end processor may be a neuromorphic semiconductor with an intelligent algorithm integrated right next to the sensor.

Compared to traditional methods that rely on a binary digital system, brain-inspired neuromorphic hardware is a promising solution. However, it still needs to be improved regarding data removal, storage, and transmission across various units [137]. From this perspective, the front-end processor may be a neuromorphic semiconductor with an intelligent algorithm integrated right next to the sensor.

Spike-Timing Dependent Plasticity (STDP), a biologically inspired learning rule, has been the focus of many neuromorphic realizations of on-chip learning. Since synaptic weight changes only occur if presynaptic spikes reach the synapse, this model is highly suitable for event-based algorithms [138, 139]. These developments are encouraging researchers to adapt neuromorphic chips for stroke rehabilitation.

J. GENERATIVE MODELING

The current wave of AI is driven by success in data availability, generative modeling, and process-intensive computing infrastructure. Recent developments have propelled several natural language applications to enable human-like conversations, such as OpenAI's ChatGPT [2]. These models are built using large Generative Pretrained Transformer (GPT) models in combination with supervised and reinforcement learning.

Large language models (LLMs) have recently been used in medicine [140]. LLMs use transformer architecture internally to build robust models. They can process human language (*i.e.*our written or verbal communication) to drive AI algorithms to develop specific models and augment medical competencies in patient care. ChatGPT is an excellent example of an LLM. Since ChatGPT is a chatbot, it can process user queries and return results in the desired style, format, and language. ChatGPT can also write code, process images, and perform machine-learning tasks. It is an emerging area of research with great potential that needs further exploration.

Furthermore, generative modeling provides many options for generating new data from existing data distributions. Health professionals and researchers often encounter imbalanced datasets and limited samples. Generative modeling can supply new data samples to address imbalanced datasets and increase the sample numbers. Nonetheless, it is imperative to conduct additional research to guarantee that the newly generated data samples adhere to medical conditions and do not infringe upon patient privacy while generating new samples.

[2]https://openai.com/research/gpt-4

K. DATA SECURITY AND PRIVACY

Health data is highly susceptible and necessitates the utmost protection against security threats. Lately, the number of data breaches has increased significantly. Data from wearable sensors must be exchanged with third-party applications, including smartphone apps and web servers, to provide information to patients, caregivers, and health professionals.

Many deep learning algorithms run on the cloud, and some lightweight algorithms run on wearable devices. These signal processing, ML, deep learning algorithms, and visualization tools require data transfer from wearable devices (edge devices) to a centralized cloud server. This presents a significant risk of leakage of sensitive health data and jeopardizes data privacy. For innovative healthcare, it is possible to achieve secure communication using wearable devices by employing identity-based systems and biometrics that authenticate user identity and blockchain-based immutable data security [141, 142].

The differential privacy technique [143] involves the transfer of patterns in wearable device data to cloud services. A global algorithm updates the appropriate ML model without utilizing individuals' raw data. Nevertheless, numerous financial, societal, and technological obstacles remain when assessing the proper degree of privacy for wearable devices, notably in medical settings.

VI. KEY FINDINGS

Figure 8 summarizes the objectives, sensors, types of feature extraction, classification models, and various performance metrics used in the literature. The **main findings** of this review are:

- Even though there have been several training interventions for stroke rehabilitation, patients still need to practice helping them get back to their pre-stroke mobility. In addition to regular care at home or in a rehabilitation center, patients must perform prescribed exercises and other possible daily activities without a physiotherapist. Remote monitoring of stroke patients' movements is necessary to ensure patients perform the recommended tasks correctly.
- The extent of remote assessment utilizing wearable sensors is contingent upon the training intervention provided. Specific training interventions, such as functional electrical stimulation (FES), are not feasible for patients to execute at home or in rehabilitation centers without the aid of professionals. Therefore, recovery from these interventions may not be remotely monitored. However, virtual reality (VR) can be employed in home-based monitoring systems without manual assistance.

This work is licensed under a Creative Commons Attribution 4.0 License. For more information, see https://creativecommons.org/licenses/by/4.0/

This article has been accepted for publication in IEEE Access. This is the author's version which has not been fully edited and
content may change prior to final publication. Citation information: DOI 10.1109/ACCESS.2024.3373910

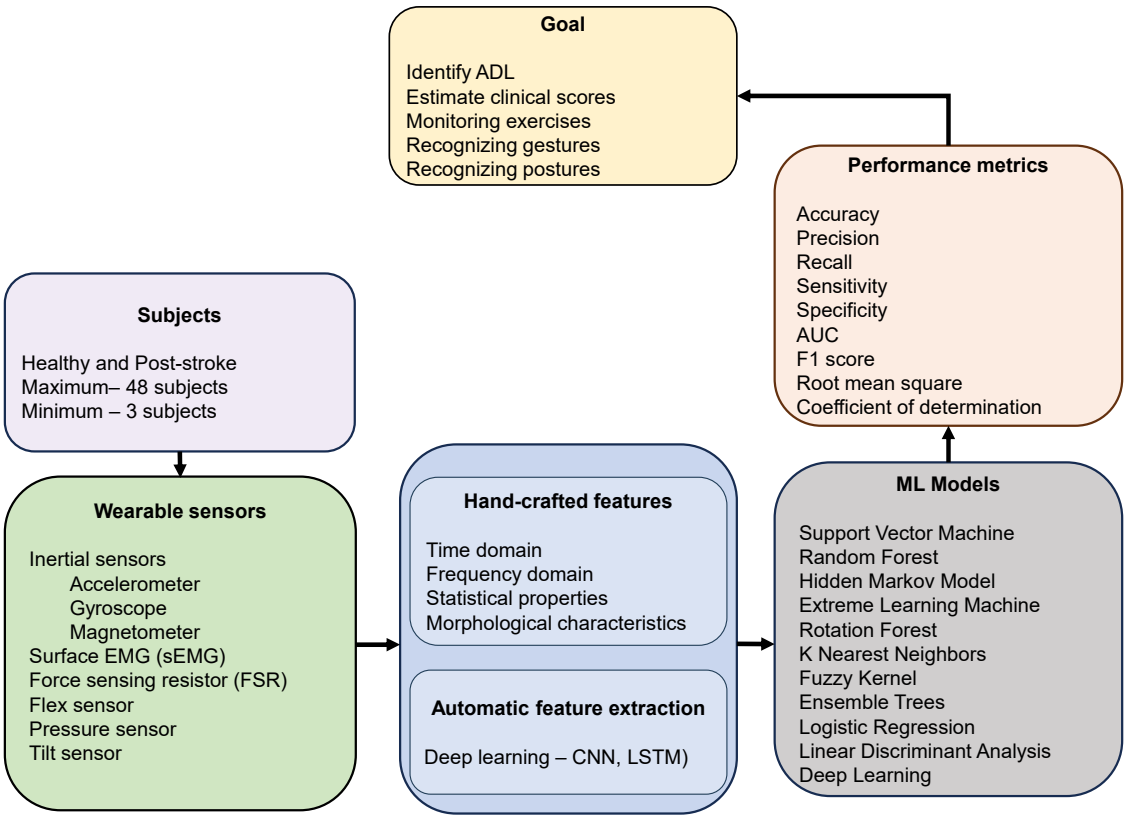


FIGURE 8: A summary of the articles presented in Table 2 including the research objectives, sensors, types of feature extraction, ML models (mostly classification models), and performance metrics used in the literature.

- The use of motor imagery (MI) is suitable for those who lack residual movement in their affected limbs. Still, it is incompatible with general motion-based wearable sensors, as no measurable movement occurs. The brain-computer interface (BCI), where electrodes are attached to the scalp, is not user-friendly, limiting its usability for patients alone, as they require additional assistance. Moreover, it is imperative to ascertain whether the brain rhythms observed during the MI process are attributed to MI.
- Accelerometers were the most frequently used sensors for evaluating body movement compared to other sensors. Nonetheless, combining data from diverse sensors enhances the monitoring quality capabilities and the prediction outcomes. Gyroscopes are better at measuring the quality of movement than accelerometers are. Additional sensors are often combined with accelerometers to obtain activity information from different modalities, enhancing the detection outcomes.
- Collecting and curating data from subjects and annotating data appropriately for use with ML models is a massive task. It is imperative to obtain training and establish guidelines to gather and manage personal health data while recognizing the privacy concerns of patients and the advancements in technology.
- Most methods generate window-based time-domain features and a few employ feature selection techniques. Nonetheless, the acquired knowledge derived from incorporating data from diverse domains will likely enhance ML models' learning and prediction outcomes. Therefore, it is essential to consider different aspects of the problem, context, and understanding before selecting valuable features. Identifying the optimal features and selecting the most suitable ones is imperative, mainly when dealing with hand-crafted features.
- Deep learning methods are becoming increasingly popular, reducing the need for expertise in hand-made feature extraction. However, understanding the intricacies of deep learning models and the availability of computational resources to build new models require human resources, time, and financial support. It also requires qualified people with theoretical and empirical domain knowledge.

VII. CONCLUSION

Recovery from stroke is a complex process that probably occurs through spontaneous and learning-dependent processes [144]. The stroke recovery process involves a series of distinct stages, and the duration of recovery and degree of distress varies according to the individual. Although we have traditional systems and methods for capturing data at high resolutions, they don't permit continuous examination,

This work is licensed under a Creative Commons Attribution 4.0 License. For more information, see https://creativecommons.org/licenses/by/4.0/

This article has been accepted for publication in IEEE Access. This is the author's version which has not been fully edited and
content may change prior to final publication. Citation information: DOI 10.1109/ACCESS.2024.3373910

as they are heavy and unportable.

Wearable devices with embedded sensors are highly appealing as they can capture high-resolution patient data while being portable, mobile, and wireless. Furthermore, wearable devices are practical for remote monitoring of stroke patients' daily activities and exercise schedules without a physician. Moreover, the ability of wearable devices to continuously collect patient data reduces missed opportunities for diagnosis and treatment.

By utilizing the predictive capability of ML models constructed using high-resolution wearable sensor data, we can develop and build systems for the automatic, remote, and precise monitoring of rehabilitation schedules, which was previously unfeasible. Specifically, combining high-resolution data from wearable devices and well-trained ML models enhances the monitoring quality in training intervention, assessment, and remote supervision.

This review has summarized the relevant research on stroke rehabilitation from 2009 to 2023 and cataloged the literature into six categories. These include (i) identifying ADL, (ii) estimating clinical scores, (iii) monitoring exercises, (iv) recognizing postures, (v) recognizing gestures, and (vi) deep learning methods. Furthermore, the review has highlighted the advantages and limitations of existing techniques and provided future research direction for adopting wearable devices in stroke rehabilitation, especially for remote monitoring.

We expect this review to benefit interested researchers in discovering the critical challenges of using wearable sensors and machine-learning techniques for post-stroke movement analysis. This review did not dig deep into individual features found in the literature due to the vast nature of the topic. We expect a future survey by the research community to complement this article and give readers additional insights. Lastly, we expect this review to provide directions and encourage the research community to explore new avenues.

REFERENCES

[1] H. Markus, A. Pereira, and G. Cloud, Stroke medicine. Oxford University Press, 2016.
[2] A. Di Carlo, "Human and economic burden of stroke," 2009.
[3] B. H. Dobkin, The clinical science of neurologic rehabilitation. Oxford University Press, 2003.
[4] S. E. Kasner, "Clinical interpretation and use of stroke scales," The Lancet Neurology, vol. 5, no. 7, pp. 603–612, 2006.
[5] World Health Organization, "International Classification of Functioning, Disability and Health (ICF)," https://www.who.int/standards/classifications/international-classification-of-functioning-disability-and-health, online; accessed 23 May 2023.
[6] A. Meigal, L. Gerasimova-Meigal, G. Rego, and D. Korzun, "Motor activity sensorics for mhealth support of human resilience in daily life," in 2022 32nd Conference of Open Innovations Association (FRUCT). IEEE, 2022, pp. 169–177.
[7] L. Mansson, P. Bäckman, F. Öhberg, J. Sandlund, J. Selling, and M. Sandlund, "Evaluation of concurrent validity between a smartphone self-test prototype and clinical instruments for balance and leg strength," Sensors, vol. 21, no. 5, p. 1765, 2021.
[8] C. L. Bentley, L. Powell, S. Potter, J. Parker, G. A. Mountain, Y. K. Bartlett, J. Farwer, C. O'Connor, J. Burns, R. L. Cresswell et al., "The use of a smartphone app and an activity tracker to promote physical activity in the management of chronic obstructive pulmonary disease: randomized controlled feasibility study," JMIR mHealth and uHealth, vol. 8, no. 6, p. e16203, 2020.
[9] F. Gür, G. C. Gür, and V. Ayan, "The effect of the erve smartphone app on physical activity, quality of life, self-efficacy, and exercise motivation for inactive people: a randomized controlled trial," European Journal of Integrative Medicine, vol. 39, p. 101198, 2020.
[10] M. Ishikawa, S. Yamada, K. Yamamoto, and Y. Aoyagi, "Gait analysis in a component timed-up-and-go test using a smartphone application," Journal of the neurological sciences, vol. 398, pp. 45–49, 2019.
[11] M. Capecci, L. Pepa, F. Verdini, and M. G. Ceravolo, "A smartphone-based architecture to detect and quantify freezing of gait in parkinson's disease," Gait & posture, vol. 50, pp. 28–33, 2016.
[12] S. Mellone, C. Tacconi, and L. Chiari, "Validity of a smartphone-based instrumented timed up and go," Gait & posture, vol. 36, no. 1, pp. 163–165, 2012.
[13] C. Dai, H. Zhang, E. Arens, and Z. Lian, "Machine learning approaches to predict thermal demands using skin temperatures: Steady-state conditions," Building and Environment, vol. 114, pp. 1–10, 2017.
[14] W.-T. Sung and S.-J. Hsiao, "The application of thermal comfort control based on smart house system of iot," Measurement, vol. 149, p. 106997, 2020.
[15] I. Boukhennoufa, X. Zhai, V. Utti, J. Jackson, and K. D. McDonald-Maier, "Wearable sensors and machine learning in post-stroke rehabilitation assessment: A systematic review," Biomedical Signal Processing and Control, vol. 71, p. 103197, 2022.
[16] S. Patel, R. Hughes, T. Hester, J. Stein, M. Akay, J. G. Dy, and P. Bonato, "A novel approach to monitor rehabilitation outcomes in stroke survivors using wearable technology," Proceedings of the IEEE, vol. 98, no. 3, pp. 450–461, 2010.
[17] S. H. Roy, M. S. Cheng, S.-S. Chang, J. Moore, G. De Luca, S. H. Nawab, and C. J. De Luca, "A combined semg and accelerometer system for monitoring functional activity in stroke," IEEE Transactions on Neural Systems and Rehabilitation Engineering, vol. 17, no. 6, pp. 585–594, 2009.
[18] J. Gubbi, A. S. Rao, K. Fang, B. Yan, and M. Palaniswami, "Motor recovery monitoring using acceleration measurements in post acute stroke patients," Biomedical engineering online, vol. 12, no. 1, p. 33, 2013.
[19] N. Friedman, J. B. Rowe, D. J. Reinkensmeyer, and M. Bachman, "The manumeter: a wearable device for monitoring daily use of the wrist and fingers," IEEE journal of biomedical and health informatics, vol. 18, no. 6, pp. 1804–1812, 2014.
[20] K. D. Nguyen, I.-M. Chen, Z. Luo, S. H. Yeo, and H. B.-L. Duh, "A wearable sensing system for tracking and monitoring of functional arm movement," IEEE/ASME Transactions on mechatronics, vol. 16, no. 2, pp. 213–220, 2011.
[21] M. Donno, E. Palange, F. D. Nicola, G. Bucci, and F. Ciancetta, "A new flexible optical fiber goniometer for dynamic angular measurements: Application to human joint movement monitoring," IEEE Transactions on Instrumentation and Measurement, vol. 57, no. 8, pp. 1614–1620, 2008.
[22] N. Lee, R. S. Goonetilleke, Y. S. Cheung, and G. M. So, "A flexible encapsulated mems pressure sensor system for biomechanical applications," Microsystem technologies, vol. 7, no. 2, pp. 55–62, 2001.
[23] E. McNamara, Z. Hudson, and S. J. Taylor, "Measuring activity levels of young people: the validity of pedometers," British medical bulletin, vol. 95, no. 1, pp. 121–137, 2010.
[24] L. Hackett, D. Reed, M. Halaki, and K. A. Ginn, "Assessing the validity of surface electromyography for recording muscle activation patterns from serratus anterior," Journal of Electromyography and Kinesiology, vol. 24, no. 2, pp. 221–227, 2014.
[25] S. Patel, H. Park, P. Bonato, L. Chan, and M. Rodgers, "A review of wearable sensors and systems with application in rehabilitation," Journal of neuroengineering and rehabilitation, vol. 9, no. 1, pp. 1–17, 2012.
[26] C. Chircov and A. M. Grumezescu, "Microelectromechanical systems (mems) for biomedical applications," Micromachines, vol. 13, no. 2, p. 164, 2022.
[27] P. Maceira-Elvira, T. Popa, A.-C. Schmid, and F. C. Hummel, "Wearable technology in stroke rehabilitation: towards improved diagnosis and treatment of upper-limb motor impairment," Journal of neuroengineering and rehabilitation, vol. 16, no. 1, p. 142, 2019.
[28] M. Kos and I. Kramberger, "A wearable device and system for movement and biometric data acquisition for sports applications," IEEE Access, vol. 5, pp. 6411–6420, 2017.

This work is licensed under a Creative Commons Attribution 4.0 License. For more information, see https://creativecommons.org/licenses/by/4.0/

This article has been accepted for publication in IEEE Access. This is the author's version which has not been fully edited and
content may change prior to final publication. Citation information: DOI 10.1109/ACCESS.2024.3373910

[29] T. Rukasha, S. I Woolley, T. Kyriacou, and T. Collins, "Evaluation of wearable electronics for epilepsy: A systematic review," Electronics, vol. 9, no. 6, p. 968, 2020.

[30] M. Khedr and N. El-Sheimy, "A smartphone step counter using imu and magnetometer for navigation and health monitoring applications," Sensors, vol. 17, no. 11, p. 2573, 2017.

[31] S. Milici, A. Lázaro, R. Villarino, D. Girbau, and M. Magnarosa, "Wireless wearable magnetometer-based sensor for sleep quality monitoring," IEEE Sensors Journal, vol. 18, no. 5, pp. 2145–2152, 2018.

[32] S. I. Lee, C. P. Adans-Dester, M. Grimaldi, A. V. Dowling, P. C. Horak, R. M. Black-Schaffer, P. Bonato, and J. T. Gwin, "Enabling stroke rehabilitation in home and community settings: A wearable sensor-based approach for upper-limb motor training," IEEE Journal of Translational Engineering in Health and Medicine, vol. 6, pp. 1–11, 2018.

[33] C. E. Lang, J. M. Wagner, D. F. Edwards, and A. W. Dromerick, "Upper extremity use in people with hemiparesis in the first few weeks after stroke," Journal of Neurologic Physical Therapy, vol. 31, no. 2, pp. 56–63, 2007.

[34] S. C. van der Pas, J. A. Verbunt, D. E. Breukelaar, R. van Woerden, and H. A. Seelen, "Assessment of arm activity using triaxial accelerometry in patients with a stroke," Archives of physical medicine and rehabilitation, vol. 92, no. 9, pp. 1437–1442, 2011.

[35] G. Uswatte, W. L. Foo, H. Olmstead, K. Lopez, A. Holand, and L. B. Simms, "Ambulatory monitoring of arm movement using accelerometry: an objective measure of upper-extremity rehabilitation in persons with chronic stroke," Archives of physical medicine and rehabilitation, vol. 86, no. 7, pp. 1498–1501, 2005.

[36] K. Saremi, J. Marehbian, X. Yan, J.-P. Regnaux, R. Elashoff, B. Bussel, and B. H. Dobkin, "Reliability and validity of bilateral thigh and foot accelerometry measures of walking in healthy and hemiparetic subjects," Neurorehabilitation and neural repair, vol. 20, no. 2, pp. 297–305, 2006.

[37] Z. Zhang, L. Liparulo, M. Panella, X. Gu, and Q. Fang, "A fuzzy kernel motion classifier for autonomous stroke rehabilitation," IEEE journal of biomedical and health informatics, vol. 20, no. 3, pp. 893–901, 2015.

[38] G. E. Hancock, T. Hepworth, and K. Wembridge, "Accuracy and reliability of knee goniometry methods," Journal of experimental orthopaedics, vol. 5, no. 1, pp. 1–6, 2018.

[39] A. Tognetti, F. Lorussi, G. D. Mura, N. Carbonaro, M. Pacelli, R. Paradiso, and D. D. Rossi, "New generation of wearable goniometers for motion capture systems," Journal of neuroengineering and rehabilitation, vol. 11, no. 1, pp. 1–17, 2014.

[40] N. Carbonaro, G. Dalle Mura, F. Lorussi, R. Paradiso, D. De Rossi, and A. Tognetti, "Exploiting wearable goniometer technology for motion sensing gloves," IEEE journal of biomedical and health informatics, vol. 18, no. 6, pp. 1788–1795, 2014.

[41] S. M. Bierma-Zeinstra, A. Bohnen, R. Ramlal, J. Ridderikhoff, J. A. Verhaar, and A. Prins, "Comparison between two devices for measuring hip joint motions," Clinical rehabilitation, vol. 12, no. 6, pp. 497–505, 1998.

[42] J. Keeley, T. G. Mayer, R. Cox, R. J. Gatchel, J. Smith, and V. Mooney, "Quantification of lumbar function. part 5: Reliability of range-of-motion measures in the sagittal plane and an in vivo torso rotation measurement technique." Spine, vol. 11, no. 1, pp. 31–35, 1986.

[43] D. Solanki and U. Lahiri, "Design of instrumented shoes for gait characterization: A usability study with healthy and post-stroke hemiplegic individuals," Frontiers in neuroscience, vol. 12, p. 459, 2018.

[44] S. K. Manna, M. H. B. Azhar, and A. Greace, "Optimal locations and computational frameworks of fsr and imu sensors for measuring gait abnormalities," Heliyon, 2023.

[45] A. Tognetti, F. Lorussi, N. Carbonaro, D. De Rossi, G. De Toma, C. Mancuso, R. Paradiso, H. Luinge, J. Reenalda, E. Droog et al., "Daily-life monitoring of stroke survivors motor performance: The interaction sensing system," in 2014 36th Annual International Conference of the IEEE Engineering in Medicine and Biology Society. IEEE, 2014, pp. 4099–4102.

[46] Z. G. Xiao and C. Menon, "Towards the development of a wearable feedback system for monitoring the activities of the upper-extremities," Journal of neuroengineering and rehabilitation, vol. 11, no. 1, p. 2, 2014.

[47] K. Aminian, B. Najafi, C. Büla, P.-F. Leyvraz, and P. Robert, "Spatio-temporal parameters of gait measured by an ambulatory system using miniature gyroscopes," Journal of biomechanics, vol. 35, no. 5, pp. 689–699, 2002.

[48] S. J. M. Bamberg, A. Y. Benbasat, D. M. Scarborough, D. E. Krebs, and J. A. Paradiso, "Gait analysis using a shoe-integrated wireless sensor system," IEEE transactions on information technology in biomedicine, vol. 12, no. 4, pp. 413–423, 2008.

[49] P. Lopez-Meyer, G. D. Fulk, and E. S. Sazonov, "Automatic detection of temporal gait parameters in poststroke individuals," IEEE Transactions on Information Technology in Biomedicine, vol. 15, no. 4, pp. 594–601, 2011.

[50] R. T. Li, S. R. Kling, M. J. Salata, S. A. Cupp, J. Sheehan, and J. E. Voos, "Wearable performance devices in sports medicine," Sports health, vol. 8, no. 1, pp. 74–78, 2016.

[51] N. Zhao, "Full-featured pedometer design realized with 3-axis digital accelerometer," https://www.analog.com/media/en/technical-documentation/technical-articles/pedometer.pdf, Analog Devices, Tech. Rep. 44-06, 06 2010, accessed: 2023-07-27.

[52] C. Vanroy, D. Vissers, P. Cras, S. Beyne, H. Feys, Y. Vanlandewijck, and S. Truijen, "Physical activity monitoring in stroke: Sensewear pro2 activity accelerometer versus yamax digi-walker sw-200 pedometer," Disability and rehabilitation, vol. 36, no. 20, pp. 1695–1703, 2014.

[53] E. F. Shepherd, E. Toloza, C. D. McClung, and T. P. Schmalzried, "Step activity monitor: increased accuracy in quantifying ambulatory activity," Journal of Orthopaedic Research, vol. 17, no. 5, pp. 703–708, 1999.

[54] D. Farina and F. Negro, "Accessing the neural drive to muscle and translation to neurorehabilitation technologies," IEEE Reviews in biomedical engineering, vol. 5, pp. 3–14, 2012.

[55] M. Al-Ayyad, H. A. Owida, R. De Fazio, B. Al-Naami, and P. Visconti, "Electromyography monitoring systems in rehabilitation: A review of clinical applications, wearable devices and signal acquisition methodologies," Electronics, vol. 12, no. 7, p. 1520, 2023.

[56] C. K. Reddy and C. C. Aggarwal, Healthcare data analytics. CRC Press, 2015, vol. 36.

[57] K. Adane, M. Gizachew, and S. Kendie, "The role of medical data in efficient patient care delivery: a review," Risk management and healthcare policy, pp. 67–73, 2019.

[58] A. Burkov, The Hundred-page Machine Learning Book. Andriy Burkov, 2019. [Online]. Available: https://themlbook.com/

[59] C. Cortes and V. Vapnik, "Support-vector networks," Machine learning, vol. 20, pp. 273–297, 1995.

[60] E. Fix and J. L. Hodges, "Discriminatory analysis. nonparametric discrimination: Consistency properties," International Statistical Review/Revue Internationale de Statistique, vol. 57, no. 3, pp. 238–247, 1989.

[61] T. Hastie, R. Tibshirani, J. H. Friedman, and J. H. Friedman, The elements of statistical learning: data mining, inference, and prediction. Springer, 2009, vol. 2.

[62] I. Goodfellow, Y. Bengio, and A. Courville, Deep learning. MIT press, 2016, http://www.deeplearningbook.org.

[63] J. MacQueen et al., "Some methods for classification and analysis of multivariate observations," in Proceedings of the fifth Berkeley symposium on mathematical statistics and probability, vol. 1, no. 14. Oakland, CA, USA, 1967, pp. 281–297.

[64] J. C. Bezdek, R. Ehrlich, and W. Full, "Fcm: The fuzzy c-means clustering algorithm," Computers & geosciences, vol. 10, no. 2-3, pp. 191–203, 1984.

[65] F. Nielsen and F. Nielsen, "Hierarchical clustering," Introduction to HPC with MPI for Data Science, pp. 195–211, 2016.

[66] M. Ester, H.-P. Kriegel, J. Sander, X. Xu et al., "A density-based algorithm for discovering clusters in large spatial databases with noise," in kdd, vol. 96, no. 34, 1996, pp. 226–231.

[67] C. M. Bishop, Pattern recognition and machine learning. Springer, 2006.

[68] I. T. Jolliffe and J. Cadima, "Principal component analysis: a review and recent developments," Philosophical transactions of the royal society A: Mathematical, Physical and Engineering Sciences, vol. 374, no. 2065, p. 20150202, 2016.

[69] O. Chapelle, B. Scholkopf, and A. Zien, "Semi-supervised learning. 2006," Cambridge, Massachusettes: The MIT Press View Article, vol. 2, 2006.

[70] C. J. Wu, "On the convergence properties of the em algorithm," The Annals of statistics, pp. 95–103, 1983.

[71] Y. Huang, D. Yu, Y. Gong, and C. Liu, "Semi-supervised gmm and dnn acoustic model training with multi-system combination and confidence re-calibration." in Interspeech, 2013, pp. 2360–2364.

[72] R. S. Sutton and A. G. Barto, Reinforcement learning: An introduction. MIT press, 2018.

This work is licensed under a Creative Commons Attribution 4.0 License. For more information, see https://creativecommons.org/licenses/by/4.0/

This article has been accepted for publication in IEEE Access. This is the author's version which has not been fully edited and
content may change prior to final publication. Citation information: DOI 10.1109/ACCESS.2024.3373910

[73] J. Schulman, S. Levine, P. Abbeel, M. Jordan, and P. Moritz, "Trust region policy optimization," in International conference on machine learning. PMLR, 2015, pp. 1889–1897.
[74] Y. Wang, H. He, and X. Tan, "Truly proximal policy optimization," in Uncertainty in Artificial Intelligence. PMLR, 2020, pp. 113–122.
[75] B. Jang, M. Kim, G. Harerimana, and J. W. Kim, "Q-learning algorithms: A comprehensive classification and applications," IEEE access, vol. 7, pp. 133 653–133 667, 2019.
[76] V. Mnih, K. Kavukcuoglu, D. Silver, A. A. Rusu, J. Veness, M. G. Bellemare, A. Graves, M. Riedmiller, A. K. Fidjeland, G. Ostrovski et al., "Human-level control through deep reinforcement learning," nature, vol. 518, no. 7540, pp. 529–533, 2015.
[77] M. Kuhn, K. Johnson et al., Applied predictive modeling. Springer, 2013, vol. 26.
[78] N. Pudjihartono, T. Fadason, A. W. Kempa-Liehr, and J. M. O'Sullivan, "A review of feature selection methods for machine learning-based disease risk prediction," Frontiers in Bioinformatics, vol. 2, p. 927312, 2022.
[79] K.-R. Müller, S. Mika, G. Rätsch, K. Tsuda, and B. Schölkopf, "An introduction to kernel-based learning algorithms," IEEE Transactions on Neural Networks, vol. 12, no. 2, pp. 181–201, 2001.
[80] C.-C. Chang and C.-J. Lin, "Libsvm: a library for support vector machines," ACM transactions on intelligent systems and technology (TIST), vol. 2, no. 3, pp. 1–27, 2011.
[81] ——, "LIBSVM – A Library for Support Vector Machines," http://www.csie.ntu.edu.tw/ cjlin/libsvm/, [Online; accessed 02-Feb-2015].
[82] L. Breiman, "Random forests," Machine learning, vol. 45, pp. 5–32, 2001.
[83] ——, Classification and regression trees. Routledge, 2017.
[84] F. Rosenblatt, "The perceptron: a probabilistic model for information storage and organization in the brain." Psychological review, vol. 65, no. 6, p. 386, 1958.
[85] D. E. Rumelhart, G. E. Hinton, and R. J. Williams, "Learning representations by back-propagating errors," nature, vol. 323, no. 6088, pp. 533–536, 1986.
[86] A. Krizhevsky, I. Sutskever, and G. E. Hinton, "Imagenet classification with deep convolutional neural networks," Advances in neural information processing systems, vol. 25, 2012.
[87] Y. LeCun, Y. Bengio, and G. Hinton, "Deep learning," nature, vol. 521, no. 7553, pp. 436–444, 2015.
[88] C. Yang, W. Jiang, and Z. Guo, "Time series data classification based on dual path cnn-rnn cascade network," IEEE Access, vol. 7, pp. 155 304–155 312, 2019.
[89] P. Ivaturi, M. Gadaleta, A. C. Pandey, M. Pazzani, S. R. Steinhubl, and G. Quer, "A comprehensive explanation framework for biomedical time series classification," IEEE journal of biomedical and health informatics, vol. 25, no. 7, pp. 2398–2408, 2021.
[90] M. Panwar, D. Biswas, H. Bajaj, M. Jöbges, R. Turk, K. Maharatna, and A. Acharyya, "Rehab-net: Deep learning framework for arm movement classification using wearable sensors for stroke rehabilitation," IEEE Transactions on Biomedical Engineering, vol. 66, no. 11, pp. 3026–3037, 2019.
[91] B. B. Nair and N. Sakthivel, "A deep learning-based upper limb rehabilitation exercise status identification system," Arabian Journal for Science and Engineering, vol. 48, no. 2, pp. 1237–1271, 2023.
[92] S. Hochreiter and J. Schmidhuber, "Long short-term memory," Neural computation, vol. 9, no. 8, pp. 1735–1780, 1997.
[93] F. A. Gers, N. N. Schraudolph, and J. Schmidhuber, "Learning precise timing with lstm recurrent networks," Journal of machine learning research, vol. 3, no. Aug, pp. 115–143, 2002.
[94] Y. Yu, X. Si, C. Hu, and J. Zhang, "A review of recurrent neural networks: Lstm cells and network architectures," Neural computation, vol. 31, no. 7, pp. 1235–1270, 2019.
[95] A. Vaswani, N. Shazeer, N. Parmar, J. Uszkoreit, L. Jones, A. N. Gomez, Ł. Kaiser, and I. Polosukhin, "Attention is all you need," Advances in neural information processing systems, vol. 30, 2017.
[96] U. Kamath, K. Graham, and W. Emara, Transformers for Machine Learning: A Deep Dive. Chapman and Hall/CRC, 2022.
[97] Y. Tian, Y. Zhang, and H. Zhang, "Recent advances in stochastic gradient descent in deep learning," Mathematics, vol. 11, no. 3, p. 682, 2023.
[98] P. Probst, A.-L. Boulesteix, and B. Bischl, "Tunability: Importance of hyperparameters of machine learning algorithms," The Journal of Machine Learning Research, vol. 20, no. 1, pp. 1934–1965, 2019.
[99] I. H. Witten, E. Frank, M. A. Hall, and C. J. Pal, Data Mining, Fourth Edition: Practical Machine Learning Tools and Techniques, 4th ed. San Francisco, CA, USA: Morgan Kaufmann Publishers Inc., 2016.
[100] M. Arif and A. Kattan, "Physical activities monitoring using wearable acceleration sensors attached to the body," PloS one, vol. 10, no. 7, p. e0130851, 2015.
[101] G. P. Sadarangani, X. Jiang, L. A. Simpson, J. J. Eng, and C. Menon, "Force myography for monitoring grasping in individuals with stroke with mild to moderate upper-extremity impairments: A preliminary investigation in a controlled environment," Frontiers in bioengineering and biotechnology, vol. 5, p. 42, 2017.
[102] M. K. O'Brien, N. Shawen, C. K. Mummidisetty, S. Kaur, X. Bo, C. Poellabauer, K. Kording, and A. Jayaraman, "Activity recognition for persons with stroke using mobile phone technology: toward improved performance in a home setting," Journal of medical Internet research, vol. 19, no. 5, 2017.
[103] T. Tran, L.-C. Chang, I. Almubark, E. M. Bochniewicz, L. Shu, P. S. Lum, and A. Dromerick, "Robust classification of functional and nonfunctional arm movement after stroke using a single wrist-worn sensor device," in 2018 IEEE international conference on big data (big data). IEEE, 2018, pp. 5457–5459.
[104] E. M. Bochniewicz, G. Emmer, A. McLeod, J. Barth, A. W. Dromerick, and P. Lum, "Measuring functional arm movement after stroke using a single wrist-worn sensor and machine learning," Journal of Stroke and Cerebrovascular Diseases, vol. 26, no. 12, pp. 2880–2887, 2017.
[105] P.-W. Chen, N. A. Baune, I. Zwir, J. Wang, V. Swamidass, and A. W. Wong, "Measuring activities of daily living in stroke patients with motion machine learning algorithms: A pilot study," International journal of environmental research and public health, vol. 18, no. 4, p. 1634, 2021.
[106] S. L. Wolf, J. P. McJunkin, M. L. Swanson, and P. S. Weiss, "Pilot normative database for the wolf motor function test," Archives of physical medicine and rehabilitation, vol. 87, no. 3, pp. 443–445, 2006.
[107] L. Yu, D. Xiong, L. Guo, and J. Wang, "A remote quantitative fugl-meyer assessment framework for stroke patients based on wearable sensor networks," Computer methods and programs in biomedicine, vol. 128, pp. 100–110, 2016.
[108] B. Oubre, J.-F. Daneault, H.-T. Jung, K. Whritenour, J. G. V. Miranda, J. Park, T. Ryu, Y. Kim, and S. I. Lee, "Estimating upper-limb impairment level in stroke survivors using wearable inertial sensors and a minimally-burdensome motor task," IEEE Transactions on Neural Systems and Rehabilitation Engineering, vol. 28, no. 3, pp. 601–611, 2020.
[109] E. Park, K. Lee, T. Han, and H. S. Nam, "Automatic grading of stroke symptoms for rapid assessment using optimized machine learning and 4-limb kinematics: clinical validation study," Journal of medical Internet research, vol. 22, no. 9, p. e20641, 2020.
[110] N. V. Chawla, K. W. Bowyer, L. O. Hall, and W. P. Kegelmeyer, "Smote: synthetic minority over-sampling technique," Journal of artificial intelligence research, vol. 16, pp. 321–357, 2002.
[111] C. Adans-Dester, N. Hankov, A. O'Brien, G. Vergara-Diaz, R. Black-Schaffer, R. Zafonte, J. Dy, S. I. Lee, and P. Bonato, "Enabling precision rehabilitation interventions using wearable sensors and machine learning to track motor recovery," NPJ digital medicine, vol. 3, no. 1, p. 121, 2020.
[112] J.-I. Pan, H.-W. Chung, and J.-J. Huang, "Intelligent shoulder joint home-based self-rehabilitation monitoring system," Int. J. Smart Home, vol. 7, no. 5, pp. 395–404, 2013.
[113] A. Yurtman and B. Barshan, "Automated evaluation of physical therapy exercises using multi-template dynamic time warping on wearable sensor signals," Computer methods and programs in biomedicine, vol. 117, no. 2, pp. 189–207, 2014.
[114] L. Yu, D. Xiong, L. Guo, and J. Wang, "A compressed sensing-based wearable sensor network for quantitative assessment of stroke patients," Sensors, vol. 16, no. 2, p. 202, 2016.
[115] E. S. Sazonov, G. Fulk, J. Hill, Y. Schutz, and R. Browning, "Monitoring of posture allocations and activities by a shoe-based wearable sensor," IEEE Transactions on Biomedical Engineering, vol. 58, no. 4, pp. 983–990, 2010.
[116] G. D. Fulk and E. Sazonov, "Using sensors to measure activity in people with stroke," Topics in stroke rehabilitation, vol. 18, no. 6, pp. 746–757, 2011.
[117] G. D. Fulk, S. R. Edgar, R. Bierwirth, P. Hart, P. Lopez-Meyer, and E. Sazonov, "Identifying activity levels and steps in people with stroke using a novel shoe-based sensor," Journal of Neurologic Physical Therapy, vol. 36, no. 2, p. 100, 2012.

This work is licensed under a Creative Commons Attribution 4.0 License. For more information, see https://creativecommons.org/licenses/by/4.0/

This article has been accepted for publication in IEEE Access. This is the author's version which has not been fully edited and
content may change prior to final publication. Citation information: DOI 10.1109/ACCESS.2024.3373910

[118] J. Cheng, X. Chen, and M. Shen, "A framework for daily activity monitoring and fall detection based on surface electromyography and accelerometer signals," IEEE journal of biomedical and health informatics, vol. 17, no. 1, pp. 38–45, 2013.

[119] S. R. Eddy, "What is a hidden markov model?" Nature biotechnology, vol. 22, no. 10, pp. 1315–1316, 2004.

[120] S. Zhong and J. Ghosh, "Hmms and coupled hmms for multi-channel eeg classification," in Proceedings of the 2002 International Joint Conference on Neural Networks. IJCNN'02 (Cat. No. 02CH37290), vol. 2. IEEE, 2002, pp. 1154–1159.

[121] F. Masse, R. Gonzenbach, A. Paraschiv-Ionescu, A. R. Luft, and K. Aminian, "Wearable barometric pressure sensor to improve postural transition recognition of mobility-impaired stroke patients," IEEE transactions on neural systems and rehabilitation engineering, vol. 24, no. 11, pp. 1210–1217, 2016.

[122] G. Yang, J. Deng, G. Pang, H. Zhang, J. Li, B. Deng, Z. Pang, J. Xu, M. Jiang, P. Liljeberg et al., "An iot-enabled stroke rehabilitation system based on smart wearable armband and machine learning," IEEE journal of translational engineering in health and medicine, vol. 6, pp. 1–10, 2018.

[123] A. Kaku, A. Parnandi, A. Venkatesan, N. Pandit, H. Schambra, and 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.

[124] S. H. Chae, Y. Kim, K.-S. Lee, and 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 and uHealth, vol. 8, no. 7, p. e17216, 2020.

[125] E. S. Sazonov, G. Fulk, J. Hill, Y. Schutz, and R. Browning, "Monitoring of posture allocations and activities by a shoe-based wearable sensor," IEEE Transactions on Biomedical Engineering, vol. 58, no. 4, pp. 983–990, 2011.

[126] H.-S. Cheng, M.-S. Ju, and C.-C. K. Lin, "Improving elbow torque output of stroke patients with assistive torque controlled by emg signals," Journal of biomechanical engineering, vol. 125, no. 6, pp. 881–886, 2003.

[127] Y.-L. Hsieh and M. F. Abbod, "Gait analyses of parkinson's disease patients using multiscale entropy," Electronics, vol. 10, no. 21, p. 2604, 2021.

[128] T. Inagaki, Y. Akiyama, S. Okamoto, T. Mayumi, and Y. Yamada, "Relationship between gait stability indices and gait parameters comprising joint angles based on walking data from 288 people," Nagoya Journal of Medical Science, vol. 85, no. 2, p. 211, 2023.

[129] A. Ekizos, A. Santuz, A. Schroll, and A. Arampatzis, "The maximum lyapunov exponent during walking and running: Reliability assessment of different marker-sets," Frontiers in physiology, vol. 9, p. 1101, 2018.

[130] P. Zhang and J. Zhang, "Deep learning analysis based on multi-sensor fusion data for hemiplegia rehabilitation training system for stoke patients," Robotica, vol. 40, no. 3, pp. 780–797, 2022.

[131] S. Qiu, Z. Wang, H. Zhao, L. Liu, and Y. Jiang, "Using body-worn sensors for preliminary rehabilitation assessment in stroke victims with gait impairment," IEEE Access, vol. 6, pp. 31 249–31 258, 2018.

[132] Y.-A. Choi, S.-J. Park, J.-A. Jun, C.-S. Pyo, K.-H. Cho, H.-S. Lee, and J.-H. Yu, "Deep learning-based stroke disease prediction system using real-time bio signals," Sensors, vol. 21, no. 13, p. 4269, 2021.

[133] S. Collange, D. Defour, and A. Tisserand, "Power consumption of gpus from a software perspective," in Computational Science–ICCS 2009: 9th International Conference Baton Rouge, LA, USA, May 25-27, 2009 Proceedings, Part I 9. Springer, 2009, pp. 914–923.

[134] M. Halpern, Y. Zhu, and V. J. Reddi, "Mobile cpu's rise to power: Quantifying the impact of generational mobile cpu design trends on performance, energy, and user satisfaction," in 2016 IEEE International Symposium on High Performance Computer Architecture (HPCA). IEEE, 2016, pp. 64–76.

[135] N. P. Jouppi, C. Young, N. Patil, D. Patterson, G. Agrawal, R. Bajwa, S. Bates, S. Bhatia, N. Boden, A. Borchers et al., "In-datacenter performance analysis of a tensor processing unit," in Proceedings of the 44th annual international symposium on computer architecture, 2017, pp. 1–12.

[136] A. Pantelopoulos and N. G. Bourbakis, "A survey on wearable sensor-based systems for health monitoring and prognosis," IEEE Transactions on Systems, Man, and Cybernetics, Part C (Applications and Reviews), vol. 40, no. 1, pp. 1–12, 2009.

[137] E. Covi, E. Donati, X. Liang, D. Kappel, H. Heidari, M. Payvand, and W. Wang, "Adaptive extreme edge computing for wearable devices," Frontiers in Neuroscience, vol. 15, p. 611300, 2021.

[138] P. U. Diehl and M. Cook, "Efficient implementation of stdp rules on spinnaker neuromorphic hardware," in 2014 International Joint Conference on Neural Networks (IJCNN). IEEE, 2014, pp. 4288–4295.

[139] G. K. Chen, R. Kumar, H. E. Sumbul, P. C. Knag, and R. K. Krishnamurthy, "A 4096-neuron 1m-synapse 3.8-pj/sop spiking neural network with on-chip stdp learning and sparse weights in 10-nm finfet cmos," IEEE Journal of Solid-State Circuits, vol. 54, no. 4, pp. 992–1002, 2018.

[140] J. Clusmann, F. R. Kolbinger, H. S. Muti, Z. I. Carrero, J.-N. Eckardt, N. G. Laleh, C. M. L. Löffler, S.-C. Schwarzkopf, M. Unger, G. P. Veldhuizen et al., "The future landscape of large language models in medicine," Communications Medicine, vol. 3, no. 1, p. 141, 2023.

[141] P. J. Soh, G. A. Vandenbosch, M. Mercuri, and D. M.-P. Schreurs, "Wearable wireless health monitoring: Current developments, challenges, and future trends," IEEE Microwave Magazine, vol. 16, no. 4, pp. 55–70, 2015.

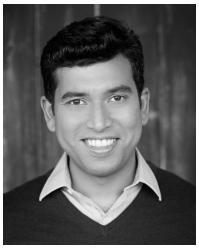
[142] M. A. Khan and K. Salah, "Iot security: Review, blockchain solutions, and open challenges," Future Generation Computer Systems, vol. 82, pp. 395–411, 2018.

[143] C. Dwork and A. Roth, "The Algorithmic Foundations of Differential Privacy," Foundations and Trends® in Theoretical Computer Science, vol. 9, no. 3–4, pp. 211–407, Aug. 2014.

[144] P. Langhorne, J. Bernhardt, and G. Kwakkel, "Stroke rehabilitation," The Lancet, vol. 377, no. 9778, pp. 1693–1702, 2011.



NANDINI SENGUPTA received the B.Tech. degree in Electronics and Communication Engineering from West Bengal University of Technology, Kolkata, India, in 2012 and the M.S. (by Research) from the Department of Electronics and Electrical Communication Engineering, Indian Institute of Technology Kharagpur, Kharagpur, India, in 2017. She is working toward a Ph.D. in the Department of Electrical and Electronic Engineering at the University of Melbourne, Parkville, Australia. Her research interests include biomedical signal processing, machine learning, and stroke rehabilitation.



ARAVINDA S. RAO (Senior Member, IEEE) received the PhD degree from the Department of Electrical and Electronic Engineering, University of Melbourne, Parkville, VIC, Australia, in 2016, where he is working as a Research Fellow with the Department of Electrical and Electronic Engineering. His research interests include the Internet of Things, computer vision, machine learning, and artificial intelligence. He is a reviewer for several journals, including IEEE Internet of Things Journal, IEEE Transactions on Multimedia, and IEEE Transactions on Circuits and Systems for Video Technology.

This work is licensed under a Creative Commons Attribution 4.0 License. For more information, see https://creativecommons.org/licenses/by/4.0/

This article has been accepted for publication in IEEE Access. This is the author's version which has not been fully edited and
content may change prior to final publication. Citation information: DOI 10.1109/ACCESS.2024.3373910



BERNARD YAN is a dual-trained Neurologist and Endovascular Neurointerventionist (a sub-specialty which treats acute stroke, aneurysms and arterio-venous malformation). He completed specialty training in Neurology in Melbourne, Australia, and a two-year subspecialty fellowship in endovascular neuro-intervention in Frankfurt, Germany. His clinical expertise is in the treatment of stroke and aneurysms. Management of acute stroke has undergone a paradigm shift in the past ten years, and in particular, acute mechanical thrombectomy (a minimally invasive technique to unblock brain arteries) is now the gold standard of care for such patients. He is the principal investigator of several international studies in cerebrovascular diseases and is the Co-principal investigator of a multicenter research DIRECT-SAFE. He is an awardee of academic grants by the National Health and Medical Research Council (NHMRC). One of his fundamental research interests is the development of portable mobile wireless sensors for monitoring patients with neurological diseases. These devices are designed to monitor the motor recovery of stroke patients and the frequency of epileptic seizures in patients with epilepsy.



MARIMUTHU PALANISWAMI (Fellow, IEEE) received the M.E. degree from the Indian Institute of Science Bengaluru, Bengaluru, India, in 1977, the M.Eng.Sc. Degree from the University of Melbourne, Parkville, VIC, Australia, in 1983, and a Ph.D. from the University of Newcastle, Callaghan, NSW, Australia, in 1987. He is a Professor at the Department of Electrical and Electronic Engineering, University of Melbourne, Melbourne, VIC, Australia. He has published over 400 refereed research articles and led one of the largest funded Australian research councils, a research network on intelligent sensors, sensor networks, and information processing programs. He has been funded by several ARC and industry grants (more than $40M) to conduct research in support vector machines, sensors, sensor networks, the IoT, machine learning, neural networks, pattern recognition, signal processing, and control areas. He represents Australia as a core partner in European Union FP7 projects, such as SENSEI, Smart Santander, the Internet of Things Initiative, and SocIoTal.

This work is licensed under a Creative Commons Attribution 4.0 License. For more information, see https://creativecommons.org/licenses/by/4.0/