Artificial Intelligence and Wearable Technologies for Upper Limb Neurorehabilitation

Ilaria Siviero, *Member, IEEE*, Nicola Valè, *Member, IEEE*, Gloria Menegaz, *Senior Member, IEEE*, Ander Ramos-Murguialday, *Member, IEEE*, and Silvia Francesca Storti, *Member, IEEE*

***Abstract*—Non-invasive neural interfaces (NIs) are increasingly investigated in upper limb neurorehabilitation, where they exploit biosignals, such as electroencephalography (EEG) and electromyography (EMG), to decode motor intentions using artificial intelligence (AI). Yet, traditional systems are complex and difficult to use outside the clinic. Wearable devices have the potential for innovative neurorehabilitation solutions thanks to their comfort, easy-to-use and long-term monitoring. However, current AI approaches require adaptation to the technical constraints of wearable devices, and the related state-of-the-art is not clearly explained and summarized. In this work, a systematic literature review on 51 studies was conducted analyzing them according to five important concepts: biosignals, wearable devices, AI-driven methods, upper limb, and clinical applications. The review highlights methodological heterogeneity, a variety of wearable sensor configurations, and open challenges related to accuracy, robustness, and clinical validation. Finally, we discuss how explainable AI (XAI) and generative AI (GenAI) may contribute to improve the interpretability and personalization of future neurorehabilitation systems.**

***Index Terms*—Artificial intelligence, wearable sensors, upper limb, rehabilitation, telemedicine.**

I. Introduction

NEUROLOGICAL conditions that affect the central nervous system (CNS), such as stroke, multiple sclerosis or traumatic brain injury, can lead to several upper limb (UL) sensorimotor dysfunctions. Stroke and traumatic brain injury have the highest incidence and prevalence, with around 11.9 million cases of stroke each year and 27 million cases of traumatic brain injury every year, globally [1], [2]. It is estimated that 94 million people live with stroke sequelae and 55.5 million people suffer from the consequences of a traumatic brain injury [1], [2], while 2.2 million people live with multiple sclerosis [3].

Rehabilitation is the process involved in managing and promoting functional recovery, restoring activity and improving social participation and quality of life. Effective intervention requires intensive, repetitive, task-oriented, and engaging exercises [4], [5], [6], aligning with the principles of experience-dependent neuroplasticity described by Kleim and Jones [7], which remarks on the key role of neuroplasticity in functional recovery.

Task-oriented approaches, like the constraint-induced movement therapy (CIMT) - and its modified versions - have shown promising results in functional recovery, boosting task-specific use of the contralesional UL through intensive and graded practice [8], [9]. However, these approaches suffer from some limitations. They are difficult to implement in home settings due to the use of constraining devices and the intensive time required. Furthermore, they are de facto limited to patients with mild to moderate levels of functional impairment [10].

For severely impaired subjects, other strategies have been explored, including the use of robotic devices, possibly integrated into virtual reality (VR) environment [5], [11], [12], functional electrical stimulation, or mental practice as in motor imagery (MI) training [13], [14]. Robot-assisted UL neurorehabilitation has been widely studied in the last two decades, showing promising results in functional recovery [15]. Rehabilitation robotics can be used in different modalities. From passive mobilization with a speed and range of motion defined according to personalized needs, to partial assistance, facilitating the user's movement in case of less impaired UL function [16]. A recent meta-analysis by Boardsworth et al. [17] suggested that the level of assistance provided by the device can impact rehabilitation outcomes. The preliminary results show that partial assistance could be more effective than total assistance. Therefore, the definition of the level of assistance provided is crucial, and, from a rehabilitation perspective, should adapt to the individual's residual functional level. However, in most of the studies included in their review, a fixed amount of force was provided towards the target of movement, irrespective of the actual effort and performance of the user [17]. While, in principle, bioelectrical signals

Received 20 January 2025; revised 1 October 2025 and 10 December 2025; accepted 30 December 2025. Date of publication 12 January 2026; date of current version 30 January 2026. This work was supported in part by the University of Verona (Italy) Ricerca di Base 2022. The work of Ilaria Siviero was supported by Recovery Assistance for Cohesion and the Territories (REACT) European Union (EU)-Programma Operativo Nazionale (PON) "Ricerca e Innovazione" 2014–2020. The work of Nicola Valè, Gloria Menegaz, and Silvia Francesca Storti was supported by NEXTGENERATIONEU (NGEU) funded by the Ministry of University and Research (MUR), National Recovery and Resilience Plan (NRRP), and project MNESYS (PE0000006) – "A Multiscale integrated approach to the study of the nervous system in health and disease" (DN. 1553 11.10.2022). *(Corresponding author: Silvia Francesca Storti.)*

Ilaria Siviero, Nicola Valè, Gloria Menegaz, and Silvia Francesca Storti are with the Department of Engineering for Innovation Medicine, University of Verona, 37134 Verona, Italy (e-mail: ilaria.siviero@univr.it; nicola.vale@univr.it; gloria.menegaz@univr.it; silviafrancesca.storti@univr.it).

Ander Ramos-Murguialday is with the Institute of Medical Psychology and Behavioral Neurobiology and the Department of Neurology and Stroke, University of Tübingen, 72074 Tübingen, Germany, also with TECNALIA, Basque Research and Technology Alliance, 20008 San Sebastian, Spain, and also with the Department of Neurology and Stroke, University of Tübingen, 72074 Tübingen, Germany (e-mail: ander.ramos@gmail.com).

This article has supplementary downloadable material available at https://doi.org/10.1109/TNSRE.2026.3651949, provided by the authors.

Digital Object Identifier 10.1109/TNSRE.2026.3651949

and dynamic data can be used to decode movement intention and control a robotic device, the use of these modalities in clinical studies is still limited due to the lack of standardized approaches and technical challenges, especially in decoding electromyography (EMG) data [18].

In the case of complete paresis, MI can be the only option to prompt movement intention. Although sensory feedback is not available when performing MI, sensorimotor learning is thought to occur through the feed-forward mechanism triggered by mental movement simulation [14].

The development of neural interfaces (NIs) represents a promising solution. They are systems that record neuronal or neuromuscular electrical signals and translate this information into external commands, providing feedback to the user or interpreting their intentions [19], [20], [21]. NIs based on brain activity recording are commonly called brain computer interface (BCI) or brain machine interface (BMI) [13], [20], [21]. From a neurorehabilitation perspective, NIs can control robotic devices or assess movement intention and execution. In both cases, decoding neuronal or neuromuscular activity through electroencephalography (EEG) and EMG improves the accuracy of the feedback provided. This allows closing the sensorimotor loop, even when voluntary movement is not possible, as in MI. NIs can control a robotic device using muscular activity through EMG signal acquisition and decoding, allowing robot-assisted rehabilitation for people who have residual muscle activation but insufficient force exertion to generate a voluntary movement [22], [23]. This aspect is crucial since these closed-loop systems are thought to promote motor recovery by inducing Hebbian plasticity and network reorganization [24], [25].

The progress in NIs research and the increasing availability and affordability of non-invasive devices to record cerebral and neuromuscular activity have been remarkable [26]. Biosignals can now be recorded using user-friendly and wearable devices that provide minimal interference with the user's movements and actions, progressively closing the gap between training and ecological settings [27]. Moreover, wearable devices can potentially enable NIs-based rehabilitation to be conducted in users' homes, rather than in hospital and research laboratory settings. Telerehabilitation - the branch of telemedicine that provides rehabilitation remotely - is rapidly emerging and is fostered by such technical innovations. It is a suitable and effective solution when reaching the clinic is complicated, either due to patient mobility impairments or residence in non-urbanized areas. However, despite being more user-friendly than systems used in clinical settings (e.g., high-density EEG), these devices still present many challenges [28]. Firstly, both patients and caregivers require training, and device preparation can be time-consuming. Secondly, limited battery runtime and memory capacity restrict long-term monitoring, which may also result in data loss.

To this end, artificial intelligence (AI) tools are rapidly spreading and being adopted in different medical contexts [29]. The application of AI is indeed transforming several aspects of rehabilitation, such as clinical and instrumental assessment, training personalization, adaptive assisted rehabilitation and recovery prediction [30]. It showed potential to significantly impact the appropriateness and efficacy of neurorehabilitation [31]. One of the most fruitful applications of AI is biosignal decoding, improving the effectiveness and user experience [18], [26]. However, although the interest in their applications is peaking in the clinical communities, several issues remain to be solved before a complete integration of AI models in NIs for rehabilitation practice [26]. Most of these applications have been limited to the research context, using prototypes or non-commercial devices and using various AI models with heterogeneous performance and generalization properties. For telerehabilitation systems to become effective and scalable, they must adapt automatically to the needs of users, who are very heterogeneous. This process should minimize the time for NIs preparation and calibration. Ideally, AI could provide real-time, personalized feedback not only during task execution but also during the preparation phase, by detecting errors early and guiding the setup process. From the literature, this aspect is not clear and needs to be analyzed.

A. Aim

This work aims to conduct a systematic literature review of AI methods for decoding UL movement using wearable devices. We assessed the state-of-the-art methodologies that integrate one or multiple data sources to better understand their applicability in clinical settings. Additionally, we identified the primary challenges and proposed potential future directions.

II. Literature Search Method

This review was conducted following the preferred reporting items for systematic reviews and meta-analyses (PRISMA) guidelines. As shown in Table I of the supplementary materials, our search query was constructed around five key concepts: *wearable devices, biosignals, AI-driven methods, upper limb, and clinical applications*. At the end of the screening process, 51 papers were in-depth analyzed and discussed. More details on the eligibility criteria, the search query and results can be found in the supplementary materials (see Fig. 1 and Table I).

III. Wearable Devices and Data Acquisition for UL Neurorehabilitation

In this Section, we introduce the technologies used to develop NIs for application in the neurorehabilitation field, discussing their main features, advantages and disadvantages. Fig. 1 shows, for each key concept of our search query, the results found in the analyzed works. Fig. 2 shows the number of included studies for years of publication, type of signals, and type of models used in their analyses. Notably, the number of studies published per year resembles a bimodal distribution.

A. EEG and EMG

EEG and EMG are biosignals used in NIs field to record brain and muscle activity. In particular, brain signals are typically acquired with EEG using conductive gel-based systems. However, the set-up procedure, which involves skin

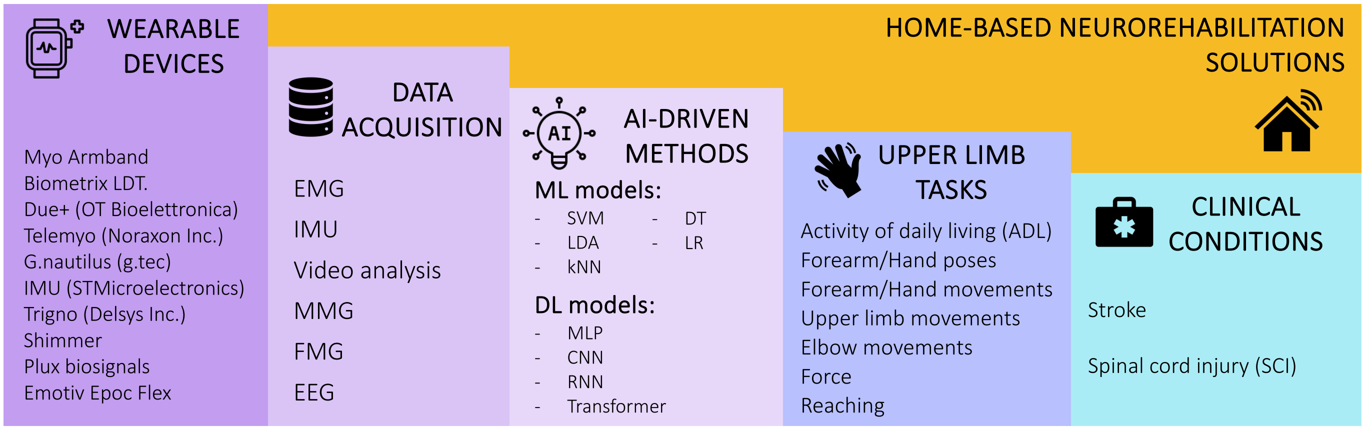


Fig. 1. Graphical outline of the review. Our analysis has been divided into five macro-categories: wearable devices, data acquisition and biosignals, AI-driven methods, upper limb tasks, and clinical conditions. In the first block on the left (wearable devices), we reported the list of devices employed in the analyzed studies. Then, we categorized the types of data, such as EEG, EMG, etc. Then, the most commonly used AI-driven methods for decoding and extracting features of upper limb tasks are listed for both machine learning (ML) and deep learning (DL) models. The upper limb tasks, for example, included activities of daily living. Stroke and spinal cord injury were the only two clinical conditions for which the proposed methods were tested. These solutions were developed with the aim of translating the methodology into rehabilitation practice. Devices, data types, methods, tasks and clinical conditions refer to the records retrieved via our search and included in this review.

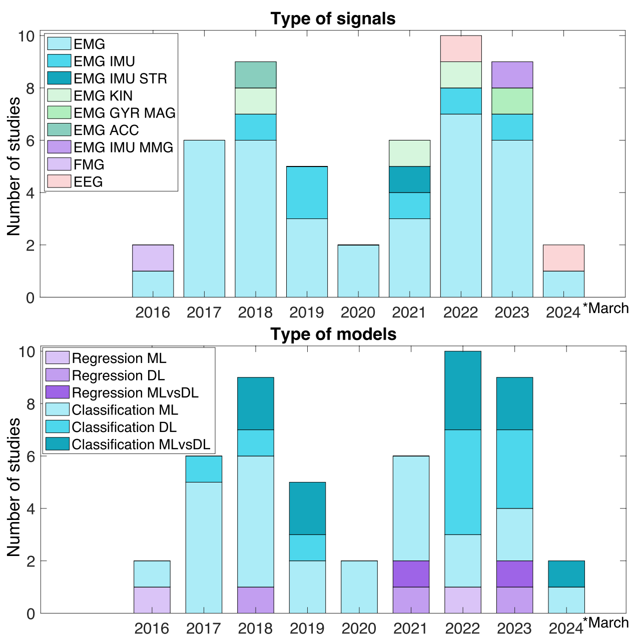


Fig. 2. Upper part: overview of the biosignals and the combination with other data sources used in the reviewed articles over the years. EMG: Electromyography; IMU: inertial measurement unit; KIN: kinematic; FMG: force myography; MMG: mechanomyography; STR: stretch sensors; GYR: gyroscope; MAG: magnetometer; ACC: acceleration. Lower part: classification and regression models used in the revized studies. The classification models are represented in different shades of blue, while the regression ones are depicted in purple. We divided studies that used only machine learning (ML), only deep learning (DL) and studies that proposed a comparison between different methodologies (ML vs. DL).

preparation with dedicated gels/pastes, is time-consuming. To this end, technological innovations, like dry or semi-dry systems, present promising alternatives for EEG acquisition. These systems, which typically include fewer channels without the need for gels, are easy-to-use, portable and wearable [32], [33]. Certainly, the reduced quality of the recorded signals (i.e., the lower signal-to-noise ratio) must be taken into account. Movement artifacts and noise associated with high electrode impedance are particularly difficult to remove and analyze. However, the trade-off between portability and data quality makes it possible to use these systems for the development of NIs.

In the context of UL motor rehabilitation, wearable EEG technologies have been limitedly used due to the difficulties of obtaining stable and reliable data. In our review, two studies proposed NIs based on data coming from dry EEG systems. In Deniz et al., two distinct experiments were conducted for predicting the lifted weight category using custom-made EEG devices [33]. The first experiment used 32 active electrodes, whereas the second experiment employed 8 dry electrodes. In [34], *g.Nautilus* headset with 16 dry channels (g.tec medical engineering GmbH, Schiedlberg, Austria) was used.

On the other hand, muscle activities can be acquired with wearable devices incorporating surface EMG (sEMG). For example, the Myo armband is a bracelet equipped with 8 sEMG channels designed to capture muscle activity data from the forearm. This device, which is connected via Bluetooth, has gained popularity for its gesture control capabilities. As shown in Table II of the supplementary materials, the Myo armband is extremely used. The advantage of these sensors, which integrate the sEMG signal, is the ability to acquire data at any time without the assistance of professional staff [35].

As shown in the upper part of Fig. 2, EMG was the technique most used to develop NIs in the analyzed studies. Moreover, it was often combined with other types of signals, such as inertial measurement unit (IMU) and Leap Motion, as well as stretch sensors, etc. Indeed, extracting robust muscle activity patterns can be challenging, especially for patients with limited motor abilities. Three of the included studies combined the acquisition of myoelectric activity with markerless video analysis [36], [37], [38]. In [37], [38], the Leap Motion system (Leap Motion Inc., San Francisco, CA, USA) was used, a commercially available device based on an infrared motion sensor designed to capture hand kinematics without the need for using markers. In [36], a Microsoft Kinect

(Microsoft, USA) was used to collect whole-body kinematics. Although a camera system is not defined as a wearable device, we have included these studies as they combined kinematic data extracted with biosignals. In any case, these small and portable cameras can be easily positioned in a room and very often transmit data via Bluetooth or Wifi.

1) Publicly Available Datasets: From the papers identified using our search query, only 5 EMG datasets publicly available were found: i) NinaPro database [39], ii) KIN-MUS UJI [40], iii) PhysioNet [41], iv) a 1- and 2-DoF forearm motions datasets [42], [43] and v) a EMG dataset with multiple time-points acquisitions of forearm and hand movements [44]. Additionally, the CapgMyo dataset was used to pre-train a deep learning (DL) model in [45]. The NinaPro DB5 dataset contains EMG signals from 10 non-disabled participants acquired through the MyoArm band device (8 EMG electrodes) [39]. In KIN-MUS UJI, 22 subjects participated in the acquisition of EMG using a Biometrics Ltd. device (7 EMG channels) [40]. PhysioNet contains data from 43 healthy subjects recorded using two devices: EMG Due+ with 8 or 6 channels (OT Bioelettronica) [41]. Finally, in [42], [43], 25 healthy participants were assessed while performing forearm motions with a total of 22-classes (1-DoF 8-classes and 2-DoF 14-classes). They also proposed 30 recording sessions of EMG from 5 healthy subjects while performing hand motions [44]. Details are reported in Table II of the supplementary materials. In addition to the datasets identified in our analysis, other recent datasets have been reported in the literature [46].

B. IMU

IMU, which consists of an accelerometer and gyroscope sensors, and sometimes a magnetometer, is often integrated into wearable devices for human motion tracking. Each sensor extracts the information in its 3-dimensional reference system (i.e. x, y and z-axis), resulting in a total of 9 degrees of freedom. Typically, multiple sensors achieve reliable tracking and capture the movement of multiple body segments. However, placing and calibrating multiple sensors on the body is time-consuming and challenging, especially for individuals with movement impairments. To calibrate sensors with the body and the global reference (world), subjects typically perform predefined poses, such as the T-pose, at the start. Gyroscopes are also affected by the so-called “gyroscopic drift” problem, which refers to the gradual accumulation of errors over time due to imperfections in the gyroscope’s construction or external factors. To solve this effect, methods that fuse multiple sensors can be used, such as the Kalman filter or the complementary filter. The Myo armband also includes an IMU to track movement and orientation. Other types of IMU are reported in Fig. 1, such as Shimmer.

C. Other Signals

Mechanomyography (MMG) is the measurement of the mechanical response of muscles [47]. A simple calibration is required, which reduces preparation time and makes the method robust to skin impedance. MMG has found applications in clinical settings to assess muscle characteristics and rehabilitation protocols, and it has been used for prosthetic control and human-robot interaction. It has also demonstrated the efficacy of combining MMG with EMG to compensate for their own limitations. For example, in [48], a sensor fusion was proposed by integrating MMG with EMG and IMU data for the development of a novel-learning based controller that decodes the onset motion to predict the final position for an assistive robot.

Force myography (FMG) is a non-invasive sensor that measures the position or movement of a limb by detecting changes in the stiffness of the associated musculotendinous complex relative to its default state. In [49], it has been used to estimate the torque exerted by the wrist through a wearable strap composed of 16 force-sensitive resistors. The user wears this strap while performing forearm pronation-supination, wrist flexion-extension and radial-ulnar movements.

IV. Upper-Limb Tasks

The retrieved studies encompassed different tasks ranging from whole UL movement to precise hand grip. Most of them did not include tasks specifically designed for rehabilitation, underlining the distance that still exists between developers and clinical users in this field. When analyzing the tasks in the included works, the main difficulty was the lack of a shared classification system for UL movements. Overall, concepts like gesture, pose, and movement were often interchanged. As shown in Table II of the supplementary materials, we propose a classification of UL tasks, distinguishing between activity of daily living (ADL), hand poses, hand movements, upper limb movements, and elbow movements.

Three studies included in their protocols ADL. In [50], tasks like drinking, using a laptop, brushing teeth, and answering a mobile phone were included. In [40], EMG and IMU data were collected during several activities, also including manipulation and precision tasks such as collecting a clip from the table or passing two buttons through their respective buttonholes using both hands. In [51], ADL were divided into 4 classes, namely non-functional, non-task-related, task-related and high-exertion, each including a group of actions involving similar muscular activity. Other studies focused on tasks that involved the whole UL [52], [53], [54], [55] or elbow flexion/extension [36], [56]. However, most studies analyzed either hand poses or hand movements. This is a relevant result from our review: although there is strong evidence that effective rehabilitation programs should include task-oriented and meaningful exercises, this notion seems not to be considered in most of the included studies. This hampers the translational potential of some included studies, ultimately affecting the development of this research field. With our proposed classification (see Table II of the supplementary materials), which distinguishes between ADL, hand poses and segmental movements, we aim to underline the importance of considering functional tasks in future studies. This aspect highlights the importance of involving engineers and clinicians in designing studies which could have an effective and concrete impact on clinical

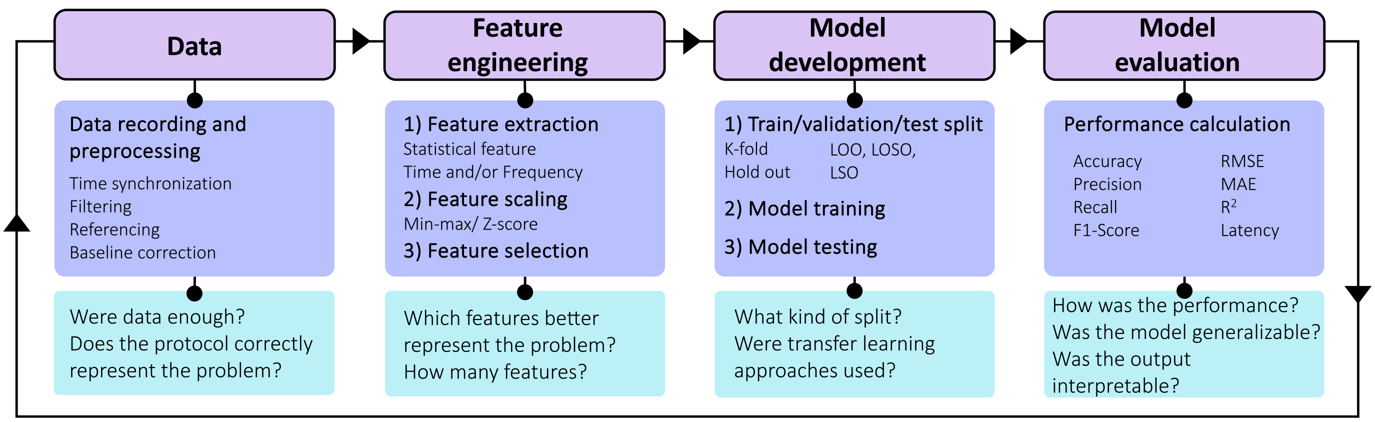


Fig. 3. A typical pipeline for developing an efficient and robust AI-driven neural interface begins with a thorough visual inspection to assess data quality. This is followed by preprocessing steps, such as filtering, baseline correction, and signal rectification, to prepare the data for analysis. In hybrid systems that integrate multiple data modalities, time synchronization is also required. Following, feature engineering allows for the extraction of meaningful and discriminative information, after which feature selection and normalization generally improve model performance. Once an appropriate model architecture is chosen (e.g., for classification or regression), training, validation and testing are performed to obtain the final model's performance (e.g. accuracy, RMSE, etc.). Adopting transfer learning methods is a useful approach to generalize and personalize the output. Below are some examples of questions to consider when developing a robust pipeline.

practice. In our synthesis, we considered a "pose" to be a static hand posture, whereas a "movement" was defined as a dynamic task. Consequently, studies that included the movement required to place the hand in the target posture were considered to address hand movements. In contrast, works where only data recorded during the final static position of the hand were analyzed, fell into the category of hand poses. Tasks included wrist flexion/extension, power grip and fist, finger abduction and pinch grip.

Most of the works that used regression models (4/7) estimated the grip using EMG signal from the forearm [57], [58], [59] or by using a force myogram and force sensing resistor placed around the participant's forearm [49]. Other works analyzed UL movements either estimating the force exerted by the end-effector (arm) on a handle attached to a robotic arm [60], or predicting the final position of the hand in a reaching task [48], or estimating the elbow joint angle during the movement [61]. Interestingly, all these studies integrated more than one data source, mostly combining EMG and IMU. Developing accurate regression models is crucial, as rehabilitative applications would ideally quantify quantitative information on continuous variables, such as joint range of motion, to assess movement quality. However, none of the studies that used regression analysis included participants with CNS dysfunctions, highlighting the difficulties in translating such approaches into real settings.

V. Data Preprocessing and Feature Engineering

Preprocessing and feature engineering are important steps in preparing the raw signals, as shown in Fig. 3. Biosignals like EEG and EMG or kinematic data are often contaminated by different sources of noise like the interference from the head or body movement, eye blinks, and other artifacts. These make the task of decoding useful information challenging. For example, in EMG data, movement artifacts can be reduced by attaching the electrodes to the skin surface using adhesives. Indeed, if the skin is not well-prepared and the electrodes are not well-attached, a noise at 50-60 Hz can be observed. Moreover, the cross-talk phenomenon typically affects EMG signals. It is caused by generating electrical potential from adjacent muscles or deep sources. Due to the nature of the recording technique, this effect tends to be more prevalent in surface EMG, potentially leading to an inaccurate interpretation of muscle activity. To minimize this, proper electrode placement, size, and inter-electrode distance are recommended.

The first step in a signal processing pipeline is to reject trials or channels with poor data quality, particularly for electrode displacements, skin sweat, and subjects' movements, where the data quality is very low. The data are subsequently cleaned using filters. For example, an EEG recorded during an exam might contain electrical interference from near equipment or eye blinks from the user, obscuring important neural activity. Other steps include correcting the temporal drift by subtracting the mean of the timeseries or, for EMG data, rectifying the data. This rectification is applied to extract, for example, the envelope modulation or power analysis. Temporal filters (e.g. bandpass, Notch or Laplace) are used to focus on the frequencies of interest and remove line-noise. For example, EMG data are typically filtered between 10/50 Hz and 500 Hz, while EEG data are filtered between 0.1-70/80 Hz. Moreover, spatial filtering (e.g. common average reference [CAR] and independent component analysis [ICA]) allows reducing EEG noise and artifacts (blinks). Without careful preprocessing, meaningful information could be lost in irrelevant data. Additionally, intra- and inter-subject variability, as well as session variability, are related to changes over time. For EMG, these characteristics are influenced by an individual's subjective properties such as skin, blood flow velocity, temperature, tissue structures, and measuring site.

Once the signals are cleaned, we need to extract the most important features. Feature extraction can occur in various domains, including time, spatial, frequency, time-frequency, as well as statistical and/or connectivity-based measures.

A. Statistical and Time Domain Feature

Statistical features are extracted from raw timeseries or the spectral representation of signals to describe them. Basic features are mean, median, standard deviation (SD), variance (VAR), maximum or minimum value, and interquartile ranges. These features enable the detection of differences between healthy and pathological outputs, as well as the identification of variations over time in biosignals. For example, the mean frequency band of EEG/EMG data can vary with task difficulties.

For more complex analysis, autoregressive (AR) models are useful to describe the input-output relation in the form of a stochastic difference equation. These models help predict the current and future values of a signal based on its past samples. For example, EEG signal can be described in small windows (e.g. 1-2 seconds) as an AR model. In the included studies, AR models were employed using a model's orders ranging from 2^{nd} to 6^{th} depending on the specific application [35], [36], [40], [43], [62], [63], [64].

In the time domain, several features can be extracted from EMG data. For example, the mean absolute value (MAV), calculated as the moving average of the full-wave rectified signal, allows the detection of muscle contraction levels. The root mean square (RMS) is an important feature modeled as an amplitude modulated Gaussian random process, typically used as a smoothing method. The waveform length (WL) is the cumulative length of the waveform over the time window, indicating the complexity of the signal. The zeros crossing (ZC) feature measures how often the signal crosses the zero. Other examples are the integrated EMG (iEMG), the sign slope changes (SSC), the Willson amplitude (WAMP), the average amplitude change (AAC) and others. Please refer to [53], [65] for further details.

Features can be used separately or combined, which can lead to the generation of redundant feature sets. However, popular and robust features for EMG pattern recognition were identified by Hudgins et al., which are MAV, SSC, WL, and ZC [66]. Moreover, not only the different combinations, but also the window size and overlap used to calculate them, are not fixed as expected, but rather depend on the data and applications. For example, a window size of 200-250 ms and an overlap ranging from 125 ms to 50 ms [34], [36], [37], [63], [67] are typical choices. Some studies have demonstrated that even using only the RMS [51], [54], [57] yields strong results in EMG applications, while others combine 4 features such as RMS, MAV, VAR and WL [48], or the combination of ZC, MAV, WL and SSC [41]. Other works extracted a complex set of more than 5 features. To avoid the inclusion of highly correlated features and to limit the computational load, these works used different approaches for feature selection (see Section V-D) [35], [36], [40], [43], [50], [53], [56], [68], [69], [70], [71], [72].

B. Frequency Domain Features

Analyzing EEG or EMG in the frequency domain is crucial to separate the oscillations in different frequency bands. For example, in [34], a filter bank was used to band-pass and filter the EEG data into specific frequency bands, achieving higher performance compared to the common spatial pattern (CSP) method for MI-BCI based applications. The band power of EEG signals was also used as a feature in [33] to measure the spectral density in predefined frequency bands. Frequency domain features are also important for the analysis of EMG data [40], [73]. In [74], the frequency root mean square envelope and signal power peak of EMG data were used for the detection of different gestures. Typically, the mean and/or median frequency (MNF/MDF) are calculated, as proposed in some studies [36], [75], and they are often combined with other time-domain features [40]. Additionally, a modified mean frequency (MMNF) was used in [76], but it did not yield better results compared to other time-domain metrics, such as the RMS. Although frequency domain analysis offers significant advantages, the Fourier transform (FT) does not provide any information about the approximate temporal location of the event of interest. However, there are other methods for solving this problem, which are presented below.

C. Time-Frequency Domain Features

Time-frequency analysis is a powerful technique, particularly when dealing with dynamic, non-stationary, and fast-transient signals, such as EEG and EMG. The statistical properties of those signals typically change over time. Unlike traditional frequency analysis, which treats signals as static, time-frequency analysis describes how the spectral content of a signal evolves over time. Decomposing the EEG signals into many well-localized features in the time-frequency domain has demonstrated the great potential of enhancing the classification error rate [77] compared to traditional methods, such as FT, AR models, etc.

The Short-Time-Fourier-Transform (STFT) is a widely used technique that provides information about when and at what frequencies a signal event occurs. However, the STFT analysis suffers from the pitfall of having a fixed combined resolution of time and frequency, limiting its precision. The wavelet transform overcame this issue. It decomposes a time-varying signal using a shifted and scaled version of a wavelet [78]. This flexibility makes wavelets ideal for analyzing rapidly changing signals like muscle activity during hand movements. Moreover, the wavelet transform is preferred to the STFT since the latter generates many features that could be difficult to integrate with other feature types, and it is relatively computationally expensive [39]. For this reason, the continuous wavelet transform (CWT) or the discrete wavelet transform (DWT) are typically used. However, the choice of which class of wavelet to use is dependent on the application. The CWT provides a continuous signal representation, offering precise time-frequency information. In contrast, the DWT aims to completely represent any discrete signals as a series of wavelet coefficients, yielding a compact and non-redundant representation. Moreover, the DWT offers the advantage of lower computational cost than the CWT.

In [39], for the first time, the CWT was used for the classification of hand gestures on EMG signal [39]. The absolute value of the CWT of a 1D-EMG signal (scalogram) was fed into a DL architecture. The performance of the CWT were compared with that of the other two models, which employed raw data and the STFT spectrogram. Differently, the work by Yang et al. proposed a denoising pipeline of EMG signals based on the CWT and the filtered data were used to extract the traditional time domain features (e.g. MAV, WL, etc.) [67]. In [33], the DWT was employed as a feature extraction to analyze EEG data using Daubechies 4 (db4) as mother wavelets and subsequently, 13 statistical features were calculated for the bands of interest. The DWT was also used as a feature extraction technique from EMG timeseries. Indeed, in [79], different gestures were classified using the extracted coefficients, while in [76], the energy of the DWT was used, outperforming other features (i.e. MNNF, WL, etc.) in terms of accuracy.

D. Feature Selection, Normalization and Scaling

Features may be extracted from the raw data through processing techniques that introduce redundancy, complexity, and poor interpretability. To reduce these effects, feature selection can help create a robust and effective pipeline, avoiding noisy features in the model. It reduces complexity, both in terms of interpretability and computational load, which is necessary, especially for real-time use. A small-dimensional feature set mitigates the overfitting (i.e., lack of generalization of the model), which is an important issue to consider, particularly when the dataset is smaller. Moreover, the presence of correlated variables in the model (collinearity problem) makes the predictions unstable. So, collinearity reduction is recommended, particularly for models that assume working with uncorrelated variables (such as logistic regression (LR)).

Despite its importance, the feature selection step remains challenging, and there is no single strategy to be followed. In clinical contexts, for instance, selecting appropriate features may rely on prior neurophysiological knowledge, such as the choice of electrode placement in EEG or EMG data acquisition [36], [76]. Multiple feature selection or projection approaches have been proposed to reduce the dimensionality, which may involve ranking or transforming the initial feature set. These methods are typically based on statistical measures, such as mutual information, analysis of variance, and χ^2 [33], [61], or projection methods such as principal component analysis (PCA) [64], [70], [71], [80], [81].

In [35], the univariate feature selection approach was employed to find the optimal feature set. They trained the model using each of the 7 features separately, and the 4 features with the highest performance were selected. After that, the remaining features were added at each iteration until they obtained the optimal set of 6 features. Notably, the univariate feature selection approach is effective only when the variables are uncorrelated. In [82], a forward selection approach was proposed. The model was run multiple times, adding a feature at each iteration and retaining it in the next iteration only if its contribution to the model is statistically significant. In [39], linear discriminant analysis (LDA) was used for feature projection, a method known for its low computational complexity and the absence of hyperparameter requirements. LDA aims at identifying a linear combination of features that best separates two or more classes. In [53], a selection of a subject-independent set of EMG features was proposed by means of information theory tools to minimize the redundancy and maximize synergistic effects [53]. They considered 2 features to be synergic if the combined information they provide about the events exceeds the information conveyed by each feature individually.

Feature selection is also important for better understanding data distribution in high-dimensional problems. For example, He et al. used the *t*-distributed stochastic neighbor embedding (t-SNE) and Davies-Bouldin index (DBI) to visualize the feature distribution and quantitatively measure the separation of the 17-classes, respectively [41]. Another work proposed a weighted cross validation feature selection (W-CVFS) method [55] for feature fusion selection. They obtained better results selecting the top 10 than using all 18 features as input. W-CVFS also outperformed other feature selection methods, namely the minimum-redundancy-maximum-relevance and infinite-latent-feature-selection methods. Along with these approaches, some works employed innovative methods for feature extraction, tailored to the specific needs of the data they dealt with. For example, in [81], the extracted features were processed using a Fibonacci-resembling spiral net ranking system [74], [83]. Typically, irrespective of the feature extraction approach, features are standardized or normalized (e.g., min-max scaling [40] or *z*-score normalization [63]), benefiting AI models which are sensitive to feature ranges, like support vector machine (SVM) or K-nearest neighbors (k-NN).

VI. AI-Driven Approaches in Motor Neurorehabilitation

Classification and regression are commonly used AI tasks. The former learns to assign a class to a new instance, and the latter to predict a target numeric value given a set of features. In neurorehabilitation, classification models can distinguish correct and incorrect movements or poses, such as an open hand or a closed hand, while regression models are suitable for assessing, for example, the quality of movement. Indeed, by predicting continuous values, it is possible to track joint angles and movement velocity during the task [57].

From a rehabilitation perspective, the main difference between these two approaches lies in the feedback provided to the user. Classification models can provide users with categorical feedback on their performance (e.g., correct/incorrect), mostly triggering error-based sensorimotor learning. This can be, for example, useful to interact with objects in virtual reality environments, where users perform predefined tasks. However, this method may not always be the most suitable from a clinical perspective, as it simplifies complex motor behaviors into discrete categories. On the other hand, the feedback provided by the regression models (e.g., movement quality) is more suitable to support reinforcement learning [84]. This consideration suggests that, in an ideal rehabilitation environment, clinicians would be able to select the type of feedback (and consequently, the type of ML model) according to the user's needs.

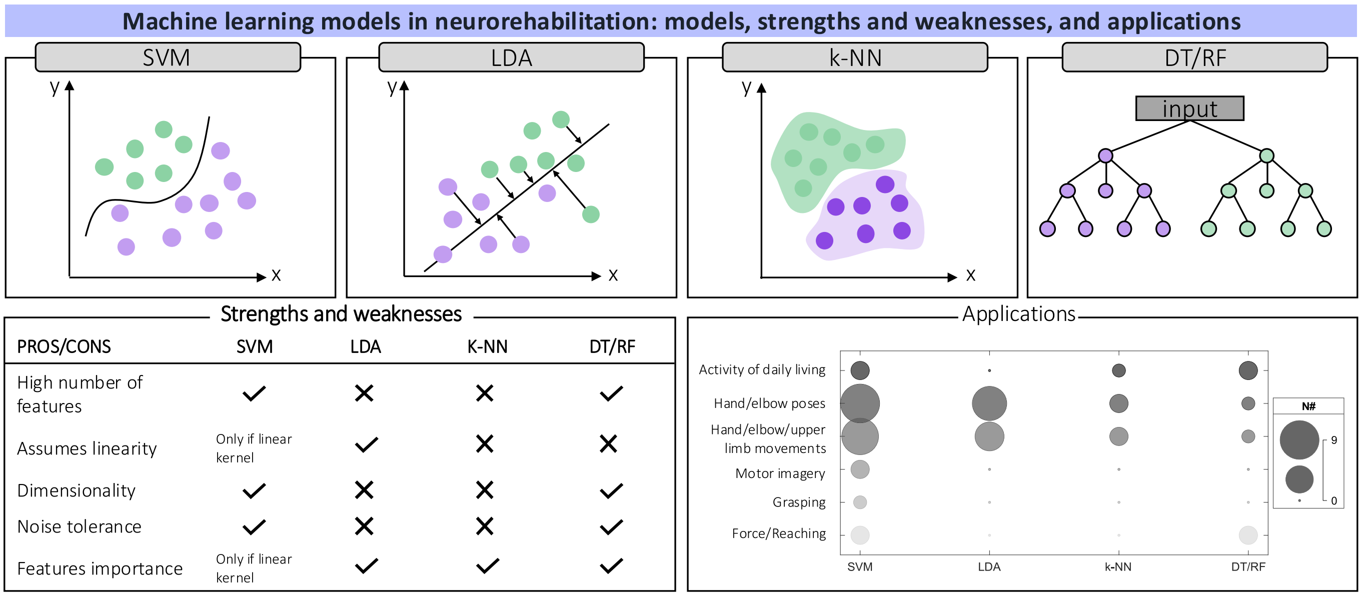
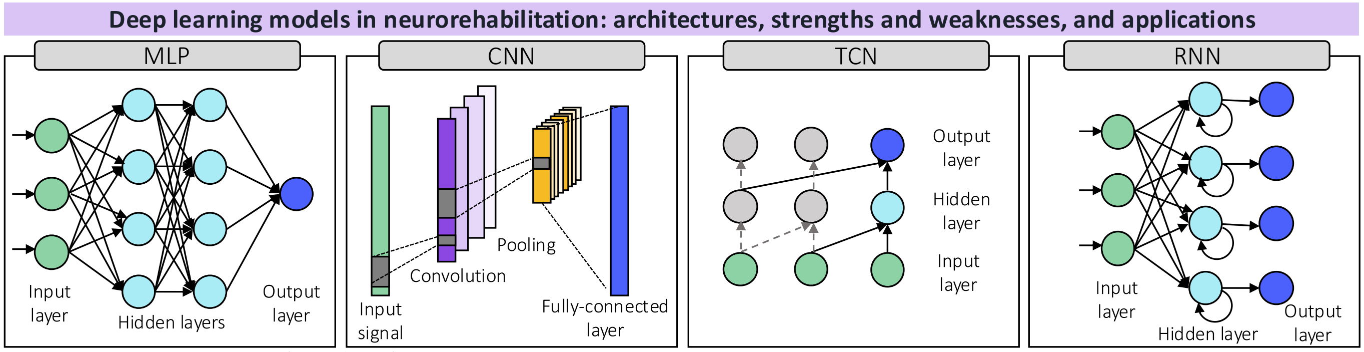


Fig. 4. Top: schematic representation of the most used machine learning models that are support vector machine (SVM), linear discriminant analysis (LDA), K-nearest neighbors (k-NN), decision tree (DT) and random forest (RF). Bottom left: we identified five key aspects to assess when developing AI-based models and neural interfaces. These aspects are: 1) "high number of features" means that the model is able to deal with high-dimensional spaces; 2) "assumes linearity" requires linearly separated data; 3) "dimensionality" refers to the ability to solve classification problems with many features in comparison to the number of samples to be classified; 4) "noise tolerance" means that it works well even in conditions where the input data is very noisy; 5) "features importance" refers to the ability of providing a score for each input feature without applying post-hoc analysis (i.e. explainable AI methods). Each model is marked with a check "v" or a cross "x" to indicate whether the respective feature is supported. Bottom right: we report the number of applications in the neurorehabilitation field. They were divided by type of task: 1) activity of daily living; 2) hand poses; 3) hand/elbow/upper limb movements; 4) motor imagery; 5) grasping and 6) force estimation and reaching. The size of the bubbles refers to the number of studies, while the gray scale refers to the type of task in each row.

Among the reviewed studies, classifiers have been adopted in 86% (44/51) of the included works (see Fig. 2). Different models were proposed to classify common ADLs like drinking, brushing and washing the face [40], [52] or the other hand movements, wrist, and thumb static or dynamic gestures like agree, open hand, pointer, flex hand [35], [37], [41], [63], [68], [75], [85], [86]. Other studies classified the changes in task weight or force while performing flexion-extension motions or repetitive grasping and lifting trials [33], [36]. In [56], 5 elbow angles during static contraction were classified. Moreover, a system focused on classifying MI tasks [34]. While these studies did not include specific rehabilitation interventions for the movements and gestures examined, the scenarios they propose are important for developing control systems. This allows users to interact with real-time biofeedback, which can be translated into rehabilitation practice.

Recent trends highlight how regression models can capture quantitative aspects of movement, such as velocity patterns and range of motion, from bioelectric signals. This information can guide exercise progression and personalized modulation according to the specific subject's needs. Notably, it can be integrated into training programs, providing patients with real-time feedback. In addition, extensive literature demonstrated that motion analyses can be used as outcome measures to assess participants' improvement after a rehabilitation intervention quantitatively. The literature search conducted for this review found a total of 7 works that used regression models [48], [49], [57], [58], [59], [60], [61]. Most of them used bioelectric signals (EMG or FMG) to predict hand force exertion [49], [57], [58], [59], [60]. Other works have aimed to predict the kinematic features of movement, such as reaching direction [48] and elbow flexion angle [61], by combining EMG and IMU data. In [61], 10 regression models were compared for the prediction of elbow flexion angle trajectory by fusing multimodal information (i.e. EMG, IMU and stretch sensors). By selecting different groups of features and algorithms, kinematics data were sufficient for detecting a quick motion intention in a small observation window; on the other hand, physiological and kinematic features were necessary to obtain a better prediction in a longer window. The results showed that using only EMG data cannot achieve the same performance as using only IMU.

Subsequent paragraphs describe the most used machine learning (ML) (paragraph VI-A) and deep learning (DL) (paragraph VI-B) models in the included studies, comparing their characteristics and performance for specific applications. In Table II of the supplementary materials, the choices of each included work are detailed. A schematic representation of some of these models is shown in Figs. 4 and 5. We identified different key strengths and weaknesses for each model. These include: number of features, linearity assumption, curse of dimensionality, noise tolerance, sequential data, and feature importance. In these Figs., we also depicted the number of papers that implemented specific models as circular markers for each task. The training and evaluating strategies are discussed in paragraph VI-C. Moreover, we introduce the use of explainable artificial intelligence (XAI) in paragraph VI-D, data augmentation and generative AI in paragraph VI-E and finally real-time applications in paragraph VI-F.



Strengths and weaknesses

PROS/CONS	MLP	CNN	RNN	TCN	Transformer
High number of features	✗	✓	✓	✓	✓
Sequential data	✗	✗	✓	✓	✓
Dimensionality	✓	✗	✓	Long data	Long data
Noise tolerance	✗	✓	✗	✓	✗
XAI	✓	✓	✓	✓	✓

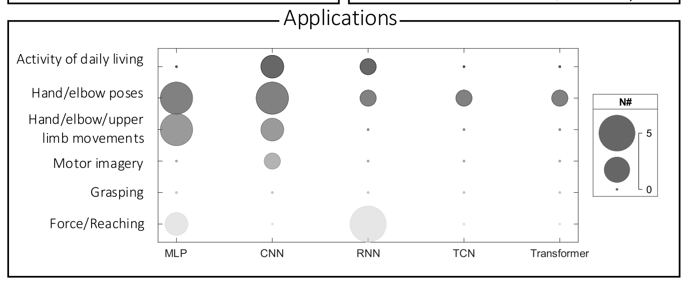


Fig. 5. Top: a representation of the most used deep learning models that are multiple layer perceptron (MLP), convolution neural network (CNN), temporal convolution network (TCN), and recurrent neural network (RNN). Bottom left: as we did for ML models in the previous figure, we identified five key strengths and weaknesses. They are: 1) "high number of features"; 2) "sequential data" means that the model is suitable to analyze sequential data such as time series; 3) "dimensionality"; 4) "noise tolerance"; 5) "XAI" refers to the possibility of applying post-hoc explainable AI analysis. Bottom right: we report the number of applications in the neurorehabilitation field. They were divided by type of task: 1) activity of daily living; 2) hand poses; 3) hand/elbow/upper limb movements; 4) motor imagery; 5) grasping and 6) force estimation and reaching. The size of the bubbles refers to the number of studies, while the gray scale refers to the type of task in each row.

A. ML Models

1) Support Vector Machine (SVM): as shown in Table II of the supplementary materials, 54.5% of the selected articles that used classification models (24/44) trained an SVM model, and 10 of them did not compare the performance with other models [34], [38], [62], [67], [71], [74], [76], [79], [85], [87]. In most cases, SVM has been adopted to classify hand poses [62], [79], [85], [87], hand movements [67], [68], [71], [76] or grasps [38] using EMG data from individual subjects, demonstrating good performance. In these studies, the number of classes ranged from 3 to 9, while the number of subjects ranged from 1 to 6, except for the work of Amirabdollahian et al., which recorded data from 25 subjects [87]. Some of these studies suggested that, overall, to achieve a good trade-off between model complexity and performance, recognizing 5 hand poses or movements using an SVM yielded good results [71], [76], [79], [85]. It is important to note that the number and position of EMG channels, as well as the information extracted from them, can significantly impact classification accuracy. Kartsch et al. [76] compared different time and frequency domain feature extraction techniques, and the combination of 4 and 7 electrodes. They showed that reducing the number of EMG signals and, consequently, the computational complexity, the recognition accuracy decreased from 96.78% to 94.02%. The SVM was also used in [50] to recognize 5 ADL, outperforming the RF model. When compared to other classical approaches to classify UL motions, SVM reached better performance as shown in [54]. They obtained an offline accuracy of 96% and a real-time accuracy of 90%, compared to artificial neural networks (ANN) (96% offline and 86% real-time) and LR (93% offline and 85% real-time). Good performance were also achieved in a subject-to-subject transfer framework combining the SVM in a semi-supervised style transfer mapping (SS-STM) approach, which was proposed in [42] and [43]. A similar framework was also presented in [71]. They combined techniques such as PCA, clustering, and SVM to train a model, which was subsequently evaluated on new users. The model achieved an average accuracy of 81% across 5 healthy subjects for classifying 5 hand movements, whereas the average accuracy was 46% using 9 movements. The deterioration of performance may be related to hand movements that present high biomechanical similarity, so their future work will focus on using other sensor modalities (i.e. inertial sensors) to improve the 9-movements classification results.

The SVM classifier was used to detect MI tasks from EEG activity [34]. In this study, MI involved hand opening and closing in which each participant performed 8 sessions. By employing an 8-fold cross-validation (CV) approach, which involves considering each session as a validation set at each iteration, the classification accuracy ranged from approximately 60% to 80% for 8 users and exceeded 80% for 3 participants. To validate their approach, in addition to their dataset acquired through wearable EEG sensors, they used two public datasets. The first was the BCI competition IV dataset IIa [88], and the second was the BCI competition III dataset IVa.

Overall, SVM demonstrates high noise tolerance and robustness to high-dimensional feature spaces, working well with datasets comprising few subjects, which is often the case in a rehabilitation context [34], [79]. However, other works pointed out that SVM was unable to achieve adequate results in certain conditions [33], [35], [39], [40], [80], [82]. Regarding the regression model, Yang et al. compared the performance of a support vector regressor (SVR) model against other architectures to predict the desired end-point position for an assistive robot [48].

2) Linear Discriminant Analysis (LDA): is used for both supervised and unsupervised problems. It allows for obtaining an interpretation of the input feature, assigning an importance score. As the SVM model, LDA model is a common choice for hand poses/movements recognition in the field of UL rehabilitation. It has been adopted in 13 selected studies. They typically involved a broader range of hand poses/hand movements, ranging between 3 and 18. Five studies chose only LDA as classifier, without making any comparison with other approaches [36], [37], [44], [69], [70].

In [70], LDA classified 9 hand poses performed by 8 non-disabled participants. Based only on EMG, LDA was able to classify them with an average accuracy across subjects of 94.18±3.63% using conventional EMG sensors and an average accuracy 90.14 ± 8.05% using the Myo armband. On the other hand, in [37], data from post-stroke patients were recorded while they performed 8 hand poses. Given the need to develop practical and effective systems that can be used for post-stroke rehabilitation, the authors investigated the fusion of EMG and kinematics data. The proposed pipeline, based on LDA classifier, was tested on 10 healthy subjects and 5 post-stroke patients, achieving an average accuracy of 96.90 ± 1.81% and 96.43 ± 3.83%, respectively.

Among the included works, the first to propose an LDA classifier for detecting changes in interaction forces during dynamic and unconstrained motions using only EMG, arm position, and movement speed was Stanbury et al. [36]. LDA was selected for its robustness and simplicity, as it does not require additional parameters. Three force levels were classified (0, −22N, + 22N), reaching a maximal accuracy of 79.17%. In [35], LDA outperformed both a multi-layers perceptron (MLP) architecture and an SVM, achieving the remarkable accuracy of 99.85 ± 0.03% and short training time (0.37 ± 0.01 s) to recognize 9 types of motions. Similarly, in [69], an LDA was used to detect user intent for assistive device control by measuring the EMG activity. The authors tested the proposed pipeline on 3 spinal cord injury (SCI) patients while performing 7 hand motions. Overall, the results suggest that LDA seems to be a robust and accurate approach for the classification of hand poses and gestures based on forearm EMG data.

3) Tree-Based Models: decision tree (DT) is a versatile and non-parametric method. Due to its straightforward interpretability and intuitive framework, DTs are frequently referred to as "white box" models. On the contrary, random forest (RF) is an ensemble and supervised learning method widely used for regression and classification. Based on the bagging method, RF creates multiple decision trees by bootstrapping the original dataset. Since RF's outcome is obtained by averaging several DT, it is considered a "black box" model, making its interpretability difficult. Nevertheless, RF has the potential to measure the relative importance of each feature in the prediction, known as RF feature importance (RFI). XGBoost is another ensemble method that uses a gradient boosting framework and parallel processing. RF was employed in 4 selected studies in comparison with other classifiers [39], [40], [50], [82]. Generally, in the included works, RF did not yield good results compared to other methods. In [40], the worst performance was obtained using RF and XGBoost (testing accuracy of 63%). Also in [50], the RF model underperformed compared to the SVM model in classifying 5 different ADLs using EMG and inertial sensor data. In a different setting, with forearm EMG data used to classify fingers' movements, the ensemble of DT did perform better in terms of accuracy than DT alone, but showed remarkable accuracy limits when compared to k-NN or SVM [68].

4) K-Nearest Neighbors (k-NN): k-NN was used in 5 studies [39], [51], [56], [68], [82]. In [51], a k-NN was trained to evaluate many combinations of activation and motion data in classifying 17 functional upper-extremity tasks divided into 4-classes [51]. The maximum accuracy was 89.2% using the EMG activation and the mean values of the acceleration and angular velocity. The k-NN was capable of combining sensor modalities to predict gesture task categories with high accuracy, selecting k= 3 based on an initial assessment of the model's performance. Moreover, k-NN requires no assumptions about the characteristics of the underlying model. In [56], EMG patterns during 5 elbow angle movements using the k-NN model were classified. They achieved good results using subject-specific EMG features. However, in these studies, data were recorded during static or isometric contractions, which casts doubt on whether k-NN is an appropriate choice for decoding dynamic movements typical of rehabilitation exercises. Last, in [82], the k-NN outperformed other models (i.e. SVM and RF).

5) Logistic Regression (LR): is a classification model designed to estimate the probability that an instance belongs to a specific class. In the analyzed studies, LR did not yield satisfactory results [54], [80]. This could be because LR works better with dichotomous data, while the studies involved multidimensional datasets with many features obtained from EMG channels [54]. While models like LDA and SVM showed high and comparable accuracy in classifying movements and gestures, their choice should align with the specific clinical goals.

B. DL Models

1) Multi-Layers Perceptron (MLP): as shown in Table II of the supplementary materials, MLP is a common classification model, especially for recognizing complex patterns from more than 5 classes [64], [86], [89]. A simple MLP was designed in [75] to recognize 5 hand movements by extracting different features from 8 EMG channels. The model consisted of an input layer, a hidden layer of 20 neurons and an output layer. Similar frameworks have been proposed in other studies, both classifying hand poses [72] and hand movements [86]. In [64], they compared different user-independent models to classify seven hand and finger poses by combining EMG and IMU data. They employed a simple MLP architecture, consisting of an input layer, two hidden layers, and an output layer. This model achieved good recognition accuracy compared to the proposed bilinear model based on a user-independent framework. The MLP also generalized well to unseen data, effectively capturing intrinsic variability between different subjects. Despite this, several limitations were identified compared to the Adaptive Least-Squares SVM model, which has the potential to outperform the others. Firstly, the structure of the MLP model, which is usually determined empirically, can be improved to enhance performance, requiring more tests and additional data. Secondly, it was sensitive to the Myo Armband sensor displacement, which is fixed but may vary among users and/or sessions. Moreover, MLP networks are affected by feature noise, resulting from users' inability to consistently replicate the same level of muscle contraction when performing the trained gestures. A similar framework was proposed in [63].

In [54], an MLP model was compared with an SVM and an LR model. They achieved an accuracy ranging from 71% to 98% for UL motion pattern recognition of 7 offline and 4 real-time movements. Interestingly, they tested a real-time pipeline for the first time, achieving higher accuracy using the SVM. The under-performance of the MLP in this task could be related to the size of their dataset, suggesting that traditional ML models could be more suitable for small-size dataset applications. Gomez-Correa et al. designed a new wearable EMG system consisting of two WyoFlex [73]. In this case, to validate their device, several tests were performed by using an ANN composed of a two-layer feedforward network with hidden sigmoid neurons and softmax output neurons [73]. Similarly, 3 tasks (i.e. rest, thumbs up, and wave out) were classified using an ANN [90].

2) Convolutional Neural Network (CNN): is commonly used in the field of myoelectric recognition due to its ability to effectively exploit spatial information in input data through convolutional operations. It is frequently employed in studies involving a large number of classes. In the revised studies, this number ranged from 7 to 26, significantly higher than what is typically handled by traditional ML methods. However, the number of classes is not the only factor driving the use of a CNN. The overall complexity of the pipeline can also increase depending on the number of input channels being analyzed. For example, in [33], a CNN-based architecture was used to distinguish 3 MI tasks from 32 EEG channels.

Salinas et al. compared ML and DL models in the classification of 26 ADLs using EMG channels and joint-angles [40]. The CNN outperformed both ML models and a DL architecture (i.e. gated recurrent unit (GRU) model). Comparing four CNN configurations, and using a hard voting approach, an accuracy of 86% was achieved. In [41], the CNN was compared with the LDA classifier and 3 DL architectures for hand poses recognition using EMG data (i.e. CNN, GRU and transformer). This work also proposed a temporal convolutional network (TCN), a variant of CNN, to capture temporal patterns from sequential data using a relatively simple structure. The DL models were good at classifying 17 tasks; but, considering model complexity, computational cost and performance, the TCN outperformed the other 3 models. Thus, they concluded that the TCN may be the best choice for hand pose recognition from the wrist. They also compared the ability of the models to recognize hand poses from the forearm or wrist activity. The results were similar when using the 4 DL models with wrist data, whereas those with forearm data were lower. However, TCN still yielded the lowest error rate in recognizing hand poses from the forearm.

CNN was also used in other studies to predict high numbers of hand movements from EMG activity. For example, Olsen et al. trained a CNN composed of five 2D convolutional layers, 3 fully connected layers, and an output layer to classify 17-classes [91]. They performed continuous recognition, in which each second of acquisition (30 samples since the features were calculated at 30 Hz) was labeled as the same class. The CNN was tested in 3 post-stroke patients in 4 conditions based on recording location: 1) non-paretic forearm EMG, 2) non-paretic wrist EMG, 3) paretic forearm EMG, and 4) paretic wrist EMG. They showed that the paretic wrist provided higher classification results (>80%) than the paretic forearm. Despite the ability to recognize many motions, the proposed model, composed of a sequence of layers, could be too complex to be implemented in the real world. In contrast, another study proposed a lightweight 1D CNN model to recognize 21 types of poses performed by 10 healthy subjects [45]. The personalized model achieved an average accuracy of 82.93%. This model can potentially be embedded into a wearable, and real-time system for hand pose recognition.

CNN obtained good performance also in [92]. Different analyses were performed to validate its reliability in intra/inter-session and between-day setups. Interestingly, no statistical differences were found comparing the CNN with a stacked sparse autoencoder with features (SSAE-f) extracted from data. Although the feature extraction step could enhance the discriminative power between classes of the input data, it requires several choices and processing time. On the other hand, the CNN outperformed the SSAE with raw data as input. Another CNN architecture was proposed and compared with an MLP model for classifying 9 hand poses, including 9 subjects who performed a total of 243 trials [72].

3) Recurrent Neural Network (RNN): the long short-term memory (LSTM) and the gated recurrent unit (GRU) are two common models used in several applications, including neurorehabilitation [93]. The former, as a solution to the vanishing gradient problem [94], is composed of different "cells" in the hidden layers. In particular, the input gate, the output gate, and the forget gate control the information flow required to predict the output. The latter is similar to the LSTM, which uses a "hidden state" to regulate the information flow. But, instead of 3 gates, it has a reset gate and an update gate. Salinas et al. [40] trained a GRU, achieving higher results compared to ML models, but lower than a CNN. One possible reason for these could be the nature of RNN models, which are powerful in extracting temporal information, but do not effectively analyze the spatial correlations in the data. It might be interesting to adopt hybrid strategies, e.g. combining CNN and RNN, to synthesize both temporal and spatial information. In the literature, this methodology has already been developed. A CNN-GRU model demonstrated the potential to decode multivariate data, outperforming traditional ML models [93]. Moreover, RNN provided good results in a regression problem in [48]. They aimed to predict motion-intention during reaching and placing tasks using EMG and IMU sensors, showing that the RNN-based model outperformed traditional regression models such as SVR, DT, and RF. An innovative approach, called NeuroPosed, based on an encoder-decoder network, a ResNet and an RNN was proposed in [58] and [59]. This aimed to estimate plausible finger pose sequences with spatial constraints across fingers, as well as temporally smooth variations over time. In particular, the LSTM was implemented to obtain real-time performance without redundant computation.

4) Transformer: is a DL model that analyzes multidimensional and sequential data all at once, eliminating the recurrent part [95], [96]. They are built around the attention mechanism, particularly self-attention, which assigns weights to different inputs relative to each other in accordance with their importance. This process emphasizes specific parts of the input sequence, thereby gaining a better understanding of them and handling long-range dependencies within sequences, which are not captured by RNN models. Transformers reduce computational costs by exploiting parallelization, scalability and flexibility. This is important not only for real-time applications, but also for solving the problem of intra-/inter-subject variability. Indeed, it has been demonstrated that transformers can capture latent temporal patterns. In [41], a transformer was adopted to classify 17 hand poses from EMG activity.

Summarizing, many models can be adopted, each with its advantages and disadvantages. However, choosing the model according to the application and complexity of the problem to be solved is good practice. As shown in Fig. 5, if we want to analyze a dynamic hand/elbow or UL movement, models like MLP or CNN are preferred, while for regression problems like force prediction or reaching tasks, RNNs are the best choice, followed by MLP. In the case of ADL classification, which typically includes a higher number of classes, CNNs or RNNs are useful.

C. Training and Evaluation Strategies

To accurately assess the performance of a model, it is essential to evaluate its robustness and generalization ability. One of the main challenges in using AI in neurorehabilitation is the heterogeneity of subjects' functional impairments. Moreover, external stimuli that may foster neuroplasticity in different ways influence the neuromuscular reorganization that sustains recovery. This leads to significant inter-subject differences in kinematic patterns and muscular and neural activation to perform the same gesture, often hampering the model's generalization ability.

To overcome this, some studies have developed approaches that leverage large datasets. In [75], both single and multi-user models' performance were tested adopting the 75-25% training/test split strategy, firstly for each user and then on the combination of 5 participants. The single-user performance for classifying 5 hand movements was higher than that of the multi-user model. They achieved a maximum accuracy of 99.2% for single and an accuracy of 88.4% for multi-users. As discussed above, using data from each participant separately poses some challenges, as it is time-consuming and user-unfriendly to create a single-subject dataset of an adequate size to calibrate and train an AI model. In [50], data from 21 healthy subjects were acquired, divided into 70% training and 30% testing. They trained an SVM and an RF model using a 5-fold CV on the training set. Achieving an accuracy of 78% using the SVM, the main limitation was the sample size and the unbalance among the different classes [50].

Alternative methods have been proposed to evaluate user-independent performance, namely leave-one-subject-out (LOSO) [82] or leave-subjects-out (LSO) CV [90]. In these approaches, data from one or multiple independent subjects are used as unseen testing data in each iteration, while the remaining subjects' data form the training set [63], [74]. Overall, a user-independent scenario can overcome the practical challenges associated with neural interfaces (NIs). In this way, calibration time would be reduced. Some issues to consider include the fact that the sensors' placement, although generally placed according to guidelines, e.g., 10-10 EEG standards or Seniam for EMG, are not exempt from experimental/anatomical variability when repeated over multiple sessions. Therefore, the same task-related physiological processes can vary between subjects and within the same subject across different sessions. Second, data from patients with limited motor abilities can be challenging to decode.

Another requirement from the clinical rehabilitation perspective is the possibility of using a particular NIs across sessions throughout the rehabilitation intervention. This means that an ideal system should perform consistently across several sessions and setups. That could be a non-trivial problem, as biosignals are extremely variable, and setup details, such as electrode placement or skin impedance, can vary significantly between sessions. These factors can cause variations in feature distribution, known as "covariate shift" [59], [97], affecting the final performance. Solving these problems requires complex models capable of analyzing time-varying and non-linear information, and larger sample data with more trials [33].

Moreover, to be translated into clinical practice, every approach must deal with the constraints of the rehabilitation context. On the one hand, the inter-session variability can be limited by designing simple setups that are as user-dependent as possible. On the other hand, specific training approaches have been used to mitigate this effect. For example, good results can be obtained if full calibration is performed in each session, thereby splitting the data into training and testing sets. However, this approach reveals significant flaws when applied in a real setting. A typical rehabilitation session lasts around 40-50 minutes, meaning that, in this case, approximately 25-35 minutes of training would be used just to record data for the calibration. This not only shrinks the proper rehabilitation time to 30% of the total session, but the repetitive process of calibration may potentially affect participants' engagement and adherence to the intervention itself. Similar strategies have been presented in some works, which split the data of each participant into 80%-20% training/testing [54], [68], [70], [72], [79]. In particular, in [70], a 5-fold CV was employed, while, in [91], 80% of the data from each participant was used as training, and the remaining 10%/10% was used for validation/testing. In [33], data were divided into 70%-30% training/testing for ML models (i.e. SVM and XGBoost) and 70%-20%-10% training/validation/testing to assess DL models (i.e. ConvNet and EEGNet) [33]. On the other hand, some studies performed only *k*-fold CV and presented the average classification accuracy over a number of validation sets ranging from 10 [37], [56], [85], 8 [34] or 5 [86]. Notably, they ensured that each gesture was presented in both the training and validation sets. In [92], different strategies to evaluate data acquisition within a single day and across multiple days were proposed by dividing the data into folds based on data availability. The within-day evaluation included within-session and between-sessions analyses. For within-session, a 10-fold CV was adopted, since there were 10 repetitions for each session, and a 2-fold CV was used for between-sessions, which included 2 sessions on the same day. Between-days analysis included 2-fold CV (pair of days) and 15-fold CV (15 days of acquisition). To overcome these problems and obtain an optimal trade-off between model performance and calibration time, Colli-Alfaro et al. proposed a sensor fusion approach (EMG and IMU), thus using different sources of data [63], [64]. Moreover, an adaptive and personalized system could be potentially obtained by re-training the model during new acquisition sessions. Thus, all these factors play a pivotal role in the development of NIs systems, enhancing the user's empowerment and the adherence to treatment.

Another important concept is transfer learning (TL). It consists of transferring knowledge gained from a pre-training model using a *source domain*, typically a large dataset, and using this as a starting point to improve the performance of a new dataset, called *target domain*, which is often smaller. In particular, the model weights are initialized to be the same as the pre-trained model and are updated in the fine-tuning stage [26]. Specific TL models were proposed in some screened works, which demonstrated the efficacy in obtaining a generalizable model [39], [43], [58], [59]. TL performs well for high-dimensional data sizes. Côté-Allard et al. acquired data from 46 subjects [39], and 12 users in [58] and [59]. In [43] two TL approaches were used: covariate shift adaptation (CSA) and semi-supervised style transfer mapping (SS-STM).

D. Explainable AI

A significant drawback of some AI approaches is their lack of transparency, the so-called "black-box problem", where the internal mechanisms of the model are not easily interpretable by humans. This issue often makes it difficult to understand which features these models relied on to make a decision or prediction, hampering their application in the medical field. Therefore, eXplainable artificial intelligence (XAI) has become an important topic recently. XAI aims to explain the *"the why and how"* of AI outcomes by trying to create a kind of bridge between human reasoning and machine intelligence [98]. By doing so, the outputs of complex models are more *interpretable* and *understandable* [99].

In the clinical field, XAI allows for a deeper understanding of input data, as well as the evaluation of stability and reproducibility of a method, providing a clearer decision-making process and helping clinicians to develop more reliable and trustworthy models. Lack of interpretability can also affect patient acceptance and adherence to rehabilitation technologies. Among the XAI models, the shapley additive exPlanations (SHAP), the local interpretable model-agnostic explanations (LIME), the layer-wise relevance propagation (LRP) and class activation mapping (CAM)-based methods are well-known XAI approaches applied in the clinical field [100]. In brain disease classification, it has been demonstrated that XAI can improve and validate the results and interpretability of AI models [101]. For example, the study in [47] employed LIME for stroke prediction using EEG signals, allowing the identification of relevant features and personalized the output.

From our analysis, it emerged that XAI approaches have been applied less frequently in the field of NIs neurorehabilitation, compared to fields such as neuroimaging. Yet, integrating explanability could allow for improving model performance (e.g., in terms of accuracy), but also to increase the adoption by clinicians and patients. For instance, in tasks such as ADLs, where different activities can be highly similar, XAI could support the AI decision-making process and provide more accurate feedback to the users. At the same time, assessing the quality of explanations remains an open challenge. Therefore, there is a clear need not only to develop new XAI methodologies in the field of NIs, but also to validate them, using metrics such as explanation goodness, user satisfaction, mental models, curiosity, trust, and human-AI performance [101].

E. Data Augmentation and Generative AI

As discussed above, model evaluation is one of the most important steps when assessing their ability of generalization. Model evaluation requires large and representative datasets to be included in the training set. However, in real-world conditions, acquiring large datasets is often difficult due to time constraints, ethical issues, cost and availability of instrumentation. Multicentric data could be a solution to increase sample sizes, but it frequently introduces heterogeneity. Assuming you have a good amount of data, data augmentation is a method to increase the size of the training set and reduce overfitting.

In the reviewed studies, data generation was performed by employing traditional methods such as sliding window or interpolation [34], [39], [45]. Côté Allard and colleagues evaluated 5 EMG data augmentation techniques (sliding window, muscle fatigue, electrode displacement, Gaussian noise, and aggregated augmentation) against the original dataset (non-augmented) [39]. Each method doubled the dataset size. Their results showed that only the sliding window approach enhanced the performance, and it was significantly different to the baseline. The advantage of using an overlapping sliding window is the creation of data that is not synthetic, as occurs for images. Further details of these specific methods for EMG augmentation are reported in Appendix B in Côté Allard et al. [39]. Few EEG channels are another challenge when using wearable sensors. To this end, Suwannarat et al. explored the feasibility of generating EEG data using the spherical spline interpolation (SSI) method to increase the spatial resolution [34]. However, this approach was not able to improve the classification performance. Thus, for classification tasks, such as motor imagery, it may be necessary to collect a sufficient number of EEG electrodes to cover the brain areas of interest.

Recently, generative AI techniques have been introduced in NIs development [102]. AI model could be trained to generate new synthetic data, both increasing temporal and spatial resolution of timeseries such as EEG or EMG. This could help in providing better generalization between subjects and also the capability of personalization. The most used models are: variational autoencoders (VAEs), generative adversarial networks (GANs), transformers and diffusion models [102]. From a practical perspective in neurorehabilitation, this approach could significantly reduce the calibration time of the NIs system, make the interface usable for a long time in view of telerehabilitation, and allow the use of wearable sensors with fewer electrodes. However, even in this case, the training of the generative model must be carefully evaluated.

The real challenge is not simply generating in-silico data, but making sure that what we generate truly mirrors the properties of real signals. Without rigorous and shared criteria, it is impossible to determine whether these data can be trusted and used in scientific or clinical practice. So far, no standardized method has been universally adopted to evaluate the quality of synthetic data. The most common approaches involve measuring how a task-related model performs with and without augmented data, or comparing generated data and original data using specific metrics, such as the L2 distance and the autocorrelation function [103]. Nonetheless, a systematic, universally applicable, model-agnostic evaluation framework is still lacking. Even with such a framework in place, several barriers still hinder its adoption in clinical and research contexts: the integration of multimodal data, compliance with privacy regulations, robustness and trustworthiness of generative models.

F. Real-Time Applications and Visual Feedback

The role of feedback in sensorimotor learning has long been considered of great importance [84], [104]. In line with this evidence, the literature on functional recovery after CNS lesion showed that closed-loop systems, i.e. devices with interfaces that provide augmented feedback, have the potential to promote motor recovery [105]. For this purpose, a real-time pipeline needs to be developed, balancing the models' performance, computational cost and data storage. Real-time applications have been proposed in some pilot studies with preliminary results [39], [45], [53], [69]. For example, Côté-Allard et al. demonstrated that real-time feedback, compared with the no-feedback framework, allows users to adapt their muscle activation strategy, reducing degradation in accuracy [39]. However, they observed a reduction in the average accuracy when comparing the offline evaluation and real-time experiment (98.31% vs. 95.42%, respectively). Indeed, in real-time scenario, there is a transition between gestures and the reaction delay of the users. This was not included during the training since it is too time-consuming. Another pilot study using a real-time classifier was proposed in [69]. In this case, participants performed 7 gestures, holding each pose for 10 s before resting and moving to the next without receiving feedback. The results failed to provide good performance. In comparison to the training, the protocol for the pilot study was modified. In particular, the classifier had 10 s to make the estimates compared to the seconds in the offline settings.

VII. Clinical Conditions

Although all the studies retrieved through our search aimed to develop a procedure for rehabilitation, none of them included an actual rehabilitation intervention, indicating that the translation of such technologies into clinical practice remains to be explored. Only the 18% (9/51) of the analyzed works included participants with neurological conditions and investigated the feasibility of the proposed approaches in people affected by stroke sequelae [37], [52], [55], [74], [81], [85] and SCI [69], while others did not specify the clinical characteristics of the participants [55], [82]. This result indicates that a significant gap remains between technical development and clinical application in this field. As previously mentioned, one of the main challenges is the intrinsic complexity of decoding movement intention in people with central nervous system lesions. In particular, the between-subjects heterogeneity, far more pronounced than what is found in healthy individuals, poses a significant challenge for training AI models that are robust for applications across different subjects. Although this is not an entirely new research topic, and previous works investigated the reliability and effectiveness of decoding motor intention and human movement using bioelectrical signals and wearable devices (e.g. [106]), with this review, we show that the use of AI models for this purpose is still under development. Overall, models' performance on impaired participants are more variable [85] or showed lower accuracy [52], [69] when compared to performance on healthy controls.

Interestingly, some of these works used generalized models trained on a sample of healthy or impaired subjects and tested on other participants [52], [74], [81], while others opted for individualized models trained and tested on the same participant [37], [69], [85]. From a general perspective, the high heterogeneity in UL muscular activation and kinematic patterns after a CNS lesion suggests that individualized models might, in theory, be more suitable for impaired users. However, this comes with the need of dedicating some time to train the model on each patient before actually starting the rehabilitation exercises. On the other hand, generalized and pre-trained models allow for more rapid set-up and use, but could show challenges in adapting to individual dysfunctions. It can be argued that these approaches are suitable for different aims: if a system is designed for assessment purposes, a generalized model should be privileged, as they require less time to set up and calibrate. In contrast, if the system is designed to be used throughout the rehabilitation phase, i.e. likely for 10-20 sessions, an initial phase of calibration could be worth considering to improve model accuracy.

Interestingly, one work reported that the fusion of EMG and kinematic data significantly improved the accuracy of their model compared to only EMG in people with stroke assessment [37]. Moreover, some analyses found that different levels of accuracy of the classification models seem to be associated with participant functional impairment [85]. Few studies analyzed only participants with neurological disorders without including a control group [55], [74], [81], [82]. Among this, in [74], [81], an approach that combined EMG and supervised learning was developed, suggesting the potential applicability for gesture recognition in paretic UL of people with stroke [74], [81]. Another study used data recorded by IMU and EMG to assess the correctness of rehabilitation exercises involving both UL and lower limb, finding a remarkable level of exercise recognition accuracy (96%) but limited automatic repetition segmentation accuracy compared to ground truth [82].

VIII. Discussion and Conclusion

In the last decade, the integration of AI-based approaches into conventional signal processing has significantly impacted research in the field of UL sensori-motor rehabilitation using wearable devices. To systematically describe and analyze the state-of-the-art, we conducted a systematic literature review of AI methods for decoding UL movement using wearable devices. In this review, 51 studies met the inclusion criteria. The analysis conducted revealed the great potential of these systems to decode UL poses. Despite the numerous published works in this field, few have been effectively tested in real-world rehabilitation scenarios. Especially in the medical field, many factors could affect the model's usability, including sensor variation, lack of task standardization, differences in patient characteristics, and their functional recovery pattern. Moreover, to effectively deploy NIs systems, technical factors need to be ensured, such as computational and memory resources, data sharing and privacy. Further research should be devoted to translating these approaches into real clinical contexts to detect possible issues and limitations.

In many cases, these factors are not taken into account, the methodological choices are not clearly explained, or are not compared to alternative strategies. Because the datasets used are different and most of them are not publicly available, direct comparison between studies is difficult. Indeed, some works' limitations are likely a consequence of the costs associated with collecting a large amount of data. Only 5 public databases were used as input to the AI models, all limited to EMG data [39], [40], [41], [42], [43], [44]. Our synthesis suggests that fostering the creation of public datasets seems to be a high priority for enriching research in this area. From an applicative perspective, most of the studies did not define a proper rehabilitation protocol, but rather a set of hand poses/movements that can be used in many other contexts. Of the included works, only 7 have validated or tested their method on patient data. This is a weak point of the actual state-of-the-art and thus requires much future work. A good starting point may be to design models that favor efficiency over complexity to analyze useful information for action recognition. The use of TL can help improve this aspect [39], [42], [44], [58], [59].

As shown in Fig. 2, classification models are predominantly used (44/51 reviewed studies). However, the development of regression models has emerged in the neural interface rehabilitation context mainly since 2021. It has the potential to push towards more advanced and personalized systems, particularly in telemedicine applications, where continuous monitoring and individualized feedback are pivotal. Of note, two attempts to use regression models were also found in a study of 2016 [49] and 2018 [60]. In the former, validation was performed on a single subject; in the latter, this information is not available. As shown in Figs. 5 and 4, the most commonly used classifiers are the SVM and LDA, which are well-accepted in this field due to their ability to classify biosignals. Despite this, some studies have started using DL models, comparing performance using different strategies. Among them, CNN and MLP performed the best. Alternative approaches have been recently proposed to address the time-variant nature of the extracted features and to analyze a large number of classes, such as RNN or transformer [40], [41]. As shown in Table II of the supplementary materials, the performance of the proposed models varied considerably. However, it can be observed that in the case of ADL recognition, the accuracy ranges between 78-89%, which is lower compared to the classification of the pose of the hand. In the latter case, the highest performance was obtained by DL models (e.g. 98%). In general, lower results were obtained for EEG decoding, suggesting that the EEG signal could be particularly complex to process in this context compared to other biosignals.

Notably, Little et al. [61] proposed a preliminary attempt to select input features based on their importance, using the mutual information index to rank the features. For this purpose, it could be interesting to explore which type of input data impacts the model's output using XAI approaches, with the potential to provide a deeper understanding of the problem. This will enhance the performance of the more complex models (e.g., when the number of classes is high or when the tasks are similar).

The included work did not discuss whether an accuracy threshold is recommended for a model to be translated into clinical practice. However, some considerations can be taken into account. Since accurate feedback on participant performance is essential to foster sensorimotor learning, in principle, the accuracy of these models should be as high as possible. Moreover, adding to the participant's errors throughout the training a non-negligible amount of trials where the system fails to classify the correct movement may result in increased frustration and lower engagement. At the state-of-the-art, the identification of a minimal recommended accuracy threshold remains an open issue that future research is called to address.

In conclusion, we provided an in-depth analysis of AI-based approaches for UL neurorehabilitation. We set the ground for future advancements and outlined the necessary steps to develop a more robust and reliable NIs system. We believe that DL-based and regression problems could be more useful in the clinical setting for remote applications. We also believe that in the coming years, due to technological development, the use of fully wearable EEG systems may provide relevant cognitive and neural insights, promoting the personalization and efficacy of rehabilitation interventions.

Acknowledgment

The authors thank Rossella Targa (Librarian of the University of Verona) for her support in the development of the search strings and the database queries.

References

[1] V. L. Feigin et al., "World stroke organization: Global stroke fact sheet 2025," *Int. J. Stroke*, vol. 20, no. 2, pp. 132–144, Feb. 2025.

[2] S. L. James et al., "Global, regional, and national burden of traumatic brain injury and spinal cord injury, 1990–2016: A systematic analysis for the global burden of disease study 2016," *Lancet Neurol.*, vol. 18, no. 1, pp. 56–87, 2018.

[3] M. T. Wallin et al., "Global, regional, and national burden of multiple sclerosis 1990–2016: A systematic analysis for the global burden of disease study 2016," *Lancet Neurol.*, vol. 18, no. 3, pp. 269–285, 2019.

[4] B. Amatya, F. Khan, and M. Galea, "Rehabilitation for people with multiple sclerosis: An overview of cochrane reviews," *Cochrane Database Systematic Rev.*, vol. 2019, no. 1, Jan. 2019, Art. no. CD012732.

[5] P. Feys and S. Straudi, "Beyond therapists: Technology-aided physical MS rehabilitation delivery," *Multiple Sclerosis J.*, vol. 25, no. 10, pp. 1387–1393, Sep. 2019.

[6] P. Langhorne, F. Coupar, and A. Pollock, "Motor recovery after stroke: A systematic review," *Lancet Neurol.*, vol. 8, no. 8, pp. 741–754, Aug. 2009.

[7] J. A. Kleim and T. A. Jones, "Principles of experience-dependent neural plasticity: Implications for rehabilitation after brain damage," *J. Speech, Lang., Hearing Res.*, vol. 51, no. 1, pp. S225–S239, Feb. 2008.

[8] E. Taub et al., "An operant approach to rehabilitation medicine: Overcoming learned nonuse by shaping," *J. Experim. Anal. Behav.*, vol. 61, no. 2, pp. 281–293, Mar. 1994.

[9] J. Liu, Z. Wang, C. Wang, and Y. Zhang, "Interventional effects of modified constraint-induced movement therapy on upper limb function in patients who had a stroke: Systematic review and meta-analysis," *BMJ Open*, vol. 15, no. 5, May 2025, Art. no. e094309.

[10] G. Kwakkel, J. M. Veerbeek, E. E. H. V. Wegen, and S. L. Wolf, "Constraint-induced movement therapy after stroke," *Lancet Neurol.*, vol. 14, no. 2, pp. 224–234, 2015.

[11] M. Gandolfi et al., "State of the art and challenges for the classification of studies on electromechanical and robotic devices in neurorehabilitation: A scoping review," *Eur. J. Phys. Rehabil. Med.*, vol. 57, no. 5, pp. 831–840, Nov. 2021.

[12] A. Pollock et al., "Interventions for improving upper limb function after stroke," *Cochrane Database Systematic Rev.*, vol. 2014, no. 11, Nov. 2014, Art. no. CD010820.

[13] N. Padfield, J. Zabalza, H. Zhao, V. Masero, and J. Ren, "EEG-based brain–computer interfaces using motor-imagery: Techniques and challenges," *Sensors*, vol. 19, no. 6, p. 1423, Mar. 2019.

[14] A. M. Ladda, F. Lebon, and M. Lotze, "Using motor imagery practice for improving motor performance—A review," *Brain Cognition*, vol. 150, Jun. 2021, Art. no. 105705.

[15] J. Mehrholz, A. Pollock, M. Pohl, J. Kugler, and B. Elsner, "Systematic review with network meta-analysis of randomized controlled trials of robotic-assisted arm training for improving activities of daily living and upper limb function after stroke," *J. NeuroEngineering Rehabil.*, vol. 17, no. 1, p.83, Dec. 2020.

[16] A. Basteris, S. M. Nijenhuis, A. H. Stienen, J. H. Buurke, G. B. Prange, and F. Amirabdollahian, "Training modalities in robot-mediated upper limb rehabilitation in stroke: A framework for classification based on a systematic review," *J. NeuroEngineering Rehabil.*, vol. 11, no. 1, p. 111, 2014.

[17] K. Boardsworth et al., "Upper limb robotic rehabilitation following stroke: A systematic review and meta-analysis investigating efficacy and the influence of device features and program parameters," *J. NeuroEngineering Rehabil.*, vol. 22, no. 1, p.164, Jul. 2025, doi: 10.1186/s12984-025-01662-4.

[18] C. R. Carvalho, J. M. Fernández, A. J. Del-Ama, F. Oliveira Barroso, and J. C. Moreno, "Review of electromyography onset detection methods for real-time control of robotic exoskeletons," *J. NeuroEngineering Rehabil.*, vol. 20, no. 1, p.141, Oct. 2023.

[19] J. R. Wolpaw et al., "Brain–computer interfaces for communication and control," *Clin. Neurophysiol.*, vol. 113, no. 6, pp. 91–767, 2002.

[20] Y. T. Lo et al., "Neural interface-based motor neuroprosthesis in poststroke upper limb neurorehabilitation: An individual patient data meta-analysis," *Arch. Phys. Med. Rehabil.*, vol. 105, no. 12, pp. 2336–2349, Dec. 2024.

[21] S. Waldert et al., "Hand movement direction decoded from MEG and EEG," *Neurosci. J.*, vol. 28, no. 4, pp. 1000–1008, Jan. 2008.

[22] F. Baldan et al., "Robot-assisted rehabilitation of hand function after stroke: Development of prediction models for reference to therapy," *J. Electromyogr. Kinesiol.*, vol. 57, Apr. 2021, Art. no. 102534.

[23] C. Vidaurre et al., "Challenges of neural interfaces for stroke motor rehabilitation," *Frontiers Human Neurosci.*, vol. 17, Sep. 2023, Art. no. 1070404.

[24] M. Grosse-Wentrup, D. Mattia, and K. Oweiss, "Using brain–computer interfaces to induce neural plasticity and restore function," *J. Neural Eng.*, vol. 8, no. 2, Apr. 2011, Art. no. 025004.

[25] N. Birbaumer and L. G. Cohen, "Brain–computer interfaces: Communication and restoration of movement in paralysis," *Physiol. J.*, vol. 579, no. 3, pp. 621–636, 2007.

[26] B. J. Edelman et al., "Non-invasive brain–computer interfaces: State of the art and trends," *IEEE Rev. Biomed. Eng.*, vol. 18, pp. 26–49, 2025.

[27] N. Valè et al., "Wearable EEG-IMU based framework for investigating neural correlates of motor-cognitive interaction in multiple sclerosis," in *Proc. 47th Annu. Int. Conf. IEEE Eng. Med. Biol. Soc. (EMBC)*, Jul. 2025, pp. 1–7.

[28] S. Federico et al., "Telerehabilitation for neurological motor impairment: A systematic review and meta-analysis on quality of life, satisfaction, and acceptance in stroke, multiple sclerosis, and Parkinson's disease," *J. Clin. Med.*, vol. 13, no. 1, p. 299, Jan. 2024.

[29] I. Senadheera et al., "AI applications in adult stroke recovery and rehabilitation: A scoping review using AI," *Sensors*, vol. 24, no. 20, p. 6585, Oct. 2024.

[30] D. T. Barry, "Adaptation, artificial intelligence, and physical medicine and rehabilitation," *PM&R*, vol. 10, no. 9S2, pp. 43–131, Sep. 2018.

[31] D. Lee and S. N. Yoon, "Application of artificial intelligence-based technologies in the healthcare industry: Opportunities and challenges," *Int. J. Environ. Res. Public Health*, vol. 18, no. 1, p. 271, Jan. 2021.

[32] G. Di Flumeri et al., "A neuroergonomic approach fostered by wearable EEG for the multimodal assessment of drivers trainees," *Sensors*, vol. 23, no. 20, p. 8389, Oct. 2023.

[33] S. M. Deniz et al., "Prediction of lifted weight category using EEG equipped headgear," in *Proc. IEEE-EMBS Int. Conf. Biomed. Health Informat. (BHI)*, Sep. 2022, pp. 1–7.

[34] A. Suwannarat, S. Pan-Ngum, and P. Israsena, "Analysis of minimal channel electroencephalography for wearable brain–computer interface," *Electronics*, vol. 13, no. 3, p. 565, Jan. 2024.

[35] G. Yang et al., "An IoT-enabled stroke rehabilitation system based on smart wearable armband and machine learning," *IEEE J. Transl. Eng. Health Med.*, vol. 6, pp. 1–10, 2018.

[36] T. K. Stanbury, J. G. C. Alfaro, S. Chinchalkar, and A. L. Trejos, "Identifying interaction forces via EMG under changing motion dynamics," in *Proc. Int. Conf. Rehabil. Robot. (ICORR)*, Jul. 2022, pp. 1–6.

[37] H. Zhou et al., "Toward hand pattern recognition in assistive and rehabilitation robotics using EMG and kinematics," *Frontiers Neurorobotics*, vol. 15, May 2021, Art. no. 659876.

[38] G. A. G. Ricardez et al., "Wearable device to record hand motions based on EMG and visual information," in *Proc. 14th IEEE/ASME Int. Conf. Mech. Embedded Syst. Appl. (MESA)*, Jul. 2018, pp. 1–6.

[39] U. Cote-Allard et al., "Deep learning for electromyographic hand gesture signal classification using transfer learning," *IEEE Trans. Neural Syst. Rehabil. Eng.*, vol. 27, no. 4, pp. 760–771, Apr. 2019.

[40] S. A. Salinas et al., "Comparison of machine learning techniques for activities of daily living classification with electromyographic data," in *Proc. Int. Conf. Rehabil. Robot. (ICORR)*, Jul. 2022, pp. 1–6.

[41] J. He, X. Niu, P. Zhao, C. Lin, and N. Jiang, "From forearm to wrist: Deep learning for surface electromyography-based gesture recognition," *IEEE Trans. Neural Syst. Rehabil. Eng.*, vol. 32, pp. 102–111, 2024.

[42] S. Kanoga, T. Hoshino, and H. Asoh, "Subject transfer framework based on source selection and semi-supervised style transfer mapping for semg pattern recognition," in *Proc. IEEE Int. Conf. Acoust., Speech Signal Process. (ICASSP)*, May 2020, pp. 1349–1353.

[43] S. Kanoga, T. Hoshino, and H. Asoh, "Semi-supervised style transfer mapping-based framework for sEMG-based pattern recognition with 1{-} or 2-DoF forearm motions," *Biomed. Signal Process. Control*, vol. 68, Jul. 2021, Art. no. 102817.

[44] S. Kanoga, A. Kanemura, and H. Asoh, "Are armband sEMG devices dense enough for long-term use?—Sensor placement shifts cause significant reduction in recognition accuracy," *Biomed. Signal Process. Control*, vol. 60, Jul. 2020, Art. no. 101981.

[45] Z. Yu, Y. Lu, Q. An, C. Chen, Y. Li, and Y. Wang, "Real-time multiple gesture recognition: Application of a lightweight individualized 1D CNN model to an edge computing system," *IEEE Trans. Neural Syst. Rehabil. Eng.*, vol. 30, pp. 990–998, 2022.

[46] D. Di Domenico et al., "Reach&grasp: A multimodal dataset of the whole upper-limb during simple and complex movements," *Sci. Data*, vol. 12, no. 1, p. 233, Feb. 2025.

[47] M. A. Islam, K. Sundaraj, R. B. Ahmad, and N. U. Ahamed, "Mechanomyogram for muscle function assessment: A review," *PLoS ONE*, vol. 8, no. 3, Mar. 2013, Art. no. e58902.

[48] S. Yang et al., "Learning-based motion-intention prediction for end-point control of upper-limb-assistive robots," *Sensors*, vol. 23, no. 6, p. 2998, Mar. 2023.

[49] M. Sakr and C. Menon, "Regressing force-myographic signals collected by an armband to estimate torque exerted by the wrist: A preliminary investigation," in *Proc. IEEE Can. Conf. Electr. Comput. Eng. (CCECE)*, May 2016, pp. 1–4.

[50] P. Benvenuti et al., "A multimodal wearable system for human activity recognition during daily activities," in *Proc. 8th Nat. Congr. Bioengineering*, 2023, pp. 1–4.

[51] M. S. Totty and E. Wade, "Muscle activation and inertial motion data for noninvasive classification of activities of daily living," *IEEE Trans. Biomed. Eng.*, vol. 65, no. 5, pp. 1069–1076, May 2018.

[52] A. Bertomeu-Motos et al., "Feasibility of an intelligent home-based neurorehabilitation system for upper extremity mobility assessment," *IEEE Sensors J.*, vol. 23, no. 24, pp. 31117–31124, Dec. 2023.

[53] E. Trigili et al., "Detection of movement onset using EMG signals for upper-limb exoskeletons in reaching tasks," *J. NeuroEngineering Rehabil.*, vol. 16, no. 1, p. 45, Dec. 2019.

[54] B. Chen et al., "Volitional control of upper-limb exoskeleton empowered by EMG sensors and machine learning computing," *Array*, vol. 17, Mar. 2023, Art. no. 100277.

[55] X. Chen, X. Pan, T. Ji, S. Yu, and Y. Sun, "Fusion classification of stroke patients' biosignals by weighted cross-validation-based feature selection (W-CVFS) method," *Biomed. Signal Process. Control*, vol. 80, Feb. 2023, Art. no. 104282.

[56] S. M. Rahman, M. A. Ali, and M. A. Al Mamun, "The use of wearable sensors for the classification of electromyographic signal patterns based on changes in the elbow joint angle," *Proc. Comput. Sci.*, vol. 185, pp. 338–344, Jun. 2021.

[57] M. Sierotowicz et al., "EMG-driven machine learning control of a soft glove for grasping assistance and rehabilitation," *IEEE Robot. Autom. Lett.*, vol. 7, no. 2, pp. 1566–1573, Apr. 2022.

[58] Y. Liu, S. Zhang, and M. Gowda, "NeuroPose: 3D hand pose tracking using EMG wearables," in *Proc. Web Conf.*, Apr. 2021, pp. 1471–1482.

[59] Y. Liu, S. Zhang, and M. Gowda, "A practical system for 3-D hand pose tracking using EMG wearables with applications to prosthetics and user interfaces," *IEEE Internet Things J.*, vol. 10, no. 4, pp. 3407–3427, Feb. 2023.

[60] P. Xiong, C. Wu, H. Zhou, A. Song, L. Hu, and X. P. Liu, "Design of an accurate end-of-arm force display system based on wearable arm gesture sensors and EMG sensors," *Inf. Fusion*, vol. 39, pp. 178–185, Jan. 2018.

[61] K. Little, B. K. Pappachan, S. Yang, B. Noronha, D. Campolo, and D. Accoto, "Elbow motion trajectory prediction using a multi-modal wearable system: A comparative analysis of machine learning techniques," *Sensors*, vol. 21, no. 2, p. 498, Jan. 2021.

[62] Z. Wu and X. Li, "A wireless surface EMG acquisition and gesture recognition system," in *Proc. 9th Int. Congr. Image Signal Process., Biomed. Eng. Informat. (CISP-BMEI)*, Oct. 2016, pp. 1675–1679.

[63] J. G. Colli-Alfaro, A. Ibrahim, and A. L. Trejos, "Design of user-independent hand gesture recognition using multilayer perceptron networks and sensor fusion techniques," in *Proc. IEEE 16th Int. Conf. Rehabil. Robot. (ICORR)*, Jun. 2019, pp. 1103–1108.

[64] J. G. Colli Alfaro and A. L. Trejos, "User-independent hand gesture recognition classification models using sensor fusion," *Sensors*, vol. 22, no. 4, p. 1321, Feb. 2022.

[65] N. Nazmi, M. Abdul Rahman, S.-I. Yamamoto, S. Ahmad, H. Zamzuri, and S. Mazlan, "A review of classification techniques of EMG signals during isotonic and isometric contractions," *Sensors*, vol. 16, no. 8, p. 1304, Aug. 2016.

[66] B. Hudgins, P. Parker, and R. N. Scott, "A new strategy for multifunction myoelectric control," *IEEE Trans. Biomed. Eng.*, vol. 40, no. 1, pp. 82–94, Jan. 1993.

[67] X. Yang, S.-C. Yeh, J. Niu, Y. Gong, and G. Yang, "Hand rehabilitation using virtual reality electromyography signals," in *Proc. 5th Int. Conf. Enterprise Syst. (ES)*, Sep. 2017, pp. 125–131.

[68] S. Singhvi and H. Ren, "Comparative study of motion recognition with temporal modelling of electromyography for thumb and index finger movements aiming for wearable robotic finger exercises," in *Proc. 3rd Int. Conf. Adv. Robot. Mechatronics (ICARM)*, Jul. 2018, pp. 509–514.

[69] J. E. Britt, M. K. O'Malley, and C. G. Rose, "Electromyographic classification to control the SPAR glove," *IFAC-PapersOnLine*, vol. 54, no. 20, pp. 244–250, 2021.

[70] I. Mendez et al., "Evaluation of the myo armband for the classification of hand motions," in *Proc. Int. Conf. Rehabil. Robot. (ICORR)*, Jul. 2017, pp. 1211–1214.

[71] L. F. Aguiar and A. P. L. Bó, "Hand gestures recognition using electromyography for bilateral upper limb rehabilitation," in *Proc. IEEE Life Sci. Conf. (LSC)*, Dec. 2017, pp. 63–66.

[72] L. Liu, N. Zhang, and K. Li, "Research on gesture recognition of surface EMG based on machine learning," in *Proc. Chin. Conf. Biometric Recognit.*, 2022, pp. 276–283.

[73] M. Gomez-Correa and D. Cruz-Ortiz, "Low-cost wearable band sensors of surface electromyography for detecting hand movements," *Sensors*, vol. 22, no. 16, p. 5931, Aug. 2022.

[74] A. Anastasiev et al., "Empirical myoelectric feature extraction and pattern recognition in hemiplegic distal movement decoding," *Bioengineering*, vol. 10, no. 7, p. 866, Jul. 2023.

[75] F. S. Sayin, S. Ozen, and U. Baspinar, "Hand gesture recognition by using sEMG signals for human machine interaction applications," in *Proc. Signal Process., Algorithms, Archit., Arrangements, Appl. Conf. (SPA)*, Sep. 2018, pp. 27–30.

[76] V. Kartsch, S. Benatti, M. Mancini, M. Magno, and L. Benini, "Smart wearable wristband for EMG based gesture recognition powered by solar energy harvester," in *Proc. IEEE Int. Symp. Circuits Syst. (ISCAS)*, May 2018, pp. 1–5.

[77] I. Siviero, L. Brusini, G. Menegaz, and S. F. Storti, "Motor-imagery EEG signal decoding using multichannel-empirical wavelet transform for brain computer interfaces," in *Proc. IEEE-EMBS Int. Conf. Biomed. Health Informat. (BHI)*, Sep. 2022, pp. 1–4.

[78] M. B. I. Reaz, M. S. Hussain, and F. Mohd-Yasin, "Techniques of EMG signal analysis: Detection, processing, classification and applications," *Biol. Procedures Online*, vol. 8, no. 1, pp. 11–35, Dec. 2006.

[79] N. Anil and S. H. Sreeletha, "EMG based gesture recognition using machine learning," in *Proc. 2nd Int. Conf. Intell. Comput. Control Syst. (ICICCS)*, Jun. 2018, pp. 1560–1564.

[80] E. Donati, M. Payvand, N. Risi, R. Krause, and G. Indiveri, "Discrimination of EMG signals using a neuromorphic implementation of a spiking neural network," *IEEE Trans. Biomed. Circuits Syst.*, vol. 13, no. 5, pp. 795–803, Oct. 2019.

[81] A. Anastasiev et al., "Supervised myoelectrical hand gesture recognition in post-acute stroke patients with upper limb paresis on affected and non-affected sides," *Sensors*, vol. 22, no. 22, p. 8733, Nov. 2022.

[82] A. Pereira, D. Folgado, F. Nunes, J. Almeida, and I. Sousa, "Using inertial sensors to evaluate exercise correctness in electromyography-based home rehabilitation systems," in *Proc. IEEE Int. Symp. Med. Meas. Appl. (MeMeA)*, Jun. 2019, pp. 1–6.

[83] K. Chen, F.-Y. Zhou, and X.-F. Yuan, "Hybrid particle swarm optimization with spiral-shaped mechanism for feature selection," *Expert Syst. Appl.*, vol. 128, pp. 140–156, Aug. 2019.

[84] D. M. Wolpert, J. Diedrichsen, and J. R. Flanagan, "Principles of sensorimotor learning," *Nature Rev. Neurosci.*, vol. 12, no. 12, pp. 739–751, Dec. 2011.

[85] F. Ryser, T. Bützer, J. P. Held, O. Lambercy, and R. Gassert, "Fully embedded myoelectric control for a wearable robotic hand orthosis," in *Proc. Int. Conf. Rehabil. Robot. (ICORR)*, Jul. 2017, pp. 615–621.

[86] A. Mongardi et al., "Hand gestures recognition for human-machine interfaces: A low-power bio-inspired armband," *IEEE Trans. Biomed. Circuits Syst.*, vol. 16, no. 6, pp. 1348–1365, Dec. 2022.

[87] F. Amirabdollahian and M. L. Walters, "Application of support vector machines in detecting hand grasp gestures using a commercially off the shelf wireless myoelectric armband," in *Proc. Int. Conf. Rehabil. Robot. (ICORR)*, Jul. 2017, pp. 111–115.

[88] M. Tangermann et al., "Review of the BCI competition IV," *Frontiers Neurosci.*, vol. 6, p. 55, Jul. 2012.

[89] S. He, C. Yang, M. Wang, L. Cheng, and Z. Hu, "Hand gesture recognition using MYO armband," in *Proc. Chin. Autom. Congr. (CAC)*, Oct. 2017, pp. 4850–4855.

[90] A. R. Ali, M. W. A. Ramadan, and A. T. Darwish, "Neural network-based classification of hand gestures using electromyography sensor data," in *Proc. Int. Telecommun. Conf. (ITC-Egypt)*, Jul. 2023, pp. 36–231.

[91] C. D. Olsen, W. C. Hamrick, S. R. Lewis, M. M. Iverson, and J. A. George, "Wrist EMG improves gesture classification for stroke patients," in *Proc. Int. Conf. Rehabil. Robot. (ICORR)*, Sep. 2023, pp. 1–6.

[92] M. Zia ur Rehman et al., "Multiday EMG-based classification of hand motions with deep learning techniques," *Sensors*, vol. 18, no. 8, p. 2497, Aug. 2018.

[93] I. Siviero et al., "Remote motor rehabilitation: EMG-IMU based deep learning model improves the estimate of wrist kinematics," in *Proc. 46th Annu. Int. Conf. IEEE Eng. Med. Biol. Soc. (EMBC)*, Jul. 2024, pp. 1–4.

[94] S. Hochreiter and J. Schmidhuber, "Long short-term memory," *Neural Comput.*, vol. 9, no. 8, pp. 1735–1780, Nov. 1997.

[95] A. Vaswani et al., "Attention is all you need," in *Proc. Adv. Neural Inf. Process. Syst.*, vol. 30, 2025, pp. 5998–6008.

[96] B. Abibullaev, A. Keutayeva, and A. Zollanvari, "Deep learning in EEG-based BCIs: A comprehensive review of transformer models, advantages, challenges, and applications," *IEEE Access*, vol. 11, pp. 127271–127301, 2023.

[97] S. Saha and M. Baumert, "Intra{-} and inter-subject variability in EEG-based sensorimotor brain computer interface: A review," *Frontiers Comput. Neurosci.*, vol. 13, p. 87, Jan. 2020.

[98] I. Boscolo Galazzo et al., "Explainable artificial intelligence for magnetic resonance imaging aging brainprints: Grounds and challenges," *IEEE Signal Process. Mag.*, vol. 39, no. 2, pp. 99–116, Mar. 2022.

[99] P. Linardatos, V. Papastefanopoulos, and S. Kotsiantis, "Explainable AI: A review of machine learning interpretability methods," *Entropy*, vol. 23, no. 1, p. 18, Dec. 2020.

[100] A. Bennetot et al., "A practical tutorial on explainable AI techniques," *ACM Comput. Surv.*, vol. 57, no. 2, pp. 1–44, Feb. 2025.

[101] N. Bibi, J. Courtney, and K. McGuinness, "Enhancing brain disease diagnosis with XAI: A review of recent studies," *ACM Trans. Comput. Healthcare*, vol. 6, no. 2, pp. 1–35, Apr. 2025.

[102] S. Eldawlatly, "On the role of generative artificial intelligence in the development of brain–computer interfaces," *BMC Biomed. Eng.*, vol. 6, no. 1, p. 4, May 2024.

[103] A. G. Habashi, A. M. Azab, S. Eldawlatly, and G. M. Aly, "Generative adversarial networks in EEG analysis: An overview," *J. NeuroEng. Rehabil.*, vol. 20, no. 1, p.40, Apr. 2023.

[104] F. Castro and K. C. Schenke, "Augmented action observation: Theory and practical applications in sensorimotor rehabilitation," *Neuropsychological Rehabil.*, vol. 34, no. 9, pp. 1327–1346, Oct. 2024.

[105] J. J. Daly and J. R. Wolpaw, "Brain–computer interfaces in neurological rehabilitation," *Lancet Neurol.*, vol. 7, no. 11, pp. 1032–1043, 2008.

[106] D. T. Bundy et al., "Contralesional brain–computer interface control of a powered exoskeleton for motor recovery in chronic stroke survivors," *Stroke*, vol. 48, no. 7, pp. 1908–1915, Jul. 2017.