



Review Article

Ann Rehabil Med 2023;47(6):444-458

eISSN: 2234-0653

<https://doi.org/10.5535/arm.23131>

AI in Rehabilitation Medicine: Opportunities and Challenges

Francesco Lanotte, PhD^{1,2*}, Megan K. O'Brien, PhD^{1,2*}, Arun Jayaraman, PT, PhD^{1,2}

¹Max Nader Lab for Rehabilitation Technologies and Outcomes Research, Shirley Ryan AbilityLab, Chicago, IL, United States

²Department of Physical Medicine and Rehabilitation, Northwestern University, Chicago, IL, United States

Artificial intelligence (AI) tools are increasingly able to learn from larger and more complex data, thus allowing clinicians and scientists to gain new insights from the information they collect about their patients every day. In rehabilitation medicine, AI can be used to find patterns in huge amounts of healthcare data. These patterns can then be leveraged at the individual level, to design personalized care strategies and interventions to optimize each patient's outcomes. However, building effective AI tools requires many careful considerations about how we collect and handle data, how we train the models, and how we interpret results. In this perspective, we discuss some of the current opportunities and challenges for AI in rehabilitation. We first review recent trends in AI for the screening, diagnosis, treatment, and continuous monitoring of disease or injury, with a special focus on the different types of healthcare data used for these applications. We then examine potential barriers to designing and integrating AI into the clinical workflow, and we propose an end-to-end framework to address these barriers and guide the development of effective AI for rehabilitation. Finally, we present ideas for future work to pave the way for AI implementation in real-world rehabilitation practices.

Keywords: Machine learning, Rehabilitation outcome, Wearable devices, Computer vision systems, Precision medicine

Received: September 18, 2023

Accepted: November 23, 2023

Correspondence:

Arun Jayaraman
Max Nader Lab for Rehabilitation
Technologies and Outcomes Research,
Shirley Ryan AbilityLab, 355 E Erie St.,
Chicago, IL 60611, United States.
Tel: +1-312-238-6875
Fax: +1-312-238-0000
E-mail: ajayaraman@sralab.org

*These authors contributed equally to this work.

INTRODUCTION

Artificial intelligence (AI) is poised to revolutionize the healthcare industry. By combining intelligent algorithms with massive volumes of training data, AI systems can solve problems and make logical inferences more quickly and reliably than humans, and they can also learn from larger and more complex datasets [1]. While early healthcare AI applications have focused on automating technical tasks (e.g., detecting arrhythmia from electrocardiograms [ECG], or segmenting and interpreting medical images), new advances are shifting AI into broader areas of

screening, diagnosis, treatment, and prevention of disease and injury, as well as clinical decision support [2,3]. Within these emerging domains, AI offers many unique and exciting opportunities in rehabilitation.

The portfolio of rehabilitation therapies and interventions has expanded significantly over the last few decades, including new strategies for high-intensity training, assistive robotics, pharmacologics, and neurostimulation. However, many clinicians struggle to identify the most effective strategies for individual patients that will maximize their functional recovery and minimize their impairments [4]. Often, rehabilitation systems favor a

one-size-fits-all approach, wherein patients receive similar therapy structures and dosages based on the best-known evidence and nationwide insurance reimbursement models. Another challenge is that patient information is often *siloes* in each care setting. That is, although a wealth of patient information is scattered throughout the healthcare system (e.g., regular check-ups with a primary care physician, or emergency room visits from previous trauma), there is limited accessibility and interoperability of information between different settings (Fig. 1A) [5,6]. As a result, clinicians may not have all the information they need to design the best treatments for a patient, which reduces the efficacy and efficiency of care.

With progressive shifts toward digitalized medicine and portable technology, there is more data available than ever before for us to understand the time course of disease and recovery for different patient cases. AI models—specifically, machine learning algorithms—can be trained to mine the existing data silos and combine complex data to identify patient- or population-specific biomarkers of disability, disease, and injury (Fig.

1). This approach has the potential to revolutionize how we assess patients, both in the clinic and out in the community, and empower us with actionable knowledge. For instance, AI tools could support the design of tailored rehabilitation programs precisely matched to specific impairments [5], timely referral decisions [7], and comprehensive care plans for clinicians and families. The result would directly address the surging demand for personalized and precision medicine in healthcare [2], enhancing patient engagement, cost-effectiveness, and the overall quality of care [8].

In this perspective, we examine the opportunities and challenges for AI in rehabilitation medicine. First, we explore the current research trends in this field, highlighting unique insights from different types of data. Second, we identify key barriers that are impeding the translation of AI into everyday clinical practice. Third, we propose an end-to-end framework for creating clinically meaningful AI models for rehabilitation applications. Lastly, we discuss potential directions for future work.

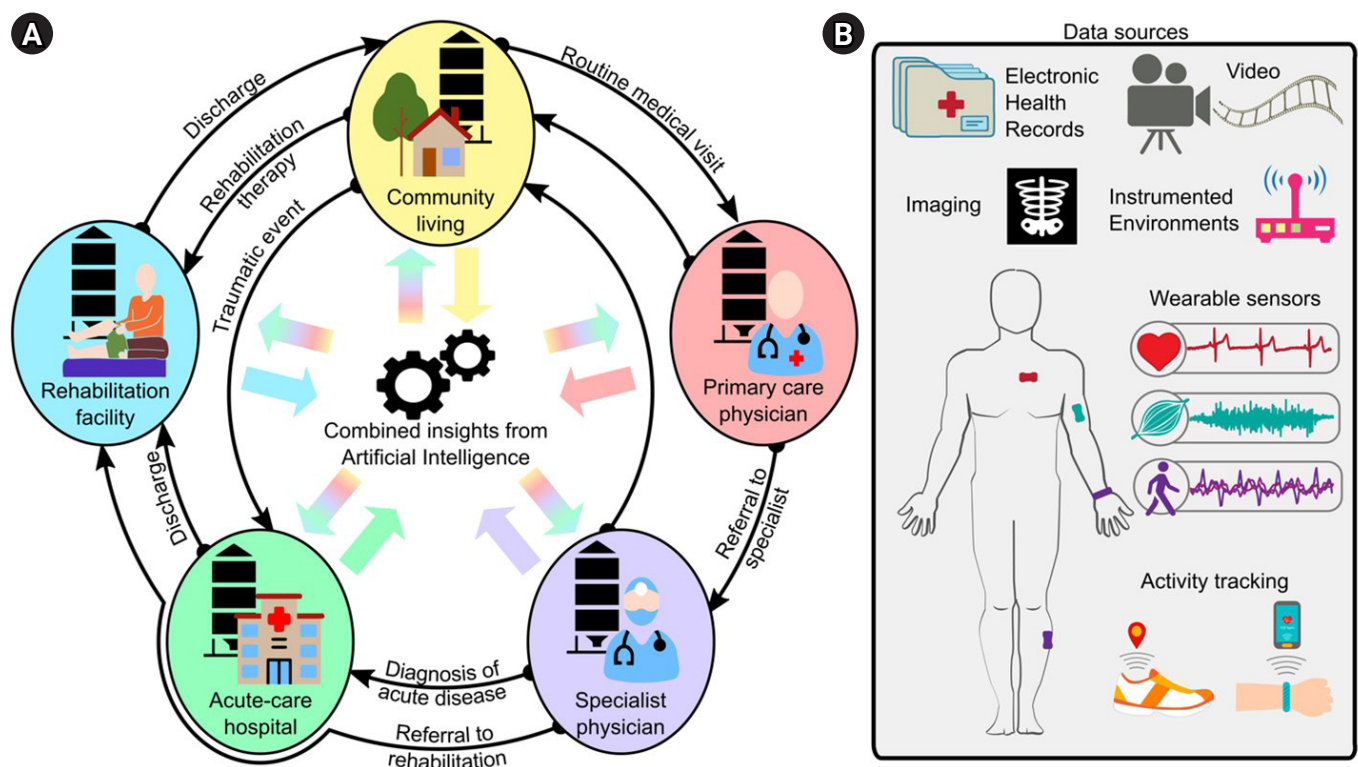


Fig. 1. Data for rehabilitation medicine. (A) Traditionally, data is siloed in the different stages of medical care (e.g., community living, primary care physician, specialist physician, acute-care hospitals, rehabilitation facilities, etc.), with limited data mobility as patients transition between care settings. Artificial intelligence (AI) can integrate this information for a tailored and comprehensive evaluation of the health status of an individual. (B) Example data sources for AI applications in rehabilitation medicine.

THE DATA LANDSCAPE: WHICH DATA ARE USED FOR AI IN REHABILITATION?

A person's health can be described by an extensive amount of data (Fig. 1). In a medical care setting, health data may include diagnostic exams, functional test scores, medications, and laboratory analyses, complemented by demographic information, medical history, comorbidities, and additional clinical notes [9,10]. These data are usually recorded in an electronic health record (EHR) for each patient, which is continuously updated and reviewed by the clinical team. Additionally, advances in wearable sensor technology have enabled continuous, high-resolution monitoring of vitals and activity during the hospital stay and, notably, in people's homes and community [11]. Instrumented environments and video recordings can also collect these data in a contactless, non-intrusive way. Each of these data sources and example applications for rehabilitation are discussed below.

Electronic health records

EHRs are a wellspring of information for AI, often serving as the *de facto* data source for models that are targeting automatic patient screening, early detection, and prognosis.

Screening and early detection are among the most popular and widespread applications of EHR-based AI in healthcare research. In contrast to traditional screening tools that require expensive, time-consuming procedures, AI models can learn subtle patterns and precursors of disease, and then automatically identify at-risk patients using longitudinal information stored in EHRs [12,13]. For instance, models trained on routine EHR data have detected autism spectrum disorder in infants as early as 30 days after birth (nearly one year earlier than standard autism screening tools) [14], and they have detected latent diseases in adults such as peripheral artery disease [15]. They have also identified individuals at high risk of falling [16], and predicted the development of pressure ulcers within the first 24 hours of admission to an intensive care unit [17]. In these examples, earlier detection and more accurate screening can help clinicians intervene with appropriate care or prevention strategies, thereby improving patient care quality in rehabilitation.

Additionally, EHR-driven AI can be used for early, data-driven prognosis, which would assist with short- and long-term care planning, patient goal setting, and identifying appropriate candidates for different treatments [18,19]. In acute stroke rehabilitation, prognostic models have recently shown promise in pre-

dicting future walking ability [20–22], functional independence [21,23], and balance [21] at an inpatient rehabilitation facility (IRF). These models have incorporated EHR data (e.g., demographic and clinical information) collected at IRF admission to predict a patient's ability at IRF discharge. Beyond the inpatient setting, longitudinal predictions of postdischarge recovery can assist with outpatient therapy planning [24]. For instance, the TWIST algorithm was designed to predict the probability of independent walking recovery up to 26 weeks poststroke using early EHR data [25], while the PREP2 combines EHR and imaging data to predict upper limb function three months poststroke [26].

Imaging

AI can be trained to segment and interpret medical images, such as from X-ray radiography or magnetic resonance imaging, to detect disease-related anomalies such as tumor masses [27,28], cardiovascular abnormalities [29], retinal glaucoma [30], and reduced grey matter [31]. Markers extracted from medical imaging have supplemented EHR data in AI models to increase diagnostic accuracy, determine disease severity, or evaluate recovery potential [32].

Wearable sensors

Today's wearable sensors can record a plethora of health-related information, spanning physiological, biomechanical, behavioral, and activity measures. Unlike EHR data, which are collected during brief patient-clinician interactions, wearable sensors can collect biometric data continuously and at much higher resolutions [33]. In the past, obtaining precise measurements of body kinematics, muscle activity, or vital signs required specialized environments, trained experts, and costly equipment. But ongoing technological advancements are bringing these and other measurements to wireless, portable, and cost-effective body-worn devices that can be deployed easily in any setting, including the community [11].

Wearable sensor data paired with AI are fostering new ways of measuring disease-specific indicators of function and impairment, complementing traditional clinical assessments [34]. For instance, inertial measurement units (IMUs), electromyography (EMG), or ECG sensors capture valuable data related to patient movement and neurological function. In acute stroke rehabilitation, we observed that sensor data improved the performance of a model predicting future walking ability, surpassing models using only EHR and other standard clinical information [35].

Similarly, sensor data during walking have been combined with clinical information to estimate dynamic balance ability in individuals with stroke, multiple sclerosis, and Parkinson's disease [36]. In the upper limb, IMUs have shown promise in assessing motor deficits in stroke and traumatic brain injury during the Wolf Motor Function Test [37], or evaluating tremor and bradykinesia for individuals with Parkinson's disease [38,39]. Novel sensors recording high-frequency vibrations can also quantify and monitor swallowing impairment for individuals with dysphagia [40]. These and other sensor-based AI tools can be used to monitor disease progression and therapy impact [41].

Consumer smartphones and smartwatches also incorporate various sensors, including IMUs to capture movement, GPS (Global Positioning System) modules to track geographic location, and optical sensors to estimate vitals such as heart rate and blood oxygenation. AI analysis of these signals can generate diverse measures of physical activity and step count [42], movement impairment [33], community mobility [43], heart rate and cardiovascular health [44], sleep [45], fall risk [46], and more. We recently applied AI to consumer wearables data to assess postoperative recovery in patients who underwent pediatric appendectomy, analyzing patterns in patient activity, heart rate, and sleep measures to detect early signs of complications [47]. Activity recognition algorithms from wearable devices can also indicate exercise adherence or changes in daily activities post-IRF discharge [48].

Instrumented environments

Instrumented environments are an emerging technique for patient monitoring. By strategically installing sensors in the patient's surroundings, this approach is a markerless and contactless method of collecting clinically-meaningful data in the hospital or at home. For example, devices emitting low-power radio signals can estimate respiration and heart rate by analyzing the signals reflected off the body [49]. Combining these data with AI could enable continuous vitals monitoring or symptom evaluation for COVID-19 [50], Parkinson's disease [51], and other conditions. Like wearable sensors, instrumented environments may offer significant economic advantages by reducing the need for regular clinic visits [52].

Videos

Another markerless data acquisition technique is human pose estimation, which uses AI to automatically detect body landmarks from videos and quantify movement, function, and im-

pairment [53]. Pose estimation is becoming more commonplace in applications like gait analysis, since it reduces dependency on costly optoelectronic motion capture equipment. Video-based gait metrics have been computed in healthy populations [54], people with Parkinson's disease [55] and stroke [56], as well as for general functional assessment [57].

More recently, AI has been applied to automatically score clinical assessments from video, such as the Movement Disorder Society Unified Parkinson's Disease Rating Scale in individuals with Parkinson's disease [58], or the General Movements Assessment in young infants [59]. In these examples, video-based AI offers a scalable solution to enhance the reliability and accessibility of valuable clinical assessments, which require specialized training and considerable practice for scoring.

BARRIERS TO AI IN CURRENT REHABILITATION PRACTICES

Despite the extensive research in AI and the burgeoning availability of healthcare-related data, integrating these models into real-world clinical practice remains a significant open challenge. We believe there are three key barriers that can fundamentally limit the translation of AI models in rehabilitation:

1. Lack of interoperability: As described above, AI can combine clinical and community data to make intelligent inferences about a patient's current, past, or future health. However, the heterogeneous nature of data reporting across different sources can lead to biased, inaccurate, or inexecutable models [7]. To mitigate this, data reporting should be standardized to ensure the *interoperability* of data that can be curated for the models [60].
2. Lack of transparency: The black-box nature of AI can also limit its widespread adoption in rehabilitation [61]. Uncertainty about the computational processes or validity of AI, paired with inevitable model errors and performance fluctuation during initial deployment and ongoing tuning, can easily generate feelings of distrust for AI decision-support tools. To make models *transparent*, developers should provide clinicians with accountable and user-friendly guidelines to interpret the goodness of AI predictions and potential sources of error.
3. Lack of actionability: Insights from AI should also be *actionable*, enabling clinicians to identify or modify care strategies to improve patient outcomes [62]. Different techniques have emerged in recent years to explain the complex

interactions across features used by the models [61,63]. Information from the most predictive or model-driving features could help clinicians design new treatment strategies for their patients. However, all models and their underlying features should be interpreted with caution since they do not always reveal causal effects.

Additional operational or performance barriers—including requirements for data storage and security, cost constraints, resource limitations, education and training challenges, regulatory and ethical complexities, and issues related to scalability and generalization—can also impede AI implementation. Overcoming these additional barriers at the site will require close collaboration between AI developers, healthcare providers, institutional regulators, and policymakers.

FRAMEWORK FOR AI DEVELOPMENT IN REHABILITATION

We offer an end-to-end framework to address the three key barriers above and guide AI development for rehabilitation. The framework identifies the high-level processes needed to bridge the gap between multidimensional data input and meaningful model output in the clinical or research setting. The framework also details the specific steps at which to incorporate the attributes of interoperability, transparency, and actionability to maximize the translational impact of AI in rehabilitation.

Defining the target output

When developing AI for any application, defining the model's output and use case is essential. In rehabilitation, the model output may be automated scores from standard clinical processes (e.g., gait speed, balance score, heart arrhythmia count), novel measures of body function and impairment not typically available in a clinic (e.g., joint kinematics, gait symmetry, muscle activation), or a prediction of a patient's outcomes (e.g., disease detection, prognosis, discharge location). The model may be intended for use in specific circumstances in the clinic or community, across or within patient groups, and/or during continuous, real-time monitoring or a snapshot in time. Clinicians and researchers familiar with the target output can provide insights on appropriate data to capture the target output based on potential confounds of disease, medication, comorbidities, and so forth. These expert collaborations are critical for robust AI development in the highly specialized rehabilitation setting, increasing the chances of successful translation.

Translating healthcare data into the target output

After defining the target output, AI developers can follow a 7-step framework to obtain the output from available healthcare data (Fig. 2). Although the attributes of interoperability, transparency, and actionability should be considered throughout the framework, we will identify the particularly relevant attributes for each step.

Step 1. Data configuration

The first step is selecting and configuring the data source(s) for the AI model (e.g., see Fig. 1B). For example, developers may need to decide the specific variables to be collected from each source, the recording parameters and placement of sensors/cameras/instrumentation, and the recording duration and frequency. These and other configuration decisions will directly impact the nature and quality of recorded data, as well as the downstream *interoperability* and *actionability* of the AI model.

When using EHR data in an AI model, developers should include EHR variables with established or hypothesized relationships to the target AI output. For instance, age, comorbidities, and medication use have all been linked to gait speed prediction [64]. Including known predictors of the target output will likely account for more inter- and intrasubject variance, thereby maximizing model performance.

When using wearable sensor data, developers should carefully consider options for different sensing modalities, body location, adhesion methods, and sampling characteristics to capture the signal of interest while minimizing problematic noise. For example, detecting gait events from IMUs is typically improved by placing a body-worn sensor closer to the point of impact with the floor [65]. Using multiple sensors on different body locations can increase the accuracy of activity detection [66] or sleep stage monitoring [45], although it may also increase power consumption and reduce patient compliance. Preliminary studies considering multi-sensor systems should begin with more complex configurations to capture enough predictors and data resolution for the target output, then streamline as necessary (see Step 6. *Streamlining*).

When using video recordings, the choice of cameras can affect the AI model performance [52]. Conventional 2D RGB cameras are generally suitable for automatic annotation and pose estimation. Depth cameras may offer improved accuracy for certain applications, such as activity recognition, particularly when combined with other sources like wearable sensors [67] or thermal cameras [68]. Video recording configurations, such as

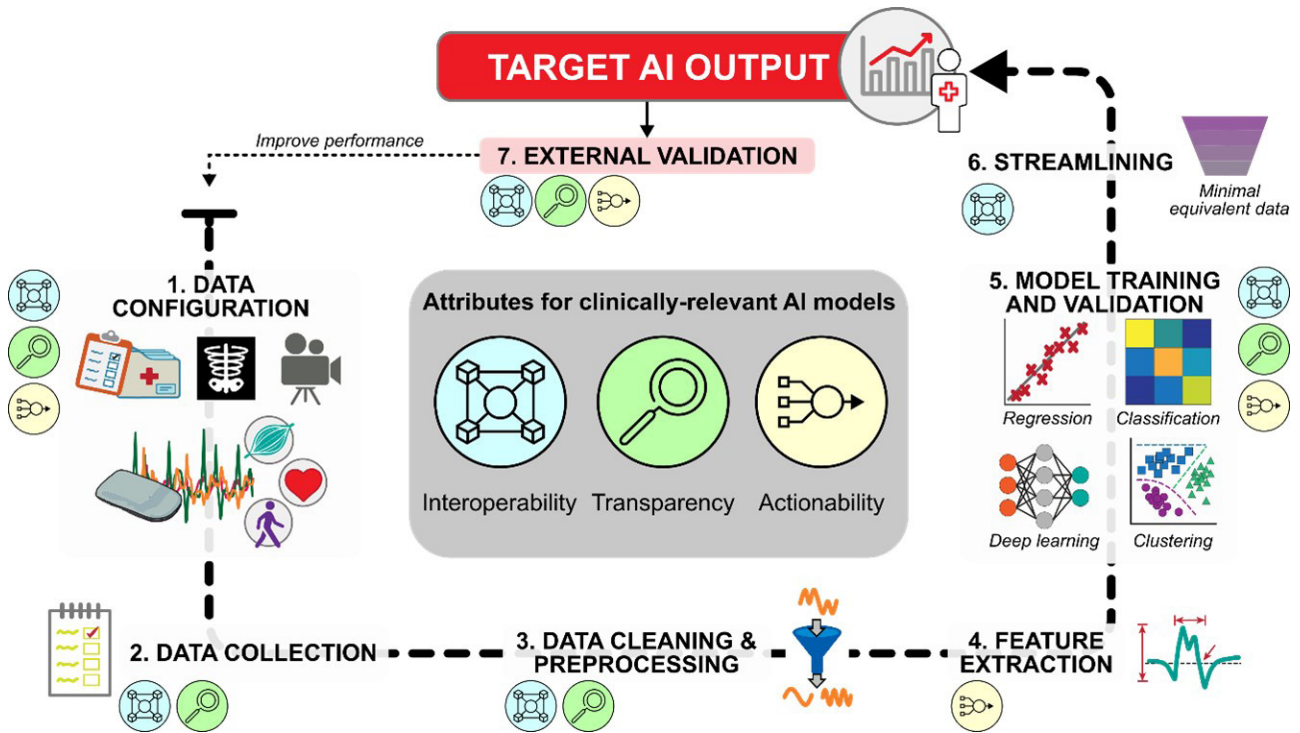


Fig. 2. Framework for developing artificial intelligence (AI) models with clinically relevant attributes. Input data, such as EHR (electronic health record), imaging, sensor data, or video recordings (see Fig. 1B), are first selected and validated for their ability to capture the desired target output. Single/multimodal data streams are collected, processed, and passed to AI algorithms for inferential and/or predictive analytics. After establishing baseline model performance, streamlined datasets can be tested to determine the minimal data needed to achieve this performance, thereby supporting practical clinical implementation. The final model should be externally validated to determine its generalizability to an independent dataset. If performance is not satisfactory, the framework can be revisited in whole or in part.

frame rate or resolution, and environmental conditions, such as lighting or occlusions, should be carefully considered, as these decisions can impact data quality, storage, and processing time.

Before using technology-acquired data as inputs to an AI model, such as measurements from sensors or video, there should be established evidence that these measurements accurately represent the true values (e.g., of motion, vitals, activity, etc.). These measurements should be validated against the “gold standard” measurement technique to assess their accuracy, reliability, and agreement level [69-71]. Measurement validation should, as best as possible, include the expected environmental conditions and patients that are representative of the model’s expected use case [48,72,73]. Protocols for measurement validation should be reported in detail to increase the later AI model’s *transparency* to potential sources of error. This will help users understand whether the input data and model can be used, or whether they should be interpreted differently, outside of the

validation conditions.

Step 2. Data collection

Data collection for AI training and testing should obtain high-quality data that are representative of the expected use case, such as during real-world clinical scenarios and from a diverse patient cohort.

Determining the appropriate sample sizes for training and testing data for robust AI is an ongoing challenge in this field. Generally, large amounts of data from diverse samples will create a more generalizable model, and multiple repetitions of the data collection protocol for each patient will account for intra-subject and intersubject variability.

Data annotations are critical to contextualize and select appropriate data during model training. Example annotations might include the activity type (e.g., walking, plantarflexion, sleep stage, etc.), clinical test scores, EHR items, use of assistive

devices or orthotics, or required assistance levels for activity completion. Failure to collect a well-distribution set of annotated data leads to imbalanced datasets and/or algorithms with inherent biases, such as to race, gender, and age [74,75]. Understanding the interactions between algorithm performance and the contextualized data, as well as their ethical implications, should guide data collection practices to mitigate potential sources of bias. This will also allow other developers to reproduce the models and integrate new data, leading to larger, more diverse datasets and enhanced AI reliability and validity.

Step 3. Data cleaning & preprocessing

Data collected in controlled laboratory or research environments is often cleaner and easier to annotate than data collected in real-world settings like clinics, homes, and communities. Data sources like EHR, wearable sensors, and videos often contain data artifacts, such as transcription errors, discrepancies in patient documentation, missing values, poor recording conditions, and technology failures. To mitigate these issues, the dataset undergoes cleaning and initial processing, such as harmonization, handling missing data, resampling, filtering, and other transformations.

Data harmonization involves expressing data into a common architecture, thereby facilitating *interoperability* between data sources and rehabilitation sites. For EHR data, harmonization might include standardizing categorical information, such as patient demographics or medical conditions, in formats or categories used in different EHR datasets [76]. For other data sources, harmonization can include using uniform measurement scales or synchronizing temporal data, such as when recording multiple time-varying signals (e.g., from sensors, video, imaging, etc.) during a single clinical task.

Missing data poses a critical challenge for AI. Missing data might arise for numerous reasons, such as incomplete data entry, data loss from devices not being worn (or worn with a depleted battery), software glitches, or patient noncompliance. The most straightforward solution is to exclude samples with missing data from the dataset, but this drastically reduces the available data for model training. Alternatively, imputation approaches, such as statistical deduction, regression, or deep learning, can fill in missing data based on existing values in the dataset.

Data resampling is often required to align data from different sources and extrapolate missing samples. Depending on the scope of the analysis, the temporal granularity of the data can be either decreased (i.e., down-sampling, to simplify the data while

keeping essential information) or increased (i.e., up-sampling, to capture finer details).

Filtering may also be required to isolate the bandwidth of interest of the signal from low-frequency noise (such as offsets and drifts), high-frequency noise (such as magnetic coupling), and interferences from other signals.

Depending on the data source, additional transformations may be necessary to improve data quality. For instance, sensors deployed to a patient in the community may not be worn in the exact orientation or placement as they were intended in the laboratory. Therefore, sensor calibration or baseline recordings can be used to “correct” the sensor signals with respect to a reference condition. Reporting transformations and other processing steps is important for *transparency*, allowing developers to generate replicable models based on the handling of the training data.

Data warping is an important concern during the cleaning and processing step. Although a primary goal is removing noise and enhancing the true target signal, an overly aggressive approach can obscure the authentic signal. This may reduce measurement resolution (e.g., aliasing), or prove impracticable with real-world datasets, ultimately compromising the performance and generalizability of an AI model.

Step 4. Feature extraction

Features are a set of values derived from processed data containing information related to the target output. This is a pivotal step to drive the *actionability* of the model, since these features can later be interpreted and acted upon.

Feature extraction varies in complexity, from extracting EHR fields to computing statistical moments from time-series signals (such as mean, standard, deviation, skewness, kurtosis from sensor or video data). Signal characteristics, including spatial, temporal, and frequency aspects, can also be engineered for clinical relevance. For example, wavelet transformations can estimate step length and stance time from IMU data [77,78] or fast Fourier transforms identify frequency bandwidths during muscle contraction in EMG signals [79]. EHR and annotation data offer additional contextual features [80], such as medications or assistive device use. For categorical or ordinal features, operations like one-hot encoding or ordinal encoding should be applied to prevent model bias [81].

Often, many extractable features are redundant (i.e., highly correlated) or irrelevant to the model. A feature set with high dimensionality increases the risk of overfitting the training

data [82]. Feature selection techniques, such as dimensionality reduction or regularization, should be considered to reduce the feature set to those more strongly associated with the target AI output. This is often done in conjunction with *Step 5. Model training and validation*. Ultimately, the complexity of feature extraction and engineering is strictly linked to the modeling technique, since not all models require predefined features before training (e.g., deep learning).

Step 5. Model training and validation

Processed data and/or their features find utility in descriptive, predictive, or prescriptive models. Descriptive models elucidate underlying data measures not readily available in clinic settings. In contrast, predictive and prescriptive models infer current or future patient outcomes and offer clinical decision support, respectively, within the clinical context.

Algorithm selection depends on the intended use case of the AI tool, and each algorithm's architecture and assumptions will affect data utilization (i.e., inductive bias [83]) and performance. For example, clustering algorithms group similar patients by measuring distance from decision boundaries [84]. Regression algorithms are used to estimate a function between input features and continuous outputs like time to hospital discharge or clinical score [85]. Classification algorithms identify discrete quantities, such as activities [48,72], or impairment and disease categories [38,40,86,87]. Deep learning models, including neural networks, directly learn patterns from data via reinforcement techniques [88]. Leading AI models can handle complex tasks, such as labeling unannotated data [89], generating new data [90], or interpreting unstructured text [91]. While the training process varies among different approaches, some common elements impact AI *interoperability*, *transparency*, and *actionability*.

Selecting appropriate validation techniques is crucial for reliable and robust AI development. Cross-validation (CV) is a process to optimize hyperparameters, scale data, select features, and evaluate model performance. In CV, a model is trained on one set of input data and evaluated on separate ("held out") testing or validation datasets. Various methods are available to separate the training, testing, and validation datasets during AI development, including a ratio-based train-test-validation split, leave-one-out, k-fold, or Monte Carlo sampling [60]. For models intended to generalize to new patients, CV should be subject-wise, meaning that the training, testing, and validation datasets contain data from different patients, with no leakage

between them [92]. For personalized models predicting outcomes for a single patient, these datasets might contain information from the same patient but recorded at different times. Improper CV can produce overly optimistic or even completely invalid models [92].

Postdevelopment, model evaluation should extend beyond traditional performance metrics like accuracy, F-scores, absolute error, or regression coefficients. Systematic analyses should examine performance variations when training or testing across different data or patient subgroups [60], as well as the consistency and importance of the features selected across iterations. These sensitivity analyses aid in understanding model stability, considering clinical use cases, and evaluating potential biases and overfitting.

Step 6. Streamlining

Comprehensive data and complex transformations may be needed to achieve the highest AI performance; however, practical considerations such as computational cost, processing time, data recording duration, data storage limits, and user burden can limit the real-world usability of AI models. In these cases, streamlining is an important step to determine the minimal equivalent data that are necessary to attain sufficient performance (Fig. 3). Example methods of streamlining are to reduce the number of data sources, reduce the data sampling frequency, or reduce the complexity of features used for the model. The goal of streamlining is to simplify the model to enhance its *interoperability* for scalable, real-world deployment.

Step 7. External validation of the target output

A sufficiently regularized model should generalize well to unseen data. External validation determines the reproducibility of the AI's performance when applied to an entirely new dataset, such as a different cohort, location, or later time point [93]. Successful external validation increases confidence in the model's performance and enhances its credibility for practical use in the desired context. At this stage, user feedback can also help developers understand the model's deployment feasibility in new clinical settings, as well as identify the frequency and consequences of potential errors.

CHALLENGES FOR FUTURE WORK

By combining large quantities of multimodal data, AI tools offer an exciting possibility to transform rehabilitation medicine

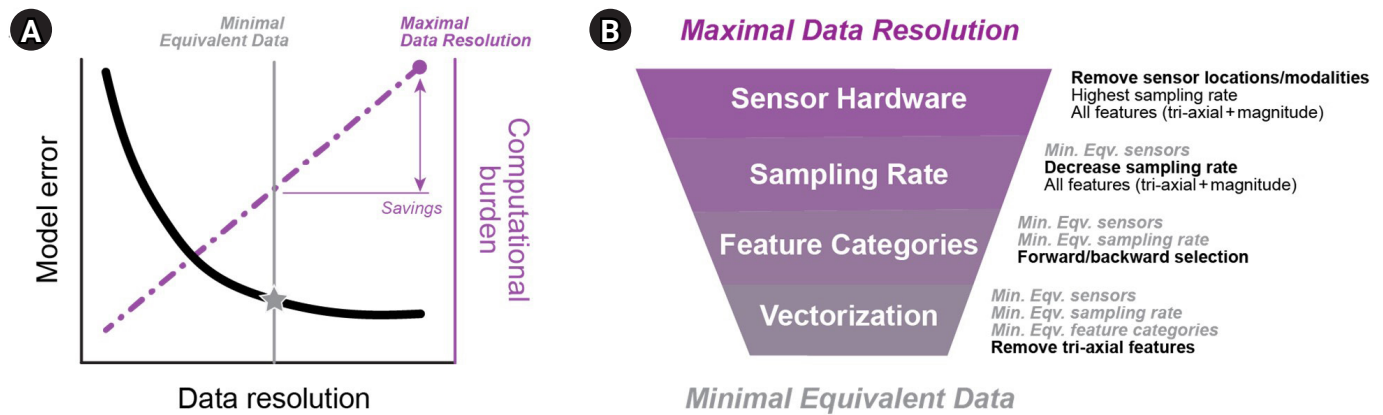


Fig. 3. Model streamlining. (A) Hypothetical illustration of the minimal equivalent data needed for a model. When paired with appropriate AI (artificial intelligence) practices, increasing the data resolution (e.g., adding data sources, increasing measurement frequency, increasing data complexity) can decrease model error (black curve) at the expense of greater computational burden (purple line). The minimal equivalent dataset (grey star) reduces the data resolution without substantial increases in error, thereby streamlining the model for more practical real-world deployment. (B) Example of a streamlining process for wearable sensor data. Descriptions on the right include data input during model training and testing, and the stepwise methods of streamlining in bold. Min. Eqv., minimal equivalent.

from a one-fits-all paradigm to personalized, precision treatments for individual patients. However, small and potentially biased datasets, as well as difficult-to-interpret models, may impede AI adoption at scale. Here, we offer some considerations for future work to address these ongoing challenges.

Large datasets are not always feasible

The quantity and quality of model training data are fundamental considerations when creating scalable, accurate AI tools for rehabilitation. Ideally, these training data would be recorded from a massive, fully representative patient cohort to capture the complete range of disease and recovery conditions that can arise for the model's target population. However, the practical challenges with collecting such datasets and the lack of data standardization mean that large, interoperable datasets are scarce in rehabilitation. Although data from large, multisite clinical trials and open-source repositories are beginning to address this gap, AI itself offers possible solutions to lengthy and costly data collection protocols.

Transfer learning is one such technique to harness knowledge from a previous AI model for a new application. Transfer learning involves adapting a model initially trained on one task or dataset to a different but related task or dataset. Importantly, this technique can drastically reduce the amount of data and computational resources needed for model training or retraining. For instance, transfer learning has been applied to classify

lower-limb movements from EMG data using a previous model that predicted joint angle [94].

Annotating datasets with precise labels is time-intensive. Self-supervised learning (SSL) addresses this issue by automatically generating annotations from unlabeled data. SSL is valuable for unstructured, large datasets where defining annotations is challenging or impractical. SSL uses contrastive learning to compare similar and dissimilar samples, identifying features for sample description and classification [89]. Recent studies show SSL models perform on par with manually annotated data, especially with multimodal data [89], and they can be generalized across external factors [95]. However, SSL-generated labels may not always be accurate, requiring continuous external supervision and comparison to exemplary annotations.

AI can also generate synthetic data to build larger, more diverse, and more representative datasets for algorithm training. Synthetic data mitigates the challenges of collecting sensitive patient information or handling imbalanced and potentially biased datasets [96]. Synthetic data can also be utilized as external datasets to further test and validate AI tools before clinical deployment [97]. However, evaluating synthetic data quality is difficult. Some state-of-the-art algorithms such as generative adversarial networks employ a generator module to produce synthetic data and a discriminator to compare it against real-world data [90], facilitating authenticity controls by developers.

AI insights are not always interpretable

To create interpretable AI tools, data in the AI workflow should be both intuitive and relevant to the intended audience (Fig. 4). Intuitive data are easily interpreted within a clinical context and offer actionable insights, while relevant data transparently align with the question of interest. These characteristics are not mutually exclusive and can be considered on separate scales—ranging from intuitive to unintuitive, and from relevant to irrelevant [98]. Intuitive and relevant data development often accompanies *Step 4. Feature extraction* in the proposed framework.

Highly relevant but unintuitive data may be less actionable for clinicians during treatment (quadrant IV, yellow). For example, if a model indicates that the “second moment of the sample

entropy” from an accelerometer during walking predicts the risk of knee injury, this may not give a clear, actionable insight for clinicians on reducing injury risks. Conversely, highly intuitive but irrelevant data may not provide the necessary information for the target AI output, rendering it even less meaningful (quadrant II, red). For example, precise measurements of skin hydration would not be useful to predict risk factors of cardiovascular disease, if there is no link between these two factors.

Complex data interpretation and novel treatment design can be supported by generative AI with large language models (LLMs) [99], similar to ChatGPT. LLMs trained on extensive EHR data show promise to reduce the administrative workload on clinicians, allowing them to generate automated reports or query specific patient information [91]. However, the safe and effective use of LLMs depends heavily on the quality of training data. Currently, these systems can still mislead users with outputs that are factually inaccurate and inconsistent with the input text [100]. Insights from experienced clinicians are vital to review and interpret AI output [101]. Developers and users should continuously assess AI models for factuality, comprehension, reasoning (i.e., by asking the model to show its reasoning process), possible harm, and bias [102].

Continued AI development after deployment

Evolving AI technologies, enhanced data infrastructure, and pervasive monitoring systems will gradually transform rehabilitation medicine, and more generally, healthcare. What are the challenges for researchers working in the field once AI tools are deployed?

Patients have multiple morbidities and a broad range of impairments that can complicate model predictions, especially for models trained under the assumption of a single and well-defined disease. As discussed, training models for every patient scenario is impractical, leading to inevitable errors. Until rigorous validation and trivial error rates are achieved, AI tools are best considered clinical support tools rather than autonomous agents, with clinicians playing the ultimate role in decision-making [103]. Consequently, we envision that there will be an increasing demand for AI implementation and usability studies in the near future. These studies may examine different user dashboards to convey AI output to clinicians, or the impact of AI tools on actual clinical practice and patient outcomes. There will also need to be additional regulations, such as the European AI Act, to define standards for a transparent, safe, secure, non-discriminatory AI-tools use before widespread

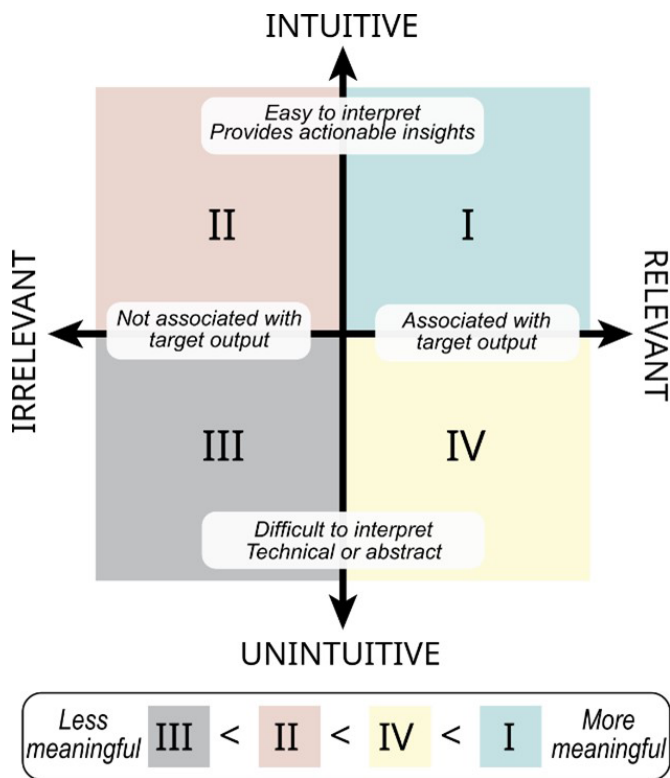


Fig. 4. Data characterization for interpretable artificial intelligence (AI). The most meaningful data for interpretable AI models are both intuitive and relevant in clinical care (quadrant I, green). Data can be highly relevant to model performance but less intuitive, thus rendering it less actionable to a clinician during treatment (quadrant IV, yellow). Other data are highly intuitive and understandable to a treating clinician but less relevant, providing little-to-no value to the model (quadrant II, red). Data that are unintuitive and irrelevant should not be included in a model (quadrant III, grey).

adoption.

AI models should regularly integrate new training data to improve their representation and performance. Integrating new data can also reduce the likelihood of data drift, which occurs when new data introduced to an AI model differs from the data used for initial training. Data drift can arise from gradual or sudden changes in data acquisition methods, clinical treatments, disease patterns, or patient characteristics [104]. Therefore, it is crucial to monitor model performance and make adjustments even after AI deployment.

CONCLUSION

Rehabilitation medicine can benefit from recent advances in new data sources and modeling techniques to transition towards customized, precise, and predictive approaches. AI can help extract meaningful clinical insights from a wealth of healthcare data, but many challenges related to the development and interpretation of these tools can limit their success in real-world settings.

We proposed a general framework to build interoperable, transparent, and actionable AI tools for rehabilitation. In this framework, training data are configured and acquired in a manner that captures the use case of the intended AI tool, with a systematic approach to validation and interpretation for the patient group(s) of interest. Decisions should be made in consultation with expert clinicians who understand the pathophysiology of the impairment or condition being studied, and who can advise on potential confounds during real-world clinical or community scenarios. AI tools that consider these factors have great potential for automatically computing measures of activity and performance, illuminating novel biomarkers of injury and disease, and predicting patient outcomes using multidimensional factors related to health.

CONFLICTS OF INTEREST

No potential conflict of interest relevant to this article was reported.

FUNDING INFORMATION

This work was supported by the National Institute on Disability, Independent Living, and Rehabilitation Research (90REG E0010).

AUTHOR CONTRIBUTION

Conceptualization: all authors. Funding acquisition: Jayaraman A. Visualization: Lanotte F, O'Brien MK. Writing – original draft: Lanotte F, O'Brien MK. Writing – review and editing: all authors. Approval of the final manuscript: all authors.

ORCID

Francesco Lanotte, <https://orcid.org/0000-0001-8978-1074>

Megan K. O'Brien, <https://orcid.org/0000-0001-8069-365X>

Arun Jayaraman, <https://orcid.org/0000-0002-9302-6693>

REFERENCES

1. Korteling JEH, van de Boer-Visschedijk GC, Blankendaal RAM, Boonekamp RC, Eikelboom AR. Human- versus Artificial Intelligence. *Front Artif Intell* 2021;4:622364.
2. Yu KH, Beam AL, Kohane IS. Artificial intelligence in healthcare. *Nat Biomed Eng* 2018;2:719-31.
3. van der Schaar M, Zame W. Machine learning for individualised medicine. In: Pearson-Stuttard J, Murphy O, editor. Annual report of the chief medical officer, 2018. Health 2040 – better health within reach. Department of Health and Social Care; 2018.
4. French MA, Roemmich RT, Daley K, Beier M, Penttinen S, Ragha-van P, et al. Precision rehabilitation: optimizing function, adding value to health care. *Arch Phys Med Rehabil* 2022;103:1233-9.
5. Agrawal R, Prabakaran S. Big data in digital healthcare: lessons learnt and recommendations for general practice. *Heredity (Edinb)* 2020;124:525-34.
6. Reda R, Piccinini F, Carbonaro A. Towards consistent data representation in the IoT healthcare landscape. Paper presented at: DH '18: Proceedings of the 2018 International Conference on Digital Health; 2018 Apr 23-26; Lyon, France. p. 5-10.
7. Bellazzi R, Zupan B. Predictive data mining in clinical medicine: current issues and guidelines. *Int J Med Inform* 2008;77:81-97.
8. Deo RC. Machine learning in medicine. *Circulation* 2015;132:1920-30.
9. Landi I, Glicksberg BS, Lee HC, Cherng S, Landi G, Danieleto M, et al. Deep representation learning of electronic health records to unlock patient stratification at scale. *NPJ Digit Med* 2020;3:96.
10. Jensen PB, Jensen LJ, Brunak S. Mining electronic health records: towards better research applications and clinical care. *Nat Rev Genet* 2012;13:395-405.
11. Xu S, Kim J, Walter JR, Ghaffari R, Rogers JA. Translational gaps and

- opportunities for medical wearables in digital health. *Sci Transl Med* 2022;14:eabn6036.
12. Brom H, Brooks Carthon JM, Ikeaba U, Chittams J. Leveraging electronic health records and machine learning to tailor nursing care for patients at high risk for readmissions. *J Nurs Care Qual* 2020;35:27-33.
 13. Delahanty RJ, Alvarez J, Flynn LM, Sherwin RL, Jones SS. Development and evaluation of a machine learning model for the early identification of patients at risk for sepsis. *Ann Emerg Med* 2019;73:334-44.
 14. Engelhard MM, Henao R, Berchuck SI, Chen J, Eichner B, Herkert D, et al. Predictive value of early autism detection models based on electronic health record data collected before age 1 year. *JAMA Netw Open* 2023;6:e2254303.
 15. Flores AM, Demsas F, Leeper NJ, Ross EG. Leveraging machine learning and artificial intelligence to improve peripheral artery disease detection, treatment, and outcomes. *Circ Res* 2021;128:1833-50.
 16. Ye C, Li J, Hao S, Liu M, Jin H, Zheng L, et al. Identification of elders at higher risk for fall with statewide electronic health records and a machine learning algorithm. *Int J Med Inform* 2020;137:104105.
 17. Cramer EM, Seneviratne MG, Sharifi H, Ozturk A, Hernandez-Boussard T. Predicting the incidence of pressure ulcers in the intensive care unit using machine learning. *EGEMS (Wash DC)* 2019;7:49.
 18. Kucukboyaci NE, Long C, Smith M, Rath JF, Bushnik T. Cluster analysis of vulnerable groups in acute traumatic brain injury rehabilitation. *Arch Phys Med Rehabil* 2018;99:2365-9.
 19. Custer MG, Huebner RA. Identifying homogeneous outcome groups in adult rehabilitation using cluster analysis. *Am J Occup Ther* 2019;73:7305205050p1-9.
 20. Bland MD, Sturmoski A, Whitson M, Connor LT, Fucetola R, Huskey T, et al. Prediction of discharge walking ability from initial assessment in a stroke inpatient rehabilitation facility population. *Arch Phys Med Rehabil* 2012;93:1441-7.
 21. Harari Y, O'Brien MK, Lieber RL, Jayaraman A. Inpatient stroke rehabilitation: prediction of clinical outcomes using a machine-learning approach. *J Neuroeng Rehabil* 2020;17:71.
 22. Henderson CE, Fahey M, Brazg G, Moore JL, Hornby TG. Predicting discharge walking function with high-intensity stepping training during inpatient rehabilitation in nonambulatory patients poststroke. *Arch Phys Med Rehabil* 2022;103(7S):S189-96.
 23. Scrutinio D, Lanzillo B, Guida P, Mastropasqua F, Monitillo V, Pusineri M, et al. Development and validation of a predictive model for functional outcome after stroke rehabilitation: the Maugeri model. *Stroke* 2017;48:3308-15.
 24. Stinear CM, Smith MC, Byblow WD. Prediction tools for stroke rehabilitation. *Stroke* 2019;50:3314-22.
 25. Smith MC, Barber AP, Scrivener BJ, Stinear CM. The TWIST tool predicts when patients will recover independent walking after stroke: an observational study. *Neurorehabil Neural Repair* 2022;36:461-71.
 26. Lundquist CB, Nielsen JF, Arguissain FG, Brunner IC. Accuracy of the upper limb prediction algorithm PREP2 applied 2 weeks post-stroke: a prospective longitudinal study. *Neurorehabil Neural Repair* 2021;35:68-78.
 27. Ehteshami Bejnordi B, Veta M, Johannes van Diest P, van Ginneken B, Karssemeijer N, Litjens G, et al. Diagnostic assessment of deep learning algorithms for detection of lymph node metastases in women with breast cancer. *JAMA* 2017;318:2199-210.
 28. Litjens G, Sánchez CI, Timofeeva N, Hermesen M, Nagtegaal I, Kovacs I, et al. Deep learning as a tool for increased accuracy and efficiency of histopathological diagnosis. *Sci Rep* 2016;6:26286.
 29. Poplin R, Varadarajan AV, Blumer K, Liu Y, McConnell MV, Corrado GS, et al. Prediction of cardiovascular risk factors from retinal fundus photographs via deep learning. *Nat Biomed Eng* 2018;2:158-64.
 30. Panwar N, Huang P, Lee J, Keane PA, Chuan TS, Richhariya A, et al. Fundus photography in the 21st century--a review of recent technological advances and their implications for worldwide healthcare. *Telemed J E Health* 2016;22:198-208.
 31. Solana-Lavalle G, Rosas-Romero R. Classification of PPMI MRI scans with voxel-based morphometry and machine learning to assist in the diagnosis of Parkinson's disease. *Comput Methods Programs Biomed* 2021;198:105793.
 32. Liew SL, Schweighofer N, Cole JH, Zavaliangos-Petropulu A, Lo BP, Han LKM, et al. Association of brain age, lesion volume, and functional outcome in patients with stroke. *Neurology* 2023;100:e2103-13.
 33. Warmerdam E, Hausdorff JM, Atrsaei A, Zhou Y, Mirelman A, Aminian K, et al. Long-term unsupervised mobility assessment in movement disorders. *Lancet Neurol* 2020;19:462-70.
 34. Adans-Dester CP, Lang CE, Reinkensmeyer DJ, Bonato P. Wearable sensors for stroke rehabilitation. In: Reinkensmeyer DJ, Marchal-Crespo L, Dietz V, editor. *Neurorehabilitation technology*. Springer; 2022. p.467-507.
 35. O'Brien MK, Shin SY, Khazanchi R, Fanton M, Lieber RL, Ghaffari R, et al. Wearable sensors improve prediction of post-stroke walking function following inpatient rehabilitation. *IEEE J Transl Eng Health Med* 2022;10:2100711.
 36. Liuzzi P, Carpinella I, Anastasi D, Gervasoni E, Lencioni T, Bertoni R, et al. Machine learning based estimation of dynamic balance and

- gait adaptability in persons with neurological diseases using inertial sensors. *Sci Rep* 2023;13:8640.
37. Adans-Dester C, Hankov N, O'Brien A, Vergara-Diaz G, Black-Schaffer R, Zafonte R, et al. Enabling precision rehabilitation interventions using wearable sensors and machine learning to track motor recovery. *NPJ Digit Med* 2020;3:121.
38. Lonini L, Dai A, Shawen N, Simuni T, Poon C, Shimanovich L, et al. Wearable sensors for Parkinson's disease: which data are worth collecting for training symptom detection models. *NPJ Digit Med* 2018;1:64.
39. Shawen N, O'Brien MK, Venkatesan S, Lonini L, Simuni T, Hamilton JL, et al. Role of data measurement characteristics in the accurate detection of Parkinson's disease symptoms using wearable sensors. *J Neuroeng Rehabil* 2020;17:52.
40. O'Brien MK, Botonis OK, Larkin E, Carpenter J, Martin-Harris B, Maronati R, et al. Advanced machine learning tools to monitor biomarkers of dysphagia: a wearable sensor proof-of-concept study. *Digit Biomark* 2021;5:167-75.
41. Hafer JF, Vitali R, Gurchiek R, Curtze C, Shull P, Cain SM. Challenges and advances in the use of wearable sensors for lower extremity biomechanics. *J Biomech* 2023;157:111714.
42. Albert MV, Deeny S, McCarthy C, Valentin J, Jayaraman A. Monitoring daily function in persons with transfemoral amputations using a commercial activity monitor: a feasibility study. *PM R* 2014;6:1120-7.
43. Kim J, Colabianchi N, Wensman J, Gates DH. Wearable sensors quantify mobility in people with lower limb amputation during daily life. *IEEE Trans Neural Syst Rehabil Eng* 2020;28:1282-91.
44. Moshawrab M, Adda M, Bouzouane A, Ibrahim H, Raad A. Smart wearables for the detection of cardiovascular diseases: a systematic literature review. *Sensors (Basel)* 2023;23:828.
45. Boe AJ, McGee Koch LL, O'Brien MK, Shawen N, Rogers JA, Lieber RL, et al. Automating sleep stage classification using wireless, wearable sensors. *NPJ Digit Med* 2019;2:131.
46. Hsieh KL, Chen L, Sosnoff JJ. Mobile technology for falls prevention in older adults. *J Gerontol A Biol Sci Med Sci* 2023;78:861-8.
47. Ghomrawi HMK, O'Brien MK, Carter M, Macaluso R, Khazanchi R, Fanton M, et al. Applying machine learning to consumer wearable data for the early detection of complications after pediatric appendectomy. *NPJ Digit Med* 2023;6:148.
48. O'Brien MK, Shawen N, Mummidisetty CK, Kaur S, Bo X, Poellabauer C, et al. Activity recognition for persons with stroke using mobile phone technology: toward improved performance in a home setting. *J Med Internet Res* 2017;19:e184.
49. Adib F, Mao H, Kabelac Z, Katabi D, Miller RC. Smart homes that monitor breathing and heart rate. Paper presented at: CHI '15: Proceedings of the 33rd Annual ACM Conference on Human Factors in Computing Systems; 2015 Apr 18-23; Seoul, Korea. p. 837-46.
50. Zhang G, Vahia IV, Liu Y, Yang Y, May R, Cray HV, et al. Contactless in-home monitoring of the long-term respiratory and behavioral phenotypes in older adults with COVID-19: a case series. *Front Psychiatry* 2021;12:754169.
51. Yang Y, Yuan Y, Zhang G, Wang H, Chen YC, Liu Y, et al. Artificial intelligence-enabled detection and assessment of Parkinson's disease using nocturnal breathing signals. *Nat Med* 2022;28:2207-15.
52. Haque A, Milstein A, Fei-Fei L. Illuminating the dark spaces of healthcare with ambient intelligence. *Nature* 2020;585:193-202.
53. Kim WS, Cho S, Baek D, Bang H, Paik NJ. Upper extremity functional evaluation by Fugl-Meyer Assessment scoring using depth-sensing camera in hemiplegic stroke patients. *PLoS One* 2016;11:e0158640.
54. Stenum J, Rossi C, Roemmich RT. Two-dimensional video-based analysis of human gait using pose estimation. *PLoS Comput Biol* 2021;17:e1008935.
55. Sato K, Nagashima Y, Mano T, Iwata A, Toda T. Quantifying normal and parkinsonian gait features from home movies: practical application of a deep learning-based 2D pose estimator. *PLoS One* 2019;14:e0223549.
56. Lonini L, Moon Y, Embry K, Cotton RJ, McKenzie K, Jenz S, et al. Video-based pose estimation for gait analysis in stroke survivors during clinical assessments: a proof-of-concept study. *Digit Biomark* 2022;6:9-18.
57. Lam WWT, Tang YM, Fong KNK. A systematic review of the applications of markerless motion capture (MMC) technology for clinical measurement in rehabilitation. *J Neuroeng Rehabil* 2023;20:57.
58. Sibley KG, Girges C, Hoque E, Foltynie T. Video-based analyses of Parkinson's disease severity: a brief review. *J Parkinsons Dis* 2021;11(S1):S83-93.
59. Adde L, Helbostad JL, Jensenius AR, Taraldsen G, Grunewaldt KH, Støen R. Early prediction of cerebral palsy by computer-based video analysis of general movements: a feasibility study. *Dev Med Child Neurol* 2010;52:773-8.
60. Chen PC, Liu Y, Peng L. How to develop machine learning models for healthcare. *Nat Mater* 2019;18:410-4.
61. Alaa AM, van der Schaar M. Demystifying black-box models with symbolic metamodells. Paper presented at: Advances in Neural Information Processing Systems 32 (NeurIPS 2019); 2019 Dec 8-14; Vancouver, Canada. p. 32.
62. Ehrmann DE, Joshi S, Goodfellow SD, Mazwi ML, Eytan D. Making machine learning matter to clinicians: model actionability in medical decision-making. *NPJ Digit Med* 2023;6:7.

63. Parimbelli E, Buonocore TM, Nicora G, Michalowski W, Wilk S, Bellazzi R. Why did AI get this one wrong? - tree-based explanations of machine learning model predictions. *Artif Intell Med* 2023;135:102471.
64. Busch Tde A, Duarte YA, Pires Nunes D, Lebrão ML, Satya Naslavsky M, dos Santos Rodrigues A, et al. Factors associated with lower gait speed among the elderly living in a developing country: a cross-sectional population-based study. *BMC Geriatr* 2015;15:35.
65. Pacini Panebianco G, Bisi MC, Stagni R, Fantozzi S. Analysis of the performance of 17 algorithms from a systematic review: influence of sensor position, analysed variable and computational approach in gait timing estimation from IMU measurements. *Gait Posture* 2018;66:76-82.
66. Atallah L, Lo B, King R, Yang GZ. Sensor positioning for activity recognition using wearable accelerometers. *IEEE Trans Biomed Circuits Syst* 2011;5:320-9.
67. Rantz M, Phillips LJ, Galambos C, Lane K, Alexander GL, Despins L, et al. Randomized trial of intelligent sensor system for early illness alerts in senior housing. *J Am Med Dir Assoc* 2017;18:860-70.
68. Luo Z, Hsieh JT, Balachandar N, Yeung S, Pusiol G, Luxenberg J, et al. Computer vision-based descriptive analytics of seniors' daily activities for long-term health monitoring. *Proc Mach Learn Res* 2018;85:1-18.
69. Gold R, Reichman M, Greenberg E, Ivanidze J, Elias E, Tsiouris AJ, et al. Developing a new reference standard: is validation necessary? *Acad Radiol* 2010;17:1079-82.
70. Fusca M, Negrini F, Perego P, Magoni L, Molteni F, Andreoni G. Validation of a wearable IMU system for gait analysis: protocol and application to a new system. *Appl Sci* 2018;8:1167.
71. van Lier HG, Pieterse ME, Garde A, Postel MG, de Haan HA, Vollebrouk-Hutten MMR, et al. A standardized validity assessment protocol for physiological signals from wearable technology: methodological underpinnings and an application to the E4 biosensor. *Behav Res Methods* 2020;52:607-29.
72. Lonini L, Gupta A, Deems-Dluhy S, Hoppe-Ludwig S, Kording K, Jayaraman A. Activity recognition in individuals walking with assistive devices: the benefits of device-specific models. *JMIR Rehabil Assist Technol* 2017;4:e8.
73. Lanotte F, Shin SY, O'Brien MK, Jayaraman A. Validity and reliability of a commercial wearable sensor system for measuring spatiotemporal gait parameters in a post-stroke population: the effects of walking speed and asymmetry. *Physiol Meas* 2023;44:085005.
74. Williams DR, Mohammed SA, Leavell J, Collins C. Race, socioeconomic status, and health: complexities, ongoing challenges, and research opportunities. *Ann N Y Acad Sci* 2010;1186:69-101.
75. Prates MOR, Avelar PHC, Lamb L. Assessing gender bias in machine translation: a case study with Google Translate. *Neural Comput Appl* 2020;32:6363-81.
76. Kohane IS, Aronow BJ, Avillach P, Beaulieu-Jones BK, Bellazzi R, Bradford RL, et al. What every reader should know about studies using electronic health record data but may be afraid to ask. *J Med Internet Res* 2021;23:e22219.
77. Zijlstra W, Hof AL. Assessment of spatio-temporal gait parameters from trunk accelerations during human walking. *Gait Posture* 2003;18:1-10.
78. Jasiewicz JM, Allum JH, Middleton JW, Barriskill A, Condie P, Purcell B, et al. Gait event detection using linear accelerometers or angular velocity transducers in able-bodied and spinal-cord injured individuals. *Gait Posture* 2006;24:502-9.
79. Nazmi N, Abdul Rahman MA, Yamamoto S, Ahmad SA, Zamzuri H, Mazlan SA. A review of classification techniques of EMG signals during isotonic and isometric contractions. *Sensors (Basel)* 2016;16:1304.
80. Herrero JG, Patricio MA, Molina JM, Cardoso LA. Contextual and human factors in information fusion. *IOS Press Ebooks*; 2010. p.79-92.
81. Yu L, Zhou R, Chen R, Lai KK. Missing data preprocessing in credit classification: one-hot encoding or imputation? *Emerg Markets Financ Trade* 2020;58:472-82.
82. Guyon I, Elisseeff A. An introduction to variable and feature selection. *J Mach Learn Res* 2003;3:1157-82.
83. Baxter J. A model of inductive bias learning. *J Artif Intell Res* 2000;12:149-98.
84. Vranas KC, Jopling JK, Sweeney TE, Ramsey MC, Milstein AS, Slatore CG, et al. Identifying distinct subgroups of ICU patients: a machine learning approach. *Crit Care Med* 2017;45:1607-15.
85. Turgeman L, May JH, Sciulli R. Insights from a machine learning model for predicting the hospital Length of Stay (LOS) at the time of admission. *Expert Syst Appl* 2017;78:376-85.
86. Lonini L, Shawen N, Ghaffari R, Rogers J, Jayaraman A. Automatic detection of spasticity from flexible wearable sensors. Paper presented at: UbiComp '17: Proceedings of the 2017 ACM International Joint Conference on Pervasive and Ubiquitous Computing and Proceedings of the 2017 ACM International Symposium on Wearable Computers; 2017 Sep 11-15; Maui, Hawaii. p. 133-6.
87. Lonini L, Shawen N, Botonis O, Fanton M, Jayaraman C, Mummidisetti CK, et al. Rapid screening of physiological changes associated with COVID-19 using soft-wearables and structured activities: a pilot study. *IEEE J Transl Eng Health Med* 2021;9:4900311.
88. Bacciu D, Chessa S, Gallicchio C, Micheli A, Pedrelli L, Ferro E, et al. A learning system for automatic Berg Balance Scale score estimation.

- tion. *Eng Appl Artif Intell* 2017;66:60-74.
89. Krishnan R, Rajpurkar P, Topol EJ. Self-supervised learning in medicine and healthcare. *Nat Biomed Eng* 2022;6:1346-52.
90. Yoon J, Drumright LN, van der Schaar M. Anonymization through data synthesis using generative adversarial networks (ADS-GAN). *IEEE J Biomed Health Inform* 2020;24:2378-88.
91. Yang X, Chen A, PourNejatian N, Shin HC, Smith KE, Parisien C, et al. A large language model for electronic health records. *NPJ Digit Med* 2022;5:194.
92. Saeb S, Lonini L, Jayaraman A, Mohr DC, Kording KP. The need to approximate the use-case in clinical machine learning. *Gigascience* 2017;6:1-9.
93. Ramspek CL, Jager KJ, Dekker FW, Zoccali C, van Diepen M. External validation of prognostic models: what, why, how, when and where? *Clin Kidney J* 2020;14:49-58.
94. Gautam A, Panwar M, Biswas D, Acharyya A. MyoNet: a transfer-learning-based LRCN for lower limb movement recognition and knee joint angle prediction for remote monitoring of rehabilitation progress from sEMG. *IEEE J Transl Eng Health Med* 2020;8:2100310.
95. Du C, Graham S, Depp C, Nguyen T. Assessing physical rehabilitation exercises using graph convolutional network with self-supervised regularization. *Annu Int Conf IEEE Eng Med Biol Soc* 2021;2021:281-5.
96. Chen RJ, Lu MY, Chen TY, Williamson DFK, Mahmood F. Synthetic data in machine learning for medicine and healthcare. *Nat Biomed Eng* 2021;5:493-7.
97. Che Z, Cheng Y, Zhai S, Sun Z, Liu Y. Boosting deep learning risk prediction with generative adversarial networks for electronic health records. Paper presented at: 2017 IEEE International Conference on Data Mining (ICDM); 2017 Nov 18-21; New Orleans, USA. p. 787-92.
98. Shen C, Wang Z, Villar SS, Van Der Schaar M. Learning for dose allocation in adaptive clinical trials with safety constraints. Paper presented at: ICML'20: Proceedings of the 37th International Conference on Machine Learning; 2020 Jul 13-18; Virtual Event. p. 8730-40.
99. Arora A, Arora A. The promise of large language models in health care. *Lancet* 2023;401:641.
100. Murphy C, Thomas FP. Generative AI in spinal cord injury research and care: opportunities and challenges ahead. *J Spinal Cord Med* 2023;46:341-2.
101. Thirunavukarasu AJ. Large language models will not replace healthcare professionals: curbing popular fears and hype. *J R Soc Med* 2023;116:181-2.
102. Singhal K, Azizi S, Tu T, Mahdavi SS, Wei J, Chung HW, et al. Large language models encode clinical knowledge. *Nature* 2023;620:172-80. Erratum in: *Nature* 2023;620:E19.
103. Vasey B, Nagendran M, Campbell B, Clifton DA, Collins GS, Denaxas S, et al. Reporting guideline for the early-stage clinical evaluation of decision support systems driven by artificial intelligence: DECIDE-AI. *Nat Med* 2022;28:924-33. Erratum in: *Nat Med* 2022;28:2218.
104. Sahiner B, Chen W, Samala RK, Petrick N. Data drift in medical machine learning: implications and potential remedies. *Br J Radiol* 2023;96:20220878.