Accuracy of machine learning algorithms for the assessment of upper-limb motor impairments in patients with post-stroke hemiparesis: A systematic review and meta-analysis

Jorge Fernando Ambros-Antemate[1,A,B,D,F], Adriana Reyes-Flores[1,B,D,F], Liliana Argueta-Figueroa[2,A,D–F], Rafael Ramírez-Ramírez[1,B,D,F], Marciano Vargas-Treviño[3,C,E,F], Jaime Gutiérrez-Gutiérrez[3,C,E,F], Eduardo Pérez-Campos Mayoral[4,B,E,F], Eduardo Pérez-Campos[4,B,E,F], Luis Angel Flores-Mejía[5,C,E,F], Rafael Torres-Rosas[6,A,D–F]

[1] School of Medicine and Surgery, Benito Juárez Autonomous University of Oaxaca, Mexico

[2] CONACyT- School of Dentistry, Benito Juárez Autonomous University of Oaxaca, Mexico

[3] School of Biological Systems and Technological Innovation, Benito Juárez Autonomous University of Oaxaca, Mexico

[4] Center for Research UNAM-UABJO, Benito Juárez Autonomous University of Oaxaca, Mexico

[5] School of Public Health, Imperial College, London, United Kingdom

[6] Immunology Laboratory, Cencer for Health and Diseases Sciences Research, School of Dentistry, Benito Juárez Autonomous University of Oaxaca, Mexico

A – research concept and design; B – collection and/or assembly of data; C – data analysis and interpretation;
D – writing the article; E – critical revision of the article; F – final approval of the article

Abstract

Background. The assessment of motor function is vital in post-stroke rehabilitation protocols, and it is imperative to obtain an objective and quantitative measurement of motor function. There are some innovative machine learning algorithms that can be applied in order to automate the assessment of upper extremity motor function.

Objectives. To perform a systematic review and meta-analysis of the efficacy of machine learning algorithms for assessing upper limb motor function in post-stroke patients and compare these algorithms to clinical assessment.

Materials and methods. The protocol was registered in the International Prospective Register of Systematic Reviews (PROSPERO) database. The review was carried out according to the Preferred Reporting Items for Systematic Reviews and Meta-Analyses (PRISMA) guidelines and the Cochrane Handbook for Systematic Reviews of Interventions. The search was performed using 6 electronic databases. The meta-analysis was performed with the data from the correlation coefficients using a random model.

Results. The initial search yielded 1626 records, but only 8 studies fully met the eligibility criteria. The studies reported strong and very strong correlations between the algorithms tested and clinical assessment. The meta-analysis revealed a lack of homogeneity ($I^2 = 85.29\%$, $Q = 48.15$), which is attributable to the heterogeneity of the included studies.

Conclusions. Automated systems using machine learning algorithms could support therapists in assessing upper extremity motor function in post-stroke patients. However, to draw more robust conclusions, methodological designs that minimize the risk of bias and increase the quality of the methodology of future studies are required.

Key words: stroke, machine learning, computer-assisted diagnosis

Introduction

According to the World Health Organization (WHO), 15 million people worldwide suffer a stroke every year.[1] Of these, approx. 5 million are left with a disability that limits their capacity to perform daily activities. They are also prone to becoming depressed or stressed due to limitations of their motor functions.[2]

Because of these conditions, patients have to participate in rehabilitation programs aimed at improving their quality of life. These programs support them in regaining motor function in the areas affected by the stroke.[3] First, it is necessary to assess the degree of impairment to properly select the best therapeutic options.[4] There are numerous motor assessment tests to evaluate the degree of upper limb disability, including the Fugl–Meyer Assessment[5] and the Wolf Motor Function Test.[6] In general, each test consists of a series of tasks to be performed by the patient, and the therapist evaluates those tasks using measures based on their observations. However, motor assessments require prior training of the examiners; therefore, in many cases, the evaluation tends to be subjective.[7] To avoid this problem, there is a great interest in the development of automated systems aimed at achieving objective and quantitative assessments for rehabilitation after strokes. Automated quantitative assessment systems can be used with home-based systems that assist patients in evaluating improvements during home-based exercise programs.

Thanks to technological advances, significant progress has been made in recent years in measuring and analyzing vital signs and human movement through artificial intelligence (AI).[8–10] Furthermore, AI has provided a technical basis for the automation of many processes,[11] such as rehabilitation[12] and evaluation of upper limb motor function.

Objectives

Based on these points, the main objective of this study was to perform a systematic review and meta-analysis of the efficacy of machine learning algorithms in assessing upper limb motor function in post-stroke patients, and compare these algorithms to clinical assessment.

Materials and methods

Study protocol and record

The systematic review was conducted according to the Preferred Reporting Items for Systematic Reviews and Meta-Analyses (PRISMA) guidelines[13] and the Cochrane Handbook for Systematic Reviews of Interventions.[14] In addition, the review protocol was published in the International Prospective Register of Systematic Reviews (PROSPERO) with the registration number PROSPERO 2021 CRD42021257217 (https://www.crd.york.ac.uk/prospero/display_record.php?ID=CRD42021257217).

Eligibility criteria, information sources and search strategy

The articles included assessed upper limb motor function in post-stroke patients through machine learning algorithms compared to standard clinical assessment. The outcomes of interest were diagnostic accuracy, specificity, and/or sensitivity. Articles were excluded if they assessed motor function to predict patient recovery time; case series and literature reviews were also excluded. The patient-intervention-comparison-outcome (PICO) strategy was used to identify the key words used (Table 1). The electronic search was performed in May 2021 and updated in October 2021. The information sources and algorithms used in each database are shown in Table 1.

Selection process

Three authors (JFAA, ARF and RRR) independently reviewed the registries obtained by the search. Duplicate records were removed using Mendeley Desktop v. 1.19.8 Reference Manager (Elsevier, Amsterdam, the Netherlands).[15] Studies that met the eligibility criteria by reading the title and abstract were retrieved in full text. Any disagreement was addressed by another reviewer (LAF) who made the final decision. The selection process is summarized in the PRISMA flowchart (Fig. 1).

Data collection process and data items

The relevant data of the included articles were collected in a standardized Microsoft Excel 2019 spreadsheet (Microsoft Corp., Redmond, USA). The data included study design, characteristics of the population, type of machine learning algorithm, data acquisition device, reference test, relative sensitivity, relative specificity, and confidence intervals. Three reviewers were responsible for data extraction (MVT, JGG and LAFM). When there were disagreements, the reviewers held discussions until reaching a consensus. The researchers of the original articles were contacted via e-mail for missing or additional details.

Assessment of risk of bias and quality of the included studies

Three reviewers (EPCM, EPC and ARF) assessed the risk of bias of the included studies following Chapter 8 of the Cochrane Handbook for Systematic Reviews of Interventions.[14] Additionally, the reviewers performed a quality assessment of the studies using the modified QUADAS-2 tool (Table 2),[16] which encompasses the following 5 domains: sample selection, index test, reference standard, flow rate, and time.

Table 1. Patient, intervention, comparison, outcome (PICO) strategy and algorithms used for the systematic review

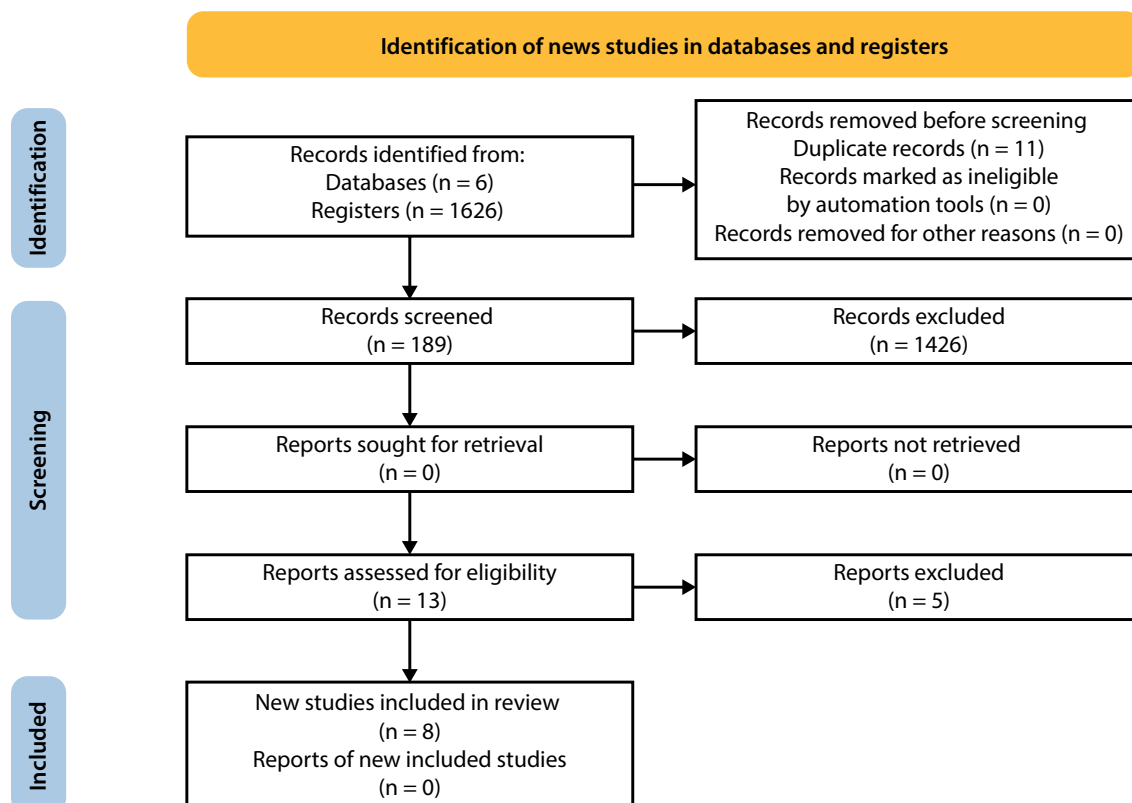| Population	Post-stroke patients with hemiparesis of the upper-limb motor impairments.
Intervention	Algorithms of machine learning.
Comparison	Clinical evaluation assessment: Fugl–Meyer assessment, Wolf Motor Function Test, Modified Ashworth Scale, Chedoke–McMaster Stroke, or Motor Assessment Scale.
Outcomes	Diagnostic: Accuracy, specificity or sensibility.
Study design	Randomized clinical trial, non-randomized clinical trial, case-control study, or cohort study.
Eligibility criteria	Studies published in English and Spanish.
Electronic databases	PubMed, IEEE Xplore, ScienceDirect, Taylor & Francis Online, Wiley Online Library, and Google Scholar.
Focused question	Is there any evidence for the application of machine learning algorithms in the assessment of upper-limb motor impairments in patients with post-stroke hemiparesis?
Number of registers found for each database	Algorithms used for search strategy adapted for each database
PubMed; October 16, 2020 32 records	("machine learning" OR "learning, machine" OR "transfer learning" OR "learning, transfer" OR "neural network" OR "deep learning" OR "knowledge bases" OR "hierarchical learning" OR "expert systems" OR "fuzzy logic" OR "computer vision" OR "artificial intelligence" OR "support vector machine") AND ("motor" OR "motor function" OR "activities, motor" OR "activity, motor" OR "motor activities") AND ("evaluation" OR "assessment" OR "quantify" OR "quantitative" OR "scoring") AND ("extremities, upper" OR "upper extremities" OR "membrum superius" OR "upper limb" OR "limb, upper" OR "limbs, upper" OR "upper limbs" OR "extremity, upper" OR "upper-limb") AND ("stroke" OR "cerebrovascular accident")
Google Scholar; October 16, 2020 1390 records	"machine learning" AND "motor function" AND ("evaluation" OR "assessment" OR "quantitative") AND "upper limb" AND "stroke"
IEEE Xplore; December 1, 2020 4 records	"machine learning" AND "motor function" AND ("evaluation" OR "assessment" OR "quantitative") AND "upper limb" AND "stroke"
ScienceDirect; December 15, 2020 39 records	"machine learning" AND "motor function" AND ("evaluation" OR "assessment" OR "quantitative") AND "upper limb" AND "stroke"
Taylor & Francis Online; December 18, 2020 29 records	"machine learning" AND "motor function" AND ("evaluation" OR "assessment" OR "quantitative") AND "upper limb" AND "stroke"
Wiley Online Library; December 22, 2020 132 records	"machine learning" AND "motor function" AND ("evaluation" OR "assessment" OR "quantitative") AND "upper limb" AND "stroke"

Fig. 1. Preferred Reporting Items for Systematic Reviews and Meta-Analyses (PRISMA) flow diagram of the selection process of the studies included in the systematic review

Table 2. QUADAS-2 modification

Domain name	Questions
Risk of bias	
1. Patient selection	1. Was a random or a consecutive sample of patients enrolled? 2. Was a case-control design avoided? 3. Were the exclusions made avoided?
2. Index test	1. Is there a sample calculation and the minimum necessary number of patients used?* 2. Was it previously specified if a threshold was used to define the positivity or negativity of the index test?
3. Benchmark	1. Is the reference test likely to correctly assess the target condition? 2. Were the reference test results interpreted without knowledge of the index test results?
4. Flow and Times	1. Was there an appropriate interval between the index test and the reference test? 2. Was the same reference test applied to all individuals? 3. Were all patients included in the analysis?
Applicability	
1. Patient selection	Is there a concern that the application or interpretation of the test being assessed does not match the review question?
2. Index test	Is there a concern that driving the index test or interpretation does not match the review question?
3. Benchmark	Is there a concern that the target condition, classified as such through the reference test, differs from the population to which the question was referred?

* This element was modified.

In case of disagreements in the assessment of risk of bias, the differences were resolved by consensus of the research group.

Summary of results

A formal narrative synthesis concerning the accuracy of the machine learning algorithms to determine the level of upper limb impairment was performed.

Meta-analysis

In order to assess the accuracy of the machine learning algorithms in determining the level of upper limb impairment, correlation coefficients were explored. The meta-analysis was performed using the metafor package (v. 3.0-2) of the R software program (R Development Core Team, 2011; R Foundation for Statistical Computing, Vienna, Austria) with the data from the correlation coefficients, using a random model. In addition, a test for funnel plot asymmetry and a likelihood ratio test for publication bias were performed using the metafor and weight packages, respectively.

Results

Selection and characteristics of the studies

The initial search yielded 1626 records. Eleven duplicate records were eliminated, leaving 1615 records that were reviewed by title and abstract. As a result of this review, 189 records related to the research question were identified. Of these, 13 full-text studies were assessed, but only 8 met the eligibility criteria (Fig. 1). All articles had an observational study design.

Results of the individual studies

Data acquisition

The researchers used different modalities for data acquisition in the included studies. Some researchers applied more than one device, while others used a single device. Of these, the most common was surface electromyography (sEMG), followed by electroencephalography (EEG), Microsoft Kinect, inertial measurement unit (IMU), accelerometer, flex sensors, and cell phone.

For sEMG, data are obtained through noninvasive electrodes, which measure the time and intensity of the electrical signals from the muscles. Among the included studies, Wang et al.,[17] Li et al.[18] and Zhou et al.[19] used this device.

Zhang et al. used EEG, which involves placing electrodes on the scalp. Each electrode sends a signal to a device called an electroencephalograph, which displays the rhythmic fluctuation of the brain's electrical activity (brain waves) in real time.[20]

An IMU is an electronic device that measures and reports velocity, orientation and, in some models, gravitational forces. Data are obtained from a combination of accelerometers, gyroscopes and magnetometers. Inertial measurement units are small devices that are placed noninvasively on the patient's skin to obtain motion data in 3 dimensions. Among the included studies, Li et al.[18] used an MPU-9250 device (InvenSense, San Jose, USA), while Zhang et al.[21] used an MPU-6050 device (Xsens Technologies, Los Angeles, USA).

Kim et al. used Microsoft Kinect (Microsoft Corp., Redmond, USA).[22] This device has cameras for motion and depth detection. It was initially developed as a video game device for Microsoft Xbox console; it tracks players' movements while they interact with a game. The Kinect consists

of an infrared light projector and a red-green-blue (RGB) video camera. The reflected infrared light is converted into depth data and calibrated with RGB data to distinguish shapes.

Yu et al. used an ADXL345 accelerometer and flex sensors.[23] An accelerometer is an electronic device that measures the vibration or acceleration of the movement of a structure. The force generated by the vibration or change in motion (acceleration) is detected, and an electrical charge is generated that is proportional to the force exerted on it. Accelerometers also play an important role in determining orientation and direction. Flex sensors are small strips composed of polymeric ink with embedded conductive particles; their function is to measure the resistivity when the sensor is flexed. Subsequently, the resistance value is converted into joint rotation angles.

Finally, Song et al. used an accelerometer and gyroscope integrated into a cell phone (iPhone 7, running an iOS 11.2.5 operating system; Apple Inc., Cupertino, USA).[24] Through this device, the researchers obtained the position and location of the hand in 3 dimensions.

Machine learning algorithms

The machine learning algorithms used for the assessment of motor function are briefly described below.

The machine learning algorithms using supervised learning included the support vector machine (SVM), which was employed by Wang et al.[17] and Zhou et al.[19] Support vector machine is a learning-based method for solving classification and regression problems. This algorithm is a decision function based on the hyperplane concept, a boundary that distinguishes several points in different classes and separates them.[25] In the same sense, Wang et al. used the backpropagation neural network (BPNN).[17] This algorithm applies the concept of gradient descent. Given an artificial neural network and an error function, this method calculates the gradient of the error function concerning the weights of the neural network. Wang et al.[17] and Zhou et al.[19] applied the random forest (RF) algorithm, which is a set of decision trees that are independent of each other. The advantage of the RF algorithm is that it can be used for both classification and regression problems, which constitute the majority of the current machine learning systems.

Continuing with supervised learning algorithms, Zhang et al. employed the convolutional neural network (CNN).[20] This type of neural network processes its layers by emulating the visual cortex of the human eye to recognize different features in the inputs. Convolutional neural network incorporates several specialized hidden layers into a hierarchy. The first layers can detect simple patterns, such as lines, curves and others; this is then specialized to deeper layers that recognize increasingly complex shapes. In the same way, Li et al. applied the least absolute shrinkage and selection operator (LASSO), which is a regression analysis method used to model the relationship between a dependent variable (which can be a vector) and one or more explanatory variables.[18] On the other hand, Kim et al. applied the artificial neural network (ANN), which is a computational learning system that uses a network of functions to understand and translate data input (usually patterns and relationships) into a desired output.[22] The concept of artificial neural network was inspired by human biology and how neurons in the human brain interconnect to understand human sensory inputs. Likewise, Zhou et al. applied linear discriminant analysis (LDA), which is based on the rule of maximum a posteriori probability and Bayesian principles, to find a linear combination of features that characterize or separate 2 or more classes of objects or events.[19]

Finally, Zhang et al. applied the K-nearest neighbor (KNN), which classifies an unknown sample by initially calculating the distance from that sample to all training samples.[21] This algorithm is used to rank values by looking for the "most similar" (closest) data points learned in the training stage and estimating new points based on that ranking. Similarly, Yu et al. used extreme learning machine (ELM).[23] This algorithm includes several hidden neurons in which the input weights are randomly assigned. In this type of network, data only go in one direction through a series of layers. It is implemented fully automatically without iterative tuning, and, in theory, no user intervention is required. Likewise, Song et al. applied the decision tree (DT).[24] This algorithm is the most frequently used in classification and regression problems, in which categorical or continuous input and output variables are used. Decision tree is composed of a root node, several internal nodes and several terminal nodes. The goal of the tree is to make the optimal choice at the end of each node. The name itself suggests that this technique uses a flowchart as a tree structure to show the predictions that result from a series of splits based on the features of the inputs.

Correlation with clinical analysis

In the included studies, the algorithms that showed a very strong correlation with the Fugl–Meyer Assessment test were CNN,[20] DT,[24] SVM,[19] and ELM.[23] The algorithms that presented a strong correlation were the framework with the union of SVM, BPNN, and RF,[17] LASSO,[18] and ANN.[22] Finally, the KNN algorithm presented a strong correlation with the Brunnstrom evaluation scale (Table 3).[21]

Risk of bias and quality assessment

At the QUADAS-2 assessment, 100% of the included studies showed a high risk of bias, whereas the applicability section showed a low risk of bias in 100% of the studies. In addition, 87.5% of the studies did not describe how the patients in the sample were enrolled; therefore, domain 1 showed a high risk of bias (Fig. 2).

Table 3. Characteristics of the individual studies and their results

Study ID	Population	Machine learning algorithm/reference test	Data acquisition device	Result
Wang et al.[17]	Patients with stroke: male n = 9 female n = 6; Healthy subjects: male n = 10 female n = 5	SVM BPNN RF FMA-UE	surface electro-myography	Correlation with FMA-UE Pearson correlation = −0.87, p < 0.0001 True positive rate: SVM = 0.9982 BPNN = 0.9749 RF = 0.9583 Sensitivity and specificity: no data
Zhang et al.[20]	Patients with chronic stroke: n = 12; Healthy subjects: n = 14	CNN FMA-UE	electroencephalo-graphy	Correlation with FMA-UE Pearson correlation = 0.9921, p < 0.0001 Sensitivity and specificity: no data
Li et al.[18]	Patients with chronic stroke: n = 18; Healthy subjects: n = 16	LASSO FMA-UE	Inertial Measurement Unit MPU-9250; surface electro-myography	Correlation with FMA-UE Pearson correlation = 0.8736, p = no data Sensitivity and specificity: no data
Kim et al.[22]	Patients with chronic stroke: n = 41	ANN FMA-UE	Microsoft Kinect	Correlation with FMA-UE Pearson correlation = 0.799, p < 0.0001 Sensitivity and specificity: no data
Zhang et al.[21]	Patients with chronic stroke: n = 21; Healthy subjects: n = 8	KNNs Brunnstrom stages	Inertial Measurement Unit MPU-6050	Correlation with the Brunnstrom stages of recovery r = 0.862, p < 0.001 Sensitivity and specificity: no data
Yu et al.[23]	Patients with chronic stroke: n = 24	ELM FMA-UE	Accelerometer ADXL345; flex sensor	Correlation with FMA-UE Coefficient of determination = 0.918, p = no data Sensitivity and specificity: no data
Song et al.[24]	Patients with chronic stroke: n = 10	DT FMA-UE	cell phone	Correlation with FMA-UE Pearson correlation = 0.97, p < 0.01 Sensitivity and specificity: no data
Zhou et al.[19]	Patients with chronic stroke: n = 6; Healthy subjects: n = 11	SVM LDA RF FMA-UE	surface electro-myography	Correlation with FMA-UE Pearson correlation = 0.93, p < 0.05 Sensitivity and specificity: no data

SVM – support vector machine; BPNN – backpropagation neural network; RF – random forest; CNN – convolutional neural network; LASSO – least absolute shrinkage and selection operator; ANN – artificial neural network; KNNs – K-nearest neighbors; ELM – extreme learning machine; DT – decision tree; LDA – linear discriminant analysis; FMA-UE – Fugl–Meyer Assessment for Upper Extremity.

Meta-analysis

The results of the meta-analysis suggest that there is a correlation between clinical assessment (Fugl–Meyer Assessment and Brunnstrom's evaluation scale) and machine learning algorithms in the evaluation of upper limb motor function (Fisher's z_r (95% confidence interval (95% CI)) = 1.62 (1.24−2.00), p < 0.001). In addition, the absence of homogeneity was observed (I^2 = 85.29%, Q = 48.15), which is attributable to the heterogeneity of the studies. The result of the test for funnel plot asymmetry was z = 1.1914, p = 0.2335, limit estimate (as sei ≥ 0): b = 0.7919 (95% CI: −0.6242−2.2080), as shown in Fig. 3A,B. In addition, a likelihood ratio test was conducted comparing the adjusted model, including the selection to its unadjusted random-effects counterpart. The Vevea and Hedges weight-function model resulted in a likelihood ratio of χ^2 = 0.3062, p = 0.58. Taken together, this suggests that there was no publication bias.

Discussion

A wide variety of machine learning algorithms are described in this systematic review. Out of the 8 included studies, 6 (75.0%) used only one algorithm to assess motor function; 3 of these presented a very strong correlation[20,23,24] and 3 showed a strong correlation[18,21,22] between the algorithms for motor assessment and clinical assessment. Two (25.0%) of the studies employed 3 algorithms: Zhou et al. showed a very strong correlation[19] and Wang et al. reported a strong correlation[17] between motor assessment algorithms and clinical evaluation. The evidence is not conclusive concerning whether better results are obtained with the exclusive use of a single algorithm or with a combination of algorithms. This is contrary to Wang et al., who are in favor of combined use.[17]

In machine learning, the number of samples required for training the machine learning model depends on the complexity of both the problem to be solved and the algorithm developed.[26] Although there is no established minimum
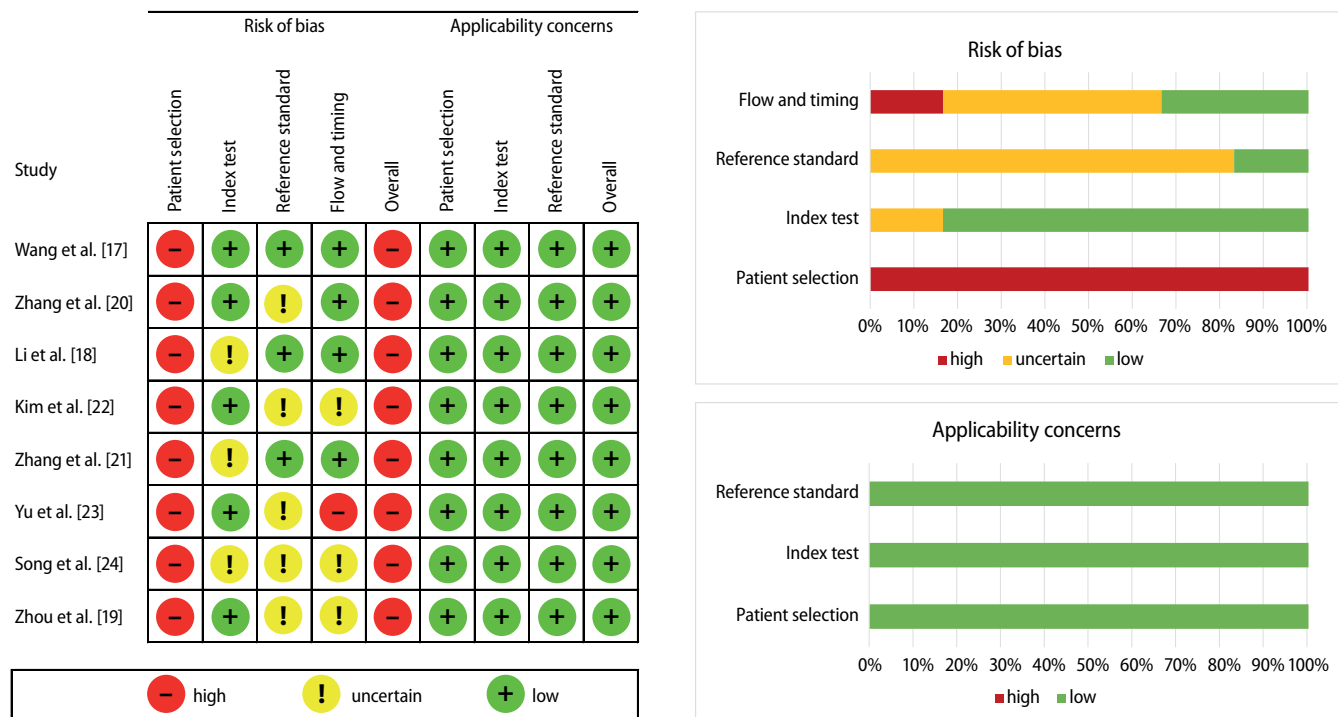
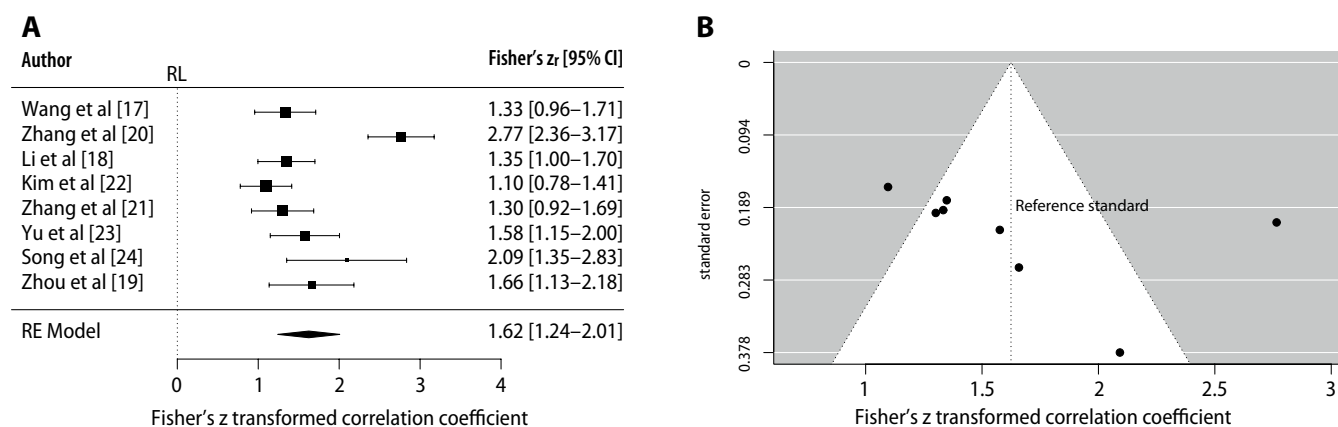**Fig. 2.** Risk of bias assessment of the included studies using QUADAS-2

Fig. 3. Meta-analysis. A. Forest plot; B. Funnel plot, reference line (RL)

RE – random effects; 95% CI – 95% confidence interval.

number of training samples,[27] there are experiments that have indicated that increasing the size of the dataset improves performance.[27–29] Therefore, once the algorithm starts to detect patterns, it is best to increase the sample size. Of the included studies, 4 did not report the number of samples used for training,[18,21,22,24] while the rest reported using 992,[19] 1080,[17] 1680,[23] and 1960[20] samples; however, they did not justify the sample size used.

Data acquisition was performed using various sensors in the training of the algorithm and the evaluation of the upper limb motor function. Some are readily available (cell phone[24] and inertial sensors[18,21,23]), while others are specialized equipment that would limit their use to healthcare units (electromyography[17–19] and electroencephalography[20]).

The gold standard for evaluating motor function is clinical assessment, for which various assessment tools are available. These tools assess either general motor function (Medical Research Council (MRC) Scale and Fugl–Meyer Assessment) or specific areas of impairment (upper limb function, trunk function, gait ability, and spasticity). In this regard, the MRC Scale, the Frenchay Arm Test, and the Action Research Arm Test (ARAT) are specific to the upper limb motor function.[30]

The Fugl–Meyer Assessment is a performance-based index of the stroke-specific impairment.[31] It is designed to assess motor functioning, sensation, balance, joint range of motion, and joint pain in patients with post-stroke hemiplegia. It is the most commonly used scale in clinical assessments of the upper limb.[32,33] Thus, 87.5% of the studies

used the Fugl–Meyer Assessment for Upper Extremity (FMA-UE). The only study that did not use the FMA-UE was the study by Zhang et al.,[21] who used the Brunnstrom's evaluation scale.[34] It rates the recovery stage of the upper and lower extremities and hands in levels. The stages are classified from I to VI of recovery, whereby I indicates that the patient has low or absent movement, and VI indicates that the patient can perform voluntary movements.

As can be seen, the reported studies present at least a strong correlation with standard clinical tests. Therefore, the proposed evaluation systems have the potential to support therapists in the objective measurement of the upper limb motor function. Although the meta-analysis found a good relationship between machine learning algorithms and clinical assessment, it also showed a high heterogeneity.

The literature proposes that home-based rehabilitation can offer potential benefits,[35–37] such as performing the exercises according to the patient's schedule, providing flexibility of location and time, and receiving remote feedback and follow-up by the therapist. The home-based rehabilitation is possible to implement by having motor function evaluation systems, such as those presented in this review.

To the best of our knowledge, there are no systematic reviews in the literature evaluating the correlation between the clinical assessment of the upper limb motor function and machine learning algorithms in post-stroke patients.

Duque et al. conducted a systematic review that included studies focused on evaluating movement analysis in patients with stroke, Parkinson's disease, spinal cord injury, Huntington's disease, multiple sclerosis, and cerebral palsy, as well as in premature infants and the elderly.[38] However, their review did not perform the risk of bias assessment or a meta-analysis. Furthermore, it only focused on describing the devices used for data acquisition and the machine learning algorithms.

There are narrative reviews regarding the use of capture sensors and machine learning to perform automated assessments in home-based rehabilitation programs.[39] Caramiaux et al. described machine learning models for motor learning and their adaptive capabilities.[40] Moon et al. conducted a scoping review to explore the use of artificial neural networks in neurorehabilitation in various pathologies, including stroke, particularly in the prediction of variables such as functional recovery and rehospitalization.[41] In the same vein, Sirsat et al. performed a narrative review about the use of machine learning in stroke patients, grouping them according to their use for the identification of associated risk factors, diagnosis, treatment, and prognosis.[42] In summary, current reviews studying the application of machine learning in stroke patients focus on its use as a plausible tool for prediction and classification of neurological and motor impairments, as well as the assessment of rehabilitation progress.

Limitations

For more than a decade, the number of publications in basic science and clinical trials has grown exponentially. Clinical trials are considered the best evidence of solving a health problem. Unfortunately, some basic science results are not necessarily reflected in clinical practice.[43] Furthermore, several publications have divergent results despite presenting characteristics that superficially seem similar, or they use different variables to measure the impact of the intervention.[44,45] Hence the importance of evidence-based medicine aimed at determining the validity and analyzing the dataset of published studies through systematic reviews.

This systematic review encountered limitations, such as small sample sizes and a risk of bias in the included studies. In addition, the results of the meta-analysis showed high heterogeneity, probably due to the diversity of the statistical tests used in the correlation and different algorithms used in the studies. In addition, this review was limited to analyzing studies focused on the evaluation of the upper limb motor function, so studies analyzing the lower limb motor function were not considered. The exclusion of studies focused on the lower limbs could lead to a limitation, since both limbs have similar ranges of mobility; however, the inclusion of both limbs may have increased the heterogeneity of the review.

Conclusions

The results of the studies included in this systematic review show strong correlations between machine learning algorithms and clinical assessment scores of the upper limb. This correlation indicates a possible application to assist therapists in improving the efficacy of individualized diagnosis of motor function in post-stroke patients. The algorithms also serve as feedback to facilitate the training process for patient rehabilitation. Finally, studies with a representative sample, low risk of bias and better methodological quality are required to reach more robust conclusions.

ORCID iDs

Jorge Fernando Ambros-Antemate
  https://orcid.org/0000-0003-1309-6147
Adriana Reyes-Flores   https://orcid.org/0000-0002-5757-6495
Liliana Argueta-Figueroa   https://orcid.org/0000-0002-1044-6757
Rafael Ramírez-Ramírez   https://orcid.org/0000-0003-3729-2150
Marciano Vargas-Treviño   https://orcid.org/0000-0002-2090-9727
Jaime Gutiérrez-Gutiérrez   https://orcid.org/0000-0002-7713-3165
Eduardo Pérez-Campos Mayoral
  https://orcid.org/0000-0002-6032-7609
Eduardo Pérez-Campos   https://orcid.org/0000-0001-6720-7952
Luis Angel Flores-Mejía   https://orcid.org/0000-0001-8157-5233
Rafael Torres-Rosas   https://orcid.org/0000-0002-5934-003X

References

1. World Health Organization, World Bank. *World Report on Disability*. Geneva, Switzerland: World Health Organization; 2011. https://apps.who.int/iris/handle/10665/44575. Accessed October 19, 2021.
2. Das J, Rajanikat GK. Post stroke depression: The sequelae of cerebral stroke. *Neurosci Biobehav Rev*. 2018;90:104–114. doi:10.1016/j.neubiorev.2018.04.005
3. Le Danseur M. Stroke rehabilitation. *Crit Care Nurs Clin North Am*. 2020;32(1):97–108. doi:10.1016/j.cnc.2019.11.004
4. Futrell M, Rozzi SL. Principles of rehabilitation. *Prim Care*. 2020;47(1):87–103. doi:10.1016/j.pop.2019.10.004
5. Gladstone DJ, Danells CJ, Black SE. The Fugl–Meyer assessment of motor recovery after stroke: A critical review of its measurement properties. *Neurorehabil Neural Repair*. 2002;16(3):232–240. doi:10.1177/154596802401105171
6. Wolf SL, Catlin PA, Ellis M, Archer AL, Morgan B, Piacentino A. Assessing Wolf motor function test as outcome measure for research in patients after stroke. *Stroke*. 2001;32(7):1635–1639. doi:10.1161/01.STR.32.7.1635
7. Tanaka K, Yano H. Errors of visual judgement in precision measurements. *Ergonomics*. 1984;27(7):767–780. doi:10.1080/00140138408963550
8. Kringle EA, Knutson EC, Engstrom C, Terhorst L. Iterative processes: A review of semi-supervised machine learning in rehabilitation science. *Disabil Rehabil Assist Technol*. 2020;15(5):515–520. doi:10.1080/17483107.2019.1604831
9. Tack C. Artificial intelligence and machine learning applications in musculoskeletal physiotherapy. *Musculoskelet Sci Pract*. 2019;39:164–169. doi:10.1016/j.msksp.2018.11.012
10. Perezcampos Mayoral C, Gutiérrez Gutiérrez J, Cano Pérez JL, et al. Fiber optic sensors for vital signs monitoring: A review of its practicality in the health field. *Biosensors (Basel)*. 2021;11(2):58. doi:10.3390/bios11020058
11. O'Neil O, Gatzidis C, Swain I. A state of the art survey in the use of video games for upper limb stroke rehabilitation. In: Ma M, Jain LC, Anderson P, eds. *Virtual, Augmented Reality and Serious Games for Healthcare 1*. Berlin, Germany: Springer Berlin Heidelberg; 2014:345–370. doi:10.1007/978-3-642-54816-1_18
12. Beristain-Colorado MDP, Ambros-Antemate JF, Vargas-Treviño M, et al. Standardizing the development of serious games for physical rehabilitation: Conceptual framework proposal. *JMIR Serious Games*. 2021;9(2):e25854. doi:10.2196/25854
13. Moher D, Liberati A, Tetzlaff J, Altman DG. Preferred reporting items for systematic reviews and meta-analyses: The PRISMA statement. *PLoS Med*. 2009;6(7):e1000097. doi:10.1371/journal.pmed.1000097
14. Higgins JPT, Thomas J, Chandler J, et al., eds. *Cochrane Handbook for Systematic Reviews of Interventions*. 2nd ed. Hoboken, USA: Wiley; 2019. doi:10.1002/9781119536604
15. Mendeley. *Reference Management Software and Researcher Network*; 2015. https://www.mendeley.com/download-reference-manager/windows. Accessed October 10, 2021.
16. Whiting PF, Rutjes AWS, Westwood ME, et al. QUADAS-2: A revised tool for the quality assessment of diagnostic accuracy studies. *Ann Intern Med*. 2011;155(8):529–536. doi:10.7326/0003-4819-155-8-201110180-00009
17. Wang C, Peng L, Hou ZG, Li J, Zhang T, Zhao J. Quantitative assessment of upper-limb motor function for post-stroke rehabilitation based on motor synergy analysis and multi-modality fusion. *IEEE Trans Neural Syst Rehabil Eng*. 2020;28(4):943–952. doi:10.1109/TNSRE.2020.2978273
18. Li Y, Zhang X, Gong Y, Cheng Y, Gao X, Chen X. Motor function evaluation of hemiplegic upper-extremities using data fusion from wearable inertial and surface EMG sensors. *Sensors (Basel)*. 2017;17(3):582. doi:10.3390/s17030582
19. Zhou Y, Zeng J, Jiang H, Li Y, Jia J, Liu H. Upper-limb functional assessment after stroke using mirror contraction: A pilot study. *Artif Intell Med*. 2020;106:101877. doi:10.1016/j.artmed.2020.101877
20. Zhang X, D'Arcy R, Menon C. Scoring upper-extremity motor function from EEG with artificial neural networks: A preliminary study. *J Neural Eng*. 2019;16(3):036013. doi:10.1088/1741-2552/ab0b82
21. Zhang Z, Fang Q, Gu X. Objective assessment of upper limb mobility for post-stroke rehabilitation. *IEEE Trans Biomed Eng*. 2015;63(4):859–868. doi:10.1109/TBME.2015.2477095
22. 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(7):e0158640. doi:10.1371/journal.pone.0158640
23. Yu L, Xiong D, Guo L, Wang J. A remote quantitative Fugl–Meyer assessment framework for stroke patients based on wearable sensor networks. *Comput Methods Programs Biomed*. 2016;128:100–110. doi:10.1016/j.cmpb.2016.02.012
24. Song X, Chen S, Jia J, Shull PB. Cellphone-based automated Fugl–Meyer assessment to evaluate upper extremity motor function after stroke. *IEEE Trans Neural Syst Rehabil Eng*. 2019;27(10):2186–2195. doi:10.1109/TNSRE.2019.2939587
25. Mechelli A, Viera S, eds. *Machine Learning: Methods and Applications to Brain Disorders*. San Diego, USA: Elsevier; 2019. ISBN: 978-012815-739-8.
26. Lantz B. *Machine Learning with R*. Birmingham, UK: Packt Publishing; 2013. ISBN: 978-1-78216-214-8.
27. Andonie R. Extreme data mining: Inference from small datasets. *Int J Comput Commun Control*. 2010;5(3):280–291. doi:10.15837/ijccc.2010.3.2481
28. Prusa J, Khoshgoftaar TM, Seliya N. The effect of dataset size on training tweet sentiment classifiers. In: *2015 IEEE 14th International Conference on Machine Learning and Applications (ICMLA)*. Miami, USA: IEEE; 2015:96–102. doi:10.1109/ICMLA.2015.22
29. Balki I, Amirabadi A, Levman J, et al. Sample-size determination methodologies for machine learning in medical imaging research: A systematic review. *Can Assoc Radiol J*. 2019;70(4):344–353. doi:10.1016/j.carj.2019.06.002
30. Cuadrado ÁA. Rehabilitation of the stroke: Evaluation, prognosis and treatment [in Spanish]. *Galicia Clín*. 2009;70(3):25–40. https://dialnet.unirioja.es/descarga/articulo/4208262.pdf. Accessed October 19, 2021.
31. Gladstone DJ, Danells CJ, Black SE. The Fugl–Meyer assessment of motor recovery after stroke: A critical review of its measurement properties. *Neurorehabil Neural Repair*. 2002;16(3):232–240. doi:10.1177/154596802401105171
32. Santisteban L, Térémetz M, Bleton JP, Baron JC, Maier MA, Lindberg PG. Upper limb outcome measures used in stroke rehabilitation studies: A systematic literature review. *PLoS One*. 2016;11(5):e0154792. doi:10.1371/journal.pone.0154792
33. Gor-García-Fogeda MD, Molina-Rueda F, Cuesta-Gómez A, Carratalá-Tejada M, Alguacil-Diego IM, Miangolarra-Page JC. Scales to assess gross motor function in stroke patients: A systematic review. *Arch Phys Med Rehabil*. 2014;95(6):1174–1183. doi:10.1016/j.apmr.2014.02.013
34. Brunnstrom S. Motor testing procedures in hemiplegia: Based on sequential recovery stages. *Phys Ther*. 1966;46(4):357–375. doi:10.1093/ptj/46.4.357
35. Chen Y, Abel KT, Janecek JT, Chen Y, Zheng K, Cramer SC. Home-based technologies for stroke rehabilitation: A systematic review. *Int J Med Inform*. 2019;123:11–22. doi:10.1016/j.ijmedinf.2018.12.001
36. Langan J, DeLave K, Phillips L, Pangilinan P, Brown S. Home-based telerehabilitation shows improved upper limb function in adults with chronic stroke: A pilot study. *J Rehabil Med*. 2013;45(2):217–220. doi:10.2340/16501977-1115
37. Anderson C, Mhurchu CN, Rubenach S, Clark M, Spencer C, Winsor A. Home or hospital for stroke rehabilitation? Results of a randomized controlled trial. II: Cost minimization analysis at 6 months. *Stroke*. 2000;31(5):1032–1037. doi:10.1161/01.STR.31.5.1032
38. Duque E, Trefftz H, Srivastava S. Objective assessments of human motor ability of the upper limb: A systematic review. *Technol Disabil*. 2021;33(1):29–44. doi:10.3233/TAD-200263
39. Liao Y, Vakanski A, Xian M, Paul D, Baker R. A review of computational approaches for evaluation of rehabilitation exercises. *Comput Biol Med*. 2020;119:103687. doi:10.1016/j.compbiomed.2020.103687
40. Caramiaux B, Françoise J, Liu W, Sanchez T, Bevilacqua F. Machine learning approaches for motor learning: A short review. *Front Comput Sci*. 2020;2:16. doi:10.3389/fcomp.2020.00016
41. Moon S, Ahmadnezhad P, Song HJ, et al. Artificial neural networks in neurorehabilitation: A scoping review. *NeuroRehabilitation*. 2020;46(3):259-269. doi:10.3233/NRE-192996
42. Sirsat MS, Fermé E, Câmara J. Machine learning for brain stroke: A review. *J Stroke Cerebrovasc Dis*. 2020;29(10):105162. doi:10.1016/j.jstrokecerebrovasdis.2020.105162

43. Manzo-Toledo A, Torres-Rosas R, Mendieta-Zerón H, Arriaga-Pizano L, Argueta-Figueroa L. Hydroxychloroquine in the treatment of COVID-19 disease: A systematic review and meta-analysis. *Med J Indones*. 2021; 30(1):20–32. doi:10.13181/mji.oa.205012

44. Castro-Gutiérrez MEM, Argueta-Figueroa L, Fuentes-Mascorro G, Moreno-Rodríguez A, Torres-Rosas R. Novel approaches for the treatment of necrotic immature teeth using regenerative endodontic procedures: A systematic review and meta-analysis. *Appl Sci*. 2021; 11(11):5199. doi:10.3390/app11115199

45. Ávila-Curiel BX, Gómez-Aguirre JN, Gijón-Soriano AL, Acevedo-Mascarúa AE, Argueta-Figueroa L, Torres-Rosas R. Complementary interventions for pain in patients with temporomandibular joint disorders: A systematic review [in Spanish]. *Rev Int de Acupunt*. 2020;14(4): 151–159. doi:10.1016/j.acu.2020.10.004