PLOS ONE

RESEARCH ARTICLE

Machine learning in predicting outcomes for stroke patients following rehabilitation treatment: A systematic review

Wanting Zu, Xuemiao Huang, Tianxin Xu, Lin Du, Yiming Wang, Lisheng Wang ⓘ *, Wenbo Nie*

School of Nursing, Jilin University, Changchun, China

* lishengwang@jlu.edu.cn (LW); niewb@jlu.edu.cn (WN)

Abstract

Objective

This review aimed to summarize the use of machine learning for predicting the potential benefits of stroke rehabilitation treatments, to evaluate the risk of bias of predictive models, and to provide recommendations for future models.

Materials and methods

This systematic review was conducted in accordance with the PRISMA statement and the CHARMS checklist. The PubMed, Embase, Cochrane Library, Scopus, and CNKI databases were searched up to April 08, 2023. The PROBAST tool was used to assess the risk of bias of the included models.

Results

Ten studies within 32 models met our inclusion criteria. The optimal AUC value of the included models ranged from 0.63 to 0.91, and the optimal R^2 value ranged from 0.64 to 0.91. All of the included models were rated as having a high or unclear risk of bias, and most of them were downgraded due to inappropriate data sources or analysis processes.

Discussion and conclusion

There remains much room for improvement in future modeling studies, such as high-quality data sources and model analysis. Reliable predictive models should be developed to improve the efficacy of rehabilitation treatment by clinicians.

Introduction

Stroke remains one of the most common diseases that causes functional impairment, especially due to the rapidly growing number of older adults [1]. Due to the increasing prevalence of

patients suffering from the effects of stroke, the importance and burden of stroke rehabilitation are high [1, 2]. In recent years, many effective stroke rehabilitation treatments have been proposed through randomized trials, such as task-oriented training treatment, functional strength training, and robot-assisted treatment [3–5]. Nonetheless, clinicians often face the challenge of choosing the most adequate rehabilitation treatment for patients since the benefits of treatments vary across individuals with different characteristics [6]. The precise prediction of rehabilitation treatment is therefore important for properly distributing rehabilitation resources and delivering patient-specific rehabilitation [7, 8].

Machine learning is a type of artificial intelligence that focuses on constructing computerized algorithms to automatically improve performance through experience. In recent decades, machine learning has shown an ability to effectively deal with high-throughput data, and it has become a popular method in many fields, ranging from biology to social science [9, 10]. Many kinds of research based on machine learning have also evolved in the medical field due to its ability to handle health care data and thus aid clinical workflows. In the stroke field, machine learning methods are currently applied in early detection, diagnosis, and outcome prediction [11, 12]. Recently, an increasing number of studies have examined machine learning methods with the aim of predicting outcomes and identifying stroke patients who might benefit from specific rehabilitation treatments. A systematic review that evaluates the quality of these studies would be beneficial for further similar studies.

Objective

This review aimed to systematically summarize studies that used machine learning methods to build models as well as externally validated studies that predicted the potential benefits of patients following stroke rehabilitation treatments. We also aimed to evaluate the risk of bias of the included models and therefore propose potential improvements, which might provide evidence for further modeling studies and thus aid the decision-making process in stroke rehabilitation clinical settings.

Materials and methods

Protocol

This review was performed in accordance with the PRISMA statement and the **CH**ecklist for critical **A**ppraisal and data extraction for systematic **R**eviews of prediction **M**odelling **S**tudies (CHARMS) [13, 14]. The CHARMS checklist was developed to support the design of systematic reviews of predictive modeling studies and provides guidance for forming the review question, study selection, and data extraction. The aim of our review was summarized into key items, as presented in Table 1. In addition, our systematic review has been registered on PROSPERO (ID number: CRD42022299195, available at https://www.crd.york.ac.uk/PROSPERO/).

Table 1. CHARMS guidelines for the formation of review question.

Key item	Definition
Intended scope of the review	Studies aimed at predicting clinical outcomes after specific rehabilitation treatment for stroke patients
Type of prediction modelling studies	Model development or validation studies
Target population to whom the prediction model applies	Post-stroke patients that had received rehabilitation treatment
Outcome to be predicted	Motor functional outcomes measured by standardized scales
Intended moment of using the model	Before rehabilitation treatment

https://doi.org/10.1371/journal.pone.0287308.t001

Table 1 shows the aim of this review according to the CHARMS guidelines.

Inclusion and exclusion criteria

Given the aim of this review, the eligibility criteria were as follows:

Inclusion criteria

· Studies focused on the development or validation of prediction models for recovery potential after stroke rehabilitation

· Models based on machine learning methods

· Patients in the primary studies must have received specific stroke rehabilitation treatment regardless of the stroke stage and age of the patients

· The predicted outcomes of the model must be motor functional outcomes assessed through standard tools

· The prediction model was designed for use before rehabilitation treatment

Exclusion criteria

· Studies aimed at identifying predictors related to outcomes rather than predicting clinical outcomes for individual patients

· Studies aimed at evaluating the impact of using predictive models in clinical settings

· Full-text article was not available

· Model methods were not reported in detail, including study protocol, conference abstracts, letters, etc.

· Reviews or comments without original research

Search strategy

Two authors independently searched the PubMed, Embase Cochrane Library, Scopus, and CNKI (China National Knowledge Infrastructure) databases up to December 15, 2021 (updated on April 08, 2023) to identify relevant studies. 'Stroke', 'machine learning', 'rehabilitation', and their synonyms were used as MeSH terms or free-text words to identify eligible studies. An example search strategy for PubMed is provided in the S1 Table. We also manually searched the reference lists and citations of the included studies as well as Google Scholar to obtain additional resources.

After removing duplicates, we selected eligible studies based on titles and abstracts in accordance with the inclusion and exclusion criteria described above. The full texts were then screened by two reviewers, and any disagreements were resolved by consulting a third reviewer.

Data extraction and quality assessment

A data extraction sheet was used to address any information that would increase the risk of bias of the models. Briefly, the extracted data included the source of data, participants, predicted outcomes, predictors, model development, model performance, and model evaluation methods, as recommended in the CHARMS checklist [14]. We extracted discrimination and calibration data as primary metrics to describe model performance. Discrimination is often estimated by the area under the receiver-operating characteristic curve (AUC-ROC) for logistic regression models and should reflect the ability of a model to distinguish between individuals with or without the outcome of prediction models. Calibration is often estimated by calculating the Hosmer–Lemeshow goodness-of-fit test with a calibration pot and should reflect the agreement between the predicted and the observed outcome [15, 16]. We entered the details into the data extraction sheet, which is provided in the S2 Table.

PROBAST (Prediction model Risk Of Bias ASsessment Tool) was used to guide the risk of bias assessment in this review [15]. The PROBAST tool was mainly designed to estimate the quality of the individual prediction model in systematic reviews. The prediction models were explicitly classified into three types in this tool and relevant signaling questions were proposed for evaluating different types of prediction models. Furthermore, the signaling questions were grouped into four domains of the potential source of bias: participants, predictors, outcome, and analysis. If one of the four domains had high risk of bias (ROB), the overall judgement would be a high ROB [17]. The unit of evaluation applied in this review was model rather than study, since some studies might develop or validate several models.

Results

Search results

The PRISMA flowchart (Fig 1) presents the selection process of eligible studies in this review. In total, 3639 records were obtained based on the search strategy. After deleting duplicates, 2289 records were then screened according to the title and abstract. The majority of studies were excluded at this stage because their aims, designs, and outcomes were outside the scope of this review. Twenty-one full-text studies that met the inclusion and exclusion criteria were then examined and excluded for the reasons shown in the flowchart. Finally, 10 studies were included in this narrative review, and 32 models were included in the risk of bias assessment.
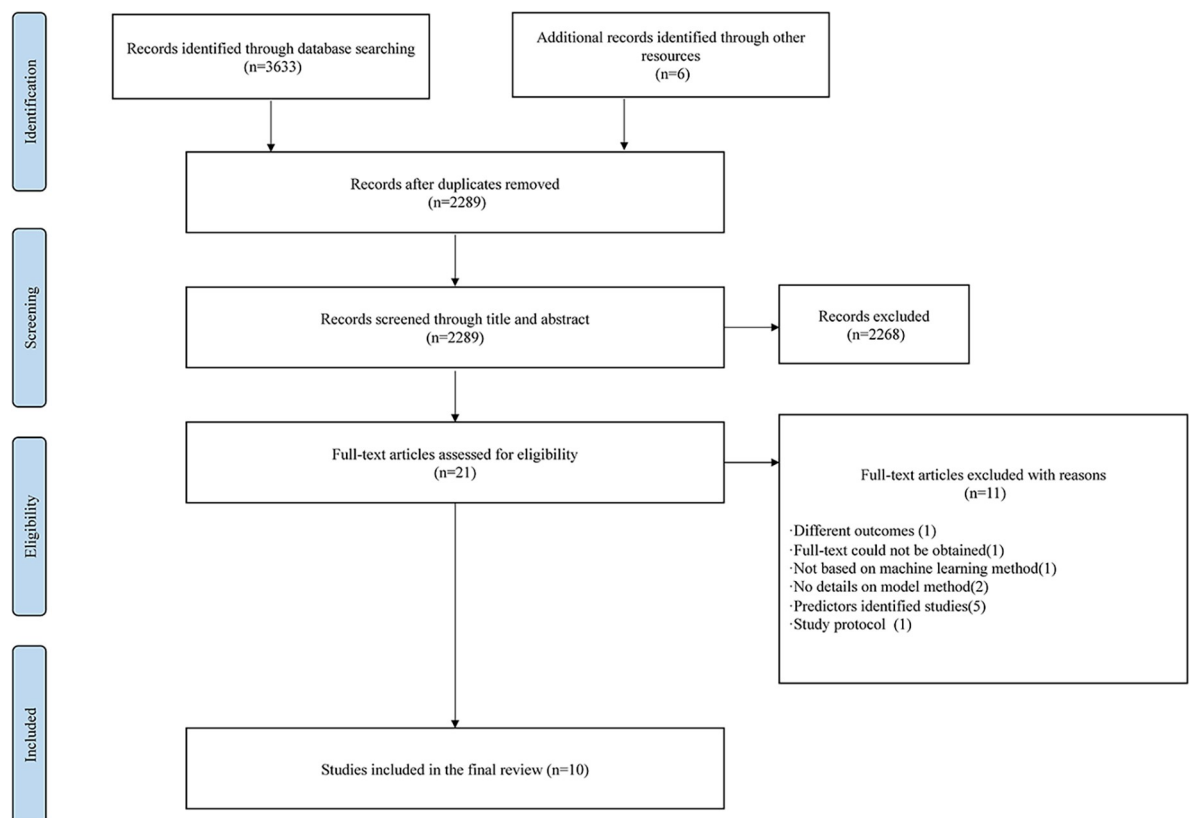

Fig 1. PRISMA flow chart. Fig 1 shows the selection process of eligible studies in this review according to PRISMA.

https://doi.org/10.1371/journal.pone.0287308.g001

Characteristics of included studies

We summarized the characteristics of the included studies in Table 2. All ten studies described model development, and one of them also considered external validation for previous models [18]. Only two studies used data from randomized trials [19, 20], whereas the majority of the studies used electronic medical records as the data source for model development or validation. Five studies utilized multicenter data within their country as data sources [19–23]. Four of the included articles were conducted in the United States [19, 22–24], four were conducted in Europe [18, 21, 25, 26], and two were conducted in Asia [20, 27].

Table 2 described the characteristics of included studies.

Regarding the participants included in the primary studies, two studies selected chronic stroke as one of the inclusion criteria for patients [19, 20], and two studies only included patients who had been admitted within 90 days of onset of stroke [18, 21]. Two studies included subacute phase stroke patients [25, 26], whereas the remaining four studies had no restrictions on the stage or type of stroke; however, one study explicitly excluded acute stroke patients [27]. Furthermore, all of the participants had completed an organized physical rehabilitation program, one of the studies involved transcranial magnetic stimulation [19], and two studies used robot-assisted rehabilitation [25, 27].

Among the included studies, regression was the most common method implemented to develop models. Specifically, logistic regression was used in six studies [18, 19, 21–23, 27], linear regression was used in two studies [25, 26], and a single study used Lasso regression [24]. Other common machine learning approaches, such as artificial neural network, k-nearest neighbors, and random forest were used in three studies [19, 20, 26], as presented in Table 2. Four studies provided models with external validation [20–23], four studies considered internal validation [19, 24–26], and a single study did not mention the validation process [27]. Moreover, one study externally validated two existing models in a previous study using data from a different country and also developed a novel model with internal validation using the same database [18]. The optimal AUC value of the included models ranged from 0.63 to 0.91. Four studies chose the R^2 value to describe the discrimination of models, and it ranged from 0.64 to 0.91 [19, 24–26]. These outcomes suggest that the discriminative ability of the included models varied.

Quality assessment of included studies

According to the PROBAST tool, all the models demonstrated an overall high (n = 30) or unclear (n = 2) risk of bias (Fig 2). This indicates that the performance and usability of each model might be overoptimistic. Nearly all the models were biased from participants and analysis domains, and the common causes of downgrading were inappropriate data sources or analysis processes. Models from the same study have a common risk of bias in terms of participant domain since they share the same data source. Among the included models, only twelve models in two studies that used randomized trail data were rated as having a low risk of bias in the participant domain. Twenty models were rated as an unclear risk of bias or high risk of bias in the outcome domain. All the models had a low risk of bias for the predictor domain, which indicates that all the predictors selected could be obtained before treatment and tend to be assessed in similar ways. However, in the analysis domain, all the remaining models were rated as having a high risk of bias, except two models were considered to have an unclear risk of bias.

Discussion

In recent years, as machine learning has emerged as an attractive approach to address big data in health care, many related studies have bene published, especially studies of stroke patients.

Table 2. Characteristics of included studies.

Study	Number of models	Study type	Data source	Sample size Stroke stage	Rehabilitation treatment	Outcome	Number and type of predictors in the final model	Model development method	Optimal model performance results
Tozlu et al., 2020	10	Model development	Multicenter randomized trial	N = 102 Chronic stroke	Repetitive TMS and physical therapy	Post-intervention UE-FMA	93 Demographic variables, clinical variables, 2 TMS-based neurophysiological measures, and structural disconnectivity measurements for 86 regions	EN ANN SVM CART RF	R^2 0.910 RMSE 0.362
						FMA changed <5.5 (nonresponder) or ≥5.5 (responder) MCID = 5.5	93 Demographic variables, clinical variables, 2 TMS-based neurophysiological measures, and structural disconnectivity measurements for 86 regions	EN ANN SVM CART RF LR	AUC 0.63
Scrutinio et al., 2017	2	Model development	Electronic medical records	Derivation Cohort (n = 717); Validation Cohort (n = 875) Acute and subacute stroke	Physical and occupational therapy	M-FIM score of >61 points at discharge	5 Age, Time since onset, Unilateral neglect, Motor-FIM score, Cognitive FIM score	LR	AUC 0.883 HL test 4.12 (P = 0.249)
						physical independence grade ≥5 (FIS)	5 Age, Time since onset, Male gender, Motor-FIM score, Cognitive FIM score	LR	AUC 0.913 HL test 1.20 (P = 0.754)
Lee et al., 2021	2	Model development	Electronic medical records	N = 86 Subacute stroke	Robot-assisted therapy using InMotion2 along with conventional therapy	FMA changed≥9 (responder) MCID = 9	2 Time since onset Hand Movement Scale	LR	AUC 0.658
				N = 58 Chronic stroke		FMA changed≥5.25 (responder) MCID = 5.25	2 Time since onset Hand Movement Scale	LR	AUC 0.739
García-Rudolph et al., 2021	2	External validation	Electronic medical records	N = 710 Acute and subacute stroke	Inpatient rehabilitation program	M-FIM score of >61 points at discharge	5 Age, Time since onset, Unilateral neglect, Motor-FIM score, Cognitive FIM score	LR	AUC 0.873 HL test 6.07 (P = 0.63)
		External validation	Electronic medical records	N = 710 Acute and subacute stroke	Inpatient rehabilitation program	physical independence grade ≥5 (FIS)	5 Age, Time since onset, Male gender, Motor-FIM score, Cognitive FIM score	LR	AUC 0.803 HL test 8.91 (P = 0.34)
		Model development	Electronic medical records	N = 710 Acute and subacute stroke	Inpatient rehabilitation program	M-FIM score of >61 points at discharge; physical independence grade ≥5 (FIS)	5 Age, Time since onset, Unilateral neglect, Motor-FIM score, Aphasia	LR	AUC 0.894 HL test 10.40 (P = 0.23) for primary outcome AUC 0.845 HL test 6.94 (P = 0.54) for secondary outcome

(Continued)

Table 2. (Continued)

Study	Number of models	Study type	Data source	Sample size Stroke stage	Rehabilitation treatment	Outcome	Number and type of predictors in the final model	Model development method	Optimal model performance results
Bates et al.,2015 PART 1	1	Model development	Electronic medical records	N = 4020 No limitation	consultative or comprehensive treatment	physical grade IV (FIM)	7 Age, clinical variables, type of rehabilitation, recovery time, acute procedure	LR	AUC 0.84 HL test P = .93
Bates et al.,2015 PART 2	1	Model development	Electronic medical records	N = 5416 No limitation	consultative or comprehensive treatment	physical grade VI (FIM)	7 Age, clinical variables, type of rehabilitation recovery time, acute procedure	LR	AUC 0.83 HL test P = .38
Harari et al, 2020	4	Model development	Electronic medical records	N = 50 No limitation	Inpatient rehabilitation program	FIM	10 Demographic information, stroke characteristics, and scores of the clinical tests	Lasso regression	R^2 0.76 MAE 7.6
						Ten-Meter Walk Test	6 Demographic information, stroke characteristics, and scores of the clinical tests		R^2 0.70 MAE 0.26
						Six-Minute Walk Test	4 Demographic information and scores of the clinical tests		R^2 0.70 MAE 73.2
						Berg Balance Scale	8 Demographic information, stroke characteristics, and scores of the clinical tests		R^2 0.77 MAE 6.4
Thakkar et al., 2020	2	Model development	Randomized controlled trials	N = 239 Chronic stroke	Contemporary task-oriented interventions	FMA changed$\geq$4 (high responders) FMA changed<4 (low responders) MCID = 4	3 Time since stroke, baseline FIM scores and FMA scores	KNN ANN	Accuracy 85.42% AUC 0.89
Goffredo et al., 2022	6	Model development	Electronic medical records	N = 66 Subacute stroke	InMotion2 based upper limb Robot-assisted Therapy (ulRT) along with conventional therapy	Post-intervention MI_{ELBOW}	4 RMK metrics	MLR	R^2 0.683 RMSE 5.547
						Post-intervention $MI_{SHOULDER}$	4 RMK metrics		R^2 0.640 RMSE 5.294
						Post-intervention MI_{UL}	4 RMK metrics		R^2 0.765 RMSE 13.859

(*Continued*)

Table 2. (Continued)

Study	Number of models	Study type	Data source	Sample size Stroke stage	Rehabilitation treatment	Outcome	Number and type of predictors in the final model	Model development method	Optimal model performance results
Gandolfi et al, 2023	2	Model development	Electronic medical records	N = 95 Subacute stroke	tailored upper limb rehabilitation treatment	UE-FMA at discharge	24 demographic and clinical domains; cognitive; body function and disability domains	MLR RF	R^2 0.807 RMSE 6.17

TMS, transcranial magnetic stimulation; MCID, minimal clinically important difference; UE-FMA, upper-extremity Fugl-Meyer Assessment; FIM, Functional Independence Measure; FIS, Functional Independence Staging system; ANN, artificial neural network; CART, classification and regression trees; EN, elastic net; RF, random forest; SVM, support vector machine; KNN k-nearest neighbors; LR, logistic regression; RMSE, root of mean squared error; AUC, area under the receiver operating characteristic curve; HL test, the Hosmer–Lemeshow test; MAE, Mean Absolute Error; MI_{ELBOW}, Motricity Index affected elbow flexion; $MI_{SHOULDER}$, Motricity Index affected shoulder abduction; MI_{UL}, Motricity Index affected Upper Limb; RMK, Robot-Measured Kinematic; MLR, multiple linear regression; SSV, Split Sample Validation; LOOCV, Leave-One-Out Cross-Validation.

https://doi.org/10.1371/journal.pone.0287308.t002

In this review, we systematically searched studies aimed at using machine learning methods to predict recovery potential following stroke rehabilitation treatments. Based on our results, we will discuss the possible bias of the included models and its impact from the four important aspects of constructing models specified by the PROBAST tool, and we will suggest future research directions.
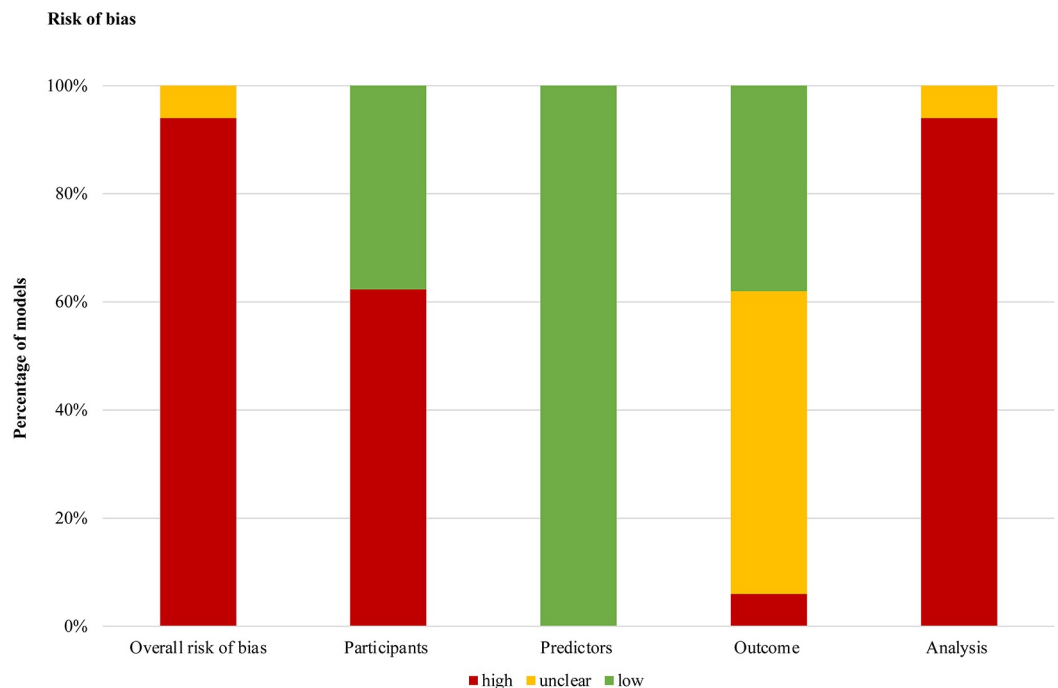

Fig 2. Risk of bias summary. Fig 2 shows the percentage risk of bias ratings for each aspect of the included models according to the PROBAST tool." high", "unclear" and "low" represent a high, unclear, and low risk of bias, respectively.

https://doi.org/10.1371/journal.pone.0287308.g002

Participants

Most of the included studies used electronic medical records (EMRs) as the data source for prediction model development; however, the inherent biases of EMRs should be noted [17]. For example, since routine care data are usually measured by general practitioners, measurement definitions may differ between individuals, particularly among multicenter practitioners [28, 29]. While data from randomized control trials are usually the gold standard for collecting data, they may not always reflect the real world due to their narrow inclusion criteria [30]. Thus, leveraging both interventional data from trials and observational data from the real world could be considered for further studies [31, 32]. Furthermore, stroke type and stage were not restricted and classified in some studies we reviewed. Although a larger scope of the target population would increase the generalizability of models, confounders could also increase to some extent. For example, an ambiguous time since stroke might influence the accuracy of prediction models because spontaneous biological recovery efficiency is not considered [33–35]. The recovery potential of a certain treatment for patients might differ between those who are fitters of the "proportion recovery rule" and nonfitters [36]. Consequently, we believe that a well-defined recruitment criterion of participants in original modeling studies should be applied and reported to enhance the model interpretation [37].

Predictors

To date, with the growing interest in predicting stroke rehabilitation outcome, variables such as age, initial motor impairment, stroke severity, biomarkers, and imaging data have been identified as significant factors for predicting stroke outcome [8, 38–40]. The candidate predictors selected in the included models varied. Demographic characteristics and clinical measures including age, sex, side of impairment, and baseline functional stages were commonly selected for analysis. Notably, treatment measures were not included as separate predictors during the variable selection process in most included studies; however, previous studies showed that a predictive model that does not include treatment as a predictor might omit intervention effects, thus leading to an inaccurate outcome [41, 42]. Although a concrete treatment strategy cannot be prospectively obtained before treatment, we recommend that a rehabilitation treatment plan tailored to patients could serve as a predictor in models to inform the potential recovery of individuals. In addition, given that the inconsistency among types of treatments for patients with stroke might increase the heterogeneity of results, we recommend that future studies report the details of structured interventions and facilitate the consistency of interventions.

Outcome

Ideally, the outcome should be independently measured without information from predictors to reduce bias [17]. Due to the natural feature of the existing data source used in the included models, it is unclear whether the measurement of outcome was blindly recorded without information on predictors. Another concern is that nearly all the models included in this review assessed the outcome at post-treatment or discharge as a single endpoint, while other researchers propose that a single endpoint could not fully account for the improvement following rehabilitation if participants were recruited in wide time windows after stroke. The discharge timepoint is also inappropriate since it is often limited by local rehabilitation resources [34, 43]. Thus, we suggest that follow-up endpoints might be obtained to detect the longer-term benefits of a treatment and to ensure that the model's predictive ability is as accurate as possible.

Model analysis

The analysis process, which is also the main source of bias in the included models according to the PROBAST tool, could be improved in several aspects. First, a sufficient sample size for developing models, especially regression models, is usually based on the events per variable (EPV), which could be calculated by the number of candidate predictors [15, 44]. Generally, EPV less than 10 is considered insufficient, while the most adequate EPV is still being debated [45, 46]. The insufficient sample size may lead to overfitting in modeling studies [47–49]. Another aspect concerns how missing data were handled in the included models. Models that excluded patients with incomplete data rather than properly handled missing data might result in a selective sample and thus overestimated model performance [17, 50, 51]. Additionally, among the reviewed models, the most frequent method used during the predictor selection process was backward selection. However, overfitting should be quantified through internal validation if the model was developed based on an insufficient sample size [16, 17]. Previously published models that used univariate analysis to determine predictors should be avoided in future studies since this approach could lead to inaccurate predictor selection [16, 52, 53]. In future studies, researchers could combine both nonstatistical methods and statistical methods to identity the candidate predictors [16, 17].

Moreover, as for the method for developing models, the most frequently used method in the included studies was logistic regression, which is consistent with a recent review, indicating a preference for logistic regression models in this specific field [54]. Other machine learning algorithms, such as support vector machines, neural networks, and nearest neighbors, have only been used in studies published in recent years. Conventional regression models and novel machine learning models each have their own advantages. For example, while regression can enhance the interpretability of a model, its predictive performance may not be as good as that of novel machine learning algorithms, and vice versa [54]. Thus, future studies could explore other interpretability methods to explain the black-box model, such as one of the included studies in our review, which used four Explainable Artificial Intelligence (XAI) approaches to interpret the results of machine learning methods [26]. Finally, as for the model performance, in addition to the discrimination and calibration that should be appropriately assessed, a validation process is also essential to examine the reliability of models. Validation can be divided into internal and external validation. The former method, such as cross-validation and bootstrapping, attempts to quantify the model bias using the same database with model development. External validation aims to quantify any model bias through a database at the new participant level (e.g., from a different country, setting, recruitment time span), which is external to the model development database [15]. Although we mentioned four studies that had conducted external validation, three of them just randomly divided a single database into a development and a validation database, which is criticized as an inefficient external validation form. In this situation, the two split databases may differ by chance, and the sample size would be reduced [17, 29]. As it is increasingly recognized that the model predictive ability might vary across countries, participants, and periods, effective external validation is always recommended to present the possibility of heterogeneity in the predictive model [14, 20, 42, 43].

Implications

With the development of machine learning in the field of medicine, there is a growing interest in the field of stroke rehabilitation. However, the number of high-quality models that meet the reporting rules and can be widely used is still limited, and future model development studies need to improve the quality of models in several ways and report the model development process according to the principle of transparency [55]. It is important to note that in the clinical

setting, predictive models can only be used as a tool to assist physicians in decision-making, and the specific rehabilitation plan for the patient needs to be developed by the physician in the context of the patient's actual condition.

Limitations

This systematic review is limited by small sample sizes and suboptimal data sources for the included models, and thus the reported model performance may be overly optimistic. Moreover, due to large heterogeneity among studies, we did not conduct a meta-analysis, nor did we use quantitative methods to detect publication bias, so the results of this review should be treated with caution. Another limitation is that the rehabilitation treatment administered to patients varies across countries and rehabilitation settings, which may reduce the generalizability of the models.

Conclusions

This review reveals potential gaps between ideal models and current models, and it is exciting to see that the included models have all shown relatively positive performances; however, existing modeling studies are constrained by small sample sizes and inconsistent results, indicating that there is still room for improvement. We believe that data sharing and coordinated efforts among countries could help future research in this area. Furthermore, as the number of proven significant predictors grows, prediction models should be dynamically updated. Applicable and reliable prediction models should help clinicians improve the implementation of patient-specific stroke rehabilitation treatment.

Supporting information

S1 Checklist. PRISMA 2020 checklist.
(DOCX)

S1 Table. Example of search strategy in PubMed. This is an example search strategy for PubMed.
(DOCX)

S2 Table. Data extraction sheet. This is the details of the data extraction sheet.
(DOCX)

Author Contributions

Data curation: Wanting Zu, Xuemiao Huang, Tianxin Xu, Lin Du, Yiming Wang.

Formal analysis: Wanting Zu, Xuemiao Huang.

Funding acquisition: Wenbo Nie.

Methodology: Xuemiao Huang, Lin Du, Yiming Wang.

Project administration: Lisheng Wang.

Software: Wanting Zu, Tianxin Xu, Yiming Wang.

Supervision: Lisheng Wang.

Validation: Wenbo Nie.

Writing – original draft: Wanting Zu.

Writing – review & editing: Lisheng Wang, Wenbo Nie.

References

1. Johnson CO, Nguyen M, Roth GA, Nichols E, Alam T, Abate D, et al. Global, regional, and national burden of stroke, 1990–2016: a systematic analysis for the Global Burden of Disease Study 2016. Lancet Neurol. 2019 May; 18(5):439–58. https://doi.org/10.1016/S1474-4422(19)30034-1 PMID: 30871944

2. Virani SS, Alonso A, Aparicio HJ, Benjamin EJ, Bittencourt MS, Callaway CW, et al. Heart Disease and Stroke Statistics—2021 Update: A Report From the American Heart Association. Circulation [Internet]. 2021 Feb 23 [cited 2022 Jan 29]; 143(8). Available from: https://www.ahajournals.org/doi/ https://doi.org/10.1161/CIR.0000000000000950 PMID: 33501848

3. Graef P, Michaelsen SM, Dadalt MLR, Rodrigues DAMS, Pereira F, Pagnussat AS. Effects of functional and analytical strength training on upper-extremity activity after stroke: a randomized controlled trial. Braz J Phys Ther. 2016 Dec; 20(6):543–52. https://doi.org/10.1590/bjpt-rbf.2014.0187 PMID: 27683837

4. Winstein CJ, Wolf SL, Dromerick AW, Lane CJ, Nelsen MA, Lewthwaite R, et al. Effect of a Task-Oriented Rehabilitation Program on Upper Extremity Recovery Following Motor Stroke: The ICARE Randomized Clinical Trial. JAMA. 2016 Feb 9; 315(6):571. https://doi.org/10.1001/jama.2016.0276 PMID: 26864411

5. Bergmann J, Krewer C, Jahn K, Müller F. Robot-assisted gait training to reduce pusher behavior: A randomized controlled trial. Neurology. 2018 Oct 2; 91(14):e1319–27. https://doi.org/10.1212/WNL.0000000000006276 PMID: 30171076

6. Teasell RW, Murie Fernandez M, McIntyre A, Mehta S. Rethinking the Continuum of Stroke Rehabilitation. Arch Phys Med Rehabil. 2014 Apr; 95(4):595–6. https://doi.org/10.1016/j.apmr.2013.11.014 PMID: 24529594

7. 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 Dec; 3(1):121. https://doi.org/10.1038/s41746-020-00328-w PMID: 33024831

8. Coupar F, Pollock A, Rowe P, Weir C, Langhorne P. Predictors of upper limb recovery after stroke: a systematic review and meta-analysis. Clin Rehabil. 2012 Apr; 26(4):291–313.

9. Jordan MI, Mitchell TM. Machine learning: Trends, perspectives, and prospects. Science. 2015 Jul 17; 349(6245):255–60. https://doi.org/10.1126/science.aaa8415 PMID: 26185243

10. Mainali S, Darsie ME, Smetana KS. Machine Learning in Action: Stroke Diagnosis and Outcome Prediction. Front Neurol. 2021 Dec 6; 12:734345. https://doi.org/10.3389/fneur.2021.734345 PMID: 34938254

11. Jiang F, Jiang Y, Zhi H, Dong Y, Li H, Ma S, et al. Artificial intelligence in healthcare: past, present and future. Stroke Vasc Neurol. 2017 Dec; 2(4):230–43. https://doi.org/10.1136/svn-2017-000101 PMID: 29507784

12. Heo J, Yoon JG, Park H, Kim YD, Nam HS, Heo JH. Machine Learning–Based Model for Prediction of Outcomes in Acute Stroke. Stroke. 2019 May; 50(5):1263–5. https://doi.org/10.1161/STROKEAHA.118.024293 PMID: 30890116

13. Liberati A, Altman DG, Tetzlaff J, Mulrow C, Gøtzsche PC, Ioannidis JPA, et al. The PRISMA Statement for Reporting Systematic Reviews and Meta-Analyses of Studies That Evaluate Health Care Interventions: Explanation and Elaboration. PLoS Med. 2009 Jul 21; 6(7):e1000100. https://doi.org/10.1371/journal.pmed.1000100 PMID: 19621070

14. Moons KGM, de Groot JAH, Bouwmeester W, Vergouwe Y, Mallett S, Altman DG, et al. Critical Appraisal and Data Extraction for Systematic Reviews of Prediction Modelling Studies: The CHARMS Checklist. PLoS Med. 2014 Oct 14; 11(10):e1001744. https://doi.org/10.1371/journal.pmed.1001744 PMID: 25314315

15. Wolff RF, Moons KGM, Riley RD, Whiting PF, Westwood M, Collins GS, et al. PROBAST: A Tool to Assess the Risk of Bias and Applicability of Prediction Model Studies. Ann Intern Med. 2019 Jan 1; 170(1):51. https://doi.org/10.7326/M18-1376 PMID: 30596875

16. Steyerberg EW. Clinical Prediction Models: A Practical Approach to Development, Validation, and Updating [Internet]. Cham: Springer International Publishing; 2019 [cited 2022 Jan 18]. (Statistics for Biology and Health). Available from: http://link.springer.com/10.1007/978-3-030-16399-0

17. Moons KGM, Wolff RF, Riley RD, Whiting PF, Westwood M, Collins GS, et al. PROBAST: A Tool to Assess Risk of Bias and Applicability of Prediction Model Studies: Explanation and Elaboration. Ann Intern Med. 2019 Jan 1; 170(1):W1. https://doi.org/10.7326/M18-1377 PMID: 30596876

18. García-Rudolph A, Bernabeu M, Cegarra B, Saurı J, Madai VI, Frey D, et al. Predictive models for independence after stroke rehabilitation: Maugeri external validation and development of a new model. NeuroRehabilitation. 2021; 49(3):415–424. https://doi.org/10.3233/NRE-201619 PMID: 34542037.

19. Tozlu C, Edwards D, Boes A, Labar D, Tsagaris KZ, Silverstein J, et al. Machine Learning Methods Predict Individual Upper-Limb Motor Impairment Following Therapy in Chronic Stroke. Neurorehabil Neural Repair. 2020 May; 34(5):428–39. https://doi.org/10.1177/1545968320909796 PMID: 32193984

20. Thakkar HK, Liao W wen, Wu C yi, Hsieh YW, Lee TH. Predicting clinically significant motor function improvement after contemporary task-oriented interventions using machine learning approaches. J NeuroEngineering Rehabil. 2020 Dec; 17(1):131. https://doi.org/10.1186/s12984-020-00758-3 PMID: 32993692

21. 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 Dec; 48(12):3308–15. https://doi.org/10.1161/STROKEAHA.117.018058 PMID: 29051222

22. Bates BE, Xie D, Kwong PL, Kurichi JE, Cowper Ripley D, Davenport C, et al. Development and Validation of Prognostic Indices for Recovery of Physical Functioning Following Stroke: Part 1. PM&R. 2015 Jul; 7(7):685–98.

23. Bates BE, Xie D, Kwong PL, Kurichi JE, Ripley DC, Davenport C, et al. Development and Validation of Prognostic Indices for Recovery of Physical Functioning Following Stroke: Part 2. PM&R. 2015 Jul; 7 (7):699–710. https://doi.org/10.1016/j.pmrj.2015.01.012 PMID: 25633635

24. Harari Y O'Brien MK, Lieber RL, Jayaraman A. Inpatient stroke rehabilitation: prediction of clinical outcomes using a machine-learning approach. J NeuroEngineering Rehabil. 2020 Dec; 17(1):71. https://doi.org/10.1186/s12984-020-00704-3 PMID: 32522242

25. Goffredo M, Proietti S, Pournajaf S, Galafate D, Cioeta M, Le Pera D, et al. Baseline robot-measured kinematic metrics predict discharge rehabilitation outcomes in individuals with subacute stroke. Front Bioeng Biotechnol. 2022 Dec 6; 10:1012544. https://doi.org/10.3389/fbioe.2022.1012544 PMID: 36561043

26. Gandolfi M, Boscolo Galazzo I, Gasparin Pavan R, Cruciani F, Vale N, Picelli A, et al. eXplainable AI Allows Predicting Upper Limb Rehabilitation Outcomes in Sub-Acute Stroke Patients. IEEE J Biomed Health Inform. 2023 Jan; 27(1):263–73. https://doi.org/10.1109/JBHI.2022.3220179 PMID: 36343005

27. Lee JJ, Shin JH. Predicting Clinically Significant Improvement After Robot-Assisted Upper Limb Rehabilitation in Subacute and Chronic Stroke. Front Neurol. 2021 Jul 1; 12:668923. https://doi.org/10.3389/fneur.2021.668923 PMID: 34276535

28. Herrett E, Gallagher AM, Bhaskaran K, Forbes H, Mathur R, van Staa T, et al. Data Resource Profile: Clinical Practice Research Datalink (CPRD). Int J Epidemiol. 2015 Jun; 44(3):827–36. https://doi.org/10.1093/ije/dyv098 PMID: 26050254

29. Riley RD, Ensor J, Snell KIE, Debray TPA, Altman DG, Moons KGM, et al. External validation of clinical prediction models using big datasets from e-health records or IPD meta-analysis: opportunities and challenges. BMJ. 2016 Jun 22; i3140. https://doi.org/10.1136/bmj.i3140 PMID: 27334381

30. Franklin JM, Schneeweiss S. When and How Can Real World Data Analyses Substitute for Randomized Controlled Trials?: Real world evidence and RCTs. Clin Pharmacol Ther. 2017 Dec; 102(6):924–33.

31. Bica I, Alaa AM, Lambert C, van der Schaar M. From Real-World Patient Data to Individualized Treatment Effects Using Machine Learning: Current and Future Methods to Address Underlying Challenges. Clin Pharmacol Ther. 2021 Jan; 109(1):87–100. https://doi.org/10.1002/cpt.1907 PMID: 32449163

32. Johansson FD, Collins JE, Yau V, Guan H, Kim SC, Losina E, et al. Predicting Response to Tocilizumab Monotherapy in Rheumatoid Arthritis: A Real-world Data Analysis Using Machine Learning. J Rheumatol. 2021 Sep; 48(9):1364–70. https://doi.org/10.3899/jrheum.201626 PMID: 33934070

33. Stinear CM. Prediction of motor recovery after stroke: advances in biomarkers. Lancet Neurol. 2017 Oct; 16(10):826–36. https://doi.org/10.1016/S1474-4422(17)30283-1 PMID: 28920888

34. Stinear CM, Lang CE, Zeiler S, Byblow WD. Advances and challenges in stroke rehabilitation. Lancet Neurol. 2020 Apr; 19(4):348–60. https://doi.org/10.1016/S1474-4422(19)30415-6 PMID: 32004440

35. Stinear CM, Byblow WD, Ackerley SJ, Smith MC, Borges VM, Barber PA. Proportional Motor Recovery After Stroke: Implications for Trial Design. Stroke. 2017 Mar; 48(3):795–8. https://doi.org/10.1161/STROKEAHA.116.016020 PMID: 28143920

36. Winters C, van Wegen EEH, Daffertshofer A, Kwakkel G. Generalizability of the Proportional Recovery Model for the Upper Extremity After an Ischemic Stroke. Neurorehabil Neural Repair. 2015 Aug; 29 (7):614–22. https://doi.org/10.1177/1545968314562115 PMID: 25505223

37. Debray TPA, Vergouwe Y, Koffijberg H, Nieboer D, Steyerberg EW, Moons KGM. A new framework to enhance the interpretation of external validation studies of clinical prediction models. J Clin Epidemiol. 2015 Mar; 68(3):279–89. https://doi.org/10.1016/j.jclinepi.2014.06.018 PMID: 25179855

38. Almenkerk S van, Smalbrugge M, Depla MFIA, Eefsting JA, Hertogh CMPM. What predicts a poor outcome in older stroke survivors? A systematic review of the literature. Disabil Rehabil. 2013 Oct; 35 (21):1774–82. https://doi.org/10.3109/09638288.2012.756941 PMID: 23350761

39. Boyd LA, Hayward KS, Ward NS, Stinear CM, Rosso C, Fisher RJ, et al. Biomarkers of stroke recovery: Consensus-based core recommendations from the Stroke Recovery and Rehabilitation Roundtable. Int J Stroke. 2017 Jul; 12(5):480–93. https://doi.org/10.1177/1747493017714176 PMID: 28697711

40. Westlake KP, Nagarajan SS. Functional Connectivity in Relation to Motor Performance and Recovery After Stroke. Front Syst Neurosci [Internet]. 2011 [cited 2022 Jan 25]; 5. Available from: http://journal.frontiersin.org/article/10.3389/fnsys.2011.00008 PMID: 21441991

41. Groenwold RHH, Moons KGM, Pajouheshnia R, Altman DG, Collins GS, Debray TPA, et al. Explicit inclusion of treatment in prognostic modeling was recommended in observational and randomized settings. J Clin Epidemiol. 2016 Oct; 78:90–100. https://doi.org/10.1016/j.jclinepi.2016.03.017 PMID: 27045189

42. Schuit E, Groenwold RHH, Harrell FE, de Kort WLAM, Kwee A, Mol BWJ, et al. Unexpected predictor–outcome associations in clinical prediction research: causes and solutions. Can Med Assoc J. 2013 Jul 9; 185(10):E499–505. https://doi.org/10.1503/cmaj.120812 PMID: 23339155

43. Stinear CM, Smith MC, Byblow WD. Prediction Tools for Stroke Rehabilitation. Stroke. 2019 Nov; 50 (11):3314–22. https://doi.org/10.1161/STROKEAHA.119.025696 PMID: 31610763

44. Peduzzi P, Concato J, Feinstein AR, Holford TR. Importance of events per independent variable in proportional hazards regression analysis II. Accuracy and precision of regression estimates. J Clin Epidemiol. 1995 Dec; 48(12):1503–10.

45. van Smeden M, de Groot JAH, Moons KGM, Collins GS, Altman DG, Eijkemans MJC, et al. No rationale for 1 variable per 10 events criterion for binary logistic regression analysis. BMC Med Res Methodol. 2016 Dec; 16(1):163. https://doi.org/10.1186/s12874-016-0267-3 PMID: 27881078

46. Ogundimu EO, Altman DG, Collins GS. Adequate sample size for developing prediction models is not simply related to events per variable. J Clin Epidemiol. 2016 Aug; 76:175–82. https://doi.org/10.1016/j.jclinepi.2016.02.031 PMID: 26964707

47. van der Ploeg T, Austin PC, Steyerberg EW. Modern modelling techniques are data hungry: a simulation study for predicting dichotomous endpoints. BMC Med Res Methodol. 2014 Dec; 14(1):137. https://doi.org/10.1186/1471-2288-14-137 PMID: 25532820

48. Peduzzi P, Concato J, Kemper E, Holford TR, Feinstein AR. A simulation study of the number of events per variable in logistic regression analysis. J Clin Epidemiol. 1996 Dec; 49(12):1373–9. https://doi.org/10.1016/s0895-4356(96)00236-3 PMID: 8970487

49. Vittinghoff E, McCulloch CE. Relaxing the Rule of Ten Events per Variable in Logistic and Cox Regression. Am J Epidemiol. 2007 Jan 12; 165(6):710–8. https://doi.org/10.1093/aje/kwk052 PMID: 17182981

50. Begg CB, Greenes RA, Iglewicz B. The influence of uninterpretability on the assessment of diagnostic tests. J Chronic Dis. 1986 Jan; 39(8):575–84. https://doi.org/10.1016/0021-9681(86)90182-7 PMID: 3090089

51. Shinkins B, Thompson M, Mallett S, Perera R. Diagnostic accuracy studies: how to report and analyse inconclusive test results. BMJ. 2013 May 16; 346(may16 2):f2778–f2778. https://doi.org/10.1136/bmj.f2778 PMID: 23682043

52. Harrell FE, Lee KL, Mark DB. MULTIVARIABLE PROGNOSTIC MODELS: ISSUES IN DEVELOPING MODELS, EVALUATING ASSUMPTIONS AND ADEQUACY, AND MEASURING AND REDUCING ERRORS. Stat Med. 1996 Feb 29; 15(4):361–87. https://doi.org/10.1002/(SICI)1097-0258(19960229)15:4<361::AID-SIM168>3.0.CO;2-4 PMID: 8668867

53. Sun GW, Shook TL, Kay GL. Inappropriate use of bivariable analysis to screen risk factors for use in multivariable analysis. J Clin Epidemiol. 1996 Aug; 49(8):907–16. https://doi.org/10.1016/0895-4356(96)00025-x PMID: 8699212

54. Bonkhoff AK, Grefkes C. Precision medicine in stroke: towards personalized outcome predictions using artificial intelligence. Brain. 2022 Apr 18; 145(2):457–475. https://doi.org/10.1093/brain/awab439 PMID: 34918041; PMCID: PMC9014757.

55. Collins GS, Reitsma JB, Altman DG, Moons KGM. Transparent Reporting of a multivariable prediction model for Individual Prognosis Or Diagnosis (TRIPOD): The TRIPOD Statement. Ann Intern Med. 2015 Jan 6; 162(1):55–63. https://doi.org/10.7326/M14-0697 PMID: 25560714