Human-Centric Intelligent Systems (2024) 4:147–160
https://doi.org/10.1007/s44230-023-00051-1

REVIEW ARTICLE

Machine Learning Algorithms for the Prediction of Language and Cognition Rehabilitation Outcomes of Post-stroke Patients: A Scoping Review

Kyriakos Apostolidis[1] · Christos Kokkotis[1] · Serafeim Moustakidis[1] · Evangelos Karakasis[1] · Paraskevi Sakellari[1] · Christina Koutra[1] · Dimitrios Tsiptsios[2] · Stella Karatzetzou[2] · Konstantinos Vadikolias[2] · Nikolaos Aggelousis[1]

Received: 4 May 2023 / Accepted: 3 November 2023 / Published online: 9 December 2023

Abstract

Stroke is one of the leading causes of long-term disabilities in motor and cognition functionality. An early and accurate prediction of rehabilitation outcomes can lead to a tailor-made treatment that can significantly improve the post-stroke quality of life of a person. This scoping review aimed to summarize studies that use Artificial Intelligence (AI) for the prediction of language and cognition rehabilitation outcomes and the need to use AI in this domain. This study followed the PRISMA-ScR guidelines for two databases, Scopus and PubMed. The results, which are measured with several metrics depending on the task, regression, or classification, present encouraging outcomes as they can predict the cognitive functionality of post-stroke patients with relative precision. Among the results of the paper are the identification of the most effective Machine Learning (ML) algorithms, and the identification of the key factors that influence rehabilitation outcomes. The majority of studies focus on aphasia and present high performance achieving up to 97% recall and 91.4% precision. The main limitations of the studies were the small subject population and the lack of an external dataset. However, effective ML algorithms along with explainability are expected to become among the most prominent solutions for precision medicine due to their ability to overcome non-linearities on data and provide insights and transparent predictions that can help healthcare professionals make more informed and accurate decisions.

Keywords Stroke · Aphasia · Cognitive · Artificial intelligence · Prognosis

Abbreviations

AI	Artificial intelligence
ML	Machine learning
DL	Deep learning
DNN	Deep neural networks
RL	Reinforcement learning
MSE	Mean squared error
RMSE	Root mean squared error
CBF	Cerebral blood flow
FA	Fractional anisotropy

✉ Nikolaos Aggelousis
nagelous@phyed.duth.gr

Kyriakos Apostolidis
kyriapos1@cs.ihu.gr

Christos Kokkotis
ckokkoti@affil.duth.gr

Serafeim Moustakidis
s.moustakidis@aideas.eu

Evangelos Karakasis
ekarakas@pme.duth.gr

Paraskevi Sakellari
psakella@phyed.duth.gr

Christina Koutra
ckoutra@phyed.duth.gr

Dimitrios Tsiptsios
dtsiptsi@med.duth.gr

Stella Karatzetzou
skaratze@med.duth.gr

Konstantinos Vadikolias
kvadikol@med.duth.gr

[1] Department of Physical Education and Sport Science, Democritus University of Thrace, 69100 Komotini, Greece

[2] Department of Neurology, School of Medicine, University Hospital of Alexandroupolis, Democritus University of Thrace, 68100 Alexandroupolis, Greece

fMRI	Functional magnetic resonance imaging
rsf-MRI	Resting state functional magnetic resonance imaging
CV	Cross-validation
RF	Random forest
SVM	Support vector machine
SVR	Support vector regression
RFE	Recursive feature elimination
SOM	Self-organizing map
AAT	Aachen aphasia test
LOOCV	Leave one out cross validation
MADP	Mean absolute deviation percentage
AUC	Area under curve
MRI	Magnetic resonance imaging
SHAP	SHapley Additive exPlanations
IQ	Intelligence quotient
WAB-R	Western Aphasia Battery-Revised
MMSE	Mini-Mental State Examination
BI	Barthel Index
FIM	Functional Independence Measure
LRS	Language Recovery Score
LUQ	Language Use Questionnaire
BNT	Boston Naming Tests
PAPT	Pyramids and Palms Trees test
PNT	Philadelphia Naming Test
fALLF	Fractional amplitude of low-frequency fluctuations

1 Introduction

1.1 Backdrop

Stroke is not only the second greatest cause of death among adults, but also the primary cause of acquired disability, with a significant negative impact on long-term functional independence of stroke survivors [1]. Given the age-related nature of the disease and the fact that nearly two-thirds of stroke patients are over 65, it is anticipated that the overall burden of stroke will significantly increase, along with the number of stroke survivors, as a result of ongoing global population growth and significant improvements in life expectancy [2].

Aphasia, defined as an acquired impairment in language production and/or comprehension, is among the most common complications of stroke affecting 21–38% of acute stroke individuals [3]. Cognition, a term referring to the mental processes involved in gaining knowledge and comprehension, such as thinking, knowing, remembering, judging, and problem-solving, is also frequently affected as rates of cognitive impairment post-stroke range from 35 to 92% [4]. Aphasia and cognitive impairment both have drawbacks that can include hampered daily living activities, lost pay, higher health care costs, loss of freedom, and social isolation [3]. A person's post-stroke quality of life can be greatly enhanced by a treatment that is specifically designed for them if language and cognition rehabilitation outcomes can be predicted early and accurately.

AI is widely used in the medical field to create precision medicine, aiming to enable personalized diagnoses and treatments for each patient. The era of big data has allowed AI algorithms to predict and diagnose diseases with equal and sometimes superior accuracy than humans. The diagnosis and treatment of stroke require a vast amount of data and multi-disciplinary approaches, making it convenient for precision medicine [5]. In this study, we focus on post-stroke cognitive functionality prediction with ML algorithms, as it is under-represented in relation to motor function recovery after stroke.

1.2 Prior Research

According to Horn et al. [6], early prediction or diagnosis is crucial, as treatment is more effective when applied early after a stroke because the first few weeks after the stroke are the most important for brain reorganization. Therefore, an accurate prediction of post-stroke language outcomes is significant in order for rehabilitation therapy to meet individual needs [7]. To do this, the relationship between the data like neuroimaging, clinical, etc. must be found. Classical models used for statistical analysis such as ANOVAs [8], t-tests [9], etc. are limited to finding only linear relationships [10]. However, some studies [11, 12] have shown that brain-behavior relationships are not always linear. The majority of ML algorithms are able to find non-linear patterns for more effective solutions, which is why AI is an essential tool in modern medicine [13].

Research in this particular field faces a considerable constraint in terms of its scope and depth, primarily due to the limited number of studies available. Moreover, these studies often grapple with a range of challenges, the most prominent being the scarcity of comprehensive and standardized datasets. These collective limitations inevitably pose a formidable hurdle when it comes to drawing unequivocal and far-reaching conclusions about the current landscape of this field. Nonetheless, it is noteworthy that certain trends and patterns have emerged within these studies that merit attention. Notably, those studies that have managed to achieve higher performance metrics appear to favor the utilization of advanced techniques such as deep learning (DL) models [14], support vector machines (SVM) [15], and random forests (RF) [10]. This preference suggests that these sophisticated algorithms hold promise in extracting valuable insights from the data available in this domain. Furthermore, it is worth highlighting the intriguing observation that surveys and investigations specifically employing neuroimaging data

exhibit a tendency to yield more favorable and promising results compared to their counterparts that rely on alternative types of data. This underscores the pivotal role played by neuroimaging images, as they seem to possess a unique capacity for revealing critical information that is particularly relevant and valuable in this field of study.

1.3 Aim of the Study

Overall, AI has the potential to greatly improve the quality of life for people who have experienced a stroke. By predicting the likelihood of stroke, making accurate diagnoses, and predicting the likelihood of complications, AI can help healthcare professionals to provide better care and support for stroke survivors, ultimately leading to improved outcomes and a better quality of life. Hence, the aim of the present scoping review is to investigate the current state-of-the-art studies that deploy ML and Deep Learning (DL) algorithms for language and cognition rehabilitation outcomes after stroke [5, 6]. Moreover, it intends to highlight the data and models used for this purpose and models' capabilities. The review will examine the use of various types of data, including neuroimaging, demographic, and clinical data, and will identify the most effective machine learning algorithms and key factors that influence rehabilitation outcomes. Additionally, this review aims to understand the need of AI in this domain, and explore the potential benefits and limitations of using AI in the prediction of language and cognitive rehabilitation outcomes in post-stroke patients. Furthermore, our scoping review distinguishes itself from existing literature in several ways. Firstly, we address a notable gap in the field by concentrating on the prediction of cognitive outcomes following strokes through the utilization of artificial intelligence. Most previous studies in this domain have primarily focused on predicting post-stroke impairments [16] or exploring the application of artificial intelligence in aphasia rehabilitation [17]. Our research not only sheds light on an underexplored area but also contributes to a more comprehensive understanding of the predictive capabilities of AI in the context of stroke recovery. Finally, we ultimately opted for conducting a scoping review due to its suitability in the context of mapping a continuously evolving and expansive research domain, as well as its effectiveness in pinpointing existing gaps and unanswered questions within the field.. Additionally, our primary aim is not centered around conducting quantitative accuracy comparisons due to the inherent diversity within the available dataset. Instead, our research is oriented towards achieving more qualitative objectives. These qualitative objectives pertain to the evaluation of various methodologies employed, the thorough examination of clinical parameters under study, and a comprehensive assessment of the conducted evaluations. In essence, our study places greater emphasis on qualitative aspects rather than quantitative accuracy comparisons, as we believe these facets offer a richer understanding of the research landscape.

1.4 Artificial Intelligence in a Nutshell

Artificial Intelligence encapsulates the realm of computer science dedicated to creating machines and systems that mimic human intelligence. By leveraging techniques like machine learning, neural networks, and natural language processing, AI enables computers to perform tasks that typically require human cognitive functions, such as problem-solving, decision-making, and language understanding.

ML is a subfield of artificial intelligence, and its main goal is to simulate human behavior using data. ML algorithms are divided into three categories: supervised, unsupervised, and reinforcement learning (RL). In the first category, the data used for training are labeled while in the second, the data are unlabeled and the algorithms look for patterns. In RL, machines are trained through trial-and-error procedures in order to take action. Supervised ML consists of two phases, the learning and testing phases. Predicting cognitive functionality uses supervised machine learning as there is a plethora of labeled data for training and algorithms find patterns to perform best in the testing phase. The most prevalent ML algorithms are classification and regression models for supervised learning and the clustering models [18] for unsupervised learning. In the classification, the ML algorithm classifies an input into specific discrete values. On the contrary, regression models predict a continuous value for a specific input. In clustering models, unlabeled data are given to the algorithm and then it learns to group the similar data points. A traditional ML system starts with the feature selection stage to extract the most useful features of the given data which are then fed into the model. Next, during the training, a validation strategy is applied for the optimization of the model's parameters.

Deep Learning is a subfield of artificial intelligence and machine learning where complex neural networks simulate the human brain's intricate learning process. Utilizing multiple layers of interconnected nodes, deep learning algorithms unearth patterns, representations, and features from vast datasets, enabling remarkable breakthroughs in image and speech recognition, natural language processing, and more. This approach's power lies in its capacity to autonomously learn from data, gradually enhancing its performance with each iteration. As technology evolves, the concise yet profound concept of deep learning continues to reshape industries, pushing the boundaries of what machines can achieve.

Numerous metrics were applied to the models to assess their performance [19]. For regression problems were mainly used the mean squared error (MSE), the root mean squared error (RMSE), and R^2. The MSE computes the average of

the squares of the difference between the estimated and actual value. The RMSE computes the root of MSE while R^2 is the difference between the samples in the dataset and the predictions made by the model and it ranges between 0 and 1. Also, Accuracy, F1 score, Recall and Precision were used for classification problems. These metrics consist of four values, True Positive (TP), True Negative (TN), False Positive (FP) and False Negative.

$$Accuracy = \frac{TP + TN}{TP + TN + FP + FN} \tag{1}$$

$$Recall = \frac{TP}{TP + FN} \tag{2}$$

$$Presision = \frac{TP}{TP + FP} \tag{3}$$

$$F1score = \frac{2 * TP}{2 * TP + FP + FN} \tag{4}$$

1.5 Rehabilitation and AI

The combination of rehabilitation and AI utilizes cutting-edge technology in the field of healthcare, improving the way individuals recover from injuries and medical conditions [20]. By integrating artificial intelligence into rehabilitation processes, personalized treatment plans can be created based on a patient's unique needs and progress [21]. AI-powered devices and systems facilitate real-time monitoring, data analysis, and adaptive adjustments to therapy regimens [22]. Whether it's using wearable sensors to track movement, virtual reality simulations for cognitive rehabilitation, or robotic assistance for physical therapy, the synergy between rehabilitation and AI holds the potential to enhance patient outcomes, accelerate recovery, and improve the quality of patients' lives.

1.6 Cognitive Functionality Indicators

The assessment of language outcomes after stroke is not an easy task; therefore, several indicators have been proposed to quantify cognitive functionality objectively. The majority of predictive models use these indicators for both training and prediction. The most used indicator was the Western Aphasia Battery-Revised (WAB-R) which is responsible for the evaluation of linguistic skills, mainly affected by aphasia. In Table 1, indicators used in the papers we studied with corresponding abbreviations are presented.

2 Materials and Methods

The 22-item Preferred Reporting Items for Systematic reviews and Meta-Analyses extension for Scoping Reviews (PRISMA-ScR) was used for the present scoping review [31]. Our study's methods were designed a priori.

2.1 Research Strategy

Structured literature research in two well-known and relative databases (PubMed and Science Direct) was conducted by two investigators up to 30 August 2023 using five criteria, to ensure a comprehensive coverage of the literature. The first includes the keywords "aphasia" or "cognitive" or "language", the second the keywords "prediction" or "prognosis", the third the keywords "rehabilitation" or "therapy" or "physiotherapy" or "speech", the fourth the keywords "machine learning" or "deep learning" or "artificial intelligence" while the fifth contains "stroke" or "brain ischemia" or "cerebral ischemi*" or "post stroke" or poststroke. The

Table 1 Abbreviations and Acronyms of indicators

Abbreviation	Acronyms	Description
WAB-R [23]	Western Aphasia Battery-Revised	WAB-R assesses language skills after aphasia
MMSE [24]	Mini-Mental State Examination	MMSE tests cognitive function including concentration, orientation, verbal memory, naming, attention, and visuospatial skills
BI [25]	Barthel Index	BI measures the ability of a subject to function independently
FIM [25]	Functional Independence Measure	FIM assesses independence for self-care
LRS [7]	Language Recovery Score	LRS is an indicator that represents the overall language impairments
LUQ [26]	Language Use Questionnaire	LUQ provides information about the history of language exposure. (For multilingual subjects)
BNT [27]	Boston Naming Tests	BNT is a psychometric tool that aims to assess if a patient can retrieve from his memory the naming of objects in pictures
PAPT [28]	Pyramids and Palms Trees test	PAPT is similar to BNT and it has the same goal
PNT [29]	Philadelphia Naming Test	PNT is a test of 175 pictures that should be named by patients
fALLF [30]	Fractional amplitude of low-frequency fluctuations	fALLF is a neuroimaging method that evaluate brain activity and function

retrieved articles were further screened for potentially relevant articles.

2.2 Selection Criteria

The present study includes only full-text original articles which published in English. Also, conferences, analyses, reviews, case reports, meeting summaries, or unpublished abstracts were excluded. There was no restriction on study design or sample characteristics. The retrieved articles were screened in depth to ensure that all relevant articles were included.

2.3 Eligibility Criteria

The applied criteria were the following: (i) papers that predict language or cognitive outcomes after stroke; (ii) studies that applied AI algorithms, including ML or DL algorithms.

2.4 Data Extraction

A custom predefined data form was created in Microsoft to record key variables from each study, aiming to minimize error and standardize our approach. We recorded authors, year of publication, category of patients, application domain (aphasia or cognition), intervention applied on patients, data used for training the algorithms, number and age of participants, indicators used to quantify impairments, AI algorithms along with feature engineering and validation techniques and the results. To enhance the reliability of our data extraction process, two independent reviewers (K.A. and C.K.) assessed the extracted data. In case of disagreement, a third reviewer (S.M.) was consulted to resolve the discrepancies.

2.5 Data Analysis

To evaluate the effectiveness of the proposed methods, we employed a variety of metrics, including Accuracy, Recall, RMSE, and MSE, which are elaborated upon in Sect. 1.1. Additionally, we have incorporated an analysis of model complexity into our survey to facilitate a more objective comparison. In Sect. 3, we provide a comprehensive overview of each study, presenting all relevant metrics alongside the number of subjects and outcome assessments. This approach aims to offer a holistic understanding while simultaneously minimizing potential biases.

3 Results

In total, 95 studies were retrieved: 47 from PubMed and 45 from Science Direct. After excluding duplicated, and irrelevant studies, 16 papers were selected. Finally, after screening the full text of the papers, 9 articles were included. Figure 1 presents the PRISMA flowchart.

The selected studies focus on aphasia severity or prediction and on cognitive functionality prediction.

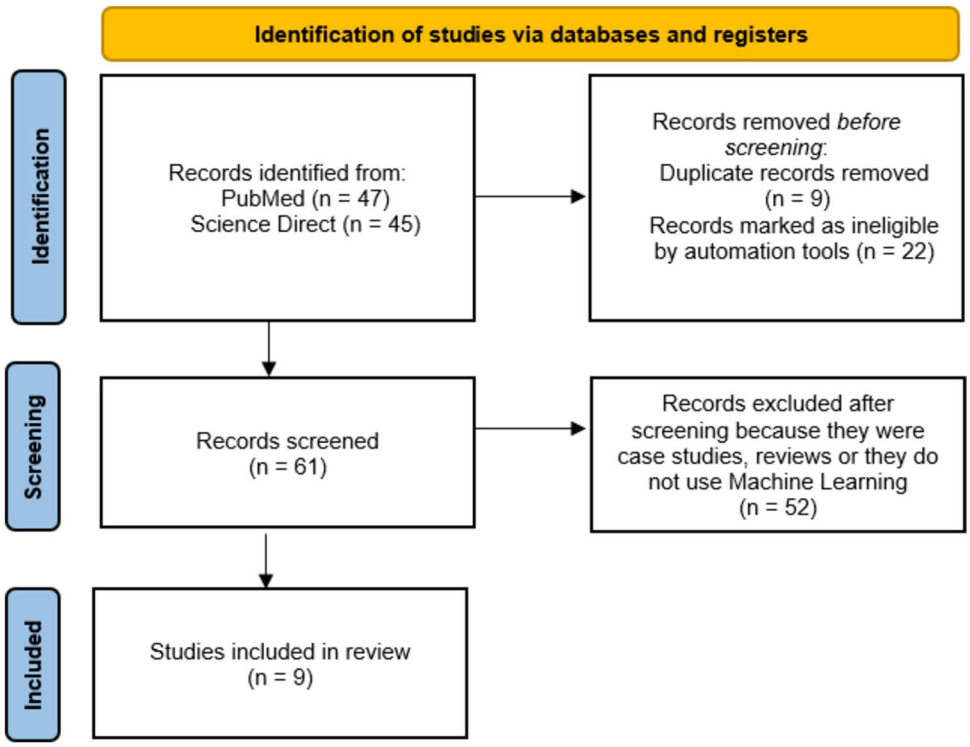

Fig. 1 PRISMA flowchart

3.1 Statistics

In this section, are presented figures with statistics that extracted from the above studies and the Table 2 which summarizes the most useful information of them. Figure 2 shows the trend of AI in prediction of cognitive functionality while Figs. 3 and 4 show the distribution of models and data respectively.

As depicted in Fig. 2, there is a growing interest in the use of ML for predicting cognitive rehabilitation outcomes following stroke. The figure shows that one paper was published in each of the years 2010, 2017, 2018, and 2020, whereas in 2021 and 2022, three and two papers were published, respectively. These findings suggest a recent surge in research efforts focused on utilizing ML to enhance stroke rehabilitation outcomes, which may be attributed to the increased availability of large datasets and advanced computational tools.

Figures 3 and 4 display the various models and types of data utilized in the 9 studies included in our review. It is worth noting that most studies employ multiple models and data types, which is why multiple models are depicted in the figures. For example, one study [32] used 10 different models to provide a comprehensive analysis of cognitive rehabilitation prediction. Additionally, we observed certain patterns among the models and types of data used. All studies that utilized SVMs also used neuroimaging data, and two out of three of these studies also included demographic data. Similarly, RF was commonly used in conjunction with these data types, suggesting that they provide valuable information that these models are well-suited to capture. These observations provide insights into the most effective combinations of models and data types for predicting cognitive rehabilitation outcomes following stroke.

Table 2 presents the results from the studies included in current review. The Table 2 is split into three parts depending on the complexity of models. The first category, "Low Complexity" contains studies that use simple ML models, while the second and the third part contains studies with "Intermediate" and "High" complexity respectively. Studies where, experienced with several ML models, were included in the part that their best model belongs. In the "Low Complexity" category belong K-Nearest Neighbors (KNN) which is a relatively simple algorithm, the ElasticNet which is a linear regression method that combines L1 and L2 regularization techniques and SOM that is an unsupervised and dimensionality reduction method. The "Intermediate Complexity" category includes classical ML models such as SVM, RF that are more complex than simple linear models. In the last category, included complex models like DNNs which consist of multiple layers and nodes. Finally, in each part of the table, studies were sorted by year.

3.2 Aphasia

Sigfus Kristinsson et al. [33] studied the prediction of language outcomes in chronic aphasia using multiple neuroimaging modalities in order to analyze further the neurobiological substrates of aphasia. 116 subjects with chronic left-hemisphere stroke took place in this study. Neuroimaging data were acquired on the Siemens 3 T MRI scanner and consisted of cerebral blood flow (CBF), diffusion-based fractional anisotropy (FA) values, functional magnetic resonance imaging (fMRI), and lesion-load data. Western Aphasia Battery (WAB) was used for the evaluation of language out-comes and aphasia severity using specific sub-scores such as Auditory Comprehension, Naming, Spontaneous Speech, Speech Repetition, Fluency, and Aphasia Quotient. They applied univariate regression analysis to reduce the features in the data and then fed it into a Support Vector Regression (SVR) model to predict language measures. ten-fold cross-validation (CV) was implemented for the evaluation of the model. Experiments were performed with three different combinations of data: (i) lesion volume alone, (ii) each neuroimaging modality alone, and (iii) a blend of all modalities. The latter per-formed the best accuracy with Pearson's correlation 0.53–0.67.

Pustina et al. [6] developed a multimodal framework named STAMP, which is based on three neuroimaging modalities (structural connectivity, lesion maps, and functional connectivity) and aims to predict four aphasia scores. The study consists of 53 left hemispheric chronic post-stroke patients. The proposed framework can be de-scribed by two main elements: Random Forest (RF) and prediction stacking. STAMP achieved high accuracy in all aphasia scores and ranges from 79 to 88%. Moreover, Recursive Feature Elimination (RFE) was applied aiming to remove irrelevant features and identify useful topological information in the brain. Finally, authors concluded that the multimodal perspective could play a significant role in translating neuroimaging research into clinical tools.

Another study that investigated the response to a language treatment was done by Billot et al. [15]. This research includes 55 patients with chronic poststroke aphasia who received 12 weeks of language treatment. Several assessments were implemented on patients, such as the WAB-Revised etc., in order to collect language measurements. They trained two ML models, Support Vector Macnine (SVM) and RF; however, SVM trained on demographics, measures of anatomic integrity, aphasia severity, and resting-state functional connectivity was the most effective achieving 94% F1 score. This study concluded that training a model with a subset of multimodal neuroimaging, behavioral, and demographic data is the most effective method. In addition, the authors stated that resting-state functional connectivity (e.g., rsfMRI), aphasia severity, and anatomical

Table 2 Results from the included studies

Author	Year	Application domain	Intervention	Data	Subjects	Outcome assessment	Feature engeeniring	Machine Learning	Validation	Best Performance
Low complexity										
Michael Iorga et al. [35]	2021	Prediction of language functionality after treatment in speech and language	Treatment in agrammatism or dysgraphia or anomia	Resting state fMRI, behavioral, language measures	57 aphasic patients (chronic)	Prediction of one of the three TSM after three months of therapy	–	ElasticNet	LOOCV	The average performance of the GICA fALFF models is (R^2 = 0.816–0.876)
Uli Grasemann et al. [34]	2021	Prediction of the response in aphasia therapy	Patients received therapy in one of their languages	Clinical and demographic data	13 bilingual aphasic patients	Prediction of the treatment response	–	SOM	LOOCV	R^2 = 0.989 for English naming and 0.974 for Spanish naming
Intermediate complexity										
Helard Becerra Martinez et al. [32]	2022	Predicting cognitive functionality improvement after therapy for post-stroke patients	–	Demographic, cognitive 24 different assessments at admission from 24 different standardized neuropsychology tests	201 patients mean age: 49.51 years	The cognitive outcome was assessed in terms of global improvement	ANOVA f-test	RF Classifier, Extra Trees Classifier K-Neighbors Classifier, XGB Classifier, Logistic Regression, Ridge Classifier, Bagging Classifier, Linear SVC, Linear Discriminant Analysis, Bernoulli NB	fivefold cv	RF classifier Recall: 70.1% F1: 63.8% Precision: 65.2% AUC: 52.6%
Anne Billot et al. [15]	2022	Response assessment on post-stroke language rehabilitation	Language treatment for 12 weeks	Demographic, behavioral and structural and functional neuroimaging data	55 (mean age = 58.8 years. (chronic)	WAB-R (after 12 weeks)	Pearson Correlation	SVM, RF	LOOCV	The best model was SVM with the optimal feature set achieving accuracy (92.7%), F1 (94.1%), precision (91.4%), and recall values (97.%)

Table 2 (continued)

Author	Year	Application domain	Intervention	Data	Subjects	Outcome assessment	Feature engeeniring	Machine Learning	Validation	Best Performance
Sigfus Kristinsson et al. [33]	2021	Prediction of specific language functions and aphasia severity	–	Neuroimaging, diffusion-based fractional anisotropy (FA)-values, lesion-load data, and cerebral blood flow (CBF)	116 Mean age: 58.5 ± 10.9 years. (chronic)	Western Aphasia Battery	Univariate regression analysis	SVR	tenfold cross validation (CV)	The multimodal prediction model yielded the most accurate prediction in all cases (r = 0.53–0.67)
Patrizio Sale et al. [36]	2018	Prediction of improvement on motor and/or cognitive after rehabilitation	Multidisciplinary intensive therapy for 3 h physiotherapy session in a daily basis	Functional and clinical data	55 (sub-acute)	T1 Cognitive FIM, T1 Motor, (FIM) T1 (BI), and T1 Total FIM	Mutual Information Criterion	SVM	(70% training, inside fivefold cv)/ testing (30%)	MADP = 17.55 RMSE = 4.28
Dorian Pustina et al. [10]	2017	Prediction the severity of aphasia after stroke	–	Neuroimaging	53 Mean age: 57.1 ± 12.3 years (chronic)	WAB, PNT	Recursive feature elimination with tenfold split	RF	tenfold CV	RMSE = 0.9 – 15.6
Dorothee Saur et al. [7]	2010	Prediction of language results half a year after stroke	–	fMRI, demographic and behavioral	21 aphasic stroke patients (acute)	Their goal was to estimate the extent of language improvement over half a year	A novel visualization technique in MRIs	SVM	LOOCV	86% accuracy
High complexity										
Kaoru Sakatani et al. [14]	2020	Predicting the cognitive functionality according to blood test items and subject's age	Rehabilitation and medication	MRI, Blood test	202 Mean age 73.48 ± 13.1 years)	Mini Mental State Examination (MMSE)	Multivariate regression analysis	DNNs	Leave-one-out cross-validation (LOOCV)	Sensitivity of 75% and specificity of 87%

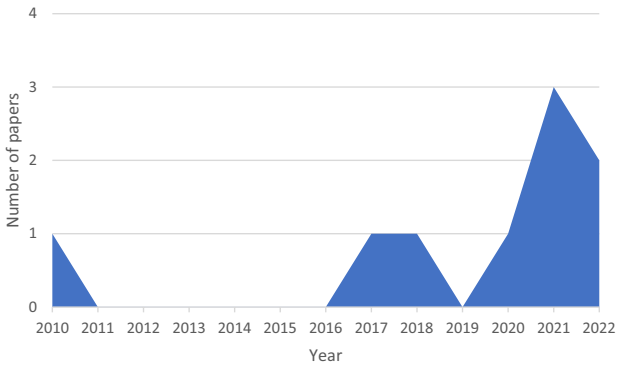

Fig. 2 The trend of AI in cognitive functionality prediction. The x-axis represents the year, while the y-axis displays the number of published studies on this topic over time

integrity were consistently critical predictors of language therapy prognosis.

Saur et al. [5] examined the utility of functional MRI (fMRI) scans to predict language recovery six months after stroke in 21 stroke patients with aphasia. Patients' assessment was held with the Aachen Aphasia Bedside Test and the subtests: naming, repetition, writing, and auditory/speaking from the Aachen Aphasia Test (AAT), An SVM was deployed in order to predict if a patient would have a good or bad language outcome after six months. Experiments performed with fMRIs and fMRIs along with LRS and age. LRS corresponds to a univariate indicator of the overall level of language impairment and allows the separation of patients into a wide range of impairments at all stages after stroke. Additionally, the authors introduced a novel methodology for visualizing the critical points of an fMRI. This helps the model pay attention to points that are most significant for the predicted outcome. The model with fMRIs achieved 76%

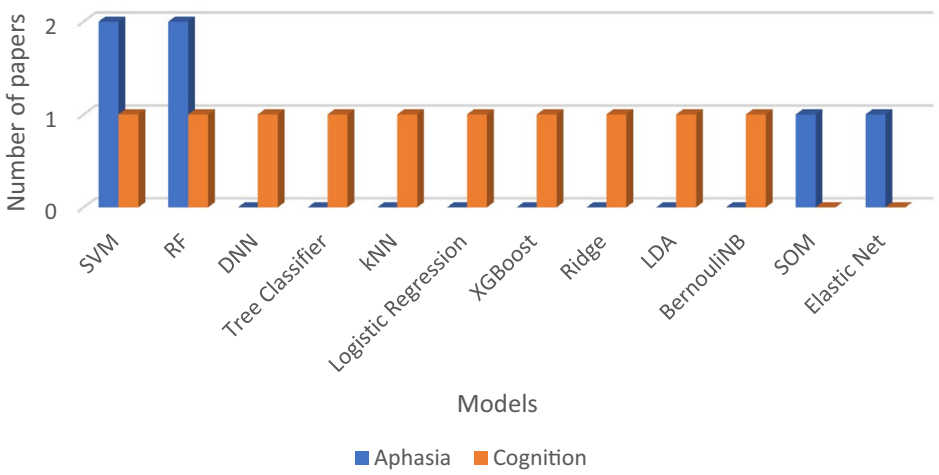

Fig. 3 The prevalence of machine learning models in the reviewed papers. The x-axis displays the various machine learning models, while the y-axis indicates the number of papers that employed each model

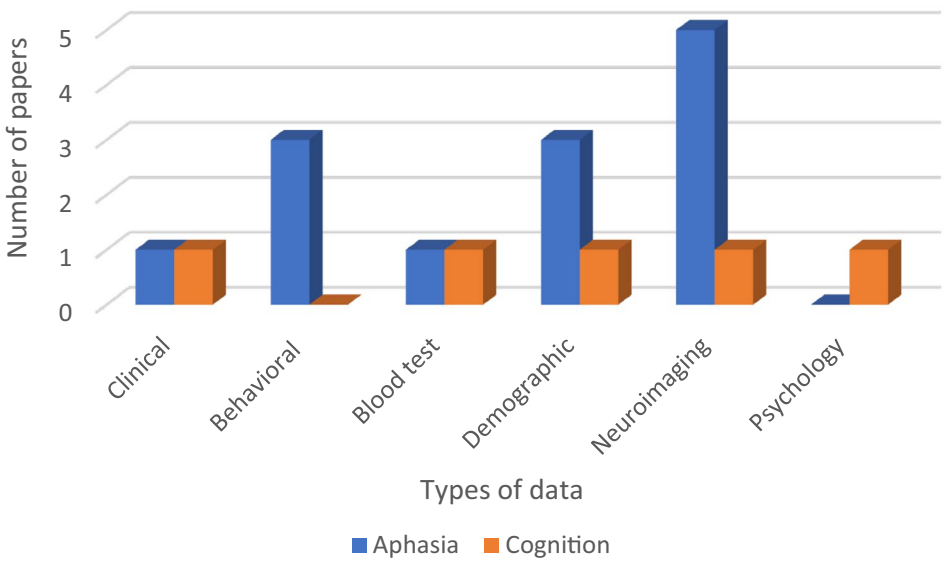

Fig. 4 Bar graph depicting the frequency of data types used in the papers reviewed. The x-axis represents the different types of data, while the y-axis shows the number of papers that utilized each type of data

accuracy, while the model with fMRI + LRS + age achieved 86%.

Graseman et al. [34] investigated the phenomenon of predicting language therapy outcomes in bilingual aphasic people, which is complicated because of manifold pre- and post-stroke factors. A BiLex model, which consisted of Self Organizing Maps (SOM), was deployed in order to simulate language deficits and treatment. The study included 13 bilingual aphasic subjects with native English and Spanish. The model aimed to predict the naming ability after rehabilitation therapy in one of the above languages. Results were encouraging achieving $R^2 = 0.989$ and $R^2 = 0.974$ in English naming and Spanish naming respectively.

Iorga et al. [35] attempted to predict language outcomes after language and speech rehabilitation treatment in one of three aphasia impairments: dysgraphia, agrammatism, or anomia. The study involved 57 chronic aphasic patients who received a three-month therapy in one of the above impairments. The performance of language outcomes was modeled using Elastic Net regression. The data they used were resting state fMRI (rsfMRI) and behavioral language measures. And then they obtained fALFF from rsfMRIs. The results showed that behavioral measures provide high performance in anomia ($R^2 = 0.948$), while fALLF provided high performance in agrammatism ($R^2 = 0.876$) and dysgraphia ($R^2 = 0.822$).

3.3 Cognitive Function

Kaoru Sakatani et al. [14] developed a Deep Neural Network (DNN) in order to predict cognitive function based on blood tests and age. The study included 202 subjects, 65 of whom were patients, who were hospitalized for rehabilitation after a stroke, while 37 and 165 sub-jects were included in the healthy group and health examination group, respectively. All patients were assessed with MMSE. The DNN had two hidden layers with 400 neurons for each layer, and the output vector predict-ed the MMSE score. The model was validated with the LOOCV method. The DNN achieved 75% sensitivity and 85% specificity in a binary classification problem (normal, MMSE score ≥ 24; cognitive impairment, MMSE scores ≤ 23).

Sale et al. [36] collected functional and clinical data from 55 stroke patients who underwent inpatient rehabilitation, so that to investigate these parameters as prognostic indicators of treatment response. The treatment concerned motor and cognitive improvements, which were measured with several techniques such as FIM and T1 BI. Thereafter, the authors employed the Mutual Information criterion for feature selection and then fed them into a linear SVM regression algorithm. Results showed that BI was not highly predictable with a 22.6 RMSE, while Cognitive FIM presented a 4.28 RMSE.

Martinez et al. [32] tried to understand and predict the cognitive improvement of stroke patients after rehabilitation therapy. This research recruited 201 ischemic stroke patients with their demographic information and applied 24 neuropsychology tests in several cognitive domains at admission. The patient's improvement was recorded with the same tests, whenever they were about to leave the rehabilitation center. Twenty ML models were trained aiming to predict the global patient's improvement, which was extracted from the indicators of improvement from the admission and discharge cognitive assessments. The most effective model was the RF classifier, which achieved 70% Recall. Furthermore, the authors emphasized the importance of interpretability in such models, and therefore they applied the SHAP method to translate the significance of each feature. Results revealed that time since injury and admission compliance were the most crucial features.

4 Discussion

This scoping review summarizes the use of AI in predicting language outcomes after stroke using various types of data, with a focus on the significance of AI in cognitive function prediction. It provides an overview of current research on using machine learning algorithms for predicting language and cognition rehabilitation outcomes in post-stroke patients, highlights the most effective algorithms and data, and identifies key factors that influence rehabilitation outcomes. The aim is to help healthcare professionals understand the potential benefits and limitations of using these algorithms and tailor re-habilitation programs to individual patients.

The trend of AI in cognitive functionality prediction over time is illustrated in Fig. 2. It can be observed that the first paper that used AI to predict language outcomes was published in 2010. However, there was a gap of several years before a systematic publication of related studies began in 2017. This suggests that while the idea of using AI for cognitive functionality prediction was present early on, it took some time for the field to gain momentum and for more studies to be conducted and published. The increase in the number of publications on this topic in recent years indicates a growing interest and investment in the use of AI for cognitive functionality prediction.

The vast majority of studies used traditional ML, while only one paper used Deep Neural Networks. The most used ML models in both regression and classification tasks, were SVM and RF and used in six out of nine studies. These two models can explore non-linear patterns, which is a major advantage when dealing with multiple datasets. Furthermore, the most prevalent validation strategies for parameter optimization were LOOCV and k-fold CV. Most of the

studies have few patients (13–57); therefore, LOOCV is widely used because it is suitable for small datasets. Figure 3 shows the distribution of AI models. However, according to [37] these validation strategies must be avoided in order to reduce over-optimistic results. A reason for the small population in studies is the exclusion/inclusion criteria were applied. Five studies used regression models, while the rest used classification models. Regression models predict a continuous value for a specific indicator for cognitive functionality. These studies show encouraging results in some tasks presenting a low error. For example, in [10], the com-prehension task achieved a 0.9 RMSE using RF model. Classification models usually predict if the cognitive improvement will be poor or sufficient. In this task, the SVM was the prominent model and it obtained the highest performance by achieving 97% Recall.

Different validation techniques, such as hold out, k-fold cross-validation, and leave-one-out cross-validation, have been used to evaluate the performance of the re-ported machine learning models. These techniques help to ensure that the model generalizes well to new unseen data, and to avoid overfitting. However, the choice of validation technique can also impact the existence of bias in the model. Hold-out was employed in one study [30], that is the most common method of validation but it can be prone to bias if the dataset is imbalanced or if the test set is not representative of the population. In contrast, k-fold cross-validation was employed in four papers [10, 25, 30, 31] and leave-one-out cross-validation in another five papers [7, 26–29], which split the data into k subsets and use each subset as a test set in turn. Cross validation can help to reduce bias by ensuring that all data points are used for both training and testing. However, these methods may not be suitable for small datasets, as they can lead to a high variance in the results. Therefore, it is important to consider the size of the dataset and the characteristics of the data when choosing a validation technique, in order to minimize bias and ensure accurate model evaluation.

All studies recruited post-stroke patients, and 6 out of 9 studies' patients were chronically aphasic. Each study provided us with the mean age of subjects, which ranged from 49.5 to 73.5. Moreover, there were studies with a large population, such as [22], which involved 202 subjects, but there were also studies with a few patients (13) [20]. Even though more data add computational complexity, they offer higher accuracy at the same time [38] which is crucial in medicine. The same conclusions were drawn in [39], where it was found that the use of larger samples leads to a more accurate representation of the population value due to their reduced susceptibility to deviation compared to smaller samples, which may deviate from the population value in either direction. On the other hand, it is worth noting that extremely large samples may exaggerate the detection of differences, highlighting statistical differences that may not be clinically significant. [40]. For these reasons, we believe that the use of relatively large but representative data sets that include patients with a range of severity levels, combined with the implementation of appropriate goodness cut-offs to ensure reliable results and explainability tools to enhance the trustworthiness of the proposed methods [41], will significantly advance precision medicine.

Every paper in this scoping review collected data from patients through various examinations in order to depict the impairments after a stroke. These tests were used to obtain a multifaceted picture of the patient's condition because relying on only one type of data was not sufficient to predict language outcomes. The data used by the papers studied in this review were a combination of neuroimaging, clinical, demographic, behavioral, and blood and neuropsychology tests. Almost all studies used multi-modal neuroimaging data because it contains rich information about the condition of the brain. These neuroimaging data included fMRI, lesion maps, and structural and functional connectivity. Figure 4 presents the distribution of data used in the studies reviewed. It is important to note that using a combination of different types of data is essential to provide a comprehensive understanding of a patient's condition, and to improve the accuracy of predictions of language outcomes. Furthermore, the use of multimodal neuroimaging data, such as fMRI and lesion maps, in particular, provides valuable information about the brain's structure and function, which can aid in the prediction of language outcomes.

Artificial intelligence models are significant for precision medicine as they are superior to traditional statistical models and they allowed clinicians to develop new therapies tailor-made for individuals. This is due to the models' ability to discover non-linear patterns for optimized solutions. However, ML and DL models are treated as black boxes as no one is able to explain why they perform so well. At the same time, surveys have questioned the robustness of AI models [42]. The combination of the above statements has led clinicians to mistrust the AI models. According to Tonekaboni et al. [43], it is important for clinicians to show which features are most informative and help the models make the right decision. In this way, they can compare their pre-diction procedure with that of the model. In this context, Martinez et al. [24] was the only study that used an Explainability tool named SHAP (SHapley Additive exPlana-tions) [44]. They used SHAP to interpret the performance of the model by computing every feature contribution. By providing more diverse and comprehensive data sets, along with the ability to understand and interpret [45] the decisions made by machine learning algorithms will lead to the safe use of AI.

Almost all studies presented good results, which is encouraging for the implementation of AI in medicine.

However, there are some contradictions. For instance, the study [26] that consists of 55 subjects presents a high recall value (97%) while the study [31], which consists of 201 subjects, presents a significantly lower recall value (70.1%). Another similar case are the studies [27, 28], where the first achieved almost perfect results ($R^2 = 0.989$) with only 13 subjects while the latter achieved $R^2 = 0.876$ with 57 subjects. The only study that utilized deep learning [29] presented moderate results, with a sensitivity of 75%. Despite having the largest subject population of 202 individuals, this study suggests that deep learning algorithms may require a larger sample size to achieve more effective results. Although the methods cannot be compared because they use different data and outcome assessments, the fact that methods with small datasets perform better raises questions. An explanation may come from the study [10, 41] where they argue that small datasets could lead to overoptimistic results. Additionally, the majority of studies used cross validation strategies. Moreover, all studies used one dataset, while one rule for realistic results is the consideration of several datasets [32].

The majority of studies that conclude this scoping review share a common limitation concerning the population of patients. All surveys involve a relatively small number of subjects, which means that the results may be over-optimistic [10, 46] as mentioned above. In addition to that, no study used an external, independent dataset to test the performance of its model, leading to doubts about the models' robustness. Another limitation is the fact that most of the studies applied feature selection methodologies, and the majority of them were looking for linear correlation among the features to keep. However, as above mentioned, cognitive functionality prediction may require non-linear correlations, and as a consequence, the authors removed features that may enhance the performance of models. Also, Martinez et al. [26] dealt with the limitation of sparse data due to the different levels of completeness of the cognitive assessment features.

Post-stroke language rehabilitation is a complex issue, with several factors that can influence the therapy outcome. Among the most frequently studied factors are the patient's educational level, premorbid intelligence, and infarct lesions. According to Connor et al. [47] educational level may affect the severity of aphasia, but not the effectiveness of language rehabilitation. However, a later study by Hillis and Tippet [48] showed a correlation between educational level and post-stroke language rehabilitation indicating that better language recovery is associated with a higher educational level. In the same study, the extent of the infarct lesion also seems to play a role. Moreover, a study by Withall et al. [49] indicated that patients with higher premorbid intelligence quotient (IQ) responded better to treatment. Nonetheless, almost all studies included in this review do not take into account these factors in their results. Only one study [28] reported that their patients had at least a high school education.

Bias is a common issue in all types of research and can have a significant impact on the validity of the findings. If not managed properly, bias can lead to incomplete or inaccurate representation of the existing literature, which can have negative implications for evidence-based decision-making. To minimize bias in scoping reviews, we followed a rigorous and transparent methodology, including clear inclusion and exclusion criteria, multiple reviewers to validate the results, and explicit reporting of any sources of bias or limitations. However, it is important to acknowledge that the current study is limited in scope, focusing only on a very specific field that is the ML-enhanced prediction of language and cognition rehabilitation outcomes after a stroke. In future work, we plan to expand the scope of our findings by conducting a wider study on cognitive functionality after a stroke, which will further enhance the readability of the findings thus providing a more balanced representation of the state of the evidence in the field.

5 Conclusions

This scoping review focuses on the prediction of cognitive functionality after a stroke using ML and DL models. Our review found a limited number of studies in this field, with only six studies addressing aphasic patients and three addressing cognitive functionality impairments. Our findings revealed that of the 9 reviewed studies, only one utilized DL algorithms, while the rest employed ML models, including SVM and Random Forest for both classification and regression tasks. SVM achieved high recall rates in language rehabilitation assessment and Random Forest achieved low root mean square error in predicting the severity of aphasia after a stroke. These models achieved promising results, with up to 97% recall in the classification task and up to 0.9 RMSE in the regression task. These results indicate that the use of AI in cognitive prediction can be beneficial for clinicians, as it can help them to apply custom therapy to patients based on their predicted conditions.

However, it is important to note that implementation of this AI in clinical practice involves a process of validation and integration with existing clinical systems. This process should include steps such as evaluating the performance of the models on re-al-world data, addressing any explainability and transparency concerns, and ensuring the models align with existing clinical workflows and regulations. Additionally, it may require collaboration between AI experts and healthcare professionals to ensure the safe and effective use of the models in patient care. With these steps in place, AI models have the potential to significantly improve the speed and accuracy of patient as-assessments and drive

advancements in precision medicine. Furthermore, future work should also focus on improving the trustworthiness of these models to facilitate their adoption and implementation in clinical practice.

Acknowledgements We acknowledge support of this work by the project "Study of the interrelationships between neuroimaging, neurophysiological and biomechanical biomarkers in stroke re-habilitation (NEURO-BIO-MECH in stroke rehab)" (MIS 5047286), which is implemented under the Action "Support for Regional Excellence", funded by the Operational Program "Competitiveness, Entrepreneurship and Innovation" (NSRFm2014-2020) and co-financed by Greece and the Euro-pean Union (European Regional Development Fund).

Author Contributions Conceptualization:[CK]; Methodology: [KA, CK and SM]; Formal analysis and investigation: [KA and CK]; Writing—original draft preparation: [KA, CK, SM, PS and CK]; Writing—review and editing: [EK, D T, SK and NA]; Funding acquisition: [KV and NA], Supervision: [KV and NA].

Funding This research was funded by the grant MIS 5047286 from Greek and European funds (EYD-EPANEK).

Data availability Not Applicable.

Declarations

Conflict of Interest The authors have no competing interests to declare that are relevant to the content of this article.

Ethical Approval Not applicable.

Consent to Participate Not applicable.

Consent for Publication Not applicable.

References

1. Karatzetzou S, Tsiptsios D, Terzoudi A, et al. Transcranial magnetic stimulation implementation on stroke prognosis. Neurol Sci. 2022;43(2):873–88.
2. Gkantzios A, Tsiptsios D, Karatzetzou S, et al. Stroke and emerging blood biomarkers: a clinical prospective. Neurol Int. 2022;14:784–803.
3. Berthier ML. Poststroke aphasia. Drugs Aging. 2005;22:163–82.
4. Sharma R, Mallick D, Llinas RH, Marsh EB. Early post-stroke cognition: in-hospital predictors and the association with functional outcome. Front Neurol. 2020;11: 613607.
5. Hinman JD, Rost NS, Leung TW, et al. Principles of precision medicine in stroke. J Neurol Neurosurg Psychiatry. 2017;88:54–61.
6. Horn SD, DeJong G, Smout RJ, et al. Stroke rehabilitation patients, practice, and outcomes: is earlier and more aggressive therapy better? Arch Phys Med Rehabil. 2005;86:101–14.
7. Saur D, Ronneberger O, Kümmerer D, et al. Early functional magnetic resonance imaging activations predict language outcome after stroke. Brain. 2010;133:1252–64. https://doi.org/10.1093/brain/awq021.
8. Zittel S, Weiller C, Liepert J. Citalopram improves dexterity in chronic stroke patients. Neurorehabil Neural Repair. 2008;22:311–4.
9. Kim E-K, Lee D-K, Kim Y-M. Effects of aquatic PNF lower extremity patterns on balance and ADL of stroke patients. J Phys Ther Sci. 2015;27:213–5.
10. Enhanced estimations of post-stroke aphasia severity using stacked multimodal predictions. https://doi.org/10.1002/hbm.23752.
11. Wang J, Marchina S, Norton AC, et al. Predicting speech fluency and naming abilities in aphasic patients. Front Hum Neurosci. 2013;7:831.
12. Saur D, Lange R, Baumgaertner A, et al. Dynamics of language reorganization after stroke. Brain. 2006;129:1371–84.
13. Hamet P, Tremblay J. Artificial intelligence in medicine. Metabolism. 2017;69:S36–40.
14. Sakatani K, Oyama K, Hu L. Deep learning-based screening test for cognitive impairment using basic blood test data for health examination. Front Neurol. 2020;11:588140
15. Billot A, Lai S, Varkanitsa M, et al. Multimodal Neural and behavioral data predict response to rehabilitation in chronic poststroke aphasia. Stroke. 2022;53:1606–14. https://doi.org/10.1161/STROKEAHA.121.036749.
16. Li X, Chen Z, Jiao H, et al. Machine learning in the prediction of post-stroke cognitive impairment: a systematic review and meta-analysis. Front Neurol. 2023;14:1211733. https://doi.org/10.3389/fneur.2023.1211733.
17. Azevedo N, Kehayia E, Jarema G, et al. How artificial intelligence (AI) is used in aphasia rehabilitation: a scoping review. Aphasiology. 2023. https://doi.org/10.1080/02687038.2023.2189513.
18. Marotta N, De Sire A, Marinaro C, et al. Efficacy of transcranial direct current stimulation (tDCS) on balance and gait in multiple sclerosis patients: a machine learning approach. J Clin Med. 2022;11:3505.
19. Dalianis H. Evaluation metrics and evaluation. In: Dalianis H (ed) Clinical text mining. Springer, 2018; pp 45–53.
20. Rathnayaka MHKR, Watawala WKCR, Manamendra MG, et al. Cognitive rehabilitation based personalized solution for dementia patients using reinforcement learning. In: 2021 IEEE International Systems Conference (SysCon). IEEE, Vancouver, BC, Canada, 2021; pp 1–6.
21. Das A, Day TW, Kulkarni V, et al. 15—Towards intelligent extended reality in stroke rehabilitation: application of machine learning and artificial intelligence in rehabilitation. In: Pillai AS, Menon B, editors., et al., Augmenting neurological disorder prediction and rehabilitation using artificial intelligence. Academic Press; 2022. p. 309–29.
22. Guo L, Zhang B, Wang J, et al. Wearable intelligent machine learning rehabilitation assessment for stroke patients compared with clinician assessment. J Clin Med. 2022. https://doi.org/10.3390/jcm11247467.
23. Kertesz A. Western Aphasia Battery-Revised (WAB-R) [Database record]. APA PsycTests, 2006.
24. Folstein MF, Folstein SE, McHugh PR. "Mini-mental state": a practical method for grading the cognitive state of patients for the clinician. J Psychiatr Res. 1975;12:189–98.
25. Prodinger B, O'Connor RJ, Stucki G, Tennant A. Establishing score equivalence of the Functional Independence Measure motor

scale and the Barthel Index, utilising the International Classification of Functioning, Disability and Health and Rasch measurement theory. J Rehabil Med. 2017;49:416–22.
26. Kastenbaum JG, Bedore LM, Peña ED, et al. The influence of proficiency and language combination on bilingual lexical access. Biling Lang Cognit. 2019;22:300–30.
27. Kaplan E, Goodglass H, Weintraub S. Boston naming test, 2nd ed. Pro-Ed; 2001.
28. Howard D, Patterson KE. The pyramids and palm trees test. A test of semantic access from words and pictures. Thames Valley Company; 1992.
29. Roach A, Schwartz MF, Martin N, et al. The Philadelphia naming test: scoring and rationale. Clinical aphasiology. 1996;24:121–33.
30. Zou Q-H, Zhu C-Z, Yang Y, et al. An improved approach to detection of amplitude of low-frequency fluctuation (ALFF) for resting-state fMRI: fractional ALFF. J Neurosci Methods. 2008;172:137–41.
31. PRISMA Extension for Scoping Reviews (PRISMA-ScR): Checklist and Explanation | Annals of Internal Medicine. https://doi.org/10.7326/M18-0850. Accessed 8 Nov 2022.
32. Martinez HB, Cisek K, García-Rudolph A, et al. Understanding and Predicting Cognitive Improvement of Young Adults in Ischemic Stroke Rehabilitation Therapy. Front Neurol. 2022;13:886477. https://doi.org/10.3389/fneur.2022.886477.
33. Kristinsson S, Zhang W, Rorden C, et al. Machine learning-based multimodal prediction of language outcomes in chronic aphasia. Hum Brain Mapp. 2021;42:1682–98. https://doi.org/10.1002/hbm.25321.
34. Grasemann U, Peñaloza C, Dekhtyar M, et al. Predicting language treatment response in bilingual aphasia using neural network-based patient models. Sci Rep. 2021;11:10497. https://doi.org/10.1038/s41598-021-89443-6.
35. Iorga M, Higgins J, Caplan D, et al. Predicting language recovery in post-stroke aphasia using behavior and functional MRI. Sci Rep. 2021;11:8419. https://doi.org/10.1038/s41598-021-88022-z.
36. Sale P, Ferriero G, Ciabattoni L, et al. Predicting motor and cognitive improvement through machine learning algorithm in human subject that underwent a rehabilitation treatment in the early stage of stroke. J Stroke Cerebrovasc Dis. 2018;27:2962–72.
37. Boulesteix A-L. Ten simple rules for reducing overoptimistic reporting in methodological computational research. PLoS Comput Biol. 2015;11: e1004191.
38. Gupta S, Sedamkar RR. Machine learning for healthcare: Introduction. In: Jain V, Chatterjee J (eds) Machine learning with health care perspective. Springer, 2020; pp 1–25.
39. Andrade C. Sample size and its importance in research. Indian J Psychol Med. 2020;42:102–3. https://doi.org/10.4103/IJPSYM.IJPSYM_504_19.
40. Faber J, Fonseca LM. How sample size influences research outcomes. Dent Press J Orthod. 2014;19:27–9. https://doi.org/10.1590/2176-9451.19.4.027-029.ebo.
41. Rasheed K, Qayyum A, Ghaly M, et al. Explainable, trustworthy, and ethical machine learning for healthcare: a survey. Comput Biol Med. 2022;149:106043.
42. Apostolidis KD, Papakostas GA. A survey on adversarial deep learning robustness in medical image analysis. Electronics. 2021;10:2132. https://doi.org/10.3390/electronics10172132.
43. Tonekaboni S, Joshi S, McCradden MD, Goldenberg A. What clinicians want: contextualizing explainable machine learning for clinical end use. In: Machine Learning for Healthcare Conference. PMLR, 2019; pp 359–380.
44. Lundberg S, Lee S-I. A unified approach to interpreting model predictions. In: NIPS'17: Proceedings of the 31st International Conference on Neural Information Processing Systems, 2017; p. 4768–77.
45. Wong J, Murray Horwitz M, Zhou L, Toh S. Using machine learning to identify health outcomes from electronic health record data. Current epidemiology reports. 2018;5:331–42.
46. Héroux ME, Taylor JL, Gandevia SC. The use and abuse of transcranial magnetic stimulation to modulate corticospinal excitability in humans. PLoS ONE. 2015;10: e0144151.
47. Connor LT, Obler LK, Tocco M, et al. Effect of socioeconomic status on aphasia severity and recovery. Brain Lang. 2001;78:254–7. https://doi.org/10.1006/brln.2001.2459.
48. Hillis AE, Tippett DC. Stroke recovery: surprising influences and residual consequences. Adv Med. 2014. https://doi.org/10.1155/2014/378263.
49. Withall A, Brodaty H, Altendorf A, Sachdev PS. Who does well after a stroke? The Sydney Stroke Study. Aging Ment Health. 2009;13:693–8. https://doi.org/10.1080/13607860902845525.