ORIGINAL ARTICLE

Predicting Clinical Outcomes in Acute Ischemic Stroke Patients Undergoing Endovascular Thrombectomy with Machine Learning

A Systematic Review and Meta-analysis

Yao Hao Teo[1] · Isis Claire Z. Y. Lim[1] · Fan Shuen Tseng[1] · Yao Neng Teo[1] · Cheryl Shumin Kow[1] · Zi Hui Celeste Ng[1] · Nyein Chan Ko Ko[1] · Ching-Hui Sia[1] · Aloysius S. T. Leow[1] · Wesley Yeung[1] · Wan Yee Kong[2] · Bernard P. L. Chan[3] · Vijay K. Sharma[1,3] · Leonard L. L. Yeo[1,3] (ID) · Benjamin Y. Q. Tan[1,3]

Abstract

Purpose Conventional predictive models are based on a combination of clinical and neuroimaging parameters using traditional statistical approaches. Emerging studies have shown that the machine learning (ML) prediction models with multiple pretreatment clinical variables have the potential to accurately prognosticate the outcomes in acute ischemic stroke (AIS) patients undergoing thrombectomy, and hence identify patients suitable for thrombectomy. This article summarizes the published studies on ML models in large vessel occlusion AIS patients undergoing thrombectomy.

Methods We searched electronic databases including PubMed from 1 January 2000 up to 14 October 2019 for studies that evaluated ML algorithms for the prediction of outcomes in stroke patients undergoing thrombectomy. We then used random-effects bivariate meta-analysis models to summarize the studies.

Results We retained a total of five studies that evaluated ML (4 support vector machine, 1 decision tree model) with a combined cohort of 802 patients. The prevalence of good functional outcome defined by 90-day mRS of 0–2 when available. Random effects model demonstrated that the AUC was 0.846 (95% confidence interval, CI 0.686–0.902). A pooled diagnostic odds ratio of 12.6 was computed. The pooled sensitivity and specificity were 0.795 (95% CI 0.651–0.889) and 0.780 (95% CI 0.634–0.879), respectively.

Conclusion ML may be useful as an adjunct to clinical assessment to predict functional outcomes in AIS patients undergoing thrombectomy, and hence identify suitable patients for treatment. Further studies validating ML models in large multicenter cohorts are necessary to explore this further.

Keywords Large vessel occlusion · Functional outcomes · Neuroimaging · Diagnostic test accuracy · mRS

The authors Y.H. Teo and I.C.Z.Y. Lim contributed equally to this work.

✉ Leonard L. L. Yeo
Leonard_ll_yeo@nuhs.edu.sg

[1] Department of Medicine, Yong Loo Lin School of Medicine, National University of Singapore, Singapore, Singapore

[2] Department of Neurology, Detroit Medical Centre, Michigan, USA

[3] Division of Neurology, Department of Medicine, National University Health System, 1E Kent Ridge Road, NUHS Tower Block Level 11, 119228 Singapore, Singapore
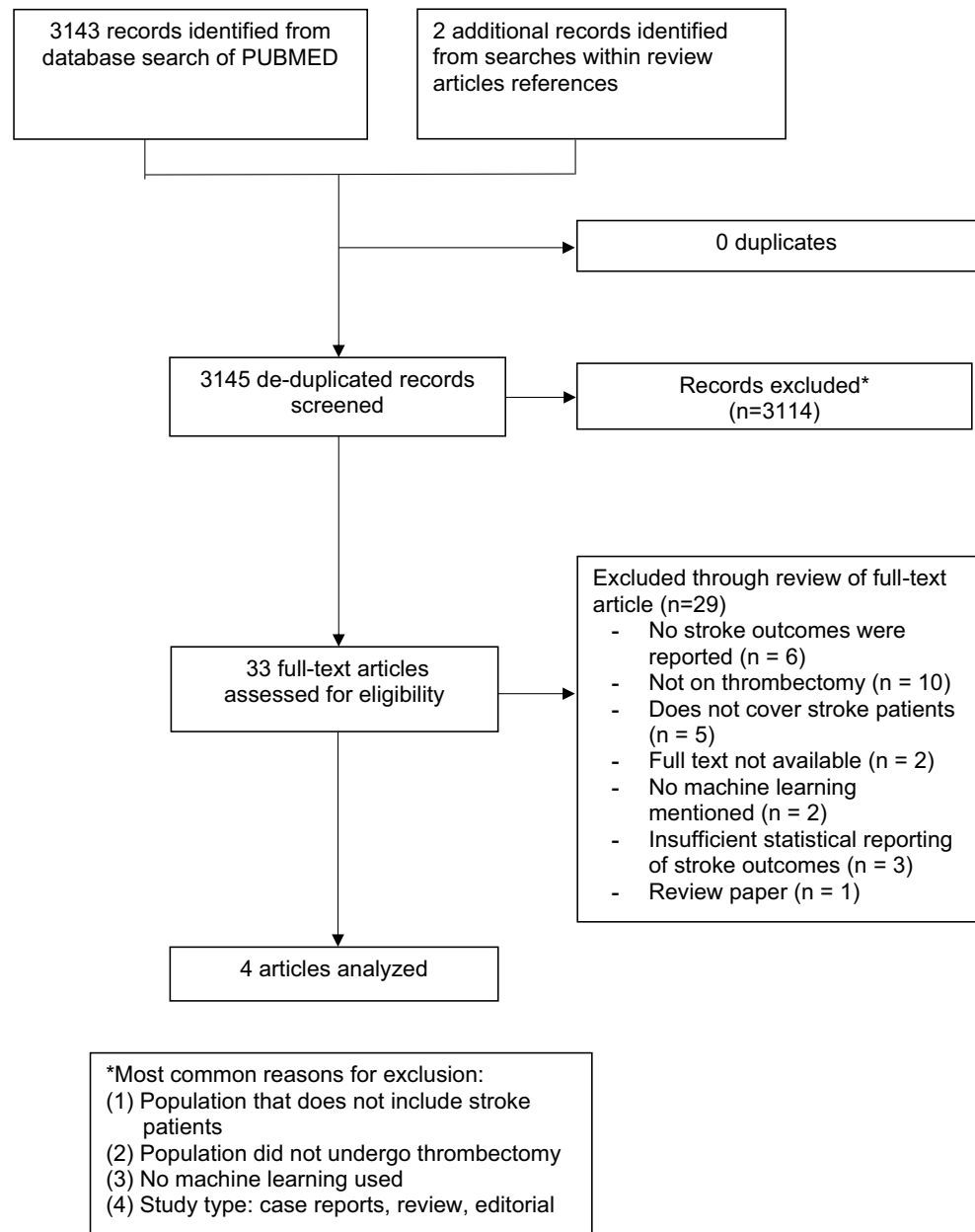
Abbreviations

AIS	Acute ischemic stroke
ASPECTS	Alberta Stroke Program Early CT Score
DC	Discharge
DOR	Diagnostic odds ratio
EVT	Endovascular thrombectomy
ML	Machine learning
mRS	Modified Rankin score
NIHSS	National Institutes of Health Stroke Scale
NPV	Negative predictive value
PICOS	Population, intervention, comparison, outcome, and study design
PPV	Positive predictive value
PRISMA	Preferred reporting items of systematic reviews and meta-analyses
ROC	Receiver operator curves
SVM	Support vector machine

Introduction

Acute ischemic stroke (AIS) is the leading cause of adult disability in Singapore and with an ageing population, the burden of stroke is expected to rise [1]. In patients with large artery occlusion in the anterior circulation presenting within 24h, mechanical thrombectomy may be indicated. Presently, clinicians use a combination of clinical judgment and neuroimaging parameters to discern if a patient will be suitable for endovascular thrombectomy (EVT) [2]. This decision-making has been previously modeled by a multivariate logistic regression model by selecting some prognostic variables, and aggregating them into a usable scale [3]; however, this classical approach is limited as it operates on the assumption that there is a linear relationship between the variables and the logarithmic odds of outcomes, and is weak to collinearity between the variables [4]. By contrast, using machine learning (ML), one can program various algorithms that are free of these linear assumptions with the benefit of being able to control collinearity by regularization hyperparameters [5].

Table 1 PICOs, inclusion criteria and exclusion criteria applied to database search

PICOS	Inclusion criteria	Exclusion criteria
Population	Patients with large vessel occlusion ischemic stroke	Non-thrombectomy interventions, alternative and traditional medicine (including chiropractic, Traditional Chinese Medicine)
	Patients who have undergone an endovascular procedure OR mechanical thrombolysis	Non-neurostroke specialties
		Pharmacy, dentistry, pathology, radiology, pediatrics, emergency medicine, obstetrics and gynecology, anesthesiology
		Allied health (dietetics, podiatry, speech therapy, midwifery, occupational and physiotherapy)
Intervention	Machine learning on large vessel occlusion interventions to understand, review and evaluate outcomes on patients	Prediction models that do not use machine learning
	Machine learning encompass various algorithms including elastic net regression or ridge regression or naive bayes or support vector machine or perceptron or neural networks or multilayer perceptron or deep learning or deep neural network or convolutional neural network or recurrent neural network or bayesian network or decision tree or random forest or gradient boosting machine or extreme gradient boosting	
Comparison	Comparisons of the various ML algorithms and its impact upon the outcome prediction in stroke patients	Traditional logistic regression for outcome prediction in stroke patients
Outcome	Modified Rankin Scale	–
	Symptomatic intracranial hemorrhage	
	Modified treatment in cerebral infarct score	
Study design	Articles in English or translated into English	Case reports and series, ideas, editorials, and perspectives
	All study designs including: – Mixed methods research, meta-analyses, systematic reviews, randomized controlled trials, cohort studies, case-control studies, cross-sectional studies, and descriptive papers – Grey literature/electronic and print information not controlled by commercial publishing	
	Year of publication: 1 January 2000–14 October 2019	
	Databases: PubMed	

Fig. 1 PRISMA flow diagram of study selection

```
┌─────────────────────────┐    ┌─────────────────────────┐
│ 3143 records identified │    │ 2 additional records    │
│ from database search of │    │ identified from searches│
│ PUBMED                  │    │ within review articles  │
│                         │    │ references              │
└─────────────────────────┘    └─────────────────────────┘
              │                            │
              └──────────────┬─────────────┘
                             │
                             │────────────►┌─────────────────────────┐
                             │             │ 0 duplicates            │
                             │             └─────────────────────────┘
                             ▼
                 ┌─────────────────────────┐     ┌─────────────────────────┐
                 │ 3145 de-duplicated      │────►│ Records excluded*       │
                 │ records screened        │     │ (n=3114)                │
                 └─────────────────────────┘     └─────────────────────────┘
                             │
                             ▼
                 ┌─────────────────────────┐
                 │ 33 full-text articles   │
                 │ assessed for            │
                 │ eligibility             │
                 └─────────────────────────┘
                             │
                             ▼
                 ┌─────────────────────────┐
                 │ 4 articles analyzed     │
                 └─────────────────────────┘
```

Excluded through review of full-text article (n=29)
- No stroke outcomes were reported (n = 6)
- Not on thrombectomy (n = 10)
- Does not cover stroke patients (n = 5)
- Full text not available (n = 2)
- No machine learning mentioned (n = 2)
- Insufficient statistical reporting of stroke outcomes (n = 3)
- Review paper (n = 1)

*Most common reasons for exclusion:
(1) Population that does not include stroke patients
(2) Population did not undergo thrombectomy
(3) No machine learning used
(4) Study type: case reports, review, editorial

Machine learning has been shown to successfully incorporate multifactorial events in various fields for clinically relevant outcomes, such as the diagnosis of acute coronary syndromes [6]. Similarly, emerging studies seem to indicate that there is great potential in introducing ML models as a clinical tool to accurately predict the suitability of an AIS patient for an EVT procedure [7–9]; however, individual studies may not be statistically powered to evaluate the robustness of the findings or may not adequately account for the small biases in each population. Hence, we conducted a systematic review and meta-analysis to evaluate the effectiveness of current ML models as a clinical tool

to predict the clinical outcome of AIS patients undergoing EVT.

Methodology

This diagnostic test accuracy (DTA) meta-analysis was conducted and reported in accordance with the Cochrane DTA handbook [10] and preferred reporting items of systematic reviews and meta-analyses (PRISMA) guidelines [11]. We searched PubMed from 1 January 2000 to 14 October 2019 for studies that evaluated ML algorithms for the prediction of outcomes with the modified Rankin

Table 2 Participant characteristics

Study	Sample size	Mean age (years, SD)	Males	Baseline NIHSS (SD/IQR)	Baseline ASPECTS (SD/IQR)	Time to procedure/min (SD)	Time to re-canaliza-tion/min (SD)	IV throm-bolysis	Outcome mRS-90d = 0–2	Outcome mRS-30d = 0–2
Alawieh (2018) [9]	36	86.0 (4)	17	15 (SD: 7)	8.4 (SD: 1.4)	NR[a]	NR	19	13	NR
	110	84.6 (4)	39	18 (SD: 16.5)	8.5 (SD: 1.4)	NR	NR	53	22	NR
Asadi (2014a) [7]	107	65.3 (13.8)	66	17.7 (SD: 7.9)	NR	82.3 (39)	340 (91.8)	44	NR	54
Asadi (2014b) [8]	50	65 (14.9)	35	20.6 (SD: 15.5)	NR	97.2 (43)	660 (374)	7	23	NR
Nishi (2019) [5]	387	76.0 (11.9)	206	19 (IQR: 14–23)	9 (IQR: 8–10)	NR	NR	214	148	NR

SD standard deviation, *NIHSS* National Institutes of Health stroke scale, *ASPECTS* Alberta stroke program early CT score, *IQR* interquartile range, *mRS-90d* modified rank scale (mRS) at 90 days, *mRS-30d* modified rank scale (mRS) at 30 days

[a]NR refers to not reported

scale (mRS), modified thrombolysis in cerebral infarction (modified TICI), symptomatic intracranial cerebral hemorrhage score, and mortality in stroke patients undergoing thrombectomy. Literature search in MEDLINE (PubMed) was performed using the following terms in combination: (large vessel occlusion OR cerebrovascular occlusion OR endovascular thrombectomy OR mechanical thrombectomy) and (artificial intelligence OR machine learning). We included all studies (randomized controlled trials, prospective/retrospective cohort studies, case-control studies), according to the PICOS (Table 1). We excluded all studies not reporting stroke outcomes following thrombectomy. The literature search and data extraction were performed independently by two reviewers, and all disagreements were resolved by mutual consensus. The corresponding authors of three papers were contacted by email to provide data not directly available in the original publication but have not responded by the time of submission, hence the papers were excluded.

Study protocol and full text articles were independently reviewed by an expert team comprising senior neurologists.

Apart from stroke outcomes of patients, we also collected data on age, sex, smoking, the presence of diabetes mellitus, hypertension, hyperlipidemia, atrial fibrillation, previous ischemic stroke and coronary artery disease. Baseline information included National Institute of Health Stroke Scale (NIHSS), Alberta Stroke Program Early CT Score (ASPECTS), modified Rankin scale (mRS), systolic blood pressure, time of admission (door), time to intravenous thrombolysis (needle), time to procedure start (puncture), time to recanalization, intravenous thrombolysis, endovascular intervention, anesthesia type (general anesthesia vs. conscious sedation), vascular occlusion site and stroke eti-

ology according to TOAST criteria [12, 13]. Outcomes collected include mRS 0–2 at discharge, 90-day follow-up, and 30-day follow-up, symptomatic intracranial hemorrhage, mortality, modified TICI score 2b/3, NIHSS at discharge and NIHSS with early clinical improvement. For the prediction models, we collected data of the accuracy metrics, including sensitivity, specificity, positive predictive value (PPV), negative predictive value (NPV), precision, accuracy, true positive rates, false positive rates, true negative rate, false negative rate, receiver operator curves (ROC), area under the curve (AUC), conclusion, and other outcomes measured. In the case of data unavailability, we requested further data by contacting the corresponding authors of the relevant studies. Quality control were performed by two independent reviewers with the Ottawa grading quality assessment [14], as shown in Appendix 1.

Statistical Analysis

In this DTA meta-analysis, the results from the five included studies were quantitatively pooled and analyzed in R version 3.6.2 (R Foundation for Statistical Computing, Vienna, Austria) using general approaches laid out by Shim et al. [15]. We used the metaprop command to pool sensitivity, specificity, PPV, NPV, and diagnostic odds ratio (DOR). The logit transformation was applied, and Clopper-Pearson exact confidence intervals were used [16]. The reitsma (Package: mada), which employs the bivariate model, was employed to generate the summary ROC curve and compute the area under curve (AUC) [17]. The random-effects model was utilized to account for between-study variance. We present between-study heterogeneity using I2 and τ2 statistics. We considered I2 of <30% to indicate low het-
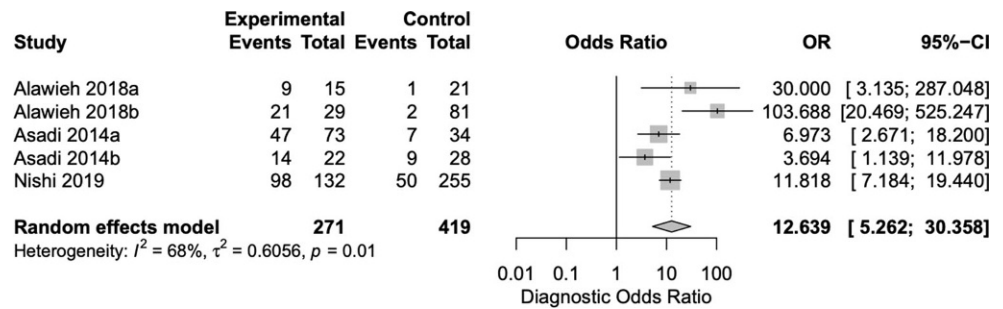
Table 3 Clinical predictors and outcomes included in the training of different ML models

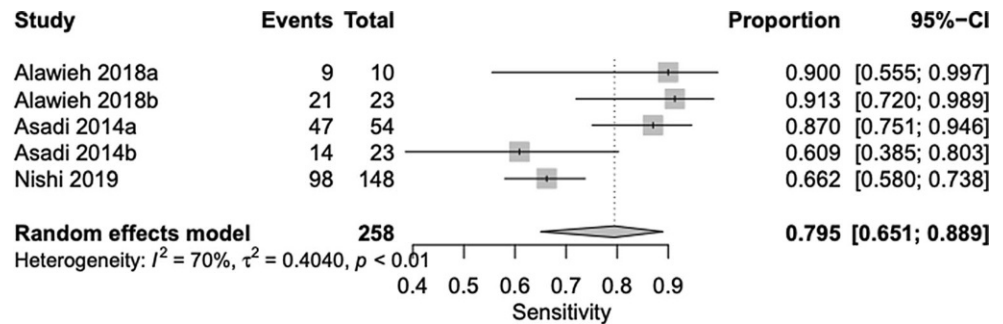| Paper	Clinical predictors and outcomes included in ML models	
Alawieh (2018) [9]	Demographic	Age, gender, race, mRS (baseline), diabetes mellitus, hyperlipidemia, hypertension, atrial fibrillation, NIHSS (baseline), IV tPA administration
	Preprocedural imaging	ASPECT score (non-contrast CT scan), MR DWI (infarction site)
	Intervention	ADAPT thrombectomy (posterior ischemic stroke patients, or anterior circulation if CT perfusion imaging showed a mismatch, e.g. presumed penumbra between relative cerebral blood volume and blood flow that correlated with their presenting NIHSS score)
	Intervention (time-specific)	Onset-to-groin time
	Procedural or post-procedural complications	Postprocedural hemorrhage, parenchymal hematoma 2 (PH2)-type symptomatic intracerebral hemorrhage
	Outcomes	mRS (90 days)
Asadi (2014a) [7]	Demographic	Age, diabetes mellitus, hypertension, hypercholesterolemia, atrial fibrillation, history of ischemic attack, neurological examination, NIHSS (baseline), IV-tPA prior to EVT, intra-arterial chemical thrombolytic agents
	Pre-procedural imaging	Initial diagnostic angiogram, occlusion site (in the case of multiple sequential occlusions, the proximal vessel was used as a data point, and depending on the extent and segments involved, the artery was categorized as first, second and third occlusion)
	Intervention	EVT (Solitaire stent-retriever, MERCI, Penumbra devices), if present, associated or post-recanalization hemodynamically significant stenoses were also treated with angioplasty or stent insertion
	Intervention (time-specific)	Procedure duration, time of onset to recanalization
	Procedural or postprocedural complications	Arterial perforation, puncture site hematoma, pseudoaneurysm, intracranial hemorrhage transformations (silent, symptomatic), number of attempts for recanalization, patient general anesthesia status
	Outcomes	mRS (90 days), TICI, neuroimaging (postprocedural CT scans 24–36 h)
Asadi (2014b) [8]	Demographic	Age, acute posterior circulation stroke, diabetes mellitus, hypertension, hypercholesterolemia, atrial fibrillation, history of ischemic heart disease, previous cerebral stroke or transient ischemic attack, neurological examination, NIHSS (baseline), IV-tPA prior to EVT, intra-arterial chemical thrombolytic agents
	Pre-procedural imaging	Initial diagnostic angiogram, occlusion site (in the case of multiple sequential occlusions, the proximal vessel was used as a data point, and depending on the extent and segments involved, the artery was categorized as first, second and third occlusion)
	Intervention	EVT (Solitaire stent-retriever, MERCI device), if present, associated or post-recanalization hemodynamically significant stenoses were also treated with angioplasty or stent insertion
	Intervention (time-specific)	Procedure duration, time of onset to recanalization
	Procedural or postprocedural complications	Arterial perforation, puncture site hematoma, pseudoaneurysm, intracranial hemorrhage transformations (silent, symptomatic), number of attempts for recanalization, patient general anesthesia status
	Outcomes	mRS (90 days), TICI, neuroimaging (postprocedural CT scans 24–36 h)
Nishi (2019) [5]	Demographic	Age, sex, mRS (baseline), hypertension, diabetes mellitus, atrial fibrillation, intake of antithrombotic drugs, pattern of stroke onset (wake-up, found, witnessed stroke), NIHSS (baseline), blood glucose level on arrival, IV-tPA, care-dependent
	Pre-procedural imaging	CT or MRI (site of occlusion), ASPECT score
	Prediction methods included	PRE score, AT2
	Intervention	EVT (stent retriever, aspiration catheter)
	Intervention (time-specific)	Onset-to-door time
	Procedural or post-procedural complications	NR[a]
	Outcomes	mRS (90 days), mTICI, SICH

NIHSS National Institutes of Health stroke scale, *ASPECTS* Alberta stroke program early CT score, *ADAPT* direct aspiration first pass technique, *mRS-90d* modified rank scale (mRS) at 90 days, *mRS* modified rank scale (mRS), *EVT* endovascular thrombectomy
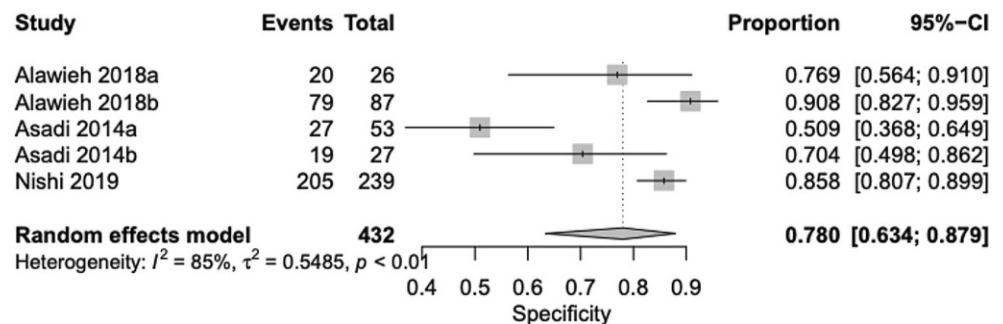
[a]NR refers to not reported

Fig. 2 Forest plot of the diagnostic odds ratio reported by studies that applied a machine learning method to predict clinical outcomes in ischemic stroke patients undergoing thrombectomy

Study	Experimental Events	Total	Control Events	Total		OR	95%-CI
Alawieh 2018a	9	15	1	21		30.000	[3.135; 287.048]
Alawieh 2018b	21	29	2	81		103.688	[20.469; 525.247]
Asadi 2014a	47	73	7	34		6.973	[2.671; 18.200]
Asadi 2014b	14	22	9	28		3.694	[1.139; 11.978]
Nishi 2019	98	132	50	255		11.818	[7.184; 19.440]
Random effects model		**271**		**419**		**12.639**	**[5.262; 30.358]**

Heterogeneity: $I^2 = 68\%$, $\tau^2 = 0.6056$, $p = 0.01$

Diagnostic Odds Ratio (0.01 0.1 1 10 100)

Fig. 3 Forest plot of the sensitivity reported by studies that applied a machine learning method to predict clinical outcomes in ischemic stroke patients undergoing thrombectomy

Study	Events	Total		Proportion	95%-CI
Alawieh 2018a	9	10		0.900	[0.555; 0.997]
Alawieh 2018b	21	23		0.913	[0.720; 0.989]
Asadi 2014a	47	54		0.870	[0.751; 0.946]
Asadi 2014b	14	23		0.609	[0.385; 0.803]
Nishi 2019	98	148		0.662	[0.580; 0.738]
Random effects model		**258**		**0.795**	**[0.651; 0.889]**

Heterogeneity: $I^2 = 70\%$, $\tau^2 = 0.4040$, $p < 0.01$

Sensitivity (0.4 0.5 0.6 0.7 0.8 0.9)

Fig. 4 Forest plot of the specificity reported by studies that applied a machine learning method to predict clinical outcomes in ischemic stroke patients undergoing thrombectomy

Study	Events	Total		Proportion	95%-CI
Alawieh 2018a	20	26		0.769	[0.564; 0.910]
Alawieh 2018b	79	87		0.908	[0.827; 0.959]
Asadi 2014a	27	53		0.509	[0.368; 0.649]
Asadi 2014b	19	27		0.704	[0.498; 0.862]
Nishi 2019	205	239		0.858	[0.807; 0.899]
Random effects model		**432**		**0.780**	**[0.634; 0.879]**

Heterogeneity: $I^2 = 85\%$, $\tau^2 = 0.5485$, $p < 0.01$

Specificity (0.4 0.5 0.6 0.7 0.8 0.9)

erogeneity between studies, 30–60% to indicate moderate heterogeneity, and >60% to indicate substantial heterogeneity. Two-sided P values of <0.05 were regarded to indicate nominal statistical significance.

Results

The PRISMA flowchart is presented in Fig. 1. MEDLINE literature search retrieved 3143 results, and bibliographic searches from review articles references uncovered two additional studies. After an initial screen, 3110 studies were excluded as they did not include stroke patients, thrombectomy patients, did not use machine learning methods, or were of an inappropriate study type. In the remaining 33 full-text papers, full text review excluded 29 papers and 4 papers were included in our analysis.

Baseline Characteristics

We retained a total of five studies that evaluated ML. Support vector machine was utilized across four studies and decision tree was utilized in Alawieh 2018 [9]. In the study by Nishi 2019 [5], regularized logistic regression, random forest, and support vector machine were utilized. To reduce heterogeneity, we analyzed the support vector machine model in Nishi 2019. In addition, the Alawieh (2018) paper included 2 separate cohorts, comprising one retrospective and one prospective. Hence, despite the PRISMA in Fig. 1 showing four included articles, the number of included studies totaled five.

The five studies (1 prospective and 4 retrospective studies) comprised a combined cohort of 805 patients, of which 690 patients (mean age: 75.4 years, 52.6% males) reported outcome assessment and was included in the analysis. A significant proportion of patients (42.1%) received
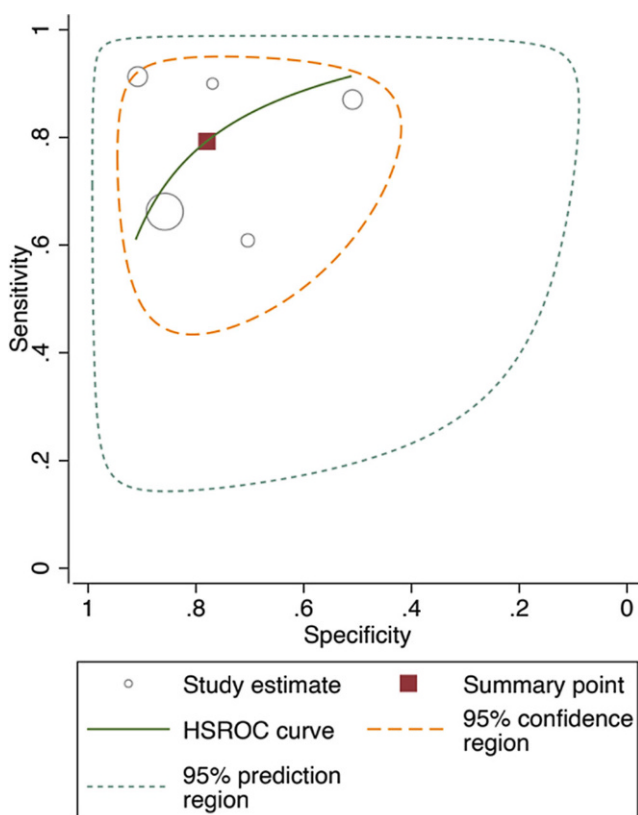
Fig. 5 Summary receiver operating characteristic curve (*ROC*). Curve is not part of the ROC. *HSROC* Hierarchical Summary Receiver Operating Characteristic

IV thrombolysis. The primary outcome measured was mRS at 90 days when available. When unavailable, mRS at the closest time point was obtained and specified. The mRS measures the degree of disability in stroke patients, with a good functional outcome defined as mRS 0–2. A good functional outcome was achieved in 37.7% of patients. As the symptomatic intracranial hemorrhage, mortality, modified treatment in cerebral infarct score 2b/3, NIHSS at discharge, and NIHSS with early clinical improvement were not reported across all studies, they were excluded from the analysis. The participant characteristics of the included studies are shown in Table 2.

Across the five cohorts, the type of machine learning modalities utilized and how they were derived are varied. A summary of the machine learning modalities, comprising machine learning model, software algorithm, training procedure, and optimizing metrics of the included studies is attached in Appendix 2. The quality of training data, comprising type of study, cohort size, class imbalance, normalization/standardization, and validation are summarized in Appendix 3. Clinical predictors and outcomes included in the five studies are summarized in Table 3.

Machine Learning Pooled Outcomes

The diagnostic odds ratio in predicting the outcome of AIS patients undergoing endovascular thrombectomy is presented in Fig. 2. Random effects model demonstrated that the odds ratio was 12.6 (95% confidence interval: 5.26–30.36, I2 = 68%).

The sensitivity is presented in Fig. 3. Random effects model demonstrated that the pooled sensitivity was 0.795 (95% confidence interval: 0.651–0.889, I2 = 70%).

The specificity is presented in Fig. 4. Random effects model demonstrated that the pooled specificity was 0.780 (95% confidence interval: 0.634–0.879, I2 = 85%).

The random effects model demonstrated that the pooled negative predictive value was 0.874 (95% confidence interval: 0.728–0.947) (I2 = 84%). The random effects model demonstrated that the pooled positive predictive value was 0.697 (95% confidence interval: 0.640–0.749, I2 = 0.0%).

The summary ROC is presented in Fig. 5. The AUC of the summary ROC was 0.846 (95% confidence interval: 0.686–0.902).

Discussion

This diagnostic test accuracy meta-analysis demonstrates the utility of ML algorithms as an adjunctive tool in identifying good candidates in acute ischemic stroke patients indicated for endovascular thrombectomy (EVT) with moderate to high AUC, sensitivity, specificity, NPV and PPV; however, there exists a large heterogeneity across ML models. The accuracy of ML will undoubtedly improve over time as algorithms are trained on larger and more robust databases, improving patient selection for endovascular thrombectomy.

Traditional predictive models include ASPECTS score, baseline NIHSS, M1 occlusion, Boston acute stroke imaging scale (M1-BASIS), and the DEFUSE-2 trial. When considering the effectiveness of traditional clinical predictive models as compared to the ML models, the AUC was utilized [18, 19]. Compared to that of the ML model (AUC for ML model = 0.846), the AUC for ASPECTS (0.730), M1-BASIS (0.721), NIHSS (0.728), and the DEFUSE-2 trial (0.730) demonstrated a lower value [18, 19]. This demonstrates a relative superiority of ML models compared to traditional predictive models in prognosticating stroke patients who are indicated for thrombectomy.

A pooled diagnostic odds ratio (DOR) of 12.6 demonstrates the potential of utilizing ML algorithms as a predictive tool for good clinical outcomes with higher discriminatory power compared to conventional models [20]; however, as the DOR is usually used as an output statistic, rather than a summary test statistic, as it can be achieved at different

combinations of sensitivity and specificity. Examining the random effects model, it demonstrated a pooled sensitivity and specificity 0.795 and 0.780, respectively, each nearing 80%. On the other hand, traditional predictive models, such as the ASPECT score, conventionally used to exclude patients for the use of reperfusion therapy, showed higher sensitivity of 0.91 and specificity of 0.88 compared to that of ML models [21]. We propose that the differences may be due to high heterogeneity across the studies analyzed. Observing the forest plots for sensitivity in Fig. 3, the first three cohorts by Alawieh and Asadi (2014a) showed much higher sensitivities of around 0.90 as compared to the remaining two cohorts. On the other hand, Nishi and Alawieh (2018b) showed much higher specificities of around 0.90 as compared to the remaining three cohorts. These are comparable to traditional models like ASPECT. Hence, we note that a large heterogeneity exists across ML models, affecting the accuracy of the pooled prognostic estimate [22].

Additionally, accounting for the prevalence of good and poor clinical outcomes amongst acute ischemic stroke patients, we derived the pooled positive predictive value and negative predictive value of 0.697 and 0.874. A high negative predictive value of 87% shows that a high proportion of patients who are predicted by ML to have poor outcomes do in fact suffer from poor outcomes postthrombectomy, hence guiding the risk-benefit decision-making process for thrombectomy [23]. Hence, we proposed that current ML models can be used as an adjunctive clinical tool with a high discriminatory value to predict the suitability of an AIS patient for an EVT procedure. Using ML algorithms, patients with predicted poor mRS-90d may be assessed to be at a higher risk for poorer outcomes, corresponding to a lower benefit from the intervention. Hence, EVT procedures may be deemed unsuitable if the risk outweighs the potential benefit.

Recent studies showed that due to its high discriminatory value, ML has the potential to be more effective compared to older regression models in predicting the clinical outcome in other diseases [24]. In stroke patients, ML has been shown to translate data like spatial regularization of diffusion-weighted index into statistical data [25], and imaging algorithms that are able to estimate the extent of potentially salvageable tissues [26]. This implies that with the correct combination of factors and algorithm, ML might be able to accurately predict outcomes of stroke patients undergoing endovascular thrombectomy.

However, we emphasize that the interplay of factors and confounders in the human body is very complex, and a significant amount of training over a large dataset needs to be applied to ML to compensate for these factors [27]. This may partially explain the observed large heterogeneity across ML models in our study. Nonetheless, the converse may also hold true. With sufficient training, ML algorithms

may be able to augment physician's decision making by accounting for relationships and interactions between different variables that the clinician may not be aware of, hence enabling clinicians to make a more informed decision about a patient's treatment.

We postulate that the implementation of a machine learning tool into clinical practice may be less challenging than previously thought, owing to its convenience, ease of use and personalization. Currently, indications for a thrombectomy include a contraindication to tissue plasminogen activator thrombolysis, a timing of within 6h onset of stroke, a large vessel intracranial vessel, and the use of conventional scoring including NIHSS ≥ 10 and ASPECTs ≥ 6. We propose that machine learning can be considered as a secondary tool to indicate thrombectomy use. On notification of a patient presenting with AIS, demographic data and medical history may be automatically extracted from the hospital database into the ML algorithm. After preliminary clinical assessment and investigations are completed, the clinician may further input relevant parameters and investigation findings into the ML algorithm to quickly determine the potential benefit and hence suitability of the patient in undergoing thrombectomy. The majority of the studies analyzed in this meta-analysis used a form of widely used machine learning called support vector machine (SVM). This is a commonly used machine learning tool, having previously been used in cancer genomics, due to the strength of the algorithm and the flexibility in the data presentation [28]. Furthermore, the primary clinical outcome is mRS, a widely used clinical outcome measure in stroke patients. This provides a relatively smoother transition to the uptake of machine learning to aid clinicians.

Limitations

Certain limitations of this study should be acknowledged. First, only a small number of studies were included in the analysis, of which the majority utilized SVM. This may reduce the generalizability of the findings across different ML models and patient cohorts.

Second, some of the utilized clinical predictors may not have been reported fully in the papers. Furthermore, among the reported clinical predictors used, while the categories of predictors are largely similar, the individual predictors included under each category differ across studies. Hence, this may introduce heterogeneity across the studies included in the meta-analysis.

Third, while machine learning has tremendous potential its "black box" limitation mandates that ML models require large databases to improve their accuracy, and the true underlying relationships between influential factors remains largely unknown to the user [29, 30].

Fourth, an mRS score of 0–2 was used to represent a good outcome post-thrombectomy. The negative outcome reported may not have been a direct result of the thrombectomy, but rather a complication of an existing comorbidity, which may affect the reliability of the SVM if sufficient variables were not accounted for.

Fifth, while the quality of training data is overall high, most of the studies ($n = 4$) were retrospective, which may introduce bias into the training of ML models. Furthermore, software algorithms differed across studies. Hence, further prospective studies are needed to improve the training of ML models.

Conclusion

The moderate to high AUC, sensitivity, specificity, negative predictive value and diagnostic odds ratio demonstrate that ML is a good adjunct clinical tool to predict the suitability of an AIS patient undergoing EVT. As seen in the heterogeneity of the studies analyzed, further development is required to improve the accuracy of various ML models. Training the algorithms over larger datasets and allocating more resources towards the refinement of algorithms may help improve its sensitivity and specificity so that it may potentially be used as a confirmatory tool in the future.

References

1. Stroke. 2018. www.singhealth.com.sg/patient-care/patient-education/stroke. Access date: 31st August 2020

2. Yew KS, Cheng E. Acute stroke diagnosis. Am Fam Physician. 2009;80:33–40.

3. Johnston KC, Connors AF Jr, Wagner DP, Knaus WA, Wang X, Haley EC Jr. A predictive risk model for outcomes of ischemic stroke. Stroke. 2000;31:448–55.

4. Ranganathan P, Pramesh CS, Aggarwal R. Common pitfalls in statistical analysis: Logistic regression. Perspect Clin Res. 2017;8:148–51.

5. Nishi H, Oishi N, Ishii A, Ono I, Ogura T, Sunohara T, Chihara H, Fukumitsu R, Okawa M, Yamana N, Imamura H, Sadamasa N, Hatano T, Nakahara I, Sakai N, Miyamoto S. Predicting Clinical Outcomes of Large Vessel Occlusion Before Mechanical Thrombectomy Using Machine Learning. Stroke. 2019;50:2379–88.

6. Iannattone PA, Zhao X, VanHouten J, Garg A, Huynh T. Artificial Intelligence for Diagnosis of Acute Coronary Syndromes: A Meta-analysis of Machine Learning Approaches. Can J Cardiol. 2020;36:577–83.

7. Asadi H, Dowling R, Yan B, Mitchell P. Machine learning for outcome prediction of acute ischemic stroke post intra-arterial therapy. PLoS One. 2014;9:e88225.

8. Asadi H, Dowling R, Yan B, Mitchell P. Intra-arterial therapy for basilar artery thrombosis: the role of machine learning in outcome prediction. EJMINT Original Article, 2014:1449000234 (2nd December 2014).

9. Alawieh A, Zaraket F, Alawieh MB, Chatterjee AR, Spiotta A. Using machine learning to optimize selection of elderly patients for endovascular thrombectomy. J Neurointerv Surg. 2019;11:847–51.

10. Macaskill P, Gatsonis C, Deeks JJ, Harbord RM, Takwoingi Y. Chapter 10: Analysing and Presenting Results. In: Deeks JJ, Bossuyt PM, Gatsonis C (editors), Cochrane Handbook for Systematic Reviews of Diagnostic Test Accuracy Version 1.0. The Cochrane Collaboration, 2010. Available from: http://srdta.cochrane.org/.

11. Liberati A, Altman DG, Tetzlaff J, Mulrow C, Gøtzsche PC, Ioannidis JP, Clarke M, Devereaux PJ, Kleijnen J, Moher D. The PRISMA statement for reporting systematic reviews and meta-analyses of studies that evaluate health care interventions: explanation and elaboration. PLoS Med. 2009;6:e1000100.

12. Adams HP Jr, Bendixen BH, Kappelle LJ, Biller J, Love BB, Gordon DL, Marsh EE 3rd. Classification of subtype of acute ischemic stroke. Definitions for use in a multicenter clinical trial. TOAST. Trial of Org 10172 in Acute Stroke Treatment. Stroke. 1993;24:35–41.

13. Mokin M, Primiani CT, Siddiqui AH, Turk AS. ASPECTS (Alberta Stroke Program Early CT Score) Measurement Using Hounsfield Unit Values When Selecting Patients for Stroke Thrombectomy. Stroke. 2017;48:1574–9.

14. Wells GA, Shea B, O'Connell D, Peterson J, Welch V, Losos M, Tugwell P. The Newcastle-Ottawa Scale (NOS) for assessing the quality of nonrandomized studies in meta-analysis. 2011. Available from: www.ohri.ca/programs/clinical_epidemiology/oxford.asp (cited 31st August 2020).

15. Shim SR, Kim SJ, Lee J. Diagnostic test accuracy: application and practice using R software. Epidemiol Health. 2019;41:e2019007.

16. Balduzzi S, Rücker G, Schwarzer G. How to perform a meta-analysis with R: a practical tutorial. Evid Based Ment Health. 2019;22:153–60.

17. Reitsma JB, Glas AS, Rutjes AW, Scholten RJ, Bossuyt PM, Zwinderman AH. Bivariate analysis of sensitivity and specificity produces informative summary measures in diagnostic reviews. J Clin Epidemiol. 2005;58:982–90.

18. Wouters A, Dupont P, Christensen S, Norrving B, Laage R, Thomalla G, Kemp S, Lansberg M, Thijs V, Albers GW, Lemmens R. Multimodal magnetic resonance imaging to identify stroke onset within 6 h in patients with large vessel occlusions. Eur Stroke J. 2018;3:185–92.

19. Yeo LL, Paliwal PR, Wakerley B, Khoo CM, Teoh HL, Ahmad A, Ting EY, Seet RC, Ong V, Chan BP, Yohanna K, Gopinathan A, Rathakrishnan R, Sharma VK. External validation of the Boston Acute Stroke Imaging Scale and M1-BASIS in thrombolyzed patients. Stroke. 2014;45:2942–7.

20. Bossuyt P, Davenport C, Deeks J, Hyde C, Leeflang M, Scholten R. Chapter 11:Interpreting results and drawing conclusions. In: Deeks JJ, Bossuyt PM, Gatsonis C (editors), Cochrane Handbook for Systematic Reviews of Diagnostic Test Accuracy Version 0.9. The Cochrane Collaboration, 2013. Available from: www.srdta.cochrane.org/.

21. Shin DH, Shin DJ, Kim JR. Do All ASPECT Score Regions have the Same Predictive Power for Functional Outcomes? J Stroke Cerebrovasc Dis. 2020;29:104516.

22. Deo RC. Machine Learning in Medicine. Circulation. 2015;132: 1920–30.

23. Lalkhen AG, McCluskey A. Clinical tests: sensitivity and specificity. Continuing Educ Anaesth Crit Care Pain. 2008;8:221–3.

24. Gibson WJ, Nafee T, Travis R, Yee M, Kerneis M, Ohman M, Gibson CM. Machine learning versus traditional risk stratification methods in acute coronary syndrome: a pooled randomized clinical trial analysis. J Thromb Thrombolysis. 2020;49:1–9.

25. Cuingnet R, Rosso C, Lehéricy S, Dormont D, Benali H, Samson Y, Colliot O. Spatially regularized SVM for the detection of brain areas associated with stroke outcome. Med Image Comput Comput Assist Interv. 2010;13:316–23.

26. Bouts MJ, Tiebosch IA, van der Toorn A, Viergever MA, Wu O, Dijkhuizen RM. Early identification of potentially salvageable tissue with MRI-based predictive algorithms after experimental ischemic stroke. J Cereb Blood Flow Metab. 2013;33:1075–82.

27. Alpaydın E. Introduction to machine learning. Massachusetts: The MIT Press Cambridge; 2010.

28. Huang S, Cai N, Pacheco PP, Narrandes S, Wang Y, Xu W. Applications of Support Vector Machine (SVM) Learning in Cancer Genomics. Cancer Genomics Proteomics. 2018;15:41–51.

29. Cabitza F, Rasoini R, Gensini GF. Unintended Consequences of Machine Learning in Medicine. JAMA. 2017;318:517–8.

30. Warwick K. March of the machines : the breakthrough in artificial intelligence 1st Illinois pbk. Urbana: University of Illinois Press; 2004.