BMJ Open

Systematic review protocol to assess artificial intelligence diagnostic accuracy performance in detecting acute ischaemic stroke and large-vessel occlusions on CT and MR medical imaging

Srinivasa Rao Kundeti ,[1,2] Manikanda Krishnan Vaidyanathan,[2] Bharath Shivashankar,[2] Sankar Prasad Gorthi[1]

To cite: Kundeti SR, Vaidyanathan MK, Shivashankar B, *et al*. Systematic review protocol to assess artificial intelligence diagnostic accuracy performance in detecting acute ischaemic stroke and large-vessel occlusions on CT and MR medical imaging. *BMJ Open* 2021;**11**:e043665. doi:10.1136/bmjopen-2020-043665

► Prepublication history for this paper is available online. To view these files, please visit the journal online (http://dx.doi.org/10.1136/bmjopen-2020-043665).

Received 10 August 2020
Revised 13 February 2021
Accepted 19 February 2021

[1]Department of Neurology, Kasturba Medical College, Manipal Academy of Higher Education, Manipal, India
[2]Philips Research, Philips Innovation Campus, Bangalore, India

Correspondence to
Srinivasa Rao Kundeti; srinivasa.rao@learner.manipal.edu

ABSTRACT

Introduction The use of artificial intelligence (AI) to support the diagnosis of acute ischaemic stroke (AIS) could improve patient outcomes and facilitate accurate tissue and vessel assessment. However, the evidence in published AI studies is inadequate and difficult to interpret which reduces the accountability of the diagnostic results in clinical settings. This study protocol describes a rigorous systematic review of the accuracy of AI in the diagnosis of AIS and detection of large-vessel occlusions (LVOs).

Methods and analysis We will perform a systematic review and meta-analysis of the performance of AI models for diagnosing AIS and detecting LVOs. We will adhere to the Preferred Reporting Items for Systematic Reviews and Meta-analyses Protocols guidelines. Literature searches will be conducted in eight databases. For data screening and extraction, two reviewers will use a modified Critical Appraisal and Data Extraction for Systematic Reviews of Prediction Modelling Studies checklist. We will assess the included studies using the Quality Assessment of Diagnostic Accuracy Studies guidelines. We will conduct a meta-analysis if sufficient data are available. We will use hierarchical summary receiver operating characteristic curves to estimate the summary operating points, including the pooled sensitivity and specificity, with 95% CIs, if pooling is appropriate. Furthermore, if sufficient data are available, we will use Grading of Recommendations, Assessment, Development and Evaluations profiler software to summarise the main findings of the systematic review, as a summary of results.

Ethics and dissemination There are no ethical considerations associated with this study protocol, as the systematic review focuses on the examination of secondary data. The systematic review results will be used to report on the accuracy, completeness and standard procedures of the included studies. We will disseminate our findings by publishing our analysis in a peer-reviewed journal and, if required, we will communicate with the stakeholders of the studies and bibliographic databases.

PROSPERO registration number CRD42020179652.

Strengths and limitations of this study

- This systematic review will address evidence gaps in the current literature by providing a comprehensive assessment of the performance of artificial intelligence (AI) models in the diagnosis of acute ischaemic stroke and large-vessel occlusion detection.
- This systematic review will focus only on the use of brain imaging data (excluding perfusion examinations) as input for AI models.
- This systematic review will compare the performance of AI models and clinicians, using a unique search strategy to validate the findings in both internal and external validation clinical settings.
- This systematic review will be limited to evidence from non-randomised trials, and the quality of the source data from each AI study will impact the quality of the results.
- Searches will not be conducted in non-English electronic databases, which may cause a language bias.

INTRODUCTION

Background

Stroke is a significant cause of global death and disability. Generally, there are two types of strokes, ischaemic (accounting for 85% of cases) and haemorrhagic/other (accounting for 15% of cases). Ischaemic stroke with large-vessel occlusion (LVO), variably defined as intracranial anterior and posterior circulation blockages, accounts for 24%–46% of all cases of acute ischaemic stroke (AIS).[1] Diagnosing AIS involves excluding acute intracranial haemorrhage, identifying AIS signs, estimating the degree of baseline ischaemic damage, and detecting LVOs[2 3] and collaterals. In this process, stroke mimics, seizures, tumours, infection and other acute

neurologic diseases are ruled out. Medical professionals must select the appropriate treatment for patients with ischaemic stroke as quickly as possible to reduce the risk of death and disability.

In 2015, five international, multicentre, randomised, controlled prospective trials were published in the *New England Journal of Medicine*.[4–8] These trials demonstrated the importance of mechanical thrombectomy (MT) for improving the prognosis of patients with anterior circulation LVO compared with standard medical care. Subsequently, the American Heart Association/American Stroke Association (AHA/ASA) guidelines[9] updated the patient selection criteria for MT based on DAWN (time from onset: 6–24 hours) and DEFUSE-3 eligibility (time from onset: 6–16 hours) and included a medical image diagnosis as part of the stroke workflow (class I; level of evidence, A).

Generally, patients with ischaemic stroke need to reach a hospital that can treat stroke within a time window to allow for urgent clinical treatment decisions. The 2018 AHA/ASA guidelines extended the time window for endovascular therapy from 6 hours to ≤24 hours in select patients.[10] Furthermore, the 2019 AHA/ASA guidelines[11] stated that patients who wake-up with stroke symptoms >4.5 hours after they were last known to be well or have unknown time of onset may benefit from IV alteplase administration within 4.5 hours of stroke symptom recognition.

However, the time window is being increasingly replaced by the tissue window for decision making.[12 13] While the observed onset of stroke symptoms defines the time window, the tissue window reflects the biological timing of dynamically evolving ischaemia. Thus, AHA/ASA guidelines[11] recommend using brain imaging (eg, MR diffusion-weighted imaging (MR-DWI), non-invasive CT angiography (CTA) or MR angiography (MRA)) in the diagnosis of AIS and detection of LVOs (class I; level of evidence, A). Imaging modalities, such as advanced MRI and CT medical imaging, have been shown to be critical in improving clinical stroke care in clinical practice. For example, Hurford *et al*[14] reported on AIS management using a CT imaging workflow. Currently used imaging techniques include non-contrast CT (NCCT), CTA, CT perfusion (CTP) and MRI. Additionally, some hospitals use MR-DWI, MRA and perfusion MR (MRP) in the diagnosis of AIS. Notably, CT and MRI stroke protocols are hospital specific, as imaging acquisition is not standardised.

The expanding role of medical imaging in the diagnosis of stroke is gradually encouraging advancements in traditional image processing and artificial intelligence (AI) algorithms to automatically extract diagnostic information and aid specialists. On the onset of stroke, AI can support health specialists with the necessary diagnostic information to accelerate the diagnosis, while ensuring precise intervention decisions in less time. Several traditional image processing techniques used in the diagnosis of AIS include shape models, such as an atlas, active contours and region growing, which can assist clinicians in extracting diagnostic information. However, limitations of conventional models include their inability to scale performance with more data, limited customisability and strong dependence on precomputed features. AI can overcome these limitations, as it automates the feature extraction process generally performed using traditional image processing techniques. It differentiates the input images, deriving complex image features by importing the feature semantics into the classifier.[15–17] Several AI models have been validated for infarct and penumbra segmentation in diagnosing ischaemic stroke on MRI in ISLES challenges (2015–2018).[18 19] For LVO detection on CTA,[16] a novel AI architecture known as Y-Net. Meijs and Manniesing[20] use a three-dimensional fully convolutional neural network (CNN) to segment cerebral vasculature. Other studies have used traditional image processing and classifiers such as a random-forest model to detect LVOs.[17 21] Furthermore, CTA collateral scoring based on AI models has been used to identify patients most likely to benefit from MT.[22]

Liu *et al*[23] reported that AI model performance is generally similar to that of healthcare (HC) professionals. However, only a few studies have externally validated results and even these are poorly reported. Murray *et al*[24] reviewed studies on the use of AI in the diagnosis of AIS and detection of LVOs published between January 2014 and February 2019 in three bibliographic databases: PubMed, Medline and Embase. Although the review by Murray *et al*[24] used loosely similar inclusion/exclusion criteria to those in our proposed study protocol, it lacked a diagnostic performance assessment of AI models compared with that of HC professionals in internal/external validation settings. Additionally, formal systematic review assessment tools, such as a bias risk assessment and summarisation of the model performance via meta-analysis, were not used. Suh *et al*[25] reviewed studies on the feasibility of using various CTP data to predict haemorrhagic transformation only (no ischaemic changes) in AIS published before 30 October 2018 in two bibliographic databases: Ovid Medline and Embase. The authors performed a meta-analysis of 15 articles, comprising over 1134 patients. The articles were heterogeneous, according to Higgins I^2 metrics, and multiple sub-group analyses were performed to solve this problem. Furthermore, Domingues *et al*[26] presented a general survey on the use of deep learning (DL) architectures in imaging contexts (both three-dimensional and two-dimensional), extensively covering 180 articles on positron emission tomography and CT modalities published from 2014 to 2019. Although the paper is mostly narrative in style and lacks the use of systematic review tools, it discusses interesting problems in the use of DL models in medical imaging, such as the non-availability of imaging data with labels, overfitting, differences in imaging acquisitions and interpretability in the HC context.

Objective

Considerable efforts have been made to implement AI models in clinical practice, and demonstration of AI models that are equal to, or exceed, human clinicians in terms of diagnostic performance always generates interest. However, a critical appraisal and thorough review of the body of evidence supporting AI-based diagnostics are required. Therefore, in this study protocol, we propose a rigorous systematic review to address the gaps identified above. Our main objective is to provide a comprehensive assessment of the performance of AI models in the diagnosis of AIS and LVO detection using brain imaging data (excluding perfusion examinations) as input. We will compare the results of AI models and clinicians using a unique search strategy to validate the findings in both internal and external validation clinical settings. We will assess potential biases and take into account issues in the study design, clinical value and reporting. Furthermore, if sufficient data are available, our secondary objective is to conduct a meta-analysis to summarise the diagnostic performance of AI algorithms in comparison to that of clinicians, other models (DL vs other models), and classifiers (such as CNN, support vector machine), random-forest and long short-term memory networks). Accordingly, we will be able to calculate pooled estimates of sensitivity and specificity. Additionally, we will perform subgroup analyses and a sensitivity analysis as part of our meta-analysis, as required.

METHODS AND ANALYSIS

Information sources

We will extract studies from eight online bibliographic databases: OvidSP, Web of Science, Scopus, TRIP, ProQuest, CINAHL, IEEE and Embase. As AI methods generally became effective in 2012, we will search for articles published from January 2012 to May 2020. We will identify additional pertinent articles by searching for the data sources referenced in the studies that meet inclusion criteria. We will also manually search the reference list of all included studies to identify missed articles.

Search strategy

Titles and abstracts will be searched using the queries indicated in table 1. Search terms such as “machine learning” and “deep learning” will be combined by using Boolean operators (eg, AND, OR) to create the queries. The results of different queries will be combined using the Boolean operator, AND. Wildcard characters (*) will be used to indicate any other character in a string (eg, patient* includes patients, scan* includes scans; see query #4 in table 1).

Inclusion and exclusion criteria

We will adhere to the Preferred Reporting Items for Systematic Reviews and Meta-analyses Protocols guidelines,[27] and reporting guidelines specific to prediction studies, as appropriate.[28] In accordance with the PICOTS (population, interventions, comparators, outcomes, timing, and setting) framework,[29] inclusion and exclusion criteria will be based on the type of patients, interventions, comparisons and outcomes, as shown in table 2. We will exclude data from non-human and duplicate studies.

Study overview

A diagnosis of AIS can be determined based on several independent imaging protocols, such as MRI/MRA with perfusion-weighted imaging, NCCT+CTA+CTP, NCCT+CTA for selected patients or NCCT and direct angiography. Perfusion studies (eg, CTP/MRP) are not within the scope of this study protocol. Although the identified articles might not have AI models built for all imaging protocols, we will consider AI models built on imaging modalities towards the study outcomes indicated in table 3.

Table 1 Search strategy applied to OvidSP, Web of Science, Scopus, TRIP, ProQuest, CINAHL, IEEE and Embase databases

Search	Query
#1	“machine learning”[Title] OR “deep learning”[Title] OR “artificial intelligence”[Title] OR AI[Title] OR “neural network*“[Title] OR “vector machine”[Title] OR bayesian[Title] OR “deep-learning”[Title] OR “reinforcement learning”[Title] OR “reinforcement learning”[Title] OR “deep neural network”[Title] OR “deep belief network”[Title] OR “convolutional neural network”[Title] OR “recurrent neural network”[Title] OR “feedforward neural network”[Title] OR “Boltzmann machine”[Title] OR “long short term memory”[Title] OR “gated recurrent unit”[Title] OR “rectified linear unit”[Title] OR autoencoder[Title] OR backpropagation[Title] OR “multilayer perceptron”[Title] OR convnet[Title] OR “convolutional learning”[Title]
#2	DICE[“Abstract”]OR “AVERAGE SYMMETRIC SURFACE DISTANCE “[“Abstract”]OR “HAUSDORFF DISTANCE”[“Abstract”]OR “symmetric surface distance “[“Abstract”]OR Jaccard[“Abstract”]OR DSC[“Abstract”] OR roc[“Abstract”]OR auc[“Abstract”]OR “goodness of fit”[“Abstract”]OR performance[“Abstract”]OR discriminate[“Abstract”]OR discrimination[“Abstract”]OR calibrate[“Abstract”]OR calibration[“Abstract”]OR accuracy[“Abstract”]OR sensitivity[“Abstract”]OR specificity[“Abstract”]OR recall[“Abstract”]OR precision[“Abstract”]OR collateral[“Abstract”]OR collaterals[“Abstract”]OR “collateral score”[“Abstract”]OR “clot burden score”[“Abstract”]
#3	“ischemic stroke”[“Abstract”]OR “large vessel occlusion”[“Abstract”]OR LVO[“Abstract”]
#4	patient*[“Abstract”]OR subject*[“Abstract”]OR scan*[“Abstract”]OR image[“Abstract”]OR volume[“Abstract”]
#5	[2012–2020]
#6	#1 AND #2 AND #3 AND #4 AND #5

Table 2 Inclusion and exclusion criteria

PICOS	Inclusion	Exclusion
P—Population	Patients with AIS, with MR and CT images	Stroke mimics (chronic disease, trauma, etc)
I—Intervention	AI/machine learning/DL algorithms using MR and CT imaging data	Non-imaging-based models
C—Comparator	1. Manual: Usual HC professionals using the standard of care, without AI intervention (HC vs AI) 2. Semi-Automatic methods/other models (AI vs others)	No comparisons
O—Outcome	Diagnosis of AIS and detection of LVOs and collaterals	Other stroke types
S—Setting	Observational studies (prospective and retrospective cohort studies, diagnostic accuracy studies, and case-control studies)	RCTs

AI, artificial intelligence; AIS, acute ischaemic stroke; DL, deep learning; HC, healthcare; LVO, large-vessel obstruction; RCT, randomised controlled trials.

Various thresholds are applied to CT and MRP images to compute and define infarct and penumbra volumes on CTP.[30 31] Furthermore, to detect infarcts on MR-DWI images, the apparent diffusion coefficient map is computed and a threshold of $<600\times1e^{-6}$ mm^2/s is applied. Similarly, in MRP cases, a threshold of <4–6 s is applied to the Tmax map to compute the penumbra volume. We will pool the data accordingly, based on the above differences, to conduct a systematic review and meta-analysis, and if required, we will perform the necessary subgroup analyses.

Detection of LVOs by quantification of the intracranial thrombus extent is measured using the clot burden score.[2 3] Additionally, the presence/absence of collaterals determines the patient's outcome for all cases of strokes. If the patient has good collaterals, there is more time to treat. Patients selected for MT based on a significant penumbra supported by collaterals, with a small core, can be treated up to 24 hours after onset and beyond. Most commonly, on CT, the Alberta Stroke Programme Early CT score (ASPECTS) and collaterals (not NCCT aspects),[32] as well as the scores of Christoforidis *et al*[33 34] and Miteff *et al*,[35] and CTA collateral scores[36] are used for collateral assessment.[37] Various AI models address these study outcomes, broadly defining the problems as classification, segmentation, scoring and correlation models (as mapped in table 3).

Data management

We will manage references from the bibliographic databases using Zotero standalone software V.5.0.88 (https://www.zotero.org/). Articles will be identified, screened and reviewed using Zotero. We will record the inclusion/exclusion status based on the predetermined eligibility criteria (table 2). Duplicate records will be detected and removed using the Zotero Duplication Detection feature or by using an Excel macro to manually compare the

Table 3 Study outcomes mapped to imaging protocols and AI models

Study outcome	Imaging protocols	AI models (type—input)
Diagnosis of AIS	NCCT and direct angiography	Scoring models—ASPECTS Segmentation models—Dense MCA sign Classification models—AIS vs non-AIS, TOAST classification, haemorrhagic vs AIS classification
	NCCT+CTA+CTP* (optional)	NCCT—same as above CTA: Classification models—AIS vs non-AIS, same as that for the 'LVO detection' outcome
	MRI/MRA+MRP*	MRI: Segmentation models—Infarct volume on DWI/ADC (in cc)
LVO detection	CTA, 4D-CTA, FD-CT, CTP*, TOF-MRI, MRA	Classification models—LVO vs non-LVO Scoring models—Clot burden score[2]
Collateral detection	Single-phase CTA, multiphase CTA, 4D-CTA, CTP*	Scoring models—Collateral score[37]

*Not within the scope of this study protocol.
ADC, apparent diffusion coefficient; AI, artificial intelligence; AIS, acute ischaemic stroke; ASPECTS, Alberta Stroke Programme Early CT Score; CTP, CT perfusion; 4D-CTA, 4-dimensional CT angiogram; DWI, diffusion-weighted imaging; FD-CT, flat-detector CT; LVO, large-vessel obstruction; MCA, middle cerebral artery; MRA, MR angiography; MRP, perfusion MR; NCCT, non-contrast CT; TOAST, Trial of Org 10 172 in Acute Stroke Treatment; TOF-MRI, time-of-flight MRI.

Table 4 Modified CHARMS data extraction form for the included studies

S. no	Data item
1	Study information (eg, author name, study date, study source)
2	Patient eligibility (eg, inclusion criteria, exclusion criteria, mean age, %LVO and %AIS)
3	Patient recruitment: (eg, no of centres, setting, phase of stroke onset)
4	Candidate predictors (eg, MR and CT medical images)
5	Type of study outcome (eg, ischaemic stroke diagnosis, LVO detection)
6	Sample size (no of patients used in training/validation/testing datasets)
7	Model development—model algorithm (eg, neural network, DL, traditional image processing, etc)
8	Model performance: 1. Classification measures (eg, sensitivity, specificity, TP, TN, FN, FP, threshold for sensitivity/specificity) 2. Segmentation measure: (eg, Dice score, accuracy, etc) 3. Scoring measures (eg, RMSE, MSE, MAE, etc) 4. Correlation measures (eg, MLV vs ALV, etc)
9	Model validation (eg, fivefold cross-validation, internal or external clinical validation, HC professional comparison available)
10	Results 1. General performance measures 2. Any alternate AI model presentations, for example, such as ASPECTS, infarct volume, clot burden scores or collateral scores
11	Interpretation: Models (confirmatory vs exploratory—ie, models useful for real-world clinical practice vs models requiring more analysis); comparisons across studies discussing generalisability; strengths and limitations

AI, artificial intelligence; AIS, acute ischaemic stroke; ALV, automatically segmented lesion volumes; ASPECTS, Alberta Stroke Programme Early CT Score; CHARMS, Critical Appraisal and Data Extraction for Systematic Reviews of Prediction Modelling Studies; DL, deep learning; FN, false negative; FP, false positive; HC, healthcare; LVO, large-vessel occlusions; MAE, mean absolute error; MLV, manually segmented lesion volumes; MSE, mean square error; RMSE, root mean square error; TN, true negative; TP, true positive.

digital object identifiers. Full-text articles will be imported into Zotero and analysed.

Study screening and data extraction

In the title/abstract review phase, two independent reviewers (BS and MKV) will perform the initial screening, manually searching the results and selecting articles for full-text retrieval. They will autonomously determine eligibility by screening the titles and abstracts of the retrieved studies and will retrieve data using a predefined data extraction form based on Critical Appraisal and Data Extraction for Systematic Reviews of Prediction Modelling Studies checklist, adjusted according to our systematic review-specific requirements (table 4). Consensus in the extracted data will be verified and any differences will be resolved by discussion and revaluation of the article. If a contradiction persists, the article will be referred to a third reviewer (SRK or SPG) for a final decision. The source imaging data used in each study will impact the results, potentially causing information bias. For example, studies may use validation through resampling, internal validation by split samples followed by cross-validation, or external validation using a separate dataset from a different geographical region.

We will use similar sample sizes and stroke phase to assess and group the source imaging data.

Quality assessment

The risk of bias will be initially evaluated by two independent reviewers (BS and SRK). A third reviewer (SPG) will then review each study using the Quality Assessment of Diagnostic Accuracy Studies (QUADAS-2) guidelines.[38] We will classify the risk of bias as 'low', 'high' or 'unclear'. We will summarise the risk of bias across individual studies in a narrative summary in the systematic review phase. We will perform the QUADAS-2 analysis using Review Manager V.5.4[39] or STATA V.16.1[40] software.

Meta-analysis

The techniques and assessments in AI-based studies vary and do not adhere to specific principles. In this study, we will evaluate the diagnostic performance of clinicians and AI models using only imaging data as input. We will base the reference standard (comparison) on studies presenting results with clinician-based external or internal validation. Some of the included studies may use other models as the reference standard. Due to a lack of a fixed reference standard, we cannot adequately compare these models. Thus, we will summarise the performance of AI models and clinicians separately in the systematic review and exclude these data from the meta-analysis.

We will perform a meta-analysis to obtain the pooled estimate of the effect if the studies provide adequate data. Two reviewers (BS and MKV) will extract the two-way confusion matrix table and calculated metrics, such as the sensitivity and specificity, with confidence intervals to perform the meta-analysis. A third reviewer (SRK or SPG) will resolve the extracted data's differences through discussion. Some included studies might not directly report all of the required data in a 2×2 table or the sensitivity and specificity, with CIs. We will compute the missing

data using existing/reported study data, such as the sensitivity, specificity or 2×2 confusion matrix table, and the number of cases, with the calculator in Review Manager V.5.4 or Microsoft excel macros. Moreover, a single study might use the same AI model or clinicians for different classification tasks. We will consider only the related tasks for this systematic review. In multiple comparisons, we will include a single 2×2 table, reporting the highest accuracy for each comparison. If sufficient data are available for a meta-analysis, we will use the hierarchical summary receiver operating characteristic (HSROC) curve to calculate pooled estimates (sensitivity and specificity with 95% CI) for the diagnostic performance of AI models and clinicians based on the 2×2 confusion matrix table. If relevant, we will perform a sensitivity analysis (ie, limiting the analysis to studies at the lowest risk of bias in a secondary analysis) to incorporate the risk of bias assessments into the synthesis. We will use STATA V.16.1 to plot the HSROC curves. Different diagnostic thresholds used in the included studies may cause heterogeneity. Hence, we will consider the studies to be heterogeneous and will conduct our analysis accordingly. In case we have sufficient data to determine heterogeneity, we will use the I^2 statistic (I^2 >60% indicates significant heterogeneity) to assess heterogeneity. We will adopt a random-effects model for the meta-analysis and use Review Manager V.5.4 and STATA V.16.1 to perform the meta-analysis. If sufficient data are available, we will use the Grading of Recommendations, Assessment, Development and Evaluations (GRADE) approach to evaluate the certainty of the evidence. Using GRADE profiler software, we will present the main findings of the systematic review as a summary of the findings table.

Subgroup analysis

Although the diagnosis of AIS encompasses all diagnostic decisions, we will perform a separate subgroup analysis for the detection of LVOs and collaterals. In the clinical workflow, an LVO is identified only after an initial suspicion of AIS. Table 3 shows the mapping of AI models to the diagnosis of AIS and detection of LVOs and collaterals. Obtaining a diagnosis within the time window of <24 hours is critical, as the patient is eligible for treatments such as MT. We will assess the performance of the AI model and standard reference as per the 'model performance' considerations defined in table 4.

Patient and public involvement

Patients were not involved in the development of the research question, outcome measures and study design.

ETHICS AND DISSEMINATION

There are no ethical considerations associated with this study protocol, as the systematic review focuses on the examination of secondary data. On completion of the analysis, we will prepare a manuscript for publication in a peer-reviewed medical journal.

LIMITATIONS AND STRENGTHS OF THE STUDY PROTOCOL

The study protocol has some limitations. First, the review will not consider the non-diagnostic benefits of AI in stroke pathways; criteria such as improving clinicians' confidence in diagnosing AIS/LVO with the help of AI clinical decision support tools and increasing the number of patients undergoing reperfusion treatments (thrombolysis or MT) are not included. Second, AI algorithms are dynamic, and improvement in their performance is dependent on the variety and degree of data. Hence, AI algorithm performance assessment on few data sources may be inconclusive and less generalisable. Third, there is a lack of random controlled trials and clinical performance trials comparing AI products or various algorithms/combinations. The primary studies are predominantly retrospective in nature and hence, the study quality will influence the quality of the systematic review. Fourthly, there is a lack of information on the source data used by AI algorithms, which constitutes an inherent bias in evaluating the performance of AI products in clinical trials/practice. Fifth, a significant constraint is that different studies on the diagnosis of AIS and LVO detection by AI report different metrics that are not generalisable between AI systems. Additionally, AI systems are evaluated against different standards, based on different data and methods, impacting the quality of the systematic review. Sixth, AI models on penumbra imaging (CTP/MRP) will be excluded, even though this is an important area in stroke care; however, perfusion imaging is optional in AIS. Finally, the exclusion of non-English studies may lead to an oversight of relevant articles.

With the development of AI models, clinicians are attempting to use AI-based software in clinical practices (eg, for the diagnosis of AIS and LVO detection). However, the accuracy metrics of these models vary significantly. Despite several studies claiming that AI models can improve stroke diagnosis and LVO detection accuracy, the proof is constrained due to a low usage of suitable systematic evaluation techniques and quality assessment tools. A systematic review evaluating the diagnostic performance of AI models should accurately reflect the current performance evidence with greater scientific rigour. Overall, this systematic review is important for assessing and critically appraising the evidence on the use of AI models in clinical practice. It could provide impactful insights into the diagnosis of AIS in the future.

Contributors SRK and SPG outlined the systematic review protocol and composed the manuscript. SRK, BS and MKV defined the concepts, search items, data extraction process and contributed to the final written manuscript. SPG critically revised the manuscript and approved the final version.

Funding The authors have not declared a specific grant for this research from any funding agency in the public, commercial or not-for-profit sectors.

Disclaimer The views of the authors do not necessarily reflect those of the NHS, NIHR or the Department of Health.

Competing interests None declared.

Patient and public involvement Patients and/or the public were not involved in the design, or conduct, or reporting, or dissemination plans of this research.

Patient consent for publication Not required.

Provenance and peer review Not commissioned; externally peer reviewed.

ORCID iD

Srinivasa Rao Kundeti http://orcid.org/0000-0003-0082-1530

REFERENCES

1 Rennert RC, Wali AR, Steinberg JA, *et al*. Epidemiology, natural history, and clinical presentation of large vessel ischemic stroke. *Neurosurgery* 2019;85:S4–8.
2 Tan IYL, Demchuk AM, Hopyan J, *et al*. CT angiography clot burden score and collateral score: correlation with clinical and radiologic outcomes in acute middle cerebral artery infarct. *AJNR Am J Neuroradiol* 2009;30:525–31.
3 Treurniet KM, Yoo AJ, Berkhemer OA, *et al*. Clot burden score on baseline computerized tomographic angiography and intra-arterial treatment effect in acute ischemic stroke. *Stroke* 2016;47:2972–8.
4 Berkhemer OA, Fransen PSS, Beumer D, *et al*. A randomized trial of intraarterial treatment for acute ischemic stroke. *N Engl J Med* 2015;372:11–20.
5 Goyal M, Demchuk AM, Menon BK, *et al*. Randomized assessment of rapid endovascular treatment of ischemic stroke. *N Engl J Med* 2015;372:1019–30.
6 Saver JL, Goyal M, Bonafe A, *et al*. Stent-retriever thrombectomy after intravenous t-PA vs. t-PA alone in stroke. *N Engl J Med* 2015;372:2285–95.
7 Campbell BCV, Mitchell PJ, Kleinig TJ, *et al*. Endovascular therapy for ischemic stroke with perfusion-imaging selection. *N Engl J Med* 2015;372:1009–18.
8 Jovin TG, Chamorro A, Cobo E, *et al*. Thrombectomy within 8 hours after symptom onset in ischemic stroke. *N Engl J Med* 2015;372:2296–306.
9 Powers WJ, Derdeyn CP, Biller J, *et al*. 2015 American heart Association/American stroke association focused update of the 2013 guidelines for the early management of patients with acute ischemic stroke regarding endovascular treatment: a guideline for healthcare professionals from the American heart Association/American stroke association. *Stroke* 2015;46:3020–35.
10 Powers WJ, Rabinstein AA, Ackerson T, *et al*. 2018 guidelines for the early management of patients with acute ischemic stroke: a guideline for healthcare professionals from the American heart Association/American stroke association. *Stroke* 2018;49:e46–110.
11 Powers WJ, Rabinstein AA, Ackerson T, *et al*. Guidelines for the early management of patients with acute ischemic stroke: 2019 update to the 2018 guidelines for the early management of acute ischemic stroke: a guideline for healthcare professionals from the American heart Association/American stroke association. *Stroke* 2019;50:e344–418.
12 Xiang X, Cao F. Time window and "tissue window": two approaches to assist decision-making in strokes. *J Neurol* 2019;266:283–8.
13 Wang D, Wang Y. Tissue window, not the time window, will guide acute stroke treatment. *Stroke Vasc Neurol* 2019;4:1–2.
14 Hurford R, Sekhar A, Hughes TAT, *et al*. Diagnosis and management of acute ischaemic stroke. *Pract Neurol* 2020;20:304–16.
15 Menze BH, Jakab A, Bauer S, *et al*. The multimodal brain tumor image segmentation benchmark (BRATS). *IEEE Trans Med Imaging* 2015;34:1993–2024.
16 Li C, Xie Y, Sun J. 3D intracranial artery segmentation using a convolutional autoencoder. *IEEE International Conference on Bioinformatics and Biomedicine (BIBM)*, 2017:714–7.
17 Meijs M, Patel A, van de Leemput SC, *et al*. Robust segmentation of the full cerebral vasculature in 4D CT of suspected stroke patients. *Sci Rep* 2017;7:15622.
18 Winzeck S, Hakim A, McKinley R, *et al*. ISLES 2016 and 2017-benchmarking ischemic stroke lesion outcome prediction based on multispectral MRI. *Front Neurol* 2018;9:679.
19 Maier O, Menze BH, von der Gablentz J, *et al*. ISLES 2015 - A public evaluation benchmark for ischemic stroke lesion segmentation from multispectral MRI. *Med Image Anal* 2017;35:250–69.
20 Meijs M, Manniesing R. Artery and vein segmentation of the cerebral vasculature in 4D CT using a 3D fully convolutional neural network. *Medical Imaging*2018;10575:105751Q.
21 Zhao S, Zhou M. An effective brain vasculature segmentation algorithm for time-of-flight MRA data. *International Conference on Virtual Reality and Visualisation (ICVRV)*, 2015:238–45.
22 Grunwald IQ, Kulikovski J, Reith W, *et al*. Collateral automation for triage in stroke: evaluating automated scoring of collaterals in acute stroke on computed tomography scans. *Cerebrovasc Dis* 2019;47:217–22.
23 Liu X, Faes L, Kale AU, *et al*. A comparison of deep learning performance against health-care professionals in detecting diseases from medical imaging: a systematic review and meta-analysis. *Lancet Digit Health* 2019;1:e271–97.
24 Murray NM, Unberath M, Hager GD, *et al*. Artificial intelligence to diagnose ischemic stroke and identify large vessel occlusions: a systematic review. *J Neurointerv Surg* 2020;12:156–64.
25 Suh CH, Jung SC, Cho SJ, *et al*. Perfusion CT for prediction of hemorrhagic transformation in acute ischemic stroke: a systematic review and meta-analysis. *Eur Radiol* 2019;29:4077–87.
26 Domingues I, Pereira G, Martins P, *et al*. Using deep learning techniques in medical imaging: a systematic review of applications on CT and PET. *Artif Intell Rev* 2020;53:4093–160.
27 Moher D, Shamseer L, Clarke M, *et al*. Preferred reporting items for systematic review and meta-analysis protocols (PRISMA-P) 2015 statement. *Syst Rev* 2015;4:1.
28 Bouwmeester W, Zuithoff NPA, Mallett S, *et al*. Reporting and methods in clinical prediction research: a systematic review. *PLoS Med* 2012;9:e1001221–12.
29 The Cochrane Collaboration. Chapter 5: Defining the review question and developing criteria for including studies. In: Higgins JPT, Green S, eds. *Cochrane Handbook of systematic reviews. version 5.0.1*, 2008.
30 Lin L, Bivard A, Krishnamurthy V, *et al*. Whole-Brain CT perfusion to quantify acute ischemic penumbra and core. *Radiology* 2016;279:876–87.
31 Røhl L, Ostergaard L, Simonsen CZ, *et al*. Viability thresholds of ischemic penumbra of hyperacute stroke defined by perfusion-weighted MRI and apparent diffusion coefficient. *Stroke* 2001;32:1140–6.
32 ASPECTS. Alberta Stroke Program Early CT score (ASPECTS)—Collateral scoring. Available: http://www.aspectsinstroke.com/collateral-scoring/ [Accessed 15 Apr 2015].
33 Christoforidis GA, Karakasis C, Mohammad Y, *et al*. Predictors of hemorrhage following intra-arterial thrombolysis for acute ischemic stroke: the role of pial collateral formation. *AJNR Am J Neuroradiol* 2009;30:165–70.
34 Christoforidis GA, Mohammad Y, Kehagias D, *et al*. Angiographic assessment of pial collaterals as a prognostic indicator following intra-arterial thrombolysis for acute ischemic stroke. *AJNR Am J Neuroradiol* 2005;26:1789–97.
35 Miteff F, Levi CR, Bateman GA, *et al*. The independent predictive utility of computed tomography angiographic collateral status in acute ischaemic stroke. *Brain* 2009;132:2231–8.
36 Menon BK, d'Esterre CD, Qazi EM, *et al*. Multiphase CT angiography: a new tool for the imaging triage of patients with acute ischemic stroke. *Radiology* 2015;275:510–20.
37 Seker F, Potreck A, Möhlenbruch M, *et al*. Comparison of four different collateral scores in acute ischemic stroke by CT angiography. *J Neurointerv Surg* 2016;8:1116–8.
38 Whiting PF, Rutjes AWS, Westwood ME, *et al*. QUADAS-2: a revised tool for the quality assessment of diagnostic accuracy studies. *Ann Intern Med* 2011;155:529–36.
39 The Cochrane Collaboration. *Review Manager (RevMan) [Computer program]. Version 5.4*, 2020.
40 StataCorp. *Stata statistical software: release 16*. College Station, TX: StataCorp LLC, 2019.