REVIEW ARTICLE

Machine Learning Algorithms to Predict the Risk of Rupture of Intracranial Aneurysms: a Systematic Review

Karan Daga[1,2] · Siddharth Agarwal[1] · Zaeem Moti[2] · Matthew B. K. Lee[3] · Munaib Din[2] · David Wood[1] · Marc Modat[1] · Thomas C. Booth[1,4]

Abstract

Purpose Subarachnoid haemorrhage is a potentially fatal consequence of intracranial aneurysm rupture, however, it is difficult to predict if aneurysms will rupture. Prophylactic treatment of an intracranial aneurysm also involves risk, hence identifying rupture-prone aneurysms is of substantial clinical importance. This systematic review aims to evaluate the performance of machine learning algorithms for predicting intracranial aneurysm rupture risk.

Methods MEDLINE, Embase, Cochrane Library and Web of Science were searched until December 2023. Studies incorporating any machine learning algorithm to predict the risk of rupture of an intracranial aneurysm were included. Risk of bias was assessed using the Prediction Model Risk of Bias Assessment Tool (PROBAST). PROSPERO registration: CRD42023452509.

Results Out of 10,307 records screened, 20 studies met the eligibility criteria for this review incorporating a total of 20,286 aneurysm cases. The machine learning models gave a 0.66–0.90 range for performance accuracy. The models were compared to current clinical standards in six studies and gave mixed results. Most studies posed high or unclear risks of bias and concerns for applicability, limiting the inferences that can be drawn from them. There was insufficient homogenous data for a meta-analysis.

Conclusions Machine learning can be applied to predict the risk of rupture for intracranial aneurysms. However, the evidence does not comprehensively demonstrate superiority to existing practice, limiting its role as a clinical adjunct. Further prospective multicentre studies of recent machine learning tools are needed to prove clinical validation before they are implemented in the clinic.

Keywords Aneurysm · Intracranial Aneurysm · Machine Learning · Subarachnoid Haemorrhage · Cerebrovascular

Introduction

The prevalence of intracranial aneurysms in the general population is approximately 3.2%. [1] Rupture risk can vary greatly depending on aneurysm morphology, for example the relative risk of rupture of larger aneurysms (>15 mm) is 15.4 in comparison to smaller (<5 mm) aneurysms [2]. Similarly, patient demographics can play a significant role as smokers have a relative risk of rupture of 1.7 in comparison to non-smokers, and female patients have a relative risk of 2.0. [2] The management of unruptured intracranial aneurysms (UIA) is contextualised by patient age, comorbidity and clinical presentation in addition to the estimated rupture risk. Cerebral aneurysms can be managed with surgical procedures such as microsurgical clipping, endovascular treatment (including coiling, stenting or intrasaccular device implantation), or they can be monitored with imaging without initial intervention.

It is essential to stratify rupture risk to guide appropriate management. Although no universally-agreed reference standard criteria exists for prognostication and defini-

✉ Thomas C. Booth
thomas.booth@kcl.ac.uk

1 School of Biomedical Engineering & Imaging Sciences, King's College London, BMEIS, King's College London. 1 Lambeth Palace Road, London, UK SE1 7EU, UK

2 Guy's and St. Thomas' NHS Foundation Trust, Westminster Bridge Road, London, UK SE1 7EH, UK

3 University College London Hospital NHS Foundation Trust, 235 Euston Rd, London, UK NW1 2BU, UK

4 Department of Neuroradiology, King's College Hospital, Denmark Hill, London, UK SE5 9RS, UK

tion, scoring systems have been devised as decision support tools. A commonly applied example is the PHASES score; 'Population, Hypertension, Age, Size of aneurysm, Earlier subarachnoid haemorrhage (SAH) from another aneurysm and Site of aneurysm' [3]. Studies assessing the PHASES score demonstrated sensitivity and specificity of 0.50 and 0.81, respectively, during a total follow-up of 3064 person-years, indicating insufficient predictive accuracy when applied independently in clinical practice [4]. Other commonly used scoring criteria are the 'Unruptured Intracranial Aneurysm Treatment Score' (UIATS), the 'International Study on Unruptured Intracranial Aneurysms' (ISUIA) score and the 'Unruptured Cerebral Aneurysm Study' (UCAS) score [5–7]. Consequently, the variety of rupture risk scoring systems has resulted in heterogeneity in clinical management planning and patient follow up. For example, UIA size is typically considered to be one of the most predictive criteria for rupture risk but is also considered by many to be inadequate in isolation given the high prevalence of rupture in aneurysms <3 mm [8, 9]. Because three-dimensional (3D) morphology is also an important risk factor, increasingly morphology has been used in decision making [10].

Given that there remains an unmet clinical need to optimise and homogenise decision making for UIAs, and given that machine learning (ML) has shown great promise in healthcare with its capability to process large multimodal datasets (resulting in a wide range of use-cases from diagnostics to management planning), it is plausible that ML-based data-driven decision tools will be added to the armamentarium of clinicians. However, whilst several recent ML models have been developed for this purpose, there has been a lack of implementation in the clinic. This study therefore aimed to systematically review and summarize the accuracy of ML models to predict UIA rupture. The review process allowed us to explore the limitations and implications of using UIA decision support tools, and provides a baseline for future research.

Materials and Methods

This systematic review is PROSPERO registered (International prospective register of systematic reviews) (CRD42023452509). The review followed the Preferred Reporting Items for Systematic Reviews and Meta-Analysis (PRISMA) guidelines, informed by the Checklist for Artificial Intelligence in Medical Imaging (CLAIM), and Cochrane review methodology for developing inclusion and exclusion criteria, search methodology and quality assessment [11–14].

Search Strategy and Selection Criteria

A 'sensitive search' strategy (Supplementary Material) was undertaken consisting of relevant search terms and subject headings, including exploded terms and Medical Subject Headings (MeSH) terms, without language restrictions [13]. The search was conducted in the following medical databases: EMBASE, MEDLINE, Web of Science and the Cochrane Register to include articles published until December 2023. Pre-prints, conference abstracts and non-peer reviewed articles were excluded (Fig. 1). Three reviewers (radiologists with 2, 2 and 7 years neurovascular research experience, respectively) all performed the literature search and selection; consensus was achieved with a fourth reviewer (15 years neurovascular research experience).

Inclusion Criteria

This review included primary studies incorporating any ML methodology to predict the rupture risk of UIAs. In a subset of studies that evaluated 'stability' of an aneurysm (as opposed to rupture), the authors defined stability as a composite outcome which included rupture as well as aneurysm growth and/or presence of symptoms.

Exclusion Criteria

Studies were excluded if epidemiological and assessing individual risk factors without developing a predictive model, or developing models that only predicted current rupture status (i.e., models performing binary classification of aneurysms as 'ruptured' or 'unruptured'). Animal studies and studies in paediatric populations (age <18 years) were also excluded.

Index Test and Reference Standard

The index test was the ML model predicting rupture risk of an aneurysm. In most studies, the reference standard was either the event of rupture or 'stability' during follow up. A few studies used an established prediction model (e.g., PHASES) as the reference standard, specified in Tables 1 and 2.

Data Extraction and Analysis

We extracted data on imaging modality used; reference standard employed; period of risk assessment/follow-up; index test ML model; model features (grouped as clinical, morphological or hemodynamic features); study population; datasets (training/validation/testing); and inclusion/exclusion criteria. We recorded information on whether testing data sets were used for either internal or external val-

Fig. 1 PRISMA flow diagram to illustrate the studies included for qualitative review
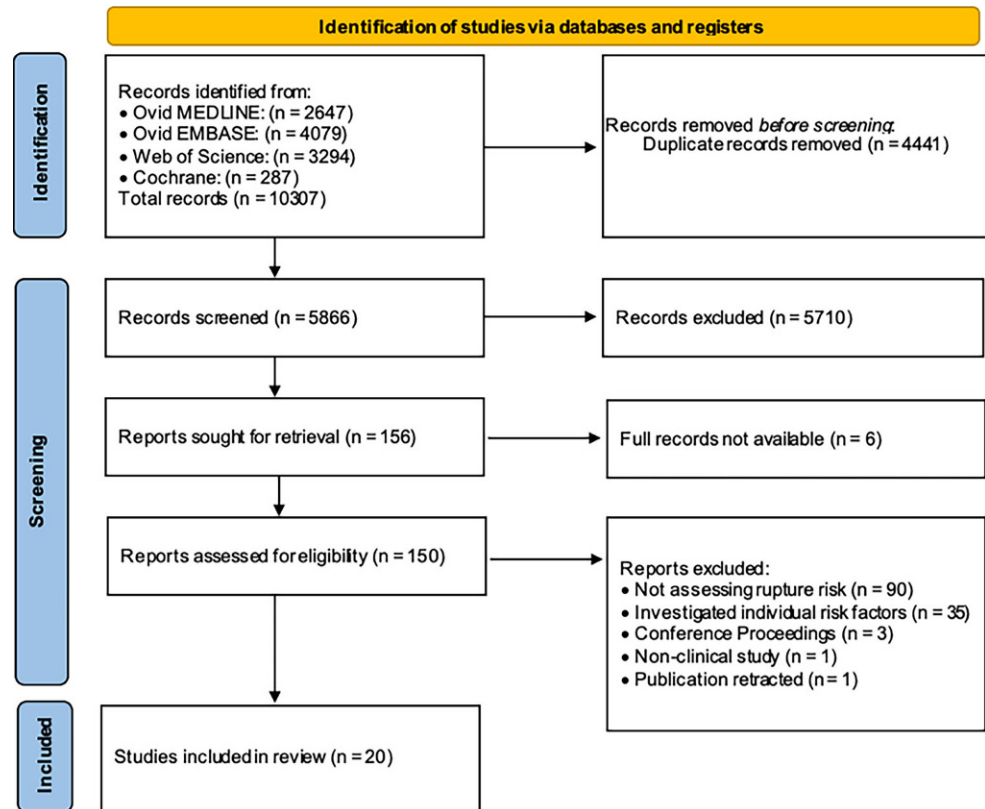
idation. Internal validation consists of testing a model on any form of internal 'hold-out' data including temporally-spaced datasets from the same institution. External validation consists of testing a model using data that was separated geographically (different institutions) from the training datasets. The performance accuracy of the best performing ML model or composite ML model were obtained, with the majority presented as receiver operating characteristic area under the curve (ROC-AUC); where possible we also calculated other measures of accuracy including sensitivity, specificity, balanced accuracy and F1 score. Study quality assessment was performed using the Prediction model of Risk Of Bias Assessment Tool (PROBAST) to assess the risk of bias and concerns regarding applicability [15].

Results

Characteristics of Included Studies

The database searches yielded 10,307 records that met the search criteria, from which 156 were identified as potentially eligible full-text articles (Fig. 1). Ninety studies were excluded as they did not assess rupture risk, 35 were excluded as they investigated correlation of UIA rupture with individual risk factors but did not predict rupture risk, the full texts were not available for 6 studies, and 5 were ex-

cluded as they were conference proceedings, non-clinical studies or were retracted. Twenty eligible studies were included (Tables 1 and 2) which spanned from 2003 to 2023, incorporating a total of 20,286 aneurysms [3, 6, 16–33]. A subset of studies included hold-out test set data with 9/20 studies (45%) only employing training data. All validation was analytical validation (i.e., within a computational research setting), no clinical validation was undertaken (i.e., not within a clinical pathway) [34]. The index test model was prospectively validated in 4/20 (20%) studies, and retrospectively validated in 8/20 (40%) studies; the remaining 8/20 (40%) were 'development only' studies. Six studies (6/20, 30%) employed the use of digital subtraction angiography (DSA) or 3D-DSA; 6/20 (30%) used computer tomography angiography (CTA); 3/20 (15%) used a combination of CTA, DSA and magnetic resonance imaging (MRA); and 5/20 (25%) did not specify the imaging modality used. All studies employed as model features a selection of clinical (e.g., age, sex, medical comorbidities), morphological (e.g., size, site, aneurysm or vessel angle, surface area) and/or hemodynamic factors (e.g., wall shear stress, oscillatory shear index, normalized pressure average) [16–33].

Table 1 Summary of findings for existing studies applying machine learning to predict unruptured intracranial aneurysm rupture risk

Publication	Study Design	Modality	Reference standard	Time period of risk assessment	Comparison to clinical practice	Index test	Model features	Hold-out test set (*n*) (or other specified dataset)	Hold-out test set performance accuracy (or performance of other specified dataset)
Weibers et al. 2003 [6]	Development only	Not stated	Comparison to result (ruptured/unruptured)	5 years	–	1. Linear regression (CPH)	(1) Morphological features	(1692 training set—no hold-out test set)	N/A
Grieving et al., 2014 [3]	Development only	CTA/DSA/MRA	Comparison to result (ruptured/unruptured)	5 years	–	1. Linear regression (CPH)	(1) Clinical features (2) Morphological features	IV: 8328 (Bootstrapping using training data, no hold-out test set per se)	AUC: 0.82
Malik et al. 2018 [16]	Development only	DSA	Comparison to risk interpretation of the UIAs was conducted by expert neurosurgeons	Duration not stated	Comparison to neurosurgeon expert opinions for reference standard	1. MLP Neural network	(1) Morphological features	(12 training set—no hold-out test set)	(Accuracy: 0.86)
Jiang et al., 2018 [17]	Development and prospective validation	CTA	Comparison to result (ruptured/unruptured)	18–27 months	–	1. Logistic regression	(1) Morphological features (2) Hemodynamic features	IV:70 (temporal validation)	Accuracy: 0.67 Balanced Accuracy: 0.69 Sensitivity: 0.71 Specificity: 0.67 AUC: 0.72 PPV: 0.19 NPV: 0.95 F1 Score: 0.30
Suzuki et al., 2019 [18]	Development and retrospective validation	Not stated	Comparison to result (ruptured/unruptured)	2 years	–	1. Logistic regression* 2. SVM	(1) Clinical features (2) Morphological features (3) Hemodynamic features	IV: 102	Accuracy: 0.82 Balanced Accuracy: 0.74 Sensitivity: 0.64 Specificity: 0.85 PPV: 0.333 NPV: 0.95 F1 Score: 0.44
Ahn, et al. 2021 [19]	Development only	3D DSA	Comparison to risk interpretation of the UIAs was conducted by two readers, disagreements arbitrated by a third reader	Duration not stated	Comparison to neuro-radiologist expert opinions for reference standard	1. Neural network	(1) Morphological features	IV: 93	Accuracy: 0.82 Balanced Accuracy: 0.82 Sensitivity: 0.82 Specificity: 0.82 PPV: 0.80 NPV: 0.83 F1 Score: 0.81

Table 1 (Continued)

Publication	Study Design	Modality	Reference standard	Time period of risk assessment	Comparison to clinical practice	Index test	Model features	Hold-out test set (n) (or other specified dataset)	Hold-out test set performance accuracy (or performance of other specified dataset)
Ou et al., 2021 [20]	Development and retrospective validation	CTA	Comparison to result (ruptured/ unruptured)	2 years	–	Combination model: 1. Logistic regression 2. LASSO 3. Ridge regression	(1) Morphological features (2) Radiomics features	IV: 122 (10-fold cross validation on training data, no hold-out test set per se)	Accuracy: 0.77 Balanced Accuracy: 0.80 Sensitivity: 0.72 Specificity: 0.88 AUC: 0.88 PPV: 0.92 NPV: 0.61 F1 Score: 0.81
Van der Kamp, et al. 2021 [21]	Development only	CTA/DSA/ MRA	Comparison to result (ruptured/ unruptured)	6 months, 1 year, 2 years	–	Combination model: 1. Linear regression (CPH) 2. Ridge regression	(1) Morphological features	(329 training set—no hold-out test set)	(AUC: 0.72)
Walther, et al. 2022 [22]	Development and retrospective validation	Not stated	Comparison to result (ruptured/ unruptured)	Duration not stated	Comparison to PHASES and UIATS	1. Gradient boosting machine	(1) Clinical features (2) Morphological features	IV: 446 (5-fold cross validation on training data, no hold-out test set per se)	Accuracy: 0.78 Balanced Accuracy: 0.78 Sensitivity: 0.86 Specificity: 0.70 AUC: 0.86 PPV: 0.74 NPV: 0.84 F1 Score: 0.80
Ou et al., 2022 [23]	Development and retrospective validation	CTA	Comparison to result (ruptured/ unruptured)	2 years	Comparison to individual 5 neurosurgeons aided by AI. (Reader + AI)	1. Neural network* 2. LASSO	(1) Clinical features (2) Morphological features	EV:120	Accuracy: 0.85 Balanced Accuracy: 0.87 Sensitivity: 0.89 Specificity: 0.84 AUC: 0.85 PPV: 0.69 NPV: 0.97 F1 Score: 0.88
Wei, et al. 2022 [24]	Development only	3D DSA	Comparison made to the PHASES score, patients divided into high and low risk group as reference standard	Duration not stated	Comparison to PHASES for reference standard	1. Logistic regression	(1) Clinical features (2) Morphological features (3) Hemodynamic features	IV: 39 (Cross validation on training data, no hold-out test set per se)	Accuracy: 0.85 Balanced Accuracy: 0.84 Sensitivity: 0.78 Specificity: 0.90 AUC: 0.85 PPV: 0.88 NPV: 0.83 F1 Score: 0.82

Table 1 (Continued)

Publication	Study Design	Modality	Reference standard	Time period of risk assessment	Comparison to clinical practice	Index test	Model features	Hold-out test set (n) (or other specified dataset)	Hold-out test set performance accuracy (or performance of other specified dataset)
Malik, et al. 2023 [25]	Development and retrospective validation	Not stated	Comparison to result (ruptured/unruptured)	(i) 5 year and (ii) life time	Comparison to PHASES, UIATS and neurosurgeons expert opinion	*Combination model:* 1. Logistic regression 2. Decision tree classifier 3. Random forest 4. AdaBoost 5. Gradient boosting 6. K-nearest neighbors 7. XGBoost	(1) Clinical features (2) Morphological features	IV: 30	Accuracy: 0.66 Sensitivity: 0.67 PPV: 0.73 F1-Score: 0.64
Xie, et al. 2023 [26]	Development and retrospective validation	CTA	Comparison to result (ruptured/unruptured)	Duration not stated	–	*Combination model:* 1. Neural network 2. LASSO 3. SVM	(1) Clinical features (2) Morphological features (3) Radiomics features	EV: 106 (3-fold cross validation)	Accuracy: 0.90 Balanced Accuracy: 0.87 Sensitivity: 0.80 Specificity: 0.94 AUC: 0.89 PPV: 0.81 NPV: 0.92 F1 Score: 0.81
Li, et al. 2023 [27]	Development only	Not stated	Comparison to result (ruptured/unruptured)	Duration not stated	–	1. Random forest* 2. Logistic regression 3. Principal component analysis	(1) Clinical features	IV: 1325 (10-fold cross validation on training data, no hold-out test set per se)	AUC: 0.73

Results to 2 decimal places

Key:

*Best Performing ML Model: Where more than one machine learning model was investigated in parallel and not in combination, the best performing model is marked by *

AUC Area under the curve of a receiver operating curve, *LASSO* Least absolute shrinkage and selection operator, *SVM* Support Vector Machine, *MLP* Multilayer Perceptron, *UIA* Unruptured Intracranial aneurysm, *CPH* Cox Proportional Hazards Model, *AI* Artificial Intelligence, *CTA* Computer Tomography Angiogram, *PHASES* Population, Hypertension, Age, Size of aneurysm, Earlier subarachnoid hemorrhage from another aneurysm and Site of aneurysm, *ELAPSS* Earlier subarachnoid hemorrhage, aneurysm Location, Age, Population, aneurysm Size and Shape, *IARS* Intracranial Aneurysm Rupture Score, *UIATS* Unruptured Intracranial Aneurysm Treatment Score, *MRA* Magnetic Resonance Angiogram, *DSA* Digital Subtraction Angiogram, *d,p* decimal places, *IV* Internal Validation, *EV* External Validation, *PPV* Positive Predicitve Value, *NPV* Negative Predictive Value

Table 2 Summary of findings for existing studies applying machine learning to predict unruptured intracranial aneurysm 'stability'. The authors defined stability as a composite outcome which included rupture as well as aneurysm growth and/or presence of symptoms

Publication	Study Design	Modality	Reference standard	Time period of risk assessment	Definition of stability	Comparison to clinical practice	Index test	Model features	Hold-out test set (n) (or other specified dataset)	Hold-out test set performance accuracy (or performance of other specified dataset)
Liu et al. 2019 [28]	Development and prospective validation	DSA	Comparison to result (stable/ unstable)	1 month stability assessment (follow-up median: 11.5 months, range: 3–26 months)	1. Remained unruptured 2. No UIA growth 3. Asymptomatic	–	1. Generalized linear model* 2. Ridge regression 3. Logistic regression	(1) Clinical features (2) Morphological features	IV: 124	AUC: 0.86
Zhu, et al. 2020 [29]	Development and retrospective validation	3D-DSA	Comparison to result (stable/ unstable)	1 month stability assessment (median follow up 15.6 months; range 5–39 months)	1. Remained unruptured 2. No UIA growth	–	1. Neural Network* 2. Random forest 3. SVM	(1) Clinical features (2) Morphological features	IV: 411	Accuracy: 0.82 Balanced Accuracy: 0.72 Sensitivity: 0.52 Specificity: 0.93 AUC:0.87 PPV: 0.71 NPV: 0.85 F1 Score: 0.60
Yang, et al. 2021 [30]	Development and prospective validation	CTA	Comparison to result (stable/ unstable)	3 years	1. Remained unruptured 2. UIA Growth ≤ 20%	Comparison made to PHASES, ELAPSS, UIATS and IARS Score	1. Neural network	(1) Clinical features (2) Morphological features (3) Hemodynamic features	IV: 37 (9-fold cross validation on training data, no hold-out test set per se)	AUC: 0.83
Liu, et al. 2022 [31]	Development and prospective validation	CTA	Comparison to result (stable/ unstable)	2 years	1. Remained unruptured 2. UIA of aneurysm <20% or <1 mm	Comparison made to PHASES and ELAPSS	1. Logistic regression	(1) Clinical features (2) Morphological features (3) Hemodynamic features	IV: 97	AUC: 0.94
Zhang, et al. 2023 [32]	Development and retrospective validation	CTA/ MRA	Comparison to result (stable/ unstable)	2 years	1. Remained unruptured 2. No IA growth 3. Asymptomatic	–	1. SVM* 2. Logistic regression 3. Adaboost	(1) Hemodynamic features	EV: 54	Accuracy: 0.83 Balanced Accuracy: 0.83 Sensitivity: 0.83 Specificity: 0.83 AUC: 0.89 PPV: 0.71 NPV: 0.91 F1 Score: 0.77

Table 2 (Continued)

Publication	Study Design	Modality	Reference standard	Time period of risk assessment	Definition of stability	Comparison to clinical practice	Index test	Model features	Hold-out test set (*n*) (or other specified dataset)	Hold-out test set performance accuracy (or performance of other specified dataset)
Irfan, et al. 2023 [33]	Development only	DSA	Comparison to risk interpretation of the UIAs was conducted by expert neurosurgeons	Duration not stated	Not defined	Comparison to neurosurgeon expert opinions for reference standard	*Combination model:* 1. Neural network 2. Decision tree classifier	(1) Morphological features (2) Hemodynamic features	IV: 141	Accuracy: 0.85 Balanced Accuracy: 0.85 Sensitivity: 0.84 Specificity: 0.86 AUC: 0.93 PPV: 0.82 NPV: 0.87 F1 Score: 0.83

Reference Standards

The reference standard for 16/20 (80%) studies was the occurrence of the stated primary endpoint during follow up: in 11/16 (68.8%) studies this was the occurrence of rupture, and in 5/16 (31.2%) this was 'stability' (the composite primary endpoint of rupture, volumetric growth, and/or presence of symptoms). The duration of follow up was stated in 13/16 (81.3%) of these studies.

The primary reference standard for 3/20 studies (15%) was the UIA risk prediction based arbitrarily on neurosurgeons/neuroradiologist expert opinion. The reference standard for 1/20 (5%) study was the PHASES score for UIA rupture risk prediction.

Index Tests

The index test for each included study involved a form of ML to predict UIA rupture risk or 'stability'. Fourteen (14/20, 70%) studies investigated UIA rupture risk prediction, and 6/20 (30%) aneurysm 'stability'. Some studies investigated the predictive performance of more than one ML model, either independently or as a combination of multiple ML algorithms. The best performing index test (presented in Tables 1 and 2) was a regression or classical ML model in 10/20 (50%) of studies, a deep learning model in 5/20 (25%) of studies, and a combination of ML algorithms in 5/20 (25%) of studies. From the available data, the hold-out test set accuracy range of the regression or classical ML models was 0.67–0.85, deep learning models 0.82–0.85 and combination models 0.66–0.90.

Test Sets

Three studies (3/20, 15%) used geographically separate datasets for training and hold-out testing (external validation) whereas 8/20 (40%) studies used hold-out test sets from the same institution (internal validation). One internal validation study (1/8, 12.5%) used a temporal split. The remaining 9/20 (45%) studies had no hold-out test set and to mitigate overfitting, six of these (6/9, 66.7%) performed either cross validation or bootstrapping.

UIA Rupture Risk Prediction

The overall accuracies of the ML models of the 14/20 (70%) studies that predicted rupture risk ranged between 0.66–0.90 (Table 1), either provided in the publication or calculated by constructing confusion matrices for the purpose of this systematic review.

UIA Stability Prediction

Of the 6/20 (30%) studies that investigated prediction of UIA 'stability', 3/6 (50%) defined stability as a composite end point of growth or rupture in a specified time period, 2/6 (33.3%) selected a composite end point of growth, rupture or presence of symptoms, and 1/6 (16.7%) did not define stability. The overall accuracies of the ML 'stability' models ranged between 0.83–0.94 (Table 2).

Comparison of Index Test to Other Reference Standards

As a secondary objective, 6/20 (30%) studies compared the ML model performance to a secondary reference standard—either the prospective predictive performance of either PHASES/UIATS scores or an expert neurosurgeon/neuroradiologist assessment of rupture risk [22, 23, 25, 30, 31, 33]. As a primary objective, 3/20 (15%) studies compared the ML model performance to the PHASES score or expert assessments, but did not compare the performance to rupture or 'stability' [16, 19, 24]. ML models had a higher performance accuracy than PHASES/UIATS scores, however, were less accurate than expert clinician predictions. In one study the performance accuracy with PHASES score, the ML model, and expert neurosurgical prediction alone were 0.50, 0.66 and 0.73 AUC-ROC respectively [25]. Another study compared the ML model, expert neurosurgical prediction alone (reader alone) and expert neurosurgical opinion aided by the ML model (reader and model), giving an AUC-ROC of 0.85, 0.88 and 0.95, respectively [23].

Bias Assessment and Concerns Regarding Applicability

An analysis of the risk of bias and concerns regarding applicability was performed for each study using the PROBAST tool (Fig. 2; [15]). Two independent reviewers performed the risk of bias assessment, with disagreements resolved by a third reviewer. The PROBAST tool categorizes prediction studies as 'development' or 'validation', which we have assigned as 'training' and 'external validation', respectively, to represent the use of ML in these prediction studies (Note: 'internal validation', whilst reasonably robust when performed temporally in particular, is not considered as 'validation' when explicitly performing PROBAST assessment; nonetheless we include granular information on both internal and external validation in Tables 1 and 2). The studies included in this review were all examples of analytical validation (i.e., within a computational research setting) [34]. Only 3/20 (15%) studies had geographically separate training sets to their testing sets (external validation).

A 'high' risk of bias or concern for applicability was identified in at least one domain in 13/20 (65%) of studies. Notably, there were high risks of bias in the 'outcome' domain (5/20, 25%) and 'analysis' domain (8/20, 40%). In terms of concerns for applicability in the 'outcome' domain, this was 'unclear' in 11/20 (55%) of studies. These primarily stemmed from studies without a defined duration of follow up to ascertain rupture or 'stability' (assigned as unclear risk), or studies that used subjective reference standards such as expert assessment of rupture risk (assigned as high risk). The inclusion and exclusion criteria were not specified in 7/20 (35%) of studies.
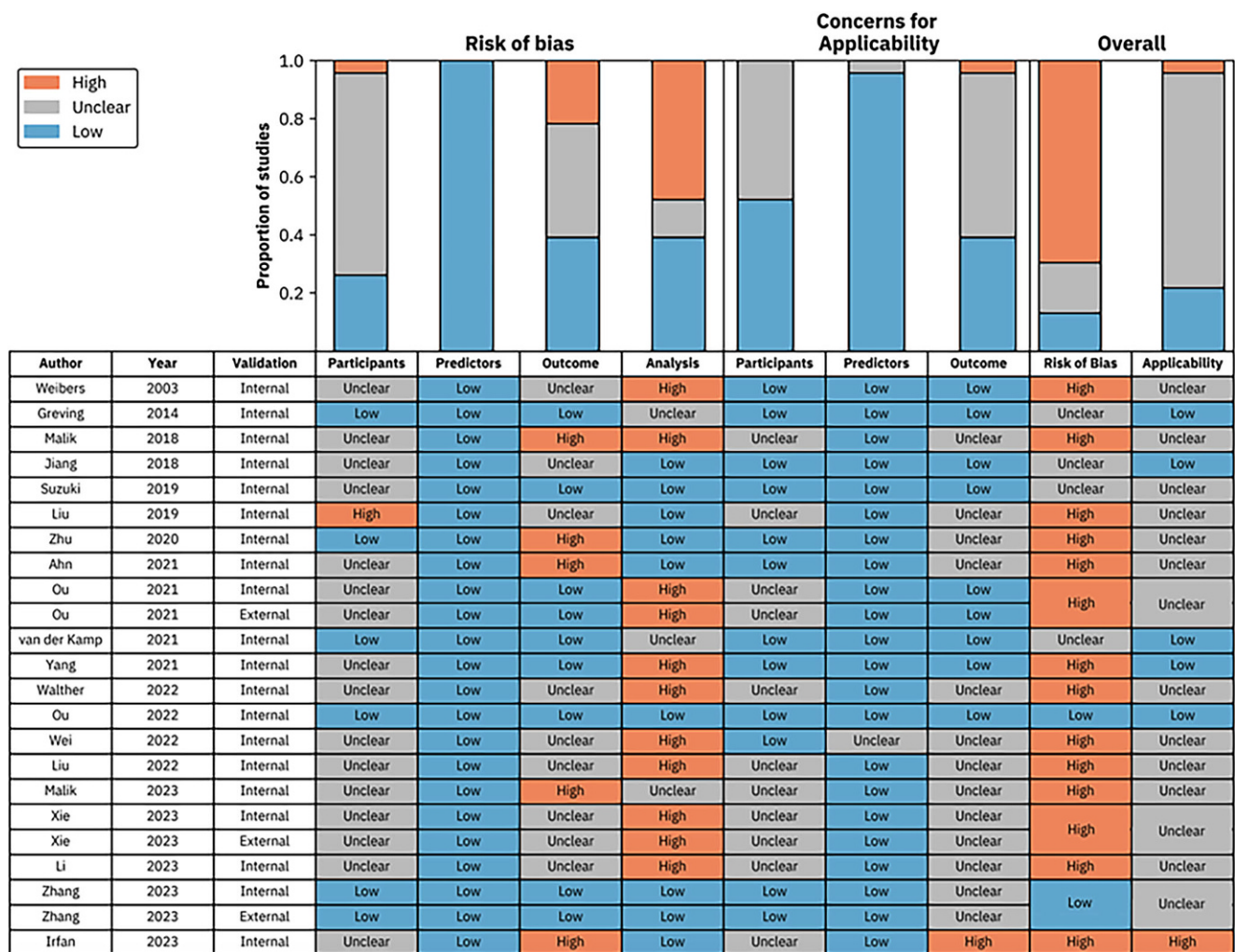
Discussion

Summary of Studies

The included studies reported high performance accuracies for predicting aneurysm rupture in hold-out tests sets ranging from 0.66–0.90 using ML Models [16–27]. However, the majority of these studies demonstrated a high risk of bias and concerns regarding applicability due to their methodology (Fig. 2), limiting the clinical applicability of their results. Six studies had a composite end-point of 'stability' which included rupture but has debatable clinical applicability. Only three studies used external test sets to validate their performance [23, 26, 32]. Some studies compared rupture risk prediction to existing clinical methods of prediction, with accuracy of ML models falling between current scoring criteria such as PHASES, and human expert risk prediction.

Current Evidence

There is one existing systematic review and meta-analysis for the use of ML in predicting aneurysm rupture risk published in 2022 by Shu et al. [35]. However, only four studies were included and three of these studies predicted current rupture status (i.e., models performing binary classification of aneurysms as 'ruptured' or 'unruptured'), not risk of UIA rupture [36–38]. Therefore, to our knowledge, this is the first review to comprehensively assess studies predicting UIA rupture risk using ML. Other studies not included in this review claim to assess the performance of ML models in predicting UIA rupture risk. However, these studies failed to meet the inclusion criteria for our review as they used the *post-rupture* imaging appearance of ruptured aneurysms to assume pre-rupture morphology, and had no subsequent validation with unruptured cases [38–43]. Furthermore, several longitudinal studies demonstrate that post-rupture morphology is not an adequate substitute to assume pre-rupture morphology when assessing UIA rupture risk as the event

Fig. 2 Summary of the risk of bias and concerns for applicability using PROBAST guidelines explicitly

Author	Year	Validation	Participants	Predictors	Outcome	Analysis	Participants	Predictors	Outcome	Risk of Bias	Applicability
Weibers	2003	Internal	Unclear	Low	Unclear	High	Low	Low	Low	High	Unclear
Greving	2014	Internal	Low	Low	Low	Unclear	Low	Low	Low	Unclear	Low
Malik	2018	Internal	Unclear	Low	High	High	Unclear	Low	Unclear	High	Unclear
Jiang	2018	Internal	Unclear	Low	Unclear	Low	Low	Low	Low	Unclear	Low
Suzuki	2019	Internal	Unclear	Low	Low	Low	Low	Low	Low	Unclear	Unclear
Liu	2019	Internal	High	Low	Unclear	Low	Unclear	Low	Unclear	High	Unclear
Zhu	2020	Internal	Low	Low	High	Low	Low	Low	Unclear	High	Unclear
Ahn	2021	Internal	Unclear	Low	High	Low	Low	Low	Unclear	High	Unclear
Ou	2021	Internal	Unclear	Low	Low	High	Unclear	Low	Low	High	Unclear
Ou	2021	External	Unclear	Low	Low	High	Unclear	Low	Low		
van der Kamp	2021	Internal	Low	Low	Low	Unclear	Low	Low	Low	Unclear	Low
Yang	2021	Internal	Unclear	Low	Low	High	Low	Low	Low	High	Low
Walther	2022	Internal	Unclear	Low	Unclear	High	Unclear	Low	Unclear	High	Unclear
Ou	2022	Internal	Low	Low	Low	Low	Low	Low	Low	Low	Low
Wei	2022	Internal	Unclear	Low	Unclear	High	Low	Unclear	Unclear	High	Unclear
Liu	2022	Internal	Unclear	Low	Unclear	High	Unclear	Low	Unclear	High	Unclear
Malik	2023	Internal	Unclear	Low	High	Unclear	Unclear	Low	Unclear	High	Unclear
Xie	2023	Internal	Unclear	Low	Unclear	High	Unclear	Low	Unclear	High	Unclear
Xie	2023	External	Unclear	Low	Unclear	High	Unclear	Low	Unclear		
Li	2023	Internal	Unclear	Low	Unclear	High	Unclear	Low	Unclear	High	Unclear
Zhang	2023	Internal	Low	Low	Low	Low	Low	Low	Unclear	Low	Unclear
Zhang	2023	External	Low	Low	Low	Low	Low	Low	Unclear		
Irfan	2023	Internal	Unclear	Low	High	Low	Unclear	Low	High	High	High

of rupture can modify the morphology and haemodynamics of an intracranial aneurysm [44–46]. Epidemiological studies were also excluded if they investigated individual risk factors for rupture but did not develop a predictive model, such as presenting odds ratios for hypertension, aneurysm size or specific morphological features, and aneurysm rupture [47–49].

The current published validated standards include, for example, the PHASES, UIATS, ISUIA or 'Earlier subarachnoid haemorrhage, aneurysm Location, Age, Population, aneurysm Size and Shape' (ELAPSS) scores to guide management of UIA [3, 5, 6, 50]. The UIATS score was developed by a Delphi consensus, and the ELAPSS score assesses aneurysm growth not rupture, hence neither were included in this review [5, 50]. The PHASES score and ISUIA criteria were developed by Cox Proportional Hazard (CPH) analysis, which is based on linear regression modelling [3–6, 51]. The PHASES model was developed from a large multicentre cohort of 8382 aneurysms. Boot-

strapping was performed on this training data to account for overfitting and achieved a concordance-statistic value of 0.82 (95% CI 0.79–0.85), however the model was not validated on a hold-out test set (i.e. no internal or external validation) [3]. The ISUIA study did not account for overfitting nor were the results validated on a hold-out test set [6]. Nevertheless, models have widely been adopted in practice [3–6, 22, 25, 52].

Role in Clinical Practice

The role of aneurysm rupture risk prediction is primarily to guide clinicians and patients towards management options. Decision making requires balancing the risks of intervention (thromboembolic ischaemia, peri-operative haemorrhage, inadequate occlusion, need for re-intervention and death) with the risk of SAH and its subsequent sequelae [6, 53–56]. The ISUIA studies reported an overall morbidity and mortality rate for open neurosurgical and endovascu-

lar procedures of 7.1–12.6%, and a meta-analysis of 2460 patients reported a permanent morbidity rate of 10.9%. [6, 54] However, it should be noted that these landmark studies were published over two decades ago and significant advances in endovascular interventions have substantially reduced morbidity and mortality [55, 56]. A recent meta-analysis of 963 aneurysms treated with an endovascular approach demonstrated a morbidity of 2.85% and mortality of 0.93% [56]. As treatment prediction morbidity and mortality data evolves, so too should rupture risk prediction to allow careful cost-benefit decision making. For now, rupture risk prediction typically involves a pragmatic multi-disciplinary team (MDT) approach incorporating, for example, PHASES, UIATS or ISUIA scores alongside neuroradiological and neurosurgical evaluation of clinical, morphological and hemodynamic factors [3, 5, 6]. Overall, this complex decision making leads to substantial heterogeneity. Here, novel ML models have the potential to automate and standardize the process of accurately identifying rupture-prone aneurysms. We have shown the high accuracy of ML models, but also have shown that they are not ready for the clinic as higher quality bias-free evidence is needed, as well as clinical validation (i.e., tested within a clinical pathway).

In six studies, the composite end-point of 'stability' was defined as a combination of the event of rupture, growth in subsequent imaging and/or presence of symptoms within a specified time period (1–39 months) [28–33]. However, several studies have suggested that UIA growth and rupture have a nonlinear and complex relationship [21, 57–59]. Therefore, employing stability-based reference standards as a surrogate for risk of rupture may be inaccurate. After all, approximately 3.2% of the general population are expected to have an intracranial aneurysm of any size which almost certainly will have grown at some point in time, but a far fewer proportion of the population experience SAH [1]. Nevertheless, there still remain several studies demonstrating a correlation between growth and rupture which is expected as there is an overlap between the growth risk factors and rupture risk factors [8, 59]. A meta-analysis of 4990 aneurysms (13,294 aneurysm-years of follow up) found that growth makes an aneurysm over 30 times more likely to rupture with an annual rupture rate of 3.1% in growing aneurysms [8]. In summary, aneurysm growth is correlated with rupture, and stability can be used as a reference standard for modelling aneurysm risk, but it is not as accurate as rupture being used as the reference standard.

Limitations

There are several limitations regarding the studies included in this review. First, only three studies demonstrated test set data that was geographically separate from training data (i.e., externally validated) [22, 26, 32]. Eight studies used an internal hold-out test set for validation and nine studies had no test set at all. Subsequently there is significant heterogeneity within the methodology of the existing literature. ML algorithms are capable of high accuracy and recall from repetitively being trained on a given data set. Studies without any testing or validation sets are at high risk of overfitting and subsequently have poor external validity. Moreover, with a lack of clinical validation, the generalisability of the results from these studies to the clinic is limited.

Second, as demonstrated by the risk of bias assessment (Fig. 2), several studies demonstrated a high or unclear risk of bias or concerns for applicability. Flaws in the methodology and analysis of several papers, for instance due to ambiguous reference standards, limit the validity of the results predicted. Five studies did not specify a duration of risk assessment or follow-up, such as specifying an annual, 2-year or lifetime risk of rupture. Of these, three studies were also at high risk of observer bias as their reference standard employed was an expert interpretation of UIA rupture risk [16, 19, 33]. The lack of standardisation between studies make it difficult to appreciate rupture risk across different cohorts. Studies with a short follow-up period are vulnerable to underestimating rupture rates in practice. Conversely, attrition to follow up of stable aneurysms may overestimate rupture risks.

Third, studies investigating prediction of UIA rupture risk are at an inherent risk of selection bias. Naturally, due to ethical concerns of leaving participants untreated when they are at high rupture risk, unstable UIA with known risk factors for rupture are less likely to be longitudinally followed up to a pre-specified endpoint. Consequently, studies excluding high risk unstable UIA that were treated may bias the ML models [19]. Additionally, several studies included specific inclusion and exclusion criteria, such that aneurysms of certain size, location or morphology were excluded. This introduces further risk of selection bias. Furthermore, even subgroup analyses based on such inclusion criteria are challenging because these specific criteria differed amongst the included studies.

Regarding limitations of our systematic review methodology, heterogeneity between study methods, reference standards and testing groups (such as follow-up duration) made it unfeasible to conduct a meta-analysis. Second, the exclusion of pre-prints and non-peer reviewed articles may contribute to publication bias. Due to the mismatch in the pace of peer review compared to developments in data-science, data-science oriented teams may be less inclined to seek publication in peer-reviewed journals than clinically-oriented teams [60].

Implications for Future Research and Clinical Practice

The systematic review serves as a baseline for future research. Despite the potential for homogenous and autonomous data driven decision-making incorporating multiple co-variates, there remains a need for further high-quality studies in this field as ML prediction models are not ready for deployment given the limited evidence, high risks of bias and concerns for applicability. Multicentre, prospectively-maintained international registries of untreated UIAs are needed to train and test ML models. Furthermore, clinical validation (i.e., tested within a clinical pathway) is paramount to establishing the capability of ML in practice. Requirements for clinical validation [34] are to run a prospective clinical study with: (1) a specified end-point (rupture is an unambiguous end-point, and the pre-defined duration would be for as long as study funding would allow); (2) a specified predictive tool (index test) already supported by analytical validation; (3) a clear purpose, including the setting, for which the predictive tool will be used (MDT assessment of intracranial aneurysms); (4) an understanding of the potential benefits and harms associated with that use (a benefit is that rupture risk can be compared to treatment risk in terms of health outcomes allowing informed decisions to be made in either routine clinical practice or for prognostic enrichment in interventional trials; a possible harm is that the predictive tool itself is not entirely proven and therefore may underestimate or overestimate rupture risk; another possible harm is that the predictive tool may lead to uncertainty because there is no risk threshold above or below which a treatment unequivocally should be triggered—therefore treatment discussions remain on a case-to-case basis informed by the risk); (5) a process for collecting and analyzing information about the performance of the predictive tool is devised and carried out (an example would be to include resources in a study to allow frequent follow up and to mitigate drop out by incorporating national epidemiological data collection tools; for example, Hospital Episode Statistics (HES) is a data warehouse containing details of all admissions to NHS hospitals in England). In higher risk aneurysms it is likely that most patients will elect for treatment meaning the recruited cohort will need to be very large.

It is noteworthy that the majority of included studies were developed by training and testing amongst Chinese populations [16–33]. As suggested by the PHASES model, ethnicity plays an integral role in risks of aneurysm rupture [3]. Therefore, studies including ethnicities other than Chinese are also needed. Additionally, ML models could be further developed to meet the needs of individual patients, such as a personalized composite of UIA rupture risk and risk of treatment akin to the UIATS score [5]. Specifically,

multi-centric prospective registries (or prospective clinical validation studies) could also include treated patients to elucidate the interaction of treatment. By including data on treated and untreated aneurysms of similar risk profiles under follow up (specified end-points would include morbidity and mortality as well as rupture), counterfactual ML prediction models could be trained to give a prediction risk for treatment and a prediction risk for no treatment.

Conclusion

In summary, this systematic review highlights innovative approaches towards rupture risk prediction of UIA using ML models to quantify rupture risk by applying features used in clinical practice. ML algorithms have the potential to automate and standardize the process of accurately identifying rupture-prone aneurysms and determining the need for an inherently high risk prophylactic intervention. However, there remains a need for further high-quality studies in this field as ML prediction models are not ready for deployment given the limited evidence, high risks of bias and concerns for applicability. Further prospective multicentre studies are needed to prove clinical validation before implementation in the clinic.

Author Contribution
1. KD: Conception and design of the review, acquisition, analysis and interpretation of the data and writing the manuscript.

2. SA: Conception and design of the review, acquisition, analysis and interpretation of the data.

3. ZM: Acquisition and analysis of the data

4. ML: Conception and design of the review, interpretation of the data.

5. MD: Conception and design of the review, acquisition of the data

6. DW: Conception and design of the review

7. MM: Conception and design of the review

8. TCB: Conception and design of the review, acquisition, analysis and interpretation of the data and writing the manuscript.

All authors have read and approved the final manuscript. TCB is the corresponding author and guarantor of the study.

Data Availability   The datasets used and/or analyzed during the current study are available from the corresponding author on reasonable request.

References

1. Vlak MH, Algra A, Brandenburg R, Rinkel GJ. Prevalence of unruptured intracranial aneurysms, with emphasis on sex, age, comorbidity, country, and time period: a systematic review and meta-analysis. Lancet Neurol. 2011;10(7):626–36.

2. Wermer MJ, van der Schaaf IC, Algra A, Rinkel GJ. Risk of rupture of unruptured intracranial aneurysms in relation to patient and aneurysm characteristics: an updated meta-analysis. Stroke. 2007;38(4):1404–10.

3. Greving JP, Wermer MJ, Brown RD, Morita A, Juvela S, Yonekura M, Ishibashi T, Torner JC, Nakayama T, Rinkel GJ, Algra A. Development of the PHASES score for prediction of risk of rupture of intracranial aneurysms: a pooled analysis of six prospective cohort studies. Lancet Neurol. 2014;13(1):59–66.

4. Juvela S. PHASES score and treatment scoring with cigarette smoking in the long-term prediction of rupturing of unruptured intracranial aneurysms. JNS. 2021;136(1):156–62.

5. Etminan N, Brown RD, Beseoglu K, Juvela S, Raymond J, Morita A, Torner JC, Derdeyn CP, Raabe A, Mocco J, Korja M. The unruptured intracranial aneurysm treatment score: a multidisciplinary consensus. Neurology. 2015;85(10):881–9.

6. Wiebers DO. Unruptured intracranial aneurysms: natural history, clinical outcome, and risks of surgical and endovascular treatment. Lancet. 2003;362(9378):103–10.

7. UCAS Japan Investigators.. The natural course of unruptured cerebral aneurysms in a Japanese cohort. N Engl J Med. 2012;366(26):2474–82.

8. Brinjikji W, Zhu YQ, Lanzino G, Cloft HJ, Murad MH, Wang Z, Kallmes DF. Risk Factors for Growth of Intracranial Aneurysms: A Systematic Review and Meta-Analysis. Ajnr Am J Neuroradiol. 2016;37(4):615–20. https://doi.org/10.3174/ajnr.A4575.

9. Daga K, Taneja M, Venketasubramanian N. Small Intracranial Aneurysms and Subarachnoid Hemorrhage: Is the Size Criterion for Risk of Rupture Relevant? Case Rep Neurol. 2020;12(1):161–8. https://doi.org/10.1159/000503094.

10. Bizjak Ž, Pernuš F, Spiclin Ž. Deep Shape Features for Predicting Future Intracranial Aneurysm Growth. Front Physiol. 2021;12:644349. https://doi.org/10.3389/fphys.2021.644349.

11. Page MJ, McKenzie JE, Bossuyt PM, Boutron I, Hoffmann TC, Mulrow CD, Shamseer L, Tetzlaff JM, Akl EA, Brennan SE, Chou R. The PRISMA 2020 statement: an updated guideline for reporting systematic reviews. Int J Surg. 2021;1(88):105906.

12. Higgins JPT, Thomas J, Chandler J, Cumpston M, Li T, Page MJ, Welch VA, editors. Cochrane Handbook for Systematic Reviews

13. of Interventions version 6.4. 2023. www.training.cochrane.org/handbook.

13. de Vet HCW, Eisinga A, Riphagen AB II. Chapter 7: searching for studies. In: Cochrane handbook for systematic reviews of diagnostic test accuracy version 0.4 The Cochrane Collaboration; 2008.

14. Mongan J, Moy L, Kahn CE Jr. Checklist for artificial intelligence in medical imaging (CLAIM): a guide for authors and reviewers. Radiol Artif Intell. 2020;2(2):e200029.

15. Wolff RF, Moons KG, Riley RD, Whiting PF, Westwood M, Collins GS, Reitsma JB, Kleijnen J, Mallett S, PROBAST Group.. PROBAST: a tool to assess the risk of bias and applicability of prediction model studies. Ann Intern Med. 2019;170(1):51–8.

16. Malik KM, Anjum SM, Soltanian-Zadeh H, Malik H, Malik GM. A framework for intracranial saccular aneurysm detection and quantification using morphological analysis of cerebral angiograms. IEEE Access. 2018;6:7970–86.

17. Jiang P, Liu Q, Wu J, Chen X, Li M, Li Z, Yang S, Guo R, Gao B, Cao Y, Wang S. A novel scoring system for rupture risk stratification of intracranial aneurysms: a hemodynamic and morphological study. Front Neurosci. 2018;5(12):596.

18. Suzuki M, Haruhara T, Takao H, Suzuki T, Fujimura S, Ishibashi T, Yamamoto M, Murayama Y, Ohwada H. Classification model for cerebral aneurysm rupture prediction using medical and blood-flow-simulation data. InICAART. 2019;2:895–9.

19. Ahn JH, Kim HC, Rhim JK, Park JJ, Sigmund D, Park MC, Jeong JH, Jeon JP. Multi-view convolutional neural networks in rupture risk assessment of small, unruptured intracranial aneurysms. JPM. 2021;11(4):239.

20. Ou C, Chong W, Duan CZ, Zhang X, Morgan M, Qian Y. A preliminary investigation of radiomics differences between ruptured and unruptured intracranial aneurysms. Eur Radiol. 2021;31:2716–25.

21. van der Kamp LT, Rinkel GJ, Verbaan D, van den Berg R, Vandertop WP, Murayama Y, Ishibashi T, Lindgren A, Koivisto T, Teo M, St George J. Risk of rupture after intracranial aneurysm growth. JAMA Neurol. 2021;78(10):1228–35.

22. Walther G, Martin C, Haase A, Nestler U, Schob S. Machine learning for rupture risk prediction of intracranial aneurysms: Challenging the PHASES score in geographically constrained areas. Symmetry. 2022;14(5):943.

23. Ou C, Li C, Qian Y, Duan CZ, Si W, Zhang X, Li X, Morgan M, Dou Q, Heng PA. Morphology-aware multi-source fusion-based intracranial aneurysms rupture prediction. Eur Radiol. 2022;32(8):5633–41.

24. Wei J, Xu Y, Ling C, Xu L, Zhu G, Jin J, Rong C, Xiang J, Xu J. Assessing rupture risk by hemodynamics, morphology and plasma concentrations of the soluble form of tyrosine kinase receptor axl in unruptured intracranial aneurysms. Clin Neurol Neurosurg. 2022;1(222):107451.

25. Malik K, Alam F, Santamaria J, Krishnamurthy M, Malik G. Toward grading subarachnoid hemorrhage risk prediction: a machine learning-based aneurysm rupture score. World Neurosurg. 2023;1(172):e19–e38.

26. Xie Y, Liu S, Lin H, Wu M, Shi F, Pan F, Zhang L, Song B. Automatic Risk Prediction of Intracranial Aneurysm on CTA Image with Convolutional Neural Networks and Radiomics Analysis. Front Neurol. 2023;14:1126949.

27. Li Y, Huan L, Lu W, Li J, Wang H, Wang B, Song Y, Peng C, Wang J, Yang X, Hao J. Integrate prediction of machine learning for single ACoA rupture risk: a multicenter retrospective analysis. Front Neurol. 2023;14.

28. Liu Q, Jiang P, Jiang Y, Ge H, Li S, Jin H, Li Y. Prediction of aneurysm stability using a machine learning model based on pyRadiomics-derived morphological features. Stroke. 2019;50(9):2314–21.

29. Zhu W, Li W, Tian Z, Zhang Y, Wang K, Zhang Y, Liu J, Yang X. Stability assessment of intracranial aneurysms using machine learn-

ing based on clinical and morphological features. Transl Stroke Res. 2020;11:1287–95.

30. Yang Y, Liu Q, Jiang P, Yang J, Li M, Chen S, Mo S, Zhang Y, Ma X, Cao Y, Cui D. Multidimensional predicting model of intracranial aneurysm stability with backpropagation neural network: a preliminary study. Neurol Sci. 2021;1:1–3.

31. Liu Q, Leng X, Yang J, Yang Y, Jiang P, Li M, Mo S, Yang S, Wu J, He H, Wang S. Stability of unruptured intracranial aneurysms in the anterior circulation: nomogram models for risk assessment. JNS. 2022;137(3):675–84.

32. Zhang M, Hou X, Qian Y, Chong W, Zhang X, Duan CZ, Ou C. Evaluation of aneurysm rupture risk based upon flowrate-independent hemodynamic parameters: a multi-center pilot study. J NeuroIntervent Surg. 2023;15(7):695–700.

33. Irfan M, Malik KM, Ahmad J, Malik G. Strokenet: an automated approach for segmentation and rupture risk prediction of intracranial aneurysm. Comput Med Imaging Graph. 2023;108:102271.

34. FDA-NIH Biomarker Working Group.. BEST (Biomarkers, EndpointS, and other Tools) Resource. 2017. https://www.ncbi.nlm.nih.gov/books/NBK464453/. Accessed 16 Nov 2020.

35. Shu Z, Chen S, Wang W, Qiu Y, Yu Y, Lyu N, Wang C. Machine learning algorithms for rupture risk assessment of intracranial aneurysms: a diagnostic meta-analysis. World Neurosurg. 2022;1(165):e137–e47.

36. Kim HC, Rhim JK, Ahn JH, Park JJ, Moon JU, Hong EP, Kim MR, Kim SG, Lee SH, Jeong JH, Choi SW. Machine learning application for rupture risk assessment in small-sized intracranial aneurysm. JCM. 2019;8(5):683.

37. Liu J, Chen Y, Lan L, Lin B, Chen W, Wang M, Li R, Yang Y, Zhao B, Hu Z, Duan Y. Prediction of rupture risk in anterior communicating artery aneurysms with a feed-forward artificial neural network. Eur Radiol. 2018;28:3268–75.

38. Tanioka S, Ishida F, Yamamoto A, Shimizu S, Sakaida H, Toyoda M, Kashiwagi N, Suzuki H. Machine learning classification of cerebral aneurysm rupture status with morphologic variables and hemodynamic parameters. Radiol Artif Intell. 2020;2(1):e190077.

39. Tong X, Feng X, Peng F, Niu H, Zhang B, Yuan F, Jin W, Wu Z, Zhao Y, Liu A, Wang D. Morphology-based radiomics signature: a novel determinant to identify multiple intracranial aneurysms rupture. Aging. 2021;13(9):13195.

40. Feng X, Tong X, Peng F, Niu H, Qi P, Lu J, Zhao Y, Jin W, Wu Z, Zhao Y, Liu A. Development and validation of a novel nomogram to predict aneurysm rupture in patients with multiple intracranial aneurysms: a multicentre retrospective study. Stroke Vasc Neurol. 2021;6(3):e480.

41. Detmer FJ, Chung BJ, Mut F, Slawski M, Hamzei-Sichani F, Putman C, Jiménez C, Cebral JR. Development and internal validation of an aneurysm rupture probability model based on patient characteristics and aneurysm location, morphology, and hemodynamics. Int J CARS. 2018;13:1767–79.

42. Yang H, Cho KC, Kim JJ, Kim JH, Kim YB, Oh JH. Rupture risk prediction of cerebral aneurysms using a novel convolutional neural network-based deep learning model. J NeuroIntervent Surg. 2023;15(2):200–4.

43. Yang H, Cho KC, Kim JJ, Kim YB, Oh JH. New morphological parameter for intracranial aneurysms and rupture risk prediction based on artificial neural networks. J NeuroIntervent Surg. 2023;15(e2):e209–e15.

44. Skodvin TØ, Johnsen LH, Gjertsen Ø, Isaksen JG, Sorteberg A. Cerebral aneurysm morphology before and after rupture: nationwide case series of 29 aneurysms. Stroke. 2017;48(4):880–6.

45. Schneiders JJ, Marquering HA, Van den Berg R, VanBavel E, Velthuis B, Rinkel GJ, Majoie CB. Rupture-associated changes of cerebral aneurysm geometry: high-resolution 3D imaging before and after rupture. AJNR Am J Neuroradiol. 2014;35(7):1358–62.

46. Cornelissen BM, Schneiders JJ, Potters WV, Van Den Berg R, Velthuis BK, Rinkel GJ, Slump CH, VanBavel E, Majoie CB, Marquering HA. Hemodynamic differences in intracranial aneurysms before and after rupture. AJNR Am J Neuroradiol. 2015;36(10):1927–33.

47. Kang H, Ji W, Qian Z, Li Y, Jiang C, Wu Z, Wen X, Xu W, Liu A. Aneurysm characteristics associated with the rupture risk of intracranial aneurysms: a self-controlled study. Plos One. 2015;10(11):e142330..

48. Detmer FJ, Chung BJ, Jimenez C, Hamzei-Sichani F, Kallmes D, Putman C, Cebral JR. Associations of hemodynamics, morphology, and patient characteristics with aneurysm rupture stratified by aneurysm location. Neuroradiology. 2019;11(61):275–84.

49. Zhong P, Lu Z, Li T, Lan Q, Liu J, Wang Z, Chen S, Huang Q. Association between regular blood pressure monitoring and the risk of intracranial aneurysm rupture: a multicenter retrospective study with propensity score matching. Transl Stroke Res. 2022;13(6):983–94.

50. Backes D, Rinkel GJ, Greving JP, Velthuis BK, Murayama Y, Takao H, Ishibashi T, Igase M, Agid R, Jääskeläinen JE, Lindgren AE. ELAPSS score for prediction of risk of growth of unruptured intracranial aneurysms. Neurology. 2017;88(17):1600–6.

51. Liu P, Fu B, Yang SX, Deng L, Zhong X, Zheng H. Optimizing survival analysis of XGboost for ties to predict disease progression of breast cancer. IEEE Trans Biomed Eng. 2020;68(1):148–60.

52. Morel S, Bijlenga P, Kwak BR. Intracranial aneurysm wall (in) stability—current state of knowledge and clinical perspectives. Neurosurg Rev. 2022;45(2):1233–53.

53. Ross IB, Dhillon GS. Complications of endovascular treatment of cerebral aneurysms. Surg Neurol. 2005;64(1):12–8.

54. Raaymakers TW, Rinkel GJ, Limburg M, Algra A. Mortality and morbidity of surgery for unruptured intracranial aneurysms: a meta-analysis. Stroke. 1998;29(8):1531–8.

55. Adamou A, Alexandrou M, Roth C, Chatziioannou A, Papanagiotou P. Endovascular treatment of intracranial aneurysms. Life. 2021;11(4):335.

56. Van Rooij SB, Sprengers ME, Peluso JP, Daams J, Verbaan D, Van Rooij WJ, Majoie CB. A systematic review and meta-analysis of Woven EndoBridge single layer for treatment of intracranial aneurysms. Interv Neuroradiol. 2020;26(4):455–60.

57. Jou LD, Mawad ME. Growth rate and rupture rate of unruptured intracranial aneurysms: a population approach. BioMed Eng OnLine. 2009;8(1):1–9.

58. Juvela S. Growth and rupture of unruptured intracranial aneurysms. JNS. 2018;131(3):843–51.

59. Backes D, Rinkel GJ, Laban KG, Algra A, Vergouwen MD. Patient-and aneurysm-specific risk factors for intracranial aneurysm growth: a systematic review and meta-analysis. Stroke. 2016;47(4):951–7.

60. Booth TC, Grzeda M, Chelliah A, Roman A, Al Busaidi A, Dragos C, Shuaib H, Luis A, Mirchandani A, Alparslan B, Mansoor N. Imaging biomarkers of glioblastoma treatment response: a systematic review and meta-analysis of recent machine learning studies. Front Oncol. 2022;31(12):799662.