

Machine Learning and Artificial Intelligence Models for Predicting Clinical Outcomes after Thoracic and Complex Endovascular Aortic Repair: A Systematic Review

Mohammad Mahadin^{1*}, Sana Shaik², Muhammad Umer Farooq Mujahid³, Leen Abuzaid⁴, Abdullah Imtiaz⁵, Leen Aburumman⁶, Shifa Mohamed Rafi², Tasneem Rashid¹, Reef Mahadin⁷

¹Prince Hamza Hospital, Amman, Jordan

²Tbilisi State Medical University, Georgia

³University of Health Sciences, Lahore, Pakistan

⁴Yarmouk University, Amman, Jordan

⁵Wah medical college, Wah cantt, Rawalpindi, Pakistan

⁶The University of Jordan, Amman, Jordan

⁷Jordan University of Science and Technology, Amman, Jordan

*Correspondence to: Mohammad Mahadin, Prince Hamza Hospital, Amman, Jordan, E-mail: mahadin.m01@gmail.com

Received: June 04, 2026; Manuscript No: JCHF-26-9903; Editor Assigned: June 06, 2026; PreQC No: JCHF-26-9903(PQ); Reviewed: June 12, 2025; Revised: June 15, 2025; Manuscript No: JCHF-26-9903(R); Published: June 30, 2026

Citation: Mahadin M, Shaik S, Mujahid MUF, Abuzaid L, Imtiaz A, et al (2026). Machine Learning and Artificial Intelligence Models for Predicting Clinical Outcomes after Thoracic and Complex Endovascular Aortic Repair: A Systematic Review. J. Cardiol. Heart Fail. Vol.2 Iss.1, June (2026), pp:64-71.

Copyright: © 2026 Mohammad Mahadin, Sana Shaik, Muhammad Umer Farooq Mujahid, Leen Abuzaid, Abdullah Imtiaz, Leen Aburumman, Shifa Mohamed Rafi, Tasneem Rashid, Reef Mahadin. This is an open-access article distributed under the terms of the Creative Commons Attribution License, which permits unrestricted use, distribution, and reproduction in any medium, provided the original author and source are credited.

ABSTRACT

Background

Thoracic endovascular aortic repair (TEVAR) and complex endovascular aortic interventions treat thoracic aortic aneurysm and Stanford type B aortic dissection, but outcomes remain heterogeneous and conventional risk tools provide limited individualized prognostication. We reviewed ML/AI prediction models after TEVAR/complex endovascular repair.

Methods

PubMed, Scopus, and Cochrane were searched from inception to December 2025 (PROSPERO CRD420251267013) per PRISMA 2020. We included adult studies reporting quantitative performance for ML/AI prognostic models and assessed risk of bias using PROBAST.

Results

Five retrospective studies (n=79-10,738) evaluated tree-based algorithms (including XGBoost), decision trees, radiomics-based ML, and deep learning using clinical and/or CT angiography features. Discrimination ranged from approximately AUC 0.70-0.99, and where compared, ML models generally outperformed conventional regression. Calibration and clinical utility were inconsistently reported; only one study used independent external validation. Heterogeneity in populations, predictors, and outcome definitions precluded meta-analysis, and overall risk of bias was moderate-to-high, mainly in the analysis domain.

Conclusion

ML/AI models show promising performance for predicting mortality, adverse events, reintervention, and aortic remodeling after TEVAR, but standardized reporting and multicenter external validation are required before clinical implementation.

Keywords: TEVAR; Thoracic Endovascular Aortic Repair; Complex Endovascular Repair; Machine Learning; Artificial Intelligence; Prediction Model; Radiomics; Deep Learning

INTRODUCTION

Thoracic endovascular aortic repair (TEVAR), approved by the FDA in 2005, has become a mainstay for treating descending thoracic aortic aneurysms and complicated Stanford type B aortic dissection, among other thoracic aortic pathologies. [1,2]

Despite advances in devices and technique, TEVAR is associated with clinically important complications, including stroke, spinal cord ischemia/paraplegia, visceral ischemia, endoleak, device migration, and a need for reintervention. [1,2] Decision-making is therefore complex, requiring integration of patient comorbidity, anatomy, procedural complexity, and anticipated durability; contemporary guidance also emphasizes shared decision-making and structured surveillance. [2,3]

Traditional risk scores and regression-based prediction models use a limited set of prespecified predictors and often provide modest discrimination or incomplete validation. [4] In contrast, ML approaches (e.g., gradient boosting, random forests, support vector machines, neural networks) can model non-linear relationships and higher-order interactions, and may improve individualized risk estimation. [5,6]

Although several ML/AI models have been proposed to predict mortality, adverse events, reintervention, and aortic remodeling after TEVAR, a structured comparison of their performance, validation strategies, interpretability, and risk of bias is needed. This systematic review summarizes available ML/AI prediction models for outcomes after TEVAR or complex thoracic endovascular aortic repair, appraises methodological quality using PROBAST, and highlights gaps for future research.

METHODS

This systematic review was conducted following the Preferred Reporting Items for Systematic Reviews and Meta-Analyses (PRISMA) 2020 guidelines. This review is registered with PROSPERO (CRD420251267013).

Search Strategy

The search was conducted in PubMed, Scopus, and the Cochrane Library from inception through December 10, 2025. The search combined two concept blocks: (1) thoracic or complex endovascular aortic repair/pathology terms ("thoracic endovascular aortic repair" OR TEVAR OR "complex EVAR" OR FEVAR OR BEVAR OR fenestrated OR branched OR "thoracoabdominal aortic aneurysm" OR TAAA OR "type B aortic dissection") and (2) ML/AI prediction terms ("machine learning" OR "artificial intelligence" OR "deep learning" OR radiomics OR XGBoost OR "random forest" OR "support vector machine" OR "neural network" OR "prediction model" OR "risk prediction"). PubMed used MeSH and title/abstract fields where available; Scopus used TITLE-ABS-KEY syntax; and Cochrane used equivalent keyword combinations adapted to its search interface. The full strategy was: PubMed: ("thoracic endovascular aortic repair"[tiab] OR TEVAR[tiab] OR "complex EVAR"[tiab] OR FEVAR[tiab] OR BEVAR[tiab] OR "thoracoabdominal aortic aneurysm"[tiab] OR TAAA[tiab] OR "type B aortic dissection"[tiab]) AND ("machine learning"[tiab] OR "artificial intelligence"[tiab] OR "deep learning"[tiab] OR radiomics[tiab] OR XGBoost[tiab] OR

"random forest"[tiab] OR "support vector machine"[tiab] OR "neural network"[tiab] OR "prediction model"[tiab]); Scopus: TITLE-ABS-KEY((TEVAR OR "thoracic endovascular aortic repair" OR "complex EVAR" OR FEVAR OR BEVAR OR TAAA OR "type B aortic dissection") AND ("machine learning" OR "artificial intelligence" OR "deep learning" OR radiomics OR XGBoost OR "random forest" OR "support vector machine" OR "neural network" OR "prediction model")); Cochrane: (TEVAR OR "thoracic endovascular aortic repair" OR "complex EVAR" OR FEVAR OR BEVAR OR TAAA OR "type B aortic dissection") AND ("machine learning" OR "artificial intelligence" OR "deep learning" OR radiomics OR XGBoost OR "prediction model").

Inclusion And Exclusion Criteria

Studies were eligible for inclusion if they enrolled adult patients (≥ 18 years) who underwent TEVAR or complex EVAR procedures, including fenestrated, branched, thoracoabdominal, arch, or other complex thoracic endovascular repairs.

Studies that reported at least one quantitative measure of predictive performance, such as the area under the receiver operating characteristic curve (AUC), accuracy, sensitivity, specificity, or calibration metrics were included. Both comparative studies (ML models versus traditional statistical models) and non-comparative ML studies were eligible, provided sufficient performance data were available.

Studies limited to open surgical repair, infrarenal EVAR without a thoracic or complex component, pediatric populations, and studies focused solely on non-predictive applications of machine learning (such as image segmentation or procedural planning without outcome prediction) were excluded. Reviews, editorials, conference abstracts without full texts, case reports, and non-English publications were also excluded. The full inclusion and exclusion criteria are summarized in the protocol.

Outcomes

The primary outcomes were the predictive performance of machine learning models for clinically relevant postprocedural endpoints. These included early (≤ 30 days) and late mortality, major complications such as stroke, spinal cord ischemia or paralysis, renal failure requiring dialysis, need for reintervention, and measures of aortic remodeling.

Selection Process and Data Extraction

Two independent authors performed the first phase of screening by title and abstract according to study inclusion criteria blindly. Duplicates were removed, and full texts of potentially relevant articles were then retrieved and assessed for eligibility. Discrepancies were resolved through discussion by a third independent author.

Data were extracted using a standardized form and included study characteristics, participant details, intervention type, ML/AI model type, predictor categories, validation strategy, reported performance metrics, and calibration or clinical utility

measures when available. Reviewers independently conducted the data extraction.

Risk of Bias Assessment

Methodological quality and risk of bias were evaluated using the Prediction Model Risk of Bias Assessment Tool (PROBAST). Each study was assessed across the domains of participants, predictors, outcomes, and analysis. Risk of bias judgments were performed independently by several reviewers, and any disagreements were resolved through group discussion to achieve consensus.

Data Synthesis

Because outcomes, modeling approaches, and reporting were heterogeneous, we synthesized findings narratively and summarized model performance in tables; quantitative pooling was not performed.

RESULTS

Study Selection

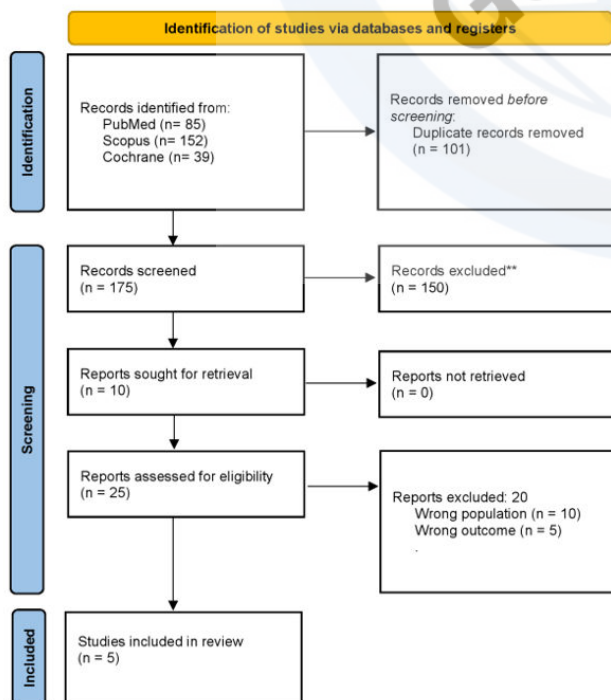
A total of 276 records were identified through database searching, including PubMed (n=85), Scopus (n=152), and the Cochrane Library (n=39). After removal of duplicates (n=101), 175 records underwent title and abstract screening, of which 150 were excluded. Twenty-five full-text reports were sought for retrieval; three could not be retrieved, leaving 22 full-text reports assessed for eligibility. Of these, 17 were excluded because they did not include a prognostic ML/AI model, focused on infrarenal EVAR only, evaluated open repair, addressed imaging segmentation or procedural planning without outcome prediction, or were non-English/abstract-only publications. Five studies were ultimately included in the qualitative synthesis (Figure 1).

Figure 1: PRISMA-Style Flow Diagram of Study Identification, Screening, Eligibility Assessment, and Inclusion.

Study Characteristics

Five retrospective prognostic studies evaluating ML or deep learning models for outcome prediction after TEVAR or complex endovascular aortic repair were included [7-11]. Studies were conducted in the United States/Canada/Singapore (Li et al.), Japan (Kano et al.), and China (Zhou et al., Dong et al., Lu et al.). Sample sizes ranged from 79 to 10,738 patients, and follow-up ranged from approximately 15 months to a median of 58 months depending on the endpoint and design. Key study characteristics, ML/AI model types, validation approaches, and performance metrics are summarized in Table 1.

Populations included degenerative thoracic aortic aneurysm and Stanford type B aortic dissection (acute/subacute or acute uncomplicated). Outcomes predicted included thoracic aorta-related late events (TALE), early mortality, adverse events, distal aortic remodeling, and reintervention.



Study	Population / setting	Outcome predicted	Model and inputs	Validation	Key performance
Li et al. 2025	10,738 elective TEVAR or complex EVAR patients from VQI; noninfrarenal aneurysm repairs	1-year TALE (death, permanent dialysis/paralysis, or stroke)	XGBoost and other ML models using preoperative, intraoperative, and postoperative registry features; logistic regression comparator	70/30 train-test split with 10-fold cross-validation in training data; no external validation	Preoperative XGBoost AUC 0.96 (95% CI, 0.95-0.97) vs logistic regression AUC 0.70; intra/postoperative AUC 0.97/0.98; Brier scores 0.09, 0.08, and 0.05
Kano et al. 2025	79 elective TEVAR patients with degenerative thoracic aortic aneurysm at a single Japanese center	Early mortality within 2 years	Decision tree analysis using 36 demographic, nutritional, comorbidity, inflammatory, immune, and procedural variables	Stratified fivefold cross-validation; no independent external validation	Accuracy 65.8%, sensitivity 81.0%, specificity 60.3%; terminal-node risk ranged from 0% to 77.7%
Zhou et al. 2024	147 acute or subacute Stanford type B aortic dissection patients undergoing proximal TEVAR at a single center	Negative distal aortic remodeling and reintervention	CTA-based CNN and point-cloud neural network (PC-NN), with and without clinical features	Internal held-out test set	PC-NN AUC 0.876 for negative remodeling; PC-NN+ clinical features AUC 0.884; PC-NN AUC 0.805 for reintervention; PC-NN + clinical features AUC 0.836
Dong et al. 2022	192 Stanford type B aortic dissection patients after TEVAR; 68 had indications for reintervention	Reintervention after TEVAR	LASSO-selected clinical, CTA, and DSA variables evaluated across eight classifiers; nomogram developed	Repeated 2:1 train-validation random splits with bootstrapping; no external validation	Logistic regression performed best, AUC 0.802; nomogram showed good calibration; risk cutoff 0.413
Lu et al. 2024	369 acute uncomplicated Stanford type B aortic dissection patients treated with initial TEVAR across two campuses	Postoperative adverse events	3D CNN-derived CTA radiomic score combined with albumin and C-reactive protein using XGBoost	Training/internal validation plus independent validation cohort from another hospital	Combined model AUC 0.990 in internal validation and 0.985 (95% CI, 0.965-1.000) in independent validation; accuracy 0.92, precision 0.92, sensitivity 0.94, specificity 0.91

Table 1: Key Characteristics and Performance Metrics of Included ML/AI Models

Model Types and Predictive Performance

Across studies, algorithms included gradient-boosted decision trees (XGBoost), decision trees (CART), logistic regression and other classical classifiers, and deep learning approaches (convolutional neural networks and point-cloud neural networks). Models incorporated clinical and perioperative variables, imaging-derived measurements, or radiomics/deep learning features from CTA.

In the largest multicenter study, a preoperative XGBoost model predicted 1-year TALE with AUC 0.96 and outperformed

logistic regression (AUC 0.70).[7] Kano et al. developed an interpretable CART model for early mortality within 2 years (sensitivity 0.81; specificity 0.60; accuracy 0.66) using fivefold cross-validation in a small cohort.[8]

Imaging-driven models targeted anatomy-dependent endpoints. Zhou et al. reported a point-cloud neural network plus clinical features for negative distal remodeling (AUC 0.884) and reintervention (AUC 0.833) on a held-out test set.[9] Dong et al. compared eight classifiers for reintervention prediction and found logistic regression performed best (AUC 0.802).[10] Lu et al. combined deep learning-based radiomics with clinical variables via XGBoost and achieved AUC 0.985 on an independent external validation cohort.[11]

Calibration, Interpretability, and Clinical Utility Reporting

Calibration and decision-analytic reporting were inconsistent. Li et al. reported Brier scores and described good calibration, while Lu et al. reported calibration and decision curve analysis [7,11] Interpretability varied: decision tree and nomogram-style approaches were transparent, whereas deep learning models were less directly interpretable.

Risk of Bias

Using PROBAST, the most frequent concerns related to the analysis domain (small samples, incomplete reporting of handling missing data, and predominant reliance on internal validation). Selection bias was a concern in studies with high exclusion due to missing follow-up imaging or restrictive eligibility criteria. Overall, risk of bias ranged from moderate to high across included studies (Table 2).

Study	Main risk-of-bias concerns	Overall judgment
Li et al. 2025	Large multicenter registry and clear test split, but no independent external validation and possible registry-related dataset shift.	Moderate
Kano et al. 2025	Small sample size, limited events, single-center design, and internal validation only.	High
Zhou et al. 2024	Single-center imaging cohort, potential selection due to follow-up CTA requirements, and internal validation only.	High
Dong et al. 2022	Selection related to availability of postoperative CTA/DSA and follow-up, single-center design, and internal validation only.	High
Lu et al. 2024	Independent validation was performed, but retrospective selection and very high model performance warrant prospective validation.	Moderate

Table 2: Summary of PROBAST Risk-of-Bias Concerns across Included Studies

DISCUSSION

Principal Findings

This systematic review identified five retrospective prognostic studies evaluating machine learning (ML) and deep learning (DL) models to predict outcomes after thoracic endovascular aortic repair (TEVAR) and complex endovascular aortic interventions [7-11]. Across heterogeneous cohorts (degenerative thoracic aneurysm and Stanford type B aortic dissection), sample sizes ranged from 79 to 10,738 patients, and endpoints included early mortality, thoracic aorta-related late events (TALE), adverse events, reintervention, and distal aortic remodeling. Reported discrimination was generally high (AUC approximately 0.80-0.99), and in studies with direct comparators, ML models outperformed conventional regression [7,10]. However, only one study incorporated independent external validation and several studies excluded large proportions of patients because of missing follow-up imaging, which may introduce selection bias and limit generalizability [11].

Why Individualized Risk Prediction Matters after TEVAR

TEVAR is a cornerstone therapy for descending thoracic aneurysm disease and type B aortic dissection, yet post-procedural trajectories remain heterogeneous. Some patients achieve durable clinical stability and favourable aortic remodeling, whereas others experience late adverse events,

persistent false lumen perfusion, or require reintervention. Contemporary aortic disease guidance emphasises longitudinal clinical and imaging surveillance after endovascular repair, [12] but surveillance schedules are often applied uniformly and may not reflect individualized risk. Better prognostic models could support shared decision-making (e.g., balancing procedural complexity against anticipated benefit), tailor surveillance intensity, and identify higher-risk patients who might benefit from earlier imaging, closer follow-up, or proactive adjunctive interventions.

Interpretation In the Context of Existing Prediction Approaches

Conventional TEVAR prediction tools have typically relied on regression-based models using a limited set of prespecified predictors. In contrast, the included studies suggest that ML—particularly gradient-boosted decision trees—may better capture non-linear associations and higher-order interactions among perioperative variables.[13] In the largest included study, Li et al. developed XGBoost models to predict 1-year TALE after elective TEVAR and complex endovascular aortic aneurysm repair; the preoperative model achieved AUC 0.96 compared with AUC 0.70 for a logistic regression comparator, and discrimination improved further when intraoperative or postoperative predictors were incorporated.[7] While these findings are promising, large apparent gains should be interpreted cautiously because optimistic performance can arise from outcome definition differences, incomplete reporting of preprocessing and tuning, or unrecognized dataset shift across sites and time periods.[14]

Multimodal and Imaging-Driven Models

Anatomy-dependent endpoints such as distal aortic remodeling and reintervention are closely linked to post-TEVAR morphology, including residual false lumen perfusion, luminal geometry, and stent-aorta interactions. DL and radiomics approaches may therefore be particularly suited for these outcomes by learning informative representations from computed tomography angiography (CTA). Zhou et al. used a point-cloud neural network and reported AUC 0.884 (95% CI 0.788–0.980) for predicting adverse distal remodeling and reintervention on a held-out test set.[9] Dong et al. compared multiple classifiers for reintervention and found that logistic regression performed best (AUC 0.802, 95% CI 0.689–0.915), highlighting that simpler models may remain competitive, particularly in smaller datasets.[10] Lu et al. integrated deep learning-derived radiomics with clinical variables (albumin and C-reactive protein) using XGBoost and achieved AUC 0.985 (95% CI 0.965–1.000) on independent external validation.[11] Given the high-dimensional nature of radiomics,[15] adherence to feature standardization initiatives and imaging reporting checklists can help reduce avoidable variation and facilitate reproducibility.[16,17]

From Accuracy to Usefulness: Calibration and Decision Impact

High discrimination alone does not guarantee reliable individualized risk estimates or clinical benefit. Calibration (agreement between predicted and observed risks) should be routinely reported alongside discrimination, and clinical utility should be evaluated using decision-analytic methods across plausible risk thresholds.[18] In our review, calibration reporting was inconsistent: Li et al. reported Brier scores and described good calibration,[7] and Lu et al. reported calibration analyses and decision curve analysis.[11] More consistent reporting of calibration-in-the-large, calibration slope, and graphical calibration would improve interpretability and support transportability assessments. In addition, model explanations (e.g., feature importance or SHAP-based explanations for tree models) may help clinicians understand whether identified predictors align with clinical knowledge, but interpretability does not substitute for external validation.[19]

Methodological Quality and Risk of Bias

Using PROBAST, the most frequent concerns related to the analysis domain: small samples with limited numbers of events (e.g., Kano et al., n=79), reliance on complete-case analyses without multiple imputation, and predominant use of internal validation only.[8,20] These limitations increase the likelihood of overfitting and optimistic performance estimates. Split-sample validation is also inefficient in modest cohorts, whereas resampling-based internal validation (bootstrap or cross-validation) provides more stable estimates. Adequate sample size planning for prediction model development is important to reduce overfitting and improve the precision of performance estimates.[21] Finally, exclusion of patients without follow-up imaging (notably in studies predicting remodeling or reintervention) may yield a selected cohort with systematically different risks and should be addressed explicitly in future studies.

Implications For Future Research and Implementation

To progress from proof-of-concept toward clinical deployment, TEVAR prediction models should be developed and reported according to established standards (TRIPOD and the updated TRIPOD+AI guidance) and appraised using tools such as PROBAST [22-23]. External validation across centres, scanners, and practice patterns is critical to address dataset shift; model performance may deteriorate when clinical practice, imaging protocols, or patient case-mix change over time [14]. Future studies should prespecify outcomes, time horizons, and clinically meaningful thresholds; evaluate calibration and net benefit; and directly compare ML to parsimonious, well-calibrated regression models. When multimodal models incorporate radiomics, standardization and transparent reporting are essential to enable replication. [16,17] Ultimately, prospective impact studies—and, where appropriate, trials reported using AI extensions such as CONSORT-AI and SPIRIT-AI—will be required to demonstrate that model-informed management improves patient-relevant outcomes rather than merely improving discrimination. [24,25]

Clinical implementation also raises workflow and regulatory issues that extend beyond model discrimination. AI-enabled risk calculators intended to guide clinical decisions may be regulated as medical device software in some jurisdictions, and therefore require a clearly defined intended use, lifecycle governance, documentation, human oversight, cybersecurity and data privacy safeguards, bias monitoring, and post-deployment performance monitoring or recalibration [27,28]. In vascular practice, models should be integrated into electronic health records, picture archiving and communication systems, or vascular quality registries in a way that minimizes manual data entry and presents calibrated absolute risks with uncertainty and actionable thresholds linked to surveillance or referral pathways. Without such workflow integration and governance, even accurate models may fail to change clinician behavior or improve patient outcomes.

Strengths and Limitations of this Review

Strengths of this review include prospective registration, adherence to PRISMA 2020 guidance, and structured assessment of prediction-model risk of bias using PROBAST [20,26]. Limitations include the small number of eligible studies, heterogeneity in populations and outcome definitions (which precluded quantitative pooling), and incomplete reporting of calibration and clinical utility in many primary studies. Publication bias is possible because prediction studies with weaker model performance may be less likely to be published, and the rapid evolution of ML methods may mean that additional studies emerge after our search date.

CONCLUSION

This systematic review shows that ML and DL models have promising predictive accuracy for outcomes following TEVAR and complex endovascular aortic repair. For composite adverse events, early mortality, reintervention, and unfavorable aortic remodeling, these models may improve individualized risk stratification by combining clinical variables with imaging-derived features such as false-lumen dimensions, residual

perfusion, and radiomic signatures. However, the evidence remains based mainly on retrospective studies, many of which had small sample sizes, incomplete reporting, and limited or no external validation. These limitations introduce possible selection bias and restrict the generalizability of the current findings beyond the populations and settings in which the models were developed.

Future work should prioritize transparent reporting, robust external validation, comprehensive calibration and decision-analytic evaluation, and prospective impact studies to determine whether these models improve surveillance strategies and clinical outcomes.

DECLARATIONS

Consent to Participate

Not applicable. This study is a systematic review and meta-analysis of previously published studies that used de-identified patient data. No individual participants were directly recruited or contacted by the authors. The Institutional Review Boards of the affiliated institutions deemed this study exempt from ethics review on this basis.

Funding Declaration

This research received no grant from any funding agency in the public, commercial, or not-specific for-profit sectors. No financial support was received for the design, conduct, or reporting of this systematic review.

Human Ethics and Consent to Participate

This study was conducted in accordance with the ethical standards of the 1964 Declaration of Helsinki and its later amendments. As a systematic review of previously published and de-identified data, the study was deemed exempt from full ethics committee review by the Institutional Review Boards of the affiliated institutions (Prince Hamza Hospital, Amman, Jordan; Tbilisi State Medical University, Georgia; University of Health Sciences, Lahore, Pakistan; Yarmouk University, Amman, Jordan; Wah Medical College, Rawalpindi, Pakistan; and the University of Jordan, Amman, Jordan). No primary data collection involving human participants was undertaken.

Competing Interests

The authors declare that they have no known competing financial interests or personal relationships that could have appeared to influence the work reported in this paper. No author has received honoraria, consultancy fees, speaker fees, travel support, or research funding from any commercial entity with a financial interest in the subject matter of this review.

Consent to Publish

Not applicable with respect to individual patient data, as this manuscript is a systematic review synthesising data from previously published studies and no identifiable participant information is presented. All co-authors have reviewed the final version of the manuscript and consent to its submission and publication.

REFERENCES

- Nation DA, Wang GJ. TEVAR: endovascular repair of the thoracic aorta. In *Seminars in interventional radiology 2015* (Vol. 32, No. 03, pp. 265-271). Thieme Medical Publishers.
- Beckman JA, Members WC. 2022 ACC/AHA guideline for the diagnosis and management of aortic disease: a report of the American Heart Association/American College of Cardiology Joint Committee on clinical practice guidelines. *Circulation*. 2022;146(24):e334.
- Chaikof EL, Dalman RL, Eskandari MK, Jackson BM, Lee WA, Mansour MA, et al. The Society for Vascular Surgery practice guidelines on the care of patients with an abdominal aortic aneurysm. *Journal of vascular surgery*. 2018;67(1):2-77.
- Patterson BO, Holt PJ, Hinchliffe R, Loftus IM, Thompson MM. Predicting risk in elective abdominal aortic aneurysm repair: a systematic review of current evidence. *European Journal of Vascular and Endovascular Surgery*. 2008;36(6):637-45.
- Handelman GS, Kok HK, Chandra RV, Razavi AH, Lee MJ, Asadi H. eD octor: machine learning and the future of medicine. *Journal of internal medicine*. 2018;284(6):603-19.
- Fouladvand S, Noshad M, Goldstein MK, Periyakoil VJ, Chen JH. Mild cognitive impairment: data-driven prediction, risk factors, and workup. *AMIA Summits on Translational Science Proceedings*. 2023;2023:167.
- Li B, Eisenberg N, Beaton D, Lee DS, Aljabri B, Al-Omran L, et al. Using machine learning to predict outcomes following thoracic and complex endovascular aortic aneurysm repair. *Journal of the American Heart Association*. 2025;14(5):e039221.
- Kano M, Nishibe T, Iwasa T, Matsuda S, Akiyama S, Iwahashi T, et al. Predicting Early Mortality after Thoracic Endovascular Aneurysm Repair: A Machine Learning-Based Decision Tree Analysis. *Annals of Vascular Diseases*. 2025;18(1):oa-25.
- Zhou M, Luo X, Wang X, Xie T, Wang Y, Shi Z, et al. Deep learning prediction for distal aortic remodeling after thoracic endovascular aortic repair in Stanford type B aortic dissection. *Journal of Endovascular Therapy*. 2024;31(5):910-8.
- Dong Y, Que L, Jia Q, Xi Y, Zhuang J, Li J, et al. Predicting reintervention after thoracic endovascular aortic repair of Stanford type B aortic dissection using machine learning. *European radiology*. 2022;32(1):355-67.
- Lu X, Gong W, Yang W, Peng Z, Zheng C, Zha Y. Deep learning-based radiomics of computed tomography angiography to predict adverse events after initial endovascular repair for acute uncomplicated Stanford type B aortic dissection. *European Journal of Radiology*. 2024;175:111468.
- Isselbacher EM, Preventza O, Hamilton Black III J, Augoustides JG, Beck AW, Bolen MA, et al. 2022 ACC/AHA guideline for the diagnosis and management of aortic disease: a report of the American Heart Association/American College of Cardiology Joint Committee on Clinical Practice Guidelines. *Journal of the American College of Cardiology*. 2022;80(24):e223-393.
- Chen T, Guestrin C. Xgboost: A scalable tree boosting system. In *Proceedings of the 22nd acm sigkdd international conference on knowledge discovery and data mining 2016*(pp. 785-794).
- Finlayson SG, Subbaswamy A, Singh K, Bowers J, Kupke A, Zittrain J, et al. The clinician and dataset shift in artificial intelligence. *New England Journal of Medicine*. 2021;385(3):283-6.
- Lambin P, Rios-Velazquez E, Leijenaar R, Carvalho S, Van Stiphout RG, Granton P, et al. Radiomics: extracting more information from medical images using advanced feature analysis. *European journal of cancer*. 2012;48(4):441-6.
- Zwanenburg A, Leger S, Vallières M, Löck S. The Image Biomarker Standardisation Initiative: standardized quantitative radiomics for

- high-throughput image-based phenotyping. *Radiology*. 2020;295(2):328-338.
17. Mongan J, Moy L, Kahn Jr CE. Checklist for artificial intelligence in medical imaging (CLAIM): a guide for authors and reviewers. *Radiology: Artificial Intelligence*. 2020 Mar 25;2(2):e200029.
 18. Vickers AJ, Elkin EB. Decision curve analysis: a novel method for evaluating prediction models. *Medical Decision Making*. 2006;26(6):565-74.
 19. Lundberg SM, Lee SI. A unified approach to interpreting model predictions. *Advances in neural information processing systems*. 2017;30.
 20. Wolff RF, Moons KG, Riley RD, Whiting PF, Westwood M, Collins GS, Reitsma JB, Kleijnen J, Mallett S. RESEARCH AND REPORTING METHODS PROBAST: A Tool to Assess the Risk of Bias and Applicability of Prediction Model Studies.
 21. Riley RD, Ensor J, Snell KI, Harrell FE, Martin GP, Reitsma JB, et al. Calculating the sample size required for developing a clinical prediction model. *Bmj*. 2020;368.
 22. Collins GS, Reitsma JB, Altman DG, Moons KG. Transparent reporting of a multivariable prediction model for individual prognosis or diagnosis (TRIPOD): the TRIPOD statement. *Journal of British Surgery*. 2015;102(3):148-58.
 23. Collins GS, Moons KG, Dhiman P, Riley RD, Beam AL, Van Calster B, et al. TRIPOD+ AI statement: updated guidance for reporting clinical prediction models that use regression or machine learning methods. *bmj*. 2024;385.
 24. Liu X, Rivera SC, Moher D, Calvert MJ, Denniston AK, Ashrafian H, et al. Reporting guidelines for clinical trial reports for interventions involving artificial intelligence: the CONSORT-AI extension. *The Lancet Digital Health*. 2020;2(10):e537-48.
 25. Rivera SC, Liu X, Chan AW, Denniston AK, Calvert MJ, Ashrafian H, et al. Guidelines for clinical trial protocols for interventions involving artificial intelligence: the SPIRIT-AI extension. *The Lancet Digital Health*. 2020;2(10):e549-60.
 26. Page MJ, McKenzie JE, Bossuyt PM, Boutron I, Hoffmann TC, Mulrow CD, et al. The PRISMA 2020 statement: an updated guideline for reporting systematic reviews. *bmj*. 2021;372.

A large, light blue watermark is centered on the page. It features a circular design with a network of interconnected nodes and lines, resembling a molecular or data structure. Overlaid on this design is the text "Galley Proof" in a bold, sans-serif font, oriented diagonally from the bottom-left towards the top-right.