

## EDITORIAL

# The utility of machine learning in the management of patients with peripheral arterial disease

Ravindhran B,<sup>1</sup> Thakker D<sup>2</sup>

1. Academic Vascular Surgical Unit, Hull York Medical School, Hull, UK
2. Centre for Responsible AI, School of Digital and Physical Sciences, University of Hull, UK

**Corresponding author:**

Bharadwaj Ravindhran  
Academic Vascular Surgical Unit,  
2nd Floor Allam Diabetes Centre,  
Hull Royal Infirmary, Anlaby Road,  
Hull HU32JZ, UK  
Email: [bharadwaj.ravindhran@nhs.net](mailto:bharadwaj.ravindhran@nhs.net)

**Received:** 4th November 2025

**Accepted:** 12th December 2025

**Online:** 19th January 2026

**Introduction**

Machine learning (ML) is moving decisively from concept to clinic-adjacent evaluation in vascular medicine. Peripheral arterial disease (PAD) has become a major focus for machine learning because it is clinically heterogeneous and carries a high health and economic burden, making it well suited to data driven approaches for earlier detection and personalised risk assessment. The core promise remains unchanged: algorithms that learn patterns across multimodal data, notes, vital signs, blood results, ultrasound or duplex waveforms, CT and MR angiography, intravascular imaging, prescriptions and longitudinal outcomes can help detect disease earlier, stratify risk more reliably and match treatment to patient and lesion characteristics. The crucial question for vascular surgeons is where that promise stands today. Across diagnostics, risk stratification and outcome prediction, AI/ML is best regarded as a maturing adjunct whose best-performing tools are either in retrospective or quasi-prospective evaluation, with a small but growing body of external validation. The field is now constrained less by raw model accuracy and more by generalisability, clinical integration, explainability and governance.

**Diagnostics**

Diagnostic applications have produced some of the most tangible early gains. Electronic health record phenotyping and natural language processing have the potential to flag probable PAD from codes, medication patterns, laboratories and free-text notes with performance that is sufficient to support targeted ankle-brachial pressure index (ABPI) or vascular laboratory testing in primary care and diabetes

clinics.<sup>1-3</sup> Real-world implementation work has started to define how such models can be integrated into care pathways; qualitative and mixed-methods studies emphasise workflow alignment, clinician trust and equity monitoring as prerequisites for sustained use.<sup>4,5</sup> In the vascular laboratory, algorithms to classify continuous wave Doppler and plethysmography waveforms are increasingly accurate in laboratory settings and can reduce inter-operator variability, although prospective evidence of the impact on diagnostic accuracy or throughput remains limited.<sup>6-9</sup> In imaging, deep learning for CT angiographic segmentation and stenosis quantification continues to advance quickly, with several groups reporting automated arterial tree segmentation, calcium scoring, runoff quantification and lesion severity classification, particularly in the femoro-popliteal and infrapopliteal beds.<sup>10-14</sup> These systems show high agreement with expert readers on retrospective datasets, but cross-scanner and cross-centre robustness and prospective clinical utility are still in the process of being established.

Intravascular imaging is benefiting from ML originally developed for coronaries: plaque component detection and calcium quantification on intravascular ultrasound and optical coherence tomography, and early peripheral applications suggest feasibility for reproducible measurements that could guide vessel preparation and device selection. However, PAD-specific clinical validation and device-agnostic performance remain works in progress.<sup>15,16</sup> Outside of imaging suites, photoplethysmography captured by smartphones and wearable devices has shown proof-of-concept for PAD screening, but large-scale prospective studies in target populations, with confirmatory testing and cost-effectiveness

**Key words:** machine learning, peripheral arterial disease, artificial intelligence, algorithms

analyses, are still awaited before routine clinical adoption can be recommended.<sup>8,17,18</sup> Taken together, diagnostic AI for PAD is moving from promising retrospective accuracy to early-stage clinical evaluation. In the near term, adoption is best directed towards tasks with outputs that are directly verifiable at the point of care such as waveform classification, stenosis measurements and structured report extraction, while simultaneously building the implementation evidence base.

### **Risk stratification**

Risk stratification for both major adverse cardiovascular events and major adverse limb events is central to PAD care, and this is an area where ML has achieved clinically meaningful performance on large registries and multi-institutional datasets. Recent studies using the Vascular Quality Initiative and other consortia demonstrate that ML can estimate 30-day and long-term risks of major adverse limb events, amputation, reintervention, poor wound healing and major adverse cardiovascular events with very good discrimination and improved calibration relative to traditional scores.<sup>19</sup> Multicentre registry analyses have trained and tested models to predict short and long-term adverse outcomes after endovascular intervention, showing reasonable performance and calibration. However, they also highlight the need for external validation and clinically actionable risk outputs. Similar models for bypass outcomes show promise, but generalisability remains challenging due to device and conduit-specific heterogeneity, centre effects, and missingness in key variables (e.g., runoff).<sup>20–25</sup> At this stage, selective deployment of validated models to inform shared decision-making and surveillance planning is appropriate provided calibration is satisfactory, probabilities are communicated transparently and local performance has been verified. Subgroup performance should be reported and shown to be acceptable in view of well-established disparities in PAD presentation and outcomes.

### **Early and delayed outcome prediction**

Beyond risk stratification, AI/ML is increasingly used for granular outcome prediction tied to procedural planning and follow-up. Peri-procedural complications such as contrast-associated kidney injury, bleeding and access site problems have been modelled with encouraging retrospective performance.<sup>26</sup> Looking further out, lesion-specific models that incorporate CTA features, duplex metrics and lesion morphology are being developed to predict primary patency and target lesion revascularisation, with several reports demonstrating strong internal performance but mixed external validation.<sup>27,28</sup> Prediction of wound healing, a particularly relevant domain for chronic limb threatening ischaemia and multidisciplinary limb salvage teams, has been studied using a mix of clinical features, perfusion measurements and wound images. While multiple groups report models with good discrimination, the literature remains fragmented by small sample sizes, heterogeneous definitions of healing and limited external testing. In the absence of interpretable models and standardised endpoints,

routine use is not yet justified.<sup>29,30</sup> There is an emerging consensus that prediction models deliver greatest value when coupled to modifiable actions – for example, earlier duplex surveillance in patients at heightened restenosis risk and targeted optimisation of perfusion and infection control when healing probability is low, with prospective impact evaluations now a priority.

### **Current work around the world**

Globally, research is converging on three practical directions. First, registry-linked modelling on major analogous datasets is yielding models that can be externally validated and benchmarked across systems, with increasing attention to calibration, decision-curve analysis and net benefit.<sup>31</sup> Second, imaging AI is being standardised, with communities working on shared tasks for CTA segmentation and lesion scoring to improve reproducibility across vendors and scanners, and on explainable overlays that let clinicians see which image regions drive a classification. Third, privacy-preserving training such as federated learning is being piloted to overcome data sharing barriers while improving model generalisability across diverse populations and devices.<sup>32</sup> Notably, in comparative evaluations, ML systems have outperformed clinicians, demonstrating higher discrimination and lower prediction error than expert assessment or conventional risk scores, moving the field a step closer to reliable adjunctive decision support at the point of care.<sup>33</sup> Operationally, clinical deployment is best underpinned by multicentre validation, adoption of standardised data models and terminology to streamline collaboration and, when feasible, federated methods to promote diversity and equity.

### **Explainable AI**

In clinical decision support, explainability means making a model's predictions transparent and clinically interpretable, both at the individual level (why this output was produced for this patient) and at the broader level (how the system behaves overall and which factors most influence its outputs). Practically, this allows us to see which variables or image regions most influenced a risk estimate, judge whether the rationale aligns with the patient's presentation, and decide when to accept, qualify or override the recommendation.<sup>34</sup> Explainability is now recognised as a safety requirement rather than a research feature. It underpins safe deployment by enabling clinicians to verify outputs, identify potential model errors and maintain accountability in decision-making. For structured (tabular) clinical prediction models, local explanations such as SHAP value summaries, Local Interpretable Model-agnostic Explanations (LIME) plots<sup>35</sup> and global feature importance analyses have become standard practice in recent high-quality studies and clinical pilots, enabling clinicians to understand why the model assigns a high risk label to a particular patient.<sup>36</sup> For imaging, saliency maps and attention overlays have matured to the point that they can highlight stenotic segments or plaque components that informed the output. Tools intended for bedside use should display their confidence and calibration characteristics, provide example-

level explanations and defer to interpretable models when they perform equivalently.<sup>31</sup> It is also recommended that ongoing performance monitoring be in place to detect drift and inequities. A 'near-miss' reporting culture around AI recommendations, akin to pharmacovigilance, is an emerging recommendation more broadly, so failures are learned from and shared.<sup>37</sup>

### AI governance

Regulatory and institutional governance has advanced materially in the past two years. This reflects a shift from voluntary principles to enforceable mechanisms for AI assurance and post-market oversight. The US FDA finalised guidance on Predetermined Change Control Plans for AI-enabled device software functions in late 2024, clarifying how iterative model updates can be managed within an approved framework, and subsequently integrated that guidance into its AI/ML SaMD (Software as a medical device) resources in 2025.<sup>38</sup> These updates clarify regulator expectations for adaptive algorithms and continuous learning systems, which have historically faced uncertainty under fixed-approval pathways. Internationally and in the UK, Good Machine Learning Practice principles jointly articulated by regulators provide direction on data management, training/validation and human factors. Together, these principles emphasise transparency, reproducibility and clinician accountability across the AI lifecycle. At the institutional level, it is recommended that model deployment be overseen by multidisciplinary committees including vascular surgery, data science, ethics, legal, IT security and patient representatives. Such committees serve as local assurance bodies, ensuring that governance responsibilities are shared rather than delegated solely to data science teams. Documentation such as model cards should specify intended use, data provenance, performance across subgroups, known failure modes and update cadence. Explicit traceability between data, model versions and deployment environments should be maintained to support audit and incident review. For continuously learning systems, versioning, change control and re-approval triggers are essential so that the tool in clinic matches the tool that was validated. It is also recommended that local monitoring assess equity, alert burden and clinical workflow effects, with the authority to pause or retire models when harms outweigh benefits.<sup>39,40</sup> Embedding these practices establishes a feedback loop between technical governance and clinical safety, aligning local oversight with emerging regulatory expectations.

### Adjunct role in practice

Clinician expertise remains central, and the most appropriate framing today is that AI/ML is an adjunct. This framing reinforces clinical accountability and aligns with current regulatory guidance that mandates human oversight for all high-risk AI systems. In diagnostics, it is recommended that algorithms pre-screen and double-check while examination, bedside Doppler and ABPI/TBPI remain foundational. Early evidence suggests that such hybrid

workflows preserve diagnostic accuracy while reducing clinician workload and variability. Beyond specialist settings, dynamic ML models embedded in community and primary care records have shown promise in identifying high risk PAD patients in real time, enabling prioritisation for assessment and treatment.<sup>41</sup> In shared decision-making, individualised risk estimates expressed as natural frequencies and simple visuals can help patients understand trade-offs between limb salvage strategies, surveillance intensity and the burden of therapy. Presenting outputs in interpretable formats also helps maintain informed consent and patient trust in AI-supported decisions. In stratification, consistent risk labels across teams, vascular surgery, podiatry, wound care, cardiology, can coordinate care so that high-risk patients receive timely interventions. Interoperability and shared data standards are key to ensuring that such risk stratification remains consistent across systems and institutions. Across the continuum, AI/ML can reinforce risk factor management by identifying patients likely to benefit from statin adherence, supervised exercise therapy, smoking cessation support, glycaemic optimisation and foot care education; it is recommended that these digital nudges be embedded within clinician-led pathways to ensure accountability and equity. In this configuration, AI functions as a precision-enabling layer within clinician-led pathways, amplifying preventive care rather than displacing professional judgement.

### Pitfalls: black box, equity, and overfitting

The most common pitfalls are now well documented. Black box models may latch onto spurious correlates such as scanner signatures, site effects or documentation quirks, which collapse when deployed elsewhere.<sup>42</sup> Overfitting to single-centre datasets remains common, as does inadvertent data leakage that inflates performance.<sup>43</sup> Equity concerns are prominent because PAD disproportionately affects patients with diabetes, chronic kidney disease and socioeconomic disadvantage; subgroup performance gaps can widen existing disparities. For these reasons, it is recommended that studies prespecify cohorts, strictly separate training and testing by patient and time, and perform external validation across sites, devices and populations. Calibration should be reported alongside discrimination, and decision-curve analyses should make net benefit explicit. Bias mitigation strategies, reweighting, balanced sampling, threshold adjustment and fairness-aware evaluation are recommended and should be transparently reported.<sup>44</sup> Automation bias and deskilling are human factor risks; it is recommended that interfaces display rationale and uncertainty and that clinicians stay in the loop for interpretation and override.<sup>45</sup> Embedding these safeguards converts technical validation into operational safety, ensuring models remain trustworthy under real-world conditions.

### Reporting and appraisal guidelines

Methodological rigour is improving, supported by new and updated guidance. Collectively, these frameworks mark a shift from ad-hoc

## KEY MESSAGES

- Machine learning has the potential to support management of peripheral arterial disease by flagging patients at risk, standardising diagnostic interpretation and providing individualised estimates of limb and cardiovascular outcomes. The most mature tools are adjuncts that fit within existing workflows. Adoption should focus on transparent verifiable outputs that link to clear actions and measurable benefits now.
- Safety and accountability depend on explainability, calibration and continuous monitoring. Tools used clinically should show why a prediction was made, expose uncertainty and perform acceptably across subgroups. Governance is essential: version control, change management, incident reporting and multidisciplinary oversight help detect drift, reduce bias and protect patients and staff alike.
- The next phase should include measured monitored deployment. Priorities include multicentre validation, common data standards, pragmatic impact studies and clinician centred design. Dynamic models in community records can prioritise high risk patients for timely assessment and treatment. When implemented responsibly, these tools have the potential to standardise care, personalise therapy and improve patient experience and outcomes.

study descriptions to structured auditable evidence standards for clinical AI. TRIPOD+AI provides harmonised reporting standards for prediction model studies regardless of whether regression or ML is used.<sup>46</sup> DECIDE-AI outlines how to report early clinical evaluation of AI decision support.<sup>47</sup> The STARD-AI guideline addresses diagnostic accuracy studies using AI.<sup>48</sup> PROBAST+AI updates the risk-of-bias and applicability assessment for prediction models built with regression or ML.<sup>49</sup> For randomised evaluations, CONSORT-AI and SPIRIT-AI remain the standards.<sup>50,51</sup> Within vascular surgery, uptake of these frameworks by authors, reviewers and editors will promote faster translation by strengthening reproducibility, comparability and clinical relevance. Their consistent adoption will also streamline regulatory submissions and evidence reviews for AI-enabled vascular tools.

## Conclusion

The state of AI for PAD in 2026 can be summarised as robust feasibility with islands of readiness. Diagnostic support tools for waveform interpretation and CTA quantification are technically mature and entering early clinical evaluation, with adoption recommended where outputs are transparent and verifiable. Risk stratification for major adverse limb events, major adverse cardiac events, amputation and reintervention is strong enough to support shared decision-making and surveillance planning in many settings, provided that external validation and calibration are documented locally. Prediction of wound healing show promise but require larger,

standardised and prospective evaluations before widespread use is recommended. Explainability, governance and equity considerations have moved to the forefront, aided by clearer regulatory pathways and stronger reporting and appraisal guidelines. The immediate trajectory is towards measured monitored deployment with clear assessment of evidence of benefit across outcomes, experience and value. With appropriate safeguards, vascular surgery services are well positioned to introduce targeted AI tools as adjuncts, helping to standardise decisions, personalise treatment and remedy gaps in diagnosis and long-term follow-up for PAD.

**Conflict of Interest:** None.

**Funding:** None.

## References

1. Afzal N, Mallipeddi VP, Sohn S, *et al.* Natural language processing of clinical notes for identification of critical limb ischemia. *Int J Med Inform* 2018;**111**: 83–9. <https://doi.org/10.1016/j.ijmedinf.2017.12.024>
2. Dev S, Zolensky A, Aridi HD, *et al.* Use of deep learning to identify peripheral arterial disease cases from narrative clinical notes. *J Surg Res* 2024;**303**: 699–708. <https://doi.org/10.1016/j.jss.2024.09.062>
3. Weissler EH, Zhang J, Lippmann S, Rusincovitch S, Henao R, Jones WS. The use of natural language processing to improve identification of patients with peripheral artery disease. *Circ Cardiovasc Interv* 2020;**13**(10):e009447. <https://doi.org/10.1161/CIRCINTERVENTIONS.120.009447>
4. Singer SJ, Kellogg KC, Galper AB, Viola D. Enhancing the value to users of machine learning-based clinical decision support tools: a framework for iterative, collaborative development and implementation. *Health Care Manage Rev* 2022;**47**(2):E21–31. <https://doi.org/10.1097/HMR.0000000000000324>
5. Wang SM, Hogg HDJ, Sangvai D, *et al.* Development and integration of machine learning algorithm to identify peripheral arterial disease: multistakeholder qualitative study. *JMIR Form Res* 2023;**7**(1):e43963. <https://doi.org/10.2196/43963>
6. Merouche S, Allard L, Montagnon E, Soulez G, Bigras P, Cloutier G. A robotic ultrasound scanner for automatic vessel tracking and three-dimensional reconstruction of B-mode images. *IEEE Trans Ultrason Ferroelectr Freq Control* 2016;**63**(1):35–46. <https://doi.org/10.1109/TUFFC.2015.2499084>
7. Luo X, Ara L, Ding H, Rollins D, Motaganahalli R, Sawchuk AP. Computational methods to automate the initial interpretation of lower extremity arterial Doppler and duplex carotid ultrasound studies. *J Vasc Surg* 2021;**74**(3):988–996.e1. <https://doi.org/10.1016/j.jvs.2021.02.050>
8. Allen J, Liu H, Iqbal S, Zheng D, Stansby G. Deep learning-based photoplethysmography classification for peripheral arterial disease detection: a proof-of-concept study. *Physiol Meas* 2021;**42**(5):054002. <https://doi.org/10.1088/1361-6579/abf9f3>
9. Kim S, Hahn JO, Youn BD. Detection and severity assessment of peripheral occlusive artery disease via deep learning analysis of arterial pulse waveforms: proof-of-concept and potential challenges. *Front Bioeng Biotechnol* 2020;**8**:720. <https://doi.org/10.3389/fbioe.2020.00720>
10. Mistelbauer G, Morar A, Scherthner R, *et al.* Semi-automatic vessel detection for challenging cases of peripheral arterial disease. *Comput Biol Med* 2021;**133**: 104344. <https://doi.org/10.1016/j.combiomed.2021.104344>
11. Dai L, Zhou Q, Zhou H, *et al.* Deep learning-based classification of lower extremity arterial stenosis in computed tomography angiography. *Eur J Radiol* 2021;**136**: 109528. <https://doi.org/10.1016/j.ejrad.2021.109528>
12. Hong W, Kang J, Kim SE, *et al.* Deep learning-based diagnosis of femoropopliteal artery steno-occlusion using maximum intensity projection images of CT angiography. *Tomography* 2025;**11**(9):104. <https://doi.org/10.3390/tomography11090104>
13. Pomozi E, Quintero-Peña C, Csore J, *et al.* Innovations in MRI and AI integration for vascular plaque evaluation and overview of deep learning techniques in peripheral vascular disease. *Methodist Debakey Cardiovasc J* 2025;**21**(5):71–80. <https://doi.org/10.14797/mdcvj.1642>

14. Salvi A, Shah R, Higgins L, Menon PG. Vision transformers for AI-driven classification of peripheral artery disease from maximum intensity projections of runoff CT angiograms. *Proceedings 2022 IEEE International Conference on Bioinformatics and Biomedicine (BIBM 2022)* 2022;3870–2. Available from: 2022 IEEE International Conference on Bioinformatics and Biomedicine (BIBM 2022) (Table of Contents)
15. Pinna A, Boi A, Mannelli L, et al. Machine learning for coronary plaque characterization: a multimodal review of OCT, IVUS, and CCTA. *Diagnostics* 2025;**15**(14):1822. <https://doi.org/10.3390/diagnostics15141822>
16. Putra RPP, Sembiring YE. Artificial intelligence and advanced vascular imaging: emerging tools for precision in peripheral arterial disease management. *J Neonatal Surg* 2025;**14**(23s):824–31. Available from: <https://www.jneonatsurg.com/index.php/jns/article/view/5195>
17. Kim KB, Baek HJ. Photoplethysmography in wearable devices: a comprehensive review of technological advances, current challenges, and future directions. *Electronics* 2023;**12**:2923. <https://doi.org/10.3390/electronics12132923>
18. Ahmad RUS, Khan WU, Khan MS, Cheung P. Emerging rapid detection methods for the monitoring of cardiovascular diseases: Current trends and future perspectives. *Mater Today Bio* 2025;**32**:101663. <https://doi.org/10.1016/j.mtbo.2025.101663>
19. Ravindhran B, Ubhi L, Nazir S, Pymer S. A systematic review and meta-analysis of machine learning models in the prediction of adverse outcomes in peripheral arterial disease. *Br J Surg* 2025;**112**(Suppl 6). <https://doi.org/10.1093/bjs/znaf042.035>
20. Flores AM, Demas F, Leeper NJ, Ross EG. Leveraging machine learning and artificial intelligence to improve peripheral artery disease detection, treatment, and outcomes. *Circ Res* 2021;**128**(12):1833–50. <https://doi.org/10.1161/CIRCRESAHA.121.318224>
21. Li B, Aljabri B, Verma R, et al. Predicting outcomes following lower extremity endovascular revascularization using machine learning. *J Am Heart Assoc* 2024;**13**(9):e033194. <https://doi.org/10.1161/JAHA.123.033194>
22. Li B, Verma R, Beaton D, et al. Predicting outcomes following open revascularization for aorticiliac occlusive disease using machine learning. *J Vasc Surg* 2023;**78**(6):1449–1460.e7. <https://doi.org/10.1016/j.jvs.2023.07.006>
23. Li B, Eisenberg N, Beaton D, et al. Using machine learning to predict outcomes following suprainguinal bypass. *J Vasc Surg* 2024;**79**(3):593–608.e8. <https://doi.org/10.1016/j.jvs.2023.09.037>
24. Li B, Eisenberg N, Beaton D, et al. Using machine learning (XGBoost) to predict outcomes after infrainguinal bypass for peripheral artery disease. *Ann Surg* 2024;**279**(4):705–13. <https://doi.org/10.1097/SLA.00000000000006181>
25. Ravindhran B, Prosser J, Lim A, et al. Tailored risk assessment and forecasting in intermittent claudication. *BJS Open* 2024;**8**(1):zrad166. <https://doi.org/10.1093/bjsopen/zrad166>
26. Cox M, Panagides JC, Di Capua J, et al. An interpretable machine learning model for the prevention of contrast-induced nephropathy in patients undergoing lower extremity endovascular interventions for peripheral arterial disease. *Clin Imaging* 2023;**101**:1–7. <https://doi.org/10.1016/j.clinimag.2023.05.011>
27. Li B, Warren BE, Eisenberg N, et al. Machine learning to predict outcomes of endovascular intervention for patients with PAD. *JAMA Netw Open* 2024;**7**(3):e242350. <https://doi.org/10.1001/jamanetworkopen.2024.2350>
28. Goffart S, Delingette H, Chierici A, et al. Artificial intelligence techniques for prognostic and diagnostic assessments in peripheral artery disease: a scoping review. *Angiology* 2025. <https://doi.org/10.1177/00033197241310572>
29. Squiers JJ, Thatcher JE, Bastawros DS, et al. Machine learning analysis of multispectral imaging and clinical risk factors to predict amputation wound healing. *J Vasc Surg* 2022;**75**(1):279–85. <https://doi.org/10.1016/j.jvs.2021.06.478>
30. Chen MY, Cao MQ, Xu TY. Progress in the application of artificial intelligence in skin wound assessment and prediction of healing time. *Am J Transl Res* 2024;**16**(7):2765–76. <https://doi.org/10.62347/MYHE3488>
31. Ravindhran B, Prosser J, Lim A, et al. Tailored risk assessment and forecasting in intermittent claudication: a proof of concept decision support tool. *EJVES Vasc Forum* 2024;**62**:S5. Available from: <http://www.ejvesreports.com/article/S2666688X24001035/fulltext>
32. Rieke N, Hancox J, Li W, et al. The future of digital health with federated learning. *NPJ Digit Med* 2020;**3**(1):119. <https://doi.org/10.1038/s41746-020-00323-1>
33. Ravindhran B, Cutteridge J, Pymer S, et al. Comparative performance of clinician and computational approaches in forecasting adverse outcomes in intermittent claudication. *Ann Vasc Surg* 2025;**120**:138–45. <https://doi.org/10.1016/j.avsg.2025.05.009>
34. Abbas Q, Jeong W, Lee SW. Explainable AI in clinical decision support systems: a meta-analysis of methods, applications, and usability challenges. *Healthcare* 2025;**13**(17):2154. <https://doi.org/10.3390/healthcare13172154>
35. Ribeiro MT, Singh S, Guestrin C. 'Why should I trust you?' Explaining the predictions of any classifier. Proceedings of the ACM SIGKDD International Conference on Knowledge Discovery and Data Mining. 2016;1135–44. Available from: <https://dl.acm.org/doi/pdf/10.1145/2939672.2939778>
36. Ali S, Abuhmed T, El-Sappagh S, et al. Explainable Artificial Intelligence (XAI): What we know and what is left to attain Trustworthy Artificial Intelligence. *Information Fusion* 2023;**99**:101805. <https://doi.org/doi.org/10.1016/j.inffus.2023.101805>
37. Shaffer Shane T. AI incident reporting: Addressing a gap in the UK's regulation of AI. 2024. Available from: <https://www.longtermresilience.org/>
38. Food and Drug Administration (FDA). Predetermined Change Control Plans for Medical Devices. 2024. Available from: <https://www.fda.gov/regulatory-information/search-fda-guidance-documents/predetermined-change-control-plan-s-medical-devices>
39. GOV.UK. Good Machine Learning Practice for Medical Device Development: Guiding Principles. 2021. Available from: <https://www.gov.uk/government/publications/good-machine-learning-practice-for-medical-device-development-guiding-principles/good-machine-learning-practice-for-medical-device-development-guiding-principles>
40. Final Document Good machine learning practice for medical device development: Guiding principles AUTHORIZING GROUP Artificial Intelligence/Machine Learning-enabled Working Group Preface. 2025. <https://www.imdrf.org/documents/good-machine-learning-practice-medical-device-development-guiding-principles>
41. Morris-Jarvis JA, Hatfield-Chetter G, Ravindhran B, et al. A dynamic neural network-based approach to identify high-risk patients with intermittent claudication. *J Vasc Surg* 2025;**81**(6):e94–5. <https://doi.org/10.1016/j.jvs.2025.03.221>
42. Nicholson Price W. Big data and black-box medical algorithms. *Sci Transl Med* 2018;**10**(471):eaa05333. <https://doi.org/10.1126/scitranslmed.aa05333>
43. Aliferis C, Simon G. Overfitting, underfitting and general model overconfidence and under-performance pitfalls and best practices in machine learning and AI. 2024;477–524. Available from: <https://www.ncbi.nlm.nih.gov/books/NBK610560/>
44. Hogan JW, Murthy VL. Model calibration, interpretability, and decision-making with AI-based risk scores. *NEJM AI* 2025;**2**(5). <https://doi.org/10.1056/Ale2500297>
45. Abdelwanis M, Alarafati HK, Tammam MMS, Simsekler MCE. Exploring the risks of automation bias in healthcare artificial intelligence applications: a Bowtie analysis. *J Safety Sci Resilience* 2024;**5**(4):460–9. <https://doi.org/10.1016/j.jnlssr.2024.06.001>
46. Collins GS, Moons KGM, Dhiman P, et al. TRIPOD+AI statement: updated guidance for reporting clinical prediction models that use regression or machine learning methods. *BMJ* 2024;**385**:e078378. <https://doi.org/10.1136/bmj-2023-078378>
47. Vasey B, Nagendran M, Campbell B, et al. Reporting guideline for the early stage clinical evaluation of decision support systems driven by artificial intelligence: DECIDE-AI. *BMJ* 2022;**377**:e070904. <https://doi.org/10.1136/bmj-2022-070904>
48. Sounderajah V, Guni A, Liu X, et al. The STARD-AI reporting guideline for diagnostic accuracy studies using artificial intelligence. *Nat Med* 2025;**31**(10):3283–9. <https://doi.org/10.1038/s41591-025-03953-8>
49. Moons KGM, Damen JAA, Kaul T, et al. PROBAST+AI: an updated quality, risk of bias, and applicability assessment tool for prediction models using regression or artificial intelligence methods. *BMJ* 2025;**388**:e082505. <https://doi.org/10.1136/bmj-2024-082505>
50. Rivera SC, Liu X, Chan AW, Denniston AK, Calvert MJ. Guidelines for clinical trial protocols for interventions involving artificial intelligence: The SPIRIT-AI Extension. *BMJ* 2020;**370**:m3210. <https://doi.org/10.1136/bmj.m3210>
51. Liu X, Cruz Rivera S, Moher D, et al. Reporting guidelines for clinical trial reports for interventions involving artificial intelligence: the CONSORT-AI extension. *Nat Med* 2020;**26**(9):1364–74. <https://doi.org/10.1038/s41591-020-1034-x>