Research Square

Preprints are preliminary reports that have not undergone peer review.
They should not be considered conclusive, used to inform clinical practice,
or referenced by the media as validated information.

Artificial Intelligence in Radiotherapy-Associated Cardiovascular Toxicity: A Systematic Review of Predictive and Imaging Applications

Vivian Salama

vivian.salama@hsc.wvu.edu

West Virginia University

Brandon Godinich

Paul L. Foster School of Medicine, Texas Tech Health Sciences Center, El Paso, Texas

Nathaniel Dunham

West Virginia University

Troy Nguyen

West Virginia University

Wesley Cox Cox

West Virginia University

Peyton Lilly

West Virginia University

Joseph Schmidlen

West Virginia University

Jeffrey Ryckman

West Virginia University

Ramon Siochi

West Virginia University

Ashkan Emadi

West Virginia University

Christopher Bianco

West Virginia University

George Sokos

West Virginia University

Raymond Raylman

West Virginia University

David Clump

West Virginia University

Mina Hanna

West Virginia University

Phillip Pifer

West Virginia University

Systematic Review

Keywords: Risk Prediction, Cardio-oncology, Machine Learning, Treatment planning, Cancer Survivorship, Deep Learning, Radiation Therapy, Cardiovascular Imaging, Cardiac Contouring

Posted Date: March 18th, 2026

DOI: https://doi.org/10.21203/rs.3.rs-9033968/v1

License: This work is licensed under a Creative Commons Attribution 4.0 International License. Read Full License

Additional Declarations: Competing interest reported. VS is a co-founder and equity holder (27%) in XURE.AI Inc., an early-stage artificial intelligence startup. The company has not provided financial support for this work and has no role in the design, execution or interpretation of this study. CMB has served on speaker bureaus for Bristol Myers Squibb and Pfizer, including disease state education related to ATTR cardiomyopathy. He has also received research support from Pfizer which is not related to this work. PMP has served as a consultant for Varian Medical Systems (speaker in brachytherapy-related activities), with salary support received, and no relation with this work. JR has served as a consultant and advisory board member for Elekta and Atrium OS. He also reports ownership of stock options in Aitrium OS. PL reports employment with West Virginia University Peer-Assisted Learning Service as Co-Leader and Tutor. No commercial relationships are reported. JAS reports employment with West Virginia University as co-leader and tutor for the peer-assisted learning service. No additional commercial or industry relationships are reported. The remaining authors have nothing to disclose.

Version of Record: A version of this preprint was published at Cardio-Oncology on June 19th, 2026. See the published version at https://doi.org/10.1186/s40959-026-00528-5.

Abstract

Background

Cardiovascular toxicity (CVT) is a major effect of radiation therapy (RT) and a contributor to morbidity and mortality among cancer survivors. Artificial intelligence (AI) may improve early detection, risk stratification, and RT planning to mitigate cardiac exposure, but the current evidence has not been comprehensively synthesized. The main objective of this study is to analyze and assess the quality of literature applied AI assessments of CVT in populations receiving RT.

Methods

PRISMA-guided systematic review of PubMed, Ovid EMBASE, Cochrane Library, and Web of Science was conducted through October 1, 2025. Eligible studies were original human research in English applying AI to cardiovascular outcomes or imaging in cancer populations receiving RT. Predictive-model studies were assessed using TRIPOD + AI for quality and PROBAST for risk-of-bias. Imaging-AI studies were assessed using CLAIM and QUADAS-2 for quality and risk-of-bias respectively.

Results

Sixty-five studies were included and clustered into two groups: (1) AI prediction of RT-associated CVT (n = 31, 48%) and (2) AI-based cardiovascular imaging (n = 34, 52%). Deep learning was the most frequent approach (45/65, 69%) especially in imaging and showed highest performance (median AUC = 0.82 & sensitivity = 0.83) in prediction. Predictive models lacked calibration assessment (3/31, 10%), and external validation (6/31, 19%). TRIPOD + AI adherence averaged 79% (SD 22.68%), while PROBAST rated 97% at high overall risk-of-bias. Imaging models demonstrated strong performance (overall median DSC = 0.85, range: 0.76–0.94) particularly for larger cardiac structures, whereas coronary artery segmentation remained challenging. CLAIM adherence averaged 71%, and QUADAS-2 judged 82% at high risk-of-bias.

Conclusion

AI approaches in radiation-associated cardio-oncology are promising but not yet implementation-ready. Future work should prioritize standardized endpoints, robust external validation, calibration and clinical utility evaluation, shared high-quality imaging annotations, and prospective integration into clinical trials.

Highlights

- Radiation therapy increases long-term cardiovascular toxicity in cancer survivors.

- AI shows promise for prediction and imaging but lacks robust validation.
- Methodological quality remains limited despite TRIPOD-AI and CLAIM guidance.
- Standardized endpoints, external validation, and prospective clinical integration are urgently needed.

Introduction/Background

Cancer therapies can damage the heart and cardiovascular system, resulting in cardiovascular toxicities (CVT) that negatively impact outcomes and survival [1]. Radiation therapy (RT) is a cornerstone modality of cancer treatment, offering substantial survival benefits across multiple malignancies including breast, lung, and thoracic cancers [2, 3]. Although RT techniques improve the survival and long-term outcomes of patients, exposure of normal structures to ionizing radiation can lead to several side effects affecting patients' outcome and quality of life (QoL) [4]. Exposure of the heart and vascular structures to thoracic radiation either as a single modality or in combination with chemotherapy can lead to a spectrum of CVTs, ranging from subclinical myocardial injury to life-threatening events such as ischemic heart disease, heart failure, arrhythmias, and pericardial disease (Fig. 1) [5–7]. The long-term cardiovascular consequences of RT have become an increasingly important clinical concern in cardio-oncology and radiation oncology domains, as they affect patients' outcome, overall survival, and QoL. Approximately 20–30% of patients with lung or esophageal cancers develop CVT after RT [8, 9]. Similarly, about 55% of patients with lymphoma develop CVT after receiving mediastinal RT, although anthracycline-containing chemotherapy by itself also increased risks of CVAE [10]. Notably, chemotherapy, regardless of receipt of RT, may accelerate vascular aging, supporting the need for intensive cardiovascular risk management during survivorship [11, 12]. Cardiovascular disease (CVD) has emerged as the leading cause of chronic health complications and non-cancer mortality in cancer survivors [13–15]. Early detection and personalized risk stratification remain challenging due to the complex interaction of patient factors, treatment parameters, and biological responses, which is one of the main focuses of cardio-oncology.

In recent years, artificial intelligence (AI) and machine learning (ML) have emerged as powerful tools capable of analyzing large, multidimensional datasets generated throughout the cancer care continuum [16–18]. Artificial intelligence methods have advanced dose prediction, automated treatment planning, auto-segmentation, outcome prediction, treatment personalization and decision making [19–22]. Digital AI techniques are promising for understanding and predicting cardiotoxicity, as they can integrate clinical factors, systemic therapy factors, dosimetric variables, imaging metrics (i.e., radiomics), and biomarkers to uncover patterns not easily captured by traditional statistical approaches. Different AI applications could be beneficial in the field of post treatment CVT, such as CVD predictions, risk stratification, image analysis including auto-segmentation, radiomic features extraction and treatment dose planning (i.e., dosiomics) as illustrated in Fig. 1. Chen et al.,[23], Loap et al.[24] and Borges et al.[25] showed the efficacy of the deep learning (DL) in imaging auto-segmentation of the cardiovascular structures for better RT planning to reduce the incidence and severity of post-radiation CVT. While Zhou et al.[26], Qiao et al.[27], Bentriou et al.[28] and Dincer et al.[29] used ML classification and regression models (e.g., Logistic

regression, Gradient boosting, and Random Forest,) for CVT prediction after RT. As data availability and computational digital capacity grow, AI-based models are increasingly positioned to support personalized risk prediction and guide cardioprotective strategies.

Despite the rapid expansion of AI in cardio-oncology and radiation oncology, its application to CVT in populations receiving RT as a component of care is largely unappreciated. Existing studies vary widely in methodology, data sources, outcome definitions, and validation strategies. While several groups have explored AI-driven cardiac contouring, dose–response modeling, and risk prediction, no comprehensive study currently exists that evaluates how AI is being used, the quality of available studies, and the gaps that remain. Understanding the current landscape is critical for clinicians, researchers, and computational scientists aiming to improve survivorship outcomes and QoL of patients.

Our scientific questions in this review were:

1. What are the key applications of AI models for identifying and evaluating cardiovascular toxicity in patients receiving cancer treatment with RT as a component of care?
2. Which AI tools are commonly used to address/predict cardiovascular treatment-associated toxicity in populations receiving RT?
3. What is the quality of studies conducted in AI in cardio-oncology among patients undergoing treatment that includes RT?
4. To what degree do the existing AI models in cardio-oncology follow the AI reporting and transparency guidelines with respect to model performance reporting?

This systematic review aimed to evaluate and summarize the existing literature on AI applications related to CVT among patients undergoing treatment that includes RT. Specifically, we assessed the types of AI models developed, their input data, prediction targets, performance metrics, and clinical applicability. By identifying strengths, limitations, and opportunities within the field, this comprehensive review provides a foundation for future research and development of AI-driven tools that can enhance early detection, optimize treatment planning, and ultimately reduce cardiovascular risk in cancer survivors.

Materials and Methods

Protocol Registration

Registration of this review was conducted in the international prospective register database of systematic reviews (PROSPERO) on September 29th, 2025 (ID number: CRD420251158012) in the context of human health care.

Search Strategy

A comprehensive systematic search of PubMed, Ovid EMBASE, Cochrane Library, MEDLINE, and Web of Science databases was conducted for publications of original human research in English up to October 1st, 2025. The concepts searched included: “artificial intelligence”, “machine learning”, “deep learning”, “neural networks”, “cardiovascular toxicity”, “cardiovascular diseases”, “radiotherapy”, “cardio-oncology”. Combination of the concepts/terms was done using OR/AND Boolean operators. The search strategy using Boolean operators is described in Supplementary File S1.

Screening Process

Articles identified from databases were uploaded into Rayyan tool: an AI-powered systematic review management platform for screening. Duplicated articles were identified through Rayyan AI and were fixed. The screening process was blindly conducted with three independent reviewers (BG, ND, and PL), then a fourth reviewer (VS) solved the conflicts. Screening of the titles and abstracts was done first, and then screening of the full text was done next for the reports for retrieval. The primary outcome of this review was to identify the applications and quality of AI approaches in either predicting, preventing or managing treatment-associated cardiovascular toxicity in patients with cancer receiving care that includes RT.

Inclusion criteria

Included studies had to meet all the following criteria: (1) full original Research, (2) Available in English, (3) involved human subjects, (4) studies applied AI or ML models, (5) cardiovascular related outcome (including cardiovascular diseases or cardiovascular imaging), (6) studies in the cancer population received radiation therapy or chemo-radiotherapy, and (7) models were tested for performance.

Exclusion criteria

Articles were excluded if they met any of the following criteria: (1) irrelevant or out of scope study, (2) not a full original study (e.g., conference abstract, review articles, letter, or editorial), (3) non-human study (e.g., animal or cell lines), (4) no AI or ML model was applied, (5) no cardiovascular related outcome, or (6) no radiation therapy received.

Data Synthesis, Extraction, Collection, and Analysis

This review followed the Preferred Reporting Items for Systematic Reviews and Meta-Analysis (PRISMA) guidelines. Data from final included articles were extracted into Excel sheets. Extracted data included study characteristics and demographics (author, year, and journal), study design, number of population/participants, cancer type, AI method/model, input and outcome variables, and AI models performance metrics (e.g., AUC, accuracy, sensitivity, specificity and others). Due to significant heterogeneity in study design, AI models, input variables, outcomes, and performance metrics, a quantitative meta-analysis was not performed. Instead, a narrative synthesis was conducted. All included studies were categorized into two main groups based on the main applications of AI: (1) studies applied AI in cardiovascular toxicity prediction and stratification in patients receiving cancer

treatment with RT as a component of care (2) studies applied AI models in cardiovascular imaging analysis in patients receiving RT. Summary statistics including median, interquartile range (IQR), and range were calculated for reported model performance metrics using JMP.Pro Edition 18.2.1 software.

Results' Presentation

Results were summarized in a tabular format to summarize study characteristics, AI methodologies, and conclusions. Visualizations (e.g. bar charts, pie charts, and flow diagrams) were generated using GraphPad prism to illustrate the trend of publication, cancer types, type of research study, TRIPOD+AI and CLAIM adherence, and risk-of-bias assessment.

Data Quality and Risk of Bias Assessment

Blind screening of the identified articles was conducted by three reviewers (BG, ND and PL). Full texts of the included articles were assessed thoroughly by three reviewers (BG, ND, and PL). Conflicts were resolved by a fourth reviewer (VS). The materials and methods of the studies and the results sections were assessed. Non-full articles, or full articles that could not be accessed were excluded.

To evaluate the overall quality of the included articles, a rigorous assessment was conducted, focusing on evaluation of both the risk of bias and adherence to AI reporting guidelines for each individual article based on the application of the AI and the type of the study. For studies applying AI or ML for predictive tasks and risk stratification, the risk of bias was assessed using the Prediction Model Risk of Bias Assessment Tool (PROBAST), which examines four domains (participants, predictors, outcomes, and analysis) to determine the overall risk of bias (checklist in Supplementary File S) [17, 18, 30, 31]. Adherence to the updated AI-Transparent Reporting of a Multivariable Prediction Model for Individual Prognosis or Diagnosis (TRIPOD+AI) guidelines was analyzed using the TRIPOD+AI checklist, which covers 27 items including overall 51 questions and subitems specified for AI models [18, 32-34]. The checklist is in Supplementary File S4. While for studies/articles applying AI in cardiovascular imaging analysis, an updated Checklist for Artificial Intelligence in Medical Imaging (CLAIM) [35] was used to assess the quality of these studies and consistent reporting of AI in medical imaging. Risk of bias for the included imaging-AI studies was evaluated using the Quality assessment of diagnostic accuracy studies (QUADAS-2) tool across 11 questions covering 4 domains: patient selection, index test, reference standard, and flow and timing [36, 37]. If any answer for any question in the domain was assessed as "*No*", the final risk-of bias of the domain was judged as high.

Formal assessment of publication bias and certainty was not performed due to the heterogeneity of study designs and absence of pooled meta-analytic estimates. However, selective reporting within studies was indirectly evaluated through adherence to reporting guidelines (TRIPOD+AI and CLAIM), which assess transparency in outcome reporting and model performance metrics. Additionally, the overall certainty of the body of evidence was evaluated qualitatively considering the risk of bias (PROBAST and QUADAS-2 results),

Results

Search and Screening Results

Our comprehensive database search resulted in a total of 1583 studies from PubMed (n = 856), Ovid EMBASE (n = 241), Cochrane Library (n = 95) and Web of Science (n = 391). After a thorough screening process following the eligibility assessment criteria, a total of sixty-five articles met all the inclusion criteria. The full search process is illustrated in the PRISMA flow chart (Figure 2).

Trend, study design, cancer type and participants

All included studies were published from 2018 to 2025, with an increase in the number of publications for AI studies in cardio-oncology research particularly with RT, from 2 per year in 2018 to 16 and 10 per year in 2024 and 2025, respectively (Figure 3.a). The most common type of cancer included in these studies was breast cancer (n=26, 39%) followed by lung cancers (n=22, 34%) (Figure 3.b). More than half of the studies included used a retrospective single institutional cohort to develop their models (n=36, 55% studies) (Figure 3.c). The median sample size to build the models among all studies was 177 (range: 9 – 148755036, IQR 54-984), while 2 articles did not specify the sample size of patients or participants. Several types of AI models or frameworks were investigated in the included studies (Figure 3.c). Neural networks (Deep learning) were the most common AI models (n=45, 69%) especially in imaging analysis (e.g., auto segmentation, calcium scores calculation, or dosiomics), followed by the regression models (n=19, 29%) then classification ensembles models (e.g., Boosting models and random forest) (n=7, 11% per model).

Studies characteristics and AI applications

The characteristics and demographics of the 65 included studies and the AI models applied in these studies for CVT prediction or imaging analysis were illustrated in Supplementary Table 1. Out of the 65 included studies, 34 (52%) studies applied AI for imaging analysis, while 31 (48%) studies applied AI or ML for prediction, risk stratification, and patient selection of CVT following RT.

1. **Artificial Intelligence models applications in cardiovascular toxicity prediction and stratification in patients receiving cancer treatment with RT as a component of care**

Models' characteristics and endpoints

Across the 31 predictive modeling studies, endpoints/outcomes can be classified into several major outcome categories: clinical cardiac events, arrhythmia, heart failure or cardiomyopathy, composite/general toxicity, single toxicity (e.g., pericardial diseases valvular diseases, ischemic or coronary diseases), survival, and dosimetric or planning metrics. Major adverse cardiac events (MACE) were the most frequently used clinical endpoint, reported in 6/31 (19%) predictive modeling studies. MACEs were defined as composite cardiovascular outcomes including myocardial infarction, coronary artery disease, heart failure, or stroke and used with grading commonly based on Common Terminology

Criteria for Adverse Events (CTCAE) ≥3 or clinically judged cardiovascular events. General cardiotoxicity or cardiac event endpoints were used in 6/31 (19%) studies. These included treatment-related cardiac events or cardiotoxicity without restriction to a single diagnosis, which were defined as development of cardiovascular disease, cardiac dysfunction, or high-grade clinically significant cardiac disease. Heart failure or cardiomyopathy–specific endpoints were reported in 3/31 (10%) studies. These included late-onset cardiomyopathy (Güntürkün et al.)[38], chemoradiation-induced heart failure defined by reduction in ejection fraction (Ansari et al.)[39], and high-grade cardiac disease including heart failure requiring treatment (Bentriou et al.). Arrhythmia-specific endpoints were used in 3/31 (10%) studies, focusing primarily on atrial fibrillation (AF). Pericardial or valvular disease endpoints were uncommon, appearing in just 2/31 (6%) studies. Niedzielski et al. evaluated radiation-induced pericardial effusion, while Chounta et al. [40], focused on radiation-induced valvular heart disease. Ischemic or coronary-specific endpoints were used in 2/31 (6%) studies. Choi et al.,[41] predicted acute coronary events, while Xie et al., [42] used coronary CT angiography to identify coronary artery impairment and subsequent MACEs. Overall survival (OS) was used as the primary endpoint in 3/31 (10%) studies. Procedure-based or health-system outcomes were uniquely evaluated in just 1/31 studies (3%), with Monlezun [43] et al., using in-hospital mortality, PCI utilization, complications, and cost as endpoints. A significant proportion of studies (8/31, 26%) used dosimetric or treatment-planning endpoints rather than clinical cardiac events. These included prediction of mean heart dose or cardiac substructure dose.

Models' validation and evaluation

Model validation mostly relied on internal validation strategies, with cross-validation used in 12/31 (38%) studies, including five-fold stratified cross-validation, repeated k-fold or bootstrap resampling and leave-one-out validation for treatment-planning decision support. Independent hold-out testing was used in 9/31 (29%) studies, typically using 60–40 or 70–30 train–test splits. External validation using separate patient cohorts was performed in 6/31 (19%) studies. Several studies incorporated model comparison or benchmarking against conventional dosimetric or radiomics approaches, showing improved performance of ML dosiomics-based methods over traditional dose-volume metrics. Interpretability-driven validation was uniquely emphasized in 4/31 (13%) studies through explainable AI techniques such as SHapley Additive exPlanations (SHAP) and Gradient-weighted Class Activation Mapping (Grad-CAM) heatmaps to confirm clinically meaningful cardiac substructure contributions (Ladbury [44] et al., and Choi et al.[41]), while 1 of 31 studies validated causal inference outputs by comparison with established medical literature and expert clinical review rather than predictive accuracy metrics (Bernasconi et al. [45]).

For evaluation of the performance of ML predictive models, most studies used discrimination metrics for classification/predictive models [e.g., Receiver Operating-Area Under the Curve (AUC), sensitivity/recall, specificity/precision, and accuracy]. Twelve out of the 31 (39%) included studies applied predictive AI/ML models for CVT prediction and risk stratification didn't report any of the discrimination metrics results raising concerns regarding these predictive models. Out of the 31 studies that used ML predictive

models, 16/31 (52%) reported the model's AUC. Across the included studies all predictive ML models demonstrated good discriminatory performance, with mean AUC generally clustering around 0.80–0.83 and with median AUC values consistently around or above 0.80 (Figure 4.a). Neural network models demonstrated the highest discriminatory performance (median AUC = 0.82; IQR: 0.74–0.92; range: 0.69–0.98), although with notable variability, indicating high but variable performance. This was followed by K-nearest neighbors (median AUC = 0.81; IQR: 0.79–0.88; range: 0.66–0.98) and boosting models (median AUC = 0.81; IQR: 0.75–0.89; range: 0.66–0.96), which showed comparable central performance. Random forest, support vector machine, and regression-based models demonstrated similar and slightly lower median performance (all median AUC ≈ 0.80), with random forests showing the most stable distribution (IQR: 0.77–0.86; SD: 0.08; range: 0.66–0.98). Naïve Bayes models exhibited a comparable median AUC (0.80) but greater dispersion (IQR: 0.74–0.94; SD: 0.11; range: 0.73–0.97). Overall, the narrow interquartile ranges across most model families suggest consistent central performance, while wider ranges highlight substantial heterogeneity in upper-bound discrimination across studies.

In terms of other discrimination metrics, 22/31 (71%) did not report the ML models' sensitivity, or specificity, and 21/31 (68%) did not report the model's accuracy. With the few reported metrics, neural networks models demonstrated the highest performance, achieving a median sensitivity of 0.83 with a narrow IQR of 0.82–0.85 and a range of 0.80–0.86, indicating consistently high true-positive detection. This was followed by regression-based models, which showed a median sensitivity of 0.82 but with wider variability (IQR: 0.73–0.90; range: 0.66–0.96). Random forest models demonstrated moderate sensitivity (median = 0.73; IQR: 0.62–0.78; range: 0.59–0.83). Boosting models, reported in a single study, showed a sensitivity of 0.78. Only one article reported the F1-Score of the AI predictive mode (Vendrame et al.,) [46].

Regarding models' specificity, random forest models demonstrated the highest specificity (median = 0.83; IQR: 0.79–0.86; range: 0.77–0.88), indicating consistently strong true-negative classification. This was followed by boosting models (specificity = 0.81), although this estimate was derived from a single study, limiting assessment of variability. Neural network models showed a median specificity of 0.80 with broader dispersion (IQR: 0.72–0.88; range: 0.67–0.91). Regression-based models demonstrated slightly lower specificity (median specificity = 0.78; IQR: 0.69–0.86; range: 0.66–0.86). Overall, tree-based ensemble methods, particularly random forests, showed the most favorable and stable specificity, whereas neural networks and regression approaches exhibited greater variability across studies.

Only 3/31 (10%) studies tested the model's calibration. Niedzielski et al. [47] used the calibration curve and tested the calibration slope/intercept for the Elastic net logistic regression model: 1.356 / 0.235 (cross-validated heart + substructures model); 1.352 / 0.174 (final fitted model) [47]. Tohidinezhad et al. [48], tested the calibration of the regression model using the calibration curve which showed good conformation of the predicted probabilities versus the actual outcomes for probabilities ranging between 0.1 and 0.9. Lee et al. assessed the multivariable logistic regression–based normal tissue complication

probability (NTCP) model using Hosmer–Lemeshow [H-L] goodness-of-fit test). = 0.565 (development cohort); P = 0.800 (validation cohort) [49].

Quality of studies

The **TRIPOD+AI** checklist assessment of studies applying AI/ML for prediction, risk stratification, and patient selection in cardiovascular toxicity after RT showed moderate overall reporting quality, with an overall average/mean adherence of 79% (SD=22.68%, 95%CI=72.36-84.99) across all items, since 9 out of 31 articles had adherence to the guidelines below 70% and 13/51 domains fell below 70%. Core scientific reporting domains, such as title, abstract, objectives, data sources, outcome definitions, participant characteristics, model development, performance reporting and discussion demonstrated high adherence (≥90-100%), reflecting strong methodological transparency in fundamental study components. However, several critical AI-specific and ethical/equity-related domains showed markedly low adherence, particularly the ethical approval (29%), outcome blind assessment and subjective interpretation (16% & 32% respectively), sample size explanation (23%), analytical handling of model performance heterogeneity (23%), model updating (65%), class imbalance and fairness (71% for each), transparency compliance with open science including sharing the protocol, registration, code sharing and patient and public involvement ((68–71%). Moreover, the usability of the model, particularly user interaction (52%) was not covered in the discussion sections of the included studies. (Figure 4.b)

Risk-of-bias assessment

Using the **PROBAST** checklist, the risk-of-bias of the included 31 predictive studies revealed substantial variability across domains. The overall risk-of bias of 30/31 (97%) was high due to the high risk in the analysis domain. Low risk-of-bias was most frequently observed in Predictors Domain 2, with 28/31 (90%) of studies rated low risk. Domain 3: Outcome, where 26/31 (84%) demonstrated adequate outcome definition and assessment. Domain 1: Participants showed moderate concerns, with 22/31 (71%) of studies rated low risk and 8/31 (26%) judged unclear, largely due to insufficient reporting of participant selection and sample size explanation. In contrast, Domain 4: Analysis exhibited a critically high risk-of-bias, with 30/31 (97%) of studies rated high risk, reflecting frequent methodological shortages such as inadequate handling of overfitting, lack of appropriate validation, model calibration and insufficient reporting of performance measures. Consequently, the overall risk-of-bias was rated as high in 97% of studies, indicating that despite relatively strong reporting of predictors and outcomes, analytical limitations substantially compromise the reliability and clinical applicability of current AI-based CVT prediction models after RT (Figure 4.c). Detailed Risk-of-Bias per study is included in Supplementary File S3.

2. **Artificial Intelligence models applications in cardiovascular imaging in patients with cancer who received RT as a component of care**

Models' characteristics

Across the 34 included studies, DL approaches are predominantly used, specifically in about 30 out of 34 studies (~ 88%). Most of these DL models were convolutional neural network (CNN)–based architectures, primarily U-Net variants (standard U-Net, 3D U-Net, nU-Net, cascaded U-Nets).

Application-wise, the included imaging studies clustered the AI applications in cardiovascular imaging for radiation treatment into three dominant use cases include: 1) cardiac or cardiopulmonary substructures segmentation was the most common application, involving about 22/34 (65%) studies. 2) Coronary or thoracic aorta calcification quantification was addressed in about 8/34 (24) studies. 3) Functional radiomics or dosiomics constituted the remaining 4/24 (11%) studies. Among the segmentation-focused studies, CNN-based auto-segmentation models were used in more than 85% of the studies. These models were primarily developed to automate whole-heart and cardiac substructure contouring on radiotherapy planning CT, aiming to improve radiation dose assessment, workflow efficiency, or downstream outcome modeling.

Models' validation, and evaluation

The most common validation approach was comparison against expert manual contours, used in about 22 (65%) studies, typically involving one to three radiation oncologists and, in several high-quality studies, multi-observer consensus (e.g., Garrett Fernandes et al.[50]; Chen et al.[23]).

Several additional validation strategies were observed such as external or independent test sets in 12/34 studies, clinical acceptability scoring by physicians in 8 studies, prospective or workflow-embedded validation in 5 studies, outcome-based validation (survival, toxicity, fitness) in ~7 studies and hybrid geometric + dosimetric validation in 10 studies.

Performance evaluation and validation were mainly reported by calculating the Dice Similarity Coefficient (DSC) to evaluate the accuracy of AI-driven auto-segmentation of structures. Among the 18–20 studies (55–60%) that reported DSC for segmentation accuracy, the overall median DSC across cardiac structures was approximately 0.85, with a range of 0.76–0.94. Performance strongly depended on structure size. Whole-heart segmentation consistently showed the highest agreement, with median DSC values of 0.93–0.95, while cardiac chambers demonstrated slightly lower but still strong performance (0.88–0.92). Great vessels achieved moderate agreement (0.80–0.88). In contrast, coronary artery segmentation was a clear outlier, with substantially lower DSC values (0.56–0.65) and occasional near-zero overlap for distal segments, reflecting the technical challenges of segmenting small tubular structures on non-contrast CT. One feasibility study (Thor et al. [51]) intentionally reported very poor performance (DSC 0.15) to highlight current methodological limitations for left anterior descending (LAD) artery segmentation on non-contrast CT.

Approximately 12/34 (35%) studies did not report DSC because segmentation overlap was not the primary endpoint or was methodologically inappropriate, particularly in calcification scoring tasks. Instead, these studies emphasized clinically oriented validation metrics, including intraclass correlation coefficients (ICC 0.85–0.99) for calcium burden quantification, Cohen's kappa (κ = 0.85–0.91) for risk

stratification, surface distance metrics for coronary localization, and dosimetric or outcome-based validation in nearly one-third of studies.

Quality of studies

Across the 34 included studies applying AI to imaging analysis, overall average/mean of reporting adherence to the **CLAIM** checklist was 71% across all domains (71%, SD=29, 95%CI=62.10-79.90), since 12/34 (35%) articles fell below 70% with overall mean=71% (SD=11.58. 95%CI= 66.61-74.69) of adherence across all the articles (median 71.5%, range 34%-91%, IQR=65.5%-71.5%). All studies (34/34, 100%) clearly reported the AI method in the title, provided an abstract summary, described the clinical or scientific background, and stated study aims. Most studies demonstrated high compliance in the methods and specified the study goal (100%), data sources (100%), image acquisition protocols (33/34, 97%), reference standard definitions (32/34, 94%), AI models' evaluation of performance metrics and statistical uncertainty (32/34, 94%), and internal validation procedures (100%). However, important methodological details were inconsistently reported. Only 8/34 (24%) of studies described data de-identification procedures, 7/34 (21%) reported handling of missing data, and 6/34 18% assessed inter- or intra-rater variability. Details regarding dataset partitioning were modestly reported, including data partition; assignment of cases to training, validation and testing (22/34, 65%) and disjointness level of separation is at which level (patient, study or image) (20/34, 59%), while intended sample size justification was provided in 21/34 (62%) of studies. Model transparency was limited, with only 11/34 (32%) reporting model initialization, 7/34 (21%) describing ensemble techniques, and 20/34 (59%) conducting robustness or sensitivity analyses. Explainability or interpretability methods were reported only in 6/34 (18%) of studies. While internal validation was reported in all studies (100%), external validation was performed in only 11/34 (32%), and very limited studies reported any clinical trials registration (2/34, 8%) revealing lacking clinical trials and real-world application. Reporting within the Results section was variable, with demographics or clinical characteristics presented in 17/34 (50%) of studies, performance metrics in 31/34 (91%), performance variability or precision in 26/34 (76%), and failure analysis in only 16/34 (47%). Although study limitations and implications for practice were consistently discussed (both 100%), only 56% of studies reported funding sources, and availability of code, models, or data was reported in 74% and 56% respectively. (Figure 5)

Risk-of-bias assessment

The overall **QUADAS-2** Risk-of-Bias among the included 34 AI-imaging studies, was high (82%), as only 6/34 (18%) studies were assessed as "Unclear-risk" (with at least one domain unclear-risk with low risk in the other domains) while no study (0%) was assessed as low-risk (with low-risk among all the 4 domains and yes answered for all 11 questions), however, 28/34 (82%) studies were assessed as high risk-of-bias (4/34 12% had one high-risk domain, 10/34 29% and 12/34 (35%) studies had 2 and 3 high-risk domains respectively, while 2/34 (6%) had the four high-risk domains). The four domains (Figure 6.e) assessment across all studies demonstrated that flow and timing domain demonstrated the lowest risk of bias, with only 10/34 29% of studies rated as high-risk, 3/34 (9%) as unclear-risk and 21/34 (62%) as

low-risk [33/34 (97%) of studies reported appropriate intervals between index tests and reference standards, 32/34 94% ensured all patients received the same reference standard, and 21/34 62% included all patients in the final analysis] (Figure 6.d). In contrast, patient selection showed only 4/34 (12%) of studies at low-risk, 11/34 (32%) at unclear-risk and 19/34 (56%) at high-risk [13/38 56% of studies enrolled consecutive or random patient samples and 9/34 (26%) avoided inappropriate exclusions, while 33/34 (97%) avoided case–control designs] (Figures 6.a). For the index test domain, 25/34 (74%) were at high-risk [only 11/34 (41%) of studies reported blinded interpretation of model outputs/results relative to the reference standard, although prespecified decision thresholds were reported in 31/34 (91%)] (Figure 6.c). In the reference standard domain, 14/34 (41%) of studies were judged at high-risk [most studies employed reference standards likely to correctly classify the target condition (33/34, 97%), however, only 14/34 (41%) reported blinded interpretation of reference annotations relative to AI predictions] (Figure 6.d). Detailed Risk-of-Bias per study is included in Supplementary File 3.

Discussion

This comprehensive review of a total included 65 studies, highlighted the rapid but methodologically heterogenous expansion of AI applications within cardio-oncology and radiation-oncology fields, specifically for the prediction, and management of RT–associated cardiovascular toxicity. Our review demonstrated that AI models addressed one of the field's most pressing needs, which is the early identification of patients at highest risk of radiation-induced cardiovascular injury and optimization of treatment planning through RT-planning imaging mitigation to reduce cardiac exposure and post-treatment injury. Despite the recent surge in the use of promising AI technologies for managing radiation-induced CVT, our review raises significant concerns regarding the quality and transparency of current studies. There is considerable variability in outcome definitions, input features, validation strategies, and performance reporting. Methodological quality is inconsistent, with frequent limitations in validation rigor, calibration assessment, and adherence to AI-specific reporting standards, underscoring a substantial gap between technical development and clinical implementation in cardio-oncology.

Two complementary AI application streams emerged: predictive risk modeling and cardiovascular imaging automation. Predictive AI models primarily targeted MACEs, heart failure, arrhythmias, and composite cardiotoxicity endpoints; outcomes that are highly relevant in long-term cancer survivorship. The integration of clinical, dosimetric, imaging, and radiomic variables through AI offered a multidimensional understanding of radiation-associated cardiac risk that extends beyond traditional dose–volume metrics. This aligns with the evolving paradigm of precision cardio-oncology, where individualized risk profiles could guide cardioprotective interventions, surveillance intensity, and RT planning decisions. Overall, these models demonstrated good discriminatory performance (median AUC about 0.80), particularly neural network–based approaches, suggesting that AI can capture complex nonlinear interactions between radiation dose distributions and patient-specific risk factors. Discrimination metrics such as AUC, reflect a model's ability to distinguish between patients who will

and will not develop cardiovascular toxicity (i.e., event, or no-event) [52]. The median AUC values observed across studies suggest that many of these models achieve clinically meaningful risk stratification between higher- and lower-risk patients following RT. However, discrimination alone does not establish clinical usefulness. The limited reporting of calibration, sensitivity, and specificity substantially weakens confidence in the readiness of these models for real-world implementation. Calibration is essential to determine whether predicted probabilities accurately reflect observed event rates [53]. From a clinical oncology perspective, discrimination alone is insufficient; well-calibrated models are essential for translating predicted risks into actionable thresholds for surveillance or preventive therapy. Interestingly, random forest models tended to demonstrate higher specificity, suggesting fewer false-positive predictions. This highlights a potential trade-off in model selection between maximizing detection of cardiovascular toxicity and minimizing unnecessary downstream testing or interventions.

AI-based cardiovascular imaging studies in radiation oncology have largely centered on automated cardiac substructure segmentation and calcification assessment, with the goal of improving the precision and reproducibility of cardiac dose estimation. Across the included literature, deep learning segmentation models have shown strong geometric performance for larger structures such as the whole heart and cardiac chambers, supporting prior evidence that automated contouring can reduce interobserver variability and enhance dosimetric consistency in treatment planning [54]. Arjmandi et al., demonstrated that DL-automated contouring approach reduced the inter-expert variability and increased dosimetric accuracy in prostate cancer RT planning, compared to the gold standard reference contouring [54]. This is particularly relevant in cardio-oncology, where more accurate cardiac delineation may strengthen dose–response modeling and improve heart-sparing strategies especially in thoracic RT for thoracic cancers (e.g., breast, lung, esophageal). However, segmentation accuracy drops substantially for small, low-contrast structures such as the coronary arteries [55, 56]. This limitation is clinically significant because ischemic heart diseases caused by coronary artery disease are a common mechanism of radiation-induced cardiac morbidity [57, 58]. Higher total RT doses with the larger irradiated cardiac volumes in addition to the presence of preexisting cardiovascular comorbidities are major risks were identified as high risk factors of coronary artery diseases after RT [10, 57, 59]. Darby et al. demonstrated that heart exposure to radiation increases the rates of ischemic heart diseases after RT in patients with breast cancer [57].

Challenges related to non-contrast planning CT, motion artifacts, small sized vessels, and limited high-quality annotations continue to hinder reliable coronary segmentation [60]. As a result, although current AI imaging tools show promise for improving global cardiac dose assessment, further methodological advances are needed before they can fully support coronary-specific risk modeling and precision cardio-protection in RT. Promising recent advances in the segmentation of small, low-contrast vascular structures, particularly the coronary arteries have been driven largely by innovations in deep learning architectures specifically designed to address limited contrast, small vessel caliber, and complex branching morphology. While enhanced U-Net–based models remain widely used as strong baselines,

current research increasingly favors approaches that incorporate structural priors, multi-scale contextual modeling, and noise-robust learning strategies [61]. Diffusion-based segmentation frameworks represent one of the most recent developments, leveraging probabilistic denoising and spatial attention mechanisms to improve boundary preservation and continuity in low signal-to-noise environments [62, 63]. In parallel, anatomy-guided models that embed prior knowledge of coronary topology or myocardial spatial relationships have demonstrated improved robustness and reduced false-positive segmentation by constraining predictions within physiologically plausible regions [64]. Geometry-aware and vectorized representations of vascular trees have also emerged as a promising direction, enabling more faithful reconstruction of continuous vessel pathways and improving downstream quantitative analysis [65]. Additionally, frequency-domain feature learning and advanced attention mechanisms are increasingly used to enhance detection of fine vessel edges and small distal branches [3]. Collectively, these approaches reflect a broader shift from purely voxel-based segmentation toward structure-aware, context-informed, and noise-resilient modeling strategies, which appear particularly well suited for delineating small, low-contrast coronary anatomy in contemporary cardiac CT imaging.

Despite encouraging technical performance of AI approaches, methodological quality remains a major barrier to clinical translation. Most predictive studies included in this review were judged high risk-of-bias, particularly in the analysis domain, due to inadequate validation, small sample sizes, and incomplete performance reporting. Imaging studies similarly showed high risk-of-bias related to patient selection, blinding, and reference standard interpretation. In the context of clinical oncology, where decisions may influence life-long cardiac monitoring or modification of cancer therapy, such studies limitations are particularly critical and consequential. Moreover, limited adherence to AI-specific reporting standards either in TRIPOD-AI or CLAIM, revealed gaps in transparency, reproducibility, and ethical considerations. Cardio-oncology populations often include older adults with comorbidity, yet few models addressed class imbalance, demographic diversity, or health equity, raising concerns about generalizability.

Clinically, the field appears to be transitioning from proof-of-concept innovation to early translational exploration but has not yet reached implementation readiness. Only a minority of studies included external validation or workflow-integrated testing, and very few linked AI outputs to clinical decision pathways. For AI to meaningfully influence cardio-oncology practice, future work must move beyond retrospective modeling toward prospective validation, multi-institutional datasets, standardized CVT definitions, and integration with cardiology surveillance frameworks.

Future directions

Importantly, future collaboration between radiation oncologists, cardiologists, radiologists, imaging scientists, and data scientists will be essential to ensure that AI tools address clinically relevant endpoints and fit within survivorship care models. In addition, future research should prioritize model calibration, clinical utility analyses, and decision-curve evaluation to determine whether AI predictions translate into improved patient outcomes. Development of shared, high-quality annotated imaging

datasets, particularly for coronary artery structures, will also be critical to advance precision cardiac dose modeling. Finally, embedding AI tools into prospective clinical trials and survivorship programs will be necessary to establish real-world effectiveness and safety before routine clinical adoption.

Limitations

This review has several limitations. First, heterogeneity existed across included studies in cancer types, cardiovascular endpoints, AI architectures, and performance metrics, which precluded quantitative meta-analysis. Second, many studies lacked complete methodological reporting, making risk-of-bias assessment dependent on available descriptions and potentially underestimating weaknesses. Third, the rapidly evolving nature of AI research means that some emerging tools or preprints may not have been captured within the search timeframe. Finally, reported performance metrics may overestimate real-world effectiveness because most studies relied on internal validation and retrospective datasets, which are prone to overfitting and spectrum bias.

Conclusion

Artificial intelligence applications in radiation-associated cardio-oncology are expanding rapidly and show promising ability to enhance cardiac risk prediction and imaging-based radiation dose optimization. However, current evidence is limited by methodological weaknesses, heterogeneity, insufficient validation, and incomplete reporting. Strengthening study design, transparency, and clinical integration will be critical for transforming AI from a research tool into a reliable component of personalized cardiovascular care for cancer survivors.

Abbreviations

3D — Three-dimensional

ACE — Acute coronary event

AF — Atrial fibrillation

AI — Artificial intelligence

AUC — Area under the curve (ROC-AUC)

CI — Confidence interval

CLAIM — Checklist for Artificial Intelligence in Medical Imaging

CNN — Convolutional Neural Network

CT — Computed Tomography

CCTA — Coronary computed tomography angiography

CTCAE — Common Terminology Criteria for Adverse Events

CVAE — Cardiovascular adverse event(s)

CVD — Cardiovascular disease

CVT — Cardiovascular toxicity

DL — Deep Learning

DSC — Dice similarity coefficient

EMBASE — Excerpta Medica database

Grad-CAM — Gradient-weighted Class Activation Mapping

H-L — Hosmer–Lemeshow (goodness-of-fit test)

ICC — Intraclass correlation coefficient

ID — Identifier

IQR — Interquartile range

k — Number of folds (k-fold cross-validation)

LAD — Left anterior descending (coronary artery)

MACE — Major adverse cardiac event(s)

MEDLINE — Medical Literature Analysis and Retrieval System Online

ML — Machine Learning

NSCLC — non-small cell lung cancer

NTCP — Normal tissue complication probability

OS — Overall survival

PCI — Percutaneous coronary intervention

PRISMA — Preferred Reporting Items for Systematic Reviews and Meta-Analyses

PROBAST — Prediction Model Risk of Bias Assessment Tool

PROBAST-AI — PROBAST extension for artificial intelligence

PROSPERO — International Prospective Register of Systematic Reviews

QoL — Quality of life

QUADAS-2 — Quality Assessment of Diagnostic Accuracy Studies, version 2

QUADAS-AI — QUADAS extension for artificial intelligence

ROC — Receiver operating characteristic

ROC-AUC — Area under the receiver operating characteristic curve

RT — Radiation therapy

SD — Standard deviation

SHAP — Shapley Additive Explanations

TRIPOD — Transparent Reporting of a multivariable prediction model for Individual Prognosis or Diagnosis

TRIPOD-AI — TRIPOD extension for artificial intelligence

TRIPOD+AI — TRIPOD+AI reporting guideline

U-Net — U-shaped convolutional neural network architecture

Declarations

Ethics approval and consent to participate

Ethical approval and consent to participate were not required for this study because it is a systematic review based exclusively on previously published data and does not involve individual patient data or direct human participation.

Consent for publication

Not applicable

Availability of data and materials

All data extracted from included studies, demographics and characteristics, data of AI models performance, PROBAST, QUADAS-2 risk-of-bias, CLAIM and TRIPOD+AI are included in Supplementary File S3. Original articles extracted from Rayyan AI tool are available from the corresponding author.

Competing interests

VS is a co-founder and equity holder (27%) in XURE.AI Inc., an early-stage artificial intelligence startup. The company has not provided financial support for this work and has no role in the design, execution or interpretation of this study. CMB has served on speaker bureaus for Bristol Myers Squibb and Pfizer, including disease state education related to ATTR cardiomyopathy. He has also received research support from Pfizer which is not related to this work. PMP has served as a consultant for Varian Medical Systems (speaker in brachytherapy-related activities), with salary support received, and no relation with this work. JR has served as a consultant and advisory board member for Elekta and Atrium OS. He also reports ownership of stock options in Aitrium OS. PL reports employment with West Virginia University Peer-Assisted Learning Service as Co-Leader and Tutor. No commercial relationships are reported. JAS reports employment with West Virginia University as co-leader and tutor for the peer-assisted learning service. No additional commercial or industry relationships are reported. **The remaining authors have nothing to disclose.**

Funding

No external fund supporting this proposed work.

Authors' contributions

Conceptualization: V.S., R.R.R., M.F.H and P.M.P.; methodology, data extraction, curation and data collection: V.S., B.M.G., N.D., T.N., J.A.S, P.M.L., W.C. and R.A.S; formal analysis, V.S., B.M.G., N.D., and T.N.; investigation, V.S., B.M.G.; resources, B.M.G.; and P.M.L.; writing original draft preparation, V.S., B.M.G., N.D., T.N; writing review and editing, J.R., R.A.S., C.M.B, G.G.S, A.E., R.R.R., D.A.C., M.F.H., P.M.P.; visualization, V.S., M.F.H., and B.M.G.; supervision, M.F.H. and P.M.P.; funding acquisition, V.S., D.A.C., A.E. and P.M.P. All authors have read and agreed to the published version of the manuscript.

Acknowledgements

We thank West Virginia University Cancer Institute for the Ignite Award for supporting VS in her projects.

References

1. Echefu G, Shah R, Sanchez Z, Rickards J, Brown SA. Artificial intelligence: Applications in cardio-oncology and potential impact on racial disparities. Am Heart J Plus. 2024;48:100479.
2. Lee VH, Yang L, Jiang Y, Kong FS. Radiation Therapy for Thoracic Malignancies. Hematol Oncol Clin North Am. 2020;34(1):109-25.
3. Cella L, Palma G. Radiation Therapy in Thoracic Tumors: Recent Trends and Current Issues. Cancers (Basel). 2022;14(11).
4. Pacelli R, Caroprese M, Palma G, Oliviero C, Clemente S, Cella L, Conson M. Technological evolution of radiation treatment: Implications for clinical applications. Semin Oncol. 2019;46(3):193-201.

5. Lenneman CG, Sawyer DB. Cardio-Oncology: An Update on Cardiotoxicity of Cancer-Related Treatment. Circ Res. 2016;118(6):1008-20.
6. Hufnagle JJ, Andersen SN, Maani EV. Radiation-Induced Cardiac Toxicity. StatPearls. Treasure Island (FL)2025.
7. Curigliano G, Cardinale D, Dent S, Criscitiello C, Aseyev O, Lenihan D, Cipolla CM. Cardiotoxicity of anticancer treatments: Epidemiology, detection, and management. CA Cancer J Clin. 2016;66(4):309-25.
8. Walls GM, Bergom C, Mitchell JD, Rentschler SL, Hugo GD, Samson PP, Robinson CG. Cardiotoxicity following thoracic radiotherapy for lung cancer. Br J Cancer. 2025;132(4):311-25.
9. Wang X, Palaskas NL, Yusuf SW, Abe JI, Lopez-Mattei J, Banchs J, Gladish GW, Lee P, Liao Z, Deswal A, Lin SH. Incidence and Onset of Severe Cardiac Events After Radiotherapy for Esophageal Cancer. J Thorac Oncol. 2020;15(10):1682-90.
10. van Nimwegen FA, Schaapveld M, Janus CP, Krol AD, Petersen EJ, Raemaekers JM, Kok WE, Aleman BM, van Leeuwen FE. Cardiovascular disease after Hodgkin lymphoma treatment: 40-year disease risk. JAMA Intern Med. 2015;175(6):1007-17.
11. Stelwagen J, Lubberts S, Steggink LC, Steursma G, Kruyt LM, Donkerbroek JW, van Roon AM, van Gessel AI, van de Zande SC, Meijer C, Grafin Zu Eulenburg CH, Oosting SF, Nuver J, Walenkamp AME, Jan de Jong I, Lefrandt JD, Gietema JA. Vascular aging in long-term survivors of testicular cancer more than 20 years after treatment with cisplatin-based chemotherapy. Br J Cancer. 2020;123(11):1599-607.
12. Shachar SS, Deal AM, Reeder-Hayes KE, Nyrop KA, Mitin N, Anders CK, Carey LA, Dees EC, Jolly TA, Kimmick GG, Karuturi MS, Reinbolt RE, Speca JC, Muss HB. Effects of Breast Cancer Adjuvant Chemotherapy Regimens on Expression of the Aging Biomarker, p16(INK4a). JNCI Cancer Spectr. 2020;4(6):pkaa082.
13. Gon Y, Zha L, Kawano T, Kudo H, Morishima T, Sasaki T, Miyashiro I, Mochizuki H, Sobue T. Specific Cardiovascular Mortality in Cancer Survivors: A Nationwide Population-Based Cohort Study in Japan. J Am Heart Assoc. 2025;14(10):e037965.
14. Zaorsky NG, Churilla TM, Egleston BL, Fisher SG, Ridge JA, Horwitz EM, Meyer JE. Causes of death among cancer patients. Ann Oncol. 2017;28(2):400-7.
15. Gon Y, Zha L, Sasaki T, Morishima T, Ohno Y, Mochizuki H, Sobue T, Miyashiro I. Heart Disease Mortality in Cancer Survivors: A Population-Based Study in Japan. J Am Heart Assoc. 2023;12(23):e029967.
16. Cheng CH, Shi SS. Artificial intelligence in cancer: applications, challenges, and future perspectives. Mol Cancer. 2025;24(1):274.
17. Salama V, Godinich B, Geng Y, Humbert-Vidan L, Maule L, Wahid KA, Naser MA, He R, Mohamed ASR, Fuller CD, Moreno AC. Artificial Intelligence and Machine Learning in Cancer Pain: A Systematic Review. J Pain Symptom Manage. 2024;68(6):e462-e90.

18. Salama V, Godinich BM, Lilly PM, Pifer PM, Duckworth AL, Hall SJ, Alabboodi M, Siochi RA, Clump DA, Emadi A. Artificial Intelligence and Machine Learning in Sexual Health and Dysfunction Across the Cancer Care Continuum: A Systematic Review. Cancers (Basel). 2025;17(18).
19. Putz F, Fietkau R. The increasing role of artificial intelligence in radiation oncology: how should we navigate it? Strahlenther Onkol. 2025;201(3):207-9.
20. Erdur AC, Rusche D, Scholz D, Kiechle J, Fischer S, Llorian-Salvador O, Buchner JA, Nguyen MQ, Etzel L, Weidner J, Metz MC, Wiestler B, Schnabel J, Rueckert D, Combs SE, Peeken JC. Deep learning for autosegmentation for radiotherapy treatment planning: State-of-the-art and novel perspectives. Strahlenther Onkol. 2025;201(3):236-54.
21. Rabe M, Kurz C, Thummerer A, Landry G. Artificial intelligence for treatment delivery: image-guided radiotherapy. Strahlenther Onkol. 2025;201(3):283-97.
22. Psoroulas S, Paunoiu A, Corradini S, Horner-Rieber J, Tanadini-Lang S. MR-linac: role of artificial intelligence and automation. Strahlenther Onkol. 2025;201(3):298-305.
23. Chen X, Mumme RP, Corrigan KL, Mukai-Sasaki Y, Koutroumpakis E, Palaskas NL, Nguyen CM, Zhao Y, Huang K, Yu C, Xu T, Daniel A, Balter PA, Zhang X, Niedzielski JS, Shete SS, Deswal A, Court LE, Liao Z, Yang J. Deep learning-based automatic segmentation of cardiac substructures for lung cancers. Radiother Oncol. 2024;191:110061.
24. Loap P, Monteil R, Kirova Y, Vu-Bezin J. Automated contouring for breast cancer radiotherapy in the isocentric lateral decubitus position: a neural network-based solution for enhanced precision and efficiency. Strahlenther Onkol. 2025;201(6):601-5.
25. Borges MG, Gruenwaldt J, Barsanelli DM, Ishikawa KE, Stuart SR. Automatic segmentation of cardiac structures can change the way we evaluate dose limits for radiotherapy in the left breast. J Med Imaging Radiat Sci. 2025;56(2):101844.
26. Zhou Y, Hou Y, Hussain M, Brown SA, Budd T, Tang WHW, Abraham J, Xu B, Shah C, Moudgil R, Popovic Z, Cho L, Kanj M, Watson C, Griffin B, Chung MK, Kapadia S, Svensson L, Collier P, Cheng F. Machine Learning-Based Risk Assessment for Cancer Therapy-Related Cardiac Dysfunction in 4300 Longitudinal Oncology Patients. J Am Heart Assoc. 2020;9(23):e019628.
27. Qiao EM, He J, Silos KD, Gasho JO, Belen P, Bitterman DS, McKenzie E, Steers J, Guthier C, Nohria A, Lu MT, Aerts H, Nikolova AP, Mak RH, Atkins KM. Baseline atrial volume indices and major adverse cardiac events following thoracic radiotherapy. Front Cardiovasc Med. 2025;12:1560922.
28. Bentriou M, Letort V, Chounta S, Fresneau B, Do D, Haddy N, Diallo I, Journy N, Zidane M, Charrier T, Aba N, Ducos C, Zossou VS, de Vathaire F, Allodji RS, Lemler S. Combining dosiomics and machine learning methods for predicting severe cardiac diseases in childhood cancer survivors: the French Childhood Cancer Survivor Study. Front Oncol. 2024;14:1241221.
29. Dincer S, Akmansu M, Akyol O. Machine learning modeling of cancer treatment-related cardiac events in breast cancer: utilizing dosiomics and radiomics. Front Oncol. 2025;15:1557382.
30. Moons KGM, Wolff RF, Riley RD, Whiting PF, Westwood M, Collins GS, Reitsma JB, Kleijnen J, Mallett S. PROBAST: A Tool to Assess Risk of Bias and Applicability of Prediction Model Studies:

Explanation and Elaboration. Ann Intern Med. 2019;170(1):W1-W33.

31. Wolff RF, Moons KGM, Riley RD, Whiting PF, Westwood M, Collins GS, Reitsma JB, Kleijnen J, Mallett S, Groupdagger P. PROBAST: A Tool to Assess the Risk of Bias and Applicability of Prediction Model Studies. Ann Intern Med. 2019;170(1):51-8.
32. Collins GS, Moons KGM, Dhiman P, Riley RD, Beam AL, Van Calster B, Ghassemi M, Liu X, Reitsma JB, van Smeden M, Boulesteix AL, Camaradou JC, Celi LA, Denaxas S, Denniston AK, Glocker B, Golub RM, Harvey H, Heinze G, Hoffman MM, Kengne AP, Lam E, Lee N, Loder EW, Maier-Hein L, Mateen BA, McCradden MD, Oakden-Rayner L, Ordish J, Parnell R, Rose S, Singh K, Wynants L, Logullo P. TRIPOD+AI statement: updated guidance for reporting clinical prediction models that use regression or machine learning methods. BMJ. 2024;385:e078378.
33. Collins GS, Dhiman P, Andaur Navarro CL, Ma J, Hooft L, Reitsma JB, Logullo P, Beam AL, Peng L, Van Calster B, van Smeden M, Riley RD, Moons KG. Protocol for development of a reporting guideline (TRIPOD-AI) and risk of bias tool (PROBAST-AI) for diagnostic and prognostic prediction model studies based on artificial intelligence. BMJ Open. 2021;11(7):e048008.
34. Collins GS, Reitsma JB, Altman DG, Moons KG. Transparent Reporting of a multivariable prediction model for Individual Prognosis Or Diagnosis (TRIPOD): the TRIPOD Statement. Br J Surg. 2015;102(3):148-58.
35. Tejani AS, Klontzas ME, Gatti AA, Mongan JT, Moy L, Park SH, Kahn CE, Jr., Panel CU. Checklist for Artificial Intelligence in Medical Imaging (CLAIM): 2024 Update. Radiol Artif Intell. 2024;6(4):e240300.
36. Whiting PF, Rutjes AW, Westwood ME, Mallett S, Deeks JJ, Reitsma JB, Leeflang MM, Sterne JA, Bossuyt PM, Group Q-. QUADAS-2: a revised tool for the quality assessment of diagnostic accuracy studies. Ann Intern Med. 2011;155(8):529-36.
37. Guni A, Sounderajah V, Whiting P, Bossuyt P, Darzi A, Ashrafian H. Revised Tool for the Quality Assessment of Diagnostic Accuracy Studies Using AI (QUADAS-AI): Protocol for a Qualitative Study. JMIR Res Protoc. 2024;13:e58202.
38. Gunturkun F, Akbilgic O, Davis RL, Armstrong GT, Howell RM, Jefferies JL, Ness KK, Karabayir I, Lucas JT, Jr., Srivastava DK, Hudson MM, Robison LL, Soliman EZ, Mulrooney DA. Artificial Intelligence-Assisted Prediction of Late-Onset Cardiomyopathy Among Childhood Cancer Survivors. JCO Clin Cancer Inform. 2021;5:459-68.
39. Ansari F, Neshasteh-Riz A, Paydar R, Mohagheghi F, Felegari S, Beigi M, Cheraghi S. Radiomics Analysis on Computed Tomography Images for Prediction of Chemoradiation-induced Heart Failure in Breast Cancer by Machine Learning Models. J Med Signals Sens. 2025;15:14.
40. Chounta S, Allodji R, Vakalopoulou M, Bentriou M, Do DT, De Vathaire F, Diallo I, Fresneau B, Charrier T, Zossou V, Christodoulidis S, Lemler S, Letort Le Chevalier V. Dosiomics-Based Prediction of Radiation-Induced Valvulopathy after Childhood Cancer. Cancers (Basel). 2023;15(12).
41. Choi BS, Yoo SK, Moon J, Chung SY, Oh J, Baek S, Kim Y, Chang JS, Kim H, Kim JS. Acute coronary event (ACE) prediction following breast radiotherapy by features extracted from 3D CT, dose, and

cardiac structures. Med Phys. 2023;50(10):6409-20.

42. Xie Y, Shen H, Xu Q, Tu C, Yang R, Liu T, Tang H, Miao Z, Zhang J. Evaluating coronary arteries and predicting MACEs using CCTA in lung cancer patients receiving chemotherapy or chemoradiotherapy. Radiother Oncol. 2024;200:110498.
43. Monlezun DJ. Percutaneous Coronary Intervention Mortality, Cost, Complications, and Disparities after Radiation Therapy: Artificial Intelligence-Augmented, Cost Effectiveness, and Computational Ethical Analysis. J Cardiovasc Dev Dis. 2023;10(11).
44. Ladbury C, Li RC, Danesharasteh A, Ertem Z, Tam A, Liu JS, Hao CR, Li RS, McGee H, Sampath S, Williams T, Glaser S, Khasawneh M, Liao ZX, Lee P, Ryckman J, Shaikh P, Amini A. Explainable Artificial Intelligence to Identify Dosimetric Predictors of Toxicity in Patients with Locally Advanced Non-Small Cell Lung Cancer: A Secondary Analysis of RTOG 0617. Int J Radiat Oncol. 2023;117(5):1287-96.
45. Bernasconi A, Zanga, A., Lucas, P.J.F., Scutari, M., Trama, A., Stella, F. A causal network model to estimate the cardiotoxic effect of oncological treatments in young breast cancer survivors. Progress in Artificial Intelligence. 2024.
46. Vendrame A CC, Chiovati P, Vinante L, Parvej M, Caroli A, Pirrone G, Barresi L, Drigo A, Avanzo M. Artificial Intelligence-Based Patient Selection for Deep Inspiration Breath-Hold Breast Radiotherapy from Respiratory Signals. Applied Sciences. 2023;13(8):4962.
47. Niedzielski JS, Wei X, Xu T, Gomez DR, Liao Z, Bankson JA, Lai SY, Court LE, Yang J. Development and application of an elastic net logistic regression model to investigate the impact of cardiac substructure dose on radiation-induced pericardial effusion in patients with NSCLC. Acta Oncol. 2020;59(10):1193-200.
48. Tohidinezhad F, Nurnberg L, Vaassen F, Ma Ter Bekke R, Jwl Aerts H, El Hendriks L, Dekker A, De Ruysscher D, Traverso A. Prediction of new-onset atrial fibrillation in patients with non-small cell lung cancer treated with curative-intent conventional radiotherapy. Radiother Oncol. 2024;201:110544.
49. Lee HI, Son J, Cho B, Goh Y, Jung J, Park JH, Chie EK, Kim KS, Kim YH, Kang HC, Yoon SM. Development and Validation of a Prediction Model for Cardiac Events in Patients With Hepatocellular Carcinoma Undergoing Stereotactic Body Radiation Therapy. Int J Radiat Oncol Biol Phys. 2025;122(5):1272-81.
50. Garrett Fernandes M, Bussink J, Stam B, Wijsman R, Schinagl DAX, Monshouwer R, Teuwen J. Deep learning model for automatic contouring of cardiovascular substructures on radiotherapy planning CT images: Dosimetric validation and reader study based clinical acceptability testing. Radiother Oncol. 2021;165:52-9.
51. Thor M, Scott JM, Reiner AS, Hong L, Yu AF, Apte A, McCormick B, Perez-Andujar A, Lee CP, Cervino L, Harrison J, Liang X, Patel P, Dauer LT, Moskowitz CS, Bernstein JL. Cardiopulmonary Substructure Doses are Not Correlated With Cardiorespiratory Fitness Among Breast Cancer Survivors Treated With Contemporary Radiation Therapy. Adv Radiat Oncol. 2024;9(12):101613.

52. Cabot JH, Ross EG. Evaluating prediction model performance. Surgery. 2023;174(3):723-6.
53. Huang Y, Jiang X, Gabriel RA, Ohno-Machado L. Calibrating predictive model estimates in a distributed network of patient data. J Biomed Inform. 2021;117:103758.
54. Arjmandi N, Mosleh-Shirazi MA, Mohebbi S, Nasseri S, Mehdizadeh A, Pishevar Z, Hosseini S, Tehranizadeh AA, Momennezhad M. Evaluating the dosimetric impact of deep-learning-based auto-segmentation in prostate cancer radiotherapy: Insights into real-world clinical implementation and inter-observer variability. J Appl Clin Med Phys. 2025;26(3):e14569.
55. Chun J, Chang JS, Oh C, Park I, Choi MS, Hong CS, Kim H, Yang G, Moon JY, Chung SY, Suh YJ, Kim JS. Synthetic contrast-enhanced computed tomography generation using a deep convolutional neural network for cardiac substructure delineation in breast cancer radiation therapy: a feasibility study. Radiat Oncol. 2022;17(1):83.
56. Chin V, Finnegan RN, Chlap P, Otton J, Haidar A, Holloway L, Thwaites DI, Dowling J, Delaney GP, Vinod SK. Validation of a Fully Automated Hybrid Deep Learning Cardiac Substructure Segmentation Tool for Contouring and Dose Evaluation in Lung Cancer Radiotherapy. Clin Oncol (R Coll Radiol). 2023;35(6):370-81.
57. Darby SC, Ewertz M, McGale P, Bennet AM, Blom-Goldman U, Bronnum D, Correa C, Cutter D, Gagliardi G, Gigante B, Jensen MB, Nisbet A, Peto R, Rahimi K, Taylor C, Hall P. Risk of ischemic heart disease in women after radiotherapy for breast cancer. N Engl J Med. 2013;368(11):987-98.
58. Taha MB. Radiation-Induced Coronary Artery Disease. Methodist Debakey Cardiovasc J. 2025;21(4):113-7.
59. Maraldo MV, Giusti F, Vogelius IR, Lundemann M, van der Kaaij MA, Ramadan S, Meulemans B, Henry-Amar M, Aleman BM, Raemaekers J, Meijnders P, Moser EC, Kluin-Nelemans HC, Feugier P, Casasnovas O, Fortpied C, Specht L, European Organisation for R, Treatment of Cancer Lymphoma G. Cardiovascular disease after treatment for Hodgkin's lymphoma: an analysis of nine collaborative EORTC-LYSA trials. Lancet Haematol. 2015;2(11):e492-502.
60. Wang Q, Xu L, Wang L, Yang X, Sun Y, Yang B, Greenwald SE. Automatic coronary artery segmentation of CCTA images using UNet with a local contextual transformer. Front Physiol. 2023;14:1138257.
61. Li Y, Wu Y, He J, Jiang W, Wang J, Peng Y, Jia Y, Xiong T, Jia K, Yi Z, Chen M. Automatic coronary artery segmentation and diagnosis of stenosis by deep learning based on computed tomographic coronary angiography. Eur Radiol. 2022;32(9):6037-45.
62. Xu R, Dai L, Wang J, Zhang L, Wang Y. SADiff: Coronary Artery Segmentation in CT Angiography Using Spatial Attention and Diffusion Model. J Imaging. 2025;11(6).
63. Mu N, Song, R., Li, X., Xu, Z., Jiang, J., Zhao, C. FAD-Net: Frequency-Domain Attention-Guided Diffusion Network for Coronary Artery Segmentation using Invasive Coronary Angiography. arXiv:250611454. 2025.
64. Huang H, Esposito, M., Zhao, C. Anatomy Guided Coronary Artery Segmentation from CCTA Using Spatial Frequency Joint Modeling. arXiv:251212539. 2025.

65. Yang X, Xu L, Yu S, Xia Q, Li H, Zhang S. Segmentation and Vascular Vectorization for Coronary Artery by Geometry-Based Cascaded Neural Network. IEEE Trans Med Imaging. 2025;44(1):259-69.

Figures

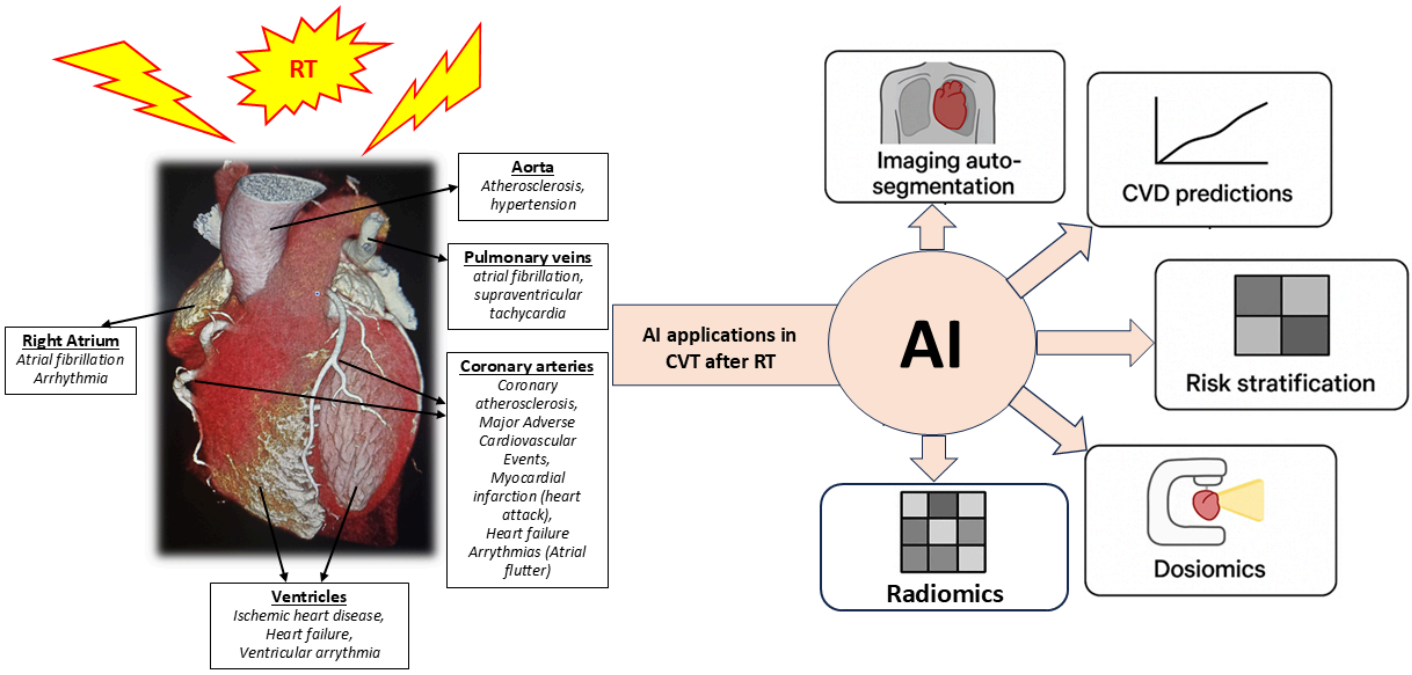


Figure 1

Illustration demonstrates post-radiation cardiovascular toxicity (CVT) and the artificial intelligence (AI) applications in the field of CVT after radiation therapy (RT).

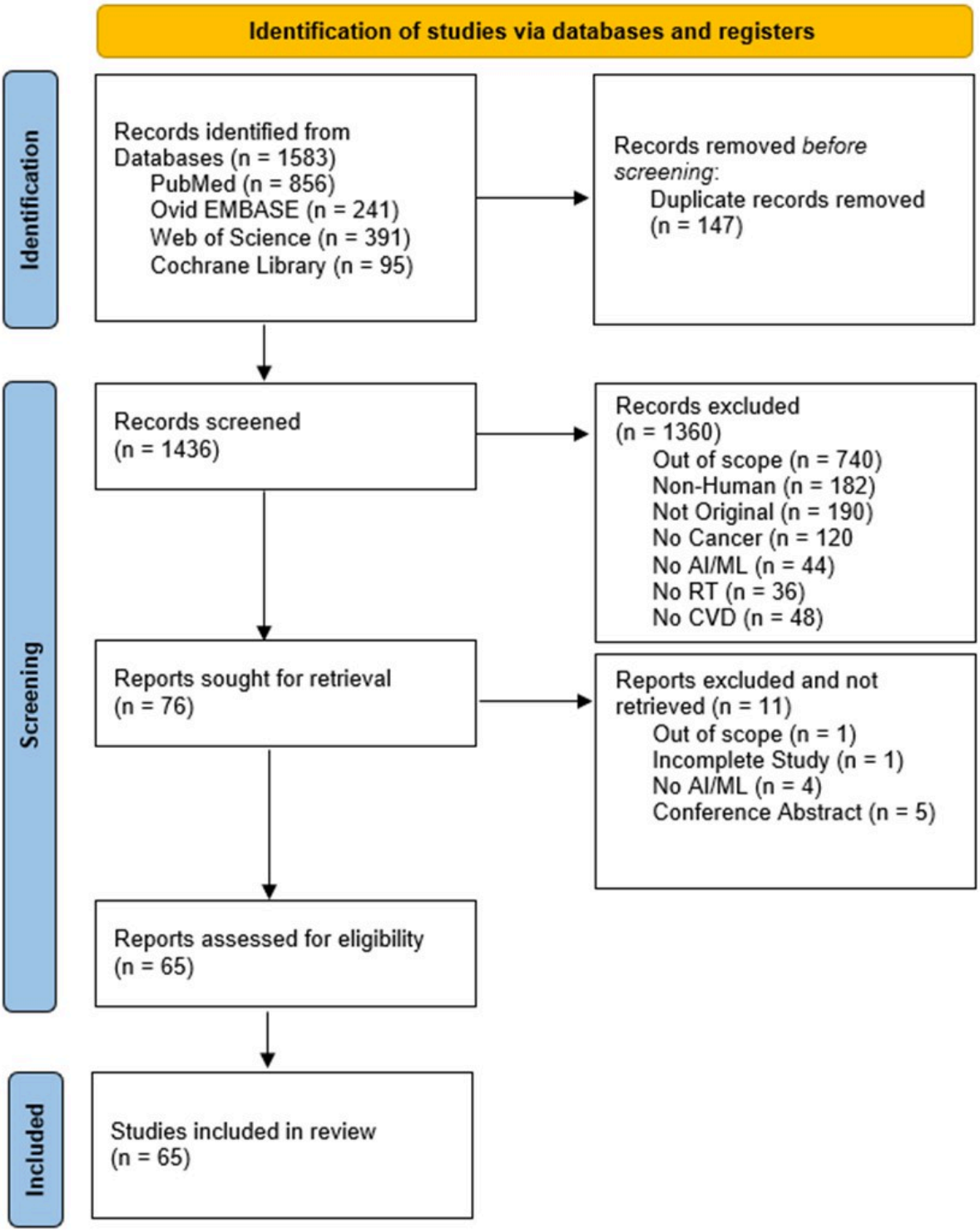


Figure 2

PRISMA flow diagram for systematic review of artificial intelligence in cardiovascular toxicity after radiation therapy.

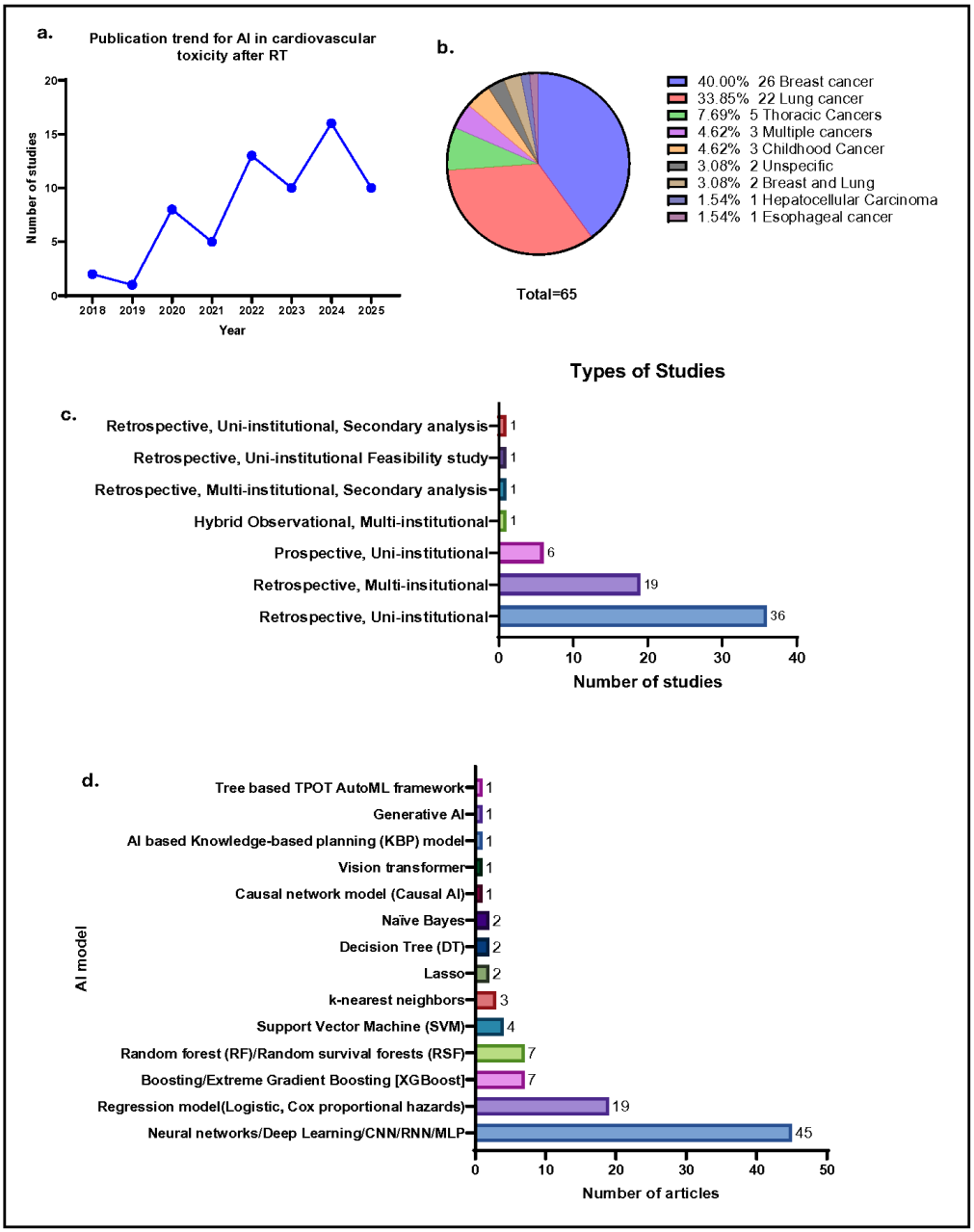


Figure 3

Publication characteristics of included articles**. a.** Publications trend for AI models applied in cardiovascular toxicity prediction or CV imaging analysis in patients with cancers who received RT. **b.** Frequency of the types of cancers studied in included articles. **c.** Types and numbers of studies across the included articles of this review. **d.** Types of AI or ML models and the number of studies investigated each model.

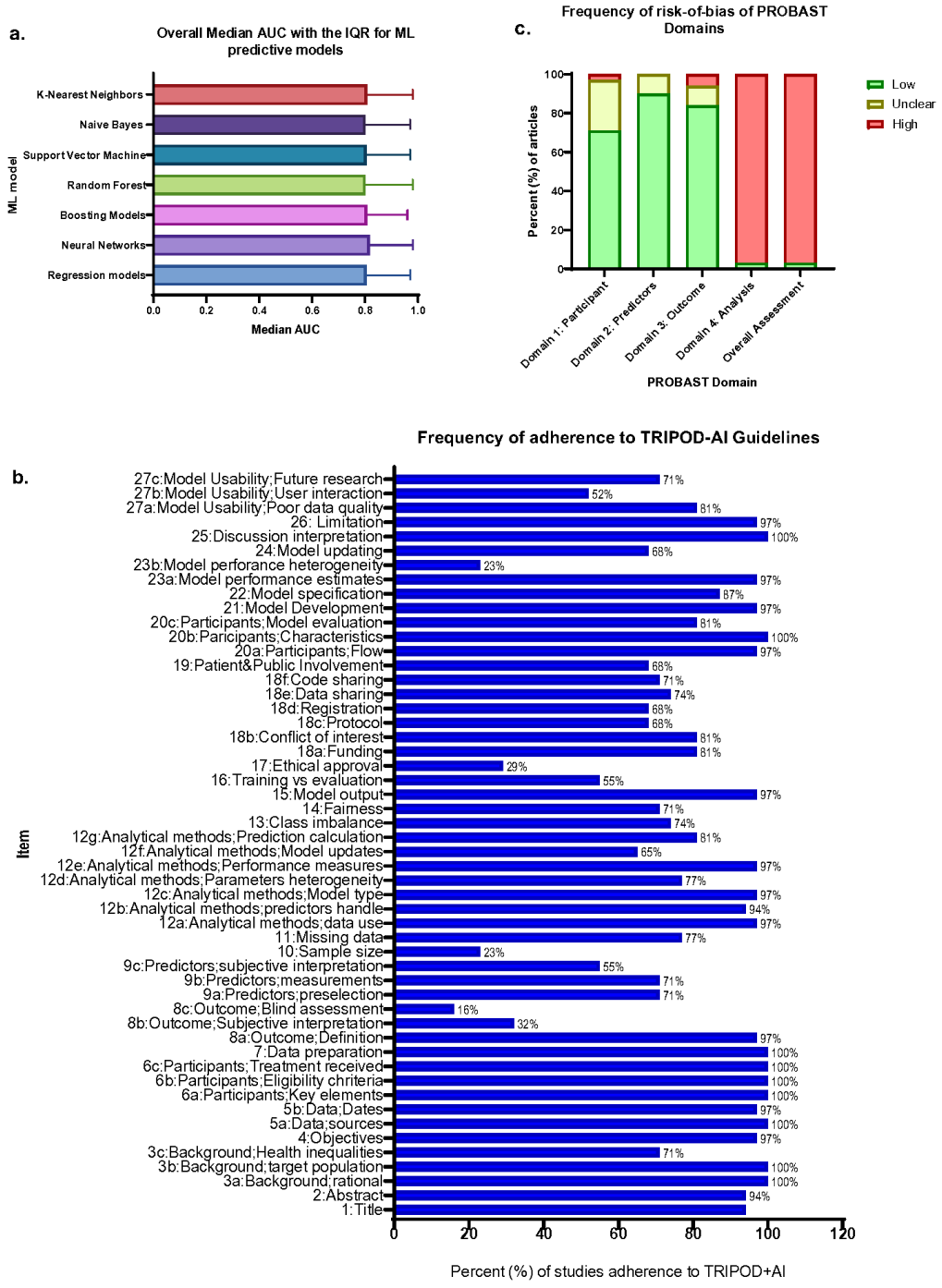


Figure 4

Characteristics, Quality and Risk-of bias analysis of included studies group applied AI in cardiovascular toxicity prediction and risk stratification in populations receiving RT. **a.**Median ROC-AUC and the Inter-Quantile Range (IQR) of AI models across the included predictive AI modeling studies. **b.** Frequency of adherence of studies applied predictive AI or ML models for cardiovascular toxicity prediction and risk stratification in populations receiving RT, to the AI specific Transparent Reporting of a multivariable

prediction Model for Individual Prognosis or Diagnosis (TRIPOD+AI) guidelines. **c.** Frequency of risk-of bias assessment of included studies applied predictive AI or ML models for cardiovascular toxicity prediction and risk stratification in populations receiving RT, using the Prediction Model Risk of Bias Assessment Tool (PROBAST).

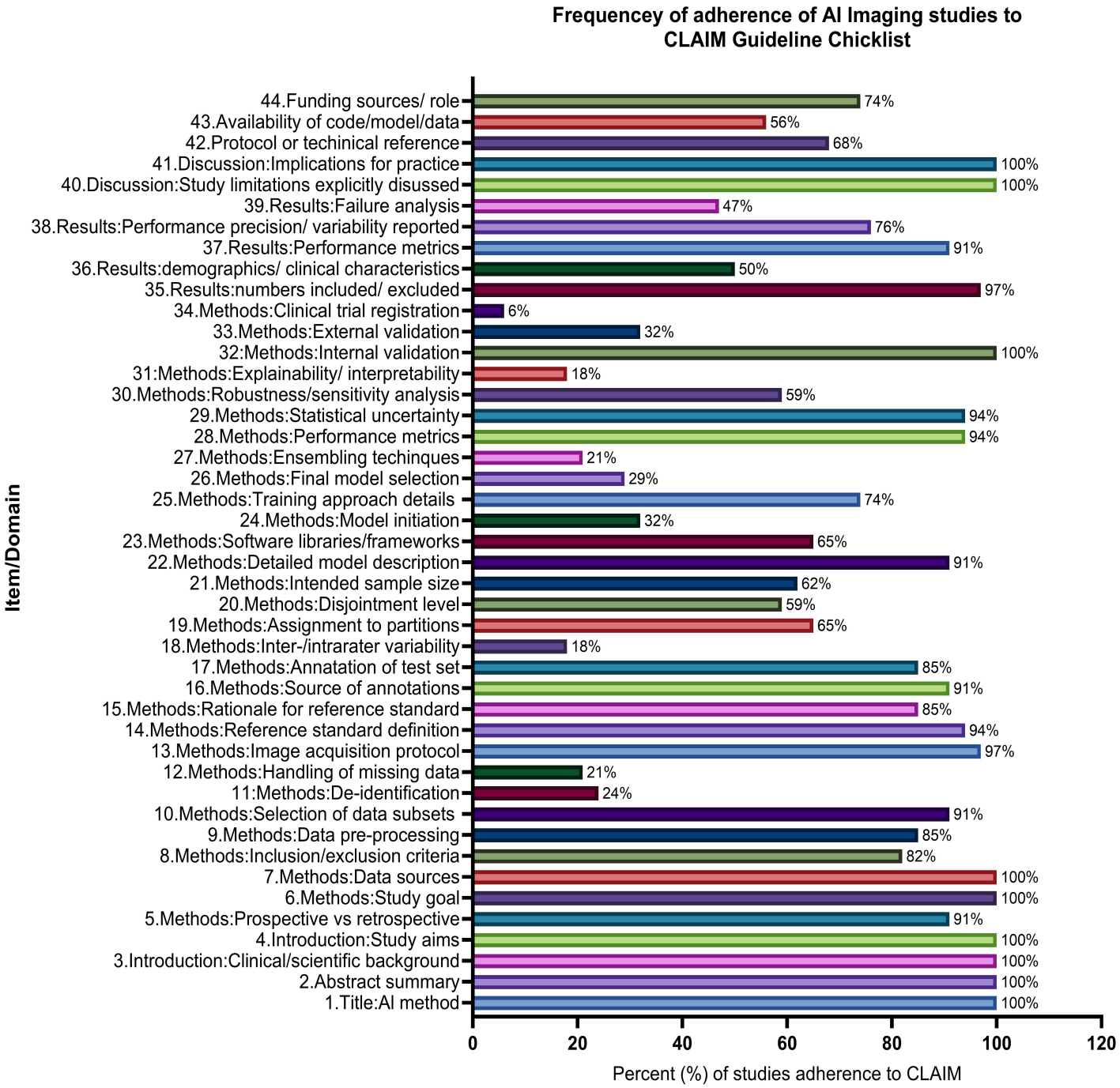


Figure 5

Frequency of adherence of studies applied AI or ML for cardiovascular imaging analysis for patients with cancer receiving RT, to the updated Checklist for Artificial Intelligence in Medical Imaging (CLAIM).

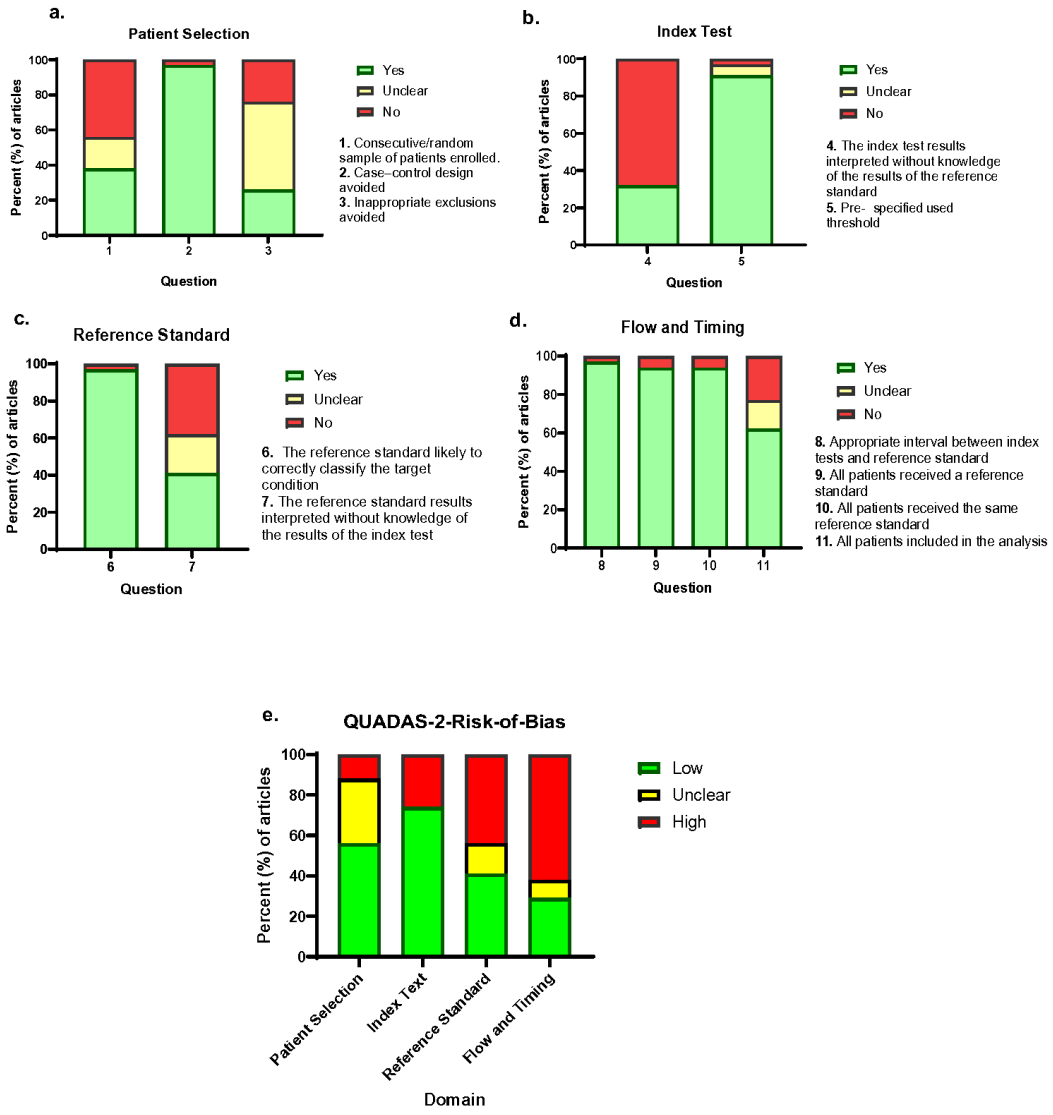


Figure 6

Frequency of Quality Assessment of Diagnostic Accuracy Studies (QUADAS-2) tool across eleven questions covering the four domains: patient selection, index test, reference standard, and flow and timing, across the 34 included AI-based cardiovascular imaging studies **a.** Frequency of answers/assessment of the three questions of the first domain; Patient Selection. **b.** Frequency of answers/assessment of the two questions of the second domain; Index Text. **c.**Frequency of

answers/assessment of the two questions of the third domain; Reference Standard. **d.** Frequency of answers/assessment of the four questions of the fourth domain; Flow and Timing. **e.** Frequency of overall risk-of-bias of the four domains of QUADAS-2 assessment tool.

Supplementary Files

This is a list of supplementary files associated with this preprint. Click to download.

- SupplementaryFile1SearchStrategyandResults.pdf
- SupplementaryFileS2PRISMAchecklistVSalama.pdf
- SupplementaryFileS3DataExtraction.xlsx
- SupplementaryTable1.docx