Natural Resources for Human Health
eISSN: 2583-1194

Original Research

Received 28 May 2026
Revised 30 June 2026
Accepted 15 July 2026

Natural Resources for Human Health 2026, Vol. 6 (12s): 884–897

KEYWORDS:
artificial intelligence; machine learning; chronic disease; diabetes mellitus; cardiovascular disease; cancer; predictive analytics; precision medicine; digital health; clinical decision support

Artificial Intelligence–Driven Prediction and Management of Chronic Diseases: A Comparative Analysis of Diabetes, Cardiovascular Disease, and Cancer

Reshmi Gopalakrishnan[1], Dr. G Venkataramana Sagar[2], Govind Asane[3], Dr. Priya Govindarajan[4], Dr. Jiten Mishra[5], Nityashree Mohapatra5, Sonali Rastogi6, Charanjeet Kaur7*

[1]Department of Microbiology, Sri Ramakrishna College of Arts and Science for Women, Coimbatore.
[2]G Pulla Reddy Engineering College (Autonomous), Kurnool 518007.
[3]Department of Pharmaceutics, SNJB Shriman Sureshdada Jain College of Pharmacy, Neminagar, Chandwad, Nashik, Maharashtra, India 423101.
[4]Department of Biochemistry, Mohamed Sathak College of Arts and Science, Affiliated to Madras University Chennai.
[5]Roland Institute of Pharmaceutical Sciences, Khodasingi, Berhampur, Ganjam, Odisha 760010.
[6] SDPC, SDGI Global University, NH-09, Delhi–Hapur Highway, Dasna, Ghaziabad, Uttar Pradesh, India.
[7]School of Allied Healthcare Sciences, Khalsa University, Amritsar. Email: brarchanni6@gmail.com
*Corresponding author: Dr. Charanjeet Kaur,
Email: brarchanni6@gmail.com

ABSTRACT: Artificial intelligence (AI) is increasingly used to convert heterogeneous clinical data into actionable predictions across chronic disease pathways, yet the maturity and clinical role of AI differ substantially among diabetes, cardiovascular disease (CVD), and cancer. This structured comparative evidence synthesis evaluated human studies and high-quality reviews addressing AI-enabled risk prediction, diagnosis, monitoring, complication forecasting, and management in these three disease domains. PubMed/MEDLINE, relevant publisher sources, and cross-referenced literature were examined up to September 2026. Animal experiments, in vivo studies, wet-laboratory-only investigations, and reports without a clinically relevant AI endpoint were excluded. Thirty-six core publications were retained for comparative synthesis. Diabetes showed its strongest implementation profile in continuous glucose monitoring, hypoglycemia forecasting, complication prediction, and automated insulin delivery; external validation remains a recurring weakness in complication models. CVD demonstrated broad AI utility across electrocardiography, imaging, wearable monitoring, and time-to-event risk prediction, with randomized evidence emerging for AI-supported echocardiographic workflows. Cancer showed the greatest dependence on high-dimensional imaging, digital pathology, and multi-omics integration; AI-assisted radiology and pathology can improve diagnostic performance, but generalizability and risk of bias remain major concerns. Across the three diseases, multimodal integration usually increased predictive capacity, whereas clinical translation was constrained by dataset shift, incomplete calibration, limited prospective validation, variable interpretability, privacy concerns, and under-

reporting of fairness. A five-domain evidence-maturity framework indicated relatively mature continuous-management applications in diabetes, signal- and imaging-based prediction in CVD, and imaging/pathology plus multi-omics applications in cancer. The comparative findings support disease-specific deployment rather than a single universal AI strategy. Clinically useful AI should be externally validated, calibrated, monitored after deployment, integrated with human decision-making, and reported using contemporary AI-specific standards.

1. INTRODUCTION

Chronic diseases account for a large proportion of long-term morbidity, healthcare utilization, and preventable mortality, but their clinical trajectories are rarely linear. Diabetes, cardiovascular disease (CVD), and cancer illustrate three distinct yet overlapping challenges: early identification of high-risk individuals, recognition of clinically important change, selection of appropriate interventions, and sustained management over time. Conventional statistical models remain essential, but they can be limited when clinical decisions depend on nonlinear interactions among laboratory values, longitudinal electronic health records (EHRs), physiological signals, imaging, pathology, treatment history, lifestyle data, and molecular profiles. Artificial intelligence (AI), particularly machine learning (ML) and deep learning (DL), provides a computational framework for extracting patterns from such heterogeneous data [1,2]. Its clinical value, however, is not defined simply by a high area under the receiver operating characteristic curve (AUROC). Models must also be calibrated, externally validated, interpretable enough for the intended task, equitable across relevant subgroups, and demonstrably useful within a clinical workflow [3-6].

The three disease domains considered in this study occupy different positions on the AI translation pathway. In diabetes, AI is closely linked to continuous glucose monitoring (CGM), automated insulin delivery, complication forecasting, and individualized phenotyping [7-16]. CVD provides abundant time-series and image data through electrocardiography (ECG), echocardiography, nuclear imaging, and wearable sensors, allowing AI to support both population-level risk assessment and point-of-care detection [17-26]. Cancer, in contrast, is characterized by extreme biological heterogeneity and increasingly multimodal data, including radiology, whole-slide pathology, genomics, transcriptomics, proteomics, and treatment-response information [27-36]. Consequently, similar algorithmic families can serve very different clinical purposes across these diseases.

A direct comparative analysis is therefore useful because it highlights not only where AI performs well, but also why the same methodological approach may translate differently among disease settings. The present article compares AI-driven prediction and management in diabetes, CVD, and cancer using a structured evidence-synthesis design. It focuses exclusively on human clinical and clinically oriented computational evidence and deliberately excludes in vivo experimentation. The objectives were to identify dominant data modalities and algorithms, summarize representative performance and clinical-use findings, compare the maturity of prediction and management applications, and define cross-cutting requirements for safe clinical implementation. A further aim was to distinguish technical model discrimination from evidence of clinical benefit, because an accurate retrospective classifier does not necessarily improve care. Particular attention was therefore given to external validation, calibration, human-AI collaboration, prospective workflow evidence, and the ability of an AI output to lead to a clinically meaningful action rather than merely reproduce an existing label.

2. MATERIALS AND METHODS

2.1 Study Design and Comparative Research Questions

This work was designed as a structured comparative evidence synthesis rather than a de novo patient-level clinical trial. The unit of analysis was the published human evidence describing AI or ML applications for prediction, diagnosis, monitoring, complication forecasting, treatment support, or workflow improvement in diabetes, CVD, and cancer. The design was selected because the three disease areas differ markedly in clinical endpoints, data modalities, and evaluation metrics, making a single pooled meta-analysis inappropriate. Instead, the study used a predefined comparative framework to map the maturity and translational relevance of AI applications across domains. The principal research questions were: (i) which data types and AI methods dominate each disease area; (ii) which clinical tasks have the strongest evidence for prediction or management; (iii) how frequently do the selected evidence sources address external validation, prospective testing, calibration, explainability, and real-world workflow integration; and (iv) what barriers are shared across all three disease areas. The analysis was informed by contemporary guidance emphasizing transparent description of prediction-model development and evaluation, particularly

TRIPOD+AI [3], and by guidance for reporting AI interventions in clinical trials [4]. The review was not intended to rank individual commercial products or produce a universal performance score. Disease-level comparison focused on evidence patterns rather than direct statistical superiority because AUROC, sensitivity, specificity, concordance index, time-in-range, workflow time, and clinical event outcomes are not interchangeable. Only clinically oriented human evidence was considered. Animal studies, in vivo disease models, cell-line-only experiments, and wet-laboratory mechanistic studies were outside the scope. This restriction ensured that all conclusions concerned patient-facing or health-system-facing AI applications. The final analytic corpus contained 36 core references, selected to provide contemporary, clinically relevant, and methodologically diverse coverage across the three disease domains and cross-cutting AI methodology. The comparison was conducted at the level of clinical use cases, not individual software products. This approach reduced the risk of treating transient commercial implementations as durable scientific evidence and allowed established and emerging applications to be assessed using the same translational questions.

2.2 Literature Search Strategy and Source Identification

Literature identification was performed primarily through PubMed/MEDLINE, supplemented by publisher-hosted guideline articles and backward citation tracing from major systematic reviews. Searches were updated through September 2026. Search concepts combined disease terms with AI methodology and clinical task terms. Diabetes searches included combinations of "diabetes," "prediabetes," "diabetic complications," "continuous glucose monitoring," "hypoglycemia," "closed-loop," "artificial intelligence," "machine learning," "deep learning," and "prediction." CVD searches included "cardiovascular disease," "coronary artery disease," "atrial fibrillation," "heart failure," "ECG," "echocardiography," "wearable," "risk prediction," and AI/ML terms. Cancer searches combined "cancer," "neoplasm," "screening," "radiology," "digital pathology," "whole-slide image," "multi-omics," "prognosis," and AI/ML terms. Methodological searches targeted reporting, explainability, and clinical deployment. Priority was given to systematic reviews, meta-analyses, externally validated prediction studies, randomized evaluations, and large multicenter investigations, while landmark earlier studies were retained when they established a clinically important AI application. Recent evidence from 2024-2026 was deliberately emphasized to reflect current practice. The search was not represented as a fully exhaustive systematic review across all bibliographic databases; therefore, no fabricated screening counts or PRISMA numerators were used. Instead, source identification was documented as a reproducible conceptual workflow (Figure 1). Publications were screened for relevance to the predefined clinical tasks and for sufficient methodological information to support comparative extraction. When multiple reviews covered similar applications, preference was given to the most recent review with transparent methods or to a study providing clinically interpretable performance, external validation, or prospective evaluation. The final reference set was frozen before synthesis to avoid selectively adding citations solely to support a desired conclusion. Search terms were iteratively expanded when a major review revealed a clinically important modality or endpoint not captured by the initial concepts. Title/abstract relevance was judged against the prespecified disease-task framework, and full-text information was used when performance, validation strategy, or population characteristics could not be established from the abstract. Duplicate conceptual evidence was minimized by prioritizing the most informative source.

2.3 Eligibility Criteria, Exclusions, and Data Extraction

Eligible publications had to address human health data or clinically deployable workflows in diabetes, CVD, or cancer and explicitly use AI, ML, DL, or closely related computational prediction methods. Eligible study types included retrospective or prospective clinical cohorts, diagnostic-accuracy studies, randomized or controlled workflow evaluations, systematic reviews, meta-analyses, and structured evidence reviews. Studies were considered relevant when the AI output informed at least one of the following: disease risk, early detection, diagnosis, prognosis, complication risk, physiologic forecasting, treatment selection, treatment adjustment, monitoring, or operational clinical workflow. Exclusion criteria were animal or in vivo experiments, cell or tissue experiments without patient-level clinical modeling, purely technical image-processing papers without a clinical endpoint, non-peer-reviewed promotional material, and AI studies unrelated to chronic-disease prediction or management. For each retained source, data were extracted into a structured evidence matrix containing disease domain, clinical task, primary data modality, algorithm family, study design, sample characteristics when available, principal performance metric, validation approach, and stated translational limitations. Additional fields captured whether the study addressed calibration, explainability, fairness/subgroup performance, prospective evaluation, and workflow integration. Because the included literature reported heterogeneous outcomes, extraction preserved the metric used by each study rather than converting all findings into a common numerical effect. Representative quantitative findings were included in the Results only when directly reported by the source. For example, sensitivity and specificity from diagnostic meta-analyses were not combined with AUROC from risk models.

Similarly, glycemic time-in-range outcomes from automated insulin-delivery studies were treated as management outcomes rather than diagnostic-performance measures. This approach prevented inappropriate cross-disease pooling and allowed the analysis to distinguish model discrimination from actual clinical utility. Extraction also recorded the comparator used by each study, such as clinician assessment, conventional regression, usual care, or a less automated treatment system. When a source reported several models, emphasis was placed on externally validated or clinically integrated results rather than the best internal-development estimate.

2.4 AI Taxonomy and Classification of Clinical Tasks

AI methods were grouped into clinically interpretable families rather than classified solely by software architecture. Classical ML included logistic-model extensions, random forests, gradient boosting, support vector machines, survival forests, and other algorithms relying primarily on structured features. DL included convolutional neural networks for imaging and signal analysis, recurrent or temporal networks, deep survival methods, and end-to-end representation learning. Multimodal AI referred to models combining two or more distinct data types, such as pathology plus genomics, imaging plus clinical variables, or physiological signals plus contextual data. Explainable AI (XAI) was recorded when a study explicitly used methods such as feature importance, SHAP-like attribution, saliency mapping, or other mechanisms designed to expose model reasoning; however, the presence of an explanation method was not assumed to guarantee clinical interpretability [5,6]. Clinical tasks were classified into four levels. Level 1 was risk prediction before overt disease or before a major complication. Level 2 was detection or diagnosis from current data. Level 3 was prognosis or prediction of future progression, recurrence, adverse events, or treatment response. Level 4 was management, defined as repeated or real-time use of AI to alter monitoring, treatment delivery, follow-up, or clinical workflow. This classification allowed meaningful comparison between, for example, CGM-based hypoglycemia forecasting in diabetes, AI-ECG risk stratification in CVD, and multimodal survival prediction in cancer. Data modalities were separately categorized as structured clinical/EHR data, physiological time-series and wearables, radiology, digital pathology, retinal imaging, or molecular/multi-omics. This taxonomy was also used to construct Figure 3, which illustrates the shared architecture from input data through AI processing to clinically actionable outputs. Hybrid systems that combined algorithmic prediction with protocolized human review were classified according to the clinical action actually supported. This distinction was important because the same neural-network architecture may function as a screening aid, a prognostic tool, or a management component depending on when it is used, what data are available, and who acts on the output.

2.5 Comparative Evaluation Framework and Evidence-Maturity Scoring

The comparative framework evaluated five dimensions: risk prediction, continuous monitoring or management, imaging/biosignal AI, multimodal or omics integration, and prospective clinical integration. Each disease domain received an ordinal evidence-maturity score from 1 to 5 for each dimension. The score was not a statistical effect size and was not intended to imply that one disease is intrinsically more suitable for AI. A score of 1 represented an early or highly experimental evidence base; 2 indicated repeated proof-of-concept evidence with limited validation; 3 indicated multiple clinical studies but important generalizability or implementation gaps; 4 indicated a comparatively mature evidence base with external or prospective support in at least part of the field; and 5 indicated broad, repeated evidence with strong clinical relevance, including external validation, large-scale studies, or real-world deployment evidence. Scores were assigned by triangulating the retained reviews and representative studies, with emphasis on validation and clinical integration rather than headline accuracy alone. For example, diabetes received a high management score because CGM forecasting and automated insulin delivery have direct repeated-use clinical applications [8,11-13], whereas cancer received a high multimodal score because pathology, radiology, and multi-omics integration are central to current precision-oncology research [33-36]. CVD received high scores for risk prediction and imaging/biosignal AI because ECG, echocardiography, nuclear imaging, and wearables have been evaluated across large cohorts and prospective workflows [19-26]. The resulting matrix is displayed in Figure 2. To maintain transparency, the article explicitly separates these structured synthesis scores from source-reported quantitative performance values. No inferential statistics were applied to the ordinal maturity matrix. To limit overinterpretation, a high score required evidence breadth plus translational support; a single study with exceptional accuracy could not by itself justify a score of 5. Conversely, an established clinical workflow could receive a relatively high score even when its numerical performance metric differed from another disease domain, because maturity reflected validation, actionability, and deployment evidence rather than a common accuracy threshold.

2.6 Quality, Bias, Reproducibility, and Ethical Assessment

Quality appraisal was integrated into the comparative synthesis through recurring methodological domains rather than by assigning a single numerical quality score to heterogeneous study types. The domains were external validation, calibration, risk of bias, dataset representativeness, fairness or subgroup analysis, interpretability, prospective evaluation, and evidence of workflow or patient-level benefit. TRIPOD+AI provides current reporting guidance for prediction models and emphasizes complete reporting, fairness considerations, open science, and subgroup evaluation [3]. CONSORT-AI similarly highlights the need to describe the AI intervention, its outputs, interaction with human decision-making, and performance errors in trial settings [4]. These principles were used as interpretive anchors. Particular caution was applied to internally validated models, studies using convenience samples, single-center imaging datasets, and high-dimensional omics models without independent testing. Explainability was considered supportive but not sufficient evidence of trustworthiness because post-hoc explanations can be unstable or misleading [5]. Privacy and data governance were considered especially important for longitudinal EHR, wearable, imaging, pathology, and genomic data. The synthesis also distinguished discrimination from calibration: a model can rank risk well while systematically over- or under-estimating absolute risk. This distinction is particularly important when AI outputs trigger screening, preventive therapy, invasive testing, or treatment modification. For management applications, the highest evidentiary weight was given to prospective or randomized evaluations that measured workflow, physiological control, or patient-important outcomes. No ethics approval was required for this article because it analyzed published literature and did not access identifiable patient-level data. Reproducibility was judged more favorably when studies described data partitions, preprocessing, model evaluation, and independent testing clearly enough to permit replication. Ethical interpretation also considered automation bias and the possibility that false reassurance or excessive alerts could alter clinician behavior. These concerns were treated as deployment issues rather than abstract algorithmic limitations.

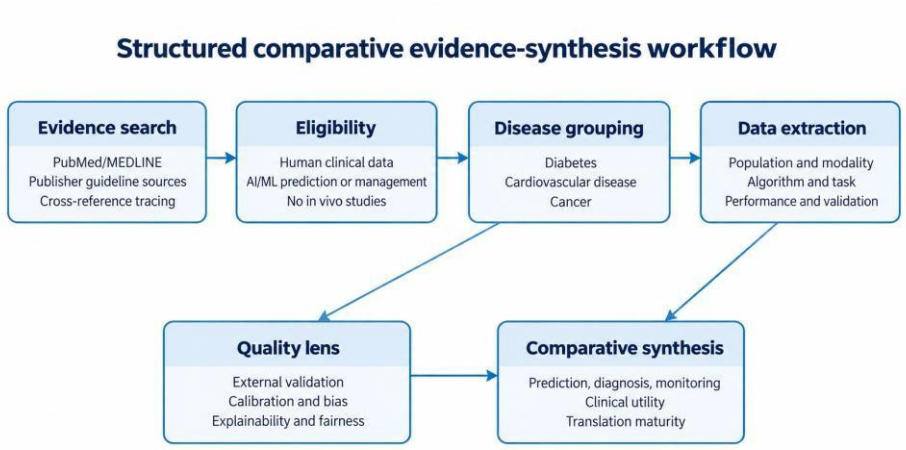


Figure 1. Structured workflow used to identify, classify, and compare clinically oriented AI evidence across diabetes, cardiovascular disease, and cancer. No in vivo or animal studies were included.

3. RESULTS

3.1 Cross-Disease Evidence Landscape

The retained literature showed a common movement from single-source prediction toward multimodal decision support, but the dominant inputs differed by disease. Diabetes research was anchored in structured clinical data, CGM time-series, retinal imaging, and automated insulin-delivery systems [7-16]. CVD research used structured risk factors together with ECG, echocardiography, SPECT, and consumer or medical-grade wearables [17-26]. Cancer research relied most heavily on radiology, digital pathology, and increasingly multi-omics data, with multimodal models attempting to combine molecular and morphological information [27-36]. Across all three domains, the strongest recent studies increasingly separated model development from independent evaluation, but external validation, calibration, and prospective clinical-utility testing remained inconsistent. Table 1 summarizes the dominant patterns.

Natural Resources for Human Health
eISSN: 2583-1194

Table 1. Comparative overview of AI applications across the three chronic disease domains.

Domain	Dominant data modalities	High-value AI tasks	Representative algorithm families	Current translational constraint
Diabetes	EHR/labs; CGM; retinal images; treatment history	Risk prediction; hypoglycemia forecasting; complication prediction; automated insulin delivery	Gradient boosting; random forest; temporal DL; control algorithms	Generalizability of complication models; false alarms; device interoperability
Cardiovascular disease	EHR; ECG; echocardiography; SPECT/CT; wearables	ASCVD risk; AF detection; LV dysfunction; CAD stratification; workflow support	Survival ML; CNN/DL; ensemble ECG models; image AI	Calibration and external validation; workflow integration; subgroup fairness
Cancer	Radiology; whole-slide pathology; clinical data; genomics/multi-omics	Screening; diagnosis; malignancy classification; prognosis; treatment-response stratification	CNNs; multiple-instance learning; multimodal fusion; survival DL	Dataset shift; high risk of bias; dependence on curated datasets; limited prospective utility studies

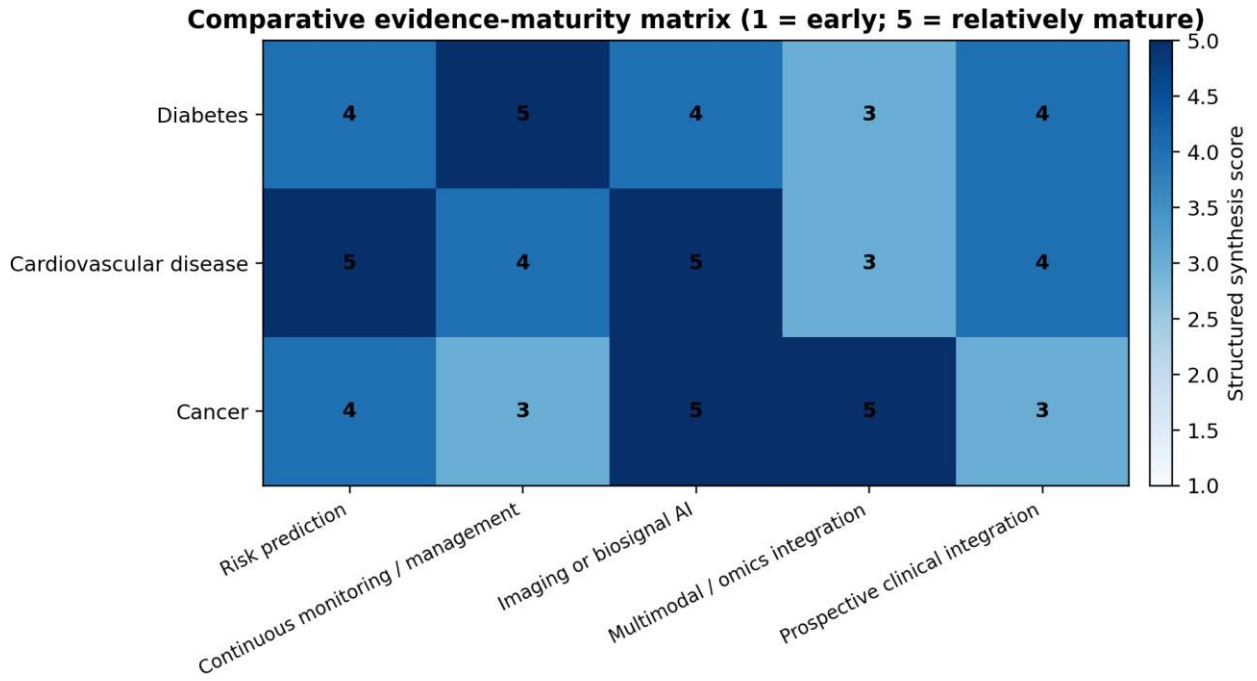


Figure 2. Comparative evidence-maturity matrix derived from the structured synthesis. Scores (1-5) summarize breadth of evidence, validation, prospective evaluation, and implementation maturity; they are not pooled effect sizes.

3.2 Diabetes: Prediction, Continuous Monitoring, and Adaptive Management

Diabetes had the clearest continuum from prediction to repeated management. A 2025 systematic review of 53 studies confirmed broad use of ML for diabetes prediction, while also highlighting variability in datasets, algorithms, training procedures, and evaluation [7]. For downstream complications, Kanbour et al. found that models predicting diabetic microvascular complications achieved a mean c-statistic of 0.79 on internal validation but 0.72 on external validation, illustrating

the typical performance drop when models move outside the development environment [9]. Diabetic kidney disease has been especially active, with systematic evidence supporting non-linear ML approaches while still emphasizing heterogeneity and the need for more rigorous external testing [10]. Retinal imaging extends the role of AI beyond retinopathy: a systematic review of AI applied to retinal images reported AUROCs ranging approximately from 0.676 to 0.971 for different diabetes-associated systemic outcomes and highlighted the shortage of longitudinal and randomized translational studies [14].

Management-oriented AI is more mature because the output can be linked directly to a repeated physiological signal. Reviews of AI-supported glycemic control describe predictive alarms, insulin-dosing support, and closed-loop control as core applications [8]. In a meta-analysis of AI-driven closed-loop systems involving 1,156 participants, closed-loop use reduced time outside the target glucose range relative to control, supporting a clinically meaningful role for algorithm-guided insulin delivery [12]. A network meta-analysis of 28 randomized trials of hybrid closed-loop systems likewise found significant improvement in time in range compared with less automated insulin strategies [11]. For short-horizon safety, a 2026 meta-analysis of CGM-driven ML models for hypoglycemia reported pooled sensitivity of 80% and specificity of 89%; adding contextual inputs such as insulin, carbohydrate intake, and physical activity could improve prediction, but false-positive alerts remain a practical issue [13]. AI can also support disease heterogeneity: a random-forest model for type 2 diabetes subtyping retained 86.3% external validation accuracy and preserved clinically meaningful differences in long-term complication risk [16]. These results position diabetes as a domain where AI can move beyond static prediction toward adaptive management, provided that models remain calibrated and device outputs are clinically interpretable.

Table 2. Representative AI evidence in diabetes.

Clinical application	Representative evidence	Key result	Interpretation
General diabetes prediction	Khokhar et al. [7]	53 studies reviewed	Strong research activity; heterogeneous datasets and evaluation methods
Microvascular complication prediction	Kanbour et al. [9]	Mean c-statistic 0.79 internal vs 0.72 external	External validation reveals meaningful performance loss
AI retinal screening for systemic complications	Yang et al. [14]	Reported AUROC range ~0.676-0.971 across outcomes	Retinal AI may support opportunistic systemic risk assessment
AI-driven closed-loop management	Wang et al. [12]	Reduced time outside target range vs controls	Demonstrates direct therapeutic-management role
CGM hypoglycemia forecasting	Quader et al. [13]	Sensitivity 80%; specificity 89%	Useful early-warning potential with false-positive trade-offs
Type 2 diabetes subtyping	Tanabe et al. [16]	86.3% accuracy in external cohort	Supports personalized phenotyping using routinely available variables

3.3 Cardiovascular Disease: Multimodal Risk Stratification and Signal-Based AI

CVD demonstrated the broadest use of routinely generated biosignals. Contemporary reviews describe AI as an extension of traditional risk estimation, especially when EHR variables are combined with ECG, imaging, genetics, or wearable data [17,18,23]. A 2024 systematic review of 33 time-to-event AI studies found frequent use of random survival forests, survival gradient boosting, penalized Cox models, and DeepSurv, but noted limited attention to social determinants and sex-stratified analysis [18]. A 2026 scoping review similarly found that only 7 of 30 prognostic CVD studies reported external validation and none reported calibration, showing that high discrimination alone has not solved the implementation problem [23].

AI-ECG illustrates the ability of DL to extract latent cardiovascular information from inexpensive tests. Attia et al. showed that an AI-enabled ECG could identify patients with atrial fibrillation despite sinus rhythm at the time of the tracing [19], and a related deep-learning ECG model identified left ventricular systolic dysfunction [20]. More recently, Awasthi et al. developed

large-scale ECG-AI models using data from more than seven million patients; models achieved AUROCs of 0.88 for high coronary calcium, 0.85 for obstructive CAD, and 0.94 for regional left ventricular akinesis, and the ensemble added risk information beyond pooled-cohort equations [21]. Imaging-based AI also shows promise: a meta-analysis of ML applied to SPECT in coronary disease found improved prognostic prediction in several studies, particularly when demographic data were combined with imaging [22]. Wearable AF detection is one of the most mature consumer-facing applications; a 2025 diagnostic meta-analysis of 26 studies and 17,349 patients reported overall sensitivity of 95%, specificity of 97%, and pooled AUC of 0.97 [25]. Importantly, clinical workflow evidence is emerging. In a blinded randomized trial, AI-based initial LVEF assessment required substantial correction in 16.8% of studies versus 27.2% after sonographer initial assessment, supporting non-inferiority and superiority for the primary workflow endpoint [26]. A systematic review of cardiovascular AI randomized trials nevertheless identified only 11 RCTs, confirming that prospective outcome evidence remains limited relative to algorithm-development literature [24].

Table 3. Representative AI evidence in cardiovascular disease.

Clinical application	Representative evidence	Key result	Clinical significance
Time-to-event CVD prediction	Teshale et al. [18]	33 studies; RSF/DeepSurv commonly used	Survival-aware AI can model censored outcomes but needs stronger subgroup analysis
AI-ECG for CAD/ASCVD risk	Awasthi et al. [21]	AUROC 0.88 CAC; 0.85 obstructive CAD; 0.94 akinesis	Low-cost ECG can contain latent structural and coronary risk information
SPECT prognosis	Cicek et al. [22]	12 retrospective studies; 73,023 participants	Imaging plus demographics may improve prognostic assessment
Smartwatch AF detection	Barrera et al. [25]	Sensitivity 95%; specificity 97%; AUC 0.97	High diagnostic potential for scalable rhythm monitoring
AI echocardiography workflow	He et al. [26]	Substantial LVEF correction 16.8% AI vs 27.2% sonographer	Prospective evidence that AI can improve measurement workflow
Randomized cardiovascular AI evidence	Barzilai et al. [24]	11 RCTs identified	Clinical trial evidence is growing but remains small relative to model literature

3.4 Cancer: Imaging, Digital Pathology, and Multi-Omics Precision Models

Cancer differed from diabetes and CVD because its AI evidence is dominated by high-dimensional diagnostic data rather than continuous physiological monitoring. A comprehensive 2025 review across 19 cancer types found growing use of AI for susceptibility and risk prediction, particularly where complex multidimensional variables exceed the practical flexibility of conventional models [27]. In clinical imaging, the most persuasive signal is not autonomous AI replacing clinicians but AI assisting them. A 2025 systematic review of multi-reader, multi-case cancer radiology studies reported pooled sensitivity and specificity of 0.67 and 0.82 for clinicians alone compared with 0.79 and 0.87 when clinicians used AI assistance [28]. This pattern supports augmentation rather than substitution. Lung cancer provides a well-studied screening example: an earlier meta-analysis of 26 studies and 150,721 imaging instances reported pooled sensitivity of 94.6% and specificity of 93.6% for AI-based imaging, but rated the overall certainty of evidence as very low [31]. More recent evidence comparing externally validated lung-cancer risk models found pooled AUC 0.82 for AI-based models versus 0.73 for traditional regression, while emphasizing high risk of bias in both model families [32].

Digital pathology is similarly promising but methodologically vulnerable. A systematic review of whole-slide image studies described AI as a potentially valuable diagnostic support tool while highlighting the complexity of whole-slide processing and real-world integration [29]. A 2024 meta-analysis across digital pathology applications reported mean sensitivity of 96.3% and

specificity of 93.3%, yet 99% of included studies had at least one high or unclear risk-of-bias or applicability domain [30]. Beyond images, precision oncology is increasingly multimodal. Reviews of multi-omics ML describe integration of genomics, transcriptomics, proteomics, metabolomics, and clinical data for patient stratification and treatment-oriented research [33,34]. In a 2026 systematic review of 48 multimodal cancer-prognosis studies combining pathology and high-throughput omics, reported concordance indices ranged from 0.550 to 0.857; most multimodal models outperformed unimodal comparators, but all studies had high or unclear risk of bias and all relied on The Cancer Genome Atlas, limiting claims of generalizability [35]. Current cancer AI therefore has substantial diagnostic and prognostic potential but requires more independent, multi-institutional prospective validation before routine autonomous use [36].

Table 4. Representative AI evidence in cancer.

Clinical application	Representative evidence	Key result	Clinical interpretation
Cross-cancer risk prediction	Felici et al. [27]	AI reviewed across 19 cancer types	Potential advantage for nonlinear, multidimensional risk profiles
AI-assisted radiology	Zhao et al. [28]	Clinician sensitivity 0.67→0.79; specificity 0.82→0.87 with AI	Best-supported role is clinician augmentation
Digital pathology	McGenity et al. [30]	Mean sensitivity 96.3%; specificity 93.3%	High reported accuracy but widespread risk-of-bias concerns
Lung cancer screening imaging	Thong et al. [31]	Sensitivity 94.6%; specificity 93.6%	Promising screening support; certainty of evidence very low
Lung cancer risk models	Leonard et al. [32]	Externally validated pooled AUC 0.82 AI vs 0.73 regression	AI may improve discrimination, especially with LDCT features
Multimodal prognosis	Jennings et al. [35]	C-index range 0.550-0.857; multimodal often better	Multi-omics/pathology fusion is promising but highly dataset dependent

3.5 Comparative Synthesis of Prediction and Management Maturity

The structured evidence-maturity matrix (Figure 2) indicated that the most mature AI use case was not identical across diseases. Diabetes scored highest for continuous monitoring and management because CGM and closed-loop insulin delivery create a feedback loop in which predictions can be translated directly into treatment adjustment [8,11-13]. CVD scored highest for risk prediction and imaging/biosignal AI because ECG, echocardiography, nuclear imaging, and wearables generate large standardized data streams with clinically established endpoints [19-26]. Cancer scored highest for imaging/pathology and multimodal/omics integration because clinical interpretation increasingly depends on complex image features and molecular heterogeneity [28-36]. The architecture underlying these applications is nevertheless similar: clinical data are transformed through feature extraction or learned representations, combined when appropriate, and converted into a risk, detection, forecast, or management output (Figure 3).

Across all domains, the gap between technical performance and deployable clinical utility was explained by five recurring factors. First, external validation commonly reduced performance or remained absent. Second, calibration was frequently under-reported, particularly in risk models. Third, retrospective single-center datasets remained common in image and omics research. Fourth, explainability methods were inconsistently evaluated for clinical reliability. Fifth, prospective evidence linking AI to patient outcomes or durable workflow benefit was much smaller than the volume of algorithm-development literature. These observations support a disease-specific implementation strategy in which the AI task, data source, clinical action, and validation pathway are designed together rather than treating model accuracy as the endpoint.

Table 5. Comparative maturity and deployment priorities derived from the structured synthesis.

Dimension	Diabetes	Cardiovascular disease	Cancer	Priority for future translation
Risk prediction	4/5	5/5	4/5	External validation, calibration, diverse populations
Continuous monitoring / management	5/5	4/5	3/5	Prospective utility, alert burden, treatment-action safety
Imaging / biosignal AI	4/5	5/5	5/5	Independent testing, reader-assistance studies, workflow integration
Multimodal / omics integration	3/5	3/5	5/5	Data harmonization, missingness handling, external multi-site cohorts
Prospective clinical integration	4/5	4/5	3/5	Randomized/pragmatic trials and post-deployment monitoring

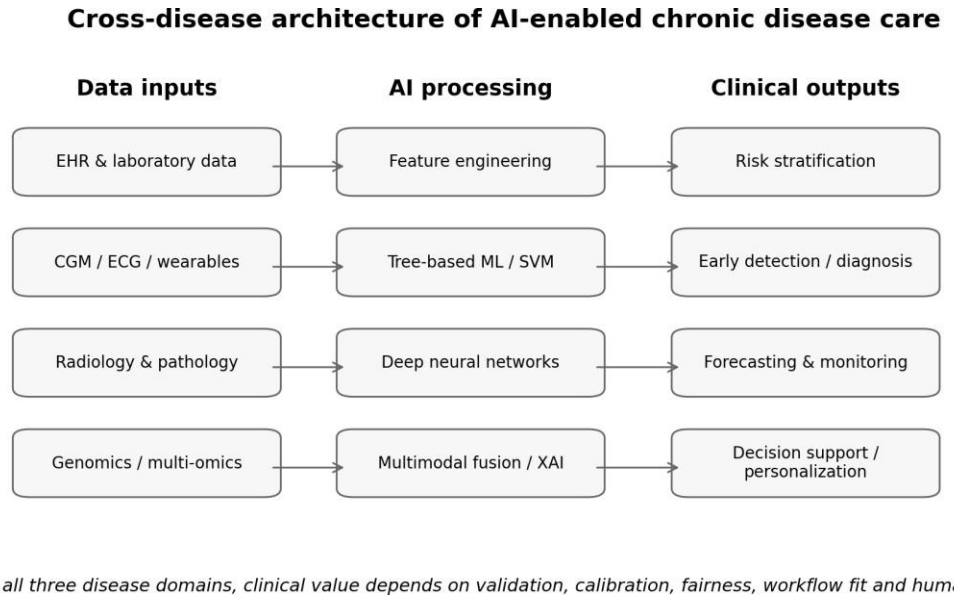


Figure 3. Cross-disease architecture of AI-enabled chronic disease care. The same computational pipeline supports different clinical endpoints depending on disease-specific data and actionability.

4. DISCUSSION

4.1 Why AI Matures Differently Across the Three Diseases

The comparison indicates that AI maturity is largely determined by the relationship between data frequency, endpoint clarity, and actionability. Diabetes produces dense longitudinal glucose data and has a measurable short-term physiological target; therefore, prediction can be linked directly to insulin delivery or behavioral action. This explains why automated management has advanced more rapidly than many complication-prediction models. CVD has a different advantage: highly standardized physiological signals such as 12-lead ECG and widely used imaging modalities. DL can extract latent patterns from these data at scale, enabling risk stratification from tests already embedded in routine care. Cancer presents a richer but more fragmented environment. Imaging and pathology are information dense, and tumor biology is inherently multimodal, but labels can be complex, molecular testing varies, treatment pathways differ by tumor type and stage, and prospective outcomes often require

long follow-up. Consequently, cancer AI can show excellent retrospective diagnostic accuracy while still facing a larger implementation gap. The central implication is that AI should be evaluated according to the clinical action it enables, not according to algorithm class alone.

4.2 Prediction Performance Is Not Equivalent to Clinical Benefit

A consistent finding across the literature is that high discrimination can coexist with weak evidence of clinical benefit. In diabetes, complication models often perform less well on external data than internal data [9]. In CVD, a recent scoping review found limited external validation and essentially absent calibration reporting among prognostic studies [23]. In cancer pathology, high mean sensitivity and specificity coexist with widespread high or unclear risk-of-bias domains [30]. These examples show why apparent accuracy can overstate readiness. A prediction model may identify rank order correctly but misestimate absolute risk, perform differently in a new hospital, or fail when acquisition protocols change. Clinical value also depends on whether the prediction changes management in a way that benefits the patient. AI-assisted radiology is illustrative: the key outcome is not that an algorithm is accurate in isolation, but whether clinicians assisted by AI perform better, work more efficiently, or make safer decisions [28]. Similarly, the randomized echocardiography trial demonstrates a workflow endpoint that is more directly relevant to implementation than retrospective image-classification accuracy [26].

4.3 Multimodal AI, Explainability, and Human-AI Collaboration

Multimodal AI is a major convergent direction. In diabetes, integrating CGM with insulin, carbohydrate intake, activity, retinal imaging, and clinical factors may improve forecasting and complication assessment [13-15]. CVD models can combine EHR variables, ECG, imaging, and wearable streams. Cancer offers the clearest rationale for multimodality because morphology, genomics, transcriptomics, and clinical context represent different layers of the same biological process [33-36]. Nevertheless, multimodal models increase the risk of overfitting, missing-data bias, and deployment complexity. Explainability can help clinicians interrogate model outputs, but explanation methods should not be treated as proof of validity. Ghassemi et al. cautioned that current XAI methods may provide false reassurance when used as a substitute for rigorous validation [5], whereas broader surveys emphasize that explanation requirements must be matched to the clinical purpose [6]. The most defensible near-term model is therefore human-AI collaboration: AI performs high-throughput pattern recognition, while clinicians integrate context, patient preferences, competing risks, and consequences of action.

4.4 Data Governance, Fairness, and Generalizability

Chronic-disease AI depends on data that are longitudinal, sensitive, and often institution specific. Diabetes wearables reveal daily behavior and physiology; CVD data include ECG and imaging fingerprints; cancer data may include genomic information with lifelong privacy implications. Governance must therefore address consent, access control, cybersecurity, auditability, and secondary use. Fairness is equally important. If development datasets underrepresent certain age groups, sexes, ethnicities, socioeconomic groups, or health systems, the model may amplify existing disparities. The CVD time-to-event review identified limited inclusion of social determinants and sex-stratified analyses [18], while TRIPOD+AI explicitly incorporates fairness and subgroup-performance considerations [3]. For cancer and diabetes, geographic and institutional variation in imaging equipment, laboratory practice, and care pathways can produce dataset shift. External validation should therefore be geographically and operationally independent whenever possible. Prospective monitoring is also required after deployment because clinical populations, devices, coding systems, and treatment standards change over time.

4.5 Implications for Research and Clinical Implementation

Future work should move from retrospective demonstration toward clinically governed implementation. Prediction studies should prespecify the intended decision, report discrimination and calibration, provide uncertainty intervals, and test performance in clinically relevant subgroups. Diagnostic AI should be evaluated in reader-assistance designs or prospective pathways rather than solely against retrospective labels. Management systems require safety constraints, fail-safe logic, human override, and monitoring for alert fatigue or automation bias. Multimodal cancer models need independent external cohorts beyond heavily reused research repositories. Diabetes complication models should demonstrate transportability across populations and health systems. CVD models should more consistently include calibration, decision-curve analysis, and comparison with established risk scores. Across all domains, AI-specific reporting standards such as TRIPOD+AI and CONSORT-AI can improve reproducibility [3,4]. The evidence also supports continuous post-deployment evaluation: AI performance should be treated as a monitored clinical property rather than a one-time certification event.

4.6 Strengths and Limitations of the Present Analysis

A strength of this study is its direct comparison of three major chronic disease domains using a common clinical framework while avoiding inappropriate pooling of fundamentally different metrics. The synthesis emphasizes recent evidence, prospective and randomized studies where available, and clinically interpretable limitations. It also explicitly excludes in vivo and animal research, maintaining focus on human clinical translation. Several limitations should be acknowledged. The work is a structured comparative synthesis rather than an exhaustive systematic review across every bibliographic database, and publication selection therefore reflects the chosen clinical scope. The 36 core references are representative rather than comprehensive. The evidence-maturity scores are ordinal interpretive summaries and should not be treated as quantitative meta-analytic estimates. Some source studies and reviews include overlapping primary cohorts, particularly in imaging and widely used public datasets. Finally, rapid evolution of AI means that regulatory clearances, model architectures, and clinical trial evidence may change after the September 2026 search cutoff.

5. CONCLUSIONS

AI is reshaping chronic-disease care, but its most credible use differs by disease context. In diabetes, the strongest clinical pathway links continuous data to forecasting and adaptive management, particularly CGM and automated insulin delivery. In CVD, AI is especially powerful for extracting latent information from ECG, imaging, and wearables and for augmenting conventional risk stratification. In cancer, AI provides major opportunities in radiology, digital pathology, and multi-omics integration, yet the heterogeneity of tumors and datasets makes generalizability a central challenge. Across all three domains, technical performance should not be equated with clinical readiness. Robust external validation, calibration, prospective evaluation, subgroup fairness, transparent reporting, data governance, and post-deployment surveillance are essential. The most defensible near-term model is not autonomous replacement of clinicians but carefully designed human-AI collaboration in which algorithms improve pattern recognition and forecasting while clinicians retain responsibility for contextual interpretation and shared decision-making. Future research should prioritize multicenter external validation, pragmatic clinical trials, interoperable multimodal data standards, and explicit measurement of patient-important outcomes. These steps can shift chronic-disease AI from high-performing prototypes toward dependable, equitable, and clinically useful systems.

Declarations

Ethics approval: Not applicable. This article is based exclusively on published literature and does not involve new patient recruitment, identifiable patient data, animal experiments, or in vivo studies.

Consent for publication: Not applicable.

Conflict of interest: The authors should disclose any relevant financial or non-financial conflicts before journal submission.

Funding: No specific funding information was provided for this manuscript.

Data availability: All data discussed in this article are derived from the cited published literature.

REFERENCES

[1] Topol EJ. High-performance medicine: the convergence of human and artificial intelligence. Nat Med. 2019;25(1):44-56. doi:10.1038/s41591-018-0300-7.

[2] Rajkomar A, Dean J, Kohane I. Machine Learning in Medicine. N Engl J Med. 2019;380(14):1347-1358. doi:10.1056/NEJMra1814259.

[3] 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. doi:10.1136/bmj-2023-078378.

[4] Liu X, Cruz Rivera S, Moher D, Calvert MJ, Denniston AK; SPIRIT-AI and CONSORT-AI Working Group. Reporting guidelines for clinical trial reports for interventions involving artificial intelligence: the CONSORT-AI extension. Nat Med. 2020;26:1364-1374. doi:10.1038/s41591-020-1034-x.

[5] Ghassemi M, Oakden-Rayner L, Beam AL. The false hope of current approaches to explainable artificial intelligence in health care. Lancet Digit Health. 2021;3(11):e745-e750. doi:10.1016/S2589-7500(21)00208-9.

[6] Markus AF, Kors JA, Rijnbeek PR. The role of explainability in creating trustworthy artificial intelligence for health care: a comprehensive survey of the terminology, design choices, and evaluation strategies. J Biomed Inform. 2021;113:103655. doi:10.1016/j.jbi.2020.103655.

[7] Khokhar PB, Gravino C, Palomba F. Advances in artificial intelligence for diabetes prediction: insights from a systematic literature review. Artif Intell Med. 2025;164:103132. doi:10.1016/j.artmed.2025.103132.
[8] Jacobs PG, Herrero P, Facchinetti A, et al. Artificial Intelligence and Machine Learning for Improving Glycemic Control in Diabetes: Best Practices, Pitfalls, and Opportunities. IEEE Rev Biomed Eng. 2024;17:19-41. doi:10.1109/RBME.2023.3331297.
[9] Kanbour S, Harris C, Lalani B, et al. Machine Learning Models for Prediction of Diabetic Microvascular Complications. J Diabetes Sci Technol. 2024;18(2):273-286. doi:10.1177/19322968231223726.
[10] Dholariya S, Dutta S, Sonagra A, et al. Unveiling the utility of artificial intelligence for prediction, diagnosis, and progression of diabetic kidney disease: an evidence-based systematic review and meta-analysis. Curr Med Res Opin. 2024;40(12):2025-2055. doi:10.1080/03007995.2024.2423737.
[11] Di Molfetta S, Di Gioia L, Caruso I, et al. Efficacy and Safety of Different Hybrid Closed Loop Systems for Automated Insulin Delivery in People With Type 1 Diabetes: A Systematic Review and Network Meta-Analysis. Diabetes Metab Res Rev. 2024;40(6):e3842. doi:10.1002/dmrr.3842.
[12] Wang X, Si J, Li Y, et al. Effectiveness and safety of AI-driven closed-loop systems in diabetes management: a systematic review and meta-analysis. Diabetol Metab Syndr. 2025;17(1):238. doi:10.1186/s13098-025-01819-0.
[13] Quader T, Pai S, Tan YM, Deshmukh H, Malabu U, Emeto TI. The efficacy of CGM driven machine learning algorithms in predicting hypoglycemia in patients with T1DM: a systematic review and meta-analysis. J Diabetes Metab Disord. 2026;25(1):2. doi:10.1007/s40200-025-01820-4.
[14] Yang Q, Bee YM, Lim CC, et al. Use of artificial intelligence with retinal imaging in screening for diabetes-associated complications: systematic review. EClinicalMedicine. 2025;81:103089. doi:10.1016/j.eclinm.2025.103089.
[15] Xie X, Wu C, Huang Z, et al. Artificial intelligence-powered prediction of diabetic complications: from clinical data to molecular omics. Brief Bioinform. 2026;27(1):bbag083. doi:10.1093/bib/bbag083.
[16] Tanabe H, Sato M, Miyake A, et al. Machine learning-based reproducible prediction of type 2 diabetes subtypes. Diabetologia. 2024;67(11):2446-2458. doi:10.1007/s00125-024-06248-8.
[17] Tiwari A, Shah PC, Kumar H, et al. The Role of Artificial Intelligence in Cardiovascular Disease Risk Prediction: An Updated Review on Current Understanding and Future Research. Curr Cardiol Rev. 2025;21(6):e1573403X351048. doi:10.2174/011573403X351048250329170744.
[18] Teshale AB, Htun HL, Vered M, Owen AJ, Freak-Poli R. A Systematic Review of Artificial Intelligence Models for Time-to-Event Outcome Applied in Cardiovascular Disease Risk Prediction. J Med Syst. 2024;48(1):68. doi:10.1007/s10916-024-02087-7.
[19] Attia ZI, Noseworthy PA, Lopez-Jimenez F, et al. An artificial intelligence-enabled ECG algorithm for the identification of patients with atrial fibrillation during sinus rhythm: a retrospective analysis of outcome prediction. Lancet. 2019;394(10201):861-867. doi:10.1016/S0140-6736(19)31721-0.
[20] Attia ZI, Kapa S, Lopez-Jimenez F, et al. Screening for cardiac contractile dysfunction using an artificial intelligence-enabled electrocardiogram. Nat Med. 2019;25(1):70-74. doi:10.1038/s41591-018-0240-2.
[21] Awasthi S, Sachdeva N, Gupta Y, et al. Identification and risk stratification of coronary disease by artificial intelligence-enabled ECG. EClinicalMedicine. 2023;65:102259. doi:10.1016/j.eclinm.2023.102259.
[22] Cicek V, Kozan Cikirikci EH, Babaoglu M, et al. Machine learning for prognostic prediction in coronary artery disease with SPECT data: a systematic review and meta-analysis. EJNMMI Res. 2024;14(1):117. doi:10.1186/s13550-024-01179-2.
[23] Pai S, Subramanian R, Krishnan L, et al. A scoping review on artificial intelligence-based tools for cardiovascular disease risk prediction. BMC Med Inform Decis Mak. 2026. doi:10.1186/s12911-026-03699-4.
[24] Hadida Barzilai D, Sudri K, Goshen G, et al. Randomized Controlled Trials Evaluating Artificial Intelligence in Cardiovascular Care: A Systematic Review. JACC Adv. 2025;4(11 Pt 1):102152. doi:10.1016/j.jacadv.2025.102152.
[25] Barrera N, Solorzano M, Jimenez Y, Kushnir Y, Gallegos-Koyner F, Dagostin de Carvalho G. Accuracy of Smartwatches in the Detection of Atrial Fibrillation: A Systematic Review and Diagnostic Meta-Analysis. JACC Adv. 2025;4(11 Pt 1):102133. doi:10.1016/j.jacadv.2025.102133.
[26] He B, Kwan AC, Cho JH, et al. Blinded, randomized trial of sonographer versus AI cardiac function assessment. Nature. 2023;616(7957):520-524. doi:10.1038/s41586-023-05947-3.
[27] Felici A, Peduzzi G, Pellungrini R, Campa D. Artificial intelligence to predict cancer risk, are we there yet? A comprehensive review across cancer types. Eur J Cancer. 2025;222:115440. doi:10.1016/j.ejca.2025.115440.

[28] Zhao D, Packer T, Jie X, Shahid M, Oke J, Plüddemann A. Diagnostic accuracy of artificial intelligence-assisted radiology assessment of cancer: a systematic review. BJR Artif Intell. 2025;2(1):ubaf016. doi:10.1093/bjrai/ubaf016.
[29] Rodriguez JPM, Rodriguez R, Silva VWK, et al. Artificial intelligence as a tool for diagnosis in digital pathology whole slide images: a systematic review. J Pathol Inform. 2022;13:100138. doi:10.1016/j.jpi.2022.100138.
[30] McGenity C, Clarke EL, Jennings C, et al. Artificial intelligence in digital pathology: a systematic review and meta-analysis of diagnostic test accuracy. NPJ Digit Med. 2024;7(1):114. doi:10.1038/s41746-024-01106-8.
[31] Thong LT, Chou HS, Chew HSJ, Lau Y. Diagnostic test accuracy of artificial intelligence-based imaging for lung cancer screening: a systematic review and meta-analysis. Lung Cancer. 2023;176:4-13. doi:10.1016/j.lungcan.2022.12.002.
[32] Leonard S, Patel MA, Zhou Z, Le H, Mondal P, Adams SJ. Comparing Artificial Intelligence and Traditional Regression Models in Lung Cancer Risk Prediction Using A Systematic Review and Meta-Analysis. J Am Coll Radiol. 2025;22(6):675-690. doi:10.1016/j.jacr.2025.02.042.
[33] Acharya D, Mukhopadhyay A. A comprehensive review of machine learning techniques for multi-omics data integration: challenges and applications in precision oncology. Brief Funct Genomics. 2024;23(5):549-560. doi:10.1093/bfgp/elae013.
[34] Li L, Sun M, Wang J, Wan S. Multi-omics based artificial intelligence for cancer research. Adv Cancer Res. 2024;163:303-356. doi:10.1016/bs.acr.2024.06.005.
[35] Jennings C, Broad A, Godson L, et al. AI in Cancer Prognosis: A Systematic Review of Multimodal Models Combining Pathology Images and High-Throughput Omics. Cancer Inform. 2026;25:11769351261434523. doi:10.1177/11769351261434523.
[36] Khosravi P, Fuchs TJ, Ho DJ. Artificial Intelligence-Driven Cancer Diagnostics: Enhancing Radiology and Pathology through Reproducibility, Explainability, and Multimodality. Cancer Res. 2025;85(13):2356-2367. doi:10.1158/0008-5472.CAN-24-3630.