Online ISSN: 3070-5320 Print ISSN: 3070-5312
American Journal of Medicine and Health Studies
Volume 3 (2026), pp. 42-59
Star Scholars Press
https://ojed.org/med
https://doi.org/10.32674/ex08x042

AI-Driven Healthcare Analytics for Early Disease Detection: A Scoping Review of Clinical Applications, Validation, and Translational Challenges

Emmanuel Agbeko Enyo
George Washington University, Washington, DC, USA
emmanuel.agbeko@gwmail.gwu.edu || emmaenyor2@gmail.com

ABSTRACT

Background: *Artificial intelligence is increasingly used to support earlier disease detection across imaging, electronic health records, physiologic signals, wearable devices, and biomarker-based platforms. However, reported performance is highly variable across diseases, data environments, and validation settings, and the translational maturity of the field remains uneven.* ***Objective:*** *To map the breadth and characteristics of empirical studies on AI-driven healthcare analytics for early disease detection and to identify patterns in model application, validation, implementation relevance, and evidence gaps.* ***Methods:*** *A scoping review was conducted using the Arksey and O'Malley framework with further refinements and reported in line with PRISMA-ScR. PubMed and Google Scholar were searched for English-language peer-reviewed studies published from 2018 to 2025. Eligibility was defined using the Population, Concept, and Context framework. Data were charted on study characteristics, data modality, target disease, analytic method, reference standard, performance metrics, and implementation relevance, and synthesized narratively.* ***Results:*** *Twenty-nine studies were included. The evidence covered sepsis and septic shock, diabetic retinopathy, breast cancer, skin cancer, colorectal polyps, pulmonary nodules and lung cancer, tuberculosis, pancreatic cancer, atrial fibrillation, and chronic kidney disease. Imaging dominated visual detection tasks, whereas EHR, physiologic, wearable, and lipidomic data were more common in temporal or multimodal prediction tasks. Most studies reported moderate-to-high discriminatory performance, but the strength of evidence depended heavily on external, temporal, or prospective validation. The most consistent gains were earlier case identification, reduced missed lesions, improved prioritization, and workflow support rather than autonomous replacement of clinicians.* ***Conclusion:*** *AI shows substantial promise as a clinically embedded decision-support tool for earlier and more targeted disease detection, but widespread adoption still depends on robust external validation, local adaptation, and stronger real-world implementation evidence.*

Keywords: artificial intelligence, early disease detection, healthcare analytics, machine learning

Corresponding Author: Emmanuel Agbeko Enyo, George Washington University, Washington, DC, USA. Email: emmanuel.agbeko@gwmail.gwu.edu, 009-0007-4854-5922

Editors: Robin Butler | Akanksha Anand | Ismatara Reena | Natalie Anthony | Reshma Gopal | Cristina Kelly

INTRODUCTION

Artificial intelligence is increasingly being proposed as an aid for earlier detection of diseases, with many high-prevalence diseases presenting after a clinically relevant delay, especially when symptoms are mild, intermittent or non-specific. Recent studies have demonstrated that AI can identify clinically important patterns in a variety of data sources, such as fundus photography for diabetic retinopathy, mammograms for breast cancer, dermo scope images for skin cancer, electrocardiograms for atrial fibrillation and chest X-rays for thoracic disease. In eye screening, Gulshan et al. (2016) and Ipp et al. (2021) demonstrated high performance of a deep learning system for detection of diabetic retinopathy, including standalone detection in non-specialist settings. In breast imaging, McKinney et al. (2020) demonstrated impressive international performance of an AI system for mammography, while Ahn et al. (2022) showed improved reader performance and efficiency with AI-aided chest radiograph interpretation. These studies suggest that AI has evolved from concept to being tested as a potentially useful clinical tool for earlier detection in a screening or diagnostic workflow.

However, existing evidence also demonstrates that high-performing models do not always translate into clinical usefulness. Tschandl et al. (2020) demonstrated that a human-AI team outperformed humans and AI alone in identifying skin cancer, which suggests that the best role for many AI systems is to assist, rather than replace, humans. Attia et al. (2019) also found that AI can detect the presence of atrial fibrillation in sinus rhythm, extending the concept of early detection from overt disease to risk factors. At the same time, these studies highlight some of the major tensions in the field. Models are often trained on highly curated data sets, and tested against expert readers, or for very specific tasks that may not account for the variability in case mix, workflow, device performance, or disease prevalence that occur in practice. Consequently, claims of diagnostic superiority must be interpreted with caution, particularly when the metrics are presented without equal consideration of false positives, transportability, and the downstream implications of clinical adoption.

These issues present the need for a critical integration of AI-driven health analytics and early disease detection. This is also highlighted by the policy landscape. The World Health Organization (2021) has developed six ethical principles for AI governance in health, which highlight that benefits have to be weighed against the risks from autonomy, transparency, accountability, inclusion, and safety. The U.S. Food and Drug Administration's (2021) action plan for artificial intelligence and machine learning highlights five broad priorities for regulatory improvement, given the challenges of assessing adaptive software using traditional device regulations. Meanwhile, the World Health Organization's (2022) tuberculosis screening handbook affirms that systematic screening should be programmatic, targeted and contextualized rather than necessarily universal. Taken together, the empirical and policy literature suggests that the central question is no longer whether AI can detect disease-related signals, but under what

conditions it can do so reliably, equitably, and in ways that genuinely improve early clinical action. This review was therefore conducted to identify the current evidence, commonalities across studies in performance and implementation, and the maturity, promise and early-stage nature of the methodological evidence.

Statement of Significance

Problem/Issue: Despite rapid advances in artificial intelligence (AI) for early disease detection, evidence regarding clinical validity, generalizability, and real-world implementation remains fragmented and inconsistent.

What is Already Known: AI models have demonstrated high diagnostic accuracy across imaging, electronic health records, and physiologic data, often under controlled or retrospective conditions. However, many studies lack robust external validation and fail to address variability in clinical settings.

What This Paper Adds: This scoping review synthesizes empirical evidence from 2018–2025, identifying patterns in data modality use, validation practices, and translational readiness. It highlights critical gaps in external validation, implementation evidence, and workflow integration.

Who Benefits: Clinicians, researchers, policymakers, and healthcare system leaders will benefit from clearer insights into where AI can meaningfully support early detection and what is required for safe clinical adoption.

RESEARCH METHOD

Study Design

Using a scoping review design, we systematically mapped the breadth and characteristics of empirical studies on AI-driven healthcare analytics for early disease detection, considering various clinical domains, care settings, and data modalities. We also identified gaps in the evidence relating to implementation, validation, and real-world applicability. This study employed the methodological framework proposed by Arksey and O'Malley (2005), which was subsequently refined to emphasize transparent reporting, clarity of purpose, and collaborative teamwork (Levac et al., 2010). The Population, Concept, and Context (PCC) framework was used to determine eligibility, in adherence to the PRISMA Extension for Scoping Reviews (Tricco et al., 2018).

Eligibility Criteria

The eligibility criteria were defined by applying the PCC approach and were consistently applied to all records (Table 1).

Table 1 Eligibility Criteria

Framework Domain	Inclusion criteria	Exclusion criteria
Population	Human participants or patient-linked clinical datasets in which AI was used to support early disease detection, early-stage diagnosis, short-horizon onset prediction, or clinically actionable case identification	Animal studies; in vitro or purely laboratory-only studies with no patient-linked diagnostic application
Concept	Empirical evaluation of AI, machine learning, or deep learning applied to healthcare analytics for early disease detection	Non-empirical AI papers; reviews, commentaries, editorials, and protocols without outcomes
Context	Any healthcare or screening context relevant to disease detection	Non-healthcare settings; studies with no clinical, screening, or diagnostic context
Study types & timeframe	Peer-reviewed empirical studies published from 2018 to 2025 in English	Grey literature only; non-English full text; studies published outside the defined timeframe

Information Source and Search Strategy

In order to identify studies focusing on the intersection of healthcare analytics, artificial intelligence, and early disease detection, a structured search strategy was developed. A combination of controlled vocabulary and free-text keywords related to "artificial intelligence," "machine learning," and "early disease detection" was used as the search terms. Boolean operators (AND, OR) were also applied to balance sensitivity and specificity. Both PubMed and Google Scholar were searched, with the results restricted to English-language articles published between 2018 and 2025 (Table 2).

Table 2 Search String

Database	Search string	Limits
PubMed	("artificial intelligence" OR "machine learning" OR "deep learning") AND ("early detection" OR screening OR diagnosis) AND (disease OR cancer OR sepsis OR retinopathy OR tuberculosis OR "atrial fibrillation")	2018–2025; English; Humans
Google Scholar	("artificial intelligence" OR "machine learning") AND ("early disease detection" OR screening OR diagnosis) AND healthcare	2018–2025; English; sorted by relevance

Data Charting and Synthesis

A data extraction form was created to extract the information of study characteristics and results in line with the PCC framework and review questions. Data extracted included study details, study context and population, data source and input features, target disease and early detection task, AI/analytics method, reference standard and performance measures, and key findings and implementation relevance. The table was tested on the first 5 studies and refined to ensure consistency. Data were charted by two independent reviewers for each study, with adjudication by a third reviewer. The results were synthesized through

descriptive mapping and narrative synthesis due to variation in the diseases, data, model, validation and performance reported. This highlighted similarities across clinical areas, variations in the use of AI for early detection, the level of internal and external validation, implementation considerations, and common evidence gaps with regard to generalizability, integration with clinical workflows, and evaluation in real-world settings.

RESULTS

Study Selection

A total of 1,065 records were identified in the database search, with 164 from PubMed and 901 from Google Scholar. Of these, 590 were duplicates, leaving 475 to screen for relevance by title and abstract. We excluded 365 records because the titles and abstracts did not meet the review objective. The remaining 110 were screened for full text eligibility. A further 34 were excluded because they did not have a patient-linked diagnostic application, 35 were excluded because they were not applied in a clinical, screening or diagnostic context, and 12 were excluded because they were grey literature or protocols with no results. After full-text review, 29 studies were included in the review.

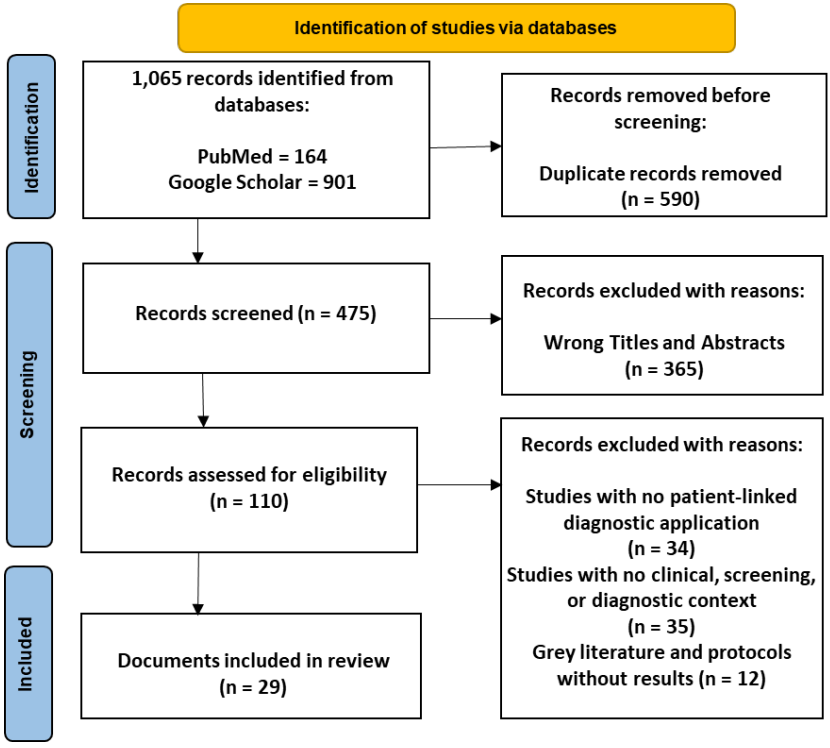

Figure 1 PRISMA-ScR Flow Diagram

Study Characteristics

The review included 29 empirical studies published between 2018 and 2025, comprising retrospective model-development studies, prospective cohort and implementation studies, feasibility studies, technical validation studies, and randomized tandem trials across

diverse clinical contexts. Collectively, the evidence covered sepsis and septic shock prediction from EHR data in acute and intensive care settings, diabetic retinopathy, colorectal polyp, pulmonary nodule/lung cancer, and tuberculosis detection from medical imaging, and serum lipidomics, atrial fibrillation detection from single-lead ECG, RR intervals, and chronic kidney disease prediction from clinical and cohort data. The included studies drew on highly heterogeneous data sources, including structured and unstructured electronic health records, claims/billing data, vital-sign time series, and wearable-device signals. Sample sizes ranged from relatively small curated image or video datasets to very large real-world cohorts and multicenter or registry-scale datasets. Methodologically, the studies most commonly applied deep learning architectures such as CNNs, ResNet-based models, and multitask networks, alongside classical machine-learning approaches such as random forest, XGBoost, logistic regression, and related ensemble frameworks. Across the included evidence, most studies reported moderate-to-high discriminatory performance and positioned AI primarily as a tool for screening support, earlier case identification, workflow enhancement, or risk stratification, rather than as a replacement for clinical judgment.

Table 3 Characteristics of Included Studies

Author & Year	Study setting & population	Data source & input features	Target disease & early detection purpose	AI/analytics method	Reference standard & key performance
Shen et al., 2019	CBIS-DDSM 2,478 mammograms from 1,249 women; INbreast 410 mammograms from 115 patients	Mammograms from digitized film and FFDM	Breast cancer detection on screening mammography	All-convolutional whole-image classifier with VGG16 and ResNet50 patch classifiers	Pathology-confirmed ROI labels and BI-RADS labels. CBIS-DDSM AUC 0.88/0.91, sensitivity 86.1%, specificity 80.1%. INbreast AUC 0.95/0.98, sensitivity 86.7%, specificity 96.1%
Bedoya et al., 2020	Adult hospitalized patients; 42,979 encounters in development/internal validation and 39,786 in temporal validation	EHR data with demographics, comorbidities, vitals, medications, and labs	Sepsis detection within 4 h of onset	Deep-learning MGP–RNN compared with RF, Cox regression, penalized logistic regression, SIRS, qSOFA, and NEWS	Sepsis defined by SIRS + blood culture + end-organ failure. C-statistic 0.88 for MGP–RNN; median advance detection 5 h
Masino et al., 2019	NICU infants hospitalized ≥48 h; 110 culture-positive, 265 clinically positive, 1,100 controls	Routine EHR data from pre-evaluation windows	Infant sepsis recognition at least 4 h before clinical recognition	8 machine-learning models with 10-fold nested cross-validation	Culture-positive and clinically positive case definitions. Best models achieved AUC 0.80–0.82 for culture-positive only and 0.85–0.87 when clinically positive cases were added
Nemati et al., 2018	ICU adults from two hospitals and MIMIC-III; roughly 27,000 development and 42,000 validation patients after exclusions	High-resolution vital-sign series and EMR data	Sepsis onset prediction 4–12 h before clinical recognition	AISE interpretable machine-learning model	Sepsis-3 outcome definition. AUROC 0.83–0.85 across 12-, 8-, 6-, and 4-hour prediction windows

Author & Year	Study setting & population	Data source & input features	Target disease & early detection purpose	AI/analytics method	Reference standard & key performance
Misra et al., 2021	45,425 inpatient visits from Geisinger systems	EHR, insurance claims, billing, and lab data	Prediction of progression from sepsis to septic shock up to 6 h from admission	Multiple models including Random Forest, XGBoost, SVM, logistic regression, and others	Septic shock defined using CMS Sepsis criteria. Best model was Random Forest, AUROC 0.9483, sensitivity 83.9%, specificity 88.1%
Dai et al., 2021	Local cohort of 666,383 fundus images from 173,346 patients; multiple external validation datasets	Fundus photographs with DR grades, DME, and lesion annotations	Diabetic retinopathy detection and grading across early to late stages	DeepDR multi-task transfer learning system	Lesion AUCs 0.901–0.967. DR grading AUCs 0.943–0.972. External grading AUCs 0.916–0.970
Oh et al., 2021	Single-center South Korean study with 11,734 DR and 1,537 healthy UWF images from 1,308 patients	Ultra-wide-field fundus photographs	Diabetic retinopathy detection using UWF images	ETDRS 7SF segmentation plus ResNet-34	ETDRS-based grading by ophthalmologist and certified grader. AUC 0.9150, accuracy 0.8338, sensitivity 0.8338, specificity 0.8341
Lam et al., 2018	Public datasets with 35,000 Kaggle images and 1,200 Messidor-1 images, plus 550 physician-verified MildDR images	Color fundus photographs	Diabetic retinopathy staging, especially mild/early disease	CNNs with transfer learning using GoogLeNet and AlexNet	Dataset grade labels and physician verification. Validation sensitivity 95%; peak accuracies 74.5% (2-ary), 68.8% (3-ary), 57.2% (4-ary)
Ahmadi Mehr & Ameri, 2022	Dermoscopic and photographic lesion datasets; dermoscopy-only analysis used 57,536 images	Lesion images plus age, sex, and anatomical site metadata	Skin cancer detection and classification	Inception-ResNet-v2 CNN with metadata input	Biopsy and histopathology as reference. Accuracy improved by ≥5% with metadata. 89.3% ± 1.1% for 4 major skin conditions and 94.5% ± 0.9% for benign vs malignant
Alam et al., 2022	HAM10000 dataset with 7 lesion classes	Skin-lesion images with normalization, resizing, and augmentation	Skin cancer classification for early identification	AlexNet, InceptionV3, RegNetY-320	HAM10000 labels as reference. Best reported performance accuracy 91%, F1-score 88.1%, ROC 0.95
Gilani et al., 2023	ISIC 2019 dataset with 3,670 melanoma and 3,323 non-melanoma images	Dermoscopy images	Melanoma vs non-melanoma classification	Deep spiking neural networks, especially spiking VGG-13	ISIC 2019 labels. Best results accuracy 89.57% and F1 90.07%
Al-Waisy et al., 2025	Public dermoscopy datasets ISIC 2019 and HAM1000	Dermoscopy images with preprocessing, segmentation, and deep feature extraction	Automated early diagnosis and classification of skin lesions	Skin-DeepNet using Mask R-CNN, GrabCut, HRNet, DBN, and ensemble fusion	On ISIC 2019, accuracy 99.65%, precision 99.51%, AUC 99.94%. On HAM1000, accuracy 100%, precision 99.92%, AUC 99.97%

Author & Year	Study setting & population	Data source & input features	Target disease & early detection purpose	AI/analytics method	Reference standard & key performance
Elhakim et al., 2024	Population-wide mammography screening cohort of 249,402 mammograms from Southern Denmark	Screening mammograms processed by Lunit INSIGHT MMG	Breast cancer screening workflow integration	Deep-learning AI used as first reader, second reader, or triage system	Histopathology or cancer-free follow-up within 24 months. AIfirst showed similar accuracy except higher arbitration. AIsecond lowered sensitivity. AItriage improved sensitivity and PPV and reduced arbitration
Chang et al., 2025	National breast screening cohort in South Korea with 24,543 women and 24,545 mammograms	Screening mammograms interpreted with and without AI-CAD	Breast cancer screening in real-world single-read setting	Prospective comparison of radiologists with vs without AI-CAD	Pathologically confirmed breast cancer. AI-CAD increased CDR from 5.01‰ to 5.70‰ with no significant change in recall rate; PPV1 12.6 vs 11.2
González-Bueno Puyal et al., 2022	Video-colonoscopy study using videos from 46 patients and the SUN polyp database	Video colonoscopy with temporal frame sequences	Colorectal polyp detection in real time	Hybrid 2D/3D CNN with ResNet-101 encoder	On SUN database, per-polyp sensitivity 97.0%, per-frame sensitivity 84.71%, specificity 84.57%, F1 83.29%
Glissen Brown et al., 2022	Multicenter U.S. tandem colonoscopy trial; 223 patients analyzed	Real-time colonoscopy with and without CADe	Adenoma and polyp detection during screening/surveillance colonoscopy	Deep-learning CADe EndoScreener	Primary outcome was adenoma miss rate. AMR 20.12% vs 31.25%, SSL miss rate 7.14% vs 42.11%, APC 1.19 vs 0.90
Lee et al., 2020	Training on 8,075 images; validated on still-image and video datasets including unaltered videos	Colonoscopy images and videos	Real-time colon polyp detection	YOLOv2 detection algorithm	Per-image sensitivity 96.7% and 90.2% on still-image datasets. In video analysis, sensitivity up to 89.9% with FP rate 6.3%. Unaltered videos showed sensitivity 89.3%
Hendrix et al., 2023	Internal and external test sets of 100 CT scans each from two Dutch hospitals	Chest CT from routine care	Detection of benign nodules, lung cancer, and metastases in non-screening CT	Deep-learning pulmonary nodule detection system	Reference standard from radiologists and cancer registry. At 1 FP/scan, sensitivity was 94.3% for benign nodules, 96.9% for primary lung cancers, and 92.0% for metastases
Jacobs et al., 2021	300 screening CT scans with 11 radiologists in observer study	Low-dose chest CT scans	Lung cancer detection on screening CT	Evaluation of top 3 Kaggle DSB2017 algorithms against radiologists	Histopathology for cancer-positive scans and ≥2 years follow-up for cancer-negative scans. AUCs 0.877, 0.902, 0.900 for algorithms and 0.917 for radiologists

Author & Year	Study setting & population	Data source & input features	Target disease & early detection purpose	AI/analytics method	Reference standard & key performance
Chen et al., 2024	External validation from hospital datasets with 250 CXRs and 105 CXRs for physician comparison	Chest radiographs for TB and abnormality detection	Pulmonary tuberculosis detection	Google Teachable Machine with MobileNet transfer learning	Clinical TB confirmed by microbiology or pathology. Validation AUCs 0.951 and 0.975. AI AUC dropped to 0.828 when abnormal non-TB CXRs were added; human-AI combination improved AUC to 0.862–0.885
Showkatian et al., 2022	Public CXR datasets from Shenzhen and Montgomery County; 2,040 training, 120 validation, 120 test images after augmentation	Chest X-ray images	Tuberculosis detection from CXR	CNN from scratch plus transfer learning models Inception_v3, Xception, ResNet50, VGG19, VGG16	Best models Xception, ResNet50, VGG16 achieved 90.0% accuracy with 91.0% precision, sensitivity, F1, and AUC
Hwang et al., 2019	Development dataset of 54,221 normal and 6,768 TB chest radiographs; external validation on 6 multicenter multinational datasets	Chest radiographs labeled and annotated by radiologists	Active pulmonary tuberculosis detection	DLAD deep CNN with image-wise classification and lesion localization	External classification AUROC 0.977–1.000 and localization AUAFROC 0.973–1.000. DLAD outperformed physicians in classification and localization
Chen et al., 2021	Optum deidentified EHR dataset with 3,322 early-stage and 25,908 late-stage pancreatic cancer cases	EHR data with diagnoses, procedures, notes, medications, and demographics	Early-stage pancreatic cancer identification	XGBoost with 70/30 split and 5-fold cross-validation	Final model AUC 0.84. At 60% sensitivity and 90% specificity, 58.3% of late-stage cases could be detected earlier, with median lead time 24 months
Dubatovka & Buhmann, 2022	Public ECG datasets including 8,528 single-lead recordings from PhysioNet/CinC Challenge 2017	Single-lead ECG split into heartbeat sequences	Atrial fibrillation detection	Deep neural network pipeline based on DeepHeartBeat variants	Best model P-DHB+RNN achieved F1 94.6 ± 1.0, sensitivity 97.4 ± 1.9, specificity 99.0 ± 0.1, AUC 99.7 ± 0.2, accuracy 98.9 ± 0.2
Hiraoka et al., 2022	79 postoperative cardiac surgery patients analyzed; 27 developed AF	Wearable PPG pulse-rate data from Apple Watch plus telemetry and 12-lead ECG	Early detection of paroxysmal AF after cardiac surgery	LightGBM gradient boosting decision tree	AF diagnosed from telemetry and 12-lead ECG. Model AUC 0.9416, sensitivity 0.909, specificity 0.838
Lown et al., 2020	Case-control validation in 415 adults >65 years, including 79 AF and 336 non-AF participants	RR intervals from wearable heart-rate monitor	AF detection using inexpensive wearable technology	Support Vector Machine on compressed Lorenz-plot images	Reference standard was 12-lead ECG. Validation sensitivity 100% and specificity 97.6%

Author & Year	Study setting & population	Data source & input features	Target disease & early detection purpose	AI/analytics method	Reference standard & key performance
Bahrami et al., 2024	MASHAD cohort, Iran; 6,855 participants used for model development	Sociodemographic, anthropometric, and laboratory variables	Early CKD prediction/detection	Gender-specific Random Forest and Neural Network models	CKD defined as GFR <60 mL/min/1.73 m². Best model was female RF with AUC 0.90; female NN 0.89; male models 0.88
Singh et al., 2022	Dataset-based CKD study	Clinical variables selected by Recursive Feature Elimination	Early detection and prediction of CKD	Proposed deep neural network compared with SVM, KNN, logistic regression, RF, and Naive Bayes	Proposed model achieved 100% accuracy and outperformed comparator models
Wang et al., 2021	Exploratory study with 333 PDAC and 262 healthy controls; validation across 1,898 participants from five hospitals, including internal, external, and prospective cohorts	Serum lipidomics with untargeted HPLC-MS and targeted MRM LC-MS assay	Pancreatic ductal adenocarcinoma detection, including early-stage disease	SVM with greedy and MS-based feature selection	Internal test set AUC 0.9444, external validation 0.9351, prospective cohort 0.9389. External early-stage PDAC detection 86.38%; prospective early-stage detection 90.91%. Model outperformed CA19-9 and CT in the prospective cohort

Thematic Synthesis

Theme 1: Data Modality Fit

The first theme was the match between the design of an AI system and the most common data stream in the clinical area. Visual studies focused on high-throughput modalities where subtle visual abnormalities are encountered frequently, such as fundus photos for diabetic retinopathy, mammograms for breast cancer, dermoscopy for skin lesions, colonoscopy images and videos for colorectal polyps, chest X-rays for tuberculosis, and chest CT for pulmonary nodules and lung cancer. In contrast, sequential or discrete data were more often used where disease progression is largely temporal, such as EHR or vital-sign data streams for sepsis and septic shock, pulse or ECG data streams from wearable devices for atrial fibrillation, and serum lipidomics or high-dimensional variables in EHR for pancreatic cancer and CKD. In all these cases, AI was not used as a generic analysis tool; rather, it was used for specific clinical tasks, such as screening, triage, first-pass lesion detection, incident-risk prediction and short-term deterioration.

Theme 2: Performance with Validation Gradients

The second theme was that most studies had good discriminatory metrics, but the evidence presented was highly dependent on the validation environment. High areas under the curve (AUCs), sensitivities and accuracies were reported for diabetic retinopathy grading, active tuberculosis detection, skin lesion classification, polyp detection and sepsis prediction, showing that AI can pick up clinically relevant signals from a wide range of data. But the most valuable studies were those that did more than internal testing. Temporal, external (multicenter) and external/prospective validation of models for sepsis detection,

tuberculosis, serum lipidomic detection of PDAC, population-wide simulations with mammograms, and prospective AI-assisted breast cancer screening demonstrated that performance can be maintained outside of the initial development set, but more often with more realistic trade-offs. Ultimately, the greatest benefits were not flawless classification, but detection earlier in the disease process, fewer misses, greater yield, or more efficient work up. This implies that it is not maximum accuracy, but robustness, that is the key translational indicator.

Theme 3: Translation and Limits

Across all studies, value of AI implementation was determined not just by the novelty of algorithms, but by the location in the workflow, quality of input, and the credibility of the reference standard. In mammography, AI outcomes depended on the point of insertion of AI. In colonoscopy, temporal filtering and video-aware architectures improved stability by reducing transient false positives. In retinal imaging, lesion-aware and field-specific approaches improved lesion detection, and in sepsis, clinical phenotypes rather than billing codes improved detection. A related theme was human-AI synergy: physician-AI reading improved tuberculosis diagnosis, while wearable AF studies identified the potential for convenience to be limited by lack of agreement with telemetry or variable rhythms in practice. However, other studies were limited by single-site, single-vendor populations, curated data, or class imbalance. Overall, the synthesis supports AI as a clinically meaningful adjunct for earlier recognition and prioritization, but not as a context-free substitute for expert judgment or rigorous validation.

DISCUSSION

This review reveals AI-based early detection is most compelling when model architecture is well suited to the most common clinical data source and the target use case. Imaging models were most common in diabetic retinopathy, breast cancer, and skin cancer, while electronic health record (EHR), physiological time series, wearable data, and serum lipidomics were more common in sepsis, atrial fibrillation, chronic kidney disease, and pancreatic cancer. In each of these areas' performance was often excellent, but what was more significant was that clinical effectiveness was determined by whether models improved early detection, reduced false negatives, or enabled more efficient triage and workup. In general, the reviewed studies suggest that AI is a complement to rather than a replacement for screening and diagnosis.

This finding is broadly in line with global evidence. In retinal disease, Gulshan et al. (2016) showed that deep learning can be used to achieve high diagnostic accuracy on fundus photographs, but Abràmoff et al. (2018) and Bellemo et al. (2019) bring us closer to real-world applications by showing autonomous or clinically verified screening performance in primary care and Africa respectively. Likewise, Esteva et al. (2017) demonstrated benchmark-level performance in skin lesion classification, but this suggests that the many subsequent skin studies continue to focus more on images than on deployment. In breast imaging, Rodríguez-Ruiz et al. (2019) and Lång et al. (2023) agree with the consensus that AI can achieve radiologist-like performance and reduce reading time, and also that benefits are dependent on workflow. The same translational point is made in colonoscopy, where Wang et al. (2019) showed that real-time CADe can increase adenoma detection in real clinical practice, and is consistent with the reviewed studies that showed the best gains came from reduced false negatives, as opposed to retrospective accuracy.

The findings of this study also align with international evidence that longitudinal and physiologic data can help predict risks earlier but with greater concern for data context and portability. Ardila et al. (2019) demonstrated that end-to-end deep learning can be used to predict lung cancer risk from low-dose CT, and Hannun et al. (2019) reported that ambulatory single-lead ECG can be used to classify rhythm at expert level. Matchaba et al. (2023) also showed the potential of national electronic medical record data to support earlier detection of pancreatic cancer. But our review indicates that the results are best read with caution. The results of the EHR, wearable, and multimodal studies were more sensitive to case definition and signal quality, health-system organisation, and proximity-to-diagnosis information than the headline results might suggest. This is particularly important for pancreatic cancer and sepsis, where high performance may depend on a high density of documentation towards diagnosis and highly specific clinical populations.

This implies that future work should focus on clinically-embedded design, alongside algorithmic innovation. Models are more likely to be useful when they are developed for a specific operational purpose and validated for their external applicability, subgroup validity, calibration, false-positive rate, and workflow impacts. The study also confirms that AI may be particularly useful where specialist time limits the ability to make timely diagnoses, such as tuberculosis screening, retinal evaluation, and/or large-scale imaging. The evidence also suggests the validity of a model trained at one site, with a particular device, or in a particular registry, may not translate to another. Re-training models and adjusting thresholds, and governance of human oversight remain critical.

Despite the findings and implications of this study, some important limitations were identified. We found a broad diversity of diseases, data types, prediction targets, reference standards and performance measures reported, which was not systematically comparable and precluded quantitative pooling. The included studies were predominantly retrospective development or technical validation studies, with some using curated public data, and inconsistent reporting of calibration, predictive values and the impact of workflows. There was also some inconsistency in how the statuses of diseases, particularly sepsis and CKD, were defined, which has implications for how performance is interpreted. As a result, the review highlights overarching trends in model performance and translation; however, it cannot be used to make claims that one algorithm class is better than another or that deployment is equally advanced across diseases.

CONCLUSION

This review demonstrates that AI-powered analytics have considerable capacity to improve early disease detection in imaging, EHR, physiologic, wearable and biomarker-based applications in healthcare. The most compelling evidence supported models that were well-matched to the clinical data source, tested beyond the dataset of origin and developed to support specific tasks in the clinical workflow such as screening, triage, lesion detection or short-term risk prediction. Across applications, AI was most often able to improve early detection, reduce false negatives and improve the efficiency of diagnosis when deployed as a decision support tool integrated into the clinical workflow. Collectively, the evidence supports AI as a promising support for earlier and more targeted detection, while underscoring that robust external validation, local adaptation, and clinically grounded deployment are essential before widespread routine use.

ACKNOWLEDGMENT

In the preparation of this manuscript, we utilized Artificial Intelligence (AI) tools for content creation in the following capacity:

☐ None

x☐ Some sections, with minimal or no editing

☐ Some sections, with extensive editing

☐ Entire work, with minimal or no editing

☐ Entire work, with extensive editing

This article incorporates content generated by Artificial Intelligence (AI) tools where applicable. The use of AI tools complied with ethical standards and guidelines for academic integrity. The final content has been thoroughly reviewed and edited to ensure accuracy, relevance, and adherence to academic standards.

Conflict of Interest

The author declares no competing interests.

Funding

Not applicable.

REFERENCES

Abràmoff, M. D., Lavin, P. T., Birch, M., Shah, N., & Folk, J. C. (2018). Pivotal trial of an autonomous AI-based diagnostic system for detection of diabetic retinopathy in primary care offices. *NPJ digital medicine*, *1*, 39. https://doi.org/10.1038/s41746-018-0040-6

Ahmadi Mehr, R., & Ameri, A. (2022). Skin Cancer Detection Based on Deep Learning. *Journal of biomedical physics & engineering*, *12*(6), 559--568. https://doi.org/10.31661/jbpe.v0i0.2207-1517

Ahn, J. S., Ebrahimian, S., McDermott, S., Lee, S., Naccarato, L., Di Capua, J. F., Wu, M. Y., Zhang, E. W., Muse, V., Miller, B., Sabzalipour, F., Bizzo, B. C., Dreyer, K. J., Kaviani, P., Digumarthy, S. R., & Kalra, M. K. (2022). Association of Artificial Intelligence-Aided Chest Radiograph Interpretation With Reader Performance and Efficiency. *JAMA network open*, *5*(8), e2229289. https://doi.org/10.1001/jamanetworkopen.2022.29289

Alam, T. M., Shaukat, K., Khan, W. A., Hameed, I. A., Almuqren, L. A., Raza, M. A., Aslam, M., & Luo, S. (2022). An Efficient Deep Learning-Based Skin Cancer Classifier for an Imbalanced Dataset. *Diagnostics (Basel, Switzerland)*, *12*(9), 2115. https://doi.org/10.3390/diagnostics12092115

Al-Waisy, A. S., Al-Fahdawi, S., Khalaf, M. I., Mohammed, M. A., Al-Attar, B., & Al-Andoli, M. N. (2025). A deep learning framework for automated early diagnosis and classification of skin cancer lesions in dermoscopy images. *Scientific reports*, *15*(1), 31234. https://doi.org/10.1038/s41598-025-15655-9

Ardila, D., Kiraly, A. P., Bharadwaj, S., Choi, B., Reicher, J. J., Peng, L., Tse, D., Etemadi, M., Ye, W., Corrado, G., Naidich, D. P., & Shetty, S. (2019). End-to-end

lung cancer screening with three-dimensional deep learning on low-dose chest computed tomography. *Nature medicine*, *25*(6), 954--961. https://doi.org/10.1038/s41591-019-0447-x

Attia, Z. I., Noseworthy, P. A., Lopez-Jimenez, F., Asirvatham, S. J., Deshmukh, A. J., Gersh, B. J., Carter, R. E., Yao, X., Rabinstein, A. A., Erickson, B. J., Kapa, S., & Friedman, P. A. (2019). An artificial intelligence-enabled ECG algorithm for the identification of patients with atrial fibrillation during sinus rhythm: a retrospective analysis of outcome prediction. *Lancet (London, England)*, *394*(10201), 861--867. https://doi.org/10.1016/S0140-6736(1931721-0)

Bahrami, P., Tanbakuchi, D., Afzalaghaee, M., Ghayour-Mobarhan, M., & Esmaily, H. (2024). Development of risk models for early detection and prediction of chronic kidney disease in clinical settings. *Scientific reports*, *14*(1), 32136. https://doi.org/10.1038/s41598-024-83973-5

Bedoya, A. D., Futoma, J., Clement, M. E., Corey, K., Brajer, N., Lin, A., Simons, M. G., Gao, M., Nichols, M., Balu, S., Heller, K., Sendak, M., & O\'Brien, C. (2020). Machine learning for early detection of sepsis: an internal and temporal validation study. *JAMIA open*, *3*(2), 252--260. https://doi.org/10.1093/jamiaopen/ooaa006

Bellemo, V., Lim, Z. W., Lim, G., Nguyen, Q. D., Xie, Y., Yip, M. Y. T., Hamzah, H., Ho, J., Lee, X. Q., Hsu, W., Lee, M. L., Musonda, L., Chandran, M., Chipalo-Mutati, G., Muma, M., Tan, G. S. W., Sivaprasad, S., Menon, G., Wong, T. Y., & Ting, D. S. W. (2019). Artificial intelligence using deep learning to screen for referable and vision-threatening diabetic retinopathy in Africa: a clinical validation study. *The Lancet. Digital health*, *1*(1), e35--e44. https://doi.org/10.1016/S2589-7500(1930004-4)

Chang, YW., Ryu, J.K., An, J.K. *et al.* Artificial intelligence for breast cancer screening in mammography (AI-STREAM): preliminary analysis of a prospective multicenter cohort study. *Nat Commun* 16, 2248 (2025). https://doi.org/10.1038/s41467-025-57469-3

Chen, C. F., Hsu, C. H., Jiang, Y. C., Lin, W. R., Hong, W. C., Chen, I. Y., Lin, M. H., Chu, K. A., Lee, C. H., Lee, D. L., & Chen, P. F. (2024). A deep learning-based algorithm for pulmonary tuberculosis detection in chest radiography. *Scientific reports*, *14*(1), 14917. https://doi.org/10.1038/s41598-024-65703-z

Chen, Q., Cherry, D. R., Nalawade, V., Qiao, E. M., Kumar, A., Lowy, A. M., Simpson, D. R., & Murphy, J. D. (2021). Clinical Data Prediction Model to Identify Patients With Early-Stage Pancreatic Cancer. *JCO clinical cancer informatics*, *5*, 279--287. https://doi.org/10.1200/CCI.20.00137

Dai, L., Wu, L., Li, H., Cai, C., Wu, Q., Kong, H., Liu, R., Wang, X., Hou, X., Liu, Y., Long, X., Wen, Y., Lu, L., Shen, Y., Chen, Y., Shen, D., Yang, X., Zou, H., Sheng, B., & Jia, W. (2021). A deep learning system for detecting diabetic retinopathy across the disease spectrum. *Nature communications*, *12*(1), 3242. https://doi.org/10.1038/s41467-021-23458-5

Dubatovka, A., & Buhmann, J. M. (2022). Automatic Detection of Atrial Fibrillation from Single-Lead ECG Using Deep Learning of the Cardiac Cycle. *BME frontiers*, *2022*, 9813062. https://doi.org/10.34133/2022/9813062

Elhakim, M. T., Stougaard, S. W., Graumann, O., Nielsen, M., Gerke, O., Larsen, L. B., & Rasmussen, B. S. B. (2024). AI-integrated Screening to Replace Double Reading of Mammograms: A Population-wide Accuracy and Feasibility Study. *Radiology. Artificial intelligence*, *6*(6), e230529. https://doi.org/10.1148/ryai.230529

Esteva, A., Kuprel, B., Novoa, R. A., Ko, J., Swetter, S. M., Blau, H. M., & Thrun, S. (2017). Dermatologist-level classification of skin cancer with deep neural networks. *Nature*, *542*(7639), 115--118. https://doi.org/10.1038/nature21056

Glissen Brown, J. R., Mansour, N. M., Wang, P., Chuchuca, M. A., Minchenberg, S. B., Chandnani, M., Liu, L., Gross, S. A., Sengupta, N., & Berzin, T. M. (2022). Deep Learning Computer-aided Polyp Detection Reduces Adenoma Miss Rate: A United States Multi-center Randomized Tandem Colonoscopy Study (CADeT-CS Trial). *Clinical gastroenterology and hepatology : the official clinical practice journal of the American Gastroenterological Association*, *20*(7), 1499--1507.e4. https://doi.org/10.1016/j.cgh.2021.09.009

González-Bueno Puyal, J., Brandao, P., Ahmad, O. F., Bhatia, K. K., Toth, D., Kader, R., Lovat, L., Mountney, P., & Stoyanov, D. (2022). Polyp detection on video colonoscopy using a hybrid 2D/3D CNN. *Medical image analysis*, *82*, 102625. https://doi.org/10.1016/j.media.2022.102625

Gulshan, V., Peng, L., Coram, M., Stumpe, M. C., Wu, D., Narayanaswamy, A., Venugopalan, S., Widner, K., Madams, T., Cuadros, J., Kim, R., Raman, R., Nelson, P. C., Mega, J. L., & Webster, D. R. (2016). Development and Validation of a Deep Learning Algorithm for Detection of Diabetic Retinopathy in Retinal Fundus Photographs. *JAMA*, *316*(22), 2402--2410. https://doi.org/10.1001/jama.2016.17216

Hannun, A. Y., Rajpurkar, P., Haghpanahi, M., Tison, G. H., Bourn, C., Turakhia, M. P., & Ng, A. Y. (2019). Cardiologist-level arrhythmia detection and classification in ambulatory electrocardiograms using a deep neural network. *Nature medicine*, *25*(1), 65--69. https://doi.org/10.1038/s41591-018-0268-3

Hendrix, W., Hendrix, N., Scholten, E. T., Mourits, M., Trap-de Jong, J., Schalekamp, S., Korst, M., van Leuken, M., van Ginneken, B., Prokop, M., Rutten, M., & Jacobs, C. (2023). Deep learning for the detection of benign and malignant pulmonary nodules in non-screening chest CT scans. *Communications medicine*, *3*(1), 156. https://doi.org/10.1038/s43856-023-00388-5

Hiraoka, D., Inui, T., Kawakami, E., Oya, M., Tsuji, A., Honma, K., Kawasaki, Y., Ozawa, Y., Shiko, Y., Ueda, H., Kohno, H., Matsuura, K., Watanabe, M., Yakita, Y., & Matsumiya, G. (2022). Diagnosis of Atrial Fibrillation Using Machine Learning With Wearable Devices After Cardiac Surgery: Algorithm Development Study. *JMIR formative research*, *6*(8), e35396. https://doi.org/10.2196/35396

Hwang, E. J., Park, S., Jin, K. N., Kim, J. I., Choi, S. Y., Lee, J. H., Goo, J. M., Aum, J., Yim, J. J., Park, C. M., & Deep Learning-Based Automatic Detection Algorithm Development and Evaluation Group (2019). Development and Validation of a Deep Learning-based Automatic Detection Algorithm for Active Pulmonary Tuberculosis on Chest Radiographs. *Clinical infectious diseases : an official publication of the Infectious Diseases Society of America*, *69*(5), 739--747. https://doi.org/10.1093/cid/ciy967

Ipp, E., Liljenquist, D., Bode, B., Shah, V. N., Silverstein, S., Regillo, C. D., Lim, J. I., Sadda, S., Domalpally, A., Gray, G., Bhaskaranand, M., Ramachandra, C., Solanki, K., & EyeArt Study Group (2021). Pivotal Evaluation of an Artificial Intelligence System for Autonomous Detection of Referrable and Vision-Threatening Diabetic Retinopathy. *JAMA network open*, *4*(11), e2134254. https://doi.org/10.1001/jamanetworkopen.2021.34254

Jacobs, C., Setio, A. A. A., Scholten, E. T., Gerke, P. K., Bhattacharya, H., M Hoesein, F. A., Brink, M., Ranschaert, E., de Jong, P. A., Silva, M., Geurts, B., Chung, K., Schalekamp, S., Meersschaert, J., Devaraj, A., Pinsky, P. F., Lam, S. C., van Ginneken, B., & Farahani, K. (2021). Deep Learning for Lung Cancer Detection on Screening CT Scans: Results of a Large-Scale Public Competition and an Observer Study with 11 Radiologists. *Radiology. Artificial intelligence*, *3*(6), e210027. https://doi.org/10.1148/ryai.2021210027

Lam, C., Yi, D., Guo, M., & Lindsey, T. (2018). Automated Detection of Diabetic Retinopathy using Deep Learning. *AMIA Joint Summits on Translational Science proceedings. AMIA Joint Summits on Translational Science*, *2017*, 147--155.

Lång, K., Josefsson, V., Larsson, A. M., Larsson, S., Högberg, C., Sartor, H., Hofvind, S., Andersson, I., & Rosso, A. (2023). Artificial intelligence-supported screen reading versus standard double reading in the Mammography Screening with Artificial Intelligence trial (MASAI): a clinical safety analysis of a randomised, controlled, non-inferiority, single-blinded, screening accuracy study. *The Lancet. Oncology*, *24*(8), 936--944. https://doi.org/10.1016/S1470-2045(2300298-X)

Lee, J. Y., Jeong, J., Song, E. M., Ha, C., Lee, H. J., Koo, J. E., Yang, D. H., Kim, N., & Byeon, J. S. (2020). Real-time detection of colon polyps during colonoscopy using deep learning: systematic validation with four independent datasets. *Scientific reports*, *10*(1), 8379. https://doi.org/10.1038/s41598-020-65387-1

Lown, M., Brown, M., Brown, C., Yue, A. M., Shah, B. N., Corbett, S. J., Lewith, G., Stuart, B., Moore, M., & Little, P. (2020). Machine learning detection of Atrial Fibrillation using wearable technology. *PloS one*, *15*(1), e0227401. https://doi.org/10.1371/journal.pone.0227401

Masino, A. J., Harris, M. C., Forsyth, D., Ostapenko, S., Srinivasan, L., Bonafide, C. P., Balamuth, F., Schmatz, M., & Grundmeier, R. W. (2019). Machine learning models for early sepsis recognition in the neonatal intensive care unit using readily available electronic health record data. *PloS one*, *14*(2), e0212665. https://doi.org/10.1371/journal.pone.0212665

Matchaba, S., Fellague-Chebra, R., Purushottam, P., & Johns, A. (2023). Early Diagnosis of Pancreatic Cancer via Machine Learning Analysis of a National Electronic Medical Record Database. *JCO clinical cancer informatics*, *7*, e2300076. https://doi.org/10.1200/CCI.23.00076

McKinney, S. M., Sieniek, M., Godbole, V., Godwin, J., Antropova, N., Ashrafian, H., Back, T., Chesus, M., Corrado, G. S., Darzi, A., Etemadi, M., Garcia-Vicente, F., Gilbert, F. J., Halling-Brown, M., Hassabis, D., Jansen, S., Karthikesalingam, A., Kelly, C. J., King, D., Ledsam, J. R., ... Shetty, S. (2020). International evaluation of

an AI system for breast cancer screening. *Nature*, *577*(7788), 89--94. https://doi.org/10.1038/s41586-019-1799-6

Misra, D., Avula, V., Wolk, D. M., Farag, H. A., Li, J., Mehta, Y. B., Sandhu, R., Karunakaran, B., Kethireddy, S., Zand, R., & Abedi, V. (2021). Early Detection of Septic Shock Onset Using Interpretable Machine Learners. *Journal of clinical medicine*, *10*(2), 301. https://doi.org/10.3390/jcm10020301

Nemati, S., Holder, A., Razmi, F., Stanley, M. D., Clifford, G. D., & Buchman, T. G. (2018). An Interpretable Machine Learning Model for Accurate Prediction of Sepsis in the ICU. *Critical care medicine*, *46*(4), 547--553. https://doi.org/10.1097/CCM.0000000000002936

Oh, K., Kang, H. M., Leem, D., Lee, H., Seo, K. Y., & Yoon, S. (2021). Early detection of diabetic retinopathy based on deep learning and ultra-wide-field fundus images. *Scientific reports*, *11*(1), 1897. https://doi.org/10.1038/s41598-021-81539-3

Qasim Gilani, S., Syed, T., Umair, M., & Marques, O. (2023). Skin Cancer Classification Using Deep Spiking Neural Network. *Journal of digital imaging*, *36*(3), 1137--1147. https://doi.org/10.1007/s10278-023-00776-2

Rodriguez-Ruiz, A., Lång, K., Gubern-Merida, A., Broeders, M., Gennaro, G., Clauser, P., Helbich, T. H., Chevalier, M., Tan, T., Mertelmeier, T., Wallis, M. G., Andersson, I., Zackrisson, S., Mann, R. M., & Sechopoulos, I. (2019). Stand-Alone Artificial Intelligence for Breast Cancer Detection in Mammography: Comparison With 101 Radiologists. *Journal of the National Cancer Institute*, *111*(9), 916--922. https://doi.org/10.1093/jnci/djy222

Shen, L., Margolies, L. R., Rothstein, J. H., Fluder, E., McBride, R., & Sieh, W. (2019). Deep Learning to Improve Breast Cancer Detection on Screening Mammography. *Scientific reports*, *9*(1), 12495. https://doi.org/10.1038/s41598-019-48995-4

Showkatian, E., Salehi, M., Ghaffari, H., Reiazi, R., & Sadighi, N. (2022). Deep learning-based automatic detection of tuberculosis disease in chest X-ray images. *Polish journal of radiology*, *87*, e118--e124. https://doi.org/10.5114/pjr.2022.113435

Singh, V., Asari, V. K., & Rajasekaran, R. (2022). A Deep Neural Network for Early Detection and Prediction of Chronic Kidney Disease. *Diagnostics (Basel, Switzerland)*, *12*(1), 116. https://doi.org/10.3390/diagnostics12010116

Tschandl, P., Rinner, C., Apalla, Z., Argenziano, G., Codella, N., Halpern, A., Janda, M., Lallas, A., Longo, C., Malvehy, J., Paoli, J., Puig, S., Rosendahl, C., Soyer, H. P., Zalaudek, I., & Kittler, H. (2020). Human-computer collaboration for skin cancer recognition. *Nature medicine*, *26*(8), 1229--1234. https://doi.org/10.1038/s41591-020-0942-0

U.S. Food and Drug Administration. (2021). *Artificial intelligence and machine learning software as a medical device action plan*. U.S. Food and Drug Administration. Available at: https://www.fda.gov/media/145022/download

Wang, G., Yao, H., Gong, Y., Lu, Z., Pang, R., Li, Y., Yuan, Y., Song, H., Liu, J., Jin, Y., Ma, Y., Yang, Y., Nie, H., Zhang, G., Meng, Z., Zhou, Z., Zhao, X., Qiu, M., Zhao, Z., Jiang, K., ... Yin, Y. (2021). Metabolic detection and systems analyses of

pancreatic ductal adenocarcinoma through machine learning, lipidomics, and multi-omics. *Science advances*, *7*(52), eabh2724. https://doi.org/10.1126/sciadv.abh2724

Wang, P., Berzin, T. M., Glissen Brown, J. R., Bharadwaj, S., Becq, A., Xiao, X., Liu, P., Li, L., Song, Y., Zhang, D., Li, Y., Xu, G., Tu, M., & Liu, X. (2019). Real-time automatic detection system increases colonoscopic polyp and adenoma detection rates: a prospective randomised controlled study. *Gut*, *68*(10), 1813--1819. https://doi.org/10.1136/gutjnl-2018-317500

World Health Organization. (2021). *Ethics and governance of artificial intelligence for health: WHO guidance*. World Health Organization. Available at: https://iris.who.int/server/api/core/bitstreams/f780d926-4ae3-42ce-a6d6-e898a5562621/content

World Health Organization. (2022). *WHO operational handbook on tuberculosis. Module 2. Screening. Systematic screening for tuberculosis disease*. World Health Organization. Available at: https://iris.who.int/server/api/core/bitstreams/0499385e-45b0-429d-942e-9bd65bbaff2d/content