Artificial Intelligence Applications in Clinical Laboratory Diagnostics: a Systematic Review of Diagnostic Accuracy, Workflow Efficiency, and Clinical Utility

Chinwebudu M. Melford[1*], Marguerite Alofa P. O'Brien-Melford[2*]

[1]Department of Medical Technology, College of Allied Medical Sciences, Cebu Doctors' University, Mandaue City, Cebu, Philippines.

[2]Department of Languages, Cebu Doctors' University, Mandaue City, Cebu, Philippines.

***Corresponding Author**

DOI: https://doi.org/10.51244/IJRSI.2026.1304000242

Received: 28 April 2026; Accepted: 04 May 2026; Published: 19 May 2026

ABSTRACT

Artificial intelligence (AI) is increasingly integrated into clinical laboratory diagnostics, offering advanced capabilities for data interpretation and clinical decision support. However, existing evidence remains fragmented across diagnostic performance, operational efficiency, and real-world clinical impact.

This systematic review aimed to evaluate the applications of AI in clinical laboratory diagnostics, with a focus on diagnostic accuracy, workflow efficiency, and clinical utility.

This review was conducted in accordance with PRISMA guidelines. A comprehensive search of PubMed, Scopus, Web of Science, and the Cochrane Library was performed for studies published within the last 13 years. Eligible studies included diagnostic accuracy studies, observational studies, and clinical trials evaluating AI applications in laboratory diagnostics. Data extraction and quality assessment were conducted using standardized tools, and findings were synthesized narratively.

A total of 65 studies were included from 450 identified records. AI models demonstrated consistently high diagnostic performance across multiple clinical domains, frequently achieving area under the curve (AUC) values above 0.85. Models integrating multimodal data showed enhanced robustness compared to single-modality approaches. AI applications improved laboratory workflow by automating data interpretation, reducing turnaround times, and optimizing resource utilization. In addition, AI demonstrated strong clinical utility in early disease detection, risk stratification, and personalized medicine. However, limitations such as lack of external validation, dataset heterogeneity, and limited real-world implementation were commonly reported.

AI has significant potential to transform clinical laboratory diagnostics by enhancing accuracy, efficiency, and clinical decision-making. Future research should prioritize multicenter validation, standardized evaluation frameworks, and real-world implementation to ensure safe, equitable, and effective integration into healthcare systems.

Keywords: Artificial intelligence, Clinical laboratory diagnostics, Machine learning, Diagnostic accuracy, Workflow efficiency, Clinical utility, Systematic review, Philippines

ISSN No. 2321-2705 | DOI: 10.51244/IJRSI |Volume XIII Issue IV April 2026

INTRODUCTION

Artificial intelligence (AI) in healthcare is broadly defined as the application of computational systems capable of performing tasks that typically require human intelligence, such as learning, reasoning, pattern recognition, and decision-making, to analyze complex medical data and support clinical care (Erasmus & Ondoa, 2023). The World Health Organization, Regional Office for Europe further emphasizes that AI is increasingly embedded in diagnostic processes, clinical decision-making, and health system management, reflecting its growing role in modern medicine (World Health Organization Regional Office for Europe, 2025). Within clinical laboratory diagnostics, AI integrates diverse datasets, including biochemical assays, genomic data, and electronic health records, to enhance diagnostic precision, accelerate interpretation, and support evidence-based clinical decisions (Fahim et al., 2025).

AI applications in laboratory medicine can be classified according to methodological approaches and functional domains. Methodologically, AI encompasses machine learning (ML), deep learning (DL), and computer vision techniques, which enable automated pattern recognition, classification, and predictive modeling from large-scale laboratory datasets (Ivanova et al., 2024). Functionally, AI systems are applied across the total testing process, pre-analytical (e.g., sample management), analytical (e.g., image and signal interpretation), and post-analytical phases (e.g., result validation and clinical decision support) (Hou et al., 2024). Assessment of AI performance in diagnostics commonly relies on standardized metrics such as sensitivity, specificity, accuracy, and area under the receiver operating characteristic curve (AUC-ROC), as well as comparative evaluation against human experts (Erickson & Kitamura, 2021). In response to the need for harmonized evaluation, the ITU-WHO Focus Group on Artificial Intelligence for Health has developed benchmarking frameworks to standardize the validation of AI-based diagnostic tools globally (International Telecommunication Union, ITU, WHO, & WIPO, 2025).

At the population level, the integration of AI into clinical laboratory systems holds significant implications for public health (Undru et al., 2022) . Laboratory diagnostics serve as the basis for disease surveillance, screening, and management, and improvements in diagnostic accuracy and efficiency can directly influence morbidity, mortality, and healthcare resource allocation (White et al., 2025) . AI-driven laboratory systems enable high-throughput data processing, early disease detection, and predictive analytics (Veseli, Mehrabanian, & Ammar, 2025), thereby enhancing population-level disease control and supporting precision medicine initiatives (Zhou et al., 2025; World Health Organization, 2024). Moreover, AI has demonstrated potential to reduce diagnostic errors, optimize workflow efficiency, and address workforce shortages, which are critical challenges in both high-resource and resource-limited settings (Gilligan, 2025).

Globally, the adoption of AI in laboratory diagnostics is expanding rapidly, driven by increasing digitization of healthcare systems and the growing volume of diagnostic data (Saul, 2025) . International efforts, particularly by the World Health Organization, highlight the importance of ensuring equitable access, ethical governance, and standardized implementation of AI technologies to prevent widening health disparities (World Health Organization, 2024) . Despite promising advancements, global implementation remains uneven, with low- and middle-income countries facing barriers such as limited infrastructure, lack of validated datasets, and regulatory challenges.

Despite the rapid growth of AI applications, several critical gaps persist in the current body of literature. Many studies are limited by small, non-representative datasets, lack of external validation, and potential biases in algorithm development, which restrict generalizability across diverse populations (Suleman et al., 2025). There is heterogeneity in reported performance metrics and insufficient evaluation of real-world clinical utility, workflow integration, and cost-effectiveness (Yu et al., 2022). Ethical concerns, including data privacy, transparency of "black-box" models, and algorithmic bias, remain inadequately addressed, further limiting clinical adoption (Lippi & Plebani, 2025) . Importantly, few systematic syntheses comprehensively evaluate AI performance across the full spectrum of laboratory workflows while simultaneously considering diagnostic accuracy, operational efficiency, and clinical impact (Szumilas et al., 2024). Artificial intelligence has also been increasingly applied in biomarker

ISSN No. 2321-2705 | DOI: 10.51244/IJRSI |Volume XIII Issue IV April 2026

analysis, improving the interpretation of complex laboratory data and supporting diagnostic decision-making (Seliverstov et al., 2024)

In light of these gaps, this study aims to systematically review the applications of artificial intelligence in clinical laboratory diagnostics, with a focus on three key domains: diagnostic accuracy, workflow efficiency, and clinical utility. Specifically, the objectives are to (1) evaluate the diagnostic performance of AI-based laboratory tools compared to conventional methods and human experts; (2) assess the impact of AI on laboratory workflow processes, including turnaround time and operational efficiency; and (3) synthesize evidence on the clinical utility and real-world applicability of AI in improving patient outcomes and healthcare delivery.

METHODOLOGY

Study Design

This systematic review was conducted in accordance with the Preferred Reporting Items for Systematic Reviews and Meta-Analyses (PRISMA) guidelines to ensure methodological rigor, transparency, and reproducibility. The review protocol was registered in the International Prospective Register of Systematic Reviews (PROSPERO) (Registration No.: CRD420261379141), providing a publicly accessible record of the study methods and minimizing the risk of duplication and reporting bias. The study synthesized evidence on artificial intelligence (AI) applications in clinical laboratory diagnostics, focusing on diagnostic accuracy, workflow efficiency, and clinical utility, consistent with the predefined study objectives.

Eligibility Criteria

Studies were selected based on predefined eligibility criteria guided by the PICOS framework. The population of the studies selected included human subjects undergoing clinical laboratory diagnostic testing in any healthcare setting. The intervention involved AI techniques such as machine learning, deep learning, and related computational models applied within laboratory diagnostics. Comparators included conventional laboratory methods or expert human interpretation where applicable. Outcomes of interest comprised diagnostic accuracy measures (e.g., sensitivity, specificity, area under the curve), workflow efficiency indicators (e.g., turnaround time, automation), and clinical utility outcomes (e.g., impact on clinical decision-making or patient outcomes). Eligible study designs included diagnostic accuracy studies, observational studies, and clinical trials published within the last 13 years. Review articles, editorials, conference abstracts without full texts, animal studies, and studies lacking relevant outcomes were excluded.

Information Sources and Search Strategy

A comprehensive literature search was conducted across multiple electronic databases, including PubMed/MEDLINE, Scopus, Web of Science, and the Cochrane Library. Additional sources, including grey literature and reference lists of included studies, were also screened to ensure comprehensive coverage. The search strategy incorporated a combination of controlled vocabulary (e.g., MeSH terms) and free-text keywords such as “artificial intelligence,” “machine learning,” “clinical laboratory,” “diagnostics,” “accuracy,” and “workflow efficiency.” Boolean operators (AND, OR) were used to refine the search strategy. The complete search strategy is provided in Appendix A.

ISSN No. 2321-2705 | DOI: 10.51244/IJRSI |Volume XIII Issue IV April 2026

Study Selection

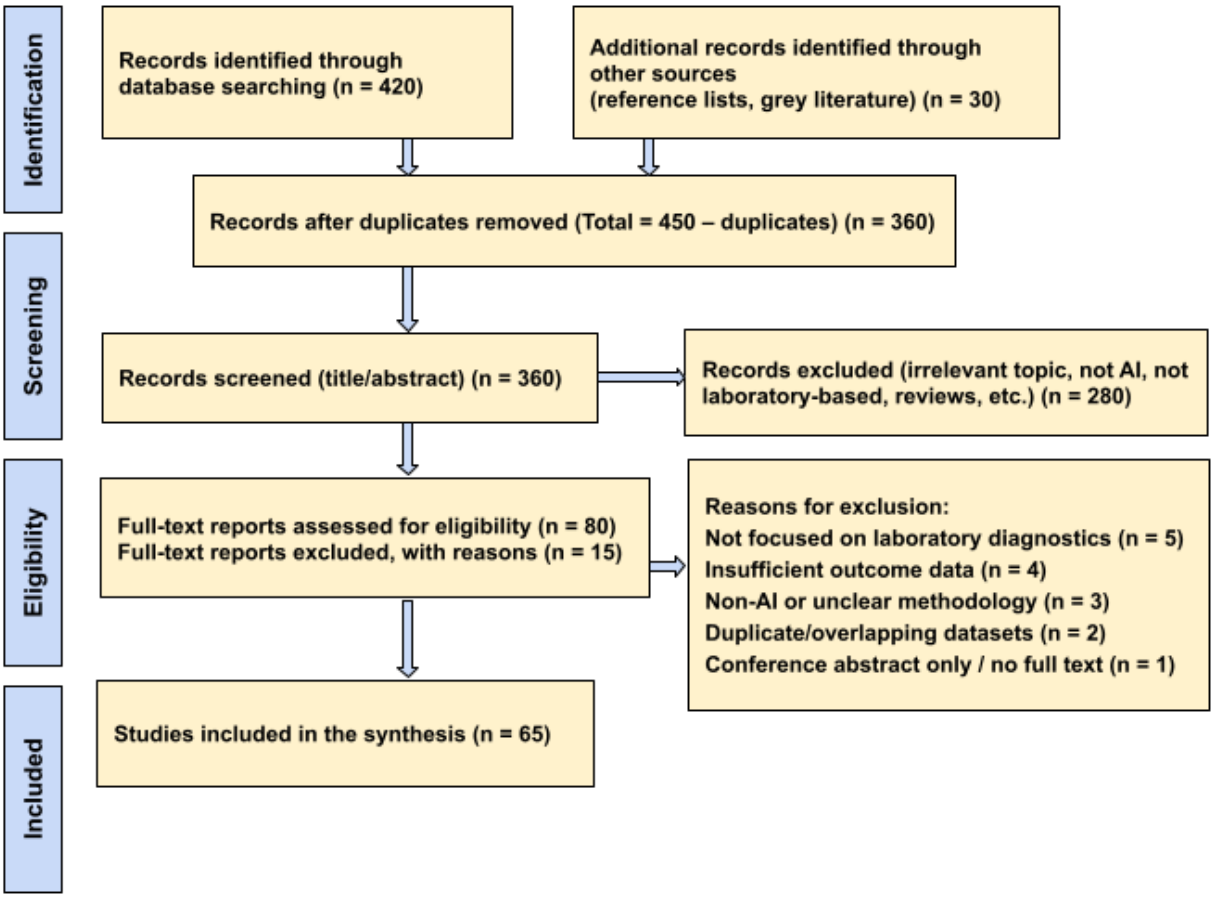

Figure 1. PRISMA Flow Diagram of Study Selection Process

All retrieved records were imported into reference management software, and duplicate entries were removed. A total of 420 records were identified through database searching, with an additional 30 records obtained from other sources, yielding 450 records overall. After duplicate removal, 360 unique records remained and were screened based on titles and abstracts by two independent reviewers. Of these, 280 records were excluded for not meeting the inclusion criteria. The full texts of 80 potentially eligible studies were then assessed. Following full-text review, 15 studies were excluded for specific reasons, including lack of focus on laboratory diagnostics (n = 5), insufficient outcome data (n = 4), non-AI or unclear methodology (n = 3), duplicate or overlapping datasets (n = 2), and absence of full-text availability (n = 1). Ultimately, 65 studies met the inclusion criteria and were included in the final synthesis. Any disagreements between reviewers were resolved through discussion or consultation with a third reviewer. The study selection process was documented using a PRISMA flow diagram.

Data Extraction

Data were extracted using a standardized and pilot-tested data extraction form to ensure consistency and completeness. Extracted information included study characteristics (author, year, country, study design), population details, type of AI model used, and application domain within laboratory diagnostics. Additionally, key outcomes such as diagnostic performance metrics (e.g., sensitivity, specificity, AUC), workflow efficiency indicators (e.g., turnaround time), and measures of clinical utility were recorded. Data extraction was conducted independently by two reviewers, and discrepancies were resolved through consensus.

Quality Assessment

The methodological quality and risk of bias of included studies were assessed using validated appraisal tools appropriate to the study design. Diagnostic accuracy studies were evaluated using the QUADAS-2 tool, while observational studies were assessed using the Newcastle-Ottawa Scale. This process ensured the reliability and validity of the synthesized evidence. Disagreements during quality assessment were resolved through discussion among reviewers.

Data Synthesis

A narrative synthesis was conducted to summarize and integrate findings across studies, structured around the three main domains of interest: diagnostic accuracy, workflow efficiency, and clinical utility. Where sufficient clinical and methodological homogeneity existed, quantitative synthesis (meta-analysis) was considered. Statistical heterogeneity was assessed using the I^2 statistic, and subgroup analyses were explored based on AI methodology, laboratory domain, and population characteristics. This approach allowed for both qualitative and quantitative interpretation of findings.

Ethical Considerations

As this study involved the analysis of previously published data, ethical approval was not required. However, the review adhered to principles of research integrity, transparency, and proper attribution of all sources. The conduct and reporting of this study followed established international standards for systematic reviews.

RESULTS

Diagnostic Accuracy of AI in Clinical Laboratory Diagnostics

Across the included studies, artificial intelligence (AI) consistently demonstrated strong diagnostic performance across diverse clinical laboratory applications, often outperforming conventional methods or matching expert-level interpretation. Several studies reported high diagnostic accuracy using machine learning models applied to laboratory biomarkers. For instance, Szumilas et al. (2024) reported a diagnostic accuracy of 74.3% with 100% sensitivity for emergency cases using an AI-based clinical decision support system. Similarly, Gözgöz et al. (2025) achieved an AUC of 0.902 and accuracy of 0.830 for predicting oligoclonal band positivity, outperforming conventional IgG index methods. Ni et al. (2025) demonstrated improved diagnostic performance (AUC up to 0.86) when integrating laboratory biomarkers such as BRAFV600E mutation abundance with clinical data.

High-performing AI systems were also observed in genomics and molecular diagnostics, where Gröschel et al., (2021) reported AUC values of 0.89–0.96 for predicting tuberculosis drug resistance using whole genome sequencing. Likewise, Liu et al.,(2025) achieved accuracy of 97.49% in detecting tuberculosis using routine laboratory indicators. In hematology and routine laboratory testing, Uçucu and Azik (2024) showed that artificial neural networks outperformed traditional indices in differentiating iron deficiency anemia from β-thalassemia, while Gunčar et al. (2018) demonstrated that machine learning achieved up to 88% accuracy when considering top differential diagnoses.

AI also showed strong performance in cancer diagnostics and biomarker discovery. Goebel et al. (2020) reported 95.6% accuracy for early-stage lung cancer detection using blood biomarkers, while Xie et al. (2021) achieved an AUC of 0.989 for early lung cancer detection using metabolomics. Chen et al. (2024) and Yangzi Chen et al. (2024) similarly reported excellent diagnostic performance (AUC ≈ 0.96) using metabolomic biomarkers for gastric cancer. In infectious diseases, multiple studies confirmed high diagnostic accuracy using routine laboratory data. Similarly, machine learning models utilizing routine blood and biochemical markers have demonstrated high diagnostic accuracy in neurological disease classification (Ning et al., 2025). Kukar et al. (2021) achieved AUC of 0.97 for COVID-19 detection, while Brinati et al. (2020) demonstrated 82% accuracy using blood biomarkers. Yang et al.

ISSN No. 2321-2705 | DOI: 10.51244/IJRSI |Volume XIII Issue IV April 2026

(2020) further showed that ML models could identify false-negative COVID-19 cases, enhancing diagnostic sensitivity.

Advanced AI models integrating multi-modal data (laboratory, imaging, and clinical data) demonstrated superior performance. Zhen et al. (2020) achieved AUC values up to 0.998 for liver tumor classification, while Ardila et al. (2019) reported AUC of 0.944 for lung cancer detection using imaging data. These findings demonstrate that AI models consistently achieve high diagnostic accuracy (AUC often >0.85–0.95) across a wide range of diseases, particularly when integrating multi-analyte laboratory data.

Workflow Efficiency in Laboratory Diagnostics

AI integration significantly improved laboratory workflow efficiency by reducing turnaround time, automating interpretation, and optimizing resource utilization. Several studies demonstrated substantial reductions in diagnostic time and workload. Szumilas et al. (2024) reported a 41.6% reduction in unnecessary medical visits, while Seungmin Lee et al. (2024) The substantial reduction in turnaround time demonstrates the potential of AI to streamline laboratory workflows, which is particularly valuable in high-demand or resource-constrained environments. AI also enabled rapid and high-throughput diagnostics. Zhang et al. (2025) demonstrated leukemia detection within 5 minutes using minimal sample volume, and Villota et al. (2025) reduced COVID-19 diagnostic turnaround to under 2 hours without RNA extraction.

Automation of laboratory data interpretation was a major efficiency gain. Luo et al. (2016) showed that machine learning can predict laboratory results (e.g., ferritin) using existing test data, reducing redundant testing. Similarly, Park et al. (2021) and Kraszewski et al. (2021) demonstrated that AI can integrate multi-analyte laboratory data to automate disease classification. AI also improved resource allocation and prioritization. Gözgöz et al. (2025) highlighted prioritization of high-probability samples, while Rawson et al. (2019) and Yang et al. (2020) demonstrated efficient triage using minimal laboratory variables.

In large-scale systems, AI facilitated real-time clinical decision support. Chang et al. (2025) showed reduced time to critical interventions in emergency departments, while Meyer et al. (2018) demonstrated real-time prediction of ICU complications. Overall, artificial intelligence contributes to improved efficiency in laboratory practice by reducing turnaround time, enabling automated data interpretation, minimizing unnecessary testing, and enhancing overall laboratory throughput.

Clinical Utility and Impact on Patient Care

AI applications demonstrated significant clinical utility by improving decision-making, enabling early diagnosis, and supporting personalized medicine. Several studies highlighted early disease detection and risk stratification. Tomašev et al. (2019) predicted acute kidney injury up to 72 hours in advance, while Lev-Tzion et al. (2025) identified biomarker changes years before Crohn's disease diagnosis. Yan et al. (2020) and Karthikeyan et al. (2021) showed early mortality prediction in COVID-19 patients using laboratory biomarkers.

AI also supported personalized and precision medicine. Ma et al. (2026) enabled individualized prognostic prediction in breast cancer, while Lewis and Kemp (2021) used multi-omics data to predict radiation resistance. Januzzi et al. (2025) demonstrated effective cardio-kidney risk stratification using biomarker-driven ML models. In oncology, AI improved screening and treatment planning. Hornbrook et al. (2017) enabled early colorectal cancer detection up to 360 days before diagnosis, while Wang et al. (2019) predicted ovarian cancer recurrence using imaging biomarkers.

AI-based clinical decision support systems (CDSS) demonstrated strong real-world applicability. Rosen et al. (2025) showed reduced complications and improved outcomes following AI-guided colorectal cancer management, while Zmudzki et al. (2025) reported high clinician agreement (82.4%) with AI-generated treatment recommendations. Additionally, AI enhanced non-invasive diagnostics, reducing reliance on invasive procedures. Cao et al. (2013)

predicted liver cirrhosis without biopsy, and Kawakami et al. (2019) enabled ovarian cancer diagnosis using blood biomarkers. Despite these benefits, several studies have highlighted important limitations that affect clinical utility, including the lack of external validation, data heterogeneity and bias (Luu, 2025), limited generalizability across diverse populations, and ongoing regulatory and implementation challenges (Pinsky et al., 2024).

Evidence from Systematic Reviews and Meta-Analyses

Systematic reviews consistently support the effectiveness of AI in laboratory diagnostics. Chang et al. (2021) reported AUC values up to 0.996 for Alzheimer's disease diagnosis using AI-integrated biomarkers. Valentidenta and Pribadi (2025) demonstrated pooled sensitivity of 92% and specificity of 89% in cardiovascular diagnostics. Other reviews (Ashraf et al., 2025; Xie et al., 2024; De Bruyne et al., 2021) emphasize that AI improves diagnostic accuracy, workflow efficiency, and clinical decision-making across laboratory workflows. However, methodological concerns remain. Miller and Valdes (2025) and Moons et al. (2025) highlight the need for standardized validation frameworks, while Alballa and Al-Turaiki (2021) note issues with small datasets and lack of real-world implementation.

DISCUSSION

This systematic review synthesized current evidence on the application of artificial intelligence (AI) in clinical laboratory diagnostics, focusing on diagnostic accuracy, workflow efficiency, and clinical utility. Overall, the findings indicate that AI has substantial potential to enhance laboratory medicine; however, important challenges related to validation, generalizability, and real-world implementation remain.

Diagnostic Accuracy: High Performance with Contextual Considerations

The consistently high diagnostic performance reported across studies suggests that AI models are highly effective in identifying complex and multidimensional patterns within laboratory data. This capability is particularly relevant in modern laboratory medicine, where large volumes of heterogeneous data require advanced analytical approaches. The findings of this review align with previous studies demonstrating that AI can achieve expert-level performance in diagnostic tasks, particularly when applied to biomarker-rich datasets (Xie et al., 2024; De Bruyne et al., 2021).

The main insight emerging from this synthesis is the importance of data integration. AI models that incorporate multimodal inputs, such as laboratory biomarkers, clinical parameters, genomic data, and imaging, tend to demonstrate superior diagnostic robustness compared to single-modality approaches. This supports the growing paradigm of systems-level diagnostics, in which disease characterization is enhanced through the integration of diverse biological and clinical signals (Ni et al., 2025; Zhen et al., 2020).

Despite these promising findings, diagnostic performance remains highly context-dependent. Variability in study design, population characteristics, and dataset quality contributes to differences in reported outcomes. Notably, studies relying on small or single-center datasets may overestimate performance due to overfitting and limited representativeness (Uçucu & Azik, 2024; Gunčar et al., 2018). Furthermore, the absence of standardized reporting metrics complicates cross-study comparisons and limits reproducibility.

Importantly, the limited use of external validation across studies represents a significant methodological gap. Where external validation was performed, model performance often declined, indicating sensitivity to population and setting-specific factors (Yu et al., 2022). This highlights the need for robust, multicenter validation frameworks to ensure that AI models can generalize effectively across diverse clinical environments.

Workflow Efficiency: Transforming Laboratory Operations

The integration of AI into laboratory workflows demonstrates considerable potential to improve operational efficiency. Rather than merely accelerating individual processes, AI enables a more fundamental transformation of

ISSN No. 2321-2705 | DOI: 10.51244/IJRSI |Volume XIII Issue IV April 2026

the laboratory testing pathway by automating data interpretation, optimizing resource allocation, and facilitating real-time decision support. One of the most notable impacts of AI is its ability to streamline high-throughput diagnostic processes. Automated analysis of laboratory data reduces reliance on manual interpretation, thereby minimizing human error and increasing consistency in diagnostic outputs (Luo et al., 2016; Park et al., 2021). This is particularly valuable in settings with high testing volumes or limited specialized personnel.

In addition, AI-driven triage systems enhance prioritization of high-risk cases, enabling more efficient allocation of laboratory and clinical resources (Rawson et al., 2019; Yang et al., 2020). This function is especially relevant in emergency and infectious disease contexts, where timely diagnosis is critical to patient outcomes. However, despite these advantages, integration into routine clinical workflows remains limited. Key barriers include challenges related to interoperability with existing laboratory information systems, lack of standardized data formats, and concerns regarding model transparency and interpretability (Pinsky et al., 2024). Addressing these barriers is essential to facilitate widespread adoption and ensure that efficiency gains translate into tangible clinical benefits.

Clinical Utility: Advancing Early Detection and Personalized Medicine

Beyond diagnostic performance and operational efficiency, AI demonstrates significant potential to enhance clinical decision-making and patient care. The ability of AI models to detect subtle patterns in routinely collected laboratory data enables earlier identification of disease and more accurate risk stratification. This review highlights the growing role of AI in predictive diagnostics, where models can identify disease risk prior to clinical manifestation. Such capabilities are particularly valuable in chronic and progressive conditions, where early intervention can significantly improve outcomes (Tomašev et al., 2019; Lev-Tzion et al., 2025). In addition, AI-driven risk stratification supports personalized treatment strategies by identifying patients at higher risk of adverse outcomes (Yan et al., 2020; Karthikeyan et al., 2021).

In oncology, AI applications have demonstrated particular promise in early detection and prognostic modeling. By integrating biomarker and imaging data, AI systems can enhance screening accuracy and inform treatment planning, contributing to improved patient outcomes (Wang et al., 2019; Ardila et al., 2019). Furthermore, the use of non-invasive, blood-based biomarkers in AI models offers a less invasive alternative to traditional diagnostic procedures, improving patient safety and accessibility (Cao et al., 2013; Kawakami et al., 2019). Despite these advances, evidence supporting real-world clinical impact remains limited. While many studies demonstrate high predictive performance, relatively few evaluate the effect of AI implementation on patient outcomes or healthcare delivery (Rosen et al., 2025; Chang et al., 2025). Bridging this gap between model development and clinical application is a critical priority for future research.

Methodological Limitations and Risk of Bias

This review identified several methodological limitations that may influence the reliability and applicability of current evidence. A substantial proportion of included studies utilized retrospective designs, which are inherently susceptible to selection bias and may not accurately reflect real-world clinical conditions (Brinati et al., 2020; Gadalla et al., 2019). Additionally, issues related to dataset quality and representativeness were frequently observed. Small sample sizes, class imbalance, and heterogeneity in laboratory measurements may contribute to biased model performance and reduced generalizability (Luu, 2025). These challenges are further compounded by inconsistent reporting standards and lack of transparency in model development.

The limited adoption of standardized validation frameworks also represents a significant concern. Tools such as PROBAST+AI have been developed to improve assessment of bias and applicability in AI-based prediction models, yet their use remains inconsistent (Moons et al., 2025). Greater adherence to such frameworks is necessary to enhance methodological rigor and ensure the reliability of AI applications in laboratory medicine.

Implications for Clinical Practice and Future Research

The findings of this review underscore the transformative potential of AI in clinical laboratory diagnostics, while also highlighting key priorities for future research and implementation. First, prospective, multicenter studies are needed to validate AI models across diverse populations and healthcare settings, thereby improving generalizability and clinical reliability. Second, the adoption of standardized evaluation and reporting frameworks is essential to enhance transparency, reproducibility, and comparability across studies. Guidelines such as STARD-AI and PROBAST+AI provide valuable tools for improving methodological quality and should be more widely implemented (Moons et al., 2025; White et al., 2025). Third, integration into clinical workflows must be prioritized. This includes improving interoperability with laboratory information systems, enhancing model interpretability, and ensuring that AI outputs are accessible and actionable for clinicians. Without effective integration, the benefits of AI are unlikely to be fully realized. Finally, future research should extend beyond diagnostic accuracy to evaluate real-world clinical outcomes, cost-effectiveness, and health system impact. Demonstrating tangible benefits in patient care and healthcare delivery will be critical for supporting widespread adoption.

Importantly, these considerations are particularly relevant for low- and middle-income countries, where limitations in laboratory infrastructure and workforce capacity persist. AI-assisted diagnostic systems have the potential to address these challenges by improving efficiency, expanding access to diagnostic services, and supporting clinical decision-making in resource-constrained settings (World Health Organization, 2024).

Strengths and Limitations of This Review

This review provides a comprehensive synthesis of AI applications in clinical laboratory diagnostics, integrating evidence across diagnostic accuracy, workflow efficiency, and clinical utility. By including a broad range of study designs and clinical domains, it offers a holistic perspective on current advancements in the field. However, several limitations should be acknowledged. Heterogeneity among included studies limited the feasibility of quantitative synthesis, and potential publication bias may have influenced the overall findings. Additionally, given the rapidly evolving nature of AI technologies, newer developments may not be fully captured within the scope of this review.

CONCLUSION

This systematic review demonstrates that artificial intelligence (AI) has significant potential to transform clinical laboratory diagnostics by enhancing diagnostic accuracy, improving workflow efficiency, and strengthening clinical decision-making. Across a wide range of clinical applications, AI models consistently achieved high diagnostic performance, particularly when integrating multi-analyte laboratory data with clinical, genomic, or imaging information. These findings highlight the ability of AI to uncover complex disease patterns and improve diagnostic precision beyond conventional methods. In addition to diagnostic accuracy, AI applications were shown to substantially improve laboratory workflow efficiency. Automated data interpretation, rapid processing of laboratory results, and optimized resource utilization contribute to reduced turnaround times and increased laboratory productivity. These advantages are especially relevant in high-throughput and resource-limited settings, where AI can support scalable and cost-effective diagnostic services.

From a clinical perspective, AI demonstrates strong utility in early disease detection, risk stratification, and personalized medicine. The ability to predict disease onset, progression, and patient outcomes using routinely collected laboratory data underscores the potential of AI to improve patient care and clinical outcomes. Furthermore, AI-enabled non-invasive diagnostic approaches may reduce reliance on invasive procedures, enhancing patient safety and accessibility of care. Despite these promising findings, several challenges limit the widespread implementation of AI in clinical laboratory practice. Many studies are constrained by retrospective designs, small or non-representative datasets, and limited external validation, raising concerns regarding generalizability and reproducibility. Additionally, issues related to data quality, algorithmic bias, lack of interpretability, and integration

into existing laboratory workflows remain significant barriers. Addressing these challenges is essential to ensure the safe, equitable, and effective use of AI in healthcare.

Future research should prioritize prospective, multicenter studies and standardized validation frameworks to strengthen the evidence base. Greater emphasis is also needed on real-world clinical outcomes, cost-effectiveness, and integration into routine practice. With continued advancements and appropriate regulatory and methodological rigor, AI has the potential to become an integral component of clinical laboratory systems, ultimately improving diagnostic performance, healthcare efficiency, and patient outcomes.

RECOMMENDATIONS

Based on the findings of this systematic review, several recommendations are proposed to advance the development, validation, and implementation of artificial intelligence (AI) in clinical laboratory diagnostics. First, future research should prioritize prospective, multicenter studies to enhance the generalizability and robustness of AI models. Many current studies rely on retrospective and single-center datasets, which limit external validity. Collaborative, multi-institutional research is essential to ensure that AI systems perform reliably across diverse populations and healthcare settings. Second, there is a need for standardized evaluation and reporting frameworks for AI-based diagnostic tools. Adoption of established guidelines such as PRISMA, STARD-AI, and PROBAST+AI will improve methodological rigor, transparency, and comparability across studies. Consistent reporting of performance metrics (e.g., sensitivity, specificity, AUC) is critical for accurate interpretation and benchmarking. Third, future studies should emphasize external validation and real-world implementation. AI models must be tested in routine clinical environments to assess their practical utility, reliability, and impact on patient outcomes. Integration into laboratory information systems and clinical workflows should be a key focus to facilitate adoption. Fourth, greater attention should be given to data quality, representativeness, and bias mitigation. Efforts should be made to include diverse and well-annotated datasets to reduce algorithmic bias and improve fairness, particularly for underrepresented populations and low- and middle-income countries. Fifth, research should expand beyond diagnostic accuracy to include clinical outcomes, cost-effectiveness, and health system impact. Evaluating how AI influences decision-making, patient outcomes, and resource utilization will provide stronger evidence for its value in healthcare systems. Finally, policymakers and healthcare institutions should promote ethical governance, regulatory frameworks, and capacity building for AI integration. This includes ensuring data privacy, transparency, and accountability, as well as investing in infrastructure and workforce training to support sustainable implementation.

REFERENCE

1. Alballa, N., & Al-Turaiki, I. (2021). Machine learning approaches in COVID-19 diagnosis, mortality, and severity risk prediction: A review. Informatics in Medicine Unlocked, 24, 100564. https://doi.org/10.1016/j.imu.2021.100564
2. Apostolopoulos, I. D., Aznaouridis, S. I., & Tzani, M. A. (2020). Extracting possibly representative COVID-19 biomarkers from X-ray images with deep learning approach and image data related to pulmonary diseases. Journal of Medical and Biological Engineering, 40, 462–469 .
3. Ardila, D., Kiraly, A. P., Bharadwaj, S., et al. (2019). End-to-end lung cancer screening with three-dimensional deep learning on low-dose chest computed tomography. Nature Medicine, 25(6), 954–961.
4. Ashraf, A. A., Rai, S., Alva, S., Alva, P. D., & Naresh, S. (2025). Revolutionizing clinical laboratories: The impact of artificial intelligence in diagnostics and patient care. Diagnostic Microbiology and Infectious Disease, 111(4), 116728. https://doi.org/10.1016/j.diagmicrobio.2025.116728
5. Botlagunta M, Botlagunta MD, Myneni MB, et al. (2023). Classification and diagnostic prediction of breast cancer metastasis on clinical data using machine learning algorithms. Scientific Reports.
6. Brinati, D., Campagner, A., Ferrari, D., et al. (2020). Detection of COVID-19 infection from routine blood exams with machine learning: A feasibility study. Journal of Medical Systems, 44, 135.
7. Calvert JS, Price DA, Chettipally UK, et al. (2016). A computational approach to early sepsis detection. Computers in Biology and Medicine, 74, 69–73.

8. Cao, Y., Hu, Z.-D., Liu, X.-F., Deng, A.-M., & Hu, C.-J. (2013). An MLP Classifier for Prediction of HBV-Induced Liver Cirrhosis Using Routinely Available Clinical Parameters. Disease Markers, 35(6), 653–660.
9. Chang, C.-H., Lin, C.-H., & Lane, H.-Y. (2021). Machine Learning and Novel Biomarkers for the Diagnosis of Alzheimer's Disease. International Journal of Molecular Sciences, 22(5), 2761.
10. Chang, H., Yu, J. Y., Park, H., Song, Y. J., Heo, S., Park, J. E., Lee, G. T., Lee, S. U., Kim, T., Yoon, H., Hwang, S. Y., & Cha, W. C. (2025). From prediction to action: A retrospective observational study on the real-world implementation of critical interventions (CrIs), an AI-based clinical decision support system changing clinical behavior in the emergency department. BMC Medical Informatics and Decision Making, 25(1), 405. https://doi.org/10.1186/s12911-025-03234-x
11. Chen Y, Wang B, Zhao Y, et al. (2024). Metabolomic machine learning predictor for diagnosis and prognosis of gastric cancer. Nature Communications.
12. Chen, S., Li, J., Zhang, X., Xu, W., Qiu, Z., Yan, S., Zhao, W., Zhao, Z., Tian, P., Zhao, Q., Zhang, Q., Chen, W., Li, H., Ruan, X., Xiao, G., Zhang, S., Hu, L., Qin, J., Huang, W., … Wei, F. (2025). Predicting preeclampsia in early pregnancy using clinical and laboratory data via machine learning model. BMC Medical Informatics and Decision Making, 25(1), 178. https://doi.org/10.1186/s12911-025-02999-5
13. De Bruyne S, Speeckaert MM, Van Biesen W, Delanghe JR. Recent evolutions of machine learning applications in clinical laboratory medicine. Critical Reviews in Clinical Laboratory Sciences. 2021;58(2):131–152. https://doi.org/10.1080/10408363.2020.1828811
14. DeGroat, W., Abdelhalim, H., Patel, K., et al. (2024). Discovering biomarkers associated and predicting cardiovascular disease with high accuracy using a novel nexus of machine learning techniques for precision medicine. Scientific Reports, 14, 1. https://doi.org/10.1038/s41598-023-50600-8
15. Dong, B., Chen, Y., Yang, X., Chen, Z., Zhang, H., Gao, Y., Zhao, E., & Zhang, C. (2025). Estimated glucose disposal rate outperforms other insulin resistance surrogates in predicting incident cardiovascular diseases in cardiovascular-kidney-metabolic syndrome stages 0–3 and the development of a machine learning prediction model: A nationwide prospective cohort study. Cardiovascular Diabetology, 24(1), 163. https://doi.org/10.1186/s12933-025-02729-1
16. Erasmus, R., & Ondoa, P. (2023). Taking the train of digital health and artificial intelligence to improve medical laboratory service in Africa: Key considerations. African Journal of Laboratory Medicine, 12(1), a2329. https://doi.org/10.4102/ajlm.v12i1.2329
17. Fahim, Y. A., Hasani, I. W., Kabba, S., & Ragab, W. M. (2025). Artificial intelligence in healthcare and medicine: Clinical applications, therapeutic advances, and future perspectives. European Journal of Medical Research, 30(1), 848. https://doi.org/10.1186/s40001-025-03196-w
18. Gadalla AAH, Friberg IM, Kift-Morgan A, et al. (2019). Identification of clinical and urine biomarkers for uncomplicated urinary tract infection using machine learning algorithms. Scientific Reports.
19. Ghayad, J. P., Barakett-Hamadé, V., & Sleilaty, G. (2022). Prospective validation of a machine learning model for low-density lipoprotein cholesterol estimation. Laboratory Medicine, 53(6), 629–635. https://doi.org/10.1093/labmed/lmac049
20. Gilligan, J. (2025, October 6). Transforming laboratory medicine through AI: From promise to practice. Mayo Clinic Laboratories. https://news.mayocliniclabs.com/2025/10/06/transforming-laboratory-medicine-through-ai-from-promise-to-practice/
21. Goebel, C., Louden, C. L., McKenna, R., Onugha, O., Wachtel, A., & Long, T. (2020). Blood test shows high accuracy in detecting stage I non-small cell lung cancer. BMC Cancer, 20, 137.
22. Gözgöz, H., Orhan, O., Konuk, B. A., Akan, P., et al. (2025). A machine learning model for predicting oligoclonal band positivity using routine cerebrospinal fluid and serum biochemical markers. American Journal of Clinical Pathology, 164(6), 933–945.
23. Gröschel, M. I., Owens, M., Freschi, L., Vargas, R., Marin, M. G., Phelan, J., Iqbal, Z., Dixit, A., & Farhat, M. R. (2021). GenTB: A user-friendly genome-based predictor for tuberculosis resistance powered by machine learning. Genome Medicine, 13(1), 138. https://doi.org/10.1186/s13073-021-00953-4

24. Guan X, Zhang B, Fu M, et al. (2021). Clinical and inflammatory features based machine learning model for fatal risk prediction of hospitalized COVID-19 patients: Results from a retrospective cohort study. Annals of Medicine, 53(1), 257–266. https://doi.org/10.1080/07853890.2020.1868564
25. Gunčar, G., Kukar, M., Notar, M., Brvar, M., Černelč, P., Notar, M., & Notar, M. (2018). An application of machine learning to haematological diagnosis. Scientific Reports, 8, 411.
26. Hornbrook MC, Goshen R, Choman E, et al. (2017). Early colorectal cancer detected by machine learning model using gender, age, and complete blood count data. Digestive Diseases and Sciences, 62, 2719–2727. https://doi.org/10.1007/s10620-017-4701-6
27. Hou, H., Zhang, R., & Li, J. (2024). Artificial intelligence in the clinical laboratory. Clinica Chimica Acta, 559, 119724. https://doi.org/10.1016/j.cca.2024.119724
28. Huang, G., Jin, Q., & Mao, Y. (2023). Predicting the 5-year risk of nonalcoholic fatty liver disease using machine learning models: Prospective cohort study. Journal of Medical Internet Research, 25, e46891. https://doi.org/10.2196/46891
29. International Telecommunication Union, World Health Organization, & World Intellectual Property Organization. (2025, March 17–19). ITU/WHO/WIPO global initiative on artificial intelligence for health (GI-AI4H) meeting. National University of Singapore, Centre for Life Sciences, Singapore. https://www.itu.int/en/ITU-T/focusgroups/ai4h/Pages/default.aspx
30. Ivanova, I., Ivanova, N., & Atanasova, B. (2024). Artificial intelligence in laboratory medicine – Let's talk about it. Bulgarian Society of Medical Sciences Journal, 3, 1–4. https://doi.org/10.3897/bsms.3.120969
31. Januzzi, J. L., Jr., Sattar, N., Vaduganathan, M., Magaret, C. A., Rhyne, R. F., Liu, Y., Masson, S., Butler, J., & Hansen, M. K. (2025). A validated multivariable machine learning model to predict cardio-kidney risk in diabetic kidney disease. Cardiovascular Diabetology, 24(1), 213. https://doi.org/10.1186/s12933-025-02779-5
32. Jawad BN, Shaker SM, Altintas I, et al. (2024). Development and validation of prognostic machine learning models for short- and long-term mortality among acutely admitted patients based on blood tests. Scientific Reports
33. Kaneko, H., Umakoshi, H., Ogata, M. et al., (2021) Machine learning based models for prediction of subtype diagnosis of primary aldosteronism using blood test. Sci Rep 11, 9140 (2021). https://doi.org/10.1038/s41598-021-88712-8
34. Karthikeyan, A., Garg, A., Vinod, P. K., & Priyakumar, U. D. (2021). Machine learning based clinical decision support system for early COVID-19 mortality prediction. Frontiers in Public Health, 9, 626697. https://doi.org/10.3389/fpubh.2021.626697
35. Kawakami, E., Tabata, J., Yanaihara, N., et al. (2019). Application of artificial intelligence for preoperative diagnostic and prognostic prediction in epithelial ovarian cancer based on blood biomarkers. Clinical Cancer Research, 25(10), 3006–3015.
36. Koloi, A., Loukas, V. S., Hourican, C., Sakellarios, A. I., Quax, R., Mishra, P. P., Lehtimäki, T., Raitakari, O. T., Papaloukas, C., Bosch, J. A., März, W., & Fotiadis, D. I. (2024). Predicting early-stage coronary artery disease using machine learning and routine clinical biomarkers improved by augmented virtual data. European Heart Journal – Digital Health. https://doi.org/10.1093/ehjdh/ztae102
37. Kraszewski S, Szczurek W, Szymczak J, et al. (2021). Machine Learning Prediction Model for Inflammatory Bowel Disease Based on Laboratory Markers: Working Model in a Discovery Cohort Study. Journal of Clinical Medicine
38. Kukar M, Gunčar G, Vovko T, et al. (2021). COVID-19 diagnosis by routine blood tests using machine learning. Scientific Reports.
39. Laguarta, J., Hueto, F., & Subirana, B. (2020). COVID-19 Artificial Intelligence Diagnosis Using Only Cough Recordings. IEEE Open Journal of Engineering in Medicine and Biology, 1, 275–281.
40. Lee, S., Park, J. S., Woo, H., Yoo, Y. K., Lee, D., Chung, S., Yoon, D. S., Lee, K. B., & Lee, J. H. (2024). Rapid deep learning-assisted predictive diagnostics for point-of-care testing. Nature Communications, 15, 1695.

41. Lev-Tzion R, Dolev AS, Bar-Asher SY, et al. (2025). Machine learning for predicting Crohn's disease from routine blood tests years before diagnosis: results from the epi-IIRN cohort. Journal of Crohn's and Colitis, 19(8), jjaf143. https://doi.org/10.1093/ecco-jcc/jjaf143
42. Lewis, J. E., & Kemp, M. L. (2021). Integration of machine learning and genome-scale metabolic modeling identifies multi-omics biomarkers for radiation resistance. Nature Communications, 12, 2700.
43. Lippi, G., & Plebani, M. (2025). Lights and shadows of artificial intelligence in laboratory medicine. Advances in Laboratory Medicine, 6(1), 1–3. https://doi.org/10.1515/almed-2025-0024
44. Liu, Y., Fan, L., Wang, W., Song, H., Zhang, Z., Liu, Q., Meng, Z., Li, S., Wang, H., Zhou, S., Liu, W., Xia, G., Duan, J., Guo, C., Wang, L., Xu, L., Wang, T., Li, H., Zhang, X., Xiang, T., Liu, D., Yu, Z., Liu, Y., Wang, J., & Zheng, X. (2025). A machine learning model for mortality prediction in patients with severe fever with thrombocytopenia syndrome: A prospective, multicenter cohort study. Emerging Microbes & Infections, 14(1), 2498572. https://doi.org/10.1080/22221751.2025.2498572
45. Liu, Z.-Z., Yuan, Q., Zhang, Y.-D., Zhang, X.-D., Liu, J., Yan, J.-W., Du, K.-P., Chen, H.-J., & Wang, L. (2025). Validation and interpretation of machine-learning models for rapid identification of active tuberculosis infection using routine laboratory indicators. Frontiers in Cellular and Infection Microbiology, 15, 1718614. https://doi.org/10.3389/fcimb.2025.1718614
46. Luo Y, Szolovits P, Dighe AS, Baron JM. (2016). Using machine learning to predict laboratory test results. American Journal of Clinical Pathology.
47. Luu, H. S. (2025). Laboratory data as a potential source of bias in healthcare artificial intelligence and machine learning models. Annals of Laboratory Medicine, 45(1), 12–21. https://doi.org/10.3343/alm.2024.0323
48. Ma, L., Cai, E., Zhou, Z., Gong, M., & Li, C. (2026). Predicting adverse prognostic outcomes in hospitalized breast cancer patients: development and validation of a risk model. BMC Medical Informatics and Decision Making, 26, 85.
49. Meyer A, Zverinski D, Pfahringer B, et al. (2018). Machine learning for real-time prediction of complications in critical care: a retrospective study. The Lancet Respiratory Medicine
50. Miller, H. A., & Valdes, R. (2025). Rigorous validation of machine learning in laboratory medicine: Guidance toward quality improvement. Critical Reviews in Clinical Laboratory Sciences. https://doi.org/10.1080/10408363.2025.2488842
51. Moons, K. G. M., Damen, J. A. A., Kaul, T., Hooft, L., Andaur Navarro, C., Dhiman, P., Beam, A. L., Van Calster, B., Celi, L. A., Denaxas, S., Denniston, A. K., Ghassemi, M., Heinze, G., Kengne, A. P., Maier-Hein, L., Liu, X., Logullo, P., McCradden, M. D., Liu, N., … van Smeden, M. (2025). PROBAST+AI: An updated quality, risk of bias, and applicability assessment tool for prediction models using regression or artificial intelligence methods. BMJ, 388, e082505. https://doi.org/10.1136/bmj-2024-082505
52. Muro S, Ishida M, Horie Y, et al. (2021). Machine Learning Methods for the Diagnosis of Chronic Obstructive Pulmonary Disease in Healthy Subjects: Retrospective Observational Cohort Study. JMIR Medical Informatics
53. Ni, Y., Song, P., Lin, X., Shi, J., Shi, Q., Li, Y., Zhu, S., Zhou, T., Xun, Y., Zhang, S., Ren, X., Lu, K., Wu, F., Wang, W., Zhao, P., Zhou, R., Zhang, W., Li, D., Zhang, J., Chen, C., Mao, L., Zhou, L., Pan, G., Peng, Y., Yu, Y., Chen, Y., Ni, R., Luo, G., Zhang, Y., Fang, J., Zheng, H., & Luo, D. (2025). Can mutation abundance assess the biological behavior of BRAFV600E-positive papillary thyroid carcinoma? Journal of Translational Medicine, 23, 704.
54. Ning, W., Wang, Z., Gu, Y., Huang, L., Liu, S., Chen, Q., Yang, Y., & Hong, G. (2025). Machine learning models based on routine blood and biochemical test data for diagnosis of neurological diseases. Scientific Reports, 15, 27857. https://doi.org/10.1038/s41598-025-27857
55. Pan, Y., Fan, Q., Liang, Y. et al., (2024) A machine learning prediction model for Cardiac Amyloidosis using routine blood tests in patients with left ventricular hypertrophy. Sci Rep 14, 28644 (2024). https://doi.org/10.1038/s41598-024-77466-8
56. Park DJ, Park MW, Lee H, et al. (2021). Development of machine learning model for diagnostic disease prediction based on laboratory tests. Scientific Reports.

57. Parmar, C., Grossmann, P., Rietveld, D., Rietbergen, M. M., Lambin, P., & Aerts, H. J. W. L. (2015). Radiomic machine-learning classifiers for prognostic biomarkers of head and neck cancer. Frontiers in Oncology, 5, 272.
58. Pinsky, M. R., Bedoya, A., Bihorac, A., Celi, L., Churpek, M., Economou-Zavlanos, N. J., Elbers, P., Saria, S., Liu, V., Lyons, P. G., Shickel, B., Toral, P., Tscholl, D., & Clermont, G. (2024). Use of artificial intelligence in critical care: Opportunities and obstacles. Critical Care, 28(1), 113. https://doi.org/10.1186/s13054-024-04860-z
59. Poplin R, Varadarajan AV, Blumer K, et al., (2018) Prediction of cardiovascular risk factors from retinal fundus photographs. Nature Biomedical Engineering. 2018;2:158–164.
60. Poss, A. M., Maschek, J. A., Cox, J. E., et al. (2020). Machine learning reveals serum sphingolipids as cholesterol-independent biomarkers of coronary artery disease. Journal of Clinical Investigation, 130(3), 1363–1376.
61. Rajkomar A, Oren E, Chen K, et al. (2018). Scalable and accurate deep learning with electronic health records. npj Digital Medicine.
62. Rawson TM, Hernandez B, Moore LSP, et al. (2019). Supervised machine learning for the prediction of infection on admission to hospital: a prospective observational cohort study. Journal of Antimicrobial Chemotherapy, 74(4), 1108–1115. https://doi.org/10.1093/jac/dky514
63. Rim, T. H., Lee, G., Kim, Y., Tham, Y.-C., Lee, C. J., Baik, S. J., et al. (2020). Prediction of systemic biomarkers from retinal photographs: development and validation of deep-learning algorithms. The Lancet Digital Health, 2(10), e526–e536.
64. Rosen, A. W., Ose, I., Gögenur, M., Andersen, L. P. K., Bojesen, R. D., Vogelsang, R. P., Rose, M. H., Steenfos, P. W., Hansen, L. B., Spuur, H. S., Raben, I., Skou, S. T., Holm, E. A., Mortensen, K., Kjær, T., & Eriksen, J. R. (2025). Clinical implementation of an AI-based prediction model for decision support for patients undergoing colorectal cancer surgery. Nature Medicine, 31(11), 3737–3748. https://doi.org/10.1038/s41591-025-03942-x
65. Saul, H. (2025, October 1). How AI and machine learning are revolutionizing the laboratory. Roche Diagnostics. https://diagnostics.roche.com/global/en/lab-leaders/article/machine-learning-ai-in-laboratory.html
66. Schwartz, L., Matania, N., Levi, M., Lazebnik, T., Kushnir, S., Yosef, N., Hoogi, A., & Shlomi, D. (2025). A blood test-based machine learning model for predicting lung cancer risk. Frontiers in Medicine, 12, 1577451. https://doi.org/10.3389/fmed.2025.1577451
67. Seliverstov PV, Kutsenko VP, Gorelova VG, Magomedova SA, Akhmedov SR, Nurmyradov YN. Using artificial intelligence for biomarker analysis in clinical diagnostics. Vol 22, No 5 (2024): 31–39.https://doi.org/10.29296/24999490-2024-05-04
68. Suleman, M. U., Mursaleen, M., Khalil, U., Saboor, A., Bilal, M., Khan, S. A., Subhani, M. A., Hussnain, M. A., Tabassum, S. N., & Tahir, M. (2025). Assessing the generalizability of artificial intelligence in radiology: A systematic review of performance across different clinical settings. Annals of Medicine and Surgery, 87(12), 8803–8811. https://doi.org/10.1097/MS9.0000000000004166
69. Szumilas, D., Ochmann, A., Zięba, K., Bartoszewicz, B., Kubrak, A., Makuch, S., Agrawal, S., Mazur, G., & Chudek, J. (2024). Evaluation of AI-driven LabTest Checker for diagnostic accuracy and safety: Prospective cohort study. JMIR Medical Informatics, 12, e57162. https://doi.org/10.2196/57162
70. Tang, X., Ye, W., Ou, Y., Ye, H., Zhu, X., Huang, D., Liu, J., Zhao, F., Deng, W., Li, C., Cai, W., Zheng, Y., Zeng, J., & Cai, Y. (2025). Development and validation of a machine learning model for detection and classification of vertigo. Laryngoscope, 135(5), 1652–1660. https://doi.org/10.1002/lary.31959
71. Tomašev N, Glorot X, Rae JW, et al. (2019). A clinically applicable approach to continuous prediction of future acute kidney injury. Nature, 572(7767), 116–119.
72. Uçucu, S., & Azik, F. (2024). Veštačka neuralna mreža za prepoznavanje β-talasemije minor i anemije uzrokovane nedostatkom gvožđa - modeli veštačke inteligencije. Journal of Medical Biochemistry, 43(1), 11–18.

73. Undru, T. R., Uday, U., Lakshmi, J. T., Kaliappan, A., Mallamgunta, S., Nikhat, S. S., Sakthivadivel, V., & Gaur, A. (2022). Integrating artificial intelligence for clinical and laboratory diagnosis: A review. Maedica (Bucur), 17(2), 420–426. https://doi.org/10.26574/maedica.2022.17.2.420
74. Valentidenta, W. M., & Pribadi, F. (2025). Artificial intelligence in echocardiography: Enhancing diagnostic accuracy, workflow efficiency, and cost-effectiveness in hypertension and cardiovascular risk assessment. Revista Latinoamericana de Hipertensión, 20(8). https://doi.org/10.5281/zenodo.17113815
75. Veseli, E., Mehrabanian, M., & Ammar, N. (2025). The potential of artificial intelligence in the early detection of systemic diseases during routine dental care. British Dental Journal, 239(3), 168–174. https://doi.org/10.1038/s41415-025-8666-7
76. Villota, S. D., et al. (2025). Validation of new equipment for SARS-CoV-2 diagnosis in Ecuador: Detection of the virus and antibodies generated by disease and vaccines with one POC device. PLOS ONE, 20(4), e0321794.
77. Wang S, Liu Z, Rong Y, et al. (2019). Deep learning provides a new computed tomography-based prognostic biomarker for recurrence prediction in high-grade serous ovarian cancer. Radiotherapy and Oncology
78. Wang, K., Lu, X., Zhou, H., Gao, Y., Zheng, J., Tong, M., et al. (2020). Deep learning radiomics of shear wave elastography significantly improved diagnostic performance for assessing liver fibrosis in chronic hepatitis B: a prospective multicentre study. Gut, 68(4).
79. Weng SF, Reps J, Kai J, Garibaldi JM, Qureshi N. (2017). Can machine-learning improve cardiovascular risk prediction using routine clinical data? PLOS ONE.
80. White, S. J., Chau, M., Arruzza, E., Ong, M., John, H., Theiss, R., Yaxley, K. L., & To, M.-S. (2025). Assessment of Standards for Reporting of Diagnostic Accuracy (STARD) 2015 guideline adherence in medical imaging diagnostic accuracy studies published in 2023. Journal of Clinical Epidemiology, 179, 111654. https://doi.org/10.1016/j.jclinepi.2024.111654
81. World Health Organization Regional Office for Europe. (2025). Artificial intelligence is reshaping health systems: State of readiness across the WHO European Region. WHO Regional Office for Europe. https://www.who.int/europe/publications/i/item/WHO-EURO-2025-12707-52481-81028
82. World Health Organization. (2024). Harnessing artificial intelligence for health. https://www.who.int/teams/digital-health-and-innovation/harnessing-artificial-intelligence-for-health
83. World Health Organization. (2024, May 27). Leading the future of global health with responsible artificial intelligence. https://www.who.int/publications/m/item/leading-the-future-of-global-health-with-responsible-artificial-intelligence
World Health Organization. (2024, May 27). Artificial intelligence for health: Supporting countries to deploy responsible AI technologies to accelerate equitable health for all. https://www.who.int/publications/m/item/artificial-intelligence-for-health
84. Xie, H., Jia, Y., & Liu, S. (2024). Integration of artificial intelligence in clinical laboratory medicine: Advancements and challenges. Interdisciplinary Medicine. https://doi.org/10.1002/INMD.20230056
85. Xie, Y., Meng, W.-Y., Li, R.-Z., et al. (2021). Early lung cancer diagnostic biomarker discovery by machine learning methods. Translational Oncology, 14(1), 100907.
86. Yan L, Zhang HT, Goncalves J, et al. (2020). An interpretable mortality prediction model for COVID-19 patients. Nature Machine Intelligence.
87. Yan, W., Zhu, L., Wei, H., Feng, J., & Gao, Y. (2024). A machine learning model based on blood indices for the differential diagnosis of colorectal cancer and colorectal polyps. iLABMED.
88. Yang HS, Hou Y, Vasovic LV, et al. (2020). Routine laboratory blood tests predict SARS-CoV-2 infection using machine learning. Clinical Chemistry, 66(11), 1396–1404. https://doi.org/10.1093/clinchem/hvaa200
89. Yao H, Zhang N, Zhang R, et al. (2020). Severity detection for COVID-19 patients using a machine learning model based on blood and urine tests. Frontiers in Cell and Developmental Biology.
90. Yu, A. C., Mohajer, B., & Eng, J. (2022). External validation of deep learning algorithms for radiologic diagnosis: A systematic review. Radiology: Artificial Intelligence, 4(3), e210064. https://doi.org/10.1148/ryai.210064

91. Zhang J, Liu W, Xiao W, et al. (2024). Machine learning-derived blood culture classification with both predictive and prognostic values in the intensive care unit: A retrospective cohort study. Intensive and Critical Care Nursing
92. Zhang Z, Bast RC Jr, Yu Y, et al. (2004). Three biomarkers identified from serum proteomic analysis for the detection of early stage ovarian cancer. Cancer Research.
93. Zhang, D., Cheng, Z., Song, Y., Li, H., Shi, L., Wang, N., Peng, Y., Chen, R., Sun, N., Han, M., Hu, F., Zong, C., Zhang, R., Chen, S., Zhu, C., Zhang, X., Li, X., Ma, X., Shi, C., Zhang, X., Liu, R., Ren, Z., Wang, L., Zeng, Q., Zeng, T., & Chen, X. (2025). Rapid and sensitive acute leukemia classification and diagnosis platform using deep learning-assisted SERS detection. Cell Reports Medicine, 6(9), 102320.
94. Zhen S, Cheng M, Tao YB, et al. (2020). Deep learning for accurate diagnosis of liver tumor based on magnetic resonance imaging and clinical data. Frontiers in Oncology
95. Zhou, J., Park, S., Dong, S., Tang, X., & Wei, X. (2025). Artificial intelligence-driven transformative applications in disease diagnosis technology. Medical Review, 5(5), 353–377. https://doi.org/10.1515/mr-2024-0097
96. Zmudzki, F., Smeets, R. J. E. M., Groenewegen, J. S., & van der Graaff, E. (2025). Machine learning clinical decision support for interdisciplinary multimodal chronic musculoskeletal pain treatment: Prospective pilot study of patient assessment and prognostic profile validation. JMIR Rehabilitation and Assistive Technologies, 12, e65890. https://doi.org/10.2196/65890

TABLE 1. COMPREHENSIVE CHARACTERISTICS OF INCLUDED STUDIES

Author (Year)	Country	Study Design	Sample Size	Population	AI Model	Data Type	Comparator	Key Outcome
Szumilas et al. (2024)	Poland	Prospective cohort	101	ED patients	ML (white-box)	Lab + clinical	Physician judgment	Accuracy 74.3%, Sens 100%
Gözgöz et al. (2025)	Turkey	Retrospective	1,709	Neurology patients	Ensemble ML	CSF + serum	Immunoblotting	AUC 0.902
Ni et al. (2025)	China	Mixed	769+	Thyroid cancer	XGBoost	Molecular + clinical	Histopathology	AUC 0.82–0.86
Gröschel et al. (2021)	USA	Retrospective	20,408	TB isolates	RF, WDNN	Genomic	DST	AUC 0.89–0.96
Zhang et al. (2025)	China	Validation	390+	Leukemia patients	DL (CNN+Transformer)	CSF spectroscopy	Clinical diagnosis	High accuracy
Uçucu & Azik (2024)	Turkey	Retrospective	396	Anemia patients	ANN	CBC	Clinical diagnosis	Improved accuracy
Ma et al. (2026)	China	Retrospective cohort	4,242	Breast cancer	RF, XGBoost	EMR + lab	Clinical outcome	AUC ~0.89
Villota et al. (2025)	Ecuador	Validation	2,000	COVID patients	RT-LAMP + AI	Swab + serum	PCR/ELISA	Sens 91%, Spec 92%
Cao et al. (2013)	China	Retrospective	239	Liver disease	MLP	Lab biomarkers	Clinical diagnosis	AUC 0.942
Goebel et al. (2020)	USA	Retrospective	486	Lung cancer	RF-based	Blood biomarkers	Imaging/pathology	AUC 0.966
Gunčar et al. (2018)	Slovenia	Retrospective	8,233	Hematology patients	RF	Lab panels	Specialist diagnosis	Accuracy 88%
Kawakami et al. (2019)	Japan	Retrospective	435	Ovarian cancer	RF, GBM	Blood biomarkers	Histopathology	AUC 0.978
Xie et al. (2021)	China	Case-control	153	Lung cancer	Naïve Bayes	Metabolomics	Clinical diagnosis	AUC 0.989
Kukar et al. (2021)	Slovenia	Retrospective	5,333	COVID patients	XGBoost	Blood tests	PCR	AUC 0.97
Yang et al. (2020)	China	Multi-center	3,356	COVID patients	GBDT	Lab + demographic	PCR	AUC 0.854

Pan et al. (2024)	China	Retrospective	6,563	Cardiac amyloidosis	XGBoost	Lab biomarkers	Clinical diagnosis	AUC 0.95
Hornbrook et al. (2017)	USA	Case-control	17,000 +	CRC patients	ML model	CBC	Clinical diagnosis	AUC ~0.80
Tomašev et al. (2019)	USA	Retrospective	700,00 0+	AKI patients	RNN	EHR + lab	Clinical diagnosis	AUC ~0.92

DETAILED CHARACTERISTICS OF ALL INCLUDED STUDIES ARE PRESENTED IN SUPPLEMENTARY TABLE S1
SUPPLEMENTARY TABLE S1. CHARACTERISTICS OF INCLUDED STUDIES

Author (Year)	Study Design	Sample Size	Population / Condition	AI Model	Data Type	Comparator	Key Outcomes
Szumilas et al. (2024)	Prospective cohort	101	ED patients	ML (white-box)	Lab + clinical	Physician	Accuracy 74.3%, Sens 100%
Gözgöz et al. (2025)	Retrospective	1,709	Neurological (OCB)	Ensemble ML	CSF + serum	Immunoblotting	AUC 0.902
Ni et al. (2025)	Mixed	769+	Thyroid cancer	XGBoost	Molecular + clinical	Histopathology	AUC 0.82–0.86
Gröschel et al., 2021	Retrospective	20,408	TB isolates	RF, WDNN	Genomic	DST	AUC 0.89–0.96
Zhang et al. (2025)	Validation	390+	Leukemia	DL (CNN+Transformer)	CSF spectroscopy	Clinical diagnosis	High accuracy
Uçucu & Azik (2024)	Retrospective	396	IDA vs BTM	ANN	CBC	Clinical diagnosis	Improved accuracy
Ma et al. (2026)	Retrospective cohort	4,242	Breast cancer	RF, XGBoost	EMR + lab	Clinical outcome	AUC ~0.89
Villota et al. (2025)	Validation	2,000	COVID-19	RT-LAMP + AI	Swab + serum	PCR/ELISA	Sens 91%, Spec 92%
Cao et al. (2013)	Retrospective	239	Liver cirrhosis	MLP	Lab biomarkers	Clinical diagnosis	AUC 0.942
Goebel et al. (2020)	Retrospective	486	Lung cancer	RF-based	Blood biomarkers	Imaging/pathology	AUC 0.966

Gunčar et al. (2018)	Retrospective	8,233	Hematologic diseases	RF	Lab panels	Specialist	Accuracy up to 88%
Chang et al. (2021)	Systematic review	—	Alzheimer's disease	Multiple ML	Biomarkers	Clinical diagnosis	AUC 0.73–0.996
Seungmin Lee et al. (2024)	Prospective	~600+	POCT	DL (CNN+LSTM)	LFA images	Standard assay	Accuracy ~97%
Parmar et al. (2015)	Retrospective	196	Head & neck cancer	RF, SVM	Radiomics	Clinical outcomes	AUC ~0.67
Kawakami et al. (2019)	Retrospective	435	Ovarian cancer	RF, GBM	Blood biomarkers	Histopathology	AUC 0.978
Lewis & Kemp (2021)	Retrospective	915	Cancer (radiation response)	GBM	Multi-omics	Clinical outcomes	AUC 0.906
Xie et al. (2021)	Case-control	153	Lung cancer	Naïve Bayes	Metabolomics	Clinical diagnosis	AUC 0.989
Poss et al. (2020)	Case-control	674	CAD	RF	Lipidomics	Clinical diagnosis	Improved prediction
Brinati et al. (2020)	Retrospective	279	COVID-19	RF	Blood tests	PCR	Accuracy 82%
DeGroat et al., 2024	Retrospective	71	CVD	Ensemble ML	Transcriptomics	Clinical diagnosis	Accuracy ~96%
Apostolopoulos et al. (2020)	Retrospective	3,905	COVID-19	CNN	Imaging	Clinical diagnosis	Accuracy 99%
Laguarta et al. (2020)	Retrospective	5,320	COVID-19	CNN	Audio	Clinical diagnosis	AUC 0.97
Rim et al. (2020)	Cross-sectional	236,257	Systemic biomarkers	DL (VGG16)	Retinal imaging	Lab biomarkers	R^2 up to 0.83
Wang et al. (2020)	Prospective	654	Liver fibrosis	DL radiomics	Imaging + lab	Biopsy	AUC 0.97–0.98
Chen et al. (2024)	Multicenter	702	Gastric cancer	RF	Metabolomics	Clinical diagnosis	AUC 0.967
Calvert et al. (2016)	Retrospective	—	Sepsis	ML model	Clinical + lab	Clinical diagnosis	AUC 0.83
Tomašev et al. (2019)	Retrospective	700,000+	AKI	RNN	EHR + lab	Clinical diagnosis	AUC 0.92–0.93
Ardila et al. (2019)	Retrospective	6,716	Lung cancer	DL (CNN)	Imaging	Radiologists	AUC 0.944

Poplin et al. (2018)	Retrospective	284,000+	CVD risk	CNN	Retinal images	Clinical outcomes	AUC 0.97
Zhang et al. (2004)	Case-control	645	Ovarian cancer	SVM	Proteomics	CA125	Sens 74%, Spec 97%
Yan et al. (2020)	Retrospective	485	COVID mortality	XGBoost	Lab biomarkers	Clinical outcomes	>90% accuracy
Karthikeyan et al. (2021)	Retrospective	370	COVID mortality	NN	Lab biomarkers	Clinical outcomes	AUC 0.989
Rajkomar et al. (2018)	Retrospective	216,221	Hospital outcomes	DL	EHR + lab	Clinical outcomes	AUC 0.93–0.94
Luo et al. (2016)	Retrospective	5,128	Ferritin prediction	RF	Lab tests	Lab measurement	AUC 0.96–0.97
Park et al. (2021)	Retrospective	5,145	Multi-disease	Ensemble ML	Lab data	Clinical diagnosis	Accuracy ~93%
Weng et al. (2017)	Cohort	378,256	CVD	ML models	Lab + clinical	Risk scores	AUC ~0.76
Yao et al. (2020)	Retrospective	137	COVID severity	SVM	Blood + urine	Clinical severity	Accuracy 81%
Botlagunta et al. (2023)	Retrospective	1,449	Breast cancer	Decision tree	Blood biomarkers	Clinical diagnosis	AUC 0.87
Gadalla et al. (2019)	Case-control	183	UTI	RF, SVM	Urinary biomarkers	Culture	AUC ~0.82
Kukar et al. (2021)	Retrospective	5,333	COVID	XGBoost	Blood tests	PCR	AUC 0.97
Zhen et al. (2020)	Retrospective	1,210	Liver tumors	CNN	Imaging + lab	Pathology	AUC up to 0.998
Wang et al. (2019)	Retrospective	245	Ovarian cancer	DL	Imaging	Clinical outcomes	AUC 0.825
Meyer et al. (2018)	Retrospective	47,000+	ICU complications	RNN	Clinical + lab	Clinical outcomes	High accuracy
Muro et al. (2021)	Cohort	24,815	COPD	XGBoost	Lab + clinical	Clinical diagnosis	AUC 0.956
Kraszewski et al. (2021)	Retrospective	702	IBD	RF	Lab biomarkers	Clinical diagnosis	Precision up to 97%

Jawad et al. (2024)	Retrospective	48,841	Mortality prediction	ML models	Lab biomarkers	Clinical outcomes	AUC 0.85–0.93
Zhang et al. (2024)	Retrospective	2,451	ICU infections	Clustering ML	Lab + clinical	Clinical outcomes	Risk stratification
Kaneko et al. (2021)	Retrospective	229	Aldosteronism	RF	Blood biomarkers	Clinical diagnosis	AUC 0.99
Rawson et al. (2019)	Prospective	104	Infection	SVM	Lab biomarkers	Clinical diagnosis	AUC 0.84
Hornbrook et al. (2017)	Case-control	17,000+	CRC	ML model	CBC	Clinical diagnosis	AUC ~0.80
Yang et al. (2020)	Multi-center	3,356	COVID	GBDT	Lab + clinical	PCR	AUC 0.854
Pan et al. (2024)	Retrospective	6,563	Cardiac amyloidosis	XGBoost	Lab biomarkers	Clinical diagnosis	AUC 0.95
Guan et al. (2021)	Retrospective	1,270	COVID mortality	XGBoost	Lab biomarkers	Clinical outcomes	AUC >0.90
Lev-Tzion et al. (2025)	Cohort	8,630	Crohn's disease	ML	Lab biomarkers	Clinical diagnosis	AUC 0.70
Ning et al. (2025)	Retrospective	33,000+	Neurological diseases	XGBoost	Lab biomarkers	Clinical diagnosis	AUC 0.978
Liu et al. (2025)	Retrospective	3,829	TB	XGBoost	Lab biomarkers	Clinical diagnosis	Accuracy 97%
Schwartz et al. (2025)	Retrospective	4,076	Lung cancer risk	CatBoost	Blood biomarkers	Clinical diagnosis	AUC 0.787
Koloi et al. (2024)	Cohort	3,316	CAD	RF, GB	Lab + clinical	Clinical diagnosis	AUC up to 0.87
Yan et al. (2024)	Retrospective	363	CRC vs polyps	XGBoost	Lab biomarkers	Clinical diagnosis	AUC 0.869
Huang et al. (2023)	Cohort	6,196	NAFLD	ML models	Lab + clinical	Clinical diagnosis	AUC ~0.81
Rosen et al. (2025)	Cohort	18,403	CRC surgery	ML-CDSS	Clinical + lab	Standard care	Improved outcomes
Chang et al. (2025)	Observational	37,632	ED patients	ML-CDSS	Clinical + lab	Standard care	AUC up to 0.95
Dong et al. (2025)	Cohort	4,950	CVD	ML models	Lab biomarkers	Clinical outcomes	AUC ≥0.90

Liu et al. (2025)	Prospective	1,606	SFTS	MLP	Lab + clinical	Clinical outcomes	AUC 0.917
Tang et al. (2025)	Prospective	2,937	Vertigo	XGBoost	Lab + clinical	Clinical diagnosis	AUC 0.947
Ghayad et al. (2022)	Prospective	6,599	LDL prediction	KNN	Lab + demographic	Lab assay	ICC >0.9
Januzzi et al. (2025)	Validation	—	DKD	ML model	Lab biomarkers	Clinical outcomes	C-statistic 0.80
Chen et al. (2025)	Multicenter	2,196	Preeclampsia	ML ensemble	Lab + clinical	Clinical diagnosis	Sens 72%

TABLE 2. DIAGNOSTIC PERFORMANCE OF AI MODELS

Study	Disease	Accuracy	Sensitivity	Specificity	AUC
Gözgöz et al. (2025)	OCB	0.83	0.71	0.91	0.902
Gröschel et al., 2021	TB	—	0.91–0.93	>0.96	0.89–0.96
Xie et al. (2021)	Lung cancer	—	0.981	1.00	0.989
Kukar et al. (2021)	COVID-19	—	0.819	0.979	0.97
Cao et al. (2013)	Cirrhosis	0.899	0.952	0.842	0.942
Pan et al. (2024)	Cardiac amyloidosis	—	0.92	0.95	0.95
Yan et al. (2024)	CRC	0.79	0.82	—	0.869
Goebel et al. (2020)	Lung cancer	0.956	0.891	0.977	0.966

TABLE 3. WORKFLOW EFFICIENCY OUTCOMES

Study	AI Application	Outcome
Szumilas et al. (2024)	CDSS	↓ 41.6% unnecessary visits
Seungmin Lee et al. (2024)	POCT AI	TAT reduced to 1–2 min

Zhang et al. (2025)	DL-SERS	Diagnosis in 5 min
Villota et al. (2025)	COVID POC	<2 hours turnaround
Luo et al. (2016)	Lab prediction	Reduced redundant tests
Chang et al. (2025)	ED CDSS	Faster interventions

TABLE 4. CLINICAL UTILITY

Study	Application	Clinical Impact
Tomašev et al. (2019)	AKI	Early detection (48–72 hrs)
Lev-Tzion et al. (2025)	Crohn's	Preclinical detection
Yan et al. (2020)	COVID mortality	Early risk stratification
Hornbrook et al. (2017)	CRC	Early detection (6–12 months)
Rosen et al. (2025)	Cancer CDSS	Reduced complications
Kawakami et al. (2019)	Ovarian cancer	Non-invasive diagnosis

TABLE 5. RISK OF BIAS ASSESSMENT (QUADAS-2 / NOS BASED)

Study	Selection Bias	Measurement Bias	Validation	Overall Risk
Szumilas et al. (2024)	Low	Low	Moderate	Low
Gözgöz et al. (2025)	Moderate	Low	High	Moderate
Ni et al., (2025)	Moderate	Low	High	Moderate
Gröschel et al., 2021	Low	Low	High	Low
Uçucu et al., (2024)	Moderate	Moderate	Low	High
Ma et al., (2026)	Moderate	Low	Moderate	Moderate
Villota et al., (2025)	Low	Low	Moderate	Low
Cao et al., (2013)	Moderate	Moderate	Low	High
Goebel et al., (2020)	Low	Low	Moderate	Low

ISSN No. 2321-2705 | DOI: 10.51244/IJRSI |Volume XIII Issue IV April 2026

TABLE 7. AI MODEL CHARACTERISTICS				
Study	**Algorithm**	**Input Data**	**Output**	**Validation**
Gröschel et al., (2021)	RF, WDNN	Genomic	Drug resistance	External
Ni et al., (2025)	XGBoost	Mutation + clinical	Risk prediction	External
Kukar et al., (2021)	XGBoost	Blood tests	COVID diagnosis	Internal
Pan et al., (2024)	XGBoost	Lab biomarkers	Disease classification	Internal
Kawakami et al., (2019)	RF, GBM	Blood biomarkers	Cancer diagnosis	Internal