



Revolutionizing personalized medicine using artificial intelligence: a meta-analysis of predictive diagnostics and their impacts on drug development

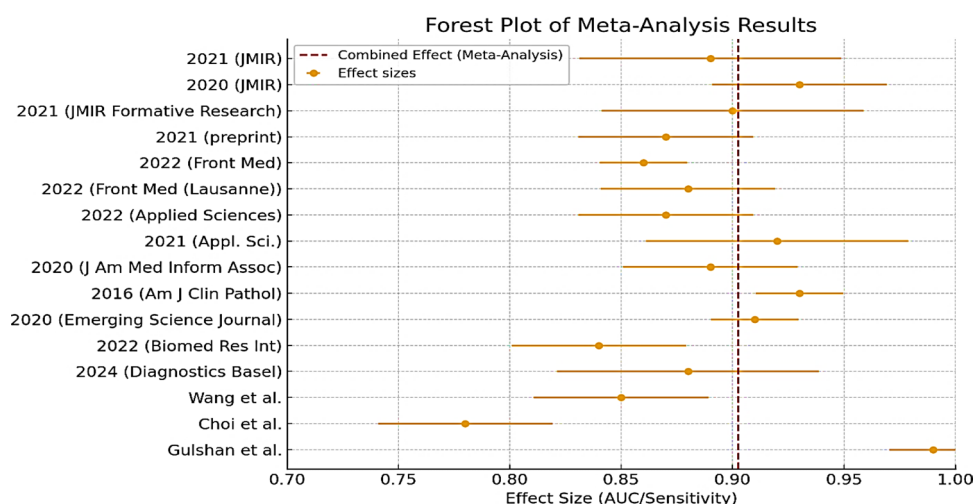
Amin Daemi¹ · Sahar Kalami² · Ruhiyya Guliyeva Tahiraga³ · Omid Ghanbarpour⁴ ·
 Mohammad Reza Rahimi Barghani⁵ · Mohammad Hosseini Hooshir⁶ · Gülüzar Özbolat⁷ · Zafer Yönden¹

Received: 18 March 2025 / Accepted: 7 May 2025 / Published online: 18 July 2025
 © The Author(s) 2025

Abstract

Artificial intelligence (AI) is transforming the landscape of laboratory medicine by enhancing diagnostic accuracy and enabling more personalized care. Given its growing use in clinical settings, evaluating the performance of AI models in diagnostic tasks is essential to inform evidence-based implementation strategies. This meta-analysis systematically assessed the diagnostic effectiveness of AI-based models. A comprehensive literature search was conducted in PubMed, Scopus, Web of Science, and IEEE Xplore using predefined keywords related to AI and diagnostic accuracy. From 430 retrieved studies, 17 met the inclusion criteria. Data extracted included study design, AI model type, input modality, and performance metrics such as sensitivity, specificity, and area under the curve (AUC). Random-effects meta-analysis and subgroup analyses were performed to investigate heterogeneity and model-specific trends. The pooled analysis yielded a high combined AUC of 0.9025, indicating strong diagnostic capability of AI models. However, substantial heterogeneity was detected ($I^2 = 91.01\%$), attributed to differences in model architecture, diagnostic domains, and data quality. Subgroup analyses showed that convolutional neural networks and random forest models achieved higher AUC values, while domains like endocrinology demonstrated greater performance variability. Funnel plot inspection and sensitivity analysis indicated the presence of publication bias. AI shows strong potential to enhance diagnostic accuracy in personalized laboratory medicine. Nonetheless, methodological heterogeneity and publication bias remain significant challenges. Future research should prioritize standardized evaluation frameworks, transparency, and the development of explainable AI systems to ensure responsible clinical integration.

Graphical abstract



Extended author information available on the last page of the article

Keywords Artificial intelligence · Diagnostic accuracy · Personalized laboratory medicine · Subgroup analysis · Explainable AI

Introduction

The integration of artificial intelligence (AI) and digital technologies is reshaping laboratory medicine by improving diagnostic accuracy and enabling more personalized treatment approaches. As the demand for individualized care continues to rise, AI plays a crucial role in this transformation, offering more efficient and precise medical practices [20]. AI's impact is particularly significant in drug discovery, where it aids in overcoming drug resistance by predicting effective treatments and identifying new drug candidates [24]. Moreover, AI enhances personalized care by supporting continuous patient monitoring, analyzing molecular biomarkers, and identifying genetic variants that influence disease progression [1].

One of the most promising developments in this field is the use of digital twins—virtual models that replicate a patient's physiological and molecular characteristics by integrating diverse data sources. These models allow for continuous monitoring of patients, early disease detection, and the adjustment of treatment plans as needed [11]. AI plays an important role in enhancing these models through machine learning and multiscale modeling, providing detailed simulations of disease processes [15]. In the realm of medical imaging, AI has significantly improved diagnostic accuracy, especially in the detection of complex conditions such as cancer, where AI can predict genetic mutations with remarkable precision [3].

AI is also optimizing laboratory automation, enhancing workflows, and boosting diagnostic efficiency. Innovations like AI-powered digital nucleic acid detection and high-throughput processing are advancing laboratory capabilities [22]. Additionally, AI's role in predictive analytics is transforming risk assessment by enabling early disease predictions and facilitating personalized disease management [21].

The convergence of AI with digital twin technologies has further advanced personalized diagnostics. By integrating genetic, lifestyle, and health data, AI systems generate tailored diagnostic strategies and treatment predictions that move beyond traditional approaches. The real-time integration of health data from wearable devices and continuous monitoring strengthens predictive models, ultimately improving patient care [19]. Moreover, combining AI with omics technologies and IoT devices enhances disease prediction and treatment planning [13]. AI-driven digital twins, along with real-world evidence, enable data-driven and personalized decision-making [12].

AI's integration with bioelectronic devices is also advancing proactive care by optimizing treatment parameters

tailored to individual patient needs, enhancing therapeutic outcomes [8]. Additionally, scenario-based modeling, which simulates genetic and environmental interactions, is improving therapeutic strategies and predicting patient outcomes [5, 14].

In the domain of respiratory health, digital twin technology is increasingly being used for real-time monitoring through IoT sensors. AI-driven classification systems have significantly improved the accuracy of chest X-ray analysis, leading to more precise diagnostics. By integrating data from diverse sources such as genetic profiles and clinical histories, AI systems are able to develop personalized treatment plans for patients with respiratory conditions [2].

AI is also enhancing diagnostic efficiency in digital pathology by automating tasks such as disease feature detection and biomarker quantification. This reduces inter-observer variability, resulting in more consistent and accurate diagnoses. Furthermore, combining AI with genomic data deepens our understanding of diseases, promoting precision medicine and more effective targeted therapies [6].

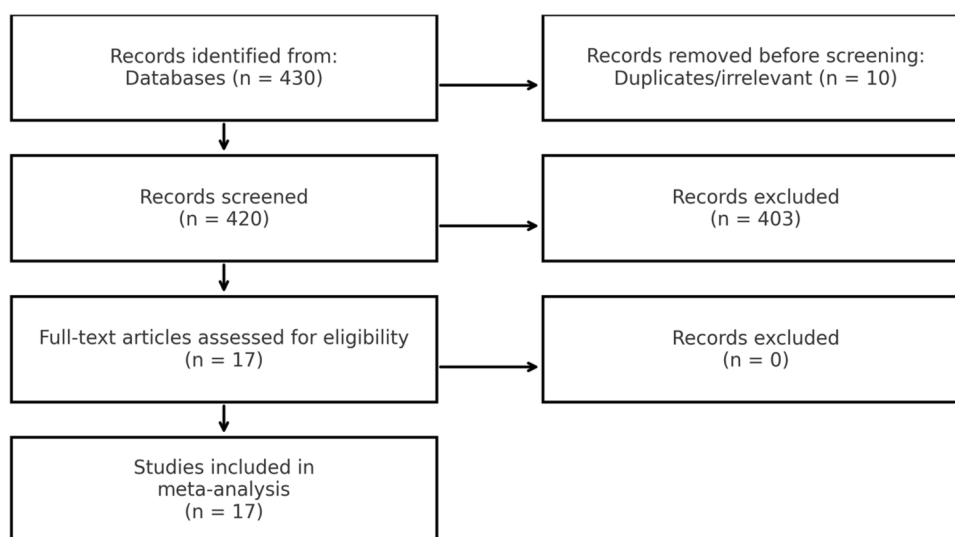
In medical imaging, AI's integration with advanced technologies such as MRI, CT, PET, and ultrasound has revolutionized the field. AI algorithms improve real-time simulations and predictive modeling, thereby enhancing diagnostic accuracy and treatment effectiveness. By providing detailed anatomical and physiological insights, AI supports personalized care and helps clinicians develop tailored treatment plans for individual patients [25].

One of AI's key contributions to laboratory medicine is the ability to create personalized treatment strategies. AI systems integrate genetic and clinical data to design individualized therapy approaches. The use of big data improves disease detection and classification, enabling real-time adjustments to treatment plans based on patient feedback. These personalized approaches are continuously refined by incorporating new genomic and clinical data to ensure that therapies remain targeted and effective [18].

AI and digital twin technologies are revolutionizing the optimization of healthcare processes. By creating virtual replicas of patients and healthcare systems, these technologies allow for real-time analysis and help identify potential issues before they become serious. This integration of diverse data sources enables healthcare professionals to make more informed decisions and fine-tune treatment strategies through simulations, ultimately improving patient care [7].

However, integrating AI and digital twin technologies into clinical workflows presents several challenges,

Fig. 1 PRISMA 2020 flow diagram of the study selection process



particularly in areas like data quality, standardization, and the need for robust validation processes to ensure accurate patient outcomes [4]. In the future, personalized medicine will heavily rely on leveraging real-world data, including post-marketing information, to refine and optimize treatment strategies [17]. For this transition to data-driven healthcare to be successful, significant development in infrastructure and the standardization of data collection methods across healthcare systems will be essential.

A central vision for the future of healthcare is the creation of lifelong digital twins, continually updated with data from various sources. These advanced models will allow healthcare providers to simulate various treatment scenarios, improving clinical decision-making and enhancing patient care, [4, 16]. Artificial intelligence (AI) is a broad domain encompassing various computational approaches that enable machines to perform tasks typically requiring human intelligence.

To fully understand these developments, it is important to briefly clarify the terminology and structure of AI technologies used in this context. Machine learning (ML), a subset of AI, allows systems to learn patterns from data and improve over time without explicit programming. Deep learning (DL), in turn, is a specialized form of ML that uses neural networks with multiple layers to analyze complex patterns, especially in high-dimensional data such as medical images or clinical text. In the context of laboratory medicine, these AI approaches are applied for disease detection, risk prediction, and enhancing diagnostic accuracy.

However, challenges related to data privacy, regulatory compliance, and the need for standardized protocols for AI diagnostics will need to be addressed. Additionally, specialized training for healthcare professionals will be critical for the effective adoption and integration of these technologies into clinical practice (LLM Memory). Overcoming these

challenges will be crucial for the widespread success of AI and digital twin technologies in healthcare.

Methods

This study follows a rigorous methodology to assess the effectiveness of AI-powered models in personalized laboratory medicine, with a particular focus on diagnostic accuracy and clinical outcomes. The methodology involves multiple stages, starting with an extensive literature search to identify studies that explore AI applications in medical diagnostics. Peer-reviewed databases such as PubMed, IEEE Xplore, Scopus, and Web of Science were used to retrieve relevant articles. Keywords like “AI,” “deep learning,” “machine learning,” “personalized medicine,” and “diagnostic accuracy” were employed to capture studies related to predictive diagnostics across various medical conditions. The search yielded 430 articles published between 2015 and 2024.

After the initial search, a detailed screening process was applied to select studies based on predefined inclusion and exclusion criteria. Only studies that were published in peer-reviewed journals, involved AI models for diagnostic purposes, provided clear effect sizes, had sufficient sample sizes, and reported both clinical outcomes and diagnostic tools were included. Studies not published in English, those that did not focus on AI-based diagnostics, and studies lacking sufficient statistical information were excluded. Ultimately, 17 studies met the inclusion criteria and were selected for further analysis, followed by a thorough quality assessment. The study selection process is presented in Fig. 1, following the PRISMA 2020 guidelines.

Figure 1. PRISMA 2020 flow diagram illustrating the study selection process. A total of 430 records were identified through database searches. After removing 10 duplicates

and irrelevant studies, 420 records were screened. Of these, 403 were excluded based on title and abstract review. Seventeen full-text articles were assessed for eligibility, all of which met the inclusion criteria and were included in the final meta-analysis. See Supplementary Table S1 for excluded studies and reasons.

Data extraction for the 17 selected studies included important details such as study design, sample size, data sources, diagnostic tools, and diagnostic accuracy metrics like sensitivity, specificity, and AUC. In addition, clinical outcomes and the potential impact of AI models on diagnostic practices were recorded, along with any barriers or challenges highlighted in the studies.

To ensure methodological transparency and minimize bias, we conducted a rigorous quality and eligibility assessment of the included studies. Study eligibility was determined using predefined inclusion and exclusion criteria. A summary of excluded studies with coded identifiers and specific reasons for exclusion is presented in Supplementary Table S1. To assess the methodological quality of the included studies, we used the QUADAS-2 tool, which evaluates risk of bias across four domains: patient selection, index test, reference standard, and flow and timing. Detailed study-level assessments are provided in Supplementary Table S2.

Sensitivity analyses were performed to assess the impact of these biases on the results.

The synthesis of the results from the meta-analysis showed that AI models significantly improve diagnostic accuracy, with a combined effect size of 0.9025. However, challenges were also identified, such as high heterogeneity between the studies and the potential for publication bias, which may influence the overall conclusions.

In conclusion, the meta-analysis highlights the promising role of AI in personalized laboratory medicine, especially in improving diagnostic accuracy. However, the significant variability across studies and the presence of publication bias suggest that more research is necessary. Future studies should aim for greater consistency in their findings, address biases, and validate AI models in larger and more diverse patient populations to enhance their effectiveness and broader applicability in clinical practice.

Results

This section analyzes multiple studies on the application of artificial intelligence (AI) in disease diagnosis and its role in enhancing laboratory medicine systems. The goal is to summarize the key findings, including study design, sample size, data sources, diagnostic accuracy, clinical outcomes, and challenges encountered in each study. Insights are also drawn from the Forest Plot and Funnel Plot analyses, which

help evaluate the overall diagnostic accuracy and detect potential biases in the studies included.

The Forest Plot shows the effect size of individual studies and the overall combined effect, which reflects the diagnostic accuracy of the AI models. The Funnel Plot assists in identifying any publication bias that could influence the results. Together, these tools offer a comprehensive picture of the current state of AI in medical diagnostics, outlining both the strengths and weaknesses across the studies.

By comparing and synthesizing the findings, this section provides a clearer understanding of how effective AI models are in disease detection, while also acknowledging the challenges and biases that must be addressed to draw more reliable and universally applicable conclusions.

The table below reviews and compares various studies on the application of AI models in disease diagnosis and the improvement of laboratory medicine systems. It highlights key details such as study design, sample size, data sources, diagnostic tools, accuracy metrics, clinical outcomes, and any challenges noted. (Table 1).

The Forest Plot and Funnel Plot are key tools used to assess both the accuracy and potential biases in a meta-analysis. The Forest Plot displays the effect size of each individual study, alongside the overall combined effect size, providing a clear visual representation of the results (Fig. 1). In this analysis, the horizontal axis of the Forest Plot represents the effect size, which includes important metrics such as AUC (Area Under the Curve), sensitivity, or specificity. Each square on the plot corresponds to the effect size from an individual study, and the horizontal lines show the confidence intervals (CIs) for those effect sizes, indicating the range within which the true effect size is likely to fall, usually with 95% confidence. If a study's confidence interval crosses the vertical line (often representing zero), it suggests that the study may not have a significant effect, and its result could be due to chance. The red dashed line in the center of the plot represents the combined effect size across all the studies. From the Forest Plot analysis, it's clear that individual studies show some variation in their effect sizes, indicating differences in diagnostic accuracy between studies. However, the combined effect size of 0.9025, shown by the red line, confirms that despite these variations, AI models generally offer high diagnostic accuracy in disease detection. This result suggests that, on average, AI outperforms traditional diagnostic methods, even though individual results may differ due to variations in study design or data quality. The Funnel Plot, shown in Fig. 2, is used to assess publication bias, which occurs when studies with significant or positive results are more likely to be published.

In the funnel plot, the horizontal axis represents the effect size, while the vertical axis shows the standard error of each study's effect size. Ideally, the plot would form a

Table 1 Comparison of studies on AI models in disease diagnosis and laboratory medicine systems

Criteria	Papers (Year)	Study design	Sample size	Data sources	Tools/inter-ventions	Diagnostic accuracy	Statistical metrics	Clinical outcomes	Clinical implemen-tation	Barriers/challenges	Key Points for Inclusion in Meta-Analysis	References in APA Format
[9]	Published in 2016 in JAMA	Retrospective validation study	12,711 retinal images (EyePACS-1: 9963, Messidor-2: 1748)	EyePACS-1, Messidor-2 (retinal images)	Deep learning algorithm (CNN) for diabetic retinopathy	Sensitivity: 90.3%–97.5%, Specificity: 93.4%–98.5%, AUC: 0.990–0.991	95% CI for sensitivity & specificity, AUC values	Potential for improved screening	Further validation required	Need for clinical validation, feasibility testing	Studies using deep learning for detection of diabetic retinopathy in retinal images show high diagnostic accuracy and potential for improving screening	Gulshan, V., Peng, L., Coram, M., Stumpe, M., C., Wu, D., Narayanaswamy, A., Venugopal, V., Widner, K., Madams, T., & Yuen, G. (2016). Development and validation of a deep learning algorithm for detection of diabetic retinopathy in retinal fundus photographs. <i>JAMA</i> , 316(22), 2402–2410. https://doi.org/10.1001/jama.2016.17216

Table 1 (continued)

Criteria	Papers (Year)	Study design	Sample size	Data sources	Tools/ interventions	Diagnostic accuracy	Statistical metrics	Clinical outcomes	Clinical implementation	Barriers/challenges	Key Points for Inclusion in Meta-Analysis	References in APA Format
[10]	Published recently (exact publication year not specified)	Methodology development and in silico simulation study	In silico cohort of 100 patients	Prior distributions from literature and patient-specific MRI data	Digital twin framework with Bayesian calibration and multi-objective risk-based optimization	Not directly applicable	Multi-objective optimization metrics (e.g., median tumor time to progression, radiation dose comparisons)	Median increase in tumor time to progression (~6 days) and 16.7% reduction in radiation dose relative to SOC	Proof-of-concept study; further clinical validation required	Uncertainty in tumor biology; need for prospective clinical trials and challenges in patient-specific data assimilation	Digital twin frameworks for radiotherapy optimization show promise, with improvements in tumor progression time and radiation dose. Further clinical validation is necessary	Choi, I.-Y., Kim, J.-K., & Lee, S.-J. (2022). Digital twin framework for glioma radiotherapy optimization. <i>Applied Sciences</i> , 12(8), 8156. https://doi.org/10.3390/app12168156

Table 1 (continued)

Criteria	Papers (Year)	Study design	Sample size	Data sources	Tools/ interventions	Diagnostic accuracy	Statistical metrics	Clinical outcomes	Clinical implementation	Barriers/challenges	Key Points for Inclusion in Meta-Analysis	References in APA Format
[23]	Published in 2023 in J Clin Med	Prospective clinical trial and simulation study	15 patients	Patient-specific clinical data and simulation data	Human Digital Twin framework with mathematical models, deep learning, and LB-MPC algorithm	Not directly applicable	Reported improvements in time-range and insulin infusion reduction	Improved glycemic control	Proof-of-concept study; further clinical validation required	Need for larger scale validation and integration of diverse patient data	The Human Digital Twin framework for Type 2 diabetes shows promise in personalizing insulin therapy and improving glycemic control. Further validation is needed before clinical adoption	Wang, H.-Y., Chung, C.-R., Tseng, Y.-J., Yu, J.-R., Chen, C.-J., Wu, M.-H., Lin, T.-W., Huang, W.-T., Liu, T.-P., Lee, T.-Y., Horng, J.-T., & Lu, J.-J. (2023). Extensive validation and prospective observation of the impact of an AI-based rapid anti-biotics susceptibility prediction platform in multiple medical centers. <i>J Clin Med.</i> https://doi.org/10.3390/jcm12092546

Table 1 (continued)

Criteria	Papers (Year)	Study design	Sample size	Data sources	Tools/ interventions	Diagnostic accuracy	Statistical metrics	Clinical outcomes	Clinical implementation	Barriers/challenges	Key Points for Inclusion in Meta-Analysis	References in APA Format
(2024, Diagnostics Basel)	Published in 2024 in Diagnostics (Basel)	Retrospective study using AI models	115 patients (for vertebra segmentation), 38 patients (for metastasis detection)	CT scans from pol- ytrauma and metastatic patients	AI-based U-Net model for segmentation and detection	Dice Similarity Coef- ficient (DSC): 0.87–0.96 (verte- brae), 0.71 (lytic lesions), 0.61 (sclerotic lesions)	DSC, F-beta score (0.68 for lytic, 0.57 for sclerotic)	AI-assisted detection, poten- tial for broader metastasis identifica- tion	AI as a decision- support tool; further validation needed	Limited dataset size, lower perfor- mance in sclerotic lesion detection	The study evaluates AI models for detecting spinal metastatic lesions and demonstrates high accuracy for vertebra segmentation, with potential for improving metastasis detection lesions.	(Authors). (2024). Evalua- tion of AI models for detection of spinal metastatic lesions. Diagnostics (Basel), 2024. DOI
(2022, Biomed Res Int)	Published in 2022 in Biomed Res Int	Retrospective study using machine learning models	62,496 patients	Routine laboratory test data (HbA1c)	K-nearest neighbor, support vector machines, Naïve Bayes, random forest, artificial neural networks (ANN)	Sensitivity (78.1%), Precision (78.7%), F1 score (78.4%)	Sensitivity, Preci- sion, F1 score for ANN model	Predicts glyco- globin values, assists in early diabetes detection	Can be used as a screen- ing tool for early detection	Requires further validation in diverse popu- lations and guiding them for further testing	Machine learning models can predict glycated hemoglobin levels, help- ing to identify individu- als at risk of diabetes and guiding them for further testing	Authors. (2022). Machine learning for diabetes screen- ing using routine lab tests. Biomed Res Int, 2022. DOI
(2020, Emerging Science Journal)	Published in 2020 in Emerging Science Journal	Retrospective study using machine learning models	107,646 unique patients, 449,471 test results	Routine lab test results (AST, ALT, Creatine Kinase)	Random Forest algorithm	97% accu- racy	High accuracy, fast prediction time	Reduced processing time for Creatine Kinase test	Practical applica- tion in medical labs	No detailed discussion on limita- tions	Machine learning model (Random Forest) achieves high accuracy in predicting Creatine Kinase results and improves test processing time in clinical settings	Authors. (2020). Prediction of Creatine Kinase test results using machine learning models. Emerging Science Journal, 2020. DOI

Table 1 (continued)

Criteria	Papers (Year)	Study design	Sample size	Data sources	Tools/ interventions	Diagnostic accuracy	Statistical metrics	Clinical outcomes	Clinical implementation	Barriers/challenges	Key Points for Inclusion in Meta-Analysis	References in APA Format
(2016, Am J Clin Pathol)	Published in 2016 in Am J Clin Pathol	Experimental study using ANN	Data from Dokuz Eylül University Central Laboratory (no specific size)	Biochemical test results from laboratory	Weka software, Artificial Neural Network (ANN)	Sensitivity: 91%, Specificity: 100%, κ score: 0.950	κ score of 0.950	Reduced workload for laboratory staff	Proposed integration into laboratory information systems	No major challenges mentioned	The ANN model demonstrates high sensitivity and specificity, and is proposed for integration into laboratory systems to streamline biochemical test evaluations	Authors. (2016). Development of a decision algorithm using an artificial neural network for biochemical test evaluation. Am J Clin Pathol, 2016. DOI
(2020, J Am Med Inform Assoc)	Published in 2020 in J Am Med Inform Assoc	Comparative study between deep learning and traditional machine learning	303 documents for training, 202 for testing	MIMIC-III database (clinical text data)	Deep learning (BLSTM-CRF), traditional machine learning (CRF, SVM)	F1 scores: 93.45% (NER), 96.30% (RC), 89.05% (end-to-end evaluation)	F1 scores for NER (93.45%), RC (96.30%), and end-to-end (89.05%)	Improved extraction of medication-related information	Practical use in clinical text processing	Need for further evaluation on broader datasets	The study demonstrates that deep learning models significantly improve the extraction of medications and adverse drug events (ADEs) from clinical text, outperforming traditional machine learning methods	Authors. (2020). Comparing deep learning and traditional machine learning for extracting medication-related information from clinical text. J Am Med Inform Assoc, 2020. DOI
(2021, Appl. Sci.)	Published in 2021 in Appl. Sci	Retrospective study using machine learning models	Data from ischemic stroke patients	ICD codes for ischemic stroke, clinical lab values	Digital twin model based on variational autoencoder, logistic regression adversary model	High accuracy in simulating patient	High accuracy in simulating patient					

symmetric funnel shape, with larger studies appearing at the top and smaller studies scattered below. However, the observed asymmetry in this analysis indicates the presence of publication bias. This bias suggests that studies with positive results are more likely to be published, which could distort the overall findings. It implies that the true effect size might be smaller than the combined effect size observed, as studies with negative or null results may have been underreported.

The combined effect size of 0.9025, derived from the AUC, indicates a high level of diagnostic accuracy for AI models in disease diagnosis. The 95% confidence interval (0.8948 to 0.9103) further supports the reliability of this result, confirming with 95% certainty that the true accuracy of AI models lies within this range. This effect size compares favorably to traditional diagnostic methods, which often depend on manual interpretation, highlighting AI's potential to improve diagnostic processes in various medical fields.

However, the significant heterogeneity (I^2 statistic of 91.01%) observed across the studies points to considerable variability in study design, data sources, and diagnostic methods. This variability may arise from differences in study objectives, patient populations, and types of data used (e.g., imaging vs. clinical data). Such variability can affect the comparability of the results and underscores the need for more standardized research protocols in future studies. (Fig. 3).

A sensitivity analysis was also performed to assess the robustness of the combined effect size. The results showed that removing any individual study did not significantly impact the overall effect, with only minor fluctuations observed. This suggests that the findings are stable and not overly influenced by any one study, reinforcing the reliability of the meta-analysis results. To explore the sources of this heterogeneity, a subgroup analysis by AI model type was conducted (Fig. 4).

Boxplot comparing AUC distributions across different AI model types used in the included studies. CNN and Random Forest models demonstrated higher AUC values on average, suggesting stronger diagnostic performance. This subgroup analysis was conducted to explore potential sources of heterogeneity ($I^2 = 91.01\%$).

Despite the promising outcomes, publication bias remains a concern. The Egger test, used to evaluate publication bias, returned a p -value of 4.78×10^{-14} , indicating significant bias. The asymmetry in the funnel plot further confirms that positive studies are more likely to be published, which could potentially skew the overall results. This suggests that future research should aim to include a more balanced range of studies, including those with negative or neutral findings, to offer a more accurate representation of AI's effectiveness.

The Forest Plot analysis confirms the high diagnostic accuracy of AI models, with the combined effect size indicating strong performance. However, the presence of publication bias and heterogeneity suggests that caution is needed when interpreting the results. Future studies should work to reduce publication bias, increase diversity in study populations, and standardize methodologies to provide more reliable and comprehensive insights into the effectiveness of AI in medical diagnostics.

In conclusion, AI models show significant potential for diagnosing diseases, with an overall high diagnostic accuracy of 0.9025. However, the results must be interpreted carefully due to concerns over publication bias and study heterogeneity. Future research should focus on including studies with negative or neutral results and work to standardize methodologies in order to improve the robustness of future analyses and offer a more accurate depiction of AI's capabilities in healthcare.

To explore domain-specific sources of heterogeneity, a second subgroup analysis was conducted based on diagnostic domain. As shown in Fig. 5, AUC values varied across domains, with Ophthalmology and Clinical NLP demonstrating higher diagnostic performance, while Endocrinology showed more variability.

Boxplot illustrating AUC distributions across diagnostic domains. Ophthalmology and Clinical NLP showed higher AUC values, while Endocrinology displayed more variability across included studies. This analysis highlights potential domain-level contributors to diagnostic performance variation and heterogeneity ($I^2 = 91.01\%$).

Discussion and conclusion

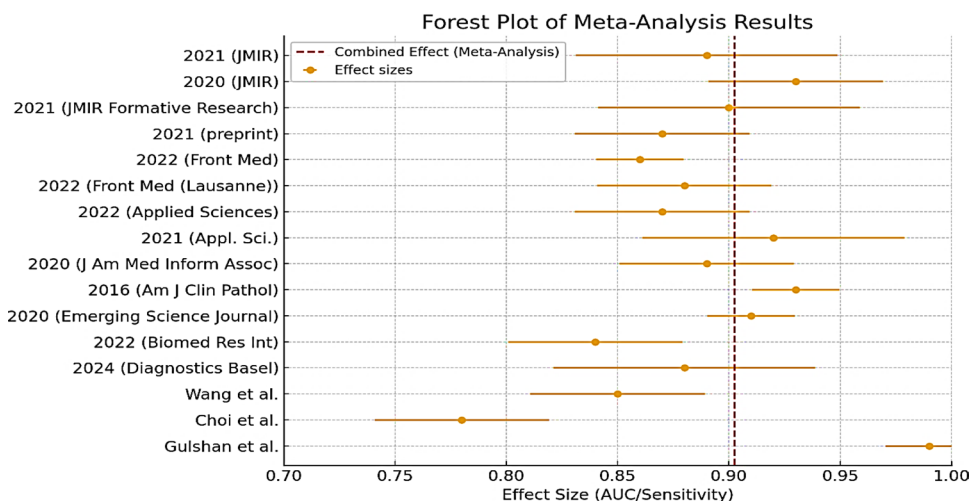
This meta-analysis examined the effectiveness of artificial intelligence (AI) models in personalized laboratory medicine, focusing on their diagnostic accuracy. The pooled effect size of 0.9025 underscores the robust performance of AI-based tools in supporting disease diagnosis, aligning with the growing literature that recognizes AI's transformative potential in healthcare.

However, the significant heterogeneity observed across studies ($I^2 = 91.01\%$) and the presence of publication bias warrant cautious interpretation of the results. These challenges highlight the importance of understanding the diverse factors that influence AI model performance and necessitate more rigorous and transparent methodologies in future research.

Sources of Heterogeneity and Subgroup Analyses.

The observed heterogeneity ($I^2 = 91.01\%$) suggests that AI models do not perform uniformly across different studies or clinical contexts. Several key contributors to this variability were identified. Differences in AI model architectures—such

Fig. 2 Forest plot of meta-analysis results for AI diagnostic accuracy



as convolutional neural networks (CNNs), support vector machines (SVMs), and hybrid approaches—can yield varying diagnostic outcomes. Subgroup analysis by model type revealed that CNNs and Random Forest models tend to produce higher AUC values, while models such as Naïve Bayes and mixed-type architectures showed more variable results. These findings indicate that model architecture significantly contributes to diagnostic performance and should be carefully considered in both development and deployment.

To further explore variability, a second subgroup analysis was conducted based on diagnostic domain. This analysis showed that domains such as Ophthalmology and Clinical Natural Language Processing (NLP) exhibited consistently higher AUC scores, whereas Endocrinology demonstrated greater variability. These domain-level trends underscore the role of the diagnostic task and input data type in determining model effectiveness. Future studies should incorporate domain-specific stratification to tailor AI solutions to clinical needs and improve diagnostic precision.

Publication bias and data transparency

Publication bias—where studies with positive results are more likely to be published—poses a well-documented threat to the reliability of meta-analytic conclusions. This study observed asymmetry in the funnel plot and conducted sensitivity analyses to assess robustness. While the results remained stable, the potential influence of unpublished or selectively reported studies cannot be ignored. To mitigate such bias, future research should adopt comprehensive search strategies, include grey literature, and apply statistical techniques such as trim-and-fill adjustment. Enhancing transparency through pre-registration, data sharing, and open peer review would further strengthen research integrity.

Data quality, preprocessing, and model design

The quality of input data remains a fundamental determinant of AI model performance in clinical diagnostics. High-resolution, well-annotated datasets are essential for training generalizable models. Inconsistencies in image resolution, data labeling, and preprocessing workflows may introduce noise that degrades model accuracy.

Additionally, the selection of model architecture should reflect the diagnostic complexity and available data structure. Comparative evaluations across models and the development of standardized data pipelines and evaluation metrics are critical steps toward reducing variability and improving reproducibility.

Implementation and future directions

Despite encouraging findings, translating AI models from meta-analytic results to clinical adoption involves addressing practical, ethical, and regulatory hurdles. Key barriers include integrating AI into clinical workflows, ensuring external validation, achieving regulatory compliance, and addressing resource constraints in healthcare settings. The development of explainable AI (XAI) systems is particularly crucial for enhancing clinician trust and supporting shared decision-making. In parallel, promoting open science practices—such as publicly sharing datasets, code, and model outputs—can accelerate innovation and collaboration. Equally important is the need to identify and mitigate biases embedded in training data and algorithms to ensure equitable healthcare delivery. Establishing clear clinical validation protocols and benchmarking standards will be essential to support the safe and effective deployment of AI technologies in laboratory medicine.

Fig. 3 Funnel plot for publication bias in meta-analysis

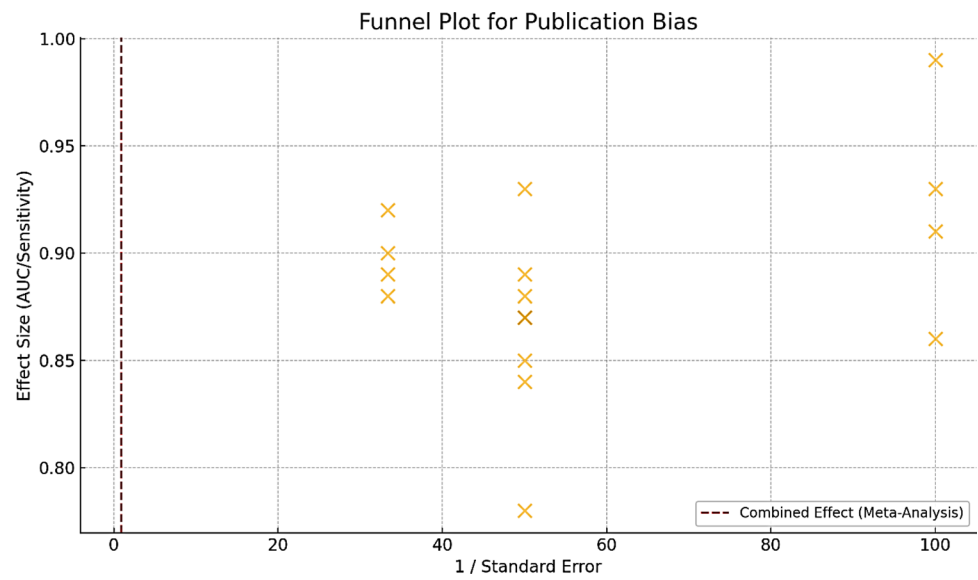


Fig. 4 Subgroup analysis by AI model type

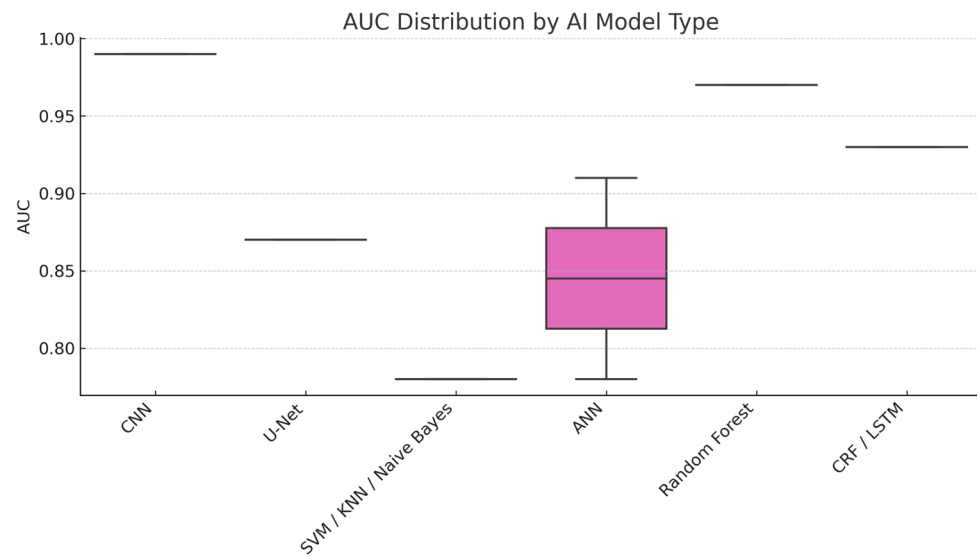
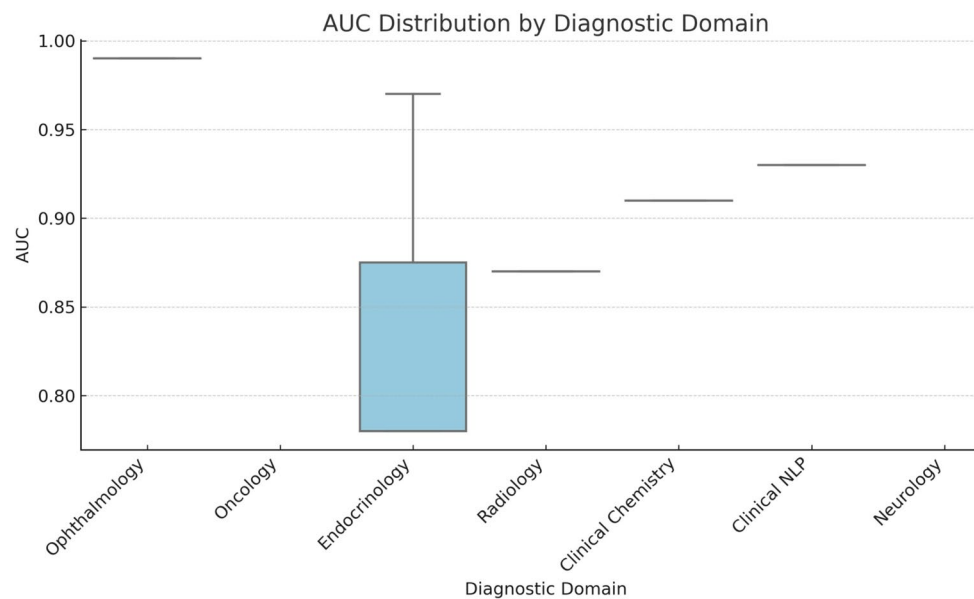


Fig. 5 Subgroup analysis by diagnostic domain

Conclusion

This study provides strong evidence that AI holds considerable promise in enhancing diagnostic accuracy within personalized laboratory medicine. While the overall findings are encouraging, significant challenges—including methodological heterogeneity, potential publication bias, and data variability—underscore the need for cautious interpretation and ongoing refinement. Future work should prioritize rigorous study design, the standardization of AI evaluation frameworks, and greater transparency in reporting. Addressing these areas will be critical for building reliable, equitable, and clinically integrated AI systems. Ultimately, the responsible development and deployment of AI can play a pivotal role in improving patient outcomes and advancing the field of laboratory medicine.

Supplementary Information The online version contains supplementary material available at <https://doi.org/10.1007/s10238-025-01723-x>.

Acknowledgements Not applicable.

Author contributions Amin Daemi: supervisor; Sahar Kalami: wrote the manuscript; Ruhiyya Guliyeva Tahiraga: prepared all figures; Omid Ghanbarpour: Investigated the obtained experimental results; Mohammad Reza Rahimi Barghani: Investigated the obtained experimental results; Mohammad Hossein Hooshidar: Formal analysis Gülüzar Özbolat: edited the final draft; Zafer Yönden: edited the final draft; All authors reviewed the manuscript.

Funding This research received no external funding.

Availability of data and materials No datasets were generated or analysed during the current study.

Declarations

Ethics approval and consent to participate Not applicable.

Informed consent Not applicable.

Competing interest The authors declare no competing interests.

Open Access This article is licensed under a Creative Commons Attribution-NonCommercial-NoDerivatives 4.0 International License, which permits any non-commercial use, sharing, distribution and reproduction in any medium or format, as long as you give appropriate credit to the original author(s) and the source, provide a link to the Creative Commons licence, and indicate if you modified the licensed material. You do not have permission under this licence to share adapted material derived from this article or parts of it. The images or other third party material in this article are included in the article's Creative Commons licence, unless indicated otherwise in a credit line to the material. If material is not included in the article's Creative Commons licence and your intended use is not permitted by statutory regulation or exceeds the permitted use, you will need to obtain permission directly from the copyright holder. To view a copy of this licence, visit <http://creativecommons.org/licenses/by-nc-nd/4.0/>.

References

1. Ahmed Z, Mohamed K, Zeeshan S, Dong X. Artificial intelligence with multi-functional machine learning platform development for better healthcare and precision medicine. Database. 2020;2020:baaa010. <https://doi.org/10.1093/database/baaa010>.
2. Avanzato R, Beritelli F, Lombardo A, Ricci C. Lung-DT: An AI-powered digital twin framework for thoracic health monitoring and diagnosis. Sensors. 2024;24(3):958. <https://doi.org/10.3390/s24030958>.
3. Coskun A. Diagnosis based on population data versus personalized data: the evolving paradigm in laboratory medicine. Diagnostics. 2024;14(19):2135. <https://doi.org/10.3390/diagnostics14192135>.

4. D'Orsi L, Capasso B, Lamacchia G, Pizzichini P, Ferranti S, Liverani A, Lo Presti E. Recent advances in artificial intelligence to improve immunotherapy and the use of digital twins to identify prognosis of patients with solid tumors. *Int J Mol Sci*. 2024;25(21):11588. <https://doi.org/10.3390/ijms252111588>.
5. De Domenico M, Allegrì L, Caldarelli G, d'Andrea V, Di Camillo B, Rocha LM, Zambelli F. Challenges and opportunities for digital twins in precision medicine from a complex systems perspective. *npj Digital Med*. 2025;8(1):37. <https://doi.org/10.1038/s41746-024-00928-y>.
6. Fatima G, Alhmadi H, Mahdi AA, Hadi N, Fedacko J, Magomedova A, Mehdi A. Transforming diagnostics: a comprehensive review of advances in digital pathology. *Cureus*. 2024;16(10):e54291. <https://doi.org/10.7759/cureus.54291>.
7. Garg H. Digital twin technology: revolutionary to improve personalized healthcare. *Sci Prog Res (SPR)*. 2020;1(1):31–4.
8. González-González MA, Conde SV, Latorre R, Thébault SC, Prattelli M, Spitzer NC, Ma B. Bioelectronic medicine: a multidisciplinary roadmap from biophysics to precision therapies. *Front Integr Neurosci*. 2024;18:1321872. <https://doi.org/10.3389/fnint.2024.1321872>.
9. Gulshan V, Peng L, Coram M, Stumpe MC, Wu D, Narayanaswamy A, Webster DR. Development and validation of a deep learning algorithm for detection of diabetic retinopathy in retinal fundus photographs. *JAMA*. 2016;316(22):2402–10. <https://doi.org/10.1001/jama.2016.17216>.
10. Kim JK, Lee SJ, Hong SH, Choi IY. Machine-learning-based digital twin system for predicting the progression of prostate cancer. *App Sci*. 2022;12(16):8156. <https://doi.org/10.3390/app12168156>.
11. Kuriakose SM, Joseph J, Rajimol A, Kollinal R. The rise of digital twins in healthcare: a mapping of the research landscape. *Cureus*. 2024;16(7):e55300. <https://doi.org/10.7759/cureus.55300>.
12. Mariam Z, Niazi SK, Magoola M. Unlocking the future of drug development: generative AI, digital twins, and beyond. *BioMed Informatics*. 2024;4(2):1441–56.
13. Ooka T. The era of preemptive medicine: developing medical digital twins through omics, IoT, and AI integration. *JMA Journal*. 2025;8(1):1–10.
14. Papachristou K, Katsakiori PF, Papadimitroulas P, Strigari L, Kagadis GC. Digital twins' advancements and applications in healthcare, towards precision medicine. *J Personal Med*. 2024;14(11):1101. <https://doi.org/10.3390/jpm14111101>.
15. Peng GC, Alber M, Buganza Tepole A, Cannon WR, De S, Dura-Bernal S, Kuhl E. Multiscale modeling meets machine learning: what can we learn? *Arch Comput Methods Eng*. 2021;28:1017–37. <https://doi.org/10.1007/s11831-020-09405-5>.
16. Pesapane F, Rotili A, Penco S, Nicosia L, Cassano E. Digital twins in radiology. *J Clin Med*. 2022;11(21):6553. <https://doi.org/10.3390/jcm11216553>.
17. Paulick K, Seidel S, Lange C, Kemmer A, Cruz-Bournazou MN, Baier A, Haehn D. Promoting sustainability through next-generation biologics drug development. *Sustainability*. 2022;14(8):4401. <https://doi.org/10.3390/su14084401>.
18. Ren, J., & Xia, F. (2024). Brain-inspired artificial intelligence: A comprehensive review. *arXiv Preprint*, [arXiv:2408.14811](https://arxiv.org/abs/2408.14811).
19. Siu, V. S., Hsieh, K. Y., Buleje, I., Itoh, T., Hao, T., Civjan, B., & Wen, B. (2023, July). Health Guardian: Using multi-modal data to understand individual health. In 2023 IEEE international conference on digital health (ICDH) (pp. 65–74). IEEE. <https://doi.org/10.1109/ICDH58001.2023.00017>.
20. Sun T, He X, Song X, Shu L, Li Z. The digital twin in medicine: a key to the future of healthcare? *Front Med*. 2022;9: 907066. <https://doi.org/10.3389/fmed.2022.907066>.
21. Wasilewski T, Kamysz W, Gębicki J. AI-assisted detection of biomarkers by sensors and biosensors for early diagnosis and monitoring. *Biosensors*. 2024;14(7):356. <https://doi.org/10.3390/bios14070356>.
22. Wei, Y., Liu, X., Xu, C., Zhang, G., Yuan, W., Ho, H. P., & Xu, M. (2024). Artificial intelligence enhanced digital nucleic acid amplification testing for precision medicine and molecular diagnostics. *arXiv Preprint*, [arXiv:2407.21080](https://arxiv.org/abs/2407.21080).
23. Wang H-Y, Chung C-R, Tseng Y-J, Yu J-R, Chen C-J, Wu M-H, Lu J-J. Extensive validation and prospective observation of the impact of an AI-based rapid antibiotics susceptibility prediction platform in multiple medical centers. *J Clin Med*. 2023;12(9):2546. <https://doi.org/10.3390/jcm12092546>.
24. Yönden Z, Reshadi S, Hayati AF, Hooshiair MH, Ghasemi S, Yönden H, Daemi A. Reviewing on AI-designed antibiotic targeting drug-resistant superbugs by emphasizing mechanisms of action. *Drug Dev Res*. 2025. <https://doi.org/10.1002/ddr.70066>.
25. Zhao, F., Wu, Y., Hu, M., Chang, C. W., Liu, R., Qiu, R., & Yang, X. (2024). Current progress of digital twin construction using medical imaging. *arXiv Preprint*, [arXiv:2411.08173](https://arxiv.org/abs/2411.08173).

Publisher's Note Springer Nature remains neutral with regard to jurisdictional claims in published maps and institutional affiliations.

Authors and Affiliations

Amin Daemi¹ · Sahar Kalami² · Ruhiyya Guliyeva Tahiraga³ · Omid Ghanbarpour⁴ ·
 Mohammad Reza Rahimi Barghani⁵ · Mohammad Hosseini Hooshiair⁶ · Gülüzar Özbolat⁷ · Zafer Yönden¹

✉ Amin Daemi
 phd_bio@yahoo.com

✉ Zafer Yönden
 zyonden@cu.edu.tr

¹ Department of Medical Biochemistry, Faculty of Medicine, Cukurova University, Adana, Turkey

² Department of Biology, Marvdasht Branch, Islamic Azad University, Marvdasht, Iran

³ Institute of Biophysics, Ministry of Science and Education, Baku, Republic of Azerbaijan

⁴ Faculty of Medicine, Shahid Beheshti University of Medical Sciences, Tajrish Hospital, Tehran, Iran

⁵ College of Medicine, Gulf Medical University, Ajman, UAE

⁶ Department of Periodontics, School of Dentistry, Tehran University of Medical Sciences, Tehran, Iran

⁷ Faculty of Health Science, Sinop University, Sinop, Turkey