AMERICAN JOURNAL OF BIOMEDICINE AND PHARMACY
https://biojournals.us/index.php/AJBP
ISSN: 2997-7177

Article

Artificial Intelligence in Interpreting Clinical Laboratory Results: A Scientific Review of its Modern Applications and Challenges

Dunya Abbas Mahmood*[1]

[1] Department of Chemistry, College of Education for Pure Sciences, University of Kirkuk, Kirkuk, Iraq
*Correspondence: dunyaabbas@uokirkuk.edu.iq

Citation: Mahmood D. A. Artificial Intelligence in Interpreting Clinical Laboratory Results: A Scientific Review of its Modern Applications and Challenges. American Journal of Biomedicine and Pharmacy. 2026, 3(4), 90-100.

Received: 10th Jan 2026
Revised: 11th Feb 2026
Accepted: 20th Mar 2026
Published: 10th Apr 2026

Copyright: © 2026 by the authors. Submitted for open access publication under the terms and conditions of the Creative Commons Attribution (CC BY) license (https://creativecommons.org/licenses/by/4.0/)

Abstract: AI and machine learning (ML) technologies are transforming clinical biochemistry throughout the analytical workflow, from QC and analytical error detection, to biomarker discovery and the interpretation of laboratory results. This review summaries recent evidence (2024–2026) in four biochemistry-centred areas: patient-based real-time QC and error detection; AI-assisted biomarker discovery from metabolomic and proteomic data; liquid biopsy and tumour biomarkers; and large language models (LLMs) for interpreting laboratory results. It also discusses the main barriers to routine use, namely limited interpretability, weak multi-centre validation, algorithmic bias and an evolving regulatory framework. We conclude that AI should be regarded as an analytical partner to the laboratory specialist rather than a replacement, and that its routine adoption requires rigorous verification standards, algorithmic transparency and regulation adapted to continuously updated models.

Keywords: Artificial intelligence; machine learning; clinical biochemistry; quality control; metabolomics; biomarker discovery; liquid biopsy; large language models; explainable AI.

Introduction

Clinical laboratories underpin most diagnostic and therapeutic decisions; more than three-quarters of clinical decisions are estimated to rely, directly or indirectly, on laboratory results [1]. At the same time, workload, analytical complexity and data volume continue to grow, particularly in biochemical and molecular testing [2]. This has created a need for analytical tools that can process large datasets quickly and accurately.

AI, including ML, deep learning and LLMs, now extends well beyond the automation of repetitive tasks. In clinical biochemistry it can detect subtle analytical errors missed by conventional QC, uncover patterns in multi-omics data and help interpret complex laboratory reports [3,4]. Because

the field is moving quickly, this review gives a concise and critical update on the most relevant biochemical applications of AI and on the limitations that still restrict their clinical use.

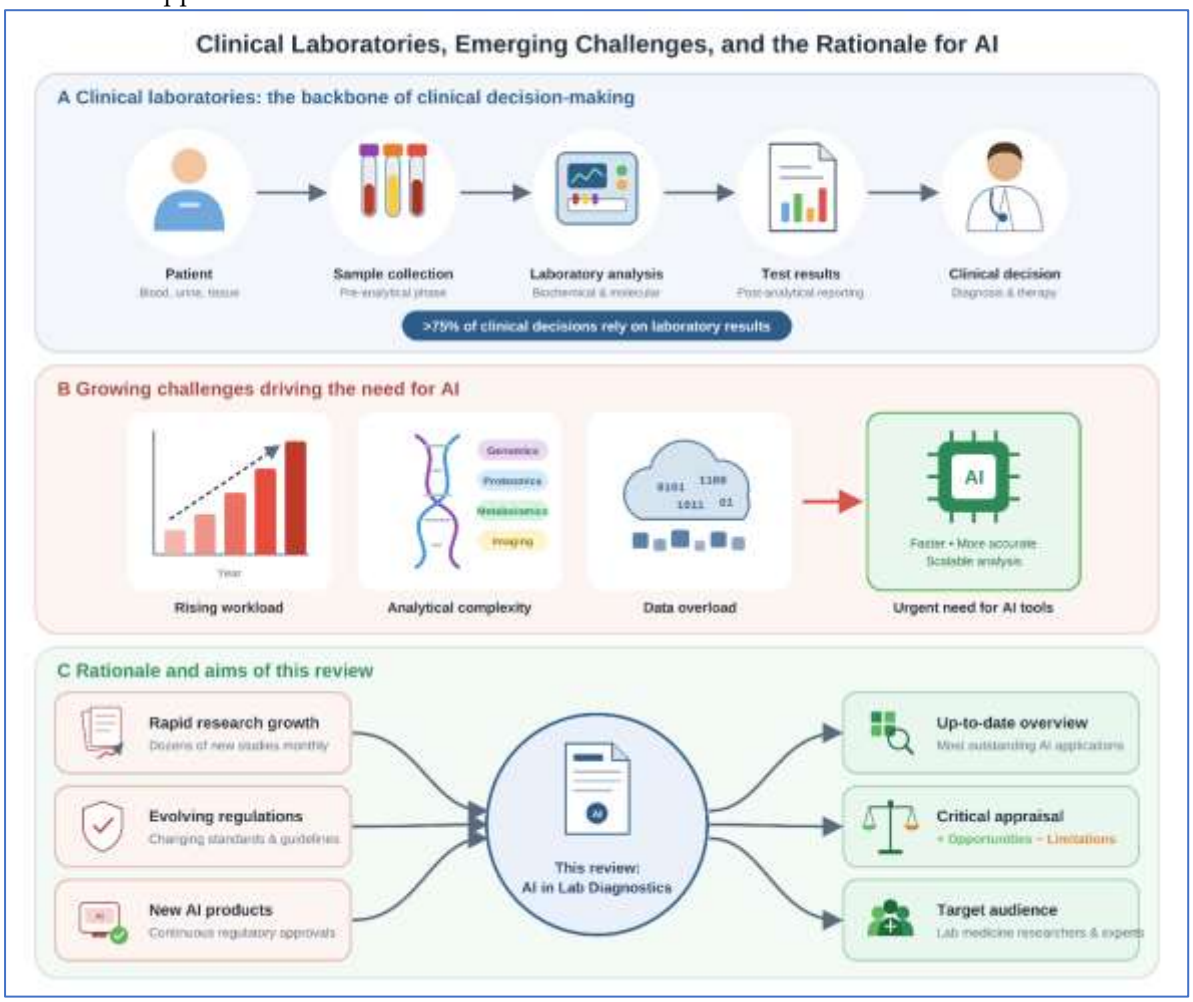

Figure 1. (A) The central role of clinical laboratories from sample collection to clinical decision. (B) Rising workload, analytical complexity and data volume driving the need for AI tools. (C) Rationale, scope and aims of the present review.

Methodology

This narrative review is based on a structured literature search covering 2020 to mid-2026, with emphasis on 2024–2026 publications. Multiple scientific databases were searched, and peer-reviewed studies directly relevant to AI applications in clinical biochemistry were prioritised. Selected systematic reviews and regulatory documents were consulted for context. As a narrative review rather than a PRISMA-based systematic review, it aims to provide a critical overview rather than a quantitative meta-analysis (Figure 2).

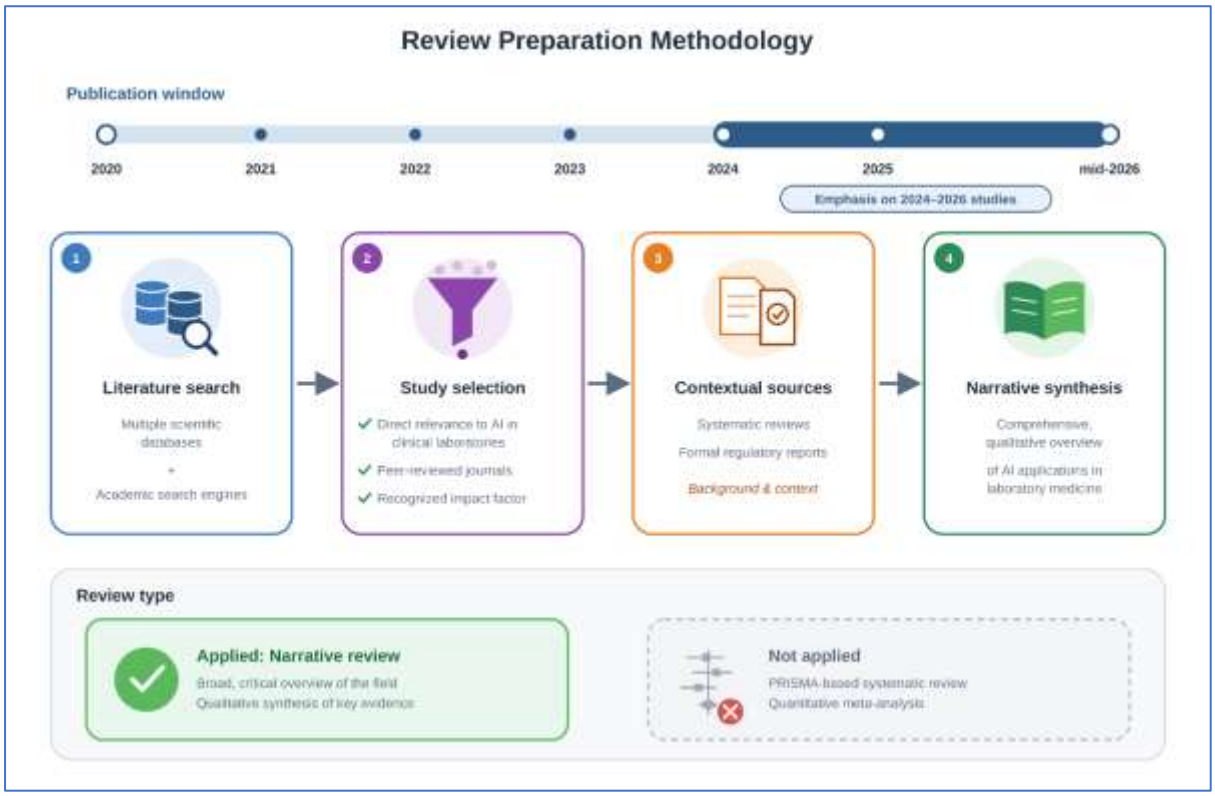

Figure 2. Review preparation: publication window, literature search, study selection criteria, contextual sources and narrative synthesis approach.

Conceptual Background: From Classical Algorithms to Generative Models

ML comprises algorithms that learn patterns from data without explicitly programmed decision rules, and includes supervised, unsupervised and reinforcement learning [5]. Deep learning, based on multi-layer neural networks, is particularly effective for complex, high-dimensional inputs [6]. Transformer-based models and LLMs add the ability to process unstructured reports and to combine text, images and numerical results within a single analytical framework [7] (Figure 3). Technological maturity, however, remains uneven: applications in imaging and pathology are closer to routine practice, whereas AI in clinical biochemistry still lacks sufficient multi-centre external validation [8,9].

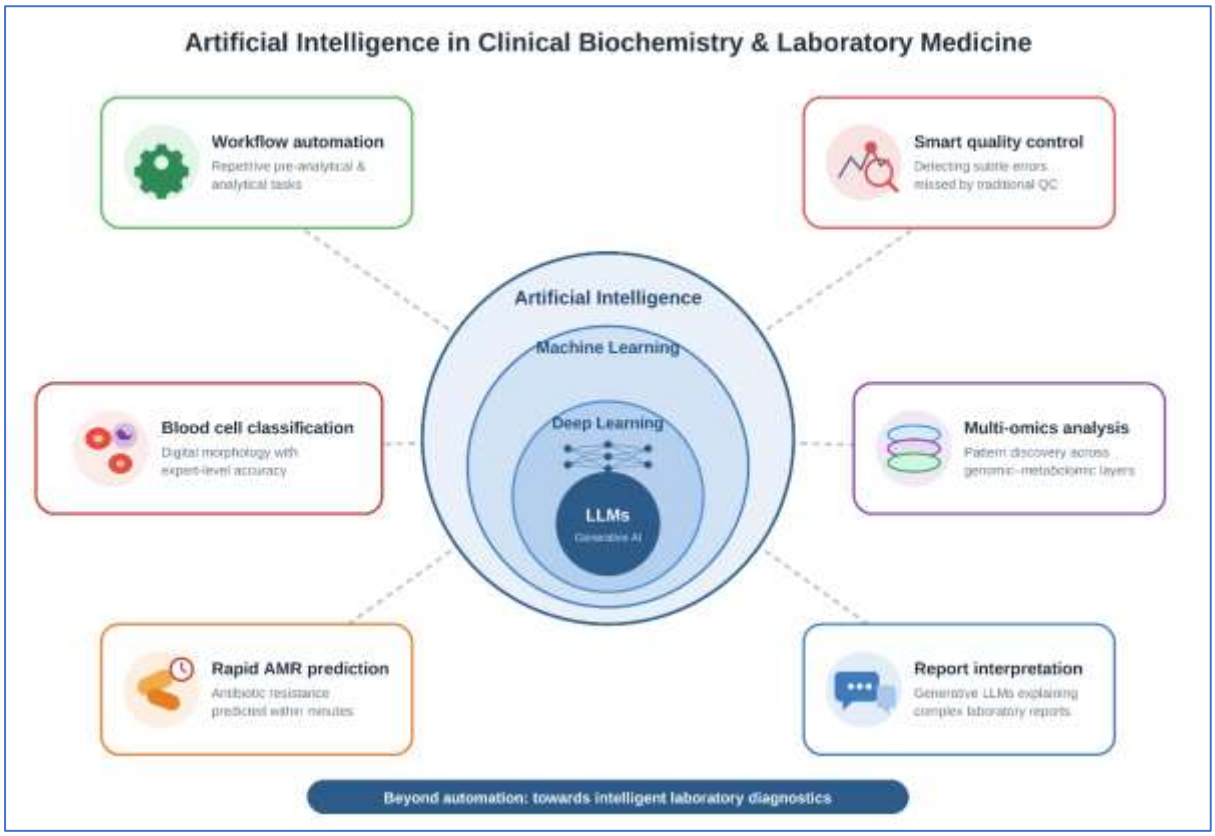

Figure 3. AI subfields and their key applications in clinical biochemistry and laboratory medicine.

AI in Quality Control and Analytical Error Detection

Patient-based real-time quality control (PBRTQC) is one of the most mature ML applications in clinical biochemistry. Unlike conventional QC, which relies on periodic analysis of control materials, PBRTQC monitors patient results continuously and can flag analytical deviations as they occur. A 2024 review described ML approaches for detecting three error types: systematic errors (bias, calibration drift); random or pre-analytical errors such as intravenous-fluid or EDTA contamination and delayed analysis; and sample misidentification [10]. Artificial neural networks have been used since 2016 to support autoverification of routine chemistry tests (sodium, potassium, creatinine and liver enzymes) using delta-check values and hemolysis–lipemia–icterus (HIL) indices [11].

Simulation studies have shown that gradient-boosting models such as XGBoost can detect drift, shift, noise and outliers in glucose, triglyceride and HDL-cholesterol data, supporting the integration of data-driven QC into automated chemistry analysers [12]. ML can also separate normal physiological variation from true analytical error, although standardisation is still needed before widespread implementation [12]. A retrospective multi-centre study further showed that ML models can detect sample misidentification errors, which are rare (about three per thousand samples) but clinically important and difficult to detect by conventional methods [13] (Figure 4).

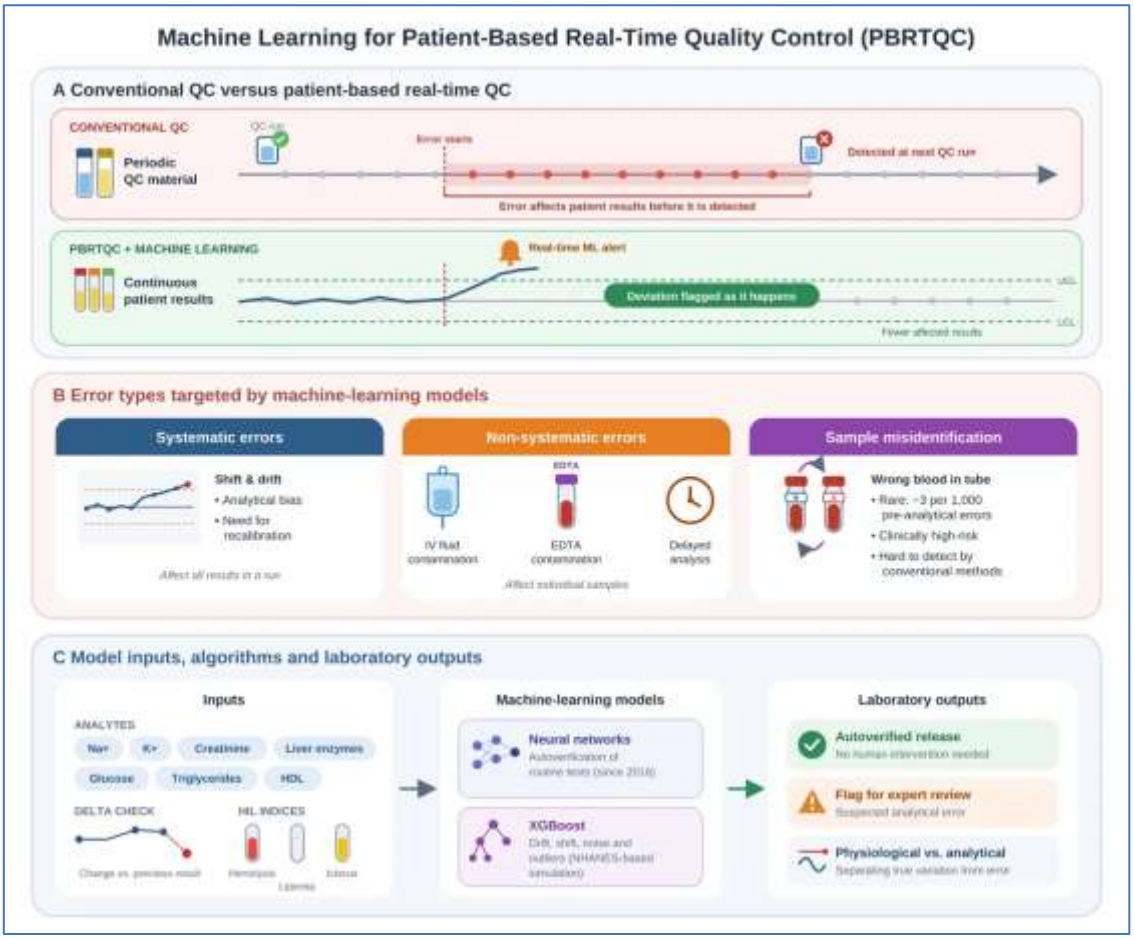

Figure 4. ML-based patient-based real-time quality control (PBRTQC). (A) Traditional periodic QC versus continuous patient-based monitoring with real-time ML alerts. (B) Error types addressed by ML models. (C) Model inputs, representative algorithms and laboratory outputs. HIL, hemolysis–lipemia–icterus.

AI-Assisted Biomarker Discovery from Omics Data

AI is increasingly used to discover biochemical biomarkers from multi-omics data, including the genome, transcriptome, proteome and metabolome. Neural networks, transformers, LLMs and feature-selection methods are the main techniques applied, although data quality, biological complexity, interpretability, reproducibility and generalisability remain major challenges [14]. These approaches have been applied in oncology, infectious, neurological and autoimmune diseases [15].

Critical perspectives are, however, growing. In clinical proteomics, ML-based biomarker discovery is often applied uncritically and is vulnerable to small sample sizes, batch effects, overfitting and data leakage, frequently yielding modest performance gains with limited interpretability [16]. Interpretable models offer a more cautious alternative: a tree-based metabolomic model distinguished individuals with Down syndrome from controls with 90.4% accuracy and an AUC of 95.9%, with SHAP analysis quantifying the contribution of each metabolite [17]. A similar combination of AutoML and explainable AI improved metabolomics-based diagnosis of hepatocellular carcinoma [18]. The main obstacles to clinical translation remain limited population diversity and insufficient independent validation [19].

Liquid Biopsy and Tumour Biomarkers

Liquid biopsy offers minimally invasive alternatives to tissue biopsy through analysis of circulating tumour DNA (ctDNA), extracellular vesicles (EVs), circulating tumour cells and tumour-educated platelets [20]. Advances in ultra-sensitive PCR, next-generation sequencing and epigenetic profiling have brought liquid biopsy into routine practice, beginning with the first FDA-approved plasma-based test for metastatic non-small cell lung cancer in 2016 [21].

EV analysis is a particularly active area. Electrochemical sensors are emerging as alternatives to ELISA, which is limited by cost, storage requirements and shelf life, and may improve test accessibility in resource-limited settings [22]. Endoscopically guided sampling of fluids adjacent to gastrointestinal and pancreatic tumours may also support early detection, although these markers are not yet fully validated [23]. ML has improved the detection of tumour-derived vesicles in very small samples; for example, EVs from ovarian cancer patients were detected with high sensitivity in no more than 10 μL of serum [24] (Figure 5).

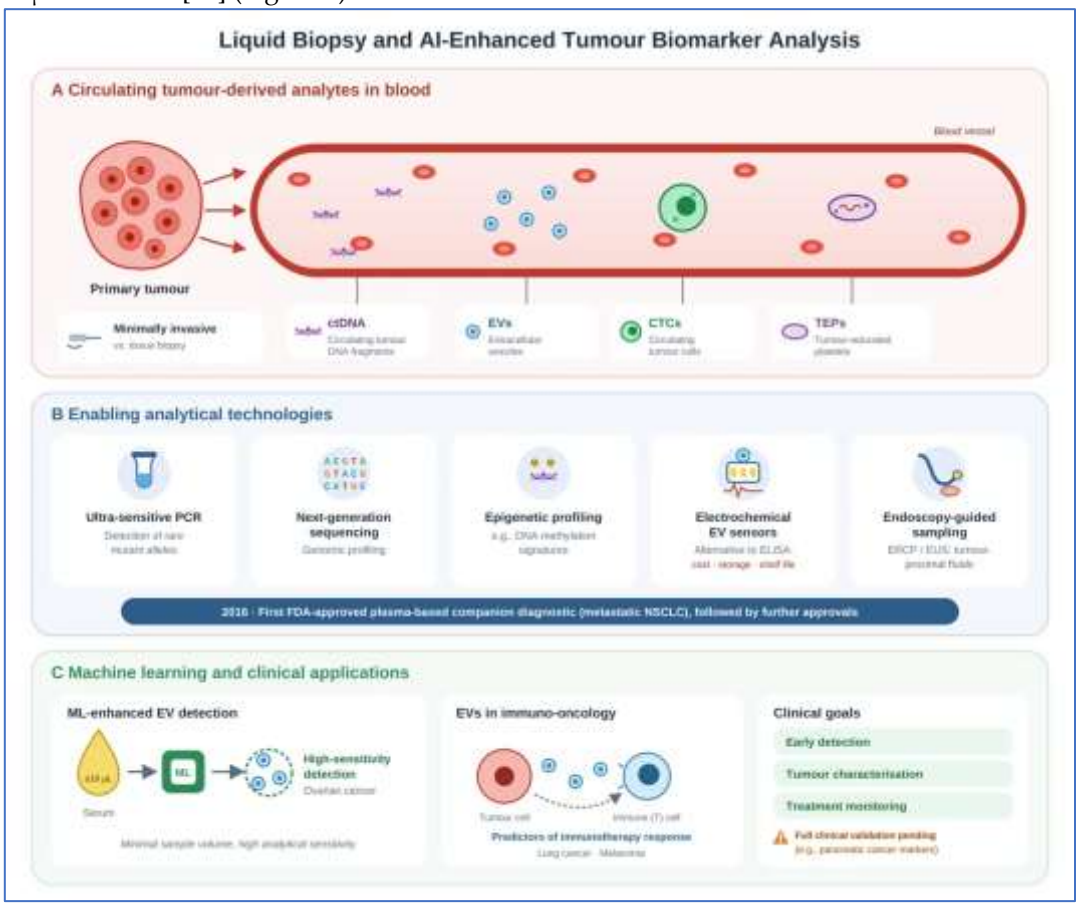

Figure 5. Liquid biopsy and AI-powered tumour biomarker analysis. (A) Circulating analytes derived from tumour tissue: ctDNA, EVs, circulating tumour cells (CTCs) and tumour-educated platelets (TEPs). (B) Enabling analytical technologies. (C) ML applications and clinical goals. ERCP, endoscopic retrograde cholangiopancreatography; EUS, endoscopic ultrasound; NSCLC, non-small cell lung cancer.

Large Language Models in Interpreting Laboratory Results

LLMs such as GPT-4 and the Claude model family can analyse unstructured clinical text together with laboratory results and patient context. Current evidence is, however, more cautious than media coverage suggests. In a study of 2,400 cases of four common abdominal conditions from the MIMIC database, open-source LLMs diagnosed less accurately than physicians, did not follow diagnostic and treatment guidelines, and could not reliably interpret laboratory results [25]. A later review also found that these models are not ready for making clinical decisions without close supervision from the physician [26].

Newer reasoning models perform considerably better. A model from the o1 series outperformed hundreds of clinicians in some diagnostic and management tasks, suggesting growing value as a decision-support tool [27]. In a retrieval-augmented generation (RAG) setting covering triage, referral and diagnosis, adding clinical information to the prompt did not significantly improve simpler models, highlighting the importance of prompt design [28]. There has also been development of a RAG framework to help interpret the recommendations of the hepatitis C management guidelines in the hospital setting [29].

Interpretability, Regulation and Methodological Challenges

The lack of transparency, "black box" problem, is a significant hurdle for the clinical use of high-performance models, and reduces trust from clinicians, especially in high stakes scenarios [30]. There are several common approaches to explainability, such as using SHAP to measure the importance of each feature in a prediction, local approximation (LIME) which approximates a complex model locally with a simpler, interpretable model, and attention mechanisms which allow to highlight the most influential inputs. A systematic review using the PRISMA framework revealed that the connection between explainability and trust is not guaranteed to be monotonic: explanations that are good helps build appropriate trust, explanations that are too simple can lead to over-reliance on the model, post-hoc explanations can only approximate the actual behaviour of the model [31]. Therefore, the concept of interpretability should be achieved through technical development, together with ethics and human-centred design [32,33] (Figure 6).

It's a dynamic environment, with regulations constantly changing. By July 2025, over 1250 medical devices had been authorized in the USA using AI; in August 2024, only 950 were approved, of which the majority were approved via the 510(k) pathway based on substantial equivalence to an existing medical device [34].In January 2025 the FDA issued draft guidance on lifecycle management of AI-enabled device software, covering model description, data sources, bias analysis and mitigation, human–AI workflow, post-market monitoring and Predetermined Change Control Plans (PCCPs), which address the continuously updated nature of AI models [35]. Regulators elsewhere, including the European Union, China, Japan and South Korea, have adopted differing approaches, while foundation models and generative AI remain largely unregulated [36]. Surveys of clinical biochemistry professionals in Turkey and Mexico reported positive attitudes towards AI but cited gaps in training, standards, trust in reliability and legal liability [37,38].

There are still a number of methodological issues. Large well annotated datasets are expensive to generate and many models have been trained and validated within a single institution and using homogeneous populations and instruments, with limited multi-center validation across analysers [12]. Under-representation of population groups can lead to algorithmic bias and a lack of accuracy for age, sex and ethnic groups, a problem that the FDA is now mandating that manufacturers solve [39]. Data leakage and overfitting are particularly problematic for high-dimensional small sample size omics data [16]. Integration with current workflows, user acceptance, data privacy, and ambiguity in legal liability for AI mistakes also hinder usage [39].

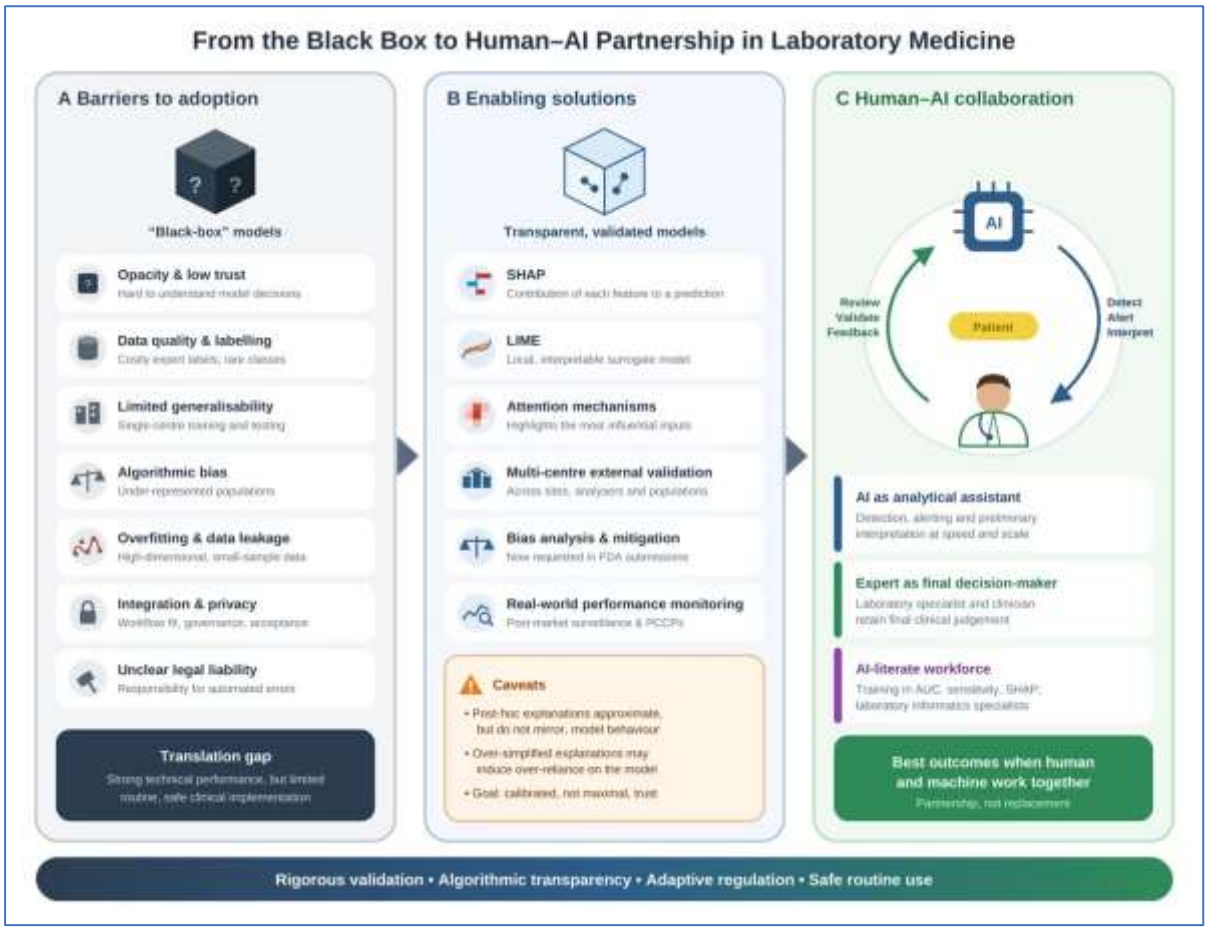

Figure 6. From the "black box" to human–AI partnership in laboratory medicine. (A) Key barriers to clinical adoption of AI. (B) Enabling solutions: Explainable AI (SHAP, LIME, Attention mechanisms), Multi-center external validation, Bias mitigation, Real-world performance monitoring. Human-AI collaboration model in which AI assists detection, alerting and preliminary interpretation, with the laboratory specialist at the end of the decision-making chain. PCCP: pre-determined change control plan.

Future Perspectives

The field will be defined by 3 trends. Multimodal models will combine laboratory data, images, genomic data and unstructured clinical text in a unified way [40] and the first foundation model to be cleared as an AI-enabled medical device in February 2025 has received regulatory acceptance of this model class. Second, Post-market monitoring of real-world performance will become a key element as the pre-market validation process will not ensure the continued safety of the developing models and explainability is now required for approval. Third, AI will not take over humans completely, and instead of that, human-AI collaboration models will be dominant, with AI assisting detection, alert and preliminary interpretation, and the laboratory specialist making the final decision. This also demands to train laboratory staff with the basic knowledge of ML and metrics for its performance (AUC, sensitivity, specificity, SHAP values) [2,37].

Conclusion

AI is revolutionizing clinical biochemistry, with data-driven and practical developments in automated QC and analytical error recognition, omics-based discovery of biomarkers, and the use of LLM to interpret lab results. Not yet at a stage of safe and reliable routine use as it will require external validation across multiple centers, interpretability, mitigation of algorithmic bias and regulation that is able to maintain pace with changes in models. The vision for the future of AI in clinical biochemistry is about collaboration, not replacement: Even though there is an intention to implement AI in the lab, the

potential of the technology does not diminish the need to rely on the clinical expertise of laboratory professionals, who will always provide the final, critical decision on clinical interpretation. The future of AI in the lab will require investment in staff training, harmonisation of verification standards, and transparency of algorithms.

REFERENCES

[1] Hicks, A. J., Carwardine, Z. L., Hallworth, M. J., & Kilpatrick, E. S. (2021). Using clinical guidelines to assess the potential value of laboratory medicine in clinical decision-making. Biochemia medica, 31(1), 74-84.

[2] Sciacovelli, L., Angelo Santini, S., & Plebani, M. (2023). Developing and enhancing all the professional activities for a successful future of Laboratory Medicine. Biochimica Clinica, 47, S63.

[3] 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.

[4] 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.

[5] Dhall, D., Kaur, R., & Juneja, M. (2019). Machine learning: a review of the algorithms and its applications. Proceedings of ICRIC 2019: Recent innovations in computing, 47-63.

[6] Sabri, O., Al-Shargabi, B., & Abuarqoub, A. (2025). The role of artificial intelligence in improving diagnostic accuracy in medical imaging: A review. Computers, Materials and Continua, 85(2), 2443-2486.

[7] Hartsock, I., & Rasool, G. (2024). Vision-language models for medical report generation and visual question answering: A review. Frontiers in artificial intelligence, 7, 1430984.

[8] Al-Antari, M. A. (2024). Advancements in artificial intelligence for medical computer-aided diagnosis. Diagnostics, 14(12), 1265.

[9] Ali, A., Akram, M. Q., Rasheed, F., Ali, A., & Jamil, I. (2025). Revolutionizing Diagnostics with Artificial Intelligence: Current Innovations, Challenges, and Future Horizons. Chronicles of Biomedical Sciences, 2(3), PID58-PID58.

[10] Liang, Y. F., Padoan, A., Wang, Z., Chen, C., Wang, Q. T., Plebani, M., & Zhou, R. (2024). Machine learning-based nonlinear regression-adjusted real-time quality control modeling: a multi-center study. Clinical Chemistry and Laboratory Medicine (CCLM), 62(4), 635-645.

[11] Demirci, F., Akan, P., Kume, T., Sisman, A. R., Erbayraktar, Z., & Sevinc, S. (2016). Artificial neural network approach in laboratory test reporting: learning algorithms. American Journal of Clinical Pathology, 146(2), 227-237.

[12] Lin, Y., Mensah, I. K., Doering, M., Shean, R. C., & Spies, N. C. (2025). Machine learning-based error detection in the clinical laboratory: a critical review. Critical Reviews in Clinical Laboratory Sciences, 62(7), 535-547.

[13] Seok, H. S., Yu, S., Shin, K. H., Lee, W., Chun, S., Kim, S., & Shin, H. (2024). Machine learning-based sample misidentification error detection in clinical laboratory tests: a retrospective multicenter study. Clinical Chemistry, 70(10), 1256-1267.

[14] Zhang, Y., Wang, T., Chen, H., & Li, X. (2021). Machine learning in biomarker discovery for precision medicine. Trends in Biotechnology, 39, 578590.

[15] Zhang, X., Rahnavard, A., & Crandall, K. A. (2025). Machine learning enhances biomarker discovery: From multi-omics to functional genomics. Medical Research Archives, 13(8).

[16] Adams, C., & Bittremieux, W. (2025). A 2025 perspective on the role of machine learning for biomarker discovery in clinical proteomics. Expert review of proteomics, 22(10), 363-374.

[17] Colak, C., Yagin, F. H., Yagin, B., Alkhateeb, A., Al-Rawi, M. B. A., Akhloufi, M. A., & Aghaei, M. (2025). Identification of metabolomics-based biomarker discovery in individuals with down syndrome utilizing kernel-tree model-enhanced explainable artificial intelligence methodology. Frontiers in molecular biosciences, 12, 1567199.

[18] Yagin, F. H., El Shawi, R., Algarni, A., Colak, C., Al-Hashem, F., & Ardigò, L. P. (2024). Metabolomics biomarker discovery to optimize hepatocellular carcinoma diagnosis: methodology integrating automl and explainable artificial intelligence. Diagnostics, 14(18), 2049.

[19] Javaid, H., Petrescu, C. C., Schmunk, L. J., Monahan, J. M., O'Reilly, P., Garg, M., ... & Martin-Herranz, D. E. (2025). The impact of artificial intelligence on biomarker discovery. Emerging Topics in Life Sciences, 8(2), 89-105.

[20] Nirob, S. R., Morol, M. K., Rahman, D., Hui, L. T., Nandi, D., Rahman, M., & Al Jubair, A. (2026). Early multi-cancer detection using liquid biopsy: emerging biomarkers and clinical strategies. Frontiers in Oncology, 16, 1877911.

[21] Moon, G. Y., Dalkiran, B., Park, H. S., Shin, D., Son, C., Choi, J. H., ... & Bu, J. (2025). Dual biomarker strategies for liquid biopsy: integrating circulating tumor cells and circulating tumor DNA for enhanced tumor monitoring. Biosensors, 15(2), 74.

[22] Shi, L., & Esfandiari, L. (2022). Emerging on-chip electrokinetic based technologies for purification of circulating cancer biomarkers towards liquid biopsy: A review. Electrophoresis, 43(1-2), 288-308.

[23] Huang, W., Zhao, M., Li, L., Chen, Q., Yang, M., Zhou, C., ... & Huang, D. (2026). Extracellular vesicle biomarkers: current status and future perspectives as novel tools in liquid biopsy. Frontiers in Immunology, 17, 1853671.

[24] Liang, Y., Lehrich, B. M., Zheng, S., & Lu, M. (2021). Emerging methods in biomarker identification for extracellular vesicle-based liquid biopsy. Journal of Extracellular Vesicles, 10(7), e12090.

[25] Hager, P., Jungmann, F., Holland, R., Bhagat, K., Hubrecht, I., Knauer, M., ... & Rueckert, D. (2024). Evaluation and mitigation of the limitations of large language models in clinical decision-making. Nature medicine, 30(9), 2613-2622.

[26] Vrdoljak, J., Boban, Z., Vilović, M., Kumrić, M., & Božić, J. (2025, March). A review of large language models in medical education, clinical decision support, and healthcare administration. In Healthcare (Vol. 13, No. 6, p. 603). MDPI.

[27] Bedi, S., Jiang, Y., Chung, P., Koyejo, S., & Shah, N. (2025). Fidelity of medical reasoning in large language models. JAMA Network Open, 8(8), e2526021.

[28] Gaber, F., Shaik, M., Allega, F., Bilecz, A. J., Busch, F., Goon, K., ... & Akalin, A. (2025). Evaluating large language model workflows in clinical decision support for triage and referral and diagnosis. npj Digital Medicine, 8(1), 263.

[29] Kresevic, S., Giuffrè, M., Ajcevic, M., Accardo, A., Crocè, L. S., & Shung, D. L. (2024). Optimization of hepatological clinical guidelines interpretation by large language models: a retrieval augmented generation-based framework. NPJ digital medicine, 7(1), 102.

[30] Liu, C. F., Chen, Z. C., Kuo, S. C., & Lin, T. C. (2022). Does AI explainability affect physicians' intention to use AI?. International journal of medical informatics, 168, 104884.

[31] Rosenbacke, R., Melhus, Å., McKee, M., & Stuckler, D. (2024). How explainable artificial intelligence can increase or decrease clinicians' trust in AI applications in health care: systematic review. Jmir Ai, 3, e53207.

[32] Mascarenhas, M., Mendes, F., Martins, M., Ribeiro, T., Afonso, J., Cardoso, P., ... & Macedo, G. (2025). Explainable AI in digestive healthcare and gastrointestinal endoscopy. Journal of Clinical Medicine, 14(2), 549.

[33] Nasarian, E., Alizadehsani, R., Acharya, U. R., & Tsui, K. L. (2024). Designing interpretable ML system to enhance trust in healthcare: A systematic review to proposed responsible clinician-AI-collaboration framework. Information Fusion, 108, 102412.

[34] Bahinski, A., Friere, M. C., McLellan, M. R., Psaty, B. M., Roden, D. M., & Steele, S. J. (2016). Scientific engagement at FDA: a report to the FDA science board from the scientific engagement subcommittee.

[35] Food, U. S., & Drug, A. (2025). Marketing submission recommendations for a predetermined change control plan for artificial intelligence-enabled device software functions. CfDEa Research, Editor.

[36] Zhang, S., Li, Y., Liu, W., Chu, Q., Wang, S., Li, J., & Chen, Y. (2025). A decade of review in global regulation and research of artificial intelligence medical devices (2015–2025). Frontiers in Medicine, 12, 1630408.

[37] İncir, S., Palaoğlu, K. E., Oğuz, O., Özgenç, A., Gören, Y., Topcu, D. İ., & Konukoğlu, D. (2026). Awareness, attitudes, and perceived barriers to artificial intelligence adoption in Turkish clinical biochemistry laboratories: An exploratory cross-sectional survey. Annals of Clinical Biochemistry, 00045632261454308.

[38] Sánchez-González, J. M., Morán-Moguel, M. C., Rivera-Cisneros, A. E., Ramírez-Barba, É. J., Sierra-Amor, R. I., Portillo-Gallo, J. H., & López-Rodríguez, M. (2025). Conceptualization of the use of artificial intelligence by clinical or research laboratory professionals: challenges for its implementation in Mexico. EJIFCC, 36(3), 366.

[39] Busch, F., Geis, R., Wang, Y. C., Kather, J. N., Khori, N. A., Makowski, M. R., ... & Bressem, K. K. (2025). AI regulation in healthcare around the world: what is the status quo?. medRxiv, 2025-01.

[40] Ran, W. H., Xi, X., Li, F., Lu, J., Jiang, J., Huang, H., ... & Li, S. (2025). Structured Semantics from Unstructured Notes: Language Model Approaches to EHR-Based Decision Support. arXiv preprint arXiv:2506.06340.

[41] Shinde, A., Deodhar, A., Gupta, V., Shinde, S., & Bava, N. (2026). Enterprise AI Agent Ecosystems: Architecture, Economics, and the Composition Gap. Economics, and the Composition Gap (April 11, 2026).