Opinion Paper

He S. Yang*, Hikmet Can Çubukçu, Fabio Del Ben, Xincen Duan, Jennifer Feng, Glynis Frans, Damien Gruson, Michael King, Javier Nieto-Moragas, Beili Wang and Fei Wang, on behalf of the International Federation of Clinical Chemistry and Laboratory Medicine (IFCC) Committee on Artificial Intelligence in Laboratory Medicine (C-AILM)

Implementation of AI systems in the clinical laboratory: insights from an expert survey and recommendations for best practice

https://doi.org/10.1515/cclm-2025-1322
Received October 7, 2025; accepted January 14, 2026;
published online February 3, 2026

Abstract

Objectives: Despite growing interest in artificial intelligence (AI) and machine learning (ML), many laboratory professionals lack experience with developing in-house AI systems or implementing those supplied by external providers. The IFCC Committee on AI in Laboratory Medicine (C-AILM) conducted a survey to collect the status of AI/ML applications, challenges, and expert perspectives on key technical considerations.

Methods: An 20-item survey was distributed to laboratory professionals experienced in AI. It covered application status (in-house or provider-supplied, with or without regulatory approval); essential information to request from AI system providers; validation or verification practices; monitoring strategies; and perceived implementation challenges.

Results: Fifty complete responses from global experts were received. AI implementation in clinical laboratories was limited and heterogeneous. Most respondents agreed that AI systems provided externally, regardless of regulatory approval status, require local verification. Key information needed from providers included performance metrics from original and external datasets, and demographics of the training/test populations. For both approved and non-approved models, high-priority verification studies were local performance analysis, confirmation of intended-use alignment, and verification of privacy and security safeguards. Top monitoring strategies were regular accuracy checks and comparison against human decision-making. Leading challenges were insufficient IT infrastructure and lack of practical implementation guidelines.

Conclusions: Although many challenges remain, clinical laboratories demonstrate strong enthusiasm for AI, particularly with the growing prevalence of commercial AI products. The timely expert insights from our survey and C-AILM recommendations for both AI system providers and clinical laboratories on essential information, verification requirements, and monitoring strategies will inform standardized guideline development.

Keywords: artificial intelligence; machine learning; laboratory medicine; AI systems; implementation; performance evaluation

***Corresponding author: He S. Yang**, Department of Pathology and Laboratory Medicine, Weill Cornell Medicine, 525 E. 68th Street, F707, 10065, New York, NY, USA, E-mail: hey9012@med.cornell.edu. https://orcid.org/0000-0001-5827-992X

Hikmet Can Çubukçu, Department of Medical Biochemistry, Sincan Training and Research Hospital, Ankara, Türkiye. https://orcid.org/0000-0001-5321-9354

Fabio Del Ben, Department Immunopathology and Cancer Biomarkers, CRO Aviano IRCCS National Cancer Institute, Aviano, Italy. https://orcid.org/0000-0002-1880-0669

Xincen Duan and Beili Wang, Department of Laboratory Medicine, Zhongshan Hospital, Fudan University, Shanghai, China. https://orcid.org/0000-0001-5167-170X (B. Wang)

Jennifer Feng, Hackensack Meridian School of Medicine, Nutley, NJ, USA

Glynis Frans, Department of Laboratory Medicine, UZ Leuven, Leuven, Belgium; and Department of Microbiology, Immunology, and Transplantation, KU Leuven, Leuven, Belgium. https://orcid.org/0000-0001-8528-5719

Damien Gruson, Department of Laboratory Medicine, Cliniques Universitaires St-Lux, Brussels, Belgium and Université Catholique de Louvain, Brussels, Belgium

Michael King, Roche Diagnostics GmbH, Penzberg, Germany. https://orcid.org/0000-0003-3201-416X

Javier Nieto-Moragas, Genetics Laboratory, Metropolitan South Clinical Laboratory, Bellvitge University Hospital, L'Hospitalet de Llobregat, Spain

Fei Wang, Department of Population Health Sciences, Weill Cornell Medicine, New York, NY, USA

Introduction

Laboratory Medicine is inherently data-rich, generating an immense volume of structured, discrete test results across a

wide range of analytes and specimen types. Most laboratory results are reported as individual numerical or categorical values, often with associated meta- and peri-data such as reference intervals and specimen details [1, 2]. The high dimensionality and complexity of the data, spanning multiple time points, test modalities, and clinical contexts, make it well suited for advanced data analytics [3]. In particular, artificial intelligence (AI) and machine learning (ML) techniques offer powerful tools for uncovering subtle patterns and interrelationships within laboratory data [4].

In recent years, there has been a rapid surge of interest in applying AI/ML to Laboratory Medicine. These AI systems have been explored for a variety of use cases [5], including but not limited to detection of pre-analytical errors [6], test result interpretation [7], optimization of test utilization [8], enhancement of laboratory efficiency [9], as well as prediction of disease onset and progression [10, 11]. Reflecting this interest, the number of publications on AI in Laboratory Medicine has grown substantially. However, despite the enthusiasm, real-world AI/ML implementations in the clinical laboratory remain quite limited. Many laboratory professionals are unfamiliar with the practical aspects of developing and validating in-house models, and often do not know what information to request from AI system providers or how to conduct local verification. A recent survey highlighted challenges such as inadequate data infrastructure and insufficient AI-related skills in clinical laboratories [12].

To better understand the current landscape, the IFCC Committee on AI in Laboratory Medicine (C-AILM) conducted an international survey targeting professionals in Clinical Pathology and Laboratory Medicine who have interest or experience in AI/ML. This survey aimed to assess the current application status of AI/ML in clinical laboratories, identify major challenges to adoption, and collect expert opinions on key implementation issues. Specifically, the survey explores what information about an AI system (whether or not approved by regulatory agencies) should be requested from AI system providers, such as vendors or manufacturers; what validation or verification studies are needed to assess performance of AI systems, and what procedures should be in place to monitor performance after deployment. In the survey itself, the term “model” was used; however, in this article, we adopt the term “AI system”, consistent with the latest European Union (EU) AI Act [13], with both terms referring to the same concept in this paper. Importantly, the survey questions focused on the post-development of AI/ML implementation, therefore model development or re-training was not discussed.

To date, there are limited guidelines or expert consensus on the best practice for AI/ML implementation in the clinical laboratory [14–16], despite the emergence of broader regulatory frameworks, such as the IVDR [17], EU AI Act [13], and the U.S. FDA’s Good Machine Learning Practice [18]. These frameworks provide high-level principles for AI governance, but they do not address the specific needs and operational contexts of laboratory medicine. Some publications have proposed that the validation of clinical AI could reference the regulatory framework for laboratory tests outlined in the Clinical Laboratory Improvement Amendments of 1988 (CLIA-88), including approaches to validation [19, 20]. However, laboratory data has distinct methodological and semantic complexities, such as method dependence, analyte-specific variability, and contextual sensitivity that extend beyond the existing general-purpose AI reporting guidelines [21]. Here, we summarize expert perspectives in the results section, and then in the discussion section, we propose a set of recommendations focusing primarily on the adoption of externally supplied AI systems. We hope our analysis of the survey and recommendations could serve as a first step in outlining practical guidelines that support safe, effective, and responsible adoption of AI/ML in laboratory medicine.

Methods

An expert survey (the entire questionnaire can be found in Supplementary Material) to collect current status and expert opinions on implementation of AI/ML applications in the clinical laboratory was generated using Qualtrics (Provo, Utah, USA). The survey covered seven domains and had a list of 20 questions in addition to the ability to add comments, as shown in Table 1.

The survey was distributed to the Association for Diagnostics & Laboratory Medicine (ADLM) Data Analytics Steering Committee and Data Science and Informatics Division, Canada Data Analytics Interest Group, European Federation of Clinical Chemistry and Laboratory Medicine (EFLM) Committee on Digitalization and Artificial Intelligence (C-AI), and laboratory professionals who were known to have experience and/or publications on AI/ML. The draft survey was reviewed by members of the IFCC Artificial Intelligence and Genomic Diagnostics Working Group and the ADLM Data Analytics Steering Committee, and was pilot-tested by our committee to ensure clarity and alignment with the survey objectives. Survey responses were collected from October 2024 to March 2025. A total of 82 individuals participated in the survey, but only 50 completed all the questions. Therefore, our analysis was based solely on the results of those 50 participants. The study was deemed exempt from institutional ethical committee approval because it involved only professional opinions regarding practice and perception, with no collection of personal identifiable private information.

Table 1: Domains and questions in the survey.

Domains	Questions	Response type
General questions	Q1. What is the name of your hospital/institution/organization and country (survey results will be analyzed anonymously)	Open text
	Q2. What is your institution type?	Multiple choic[a]
	Q3. What is your professional role?	Multiple choice[a]
	Q4. What is your specialty?	Multiple choice[a]
	Q5. What laboratory information system (LIS) is currently used in your laboratory?	Multiple choice[a]
	Q6. What electronic health records (EHR) system is currently used in your institution?	Multiple choice[a]
Implementation status	Q7. Does your EHR/LIS system or software used in your laboratory incorporate any AI/ML application?	Yes/no/not sure
	Q8. What tool/software used in your lab provides AI/ML application? If none, please type "N/A"	Open text
	Q9. What is the ML model used for? If there are multiple ML models, please select all their functions; If none, please select "N/A"	Multiple choice[a]
	Q10. Is the ML model used in your laboratory provided by an external provider or developed internally by the laboratory?	Multiple choice[a]
Opinions on models approved by regulatory agencies	Q11. If an ML model is developed by an external provider and has received clearance from a regulatory agency, such as FDA, what information regarding training and testing of the model should be released by the provider? Please provide your opinion even if you haven't adopted any ML model in your laboratory yet.	Multiple choice[a] + open text
	Q12. If an ML model has been approved by a regulatory agency, do you think it should be verified or re-trained before implementing it?	Multiple choice[a] + open text
	Q13. What studies do you think are necessary to evaluate an approved ML model provided externally? Please also provide a rank order.	Multiple choice[a] + ranking + open text
Opinions on models not approved by regulatory agencies	Q14. If the ML model is developed by an external provider and it has not been approved by a regulatory agency, what information regarding training and testing of the model should be released by the provider?	Multiple choice[a] + open text
	Q15. If the non-approved model is provided externally, what verification or validation studies do you think are necessary to evaluate the model? Please also provide a rank order.	Multiple choice[a] + ranking + open text
Strategies to evaluate and monitor ML models	Q16. Please elaborate on any specific procedure or criteria you use to evaluate model performance	Open text
	Q17. How would you monitor the performance of AI/ML tools which are currently used in clinical laboratories? Please provide your opinion on monitoring strategies even if you haven't adopted any ML model in your laboratory yet.	Multiple choice[a] + open text
Potentials and challenges	Q18. What are the challenges of adopting AI/ML in clinical laboratories	Multiple choice[a] + open text
	Q19. What do you see as the greatest potential for AI/ML tools to enhance your laboratory efficiency and patient care?	Open text
Information needs	Q20. What information would be most useful to you for adopting AI/ML into your laboratory?	Open text

[a]Indicates multiple selections are allowed.

Descriptive statistics were used to summarize and analyze the collected survey data. Excel and R were used for data processing and analysis. The word cloud figure was created in Python using the Wordcloud Package [22]. Additionally, the Borda count method was applied to rank voting-based results [23]. The Borda count method makes use of the full ranking information provided by each respondent rather than only the top choice and it yields an easily interpretable score that reflects both how often an option is ranked highly and how consistently it is prioritized across respondents. Participants were asked to rank multiple options in order of preference. For each voter, points were assigned in descending order based on the ranking position: if there are n candidate answers, the top-ranked option receives n−1 points, the second-ranked receives n−2 points, and so on, with the least preferred candidate receiving 0 points. The total Borda count for each candidate was calculated by summing the points across all voters. The candidate with the highest cumulative score was considered the most preferred overall.

Results

1. Specialty and professional role of the participants, their institute type, and electronic health record (EHR)/laboratory information system (LIS) used in their institutes (Q1-Q6)

Among the 50 responses received, 18 were from Europe, 18 from the United States, seven from Canada, six from Asia, and one from Africa. Figure 1 shows the distribution of participants globally. Among the 50 participants, 39 identified their institution as an academic medical center. This group included three participants who also selected a reference laboratory, three also selected community public laboratory services, and two selected both. Additionally, one participant identified their institution solely as a reference laboratory, four as independent laboratories, four as community hospitals, one as a public hospital laboratory, and one as consultancy.

Of the 50 respondents, 12 were laboratory medical directors, four informatics pathologists, 14 laboratory managers or specialists, and one research scientist. Others reported mixed roles, including combinations such as a laboratory medical director with informatics or research roles (n=6), laboratory director/scientist (n=1), and laboratory manager/research scientist (n=2). Additional individual roles included chemical biochemist (n=3), resident or trainee (n=1), pathologist (n=2), medical biologist specialist (n=1), chemical pathologist (n=1), consultant (n=1), and specialty student (n=1).

Among them, 25 specialized solely in clinical chemistry, while others reported expertise in clinical microbiology (n=1), hematopathology (n=1), hematology/coagulation (n=1), and informatics (n=2). The remaining participants listed multiple specialties, most commonly combinations involving clinical chemistry and informatics (n=10), as well as broader overlaps across clinical chemistry, microbiology, hematology/coagulation, cytopathology, molecular pathology, and anatomical pathology.

The participants' institutions used a highly diverse range of LIS and EHR systems. Among LIS, the most frequently used were EPIC Beaker (20 %) and Cerner Millennium (20 %), while EPIC was the most commonly reported EHR (34 %). The remaining systems were heterogeneous and varied widely.

2. Implementation status of AI and ML tools in the clinical laboratory

The survey first asked what tools or software in their laboratories provide AI/ML functionality (Q7-8). Here, "tools" refer to the specific implementation mechanisms, such as commercial software modules in the LIS/EHR, standalone applications, middleware, or custom scripts developed by the laboratory. We then asked what these AI/ML models are used for (Q9), meaning the functional applications or tasks

Figure 1: Global distribution of survey participants. Participant locations were determined using the latitude and longitude from their IP addresses. Since some cities had multiple participants, the different colored pins indicate how many participants were in each city. Blue, red, yellow and green represent 1, 2, 3, and 6, respectively.

the models perform. We also asked whether these models were developed in-house or provided externally (Q10).

Using this distinction, 34 % (17/50) of respondents reported employing 22 AI-enabled tools, which were implemented either through middleware solutions (8/22) or through custom scripting approaches (14/22). By scripting approaches, it is assumed that AI/ML models were not integrated into commercial laboratory software or modular EHR/LIS components but instead operated as standalone scripts or lightweight utilities executed outside the routine workflow. Only one respondent referred to using a LIS with built-in AI capabilities, though no details of those functions were provided and was not included in the final count.

Across these tools, 17 respondents described 41 applications. Most applications fell into two main categories: image-based and numerical tasks, the latter being more numerous. The most common use cases were image classification and discrimination that mainly referred to slide review and cell morphology analysis (10/41, 24 %). These were followed by the flagging abnormal laboratory results (7/41, 17 %), interpreting complex laboratory panels (7/41, 17 %) and monitoring quality controls (7/41, 17 %). Additional use cases included identifying pre-analytical errors (e.g., wrong blood in tube, IV fluid contamination, hemolysis) (6/41, 15 %), predicting the accuracy of measured results (3/41, 7 %), and predicting onset and progression of disease (1/41, 2 %). Regarding provenance, 10 respondents reported using vendor-supplied regulatory-approved ML models, while 12 reported using models developed internally or by another institution without regulatory approval. Among them, five selected both options, indicating the use of both internally and externally supplied models.

3. Key information required for externally developed AI systems

For AI systems developed by external developers, the respondents were asked what information regarding training and testing of the model should be disclosed or requested. The questions covered two scenarios: whether a model is approved (Q11) or not approved (Q14) by a regulatory agency. All respondents, regardless of whether they had actually implemented an AI system in their laboratory, were invited to provide their views for all the questions on this topic.

The responses for both approved and non-approved models showed similar prioritization patterns (Table 2). The top three most frequently selected responses were “performance metrics used to evaluate the model in the original dataset” (91 % & 87 %), “demographics of the patient population used for model training and testing” (89 % & 87 %), and “performance metrics if the model is tested externally” (81 % & 81 %). Notable free text comments under “Other” pointed out the importance of model stability & change control, as well as details about parameters that went into the model (e.g., reagents, vendors, and other technical components used to measure the parameters).

4. Studies needed to locally verify the performance of AI systems provided externally

When asked whether verification or retraining is needed before implementing approved ML models (Q12), 41 respondents (82 %) answered that the model should first be verified with local data, while 23 (46 %) selected retraining or customization using local data. Some respondents selected both options. Despite recognizing the need for validation, three respondents (6 %) admitted they didn’t know how to validate a model. Another three (6 %) believed the model could be used exactly as approved, without any additional checks. Five respondents (10 %) selected “it depends,” and provided additional considerations, including (1) evaluation of potential bias on the training set; (2) case-by-case evaluation on retraining, fine-tuning, and recalibration/adaptation after thorough discussion among diverse stakeholders;

Table 2: Key information required for externally developed ML tools, either approved or non-approved by regulatory agencies.

Answer (multiple answers possible)	Approved ML tools Number of respondents (percentage of respondents)	Non-approved ML tools Number of respondents (percentage of respondents)
Performance metrics used to evaluate the model in the original dataset	45 (91 %)	43 (87 %)
Demographics of the patient population used for model training and testing	44 (89 %)	43 (87 %)
Performance metrics if the model is tested externally	40 (81 %)	40 (81 %)
Model architecture and features incorporated in the model	37 (75 %)	40 (81 %)
Fairness/bias analysis (if applicable)	35 (71 %)	39 (79 %)
Explainability of the decision made by the model	31 (63 %)	34 (69 %)
Other	9 (18 %)	3 (6 %)

(3) risk-based necessity, i.e. high-impact predictive uses vs. low-risk applications; (4) retraining if the model fails local verification; and (5) alignment with ISO 15189 style verification of manufacturer-validated tools.

For both approved and non-approved ML models, respondents ranked the priorities of verification studies in a similar order (Q13, Q15), indicating consistent preferences across regulatory status (Figure 2). In both scenarios, local performance analysis garnered the highest Borda score and the greatest proportion of first-place rankings. Verification of alignment with the model's intended use in the local setting and assessment of privacy and security of the algorithm's implementation were also ranked highly. Mid-tier priorities included external benchmarking, fairness, and interpretability. External reference systems can act as trusted comparators or gold standards against which the model's performance is evaluated, providing users a more objective understanding of the model's accuracy and reliability. Meanwhile, the nearly equal weighting of non-discrimination, fairness and interpretability signals an awareness of ethical-legal imperatives, but also suggests these considerations are often deferred until core technical soundness and security are established. At the opposite end, the two lowest-ranked verification studies were patient-outcome trials in the local population and, ranked last, simply checking the model against values supplied by the external developers, e.g., manufacturer's claims.

5. Studies and criteria to evaluate and monitor an AI system's performance implemented in the clinical laboratory

In response to the open-ended question Q16, 13 participants provided free-text comments. Respondents emphasized the need for rigorous analytical validation, large and representative training and testing populations to minimize bias, and sufficient sample sizes to ensure statistical significance. They recommended comparing algorithm predictions with gold-standard methods (e.g., manual microscopy, traditional diagnostic tests) or expert clinical assessments. Local validation with institution-specific data was the preferred starting point, with external validation valued when datasets are sufficiently similar to the intended deployment population. Their feedback underscored that the evaluation of ML models in laboratory medicine should follow principles analogous to traditional assay validation/verification and adhere to established guidelines, such as those from CLSI [24]. In addition, respondents stressed that model evaluation should extend far beyond simple statistical measures of performance (e.g., accuracy, sensitivity, specificity, F1 score). Broader, application-specific performance evaluations were recommended, along with interdisciplinary evaluation teams to assess the model's alignment with its intended use case and the tangible clinical or operational outcomes it may influence. These teams should also

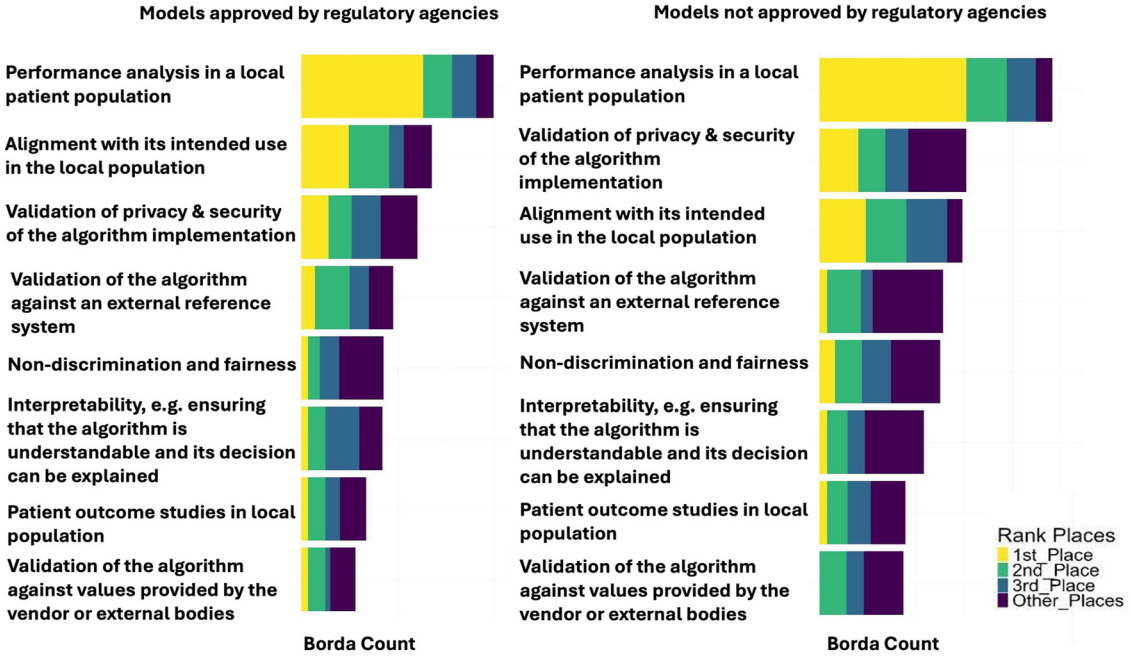

Figure 2: Prioritization of verification studies for both approved and non-approved machine learning models. Each bar represents the cumulative Borda count for a given priority, with color indicating the ranking places: yellow=1st place, green=2nd place, blue=3rd place, and purple=other places.

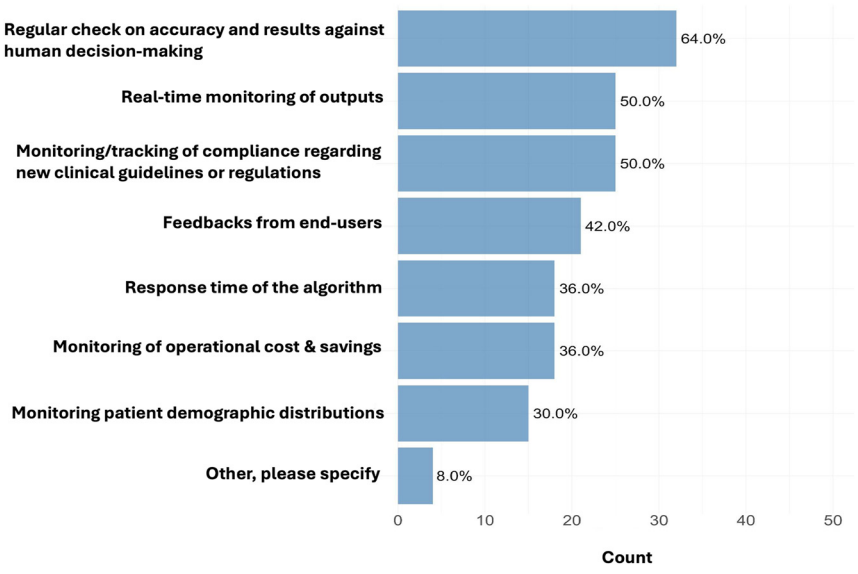

Figure 3: Summary of a multiple-choice question regarding the strategies to monitor model performance after deployment. Bars represent the proportion of respondents select each method, expressed as percentage.

evaluate the availability and quality of alternative solutions, whether a fully automated or human-in-the-loop approach is preferable, and how the model will be monitored and updated over time. Regardless of the model's regulatory approval status, the clinical setting for its application should align with that in which it was originally developed, and the underlying clinical workflow should be carefully examined before implementation. Furthermore, regulatory compliance with frameworks such as IVDR, EU AI Act, and FDA requirements was viewed as essential. Several participants noted that evaluation strategies should be proportional to the model's risk level and interpretability, rejecting a one-size-fits-all approach.

When asked how clinical laboratories should monitor the performance of ML models (Q17), as shown in Figure 3, most respondents (64 %) selected "regular checks of accuracy and results against human decision-making", a process viewed as essential for maintaining accuracy and reliability, analogous to quality control procedures for conventional laboratory tests. Half of the respondents (50 %) selected "real-time monitoring of outputs" and "tracking compliance with new clinical guidance or regulation", followed by "feedback from end-users" (42 %) as another important monitoring approach. Fewer respondents chose measures such as "algorithm response time" or "monitoring operational cost and savings", which focus on operational impact.

6. Potentials and challenges of implementing AI systems in clinical laboratories

Regarding the challenges of adopting AI/ML in clinical laboratories (Q18), as shown in Figure 4, the most frequently selected item (88 %) was "lack of adequate IT infrastructure to support ML applications". In addition, 78 % of the respondents also pointed out "lack of guidelines on ML implementation in the laboratory medicine field" and "lack of expertise in ML implementation" as major concerns. Furthermore, 66 % selected "challenges to integrate ML into workflows". In contrast, fewer respondents viewed trust issues or difficulties in collecting local data for validation as major barriers to ML adoption. A smaller number also mentioned additional challenges in the "Others" category as free text, including concerns related to data privacy, regulatory compliance, limited reimbursement pathways for quality-focused applications, and the lack of interoperability across health information systems.

We then asked the respondents a free-text question (Q19) about the greatest potential of AI/ML to enhance laboratory efficiency. The responses were categorized into topics for easier interpretation (a word cloud of detailed topics is shown in Figure 5). The greatest perceived potential was identified in workflow improvement & automatization, cited by 34 % of respondents, encompassing the automatization of routine tasks, avoidance of unnecessary procedures, and identification of process bottlenecks. The second most cited area was testing result interpretation and reporting (30 %), such as assisting result interpretation, and ensuring consistency and quality of results sent to clinicians. This was followed by patient safety applications (20 %), such as error detection, real-time alerts of critical patients, and identification of patient risks.

Other closely related topics included quality control (18 %) and error detection (14 %), which aim to improve test quality through patient-based real-time QC (PBRTQC), enhancing auto-verification, and detecting pre-analytical

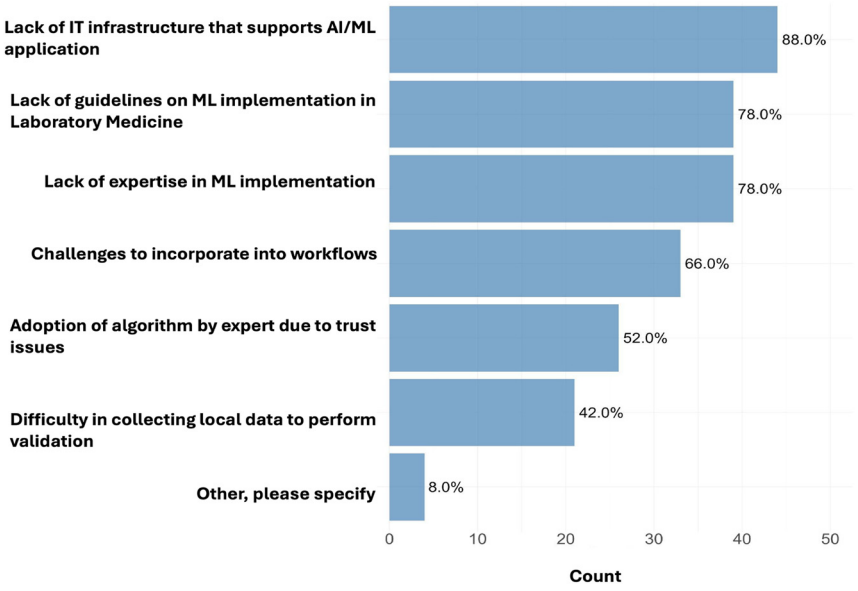

Figure 4: Summary of a multiple-choice question regarding the challenges of adopting AI/ML in clinical laboratories. Bars represent the proportion of respondents selecting each challenge, expressed as percentages.

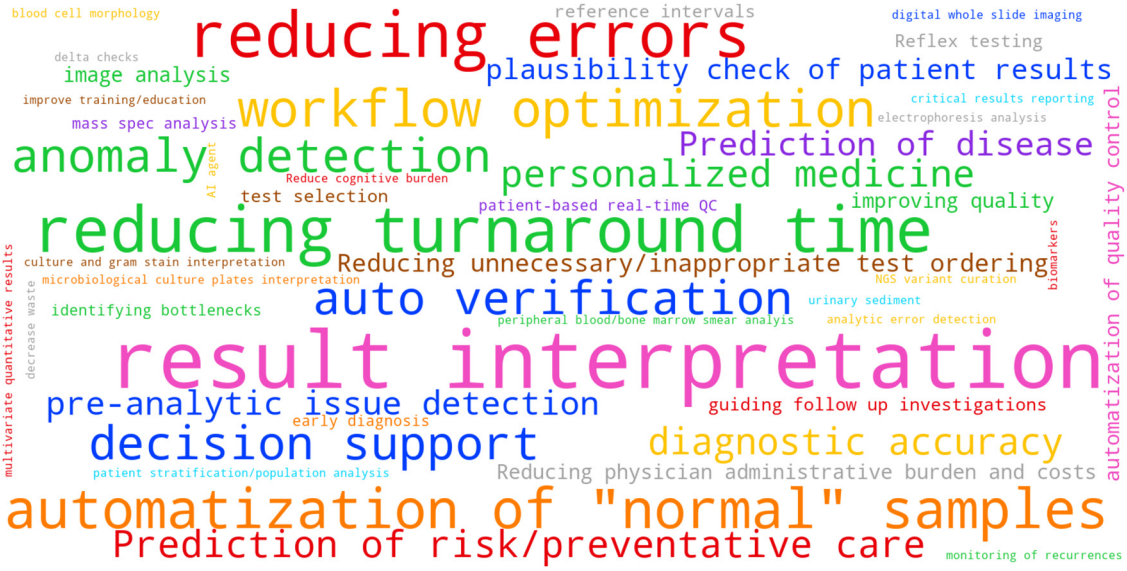

Figure 5: Word cloud summarizing perceived potential potentials of AI/ML to enhance laboratory efficiency. Keywords were extracted and categorized from free-text responses, with font size reflecting frequency of mention by respondents.

errors. The respondents also saw potential for AI/ML to enhance diagnostic accuracy and testing (18 %), with examples including consistent interpretation of culture and Gram stains, multivariate quantitative analysis, mass spectrometry data analysis, and image analysis (e.g., urinary sediment, blood cell morphology, microbiological culture).

Furthermore, respondents saw opportunities in clinical decision support & personalized medicine (18 %), as well as proactive patient care & preventive medicine (18 %), through improved test ordering guidance, interpretation support, and prediction of disease risks, outcomes, and early diagnosis. Smarter testing (14 %), such as optimizing test reflexing and reducing unnecessary or inappropriate orders, were also noted. Lastly, guidance and assistance applications (16 %) were mentioned, including reducing administrative burden, supporting training and education, and guiding users through processes, interpretations, and investigations.

Last, we asked respondents the free-text question Q20 about the information needed for adopting AI/ML in the laboratory (Q20). The most frequently encountered needs were clear guidelines (38 %) and regulatory requirements (18 %). These included guidance on model selection, validation in the general population, and verification and

implementation in laboratory settings, as well as clarity on regulatory requirements to ensure patient safety and define responsibility for errors. For adoption, they emphasized the need for information on model development, including analytical performance, training data characteristics (e.g., demographics), and a clear business case outlining anticipated benefits.

For implementation, respondents highlighted technical requirements such as data preparation procedures, and specifications for the type, volume, and quality of data necessary to ensure accuracy and reproducibility, along with vendor-provided support and training. Finally, participants noted the importance of guidance on integrating ML tools with existing systems (LIS, EHR, middleware), for example, through vendor-provided access points, and for deployment solutions embedded within existing platforms.

Discussion

Our survey revealed that, even among laboratory professionals interested in AI, implementation remains limited, and applications are highly heterogeneous. Despite the growing enthusiasm for AI, most clinical laboratories currently lack the in-house expertise to develop their own models. At the same time, more external providers are producing AI systems designed for integration into clinical workflows through the EHR and LIS. The emerging trend is toward adopting external models, much like laboratories now use commercial immunoassays, i.e. laboratories are not expected to develop immunoassays themselves; rather, they follow established verification protocols. A similar paradigm is likely to evolve for AI/ML systems, making verification of external models and continuous performance monitoring critical topics. Through our survey, we collected expert perspectives on these topics and, based on their insights, we propose several recommendations for consideration.

1. What specific information should laboratories request from AI system providers for clinical use?

The American Medical Informatics Association (AMIA) has emphasized the importance of transparency in clinical AI through guidelines that recommend clear labeling of an AI system's intended purpose, characterization of the training data population, and testing across diverse user groups to promote equity [25]. Similarly, the minimum information about clinical artificial intelligence modeling (MI-CLAIM) checklist was developed to improve transparency in model development, evaluation, and reporting [26]. In the clinical laboratory setting, we propose that a standardized model specification sheet, analogous to "Instruction for Use (IFU)" documents for clinical laboratory assays, should be offered by providers. Such a specification sheet should clearly define the model's intended use in the laboratory, including its purpose, expected outcomes, and whether human-in-the--loop is required. It describes the training, testing, and validation populations (including patient demographics), and reports performance metrics on both the original development set and independent external validation sets, together with documented safeguards, known limitations, and risk-management strategies of how hazards are identified, evaluated, and controlled across the model life cycle. In addition, the document is expected to outline system operational requirements for routine use, including the nature and format of input data, required preprocessing steps, anticipated data volumes and throughput, system dependencies, and the model's operating ranges and limits, i.e. the conditions under which outputs are accurate and reliable, the expected values or ranges of its output to enable verification and monitoring, and the scenarios where performance degrades or fails. Examples of such model facts labels have been published in literature or online [27, 28].

To further facilitate appropriate interpretation and use, transparent descriptions of the model's computational approach and practical guidance on output interpretation should be provided. Performance metrics of the model should be selected to align with the model's intended clinical application in laboratory medicine [29]. Recent guidance recommends reporting across five key performance domains: discrimination, calibration, overall performance, classification, and clinical utility, with area under the receiver operating characteristic curve (AUROC), calibration plots, clinical utility measures (e.g., net benefit with decision curve analysis), and outcome-stratified probability distribution as essential elements [30]. Information on long-term stability (e.g., seasonality effects, expected model drift, retraining triggers) and transferability (data of external validation, applicability across sites and work areas) should also be included [31, 32], together with details of validation cohorts and, where appropriate, comparisons against established systems or manual methods [33]. Finally, providers may offer recommendations tailored to local implementation, such as how to construct local test datasets, preprocessing steps, and monitoring metrics.

AI systems that utilize laboratory results are influenced by the total testing process, including the pre-analytical phase, analytical phase, and post-analytical phase. Measurement uncertainty introduces additional uncertainty in

the output of AI systems [34, 35], while systematic bias in input measurements can further distort the output [36]. Therefore, AI system providers should clearly report the conditions of their development environment, specifying pre-analytical handling procedures, the analytical performance characteristics of the measurement systems used for each measurand's analysis, and post-analytical considerations. When an AI system uses a non-harmonized measurand, providers should demonstrate the applicability of their system across different measurement systems, considering that inter-method bias arises from the lack of standardization or harmonization and its impact on output of the AI system. Data-driven approaches to harmonize results across different measurement systems, as reported previously [37], can be applied to address this challenge.

2. What approaches and studies are needed to verify the performance of models provided externally in the local laboratory setting?

Most survey respondents viewed regulatory approval as a starting point rather than a green light. Their expectations for the verification studies were nearly identical for both approved and non-approved models, suggesting that laboratory professionals evaluate both categories through the same lens. The top-ranked verification studies reflect a risk-management perspective in which decision-makers first ensure that the algorithm performs accurately in their own environment, is applied to the task for which it was designed and does not compromise sensitive health information. Together, they form a tightly linked "validation triad" requiring empirical evidence that the model works on local data, addresses the intended clinical question, and safeguards patient confidentiality and system integrity.

In laboratory medicine, the terms "validation" and "verification" have distinct meanings. According to ISO/IEC 17000: 2020 [38], validation refers to conformation of plausibility for a specific intended use through objective evidence and typically involves a comprehensive assessment of performance characteristics for newly developed, modified, or laboratory-developed methods by the provider. Verification, in contrast, refers to confirmation that a method or system that has already undergone validation performs as expected in the local laboratory environment. Accordingly, when adopting externally provided AI models in clinical laboratories, independent local verification is essential to confirm that performance aligns with the claims stated in the model specification sheet under the intended conditions of use. Variations in patient populations, disease prevalence, instrumentation, middleware, and pre-analytical processes may substantially affect model behaviors, underscoring the need for local assessment prior to deployment.

The verification process should begin with confirmation that the AI system is technically compatible with the laboratory's existing hardware, software, and system versions. This can be verified using a reference or standard dataset provided by the developer to confirm that the model performs as promised under controlled conditions. The second step is to implement the model in the laboratory environment and test it on a representative local dataset that is aligned with the model's intended clinical use [15]. Process validation concepts applied in the verification of instruments and equipment, such as installation qualification (IQ), operational qualification (OQ), and performance qualification (PQ), can also be applied to the implementation and verification of the model [39]. Observed performance metrics should be compared with user requirements defined by a user requirement specification (URS) document and those reported in the provider's specification sheet. This verification helps determine whether the model achieves the promised level of accuracy and reliability when applied to local data.

AI system providers should actively support the implementation by confirming the correct system operation and, where possible, assist in evaluating fairness and detecting potential biases in the local dataset. The joint effort between the provider and laboratory allows for a transparent assessment of the model's generalizability and fitness for local use. Importantly, the laboratory must critically evaluate whether the metrics reported in the specification sheet are clinically meaningful and relevant for the specific local environment.

Finally, our survey indicated that 46 % of respondents were in favor of retraining or customizing an AI system. Laboratories should be aware, however, that under the IVDR and EU AI Act, retraining a commercial, approval AI system may constitute a "substantial modification", thereby transforming the laboratory from a "user/deployer" to a legal "manufacturer", with corresponding responsibilities and obligations, which most clinical laboratories do not have resource to manage [40]. This regulatory reality reinforces the importance of clearly distinguishing between local verification and model modification when deploying AI systems in routine clinical laboratory practice.

3. What strategies should be implemented to monitor and ensure sustained performance of AI systems following deployment in clinical practice?

A clinical laboratory should monitor AI system's performance through both provider-managed change control and laboratory-run surveillance. On the provider side, much like recalibrating hardware after a sensor swap, AI system providers must have mechanisms to ensure that any software patch, component substitution, or algorithm update does not alter the model's expected behavior. Routine integrity checks such as checksums, hash-based fingerprinting, and version-control verification should confirm that the model remains unchanged. Providers are expected to supply release notes with detailed documentation of version history, update log, and side-by-side comparative performance metrics whenever a modification is made. Importantly, when upstream systems such as the EHR or LIS undergo upgrades (e.g., changes in field names, communication protocols, or data formats), providers are expected to support laboratories in promptly re-verifying each AI/ML application to ensure continued correct operation.

On the laboratory side, continuous local monitoring is critical. Laboratories should routinely inspect input data distributions, such as the proportion of samples exhibiting specific features (i.e. feature positive rate) and the distribution of predicted categories (i.e. class frequency) and compare them against historical baselines. Monitoring changes in input data distribution, which are often referred to as data drift, can indicate conditions under which the model may no longer perform as originally validated [41]. Data drift can lead to emerging bias in model output or gradual, unrecognized degradation of model performance. QC procedures tailored to application type should be implemented: for image-based models, such as morphology analyzers, regular blinded expert review of representative sample sets can detect systematic bias or drift. For laboratory numeric value-based models (e.g., flagging abnormal results, predictive scores), synthetic or curated test cases with known expected outputs can be used in regression testing to confirm consistent behaviors. In addition, a continuous model monitoring framework, analogous to PBRTQC in laboratory testing [42, 43], could provide an advanced layer of surveillance by automatically tracking input and output distributions and detecting shifts in distribution caused by unexpected deviation, rather than relying on periodic retrospective audits [44]. This continuous monitoring is particularly valuable for detecting subtle or sudden errors, such as those caused by an update in database or data transfer protocols. Furthermore, external quality assurance (EQA) programs, where identical test datasets or standardized cases are distributed across laboratories, can provide an independent benchmark of performance, reveal inter-laboratory variability, and uncover systemic issues that may not be visible within a single site. Overall, the integration of application-specific QC, patient-based real-time monitoring, and EQA into routine surveillance ensures that performance issues are detected and addressed promptly, minimizing the risk of prolonged undetected errors in clinical decision support.

The survey results indicated that, while laboratories recognize the value of operational metrics, clinical performance remains the primary priority when monitoring AI/ML systems after deployment. Respondents consistently emphasized that ongoing evaluation must be risk-adapted, clinical focused, and capable of detecting model drift, an approach that aligns with the long-standing culture of laboratory medicine, where safeguarding accuracy, reliability, and clinical validity is paramount. Accordingly, high-impact applications (e.g., error detection, critical value prediction, diagnostic support) require more intensive and continuous monitoring than low-risk operational tools.

The risk-stratified monitoring approach is highly consistent with the emerging regulatory framework governing high-risk AI systems in laboratory medicine. Both the EU AI Act [13] and IVDR [17] mandate comprehensive, lifecycle risk management and post-market surveillance for high-risk AI-based medical device software [45, 46]. Within these frameworks, responsibilities are clearly delineated between AI system providers and laboratories as deployers. Providers (e.g., manufacturers) are required to establish post-market monitoring, report serious incidents, and document field safety corrective action (EU AI Act Arts. 72–73; IVDR Art. 82). Laboratories, considered as "users" under the IVDR, are obligated to monitor system performance during routine use, suspend operation when incidents or serious risks arise, notify providers and competent authorities (Eu AI Act Art. 26(5); IVDR vigilance Art. 82), retain system-generated logs for ≥6 months (Eu AI Act Art. 26(6)), and manage nonconforming work and adverse incidents in accordance with ISO 15189:2022 requirement [47]. Together with similar lifecycle-oriented elements in the U.S. FDA guiding principles for Good Machine Learning Practice (GMLP) in medical device development [48], these frameworks signal a clear regulatory convergence toward risk-adapted, post-deployment monitoring as a core component of AI governance in laboratory medicine. This trajectory is

further reinforced by the forthcoming ISO/CD 24051 standard, which will define general principles for the application of AI in clinical laboratories (Part 1 [49]) and in digital pathology and AI-based image analysis (part 2 [50]).

Limitations

We acknowledge that this study has limitations. First, the survey posed challenging questions that required substantial thoughts and relevant experience, which have deterred some individuals from completing it. Those less familiar with or skeptical of AI/ML were less likely to respond, limiting generalizability to all institutions. Therefore, there is a selection bias in the survey population. Nonetheless, the responses we collected were valuable and meaningful, which reflected expert perspectives rather than the broader spectrum of clinical laboratories. Importantly, this survey was not designed for hypothesis-testing nor subject to formal psychometric validation; instead, it was purposefully constructed as an expert-opinion survey to capture informed judgement in a rapidly evolving field where empirical implementation data remain limited. Accordingly, the ranked priorities and perceived importance reported here should be interpreted as expression of expert opinions and professional judgement rather than statistically generalizable estimates.

Second, respondents were predominantly from academic centers, which typically have greater resources, dedicated research teams, and early-adopter cultures. Consequently, the results may overestimate feasibility and underrepresent challenges faced by community hospitals, private laboratories, and under-resourced settings. In addition, there was a geographic imbalance in the collected responses, with most from North America and European countries. Finally, the participant pool was weighted toward clinical chemistry, with comparatively few voices from hematology, microbiology, molecular diagnostics, and other specialties, leading to underrepresentation of discipline-specific issues and constraining cross-specialty comparisons of implementation practices.

Conclusions

Despite the heterogeneous and still limited implementation of AI/ML in clinical laboratories, our survey reveals a clear convergence of expert perspective on what is needed to support safe and effective adoption. The advance of AI in laboratory medicine is an unstoppable trend, with tremendous potential for improving laboratory efficiency, enhancing testing accuracy, and ultimately delivering better patient care. Across respondents, three priorities consistently emerged: transparency from providers regarding model design and training data, structured local verification before deployment, and ongoing performance monitoring capable of detecting drift. These shared expectations reflect a growing recognition that AI systems must be evaluated with the same rigor applied to traditional laboratory methods, but with additional safeguards to address data dependency, model evolution, and contextual sensitivity. These findings should be viewed as expert-informed priorities intended to guide discussion and framework development, rather than universally applicable standards. Collectively, these findings suggest that future guidelines should prioritize defining the minimum information vendors must provide, outlining verification frameworks tailored to laboratory workflows, and establishing risk-based monitoring strategies that integrate both technical and clinical oversight. As commercial AI products become more prevalent, laboratories will increasingly rely on clear standards that delineate responsibilities between developers and users. Our work provides an initial evidence base to inform such guidance and highlights the specific areas where professional societies and regulatory bodies can offer the most meaningful support. We hope these insights will catalyze further consensus-building and contribute to the development of practical, harmonized recommendations for AI implementation in laboratory medicine.

Acknowledgments: We would like to thank the ADLM Data Analytics Steering Committee for reviewing and editing the survey questions, and thank Silvia Colli Lanzi for her assistance.
Research ethics: Not applicable. The responses were anonymized and collected only professional opinions regarding practices and perceptions; no personal identifiable private information was collected.
Informed consent: Not applicable.
Author contributions: HSY: conceptualization, methodology, data acquisition and analysis, writing original draft, review and editing, supervision. HCÇ, FDB, XD, MK, JNM: data analysis, writing original draft, review and editing. JF: data analysis. GF, DG, BW, FW: review and editing manuscript. All authors have accepted responsibility for the entire content of this manuscript and approved its submission.
Use of Large Language Models, AI and Machine Learning Tools: None declared.
Conflict of interest: Michael King is an employee of Roche Diagnostics GmbH, Germany. All other authors state no conflict of interest.
Research funding: None declared.
Data availability: Data can be shared upon request.

References

1. Del Ben F. Beyond test results: the strategic importance of metadata for the integration of AI in laboratory medicine. Clin Chem Lab Med 2025;63:653–5.
2. Padoan A, Cadamuro J, Frans G, Cabitza F, Tolios A, De Bruyne S, et al. Data flow in clinical laboratories: could metadata and peridata bridge the gap to new AI-based applications? Clin Chem Lab Med 2025;63: 684–91.
3. Blatter TU, Witte H, Nakas CT, Leichtle AB. Big data in laboratory medicine-FAIR quality for AI? Diagnostics 2022;12:1923.
4. Yang HS, Rhoads DD, Sepulveda J, Zang C, Chadburn A, Wang F. Building the model: challenges and considerations of developing and implementing machine learning tools for clinical laboratory medicine practice. Arch Pathol Lab Med 2022;147:826–36.
5. Cubukcu HC, Topcu DI, Yenice S. Machine learning-based clinical decision support using laboratory data. Clin Chem Lab Med 2024;62: 793–823.
6. Farrell CJ, Makuni C, Keenan A, Maeder E, Davies G, Giannoutsos J. A machine learning model for the routine detection of “wrong blood in complete blood count tube” errors. Clin Chem 2023;69:1031–7.
7. Wilkes EH, Rumsby G, Woodward GM. Using machine learning to aid the interpretation of urine steroid profiles. Clin Chem 2018;64:1586–95.
8. Yang HS, Pan W, Wang Y, Zaydman MA, Spies NC, Zhao Z, et al. Generalizability of a machine learning model for improving utilization of parathyroid hormone-related peptide testing across multiple clinical centers. Clin Chem 2023;69:1260–9.
9. Seok HS, Yu S, Shin KH, Lee W, Chun S, Kim S, et al. Machine learning-based sample misidentification error detection in clinical laboratory tests: a retrospective multicenter study. Clin Chem 2024;70:1256–67.
10. Yang HS, Hou Y, Vasovic LV, Steel PAD, Chadburn A, Racine-Brzostek SE, et al. Routine laboratory blood tests predict SARS-CoV-2 infection using machine learning. Clin Chem 2020;66:1396–404.
11. Than MP, Pickering JW, Sandoval Y, Shah ASV, Tsanas A, Apple FS, et al. Machine learning to predict the likelihood of acute myocardial infarction. Circulation 2019;140:899–909.
12. Cadamuro J, Carobene A, Cabitza F, Debeljak Z, De Bruyne S, van Doorn W, et al. A comprehensive survey of artificial intelligence adoption in European laboratory medicine: current utilization and prospects. Clin Chem Lab Med 2025;63:692–703.
13. The European Parliament and the Council of the European Union. Regulation (EU) 2024/1689 of the European Parliament and of the council of 13 June 2024 (Artificial Intelligence Act). Off J Eur Union; 2024. https://eur-lex.europa.eu/eli/reg/2024/1689/oj [Accessed on 19 Dec 2025].
14. Master SR, Badrick TC, Bietenbeck A, Haymond S. Machine learning in laboratory medicine: recommendations of the IFCC working group. Clin Chem 2023;69:690–8.
15. Spies NC, Farnsworth CW, Wheeler S, McCudden CR. Validating, implementing, and monitoring machine learning solutions in the clinical laboratory safely and effectively. Clin Chem 2024;70:1334–43.
16. Cubukcu HC, Boursier G, Linko S, Bernabeu-Andreu FA, Mesko Brguljan P, Tosheska-Trajkovska K, et al. Regulating the future of laboratory medicine: European regulatory landscape of AI-driven medical device software in laboratory medicine. Clin Chem Lab Med 2025;63:1891–914.
17. The European Parliament and the Council of The European Union. Regulation (EU) 2017/746 of the European Parliament and of the Council of 5 April 2017 on in vitro diagnostic medical devices. Off J Eur Union; 2017. https://eur-lex.europa.eu/eli/reg/2017/746/oj/eng [Accessed on 22 Sep 2025].
18. The U.S. Food and Drug Administration, Health Canada, and the United Kingdom’s Medicines and Healthcare Products Regulatory Agency. Good machine learning practice for medical device development: guiding principles; 2021. https://www.fda.gov/medical-devices/software-medical-device-samd/goodmachine-learning-practice-medical-device-development-guidingprinciples [Accessed on 19 Dec 2025].
19. Jackson BR, Sendak MP, Solomonides A, Balu S, Sittig DF. Regulation of artificial intelligence in healthcare: clinical laboratory improvement amendments (CLIA) as a model. J Am Med Inf Assoc 2025;32:404–7.
20. Herman DS, Reece JT, Weissman GE. Lessons for local oversight of AI in medicine from the regulation of clinical laboratory testing. npj Digit Med 2024;7:359.
21. Carobene A, Cadamuro J, Frans G, Goldshmidt H, Debeljak Z, De Bruyne S, et al. EFLM checklist for the assessment of AI/ML studies in laboratory medicine: enhancing general medical AI frameworks for laboratory-specific applications. Clin Chem Lab Med 2026;64:27–40.
22. Oesper L, Merico D, Isserlin R, Bader GD. WordCloud: a cytoscape plugin to create a visual semantic summary of networks. Source Code Biol Med 2011;6:7.
23. Emerson P, editor. Designing an all-inclusive democracy: consensual voting procedures for use in parliaments, councils and committees. Berlin, Heidelberg: Springer; 2007:16–17 pp.
24. Clinical and Laboratory Standards Institute (CLSI). A framework for using CLSI documents to evaluate medical laboratory test method. In: CLSI guideline EP19, 3rd ed. Wayne, PA: Clinical and Laboratory Standards Institute; 2022.
25. American Medical Informatics Association. Ensuring responsible AI integration in healthcare; 2021. https://brand.amia.org/m/664b5e658f3a4a87/original/4-c-Final-AMIA-Factsheet_Responsible-AI-Integration-April-20241.pdf [Accessed on 19 Dec 2025].
26. Norgeot B, Quer G, Beaulieu-Jones BK, Torkamani A, Dias R, Gianfrancesco M, et al. Minimum information about clinical artificial intelligence modeling: the MI-CLAIM checklist. Nat Med 2020;26:1320–4.
27. Sendak MP, Gao M, Brajer N, Balu S. Presenting machine learning model information to clinical end users with model facts labels. npj Digit Med 2020;3:41.
28. Mitchell M, Wu S, Zaldivar A, Barnes P, Vasserman L, Hutchinson B, et al. Model cards for model reporting. In: Proceedings of the conference on fairness, accountability, and transparency 2019. Atlanta, GA, USA: Association for Computing Machinery; 2019:220–9 pp.
29. Spies NC, Ng DP. Performance metrics for machine learning solutions in laboratory medicine. Lab Med 2025;56:597–607.
30. Van Calster B, Collins GS, Vickers AJ, Wynants L, Kerr KF, Barrenada L, et al. Evaluation of performance measures in predictive artificial intelligence models to support medical decisions: overview and guidance. Lancet Digit Health 2025;7:100916.
31. Haymond S, Master SR. How can we ensure reproducibility and clinical translation of machine learning applications in laboratory medicine? Clin Chem 2022;68:392–5.
32. Yang J, Soltan AAS, Clifton DA. Machine learning generalizability across healthcare settings: insights from multi-site COVID-19 screening. npj Digit Med 2022;5:69.
33. Cabitza F, Campagner A, Soares F, Garcia de Guadiana-Romualdo L, Challa F, Sulejmani A, et al. The importance of being external. Methodological insights for the external validation of machine learning models in medicine. Comput Methods Programs Biomed 2021;208: 106288.

34. Campagner A, Famiglini L, Carobene A, Cabitza F. Everything is varied: the surprising impact of instantial variation on ML reliability. Appl Soft Comput 2023;146:110644.
35. Martens RJH, van Doorn W, Leers MPG, Meex SJR, Helmich F. Unraveling uncertainty: the impact of biological and analytical variation on the prediction uncertainty of categorical prediction models. J Appl Lab Med 2025;10:339–51.
36. Yesil MR, Talli I, Pelloso M, Cosma C, Pangrazzi E, Plebani M, et al. Impact of analytical bias on machine learning models for sepsis prediction using laboratory data. Clin Chem Lab Med 2025;63:2022–30.
37. Cubukcu HC, Zengi O, Alp HH, Cihan M, Ucar KT, Thelen M, et al. A proof-of-concept practical approach for achieving equivalent results from non-harmonized measurement methods while awaiting harmonization: the CA 125 example. Clin Chim Acta 2025;575:120355.
38. International Organization for Standardization. ISO 17000:2020 Conformity assessment - vacabulary and general principles. Geneva, Swizerland. https://www.iso.org/standard/73029.html [Accessed on 16 Dec 2025].
39. The U.S. Food and Drug Administration. Department of health and human services. 21 CFR part 820: quality system regulation. https://www.ecfr.gov/current/title-21/part-820 [Accessed on 16 Dec 2025].
40. Joint Artificial Intelligence Board and Medical Device Coordination Group Document. AIB 2025-1 MDCG 2025-6 interplay between the medical devices regulation (MDR) & in vitro diagnostic medical devices regulation (IVDR) and the artificial intelligence act (AIA). https://health.ec.europa.eu/document/download/b78a17d7-e3cd-4943-851d-e02a2f22bbb4_en?filename=mdcg_2025-6_en.pdf [Accessed on 18 Dec 2025].
41. Hoens TR, Polikar R, Chawla NV. Learning from streaming data with concept drift and imbalance: an overview. Prog Artif Intell 2012;1:89–101.
42. Badrick T, Bietenbeck A, Cervinski MA, Katayev A, van Rossum HH, Loh TP, et al. Patient-based real-time quality control: review and recommendations. Clin Chem 2019;65:962–71.
43. Duan X, Zhang M, Liu Y, Zheng W, Lim CY, Kim S, et al. Next-generation patient-based real-time quality control models. Ann Lab Med 2024;44:385–91.
44. Feng J, Phillips RV, Malenica I, Bishara A, Hubbard AE, Celi LA, et al. Clinical artificial intelligence quality improvement: towards continual monitoring and updating of AI algorithms in healthcare. npj Digit Med 2022;5:66.
45. Busch F, Kather JN, Johner C, Moser M, Truhn D, Adams LC, et al. Navigating the European union artificial intelligence act for healthcare. npj Digit Med 2024;7:210.
46. Lubbers BR, Schilhabel A, Cobbaert CM, Gonzalez D, Dombrink I, Bruggemann M, et al. The new EU regulation on in vitro diagnostic medical devices: implications and preparatory actions for diagnostic laboratories. HemaSphere 2021;5:e568.
47. International Organization for Standardization. ISO 15189:2022 Medical laboratory - Requirement for quality and competence. Geneva, Switzerland. https://www.iso.org/standard/76677.html [Accessed on 18 Dec 2025].
48. The U.S. Food and Drug Administration. Health Canada, and the United Kingdom's Medicines and Healthcare Product Regulatory Agency. Good Machine Learning Practice for Medical Device Development: Guiding Principles; 2021. https://www.fda.gov/medical-devices/software-medical-device-samd/goodmachine-learning-practice-medical-device-development-guidingprinciples [Accessed on 12 Dec 2025].
49. International Organization for Standardization. ISO/CD 24051-1.2. Medical laboratories Part 1: general principles for the application of artificial intelligence in medical laboratories. https://www.iso.org/standard/87673.html [Accessed on 18 Dec 2025].
50. International Organization for Standardization. ISO/CD 24051-2. Medical laboratories Part 2: digital pathology and artificial intelligence (AI) - based image analysis. https://www.iso.org/standard/87674.html [Accessed on 18 Dec 2025].

Supplementary Material: This article contains supplementary material (https://doi.org/10.1515/cclm-2025-1322).