Letter to the Editor

Implementing Machine Learning in the Clinical Laboratory: Opportunities and Challenges

He Sarina Yang[1*]

[1]*Department of Pathology and Laboratory Medicine, Weill Cornell Medicine, New York, USA*

Article Info

[*]***Corresponding Author:***
He Sarina Yang, PhD, MBBS, DABCC
Associate Professor
Department of Pathology and Laboratory Medicine
Weill Cornell Medicine
525 E. 68th St, New York, NY, 10065
E-mail: hey9012@med.cornell.edu

Keywords

Machine learning, clinical laboratory, generalizability, reproducibility, fairness

In recent years, the Laboratory Medicine field's strong interest in the potential application of artificial intelligence (AI) and machine learning (ML) has been reflected in the substantial growth of publications. A number of reported studies have explored the applications of AI/ML across the four phases of the total testing process, such as improving test utilization [1], detecting pre-analytical errors [2, 3], enhancing analytical efficiency [4], and interpreting complex laboratory panels [5-7]. In addition, machine learning-based clinical decision supporting tools have been used to predict the onset, progression, subtypes, and outcome of diseases, such as sepsis [8], acute kidney injury [9], and COVID-19 [10]. Compared to the traditional clinical algorithm or scoring systems, ML can ingest far larger, richer dataset, automatically flag subtle variables that manual scores miss, and generate more accurate predictions of disease diagnosis and prognosis.

Data generated from the clinical laboratory is largely composed of discrete numerical or categorical values in a structured format, and the number of tests ordered for each patient makes these datasets inherently high-dimensional [11]. Clinical laboratory data has several noteworthy characteristics: first, test values can vary across instrument platforms and analytical methodologies, and their corresponding reference ranges can differ as well. Therefore, integrating data from different sources must begin with careful normalization. Second, clinical datasets may contain missing values and outliers. Missing values stem from systematic missingness and random missingness. Systematic missingness, i.e., laboratory tests are not always ordered on a regular basis for all patients in clinical practice, especially with different institutional practice, reflecting institution-specific ordering patterns. On the other hand, random missingness could occur when a test is ordered but cannot be completed due to pre-analytical or analytical errors. When processing a clinical dataset, we should determine if it is reliable to impute the missing values from the remaining results using imputation techniques. In addition, outliers in laboratory data can occur due to analytical or pre-analytical errors, or true pathological reasons. The causes of outliers should be investigated to determine whether they should be included or excluded from the data analysis. Informative outliers can be accommodated with techniques like robust loss functions, whereas erroneous ones should be removed. Last but not least, unreported data, such as hemolysis, icterus, and lipemia index, are kept in the middleware, but not transmitted to EHR, yet these data may be valuable for ML analysis.

A successful ML pipeline in Laboratory Medicine begins with robust data collection: assemble a large, high-quality dataset that represents the target population; Next comes data preprocessing and method development, which includes normalization, handling of missing values and outliers, feature engineering, optimization of hyperparameters, and selection of appropriate model architecture. Once a candidate model is locked, rigorous evaluation of its performance on an internal hold-out test set and independent external datasets are needed to detect overfitting [12]. If a model is successfully developed and demonstrates strong promise during retrospective analysis, several checkpoints are still needed before it can be integrated into clinical workflow:

1. Assess a model's generalizability and transportability – the model should perform reliably on independent, unseen data collected from different geographic or demographic patient populations or different hospital settings. In the laboratory setting, factors, such as sample handling protocol, instrument platforms, methodologies, send-out laboratories, and regional patient characteristics, can shift data distributions and impact a model's performance. External validation on datasets that capture these variations is therefore indispensable as it tests a model's real-world utility, uncovers hidden biases, and confirms a model's robustness [13].
2. Demonstrate a model's interpretability. Physicians and laboratory professionals are more likely to accept a model that is comprehensible and aligned with their clinical knowledge. Most conventional ML models, such as logistic regression and decision tree, are self-interpretable due to their linear or rule-based nature. In contrast, more complicated ML models, such as deep learning networks, are largely "black boxes". While some research has shown that accuracy sometimes trumps interpretability, the assumption is that such accuracy should be widely tested and proven to be generalizable, which is challenging in many clinical scenarios. For "black box models", post-hoc methods can translate its input-output behavior into a simpler surrogate (e.g., a decision tree) or quantify each feature's influence. The Shapley additive explanations (SHAP) technique, for instance, decomposes the model's prediction for each sample as an additive integration of the contributions from each individual feature, and thus, can show how features act as force to push the model to make positive or negative predictions.
3. Evaluate a model's fairness. Since clinical decisions guided by ML may influence diagnostics, treatments, and resource allocation, we should make sure that a model does not systematically over- or under- predict risk for certain races, sexes, or socioeconomic groups. A model's fairness can be evaluated using metrics including disparate impact (the probabilities of the positive outcome for the two groups), equal opportunity difference (the difference in true positive rates between two groups), the predictivity parity difference (the difference in positive predictive values), with the White race being the privileged group. A model's performance should be assessed retrospectively and prospectively across demographic groups to detect any disparities that could introduce or perpetuate bias.
4. Involve clinical teams in the development and deployment pipeline. Evaluating an ML model is not just a data-science exercise, it requires frontline clinical insight [14]. Clinical teams should be involved and provide insights on how, where, and by whom the tool would be used – alert fatigue, hand-off timing, and documentation burden often make or break adoption. The clinicians can map model outputs to actionable decisions, e.g. ordering confirmatory tests or adjusting medication, and estimate the downstream impact on patient outcomes.

Despite the buzz around AI, clinically deployed AI/ML tools in the clinical laboratory remain limited. Compared to Radiology which has over 700 AI-based technologies approved by FDA, Laboratory Medicine only has a combined total of 37 AI applications across clinical chemistry, immunology, hematology and microbiology [15].

Translating a proof-of-concept model into measurable clinical benefit requires rigorous validation and a well-designed implementation strategy, as well as interdisciplinary collaboration. Several huddles slow the pipeline from model development to deployment [16]:

1. lack of data infrastructure that supports ML validation implementation;
2. lack of informatics personnel to conduct data extraction, model deployment, and ongoing maintenance;
3. limited knowledge and experience among laboratorians, leaving many unfamiliar with how to validate, monitor, and trouble-shoot ML outputs,
4. lack of guidelines or expert consensus on best practice for ML implementation in the clinical laboratory.

Our field urgently needs more success stories of integrating ML into clinical workflow, along with practical guidelines on how to validate and implement various types of AI/ML applications. Machine learning won't replace seasoned laboratorians, but laboratories that harness ML thoughtfully will outperform those who don't – catching errors earlier, improving laboratory efficiency, freeing humans for higher-order tasks, and delivering more precise care. The opportunity is huge; the challenge is to move beyond proof-of-concepts into regulated, reproducible, and equitable practice. With the right guidelines and a collaborative mindset, the clinical laboratory can become a flagship environment where AI proves its value.

References

1. H.S. Yang, W. Pan, Y. Wang, M.A. Zaydman, N.C. Spies, Z. Zhao, T.A. Guise, Q.H. Meng, F. Wang, Generalizability of a Machine Learning Model for Improving Utilization of Parathyroid Hormone-Related Peptide Testing across Multiple Clinical Centers, Clin Chem 2023;69(11): 260-1269.
2. C.J. Farrell, C. Makuni, A. Keenan, E. Maeder, G. Davies, J. Giannoutsos, A Machine Learning Model for the Routine Detection of "Wrong Blood in Complete Blood Count Tube" Errors, Clin Chem 2023;69(9):1031-1037.
3. B.V. Graham, S.R. Master, A.E. Obstfeld, R.B. Wilson, A Multianalyte Machine Learning Model to Detect Wrong Blood in Complete Blood Count Tube Errors in a Pediatric Setting, Clin Chem 2025;71(3): 418-427.
4. H.S. Seok, S. Yu, K.H. Shin, W. Lee, S. Chun, S. Kim, H. Shin, Machine Learning-Based Sample Misidentification Error Detection in Clinical Laboratory Tests: A Retrospective Multicenter Study, Clin Chem 2024;70(10):1256-1267.
5. E.H. Wilkes, G. Rumsby, G.M. Woodward, Using Machine Learning to Aid the Interpretation of Urine Steroid Profiles, Clin Chem 2018;64(11):586-1595.
6. E.H. Wilkes, E. Emmett, L. Beltran, G.M. Woodward, R.S. Carling, A Machine Learning Approach for the Automated Interpretation of Plasma Amino Acid Profiles, Clin Chem 2020;66(9):1210-1218.
7. F. Chabrun, X. Dieu, M. Ferre, O. Gaillard, A. Mery, J.M. Chao de la Barca, A. Taisne, G. Urbanski, P. Reynier, D. Mirebeau-Prunier, Achieving Expert-Level Interpretation of Serum Protein Electrophoresis through Deep Learning Driven by Human Reasoning, Clin Chem 2021;67(10):1406-1414.
8. H.S. Yang, Machine Learning for Sepsis Prediction: Prospects and Challenges, Clin Chem 2024;70(3):465-467.
9. A.U. Rehman, J.A. Neyra, J. Chen, L. Ghazi, Machine learning models for acute kidney injury prediction and management: a scoping review of externally validated studies, Crit Rev Clin Lab Sci 2025:1-23.
10. H.S. Yang, Y. Hou, L.V. Vasovic, P.A.D. Steel, A. Chadburn, S.E. Racine-Brzostek, P. Velu, M.M. Cushing, M. Loda, R. Kaushal, Z. Zhao, F. Wang, Routine Laboratory Blood Tests Predict SARS-CoV-2 Infection Using Machine Learning, Clin Chem 2020;66(11):1396-1404.
N.C. Spies, C.W. Farnsworth, S. Wheeler, C.R. McCudden, Validating, Implementing, and Monitoring Machine Learning Solutions in the Clinical Laboratory Safely and Effectively, Clin Chem 2024;70(11): 1334-1343.
12. H.S. Yang, D.D. Rhoads, J. Sepulveda, C. Zang, A. Chadburn, F. Wang, Building the Model, Arch Pathol Lab Med (2022).
13. F. Cabitza, A. Campagner, F. Soares, L. Garcia de Guadiana-Romualdo, F. Challa, A. Sulejmani, M. Seghezzi, A. Carobene, The importance of being external. methodological insights for the external validation of machine learning models in medicine, Comput Methods Programs Biomed 2021;208:106288.
14. F. Wang, A. Beecy, Implementing AI models in clinical workflows: a roadmap, BMJ Evid Based Med (2024).
15. F. Del Ben, Beyond test results: the strategic importance of metadata for the integration of AI in laboratory medicine, Clin Chem Lab Med 2025;63(4):653-655.
[16] J. Cadamuro, A. Carobene, F. Cabitza, Z. Debeljak, S. De Bruyne, W. van Doorn, E. Johannes, G. Frans, H. Ozdemir, S. Martin Perez, D. Rajdl, A. Tolios, A. Padoan, C. European Federation of Clinical, I. Laboratory Medicine Working Group on Artificial, A comprehensive survey of artificial intelligence adoption in European laboratory medicine: current utilization and prospects, Clin Chem Lab Med 63(4) (2025) 692-703.