Opinion Paper

Andrea Padoan* and Mario Plebani

Flowing through laboratory clinical data: the role of artificial intelligence and big data

https://doi.org/10.1515/cclm-2022-0653
Received July 7, 2022; accepted July 8, 2022;
published online July 18, 2022

Abstract: During the last few years, clinical laboratories have faced a sea change, from facilities producing a high volume of low-cost test results, toward a more integrated and patient-centered service. Parallel to this paradigm change, the digitalization of healthcare data has made an enormous quantity of patients' data easily accessible, thus opening new scenarios for the utilization of artificial intelligence (AI) tools. Every day, clinical laboratories produce a huge amount of information, of which patients' results are only a part. The laboratory information system (LIS) may include other "relevant" compounding data, such as internal quality control or external quality assessment (EQA) results, as well as, for example, timing of test requests and of blood collection and exams transmission, these data having peculiar characteristics typical of big data, as volume, velocity, variety, and veracity, potentially being used to generate value in patients' care. Despite the increasing interest expressed in AI and big data in laboratory medicine, these topics are approaching the discipline slowly for several reasons, attributable to lack of knowledge and skills but also to poor or absent standardization, harmonization and problematic regulatory and ethical issues. Finally, it is important to bear in mind that the mathematical postulation of algorithms is not sufficient for obtaining useful clinical tools, especially when biological parameters are not evaluated in the appropriate context. It is therefore necessary to enhance cooperation between laboratory and AI experts, and to coordinate and govern processes, thus favoring the development of valuable clinical tools.

Keywords: artificial intelligence; big data; clinical laboratories; flow of information; laboratory medicine.

Clinical laboratories are facing a new transformation

In the last few years, medical laboratories have undergone a substantial change, moving toward the consolidation of clinical testing in megastructures in order to enhance volumes and reduce costs [1]. This paradigm of a business model based on laboratory tests outsourcing has fallen short of expectations; indeed, it has not been proven to significantly reduce overall costs [2]. It has been demonstrated that, up to a threshold of one million test per year, high costs correlate with volumes. However, over this threshold, the association between volumes and costs is not linear, as laboratory organization, rather than test volume, has a more significant effect on final costs [2]. Importantly, it was postulated that the correlation between costs and test volumes was mainly sustained by the "traditional laboratory test", widely requested in the last few decades, and well known by physicians, who were capable of correctly interpreting their values [3]. Fortunately, with the technological and organizational advancements, which progressed parallel with improved knowledge of the pathophysiology of human diseases, different reasons have led to a change to the vision of clinical laboratory as commodities. Yet, laboratories have undergone a progressive transition from the so-called "silos" model to a more integrated and patient-centered vision [3]. Different drivers underpin this change, positive drivers being diagnostic stewardship, advances in molecular/genotypic testing, and innovative technology, which pave the way for more personalized laboratory results [1, 3]. On the other hand, the miniaturization of laboratory-based point of care (POC) testing devices and their implementation in telemedicine, pinpoints the pivotal role of the laboratory in "near patient" testing, and points to a new future role for

***Corresponding author: Prof. Andrea Padoan**, Department of Laboratory Medicine, University-Hospital of Padova, Via Giustiniani 2, 35128, Padova, Italy; and Department of Medicine-DIMED, University of Padova, Padova, Italy, E-mail: andrea.padoan@unipd.it. https://orcid.org/0000-0003-1284-7885
Mario Plebani, Department of Laboratory Medicine, University-Hospital of Padova, Padova, Italy; and Department of Medicine, University of Padova, Padova, Italy. https://orcid.org/0000-0002-0270-1711

clinical laboratories. Another change revolutionizing clinical laboratories depends on the digitalization of processes, which are changing all healthcare sectors, including medical laboratories. This force is fueled by several drivers, such as the easy management of enormous volumes of data, the support of new technological systems to data integration and, last but not least, the wide, efficient utilization artificial intelligence (AI). Since AI is not new to computer science, in recent years its application has increased, mainly thanks to the computational power needed to harvest the potential of these methods [4]. However, within hospitals and healthcare services in general, every day clinical laboratories produce a huge amount of mainly high-quality information (e.g. patients' results), that is a "gold mine" for teaching computer-based algorithms, such as AI tools.

Clinical laboratories' roles in the explosion of big healthcare data

Traditionally, clinical studies have always included medical laboratory results, in addition to demographical information. Despite being conducted following rigorous criteria, these studies were often limited by, for example, difficulty in accruing patients or the scarce resources among the medical staff. In the last years, with the widespread diffusion of laboratory information systems (LIS) allowing the rapid retrieval of patients' results, enormous quantities of retrospective data were made readily available without incurring additional costs to Researchers. Thus, some types of clinical studies are changing their face, and several models are rapidly evolving using data obtained from LIS, especially when the research objective is the association between results and positive or negative biochemical phenotypes [5]. Furthermore, the digital transformation of healthcare, currently allows integrated information of data to be obtained from different disciplines and different patients. The creation of integrated data warehouses in healthcare facilities, recognized sources of "big data", have paved the way for the use of AI tools for data analyses and other computational tools, such as natural language processing (NLP), which could, in-turn, generate further new resources for the care of patients.

Uncovering the flow of information in modern clinical laboratories

The delivery of laboratory testing postulated 50 years ago by G. D. Lundberg [6], who coined the term "brain to brain turnaround time loop" was revisited in 2011 by Plebani et al. [7]. The years since Lundberg first expressed this concept have seen a constant increase in the type and quantity of information generated for each patient during laboratory testing. In addition to clinical results and demographical parameters, LIS may include other "relevant" compounding data, such as internal quality control (instrumental) or external quality assessment (EQA) results, as well as timing of test requests and of blood collection or exam results transmission. However, a more detailed search reveals a greater volume of information generated daily by the clinical laboratory, even if most of the information is only partially recorded in LIS. The information might include pre-analytical features (e.g. retesting intervals, additional dietary information, lifestyle, sample storage conditions or handling, presence of hemolysis), analytical features (e.g. technical and medical validations, automatic check for sample quality or sample mismatch, instrumental self-diagnosis or complex check on data elaboration) and post-analytical features (e.g. suggestions from clinical decision support systems based on patients' test results, physicians visualization of results, time for communicating critical test results (Figure 1).

These data type has peculiar characteristics, such as volume, velocity, variety, and veracity but interestingly can finally generate value for patients' care, for laboratory personnel and manufacturers' technicians. Since all these features are attributed to big data, it is reasonable to state that every laboratory, produces big data daily [8]. Furthermore, laboratories regularly maintain the lab test catalog, which is available for laboratory staff, hospital physicians and personnel (Figure 1).

Currently, several tools (some based on simple "if-then" algorithms, others sophisticated and based on AI) have been successfully realized and, in some cases, integrated in the pre-analytical, analytical or post-analytical phases of the brain-to-brain loop. For example, different informatic demand management tools that have been produced and applied, ensure the appropriate test request in specific context [9, 10]. In COVID-19 rapid testing, AI was found to be useful for the interpretation of test results [11]. Another example of AI application in laboratory medicine is the detection of sample mix-up with delta check by ML tools, which outperform the use of a single hematological parameter, such as the MCV [12]. In a different study, AI based rules for auto-validation of laboratory results have shown to reach an high agreement with laboratory professionals [13], while other ML approaches have been used for predicting thromboembolism in cancer patients, thus facilitating the collaborative efforts between clinicians and

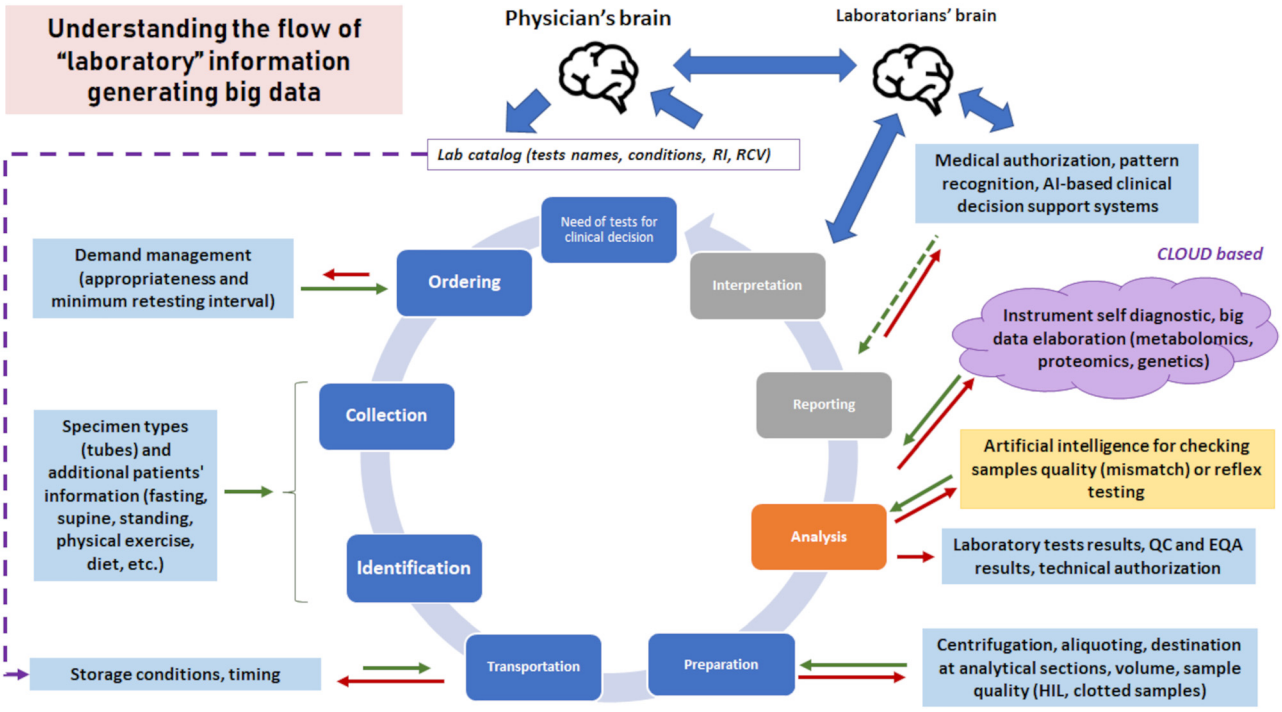

Figure 1: A modified brain to brain loop diagram, showing the flow of information within a clinical laboratory.
Arrows represent the flow of information toward (green) the laboratory information system (LIS) and, oppositely (red), from LIS to other external informatic tools, which might include artificial intelligence (AI) algorithms.

laboratories in the identification of patients to be monitored for risk of disease [14].

Advances in the capability of laboratories in detecting molecular genotypic abnormalities, detecting metabolomic alterations and conducting proteomic tests (e.g. MALDI-TOF/MS analyses for detecting amyloidosis) have opened the possibility of achieving greater personalization of laboratory results, but call for the implementation of new workflows. First, the enormous amount of information obtained calls for adequate infrastructures for storing data for as long as required in national regulations. Second, since results are not always immediately intelligible, they call for software that may include AI tools and that is often cloud based, thus incurring further issues regarding data ownership and cloud-storage time. Third, other web-based resources (e.g. tools for checking and describing rare point mutations in genetics or tools for detecting proteins from fragmentation patterns in proteomics) are often used by specialists in laboratory medicine using these technologies.

AI meet big-laboratory data

AI have been suggested as a proponent of a compelling pitch, moving from a traditional approach focused on comparing effectiveness in the average person, to strategies tailored to the individual. This revolution has been fueled by the availability of accessible longitudinal data, combined with the application of flexible ML approaches [15]. Therefore, the main driver of recently increased expectations in the AI field is the ready availability of big volumes of healthcare data [16].

In the laboratory, advancements in LIS have allowed large numbers of results to be collected in a limited timespan, using relatively few resources [17]. However, in most cases, for ML applications it is necessary to integrate laboratory data with additional clinical data from patients' (e.g. diagnosis, disease recurrence, and comorbidities). Laboratory and clinical data should be integrated both horizontally and vertically, the first dimension being involved in the longitudinal view of data, obtained from different testing and examinations conducted on patients over time. Vertical integration includes medical records from the laboratory, clinical records from different medical specialties (e.g. cardiology, and radiology) and from patients themselves (e.g. using wearables devices or connected diagnostic lab-on-skin tests) [18, 19]. The potential utility of integrating patients' records', which are currently fragmented, being held in different parts of the healthcare data warehouse, is compromised due to the existence of

different data structures. Structured data sources (e.g. laboratory tests, patients' admission and discharge dates) can often be easily combined by appropriate software, while unstructured sources of data (e.g. clinical notes and observations collected during hospitalization) [20] tend to be heterogeneous, comprising multiple types of data without any inherent structure. Regarding unstructured data, tools such as NLP techniques, which are often cloud-based, can be helpful in extracting information from free-form text of healthcare records, but these systems have not yet been evaluated in a setting that simulates clinical workflow [21]. Finally, commercially established classes of wearable medical devices are achieving technological levels enabling them to be used as lab-on-a-chip [22], not only for the continuous follow-up of hospitalized patients [e.g. portable point-of-care (POC) systems], but also for outpatients requiring close monitoring and a prompt clinical decision in the presence of vital parameter modifications [19]. In particular settings, already available wearable systems (e.g. consumer wearable device) that continuously monitor heart rate, body temperature, electrodermal activity and movements, can be used to evaluate vital signs. As demonstrated, these data are used for teaching AI algorithms in the prediction of variations in clinical laboratory test results, thus enabling the detection of deterioration in a patient's vital status [23].

Challenges and pitfalls of AI implementation in laboratory medicine

Two recent studies showed how methods and models combining data analytics and AI have been progressively applied to laboratory medicine and other medical specialties [5, 24]. The results obtained revealed a growing increase in interest in AI, documented by an increased number of publications between 2017 and 2021, but only 8/44 (18.2%) of the articles considered were from laboratory medicine groups, the others being produced by researchers of other disciplines. Indeed, several limitations have been described by many authors [18, 25–27]. From a "local" point-of-view, the first pitfall might regard specific knowledge of AI in the laboratory community, which is pivotal for motivating new researches and for addressing implementation challenges [28]. Indeed, in a recent web-based survey among stakeholders in laboratory medicine in the United States, it was underlined that most of the participants believed AI would be valuable to them in the near future, although vital prerequisites still need to be forthcoming, and specific knowledge of AI in the medical community at large is poor [28]. This point could be ascribed to a wider concept embracing the digital revolution of the laboratory, a theme that also includes new specialties such as clinical bioinformatics, communication enhancement and interactive skills, and informatics skills [3, 29, 30]. Training aiming to support this cohort of laboratory medicine specialists by providing digital skills will call for the acknowledgment, and bridging, of the gap in a partnership between teachers and students, an important role in this approach also being played by Scientific Societies [29]. The second limitation could be the provision of easy access to patients' clinical records (e.g. diagnosis, comorbidities, clinical parameters and therapeutic drugs) [16]. To be of clinical utility, algorithms for AI and ML should benefit from the multiple reusability of data and clinical results for rapid learning [20]. This differs from what usually happens in practice when, in a clinical study, laboratory results are matched once with patients' clinical data (e.g. disease), after which ML is applied. This paradigm must be overcome, in order to facilitate the application of ML, and also because ML algorithms might benefit from a continuous refinement using further data, evolving over time [20].

From a "global view-point", relevant limitations to AI implementation are data quality and results standardization, legal and privacy issues and IT security. Currently, AI tools call for huge datasets and often results cannot be obtained by a single laboratory. It might be difficult to exchange electronic patient records or other IT infrastructures (e.g. laboratory test results): even when data have been mapped according to common data structures, records it is not necessarily possible to merged and analyze data together [18]. On considering different labs, only a limited set of analytes is found to be sufficiently standardized to ensure compatibility [18, 31]. To this end, universal coding of laboratory test names, achievable by the logical observation of identifiers' names and codes (LOINC), is of utmost importance [25]. Besides the laboratory analysis, the result unit must be machine-readable for interconnection and the exchange of laboratory results [25]. Furthermore, despite the efforts made to standardize and harmonize, the measuring devices used considerably influence some results, and the use of device-specific target values derived by EQA schemes can only be efficient if shared in a predisposed database, such as the EUDAMED [25]. Interpretative comments, which are a crucial part of laboratory reports, should be structured and grouped into different categories for their efficient summarization and sharing [25]. Ethical and normative issues are of prime importance. For example, patients' consent for the use of

their personal health data for receiving certain treatments are not valid in case if used for AI application [32]. Furthermore, norms must be consistent with existing regulations (e.g. national regulations), with the existing regulations on data protection, including the General Data Protection Regulation (GDPR) act 2018, with the European Union charter of fundamental rights and other relevant clauses [32].

Finally, is should be emphasized that even if data scientists are can certainly design and produce excellent AI algorithms, the active contribution of laboratory professionals is the key to providing accurate data analysis and interpretation across the entire process. Indeed, the theoretical mathematical postulation of algorithms is not sufficient for obtaining clinical useful algorithms, especially when the biological parameters are not evaluated in the appropriate context, with the consideration of relevant laboratory medicine concepts (e.g. biological variability, analytical goals and analytical variability).

Conclusions

Laboratory professionals play a major role in all the medical specialties, not only in assisting clinicians, but also in selecting the right test for the right patient at the right time. This concept, identified in "clinical laboratory stewardship", encompasses all phases of the total testing process and is one of the most important drivers of the forthcoming changes in laboratory medicine: the shift toward a patient centered model [3]. In parallel, the technological advancements and the digital revolution are providing increasing support to precision medicine, which accurately associates patients with their specific profiles, based on personal exams and laboratory tests, to their clinical endpoints. This change in the paradigm of laboratory medicine, from "standalone factories", which are distant from diagnostic and therapeutic pathways, to key assets for making an early diagnosis, establishing a prognosis and providing personalized treatment could be well supported by AI and their tools, such as ML [1], technologies fueled by the vast availability of big data in healthcare, including laboratory data, but now requiring new professionals in data science who are able to transform raw information into improvement in patients' care. Although competences and skills of laboratory medicine experts probably will never fully cover the sophisticated mathematical theoretical application of ML, there is an urgent need to enhance cooperation between laboratory and AI experts, to coordinate and govern the processes, and to favor the implementation of appropriate technologies. Otherwise, there is a risk of obtaining a barren scenario, riddled with sophisticated technologies that are of little use to either the laboratory or the patient.

Research funding: This research received no external funding.
Author contributions: All authors have accepted responsibility for the entire content of this manuscript and approved its submission.
Competing interests: Authors state no conflict of interest.
Informed consent: Not applicable.
Ethical approval: Not applicable.

References

1. Plebani M. Quality and future of clinical laboratories: the Vico's whole cyclical theory of the recurring cycles. Clin Chem Lab Med 2018;56:901–8.
2. Plebani M. Clinical laboratory: bigger is not always better. Diagnosis 2018;5:41–6.
3. Plebani M, Laposata M, Lippi G. A manifesto for the future of laboratory medicine professionals. Clin Chim Acta 2019;489: 49–52.
4. Cadamuro J. Rise of the machines: the inevitable evolution of medicine and medical laboratories intertwining with artificial intelligence—a narrative review. Diagnostics 2021;11:1399.
5. Ronzio L, Cabitza F, Barbaro A, Banfi G. Has the flood entered the basement? A systematic literature review about machine learning in laboratory medicine. Diagnostics 2021;11:372.
6. Lundberg GD. Acting on significant laboratory results. JAMA J Am Med Assoc 1981;245:1762.
7. Plebani M, Laposata M, Lundberg GD. The brain-to-brain loop concept for laboratory testing 40 Years after its introduction. Am J Clin Pathol 2011;136:829–33.
8. Guerranti R, Padoan A, Angeletti D, Marco F, Tommaso T. Introduction to big data and artificial intelligence in laboratory medicine. Biochim Clin 2021;45:57.
9. Ferraro S, Panteghini M. The role of laboratory in ensuring appropriate test requests. Clin Biochem 2017;50:555–61.
10. Pelloso M, Basso D, Padoan A, Fogar P, Plebani M. Computer-based-limited and personalised education management maximise appropriateness of vitamin D, vitamin B12 and folate retesting. J Clin Pathol 2016;69:777–83.
11. Mendels DA, Dortet L, Emeraud C, Oueslati S, Girlich D, Ronat JB, et al. Using artificial intelligence to improve COVID-19 rapid diagnostic test result interpretation. Proc Natl Acad Sci USA 2021;118:3–5.
12. Zhou R, Liang Y, Cheng HL, Wang W, Huang D, Wang Z, et al. A highly accurate delta check method using deep learning for detection of sample mix-up in the clinical laboratory. Clin Chem Lab Med 2021. https://doi.org/10.1515/cclm-2021-1171 [Epub ahead of print].
13. Demirci F, Akan P, Kume T, Sisman AR, Erbayraktar Z, Sevinc S. Artificial neural network approach in laboratory test reporting. Am J Clin Pathol 2016;146:227–37.
14. Ferroni P, Zanzotto FM, Scarpato N, Riondino S, Guadagni F, Roselli M. Validation of a machine learning approach for venous thromboembolism risk prediction in oncology. Dis Markers 2017; 2017:8781379.

15. Wilkinson J, Arnold KF, Murray EJ, van Smeden M, Carr K, Sippy R, et al. Time to reality check the promises of machine learning-powered precision medicine. Lancet Digit Heal 2020;2:e677–80.
16. Gruson D. Big data, artificial intelligence and laboratory medicine: time for integration. Adv Lab Med 2021;2:1–3.
17. Kammergruber R, Durner J. Laboratory information system and necessary improvements in function and programming. J Lab Med 2018;42:277–87.
18. Ganslandt T, Neumaier M. Digital networks for laboratory data: potentials, barriers and current initiatives. Clin Chem Lab Med 2019;57:336–42.
19. Liu Y, Pharr M, Salvatore GA. Lab-on-skin: a review of flexible and stretchable electronics for wearable health monitoring. ACS Nano 2017;11:9614–35.
20. Damiani A, Masciocchi C, Lenkowicz J, Capocchiano ND, Boldrini L, Tagliaferri L, et al. Building an artificial intelligence laboratory based on real world data: the experience of gemelli generator. Front Comput Sci 2021;3:1–20.
21. Chi EA, Chi G, Tsui CT, Jiang Y, Jarr K, Kulkarni CV, et al. Development and validation of an artificial intelligence system to optimize clinician review of patient records. JAMA Netw Open 2021;4:1–12.
22. Komatireddy R, Topol EJ. Medicine unplugged: the future of laboratory medicine. Clin Chem 2012;58:1644–7.
23. Dunn J, Kidzinski L, Runge R, Witt D, Hicks JL, Schüssler-Fiorenza Rose SM, et al. Wearable sensors enable personalized predictions of clinical laboratory measurements. Nat Med 2021; 27:1105–12.
24. Cabitza F, Banfi G. Machine learning in laboratory medicine: waiting for the flood? Clin Chem Lab Med 2018;56:516–24.
25. Bietenbeck A, Streichert T. Preparing laboratories for interconnected health care. Diagnostics 2021;11:1487.
26. Martin MK. No lab is an island: universal coding of laboratory test names. J Vet Diagn Invest 2021;33:415–8.
27. Lippi G. Machine learning in laboratory diagnostics: valuable resources or a big hoax? Diagnosis 2019;2:133–5.
28. Paranjape K, Schinkel M, Hammer RD, Schouten B, Nannan Panday RS, Elbers PWG, et al. The value of artificial intelligence in laboratory medicine. Am J Clin Pathol 2021;155: 823–31.
29. Wieringa G. Teaching the pony new tricks: competences for specialists in laboratory medicine to meet the challenges of disruptive innovation. Clin Chem Lab Med 2019;57:398–402.
30. Plebani M, Laposata M, Lippi G. Driving the route of laboratory medicine: a manifesto for the future. Intern Emerg Med 2019;14: 337–40.
31. Ceriotti F. Is there a classical role for the clinical laboratory in digital health? Clin Chem Lab Med 2019;57:353–8.
32. Pennestrì F, Banfi G. Artificial intelligence in laboratory medicine: fundamental ethical issues and normative key-points. Clin Chem Lab Med 2022. https://doi.org/10.1515/cclm-2022-0096 [Epub ahead of print].