REVIEW

Rise of the Machines: Artificial Intelligence and the Clinical Laboratory

Shannon Haymond[a,b] and Christopher McCudden[c,*]

Background: Artificial intelligence (AI) is rapidly being developed and implemented to augment and automate decision-making across healthcare systems. Being an essential part of these systems, laboratories will see significant growth in AI applications for the foreseeable future.

Content: In laboratory medicine, AI can be used for operational decision-making and automating or augmenting human-based workflows. Specific applications include instrument automation, error detection, forecasting, result interpretation, test utilization, genomics, and image analysis. If not doing so today, clinical laboratories will be using AI routinely in the future, therefore, laboratory experts should understand their potential role in this new area and the opportunities for AI technologies. The roles of laboratorians range from passive provision of data to fuel algorithms to developing entirely new algorithms, with subject matter expertise as a perfect fit in the middle. The technical development of algorithms is only a part of the overall picture, where the type, availability, and quality of data are at least as important. Implementation of AI algorithms also offers technical and usability challenges that need to be understood to be successful. Finally, as AI algorithms continue to become available, it is important to understand how to evaluate their validity and utility in the real world.

Summary: This review provides an overview of what AI is, examples of how it is currently being used in laboratory medicine, different ways for laboratorians to get involved in algorithm development, and key considerations for AI algorithm implementation and critical evaluation.

INTRODUCTION

Artificial intelligence (AI) is a decades-old concept that is becoming a new practical reality in medicine. AI algorithms are being developed at a rapid pace and now becoming available for diagnosis, treatment, and prognosis. AI is broadly defined as computers imitating human thinking but can be further subdivided as shown in Fig. 1 into Artificial General Intelligence (AGI), Artificial Narrow Intelligence (ANI), and Artificial Superintelligence (ASI). AGI, also called "strong AI," refers to full mimicry of human reasoning, learning, and decision-making. For example, AGI would describe the ability to solving any real-world problem that humans encounter daily. AGI is variably estimated to be either decades away or never to occur because of the extent of progress needed (1).

[a]Department of Pathology, Northwestern University Feinberg School of Medicine, Chicago, IL; [b]Ann & Robert H. Lurie Children's Hospital of Chicago, Chicago, IL; [c]Department of Pathology & Laboratory Medicine, University of Ottawa, Canada, The Ottawa Hospital, and the Eastern Ontario Regional Laboratory Association, Canada.
*Address correspondence to this author at: Division of Biochemistry, The Ottawa Hospital, Department of Pathology & Laboratory Medicine, University of Ottawa, 501 Smyth Rd, Ottawa ON, K1H 8L6 Canada. E-mail cmccudde@uottawa.ca.
Received April 7, 2021; accepted June 8, 2021.
DOI: 10.1093/jalm/jfab075

IMPACT STATEMENT

Artificial intelligence (AI) is area of rapid growth and a fact of life when it comes to "big" data. Healthcare and laboratory medicine in particular have numerous rich potential applications including how instruments work, how data is interpreted, and the generation of entirely new predictions. Whatever the application, it is important to understand how AI algorithms are developed, how to assess their validity, how to implement them, and where laboratorians fit in to this emerging area. This review provides a high-level overview of the many considerations of AI with a focus on the perspective of laboratorians.

ASI is extension of AGI, where computers can reason and think but with orders of magnitude more speed and complexity than humans. ASI is the foundation for many science-fiction novels and has no reasonable expectation of occurring anytime soon. ANI, also known as "weak AI," refers to solving specific targeted problems, for which algorithms are expressly designed and validated. ANI is where there has been significant progress in the development uptake and utility of practical applications. ANI in combination with large datasets holds tremendous promise in laboratory medicine and is the focus of this review; AI will heretofore be synonymous with artificial narrow intelligence.

AI (ANI) methodologies include natural language processing, computer vision, robotics, expert systems, and machine learning. Natural language processing describes the ability of computers to analyze and understand text; for example, using an automated phone tree to reach a unit in the hospital. Computer vision is acquiring, analyzing, and understanding images or video; for example, face recognition. Robotics is a large field, but includes control systems, sensors, autonomous vehicles, and reinforcement learning among many others (e.g., a robotic liquid handler). Expert systems rely on human knowledge to create decision rules based on inputs; for example, if/then/else rules for complex decision-making. Finally, machine learning is a diverse set of algorithm creation methods that can be used to classify, group, or make predictions.

Supervised and unsupervised learning are the most commonly used types of machine learning. In supervised learning, the inputs and outputs (i.e., values or classes to be predicted) are known and the computer attempts to learn the mathematical function that best describes the relationship between the inputs and outputs. Regression and classification are forms of supervised learning. If data are not labeled, unsupervised approaches can be used to associate or cluster data into different groups. Unsupervised learning can be useful to subcategorize groups with unknown characteristics, for example into different types of genomic markers in cancer. This is a discovery type of activity that can lead to improved understanding of why or how subgroups exist. Unsupervised learning includes dimensionality reduction, such as principal components analysis, which can be used to compress data into linear combinations of features that explain its variance. This is useful to increase analysis speed and reduce storage needs where there are many variables and large datasets that have high computation demands.

Supervised learning is the prototypical method of using known (classified/labeled) observations to develop a model used to classify future unknowns.

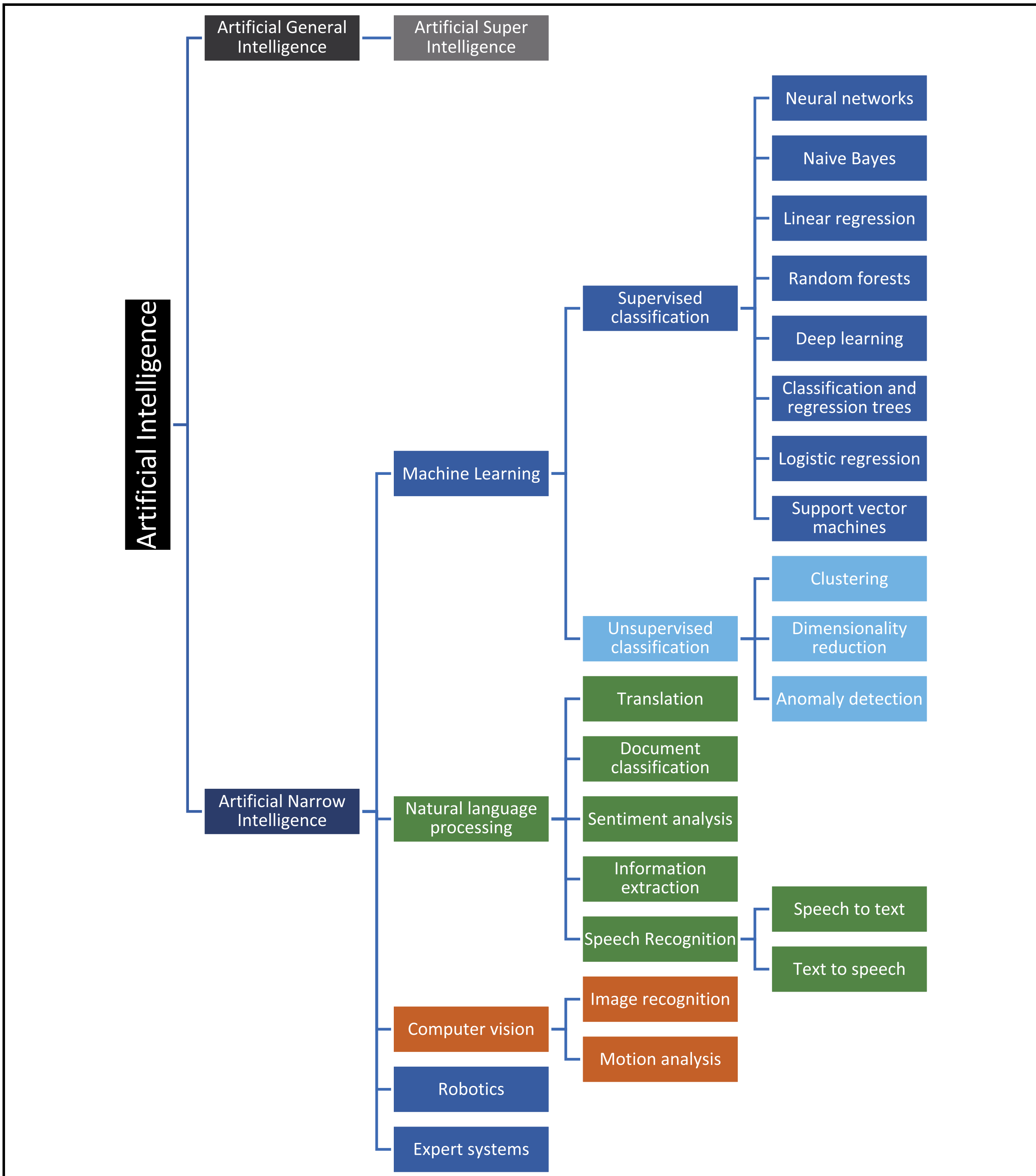

Fig. 1. Hierarchy of artificial intelligence concepts and tools.

For example, predicting acute kidney injury (AKI) from laboratory results using a logistic regression trained on data from patients with and without AKI. Supervised machine learning is the branch of AI where most of the development and practical applications are evolving in laboratory medicine and will be the bulk of the focus of discussion herein. Supervised machine learning includes an array of techniques, such as support vector machines (SVM), linear regression, neural networks, deep learning, and classification and regression trees. At least a few of these, such as linear regression, will be immediately recognizable as a common statistics method. It is worth highlighting here the nuance between machine learning and a concept of statistical learning (2). In machine learning, the focus is largely on getting predictions correct. In statistical learning, there is a further goal of better understanding how the models work. Accurate predictions still matter, but the model itself is of interest as it helps elucidate relationships between variables and predictions. In short, statistical learning is focused on making predictions and inferences, whereas machine learning is focused on predictions. This distinction is largely one of intent, but different algorithms contribute to interpretability. For example, a linear regression model, with familiar concepts of optimizing coefficients or weights to minimize error in 2 dimensions, is easier to comprehend than that of SVM, which involves hyperplanes in *N*-dimensional space.

Key AI concepts for laboratorians are borne from questions that laboratories might ask of AI, such as, "What is the role of the laboratory? What is required to use an AI algorithm in production? How should laboratorians participate in the development of new algorithms? How can laboratorians assess the validity and utility algorithms?" This review will provide an examples of AI use in the laboratory, how and why laboratorians should participate in the use and development of AI, and an overview of how to evaluate and implement AI algorithms.

MACHINE LEARNING APPLICATIONS IN THE LABORATORY

There are many opportunities to apply machine learning in laboratory medicine across all laboratory medicine specialties, yielding a plethora of examples, some of which have received regulatory approval for use as clinical diagnostics (3). Machine learning applications to identify clinical populations at risk for negative outcomes and to aid with diagnosis and prognosis are outside of the examples discussed in this section as they do not typically originate from laboratory medicine or guide its decisions. Many of these patient-level risk prediction models include laboratory values or engineered derivatives as features so they are relevant to other sections of this review. Here, we focus our attention on examples that aim to improve the quality and efficiency of operations and service in clinical laboratories.

Test Utilization

Test stewardship or utilization remains a major emphasis for clinical laboratories. Machine learning has been used to predict test results from other available data in an effort to minimize unnecessary testing. Retrospective, integrated data sets that contain related laboratory values, patient demographics, and clinical labels from diagnosis coding or provider notes, have been mined to investigate the utility of individual components of multianalyte panel tests, specifically those focused on an organ system (e.g., liver panel) or physiologic process (e.g., iron deficiency panel) (4, 5). Beyond multianalyte tests, others have conducted similar studies to predict the expected diagnostic value of a test based on results from related tests.

Zhang et al. demonstrated that patient history of malignancy with results from CBC and differential tests could be used to predict if peripheral blood flow cytometry results would be abnormal, showing potential to decrease unnecessary utilization of peripheral blood flow cytometry by 35–50% (6). In another example, results of HBV surface antigen and anti-HCV antibody immunoassays can be predicted from patient demographics and other laboratory results (7). These examples suggest machine learning approaches may be used to reveal the complex nature of such relationships and may offer support for reconsidering the routine use of multianalyte or related tests.

Roy et al. found that low yield laboratory tests were common in their institution and proposed that rule-based algorithms and machine learning methods could be used to aid in test utilization (8). A follow up study evaluated their approach with additional data and investigated the transference of trained models for use at 2 additional institutions (9). Although the cross-site performance of the models to predict normal test results decreased relative to local (i.e., where the models were trained) performance, the prediction power [as area under the receiver-operator curve (AUROC)] remained reasonably high at both external sites.

Automated Result Review

Autoverification of test results has become routine practice in clinical laboratories, particularly in those with high throughput. These systems apply rules-based logic frameworks to automate the review and release of results, based on single or multiple parameters that assess result quality. The rules are often created based on laboratorian experience or simple signals commonly observed in a limited set of results (typically 1–5 analytes considered simultaneously) and are deployed as a chained series of decisions. Thus, there have been several reports using machine learning methods to potentially discern complex relationships in more comprehensive data sets and improve the efficiency and effectiveness of autoverification or quality review (10–12).

Test Interpretation

As with radiology and pathology, there is interest in applying machine learning to augment result interpretation workflows in laboratory medicine. Test panel interpretation requires recognition of patterns in multivariate data that are interrelated—a task that is well suited to machine learning methods. Two recently published examples demonstrate this proof of principle for classification of urine steroid and amino acid profiles (13, 14). These attempts show great potential in addressing an important problem for clinical laboratories, particularly given the limited number of experts who are trained to provide interpretations for these specialized tests. As noted in an accompanying editorial, the amino acid example also highlights the significant challenge in applying machine learning to rare disease diagnosis (15). Because there were not enough cases of several types of individually rare disorders, these were grouped into a single class called "rare inherited metabolic diseases" whose performance had relatively poor accuracy. While these early reports do not indicate that such solutions are ready as standalone diagnostics, they do show promise as a way to triage results that may be less clear and require human review. Similar work done some time ago demonstrated the use of neural networks for interpretation of serum protein electrophoresis (16). One could conceive of such approaches being applied to the review and interpretation of other profile-based assays, including urine organic acids and hemoglobin electrophoresis.

Machine learning algorithms have also been applied as an alternative to measurement of 6-thioguanine nucleotide metabolites for optimizing thiopurine therapy for inflammatory bowel disease, reducing total expenditures on send-out thiopurine metabolite testing by $75 000 per year (17).

GENOMICS

Genomics is another area of laboratory medicine with significant needs for assistance to process high volumes of data into clinical information. Clinical molecular laboratories employ bioinformatic pipelines to process the large and complex data resulting from modern sequencing methods. Human review steps often follow computational workflows and are time-consuming, costly, not standardized, and lack reproducibility. As constitutional and somatic genetic testing moves toward panels with 100 s–1000 s of genes and even whole exome or genome coverage, assays return variant files whose complexity and size easily surpass human capacity for efficient or effective review. The task becomes even more complicated as interpretation of genetic variation also requires synthesis of data from electronic health records and databases containing information on population-level allele frequencies (e.g., gnomAD) and reported relationships between genomic variation and human health (e.g., ClinVar) or cancer (e.g., CIViC and COSMIC). Thus, machine learning methods may be employed to improve the scalability of genomic analyses and improve their cost, efficiency, and consistency (18–20). For example, Wu et al. found that a random forest classifier could systematically distinguish valid somatic variants from sequencing artifacts in pediatric non-FFPE tumors, reducing the required time to review variants and increasing review capacity by 42% (21).

Microscopic Image Analysis

There has been significant progress in applying machine learning to medical image analysis. Beyond digital pathology, this includes the automation of cell or particulate classification and counting from microscopic images. The US FDA has cleared several such methods for use as clinical diagnostics. For example, the CellaVision DM96 and DM1200 are machine learning-based applications for automated image analysis and cell classification and counting of white blood cells, red blood cells, and platelets in blood and body fluids (22, 23). Neural network approaches have also been used to automate urine sediment analysis on instruments such as the Iris Diagnostics automated microscopy system (iQ200) (24). Both of these examples of AI-automated image analysis are routinely used in clinical laboratories today.

AI and Laboratorians

With the above examples in mind, when it comes to AI algorithms, laboratorians have a variety of potential roles to play (Table 1). These roles include consuming information from algorithms as users, developing algorithms, serving as subject matter experts, evaluating AI work as peer reviewers and administrative decision makers, and being stewards of a raw data source. The level of involvement in these roles ranges from essentially no effort as a consumer or data source to developing significant expertise for direction of algorithm development or its evaluation. The benefits to these roles are largely commensurate with the level of investment and involvement, but whichever path laboratorians take, it is important to understand each.

LABORATORY AS DATA SOURCE

Foundationally, the fuel of AI is data. Laboratory data comprises the largest transactional volume of information in healthcare, such that it is no surprise that laboratories play an important part in clinical predictive models. Indeed, serving as a data source is a current reality as laboratory results are transmitted, extracted, interfaced, warehoused, and cataloged across healthcare entities and research databases. Simply persisting at producing quality results, as laboratorians always have, will enable AI research and development to have an ever-growing fuel source.

Table 1. Potential roles for laboratorians related to AI technologies.
Data source steward
Accurate, high volume, and consistently produced laboratory result data is used in AI today
Variations in availability and quality of some laboratory data can be problematic
Ancillary laboratory-related data sources have great potential as AI inputs
Subject matter expert
Operational knowledge: regulatory, economic, workflow constraints, and considerations
Clinical knowledge: nuance and specifics of specimens, test methods, and results
Method developer and evaluator
Involvement ranges from independent to collaborative and as drivers to consumers
Need for familiarity with at least basic principles of AI
Critical review and participation key to safe, effective, and sustained use of AI

An important consideration for AI is data quality (25, 26). High-quality data will yield better predictions than large amounts of low-quality data. Simply put, with AI, "garbage in garbage out" still applies. Laboratory data is typically of consistent structure and accurate, making it of use with relatively little to no manipulation. Contrast this with other sources, such as diagnostic codes or administrative health data that may be riddled with inaccuracies (27, 28) due to manual entry, seriously limiting usage or requiring herculean efforts to clean and resolve errors.

Beside laboratory results, laboratorians will be familiar with other hidden data sources. These are hidden in the sense that they may not be routinely captured or analyzed, exist on hospital networks, or be supported by hospital IT groups, and are not part of the patient health record. These include quality assurance data, external proficiency testing, procedures, and instrument "metadata." Metadata includes pump pressures, currents, reaction rates, probe usage counts, etc., which are unseen with results, but potential predictors for algorithm development. Manufacturers already use this data to identify potential instrument failures, and they could provide a window into specimen quality, individual test performance, and the accuracy of patient results. While useful, metadata may be difficult to capture because it may not be connected to organizational networks.

While laboratory data is largely accurate and consistent, AI algorithm developers must consider overall data availability, missing values, outliers, and time scales. Outliers and missing data can be managed in data cleaning and preparation steps, but when there are too many problems it can lead to bias. This is particularly true for data that are missing not at random. In this case, there is a relationship between the tendency of a value to be missing and its values. This is a very common occurrence with laboratory data where test orders are often strongly associated with each other and the disease state and measurement methods have limits that censor very high or low values.

In general, availability of data is one of the most common barriers for deploying models in clinical settings. Clinical data are highly distributed and may not be well connected within computing networks. Another problem comes from differences in the time scales in retrospective and production data. Retrospective data sets often have data at much greater frequency than may be available in production data sources. Model development is often conducted using retrospective data that are accessed manually or in an ad hoc manner. When a model is deployed, however, the data must be available at a frequency relevant to the model and may need to be available in an automated way. This is why it is often advisable to rely on accessing or building database structures (data marts, data warehouses) in modeling pipelines. For example, a retrospective data set may have values from every hour or minute of the day, whereas a production data source may update only once per day. Although one could develop a well-performing model based on the hourly trends of data, if the production data source only refreshes every day, necessary data inputs will not be available to make predictions. Overall, laboratory-generated, consistent, high volume, high-quality data will continue to be an important resource for AI in medicine whomever develops the algorithm, but there are many factors that need considered to use it effectively.

SUBJECT MATTER EXPERTS

While laboratories will continue to provide a rich source of data, laboratorians can do much more. As with many aspects of healthcare, laboratorians can improve understanding and value of laboratory results for AI by serving as subject matter experts in clinical testing. Subject matter experts provide domain-specific knowledge that AI programmers and developers need to be successful in creating effective algorithms. Laboratorians possess strong understanding of constraints and considerations due to regulatory, economic, and workflow aspects of clinical laboratories and health systems. In addition, laboratorians have knowledge of important and generally unknown aspects of laboratory data and tests, such as reference intervals, method and instrument changes, workflows, interferences, lot-to-lot variation, ordering biases, imprecision, preanalytical errors, interpretative comments, data quality, and laboratory metadata.

Of these, instrument changes are perhaps the most illustrative of the need for domain-specific knowledge. When instrumentation changes over time, laboratorians will be able to provide detailed explanation as to when instruments changed, how they compare with previous analysis, and how they might be normalized. This could help either subselect results in different time ranges or tune and retrain the algorithm to consider these details. Without that inside knowledge, predictors may fail and potentially result in throwing away useful information. This scenario also applies broadly to result flags (e.g., high, low, critical), which may no longer serve as the same benchmark for comparing results as elevated or decreased.

Augmenting the knowledge of results from their own laboratory, laboratorians are also able to provide perspective on variations in results, which may be relevant when developing an AI algorithm or implementing one from an outside institution. Clinicians using laboratory data may need guidance on how to integrate and use results from many different laboratories or versions of similar laboratory tests (i.e., laboratory vs point-of-care glucose). This requires knowing the differences in performance and bias between laboratories, characteristics of the analytes or tests of interest, how operational changes have occurred over time, and an overall understanding of data quality. As a simple example, a laboratorian could readily provide reality checks for analytes that are known to be standardized vs those that would be expected to differ (e.g., HbA1c vs TSH) or help map laboratory values or guide retraining based on known instrument bias. In more advanced use cases, knowledge of seasonal variation and circadian rhythms could be the difference between a useful prediction and excluding the entire dataset. There are many such examples, which illustrate how intrinsic knowledge to laboratorians may be critical components of feature selection and engineering in AI model development or performance monitoring. The ability to choose predictors and identify their limitations takes AI algorithms from black boxes to useful models that can potentially identify new connections between variables (i.e., statistical learning). Another key attribute of laboratorians as subject matter experts is the ability to classify or label data to develop algorithms. This is an essential part of any algorithm development where the true results need to be known to train, test, and validate predictions (29, 30). As subject matter experts in clinical testing, laboratorians can guide AI algorithm development to maximize the utility of the data inputs.

Further to clinical test knowledge is operational understanding. This type of subject matter expertise can help define how AI will be used, who will use it, and how it will drive patient care or operations. Important operational considerations include the feasibility of a given prediction and intervention strategy and associated details, such as its timing or time scale, the level of specificity needed, the allowable error in decision-making, and who makes the decision and when or in what workflow. These factors may dictate how the problem is formulated, what is predicted, and ultimately how to best deploy the model. The most useful models in laboratory medicine will be developed with the support of laboratorians who provide input and feedback throughout the development process. This will help drive acceptance of the selected model and shape expectations about what the model does and does not do. Reiterating the need to consider time scale, operational understanding is important. Consider that a retrospective analysis might attempt to use a send-out laboratory test, but it cannot be effective if it is not available at the time the algorithm needs it.

Collectively, as subject matter experts, laboratorians have the ability to guide selection of predictors, classify outcomes essential to train algorithms, and focus efforts at realistic and achievable predictions.

Algorithm Development and Evaluation

Beyond providing data and subject matter expertise, is algorithm development and evaluation. The basic process of AI method development is shown in Fig. 2. Like most projects, it starts with defining the goals. Specifically, what is the problem and how might AI be the solution? Once the problem is defined, this frames the selection of inputs (i.e., data and features), potential algorithms, and processes for validating and choosing a model for implementation. Starting with defining the problem, laboratorians are uniquely situated to appreciate and benefit from AI applications.

To develop algorithms, there are several paths for laboratorians, ranging from completely independent work to fully collaborative interdepartmental or interorganizational projects. For some applications, laboratorians may choose to implement a commercially available algorithm. In any of these scenarios, the laboratorian serves as both the evaluator and consumer of the algorithm and a built-in subject matter expert as described above. The specifics of how to develop a given algorithm are beyond the scope of this review, but there are numerous high-quality resources available for free online through universities and other educational sites (2, 31–34).

Similarly, the details of evaluating AI technologies are outside the scope of this review and have been discussed elsewhere (35–38). With the increase in AI-related publications and commercial programs using AI, it is important that laboratorians become familiar with the basic principles of these technologies and participate in their critical review. It is worth noting that laboratorians are accustomed to evaluating diagnostics and many of the same principles apply to the assessment of AI applications. With any new test or tool, a consumer would want to know about the methods used to develop it, what is needed to use it, where it can be applied, and how well it performs at the task.

Poor reproducibility and inadequate descriptions of methods have led to recommendations and guidelines for clear and comprehensive reporting of machine learning development and analyses (39–43). If followed, transparent method descriptions and results should enable critical review, of which some key steps are shown in Table 2.

It is important to note that model performance is not the same as its impact or value. The evaluation of advanced analytic models typically focuses on metrics such as AUROC and mean-squared errors. These may provide a convenient quantitative summary of the performance of the model and describe its ability to represent the data, but they do not reflect the consequences of taking action based on the model's output. Liu et al. offer a framework for understanding a model's value, which is described as the tradeoff between the resources used and benefits gained by developing and deploying the model (44). Although presented in the context of predictions for clinical decision support, the idea of demonstrating model impact or value should be considered more broadly to ensure the effective and sustainable use of machine learning methods. If an algorithm is to be a diagnostic test it also requires clinical validation, which can be a significant barrier to implementation. The US FDA recently issued guidance for validation of software as a medical device and has

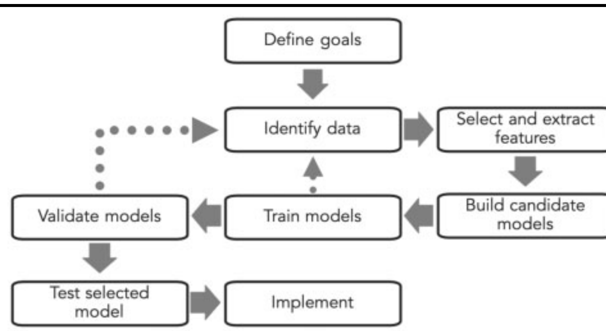

Fig. 2. Overview of the predictive modeling process. Defining the problem or question is an essential starting point.

Table 2. Key points and questions to consider when evaluating AI applications.

Are the purpose and goals for model clearly defined? Does it address a valid problem? Is AI a feasible solution?
Assess the relevance and adequacy of the data, including collection, cleaning, and annotation/labeling procedures. Are potential sources of bias identified and mitigated? Are there sufficient numbers of observations? For classifications, are the subgroups or classes balanced? Is there potential for data leakage?
Are key components of the model (e.g., algorithm type, features, weights, model-specific hyperparameters) described in appropriate detail? Are selection criteria specified? Is the level of interpretability suitable for the problem?
Are features and their relative contributions consistent with available knowledge of the system or workflow?
Review the rigor of the methods for replication, testing, or external validation. Will the results support the generalization of the model?
Does the model demonstrate acceptable performance on a test or external validation set? Are the selected error metrics meaningful for the specific problem? Is there evidence of issues with overfitting? Recognize that training error is usually less than test error and is not a reliable indicator of model performance on new data.
Are data sources adequately described or available? Are software methods and versions defined and is source code available?

cleared several machine learning-based diagnostics through this and other pathways (45). Algorithms developed and implemented as diagnostics at a single laboratory may also be used as Laboratory Developed Tests. Other algorithms may be used as clinical decision support tools.

AI ALGORITHM IMPLEMENTATION

Despite their promise, there remains a gap in translating AI algorithms into production. The lack of translation may help drive a common perception that the machine learning code is the greatest component of a ML-based application and that a successful model equates to a useful application. In reality, for production systems, the ML code is a relatively small component in the overall system, although a very critical one around which other functionality is assembled or architected (46). Barriers to implementation are more often related to data, human factors, and infrastructure requirements.

Computational (or Informatic) Infrastructure

Infrastructure for machine learning consists of several connected IT components and the underlying code base constructed into a pipeline, as is shown in Fig. 3, A. This graphic depicts the deployment of a static model—one whose specification does not change for a while. In this case, the training and testing steps are shown in gray and are performed offline to find the best model, which is selected and implemented. Further retraining may be necessary over time due to changes in input data or workflows, much like routine calibration performed for laboratory instrumentation. The process begins on the left with data ingestion from a raw data store. In a clinical laboratory, this may be the database underlying the LIS or EHR, for example. Machine learning pipelines often rely on relational databases and other aggregated data sources. The system may be designed to include data preparation steps that would generate commonly used features or those designed for specific models. Calculating and saving a field for age, based on the raw output of date of birth, is an example of a feature commonly of interest when developing clinical models. Calculating the weekly lagged volume for a test or a scaled result value are examples of features that may be created and stored to support specific, implemented models. The processed data are fed to the implemented model automatically or at a defined interval—depending on whether the system is running in continuous or batch mode. An output (also called a score) is calculated using input data and any required parameters specified in the model formulation. The score is displayed to the user in the designated interface and point in their workflow. The output and any required metadata may be stored in a database that can also be used for monitoring model performance over time (Fig. 3, B). Although shown as individual entities in this example, the data stores in an analytics pipeline may also be components of a single database. This may be the case if data from an LIS is used in an algorithm and the result is sent back to the LIS.

In its simplest form, the processes in Fig. 3 may be conducted on a single computer, including a personal workstation or server. Basic machine learning development work is often performed using local workstations, such as laptops, personal computers, and virtual desktop machines. More advanced efforts may involve production-level pipelines using servers configured for distributed computing [either on premises (local IT) or in the cloud]. Cloud computing is the web-based delivery of servers, storage, databases, software, and other applications in a flexible manner dependent on user demand. Distributed systems maximize efficiency by using multiple, connected computers to work together on the same task, enabling computing processes to occur independently and in parallel.

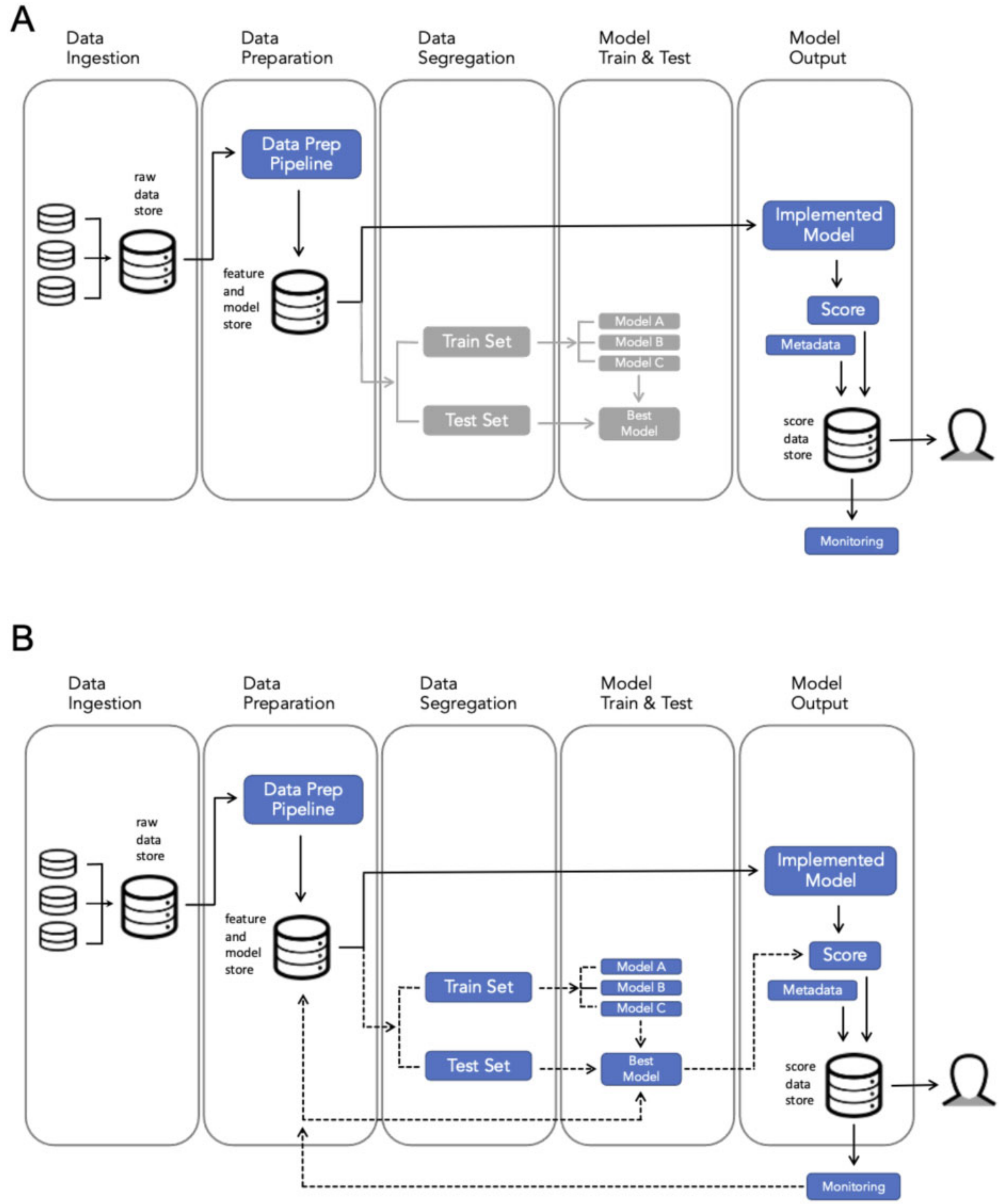

Fig. 3. Machine learning pipeline. (A), Static algorithm. (B), Adaptive algorithm.

Machine learning workflows are reliant on various software frameworks and computer programming scripts that are commonly connected through application programming interfaces (APIs). The most popular coding languages for data science and machine learning are R and Python. Both are available in free, open-source versions. These languages are often used with software libraries, toolkits, or user interfaces that aid in executing the machine learning process and facilitating reproducible workflows. Amazon, Microsoft, and Google also have machine learning applications that provide user interfaces for building, training, testing, and deploying models by leveraging APIs to and from data sources, distributed computing platforms, machine learning libraries, and other resources.

The management and scaling of the system may be automated using workflow orchestration software to control the scheduling and execution or resource allocation (47). Models are increasingly being deployed using container- and cloud-based technologies, which require additional expertise and configuration (48, 49). In addition to the computing and IT infrastructure components, there are other, human resources, that may be new to healthcare organizations, such as data scientists, machine learning engineers, data architects, and solution architects. For comparison, Fig. 3, B shows an adaptive or online model, which would require additional configuration to facilitate the live training and selection of models and features based on continued data inputs and feedback from outputs. This is an even more complex system to deploy and manage.

The choices about where or how to implement a model typically depend on whether it is a custom model or one available from a vendor and if it will be deployed within the EHR or a different clinical system. Still undergoing rapid development, EHR vendors have invested significant resources in enabling advanced analytics within medical records. This includes standing up cloud-based machine learning platforms that integrate with EHRs to score models and provide results at the point of care. These environments facilitate implementation of vendor-provided models for predicting risk of sepsis, AKI, and readmission, for example, and with more work, also can support custom-built models or those from third parties. Making use of the rule builder or scoring systems within EHRs is limited to less complex models, such as linear and logistic regression that can be specified with basic algebraic formulations. Interfacing standards and frameworks, such as FHIR (Fast Healthcare Interoperability Resource) may also be leveraged to connect third-party AI applications to the EHR. Alternatively, an external mechanism such as the example depicted in Fig. 3 may be developed to ingest data, run the algorithm, score the model, and send the results back to the EHR. These same types of external setup may be used to implement models outside of the EHR (i.e., LIS, other laboratory interfaces). Basic models may also be deployed using commonly available business applications, such as Excel and PowerBI or Tableau. Finally, there is future potential for laboratory middleware programs to deploy analytic models.

SUMMARY

AI is changing healthcare. It is adding information, integrating and interpreting existing data, and will be an area of high growth for the foreseeable future. Laboratories are well-positioned for AI because they are a rich source of actionable data and have a unique and valuable discipline expertise. For laboratorians, AI offers potential improvements to testing methods, quality assurance, operations, and diagnostic interpretation and reporting workflows. Implementation of AI algorithms is analogous to bringing in new testing, with both benefits and challenges. The roles laboratories play will be important both for the utility of the algorithms and the value of laboratorians themselves.

Author Contributions: *All authors confirmed they have contributed to the intellectual content of this paper and have met the following 4 requirements:* (a) *significant contributions to the conception and design, acquisition of data, or analysis and interpretation of data;* (b) *drafting or revising the article for intellectual content;* (c) *final approval of the published article; and* (d) *agreement to be accountable for all aspects of the article thus ensuring that questions related to the accuracy or integrity of any part of the article are appropriately investigated and resolved*.

Authors' Disclosures or Potential Conflicts of Interest: *No authors declared any potential conflicts of interest*.

REFERENCES

1. Ramamoorthy A, Yampolskiy R. Beyond mad? The race for artificial general intelligence. ITU Journal. 2018;8:1–8.
2. Hastie T, Tibshirani R, Friedman J. The elements of statistical learning: data mining, inference, and prediction. Springer Science & Business Media; 2013.
3. Topol EJ. High-performance medicine: the convergence of human and artificial intelligence. Nat Med 2019;25: 44–56.
4. Lidbury BA, Richardson AM, Badrick T. Assessment of machine-learning techniques on large pathology data sets to address assay redundancy in routine liver function test profiles. Diagnosis (Berl) 2015;2:41–51.
5. Luo Y, Szolovits P, Dighe AS, Baron JM. Using machine learning to predict laboratory test results. Am J Clin Pathol 2016;145:778–88.
6. Zhang ML, Guo AX, Kadauke S, Dighe AS, Baron JM, Sohani AR. Machine learning models improve the diagnostic yield of peripheral blood flow cytometry. Am J Clin Pathol 2020;153:235–42.
7. Richardson AM, Lidbury BA. Infection status outcome, machine learning method and virus type interact to affect the optimised prediction of hepatitis virus immunoassay results from routine pathology laboratory assays in unbalanced data. BMC Bioinformatics 2013;14: 206.
8. Roy SK, Hom J, Mackey L, Shah N, Chen JH. Predicting low information laboratory diagnostic tests. AMIA Jt Summits Transl Sci Proc 2018;2017:217–26.
9. Xu S, Hom J, Balasubramanian S, Schroeder LF, Najafi N, Roy S, et al. Prevalence and predictability of low-yield inpatient laboratory diagnostic tests. JAMA Netw Open 2019;2:e1910967.
10. Demirci F, Akan P, Kume T, Sisman AR, Erbayraktar Z, Sevinc S. Artificial neural network approach in laboratory test reporting: learning algorithms. Am J Clin Pathol 2016; 146:227–37.
11. Yu M, Bazydlo LAL, Bruns DE, Harrison JH. Streamlining quality review of mass spectrometry data in the clinical laboratory by use of machine learning. Arch Pathol LaboratoryMed 2019;143:990–8.
12. Rosenbaum MW, Baron JM. Using machine learning-based multianalyte delta checks to detect wrong blood in tube errors. Am J Clin Pathol 2018;150:555–66.
13. Wilkes EH, Rumsby G, Woodward GM. Using machine learning to aid the interpretation of urine steroid profiles. Clin Chem 2018;64:1586–95.
14. Wilkes EH, Emmett E, Beltran L, Woodward GM, Carling RS. A machine learning approach for the automated interpretation of plasma amino acid profiles. Clin Chem 2020;66:1210–8.
15. Ganetzky RD, Master SR. Machine learning for the biochemical genetics laboratory. Clin Chem 2020;66: 1134–5.
16. Ognibene A, Motta R, Caldini A, Terreni A, Dea ED, Fabris M, et al. Artificial neural network-based algorithm for the evaluation of serum protein capillary electrophoresis. Clin Chem LaboratoryMed 2004;42:1451–2.
17. Waljee AK, Sauder K, Patel A, Segar S, Liu B, Zhang Y, et al. Machine learning algorithms for objective remission and clinical outcomes with thiopurines. J Crohn's Colitis 2017;11:801–10.
18. Zou J, Huss M, Abid A, Mohammadi P, Torkamani A, Telenti A. A primer on deep learning in genomics. Nat Genet 2019;51:12–8.
19. Ainscough BJ, Barnell EK, Ronning P, Campbell KM, Wagner AH, Fehniger TA, et al. A deep learning approach to automate refinement of somatic variant calling from cancer sequencing data. Nat Genet 2018;50:1735–43.
20. Zomnir MG, Lipkin L, Pacula M, Dominguez Meneses E, MacLeay A, Duraisamy S, et al. Artificial intelligence approach for variant reporting. JCO Clin Cancer Inform 2018;2:1–13.
21. Wu C, Zhao X, Welsh M, Costello K, Cao K, Abou Tayoun A, et al. Using machine learning to identify true somatic variants from next-generation sequencing. Clin Chem 2020;66:239–46.
22. Swolin B, Simonsson P, Backman S, Löfqvist I, Bredin I, Johnsson M. Differential counting of blood leukocytes using automated microscopy and a decision support system based on artificial neural networks - evaluation of DiffMaster ™ Octavia: Leukocyte differential by neural networks. Clin Laboratory Haematol 2003;25:139–47.
23. Kratz A, Bengtsson H-I, Casey JE, Keefe JM, Beatrice GH, Grzybek DY, et al. Performance evaluation of the CellaVision DM96 system: WBC differentials by automated digital image analysis supported by an artificial neural network. Am J Clin Pathol 2005;124: 770–81.
24. Brunzel NA. Fundamentals of urine & body fluid analysis. 4th Ed. St. Louis (MO): Elsevier; 2018.
25. Pilászy I, Tikk D. Recommending new movies: even a few ratings are more valuable than metadata. Proceedings of the third ACM conference on Recommender systems. 2009. p. 93–100.
26. Car J, Sheikh A, Wicks P, Williams MS. Beyond the hype of big data and artificial intelligence: building foundations

for knowledge and wisdom. BMC Med 2019;17:143.
27. van Walraven C, Austin P. Administrative database research has unique characteristics that can risk biased results. J Clin Epidemiol 2012;65:126–31.
28. Lucyk K, Tang K, Quan H. Barriers to data quality resulting from the process of coding health information to administrative data: a qualitative study. BMC Health Serv Res 2017;17:766.
29. Teng T-H, Tan A-H, Zurada JM. Self-organizing neural networks integrating domain knowledge and reinforcement learning. IEEE Trans Neural Netw Learn Syst 2015;26:889–902.
30. Li R, Yin C, Yang S, Qian B, Zhang P. Marrying medical domain knowledge with deep learning on electronic health records: a deep visual analytics approach. J Med Internet Res 2020;22:e20645.
31. Coursera. Build skills with online courses from top institutions. 2021. https://www.coursera.org/ (Accessed April 2021).
32. edX. Free Online Courses by Harvard, MIT, and more. 2021. https://www.edx.org/ (Accessed April 2021).
33. MIT OpenCourseWare. Free online course materials. 2021. https://ocw.mit.edu/index.htm (Accessed April 2021).
34. University of Michigan. Free online courses and MOOCs: class central. 2021. https://www.classcentral.com/university/umich (Accessed April 2, 2021).
35. Liu Y, Chen P-HC, Krause J, Peng L. How to read articles that use machine learning: users' guides to the medical literature. JAMA 2019;322:1806–16.
36. Doshi-Velez F, Perlis RH. Evaluating machine learning articles. JAMA 2019;322:1777–9.
37. Scott I, Carter S, Coiera E. Clinician checklist for assessing suitability of machine learning applications in healthcare. BMJ Health Care Inform 2021;28:e100251.
38. Nagendran M, Chen Y, Lovejoy CA, Gordon AC, Komorowski M, Harvey H, et al. Artificial intelligence versus clinicians: systematic review of design, reporting standards, and claims of deep learning studies. BMJ2020;368:1–12. https://www.bmj.com/content/368/bmj.m689
39. Luo W, Phung D, Tran T, Gupta S, Rana S, Karmakar C, et al. Guidelines for developing and reporting machine learning predictive models in biomedical research: a multidisciplinary view. J Med Internet Res 2016;18:e323.
40. Stevens LM, Mortazavi BJ, Deo RC, Curtis L, Kao DP. Recommendations for reporting machine learning analyses in clinical research. Circ Cardiovasc Qual Outcomes 2020;13:e006556.
41. Pineau J, Vincent-Lamarre P, Sinha K, Larivière V, Beygelzimer A, d'Alché-Buc F, et al. Improving reproducibility in machine learning research (a report from the NeurIPS 2019 reproducibility program). Preprint at arXiv 200312206 [cs, stat]. http://arxiv.org/abs/2003.12206 (2020).
42. Pineau J. The machine learning reproducibility checklist. 2020. https://www.cs.mcgill.ca/~jpineau/ReproducibilityChecklist.pdf (Accessed April 2021).
43. Collins GS, Moons KGM. Reporting of artificial intelligence prediction models. Lancet 2019;393:1577–9.
44. Liu VX, Bates DW, Wiens J, Shah NH. The number needed to benefit: estimating the value of predictive analytics in healthcare. J Am Med Inform Assoc 2019;26:1655–9.
45. United States Food and Drug Administration. Software as a Medical Device (SAMD): clinical evaluation guidance for industry and Food and Drug Administration Staff. 2017. https://www.fda.gov/media/100714/download (Accessed April 2021).
46. Sculley D, Holt G, Golovin D, Davydov E, Phillips T, Ebner D, et al. Hidden technical debt in machine learning systems. In: Cortes C, Lawrence N, Lee D, Sugiyama M, Garnett R, editors. Advances in neural information processing systems.Curran Associates, Inc.; 2015. https://proceedings.neurips.cc/paper/2015/file/86df7dcfd896fcaf2674f757a2463eba-Paper.pdf
47. García ÁL, Fernández-del-Castillo E, Fernández PO, Plasencia IC, de Lucas JM. Resource provisioning in Science Clouds: requirements and challenges. Softw Pract Exper 2018;48:486–98.
48. Pahl C, Brogi A, Soldani J, Jamshidi P. Cloud container technologies: a state-of-the-art review. IEEE Trans Cloud Comput 2019;7:677–92.
49. García ÁL, Lucas JMD, Antonacci M, Castell WZ, David M, Hardt M, et al. A cloud-based framework for machine learning workloads and applications. IEEE Access 2020;8:18681–92.