Proceedings of the Association for Pathology Informatics

Bootcamp 2022

Amrom E Obstfeld  
*University of Pennsylvania*

Victor Brodsky  
*Washington University School of Medicine in St. Louis*

Alexis B Carter  
*Children's Healthcare of Atlanta*

Peter Gershkovich  
*Yale University*

Shannon Haymond  
*Northwestern University*

*See next page for additional authors*

Follow this and additional works at: [https://digitalcommons.wustl.edu/oa_4](https://digitalcommons.wustl.edu/oa_4)

Part of the Medicine and Health Sciences Commons

Please let us know how this document benefits you.

**Recommended Citation**

Obstfeld, Amrom E; Brodsky, Victor; Carter, Alexis B; Gershkovich, Peter; Haymond, Shannon; Levy, Bruce; Sinard, John; Sellers, Devereaux; Stoffel, Michelle; and Jackups, Ronald, “Proceedings of the Association for Pathology Informatics Bootcamp 2022.” Journal of Pathology Informatics. 14, 100331 (2023).  
[https://digitalcommons.wustl.edu/oa_4/2960](https://digitalcommons.wustl.edu/oa_4/2960)

This Open Access Publication is brought to you for free and open access by the Open Access Publications at Digital Commons@Becker. It has been accepted for inclusion in 2020-Current year OA Pubs by an authorized administrator of Digital Commons@Becker. For more information, please contact vanam@wustl.edu.
Proceedings of the Association for Pathology Informatics Bootcamp 2022

Amrom E. Obstfeld a,b,⁎, Victor Brodsky b, Alexis B. Carter c, Peter Gershkovich d, Shannon Haymond e, Bruce Levy f, John Sinard d, Devereaux Sellers g, Michelle Stofel h, Ronald Jackups b

a Department of Pathology and Laboratory Medicine, University of Pennsylvania, Philadelphia, PA, USA
b Department of Pathology and Immunology, Washington University School of Medicine in Saint Louis, St. Louis, MO, USA
c Department of Pathology and Laboratory Medicine, Children’s Healthcare of Atlanta, Atlanta, GA, USA
d Department of Pathology, Yale School of Medicine, New Haven, CT, USA
e Department of Pathology, Ann & Robert H. Lurie Children’s Hospital of Chicago, Northwestern Feinberg School of Medicine, Chicago, IL, USA
f Department of Pathology, Geisinger Medical Center, Danville, PA, USA
g Department of Pathology, MetroHealth Medical Center, Cleveland 44109, OH, USA
h Department of Laboratory Medicine and Pathology, University of Minnesota, Minneapolis, Minnesota, USA

ABSTRACT

Keywords:
Pathology informatics
Data science
Data analytics

The Pathology Informatics Bootcamp, held annually at the Pathology Informatics Summit, provides pathology trainees with essential knowledge in the rapidly evolving field of Pathology Informatics. With a focus on data analytics, data science, and data management in 2022, the bootcamp addressed the growing importance of data analysis in pathology and laboratory medicine practice. The expansion of data-related subjects in Pathology Informatics Essentials for Residents (PIER) and the Clinical Informatics fellowship examinations highlights the increasing significance of these skills in pathology practice in particular and medicine in general. The curriculum included lectures on databases, programming, analytics, machine learning basics, and specialized topics like anatomic pathology data analysis and dashboarding.
Introduction

The Pathology Informatics Bootcamp, presented yearly at the annual Pathology Informatics Summit, is a program specifically designed for pathology trainees to learn essential concepts in the rapidly evolving field of Pathology Informatics. The bootcamps are composed of eight 45-min lectures presented by leaders in the field. Topics for these lectures generally revolve around a single unifying theme representing a core topic in the field of informatics. Previous bootcamp topics have included fundamentals of digital pathology, laboratory information systems, and clinical informatics. The main focus of the 2022 bootcamp was data analytics, data science, and data management.

Pathology and laboratory medicine is amongst the most data-driven fields in medicine, likely because of the well-structured nature of data in laboratory information systems versus the electronic health record which is dominated by free text. As data analysis plays a large role in pathology practice in the present which will only grow in the future, the field must develop didactic programs that teach data skills, appropriate analytical methodology, and analytical modes of thinking. Increased recognition of the importance of this subject is evident by the larger footprint for the topic in the current version of Pathology Informatics Essentials for Residents (PIER).1,2 and the subject has been recently expanded significantly within the latest version of Clinical Informatics Board Examination core content.3,4

In recognition of these important trends, the Association for Pathology Informatics (API) organized the bootcamp at its annual summit around these important topics. Morning sessions addressed general topics in this space including the fundamentals of databases, programming, data analytics, and the basics of machine learning (ML). Afternoon sessions covered selected focused topics such as the application of data analytics to anatomic pathology data, data analytics from the viewpoint of a large healthcare organization, dashboarding, and the practical applications of ML. Table 1 lists the sessions, the major content discussed in each section, as well as a proposed mapping to PIER. We present the proceedings of this bootcamp on pathology and clinical laboratory analytics in order to draw attention to the importance of this topic for pathology practice.

Findings

Databases—Presenter: Peter Gershkovich, MD MHA

Database management systems are omnipresent. They organize and store data for most applications used worldwide. Databases remain critical in laboratory information systems as well. Introduction to databases—the first lecture in the Pathology Informatics Bootcamp schedule—covered the following:

1. the need for secure data management and storage,
2. specific applications for laboratory systems,
3. milestones in the history of database management systems,
4. descriptions of various forms of databases including relational databases (schema, relational principles, common models, SQL, and laboratory schema examples) as well as NoSQL (e.g., document-based) databases and mixed model databases, key/value pairs, graph, column, array/matrix, hierarchical, and network databases.

Using specific examples, the talk demonstrated how significant the choice of databases was, in particular in the implementation of systems for very complex data models (e.g., tumor DNA sequencing software).5 Database choice is essential in 2 key respects—data integrity and development speed. The presentation pointed out relational databases’ limitations. While they demonstrated their advantages in the 90s and are still relevant and widely used today, they typically have a rigid schema and are challenging and expensive to scale. They slow the development process when the underlying data structure and complexity changes rapidly (e.g., in genomics) and the velocity/volume of data is high. Mixed data models (e.g., PostgreSQL) offer benefits that both relational and document-based models provide.

Understanding databases gives Pathology Informaticians the ability to assess the capabilities of the entire laboratory systems—their complexity, agility, interoperability, and the structure of the domain model captured by the LIS. The number of databases continues to grow, and their capabilities are increasing. There is a tendency among database vendors to move towards mixed model systems to expand the clientele, however, these claims of broad functional coverage should be carefully verified. Having specialized databases that focus on a particular task had major advantages and contributed to their popularity. It is not necessary that broader inclusion of database models in one system will lead to better performance. Informaticians should continue to monitor database development progress, especially in the Array/Matrix area where the model directly supports artificial intelligence (AI).

Fundamentals of computer programming—Devereaux Sellers, MD, MBA

The main subjects of the presentation “The Fundamentals of Computer Programming” were programming languages and their structure, version control, testing, and the software development lifecycle. In addition to being a great introduction to programming and software, the topics also touch on areas of informatics that are often overlooked for more trendy topics. Understanding the basics and terminology is the most important thing about these topics. Pathologists are not likely to be regularly writing code but may be involved in discussions with software developers (in-house or outside) where the conversations may revolve around testing, the development life cycle, or versioning.6 Knowing the terms and understanding the basics can prove very useful and enable pathologists to understand and speak their language.

In the near future, we should see most of these topics remain quite stable especially with respect to programming languages and their structure. It is possible that a new language or 2 will be developed, but nearly all languages follow a similar structure, contain variables, and have control structures. The main difference one would expect to see is with the syntax. For version control and testing, we would not anticipate any significant changes, as there are some fundamental methods for version control and testing, and anything new would typically be built upon one of the existing models. There are many different models for the software development life cycle—and it is possible that in the near future, a new model may be developed, though it will most likely be a combination or variation of one of the main existing models. Pathologists will likely become more involved in the informatics space in the near to medium-term, driving development on the cutting edge of both anatomical and clinical pathology.

Fundamentals of analytics: Converting data to knowledge—Michelle Stoffel MD PhD

The major topics for this session were to define general data and analytics concepts, describe data types and characteristics, explore common analytics approaches, and understand the types of limitations which could pose a barrier to deriving meaningful knowledge from data.7 Analytics approaches were explored using specific use cases from a real practice setting (academic and community hospital system). Use cases were chosen in the context of demonstrating analytics efforts that guide decisions, predict needs, provide monitoring, or gain insights. Highlighting quantitative and qualitative methods, use case examples included defining and cleaning molecular pathology testing data for test volumes reporting, identifying the correct patient subset for laboratory test corrections following a chemistry reagent recall, and using qualitative workflow analysis to understand phlebotomy “pain points” to optimize blood collection sample utilization.

The topic of how to select and apply appropriate analytics tools for pragmatic laboratory data needs is important because the validity of laboratory analytics largely depends on understanding the limitations of data quality and the importance of domain knowledge when initially designing analytical approaches. In the laboratory, we are accustomed to thinking about pre-analytical error and sample quality when it comes to specimen analysis, but pre-analytical errors and low-quality data impact our analyses too. A
### Table 1

<table>
<thead>
<tr>
<th>Session title</th>
<th>Presenter</th>
<th>Content</th>
<th>PIER mapping</th>
</tr>
</thead>
<tbody>
<tr>
<td><strong>Databases and Data Management</strong></td>
<td>Peter Gershokovich, MD</td>
<td>• Definition of databases and their role in the laboratory environment&lt;br&gt;• Productivity improvement through databases&lt;br&gt;• The key characteristics that define and distinguish relational databases.&lt;br&gt;• Typical database schemas of Laboratory Information Systems&lt;br&gt;• The role of Query Language (e.g., SQL) in the lab&lt;br&gt;• The principles and ideas behind modern NoSQL databases&lt;br&gt;• Introduction to programming languages&lt;br&gt;• Control structures as fundamental blocks of code facilitating the flow of the program’s execution based on conditions and loops&lt;br&gt;• Version control systems for tracking changes to code over time&lt;br&gt;• Software testing for the evaluation of software functionality and detection of defects&lt;br&gt;• Software development life cycles as frameworks that define the process used to build application</td>
<td>Essential 1 - Topic 2: Subtopic 3: Database architectures&lt;br&gt;Essential 1 - Topic 2: Subtopic 4: Uses of databases in pathology and medicine&lt;br&gt;Essential 4 - Topic 3: Subtopic 1: Data integration, including extract, transformation, and load principles</td>
</tr>
<tr>
<td><strong>Fundamentals of Computer Programming</strong></td>
<td>Devereaux Sellers, MD</td>
<td>• The role of variables in programming languages and their structure&lt;br&gt;• Control structures as fundamental blocks of code facilitating the flow of the program’s execution based on conditions and loops&lt;br&gt;• Version control systems for tracking changes to code over time</td>
<td>Essential 1 - Topic 2: Subtopic 1: Computer hardware and software&lt;br&gt;Essential 3 - Topic 2: Subtopic 2: LIS testing and training</td>
</tr>
<tr>
<td><strong>Data Analytics: Converting Data into Knowledge</strong></td>
<td>Michelle Stoffel, MD, PhD</td>
<td>• Data types and characteristics&lt;br&gt;• Analytical methodologies pertinent to lab medicine &amp; pathology&lt;br&gt;• Review of practical approaches to laboratory analytics education&lt;br&gt;• Limitations in data analysis such as data quality and pre-analytical errors</td>
<td>Essential 1 - Topic 3: Subtopic 1: Structured versus unstructured data&lt;br&gt;Essential 1 - Topic 3: Subtopic 3: The “V”s (volume, velocity, variety, veracity, value) of Big Data&lt;br&gt;Essential 3 - Topic 4: Subtopic 2. Data warehousing (e.g., data lakes, data marts, and Unified Data Architecture (UDA))</td>
</tr>
<tr>
<td><strong>The Basics of Artificial Intelligence and Machine Learning</strong></td>
<td>Alexis Carter, MD</td>
<td>• Fundamental definitions in artificial intelligence and machine learning&lt;br&gt;• Comparison between artificial intelligence systems and other related systems&lt;br&gt;• Artificial intelligence uses and benefits&lt;br&gt;• Review of machine learning development process&lt;br&gt;• Major algorithms in machine learning&lt;br&gt;• Major challenges in artificial intelligence and machine learning</td>
<td>Essential 1 - Topic 3: Subtopic 5: Artificial intelligence and machine learning&lt;br&gt;Essential 4 - Topic 4: Subtopic 4: Machine learning techniques (e.g., &quot;artificial intelligence&quot;)</td>
</tr>
<tr>
<td><strong>Data Analytics in Anatomic Pathology</strong></td>
<td>John Sinard, MD, PhD</td>
<td>• Data analytics’ potential in anatomic pathology&lt;br&gt;• Challenges in acquiring reliable, consistent anatomic pathology data&lt;br&gt;• Complexities with anatomic pathology data in lab information systems&lt;br&gt;• Enhancing data analysis with synoptic reporting and case ‘flags’</td>
<td>Essentials 2 - Topic 5 - Subtopic 3: AP and CP LIS similarities and differences&lt;br&gt;Essentials 3 - Topic 3 - Subtopic 3: Workflow management (e.g., tracking)&lt;br&gt;Essentials 3 - Topic 3 - Subtopic 4: Error tracking and reduction&lt;br&gt;Essentials 3 - Topic 3 - Subtopic 5: Quality metrics and monitoring (e.g., TAT)</td>
</tr>
<tr>
<td><strong>Enterprise Data Analytics: Adding Value to the System</strong></td>
<td>Bruce Levy, MD</td>
<td>• Healthcare analytics at the enterprise level&lt;br&gt;• Data analytics as a hallmark of healthcare industry frontrunners&lt;br&gt;• Health systems’ data challenges and solutions&lt;br&gt;• Pathology data as a vital asset to enterprise data efforts&lt;br&gt;• A hub and spoke model for data analytics to balance enterprise and pathology demands</td>
<td>Essentials 1 - Topic 1 - Subtopic 4: Practices of and interactions between clinical and pathology informatics&lt;br&gt;Essentials 2 - Topic 5 - Subtopic 2: How and why pathology shares data within the healthcare information ecosystem</td>
</tr>
<tr>
<td><strong>So You Want to Make a Dashboard?</strong></td>
<td>Victor Brodsky, MD</td>
<td>• Evolution and purpose of dashboards&lt;br&gt;• Composition and accessibility of dashboards&lt;br&gt;• Dashboard software solutions&lt;br&gt;• Infrastructure and management requirements&lt;br&gt;• Limitations and advancements needed in dashboard utilization</td>
<td>Essentials 3 - Topic 4 - Subtopic 1: Data analytics capabilities and shortcomings of the LIS&lt;br&gt;Essentials 3 - Topic 4 - Subtopic 3: Data analytics and visualization tools in health care and the lab&lt;br&gt;Essentials 3 - Topic 4 - Subtopic 4: Essentials of dashboarding&lt;br&gt;Essentials 1 - Topic 3 - Subtopic 5: Artificial intelligence and machine learning&lt;br&gt;Essential 4 - Topic 3 - Subtopic 4: Machine learning techniques (e.g., &quot;artificial intelligence&quot;)</td>
</tr>
<tr>
<td><strong>Practical Applications of Machine Learning in Pathology</strong></td>
<td>Shannon Haymond, PhD</td>
<td>• Definitions of reproducibility, practicality, and generalizability as they relate to the implementation of ML models.&lt;br&gt;• Standards and best practice guidance in the development of ML models&lt;br&gt;• Importance of educating laboratory professionals in the fields of AI and ML&lt;br&gt;• Clinical laboratorians are critical for ML development and implementation decisions</td>
<td>Introduction to machine learning in pathology and laboratory medicine—Presenter: Alexis Carter MD</td>
</tr>
</tbody>
</table>

This potential didactic approach for this subject might involve a smaller group or workshop-style format, rather than a traditional instructional approach. By providing learners with practical scenarios and encouraging them to brainstorm solutions within an intimate group setting, innovative analytical methods may emerge. This interactive, reciprocal learning experience can prove advantageous for all participants, including the instructors.
characteristics must be understood and addressed before incorporating them into clinical care. The presentation's primary objectives were to: (1) demystify AI and ML, (2) to facilitate basic understanding of the different categories of ML and of appropriate processes for the development of ML tools, and (3) to lay the groundwork for Dr. Shannon Hamond’s subsequent presentation on implementing these tools in the laboratory.

The presentation discussed the foundational concepts and methodologies of AI and ML, many of which are new or easily confused with already existing terms in the medical domain that have different definitions. The differences between ML and traditional programming and statistical methods were highlighted. Examples of AI and ML applications in the laboratory and possible future uses were presented along with the challenges and barriers associated with implementing these technologies in medicine. Ensuring high data quality and quantity, adhering to good scientific research principles, and addressing cybersecurity risks and ethical concerns are crucial aspects of AI and ML development. ML models can be unexpectedly brittle, meaning that very small changes in input (e.g., drift in laboratory results used as input) can cause the model to not perform as expected. Investigators at the Massachusetts Institute of Technology demonstrated that intentionally adversarial changes to a few pixels in images could cause an image classification system to grossly misinterpret the content of an image with a high degree of certainty.10,11 Another critical challenge is the “black box problem”, a term that reflects the difficulty in determining the cause of a model’s erroneous predictions. Other critical challenges include the ethics of the use of AI without a person’s (patient’s) knowledge and determining legal accountability for AI performance gaps. In addition, medicine generally lacks personnel with expertise in this area, strategy for incorporation of these tools, and financial reimbursement mechanisms. Fortunately, some international guidelines are starting to be published regarding the safe use of AI and ML in medicine.12 Specific ML algorithms, the ML model development process, and resources to encourage the audience to explore AI and ML further through independent reading and small projects were also covered. This knowledge is essential for pathologists making informed decisions on adopting AI and ML tools for patient care in their laboratories.

Data analytics in anatomic pathology—Presenter: John Sinard, MD PhD

When one thinks of “data analytics” in pathology, one typically first thinks about laboratory medicine, where the test results are typically numeric, and one can monitor trends and variance within the population. However, there are also many opportunities for data analytics in anatomic pathology. The primary challenge in doing analytics on clinical data is in getting good data—the analytics itself is the easy part. Data acquisition is complicated by restrictions on data access, getting electronic data, standardizing data entered by multiple individuals over time, and dealing with missing or conflicting data. Anatomic pathology data adds the additional challenge that most anatomic pathology laboratory information systems (AP-LISs) treat pathology specimens as if they were all one “test”—but in fact, there are vast differences between specimens with respect to complexity, purpose of evaluation, and required details of the reports.13 Additionally, most of the meaningful data in anatomic pathology systems is stored as free text, making aggregation, and analysis difficult. Finally, because of the growing prevalence of subspecialty practices in anatomic pathology, one needs to be very careful when comparing any metric across pathologist or pathology services to be sure that one is comparing “apples to apples”.

Anatomic pathology analytics can be roughly divided into “Operational Analytics” (e.g., specimen volumes, case types, frozen section arrival times, distribution of grossing workloads, etc.) and “Performance Analytics” (e.g., turnaround time, malignancy rates, histotechnologist throughput, etc.).14 For the most part, since AP-LISs usually have limited built-in capabilities for data analysis and presentation, it is usually best to extract the data from the information system and do the analysis and presentation in software tools designed for this purpose. A number of examples of data analytics metrics were discussed in the course, including case turnaround time (and the complexities of doing that in a meaningful way), case volumes, case distribution across subspecialties (and the decision about whether to count cases, parts, blocks, or slides), and frozen section staffing needs analysis. Synoptic reporting often allows access to discrete data elements, enhancing opportunities for data analysis. Finally, addition of “flags” to cases can be used to create new data for analysis, such as monitoring one’s disagreement rate with outside diagnosis on referral cases.

With increasing use of synoptic reporting and other forms of discrete data storage, the opportunities for data analytics in anatomic pathology will only increase.

Enterprise data analytics—Presenter: Bruce Levy MD

The healthcare industry has been slow to recognize data as a strategic asset. The cause of this is multifactorial including managerial and cultural issues, ethical considerations, absence of a clear return-on-investment (ROI), regulatory restrictions, and a lack of trust of these analytics tools, especially involving AI algorithms. In addition, data and analytics often consist of siloed data repositories and is scattered throughout clinical and business units, resulting in inefficiencies and duplication of efforts. Instead, health systems need to strive towards a system where high-quality data analytics are readily available and seamlessly fit into each persons’ workflows.

Health systems should work towards a system of readily available high-quality data analytics that seamlessly fit into individual workflows by adopting 6 goals:

1. Develop a vision for data and analytics and connect it to the strategic priorities of the organization.
2. Define an organizational structure including roles and responsibilities.
3. Manage the institution’s data assets.
4. Implement a robust data governance program.
5. Establish analytics processes to standardize visualization and delivery of data.
6. Promote the thoughtful implementation and rigorous evaluation of institutional programs and initiatives.

Pathology plays a crucial role in health data and analytics, as laboratories generate significant amounts of health data. Pathologists’ expertise in diagnostic medicine places them in a strong position to analyze lab data effectively, but this can sometimes be blurred with data ownership and control over its analysis. Rather, this represents an opportunity for pathology to share in the leadership of how data and analytics function within the context of patient care and health system operations. The increasing adoption and rise of digital pathology, genomic data, and computational pathology presents new opportunities and challenges including ethical questions, for example, the predictive and prescriptive models within the framework of computational pathology raise serious ethical questions regarding the use of patient data. Pathologists must also adopt to being the subject of data analytics as laboratory processes and workflows become increasingly digitized, enabling monitoring of productivity, workflow optimization, and the impact of the laboratory on the total cost of patient care. By addressing these challenges and embracing a data-driven approach, pathology can contribute significantly to patient care and health system operations.

So you want to make a dashboard?—Presenter: Victor Brodsky

The word “dashboard” was initially used to describe a wooden board at the front of carriages protecting drivers and passengers from mud and rocks being splashed (or “dashed”) by the horses that pulled them. Today, it refers to a graphical user interface with an “at a glance” view of key performance indicators. Dashboards may be used to consolidate a visual presentation of key performance metrics, enable forecasting, inform strategic decisions, and to facilitate early detection of trends, outliers, and correlations. Mostly, dashboards fall into one or more of the following categories: strategic (displaying snapshots in time), analytical (detailed),
operational (live), and balanced score cards (showing metrics for progress towards specific strategic objectives). 

Dashboards are generally composed of data visualizations such as scatter and bubble charts for demonstrating relationships; histograms, scatter, 3D area, or violin charts for distributions; line or bar charts for comparisons over time; and pie, waterfall, and stacked bar or area charts for the composition of groups. A table or grid on the other hand is optimal if individual exact values are important, if magnitudes, units, or time periods are not comparable, if an interactive sortable list is convenient, or if data input is expected. To be inclusive, chart color palettes suitable for the color-blind and compliance with the “WA1-ARIA” (Web Accessibility Initiative—Accessible Rich Internet Applications) standard within web pages to enable screen reader software are good choices. Dashboard data should be presented in a simplified manner, allowing for quick assessment with minimal user interaction.

Among commercial dashboard software solutions, Tableau and Microsoft’s Power BI together appear to currently have over 50% of the market share, while the remainder is split between more than 30 different commercial and open-source packages. For the more adventurous, development of dashboards from scratch can be accelerated by relying on open-source data visualization packages. Having taken our own advice at Washington University School of Medicine in St. Louis, we have produced a web-based dashboard platform framework, we plan to open source in the near future. Once available, it will be accessible at https://github.com/victorbrodsky to enable anyone interested in dashboards and coding to proceed directly to building specific charts of interest while relying on existing boilerplate code for this software for the rest of the described functionality.

A reliable dashboard that stands the test of time requires considerable effort in developing reliable infrastructure and data streams. One essential aspect involves implementing role-based permissions and user management to ensure that specific data are only visible to authorized users. System administrators play a crucial role in managing access, monitoring uptime, and conducting tests after upgrades. To achieve optimal application responsiveness, it may be necessary to add suitable database indexes and periodically pre-generate visualizations.

It is also important to acknowledge the limitations of dashboards. Manually filtering dashboard charts by numerous permutations of criteria for pattern identification is challenging. Anomaly detection algorithms streamline this process by continuously highlighting relevant data subsets or triggering notifications. This reduces manual analysis and promotes automated, actionable alerts. Everyone wants a dashboard until they get one. With it, comes the realization that a timely notification from someone else, preferably someone vigilant, who is continuously monitoring said dashboard for expected (and unexpected) anomalies would be infinitely better.

Practical applications of machine learning in pathology—Presenter: Shannon Haymond PhD

Though the scientific literature is filled with reports of ML methods, there is a gap in the translation of these potential innovations to patient care or clinical operations. There are likely many reasons for why this is the case. Perhaps these reports do not describe tools that are solving relevant problems as researchers are predicting what they can rather than what is clinically useful. In other cases, the methods may contain pitfalls leading to overly positive results, limiting their generalizability for broad implementation or potentially creating harm. Some reported methods may not be appropriately described to be reproduced or transferred within varied clinical workflows. There are additional technical difficulties with implementing even robust ML methods, especially those using highly complex algorithms, within hospital or laboratory infrastructures. Interoperability challenges, limited data access, closed informatics systems, and a lack of experience and availability of computational resources may hinder or impede the implementation process.

To realize the promise of ML in clinical laboratories, solutions must be reproducible and practical. Work is reproducible if consistent results can be obtained when using the same input data, steps, methods, code, and conditions. Applications are practical if they are generalizable and clinically useful. Generalizability describes the extent to which results of a method apply in other contexts or populations that differ from the one initially validated. Clinically useful applications are those that can be implemented to aid decision-makers by improving patient care or clinical service or operations.

This will be achieved through establishing standards or best practices, providing education to healthcare professionals, and involving them as subject matter experts in the development and reporting of these applications. Checklists and other guidance are available to describe best practices for developing and reporting ML methods. These are useful for both the validation and the critical review of ML-based methods. Because of the great interest in ML in nearly every industry, high quality educational resources are easily accessible for a variety of skill levels and in different formats. Those wanting to acquire more advanced skills may even consider completing a formal degree or a certificate program. Laboratory medicine-focused offerings are also becoming available from professional societies such as API, AACC, MSACL, and others. To best support human decision-making, clinical laboratory professionals, as subject matter experts, should be involved in method development and implementation decisions. This will help to: (i) avoid negative consequences such as alert fatigue and critical misses, (ii) build trust through model interpretability and familiarity with results, and (iii) ensure ML is integrated into clinical workflows in an effective manner.

Conclusion

The ability to effectively use data is transitioning from being a desirable skill to a necessary one for healthcare professionals in general and pathologists in particular. Although pathology-operated laboratories are just one of the data sources that constitute the electronic health record, they provide a disproportionately large amount of the discrete and readily computable data. It has been noted that there is a shortage of pathologists with these skills, a troubling fact given the predicted increase in the need for these skills in the workforce. Addressing this gap is crucial to ensure a sustainable workforce and meet the emerging needs of the field of pathology as it increasingly evolves into a fully digital discipline.

The goal of this bootcamp was to establish a mental framework for pathology trainees in the fundamentals of data management and data science. Data management at its core involves the organization and storage of data throughout its lifecycle. Data science is an interdisciplinary field that involves using statistical, computational, and ML techniques to extract insights and knowledge from data. In combination, these fields can be leveraged to provide a range of data resources from simple reports to useful predictions of patient outcomes. The bootcamp was designed to contribute significantly to the educational landscape of pathology trainees by establishing a baseline of knowledge in these fields both in the live sessions and in this summary for the scientific literature.

The API continues to offer educational opportunities and resources to its members and the Pathology and Laboratory Medicine community at large. The 2023 bootcamp focused on ethical and regulatory concerns in data management and analytics. Future programming regarding other topics in Pathology Informatics such as Digital Pathology, Clinical Informatics, and important standards and terminologies in Pathology practice is planned in order to fulfill API’s educational mission.

Conflict of Interests

The authors declare the following interests: Editorial board of the Journal of Pathology Informatics (ABC). Paid faculty member for the American Medical Informatics Association Clinical Informatics Board Review Course in which I teach artificial intelligence and machine learning (ABC).
References


