ORIGINAL ARTICLE

A Road Map for Translational Research on Artificial Intelligence in Medical Imaging: From the 2018 National Institutes of Health/RSNA/ACR/The Academy Workshop

Bibb Allen Jr, MD[a], Steven E. Seltzer, MD[b,c], Curtis P. Langlotz, MD, PhD[d], Keith P. Dreyer, DO, PhD[e], Ronald M. Summers, MD, PhD[f], Nicholas Petrick, PhD[g], Danica Marinac-Dabic, MD, PhD, MMSC[h], Marisa Cruz, MD[i], Tarik K. Alkasab, MD, PhD[e], Robert J. Hanisch, PhD[j], Wendy J. Nilsen, PhD[k], Judy Burleson, BSW, MHSA[l], Kevin Lyman, BS[m], Krishna Kandarpa, MD, PhD[n]

Abstract

Advances in machine learning in medical imaging are occurring at a rapid pace in research laboratories both at academic institutions and in industry. Important artificial intelligence (AI) tools for diagnostic imaging include algorithms for disease detection and classification, image optimization, radiation reduction, and workflow enhancement. Although advances in foundational research are occurring rapidly, translation to routine clinical practice has been slower. In August 2018, the National Institutes of Health assembled multiple relevant stakeholders at a public meeting to discuss the current state of knowledge, infrastructure gaps, and challenges to wider implementation. The conclusions of that meeting are summarized in two publications that identify and prioritize initiatives to accelerate foundational and

[a]Department of Radiology, Grandview Medical Center, Birmingham, Alabama.

[b]Radiology Department, Brigham and Women's Hospital, Boston, Massachusetts.

[c]Radiology, Harvard Medical School, Boston, Massachusetts.

[d]Department of Radiology, Stanford University, Palo Alto, California.

[e]Department of Radiology, Massachusetts General Hospital, Boston, Massachusetts.

[f]Radiology and Imaging Sciences, National Institutes of Health Clinical Center, Bethesda, Maryland.

[g]Center for Devices and Radiological Health, US Food and Drug Administration, Silver Spring, Maryland.

[h]Division of Epidemiology, Center for Devices and Radiological Health, US Food and Drug Administration, Silver Spring, Maryland.

[i]Digital Health Unit, Center for Devices and Radiological Health, US Food and Drug Administration, Silver Spring, Maryland.

[j]Office of Data and Informatics, Material Measurement Laboratory, National Institute of Standards and Technology, Gaithersburg, Maryland.

[k]National Science Foundation, Division of Information and Intelligent Systems, Alexandria, Virginia.

[l]American College of Radiology, Department of Quality and Safety, Reston, Virginia.

[m]Enlitic, San Francisco, California.

[n]Research Sciences & Strategic Directions, Office of the Director, National Institute of Biomedical Imaging and Bioengineering, The National Institutes of Health, Bethesda, Maryland.

Corresponding author and reprints: Bibb Allen Jr, MD, Department of Radiology, Grandview Medical Center, 390 Grandview Pkwy, Birmingham, AL 35243; e-mail: bibb@mac.com.

Bibb Allen Jr, MD, discloses Chief Medical Officer, ACR Data Science Institute, Reston, Virginia. Steven E. Seltzer, MD, discloses GE Healthcare: Dr Seltzer is a member of the Board of Review for the GERRAF program. He receives travel expense reimbursement for one board meeting per year. Curtis P. Langlotz, MD, PhD, reports personal fees, nonfinancial support, and other from Montage Healthcare Solutions; nonfinancial support and other from Nines.ai; personal fees, nonfinancial support, and other from whiterabbit.ai; nonfinancial support and other from GalileoCDS, Inc; nonfinancial support and other from Bunker Hill, Inc, outside the submitted work; and board of directors, RSNA. Keith J. Dreyer, DO, PhD, discloses Chief Science Officer, American College of Radiology Data Science Institute, Reston, Virginia. Ronald M. Summers, MD, PhD, reports royalties from PingAn, ScanMed, Philips, iCAD; research support from PingAn, and GPU card donations from NVIDIA, outside the submitted work; Dr. Summers' work is supported by the Intramural Research Program of the National Institutes of Health Clinical Center. Tarik K. Alkasab, MD, PhD, reports his department has a consulting relationship with Nuance Communications, Inc. Kevin Lyman is CEO of Enlitic. The other authors state that they have no conflict of interest related to the material discussed in this article. This work was supported in part by the Intramural Research Program of the National Institutes of Health Clinical Center.

© 2019 Published by Elsevier on behalf of American College of Radiology
1546-1440/19/$36.00 ■ https://doi.org/10.1016/j.jacr.2019.04.014

translational research in AI for medical imaging. This publication summarizes key priorities for translational research developed at the workshop including: (1) creating structured AI use cases, defining and highlighting clinical challenges potentially solvable by AI; (2) establishing methods to encourage data sharing for training and testing AI algorithms to promote generalizability to widespread clinical practice and mitigate unintended bias; (3) establishing tools for validation and performance monitoring of AI algorithms to facilitate regulatory approval; and (4) developing standards and common data elements for seamless integration of AI tools into existing clinical workflows. An important goal of the resulting road map is to grow an ecosystem, facilitated by professional societies, industry, and government agencies, that will allow robust collaborations between practicing clinicians and AI researchers to advance foundational and translational research relevant to medical imaging.

J Am Coll Radiol 2019;■:■-■. *© 2019 Published by Elsevier on behalf of American College of Radiology*

BACKGROUND

Fueled by advances in computing power, data availability, and machine learning techniques, applications of artificial intelligence (AI) are rapidly increasing in many industries, including health care and medical imaging. Although initial hype suggested that AI algorithms might soon replace portions of our diagnostic imaging workforce [1–3], most experts believe, if properly applied, AI will augment the care being provided by radiologists and improve patient outcomes [4,5]. But with these promises also come perils. The Department of Defense, understandably interested in advances in AI technologies, cosponsored the 2017 Jason Study on AI relevant to national defense, which in part highlighted the unintended consequences of certain computational approaches to data evaluation and how easily algorithms can be misled such that the slightest intentional introduction of what a human would see as noise can significantly change the result or confidence level of the algorithm inference [6]. That brittle nature of AI was not acceptable to the Department of Defense, and neither should it be acceptable for medical imaging. If AI for medical imaging is not developed and advanced in a way that robustly addresses its current frailties and ensures that algorithms are useful, safe, effective, and easily integrated into physicians' workflows, AI models in health care could produce misinformation, could introduce unintended bias, and may actually hinder care rather than enhance it [7]. Interconnectivity, interoperability, and cybersecurity challenges will need to be resolved parallel to advances in AI technology to ensure integrity of training data and effective implementation of AI in clinical practice. Additionally, end users with a broad level of expertise must be involved in algorithm testing to ensure algorithms perform as anticipated.

The stakeholder community for AI in health care is very complex. To develop and maintain an ecosystem for AI in medical imaging, both software development and health care stakeholders must be considered. Although data science researchers and developers of AI applications may be more acquainted with the software development ecosystem, they may be unfamiliar with the numerous idiosyncrasies of the health care ecosystem. For the health care system to become a viable market for AI applications related to medical imaging, the diagnostic imaging community, including both researchers and practicing physicians, must take on a leading role in working with data science researchers, AI developers, and standards bodies to ensure effective pathways for both foundational and translational research to ensure reliable deployment of AI into clinical practice.

On August 2, 2018, a 2-day international workshop convened by the National Institute of Biomedical Imaging and Bioengineering and cosponsored by RSNA, ACR, the Academy of Radiology and Biomedical Research (The Academy), and the National Institutes of Health (NIH) was held to bring stakeholders from academia, industry and government, including the US FDA, the National Institute of Standards and Technology (NIST), National Science Foundation (NSF), and NIH, together to discuss current knowledge and research gaps to identify and prioritize future initiatives for foundational and translational research in AI for medical imaging. Presenters included an international group of data science researchers, radiologists, industry developers, medical specialty society representatives, and government officials. The audience included hundreds of on-site and virtual attendees representing all stakeholders, many of whom provided valuable insights to the workgroup. As a result of the workgroup, a road map for foundational research has been recently published [8]. This article will summarize the translational research aspects of those discussions. All sessions were recorded and are available at the NIH website [9,10].

OVERVIEW OF THE CURRENT STATE

Concept to Market

AI development in health care (Fig. 1), as in any industry, will have a cycle from concept to deployment that begins as a

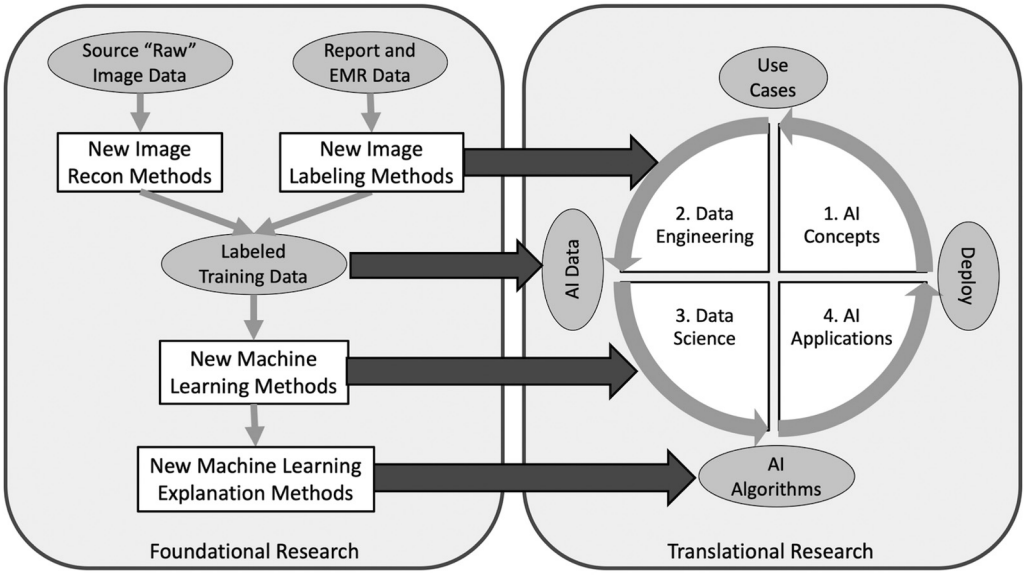

Fig 1. As in other industries, AI development in medical imaging includes both foundational and translational research activities. The foundational portion of the National Institutes of Health Workshop considered research priorities to accelerate and improve the development of AI algorithms for medical imaging [8]. The translational portion of the workshop considered medical imaging use cases for algorithm development and how these applications will be validated, deployed, and monitored in routine clinical practice. The diagram shows how foundational and translational research activities are connected. Foundational research leads to new image reconstruction and labeling methods, new machine learning algorithms, and new explanation methods, each of which enhance the data sets, data engineering, and data science that lead to the successful deployment of AI applications in medical imaging. AI = artificial intelligence; EMR = electronic medical record; Recon = reconstruction. The figure was developed by the authors for publication in both *Radiology* and *JACR*. This figure also published in reference 8.

concept of what humans want an AI algorithm to do—that is, establish a clinical need (Quadrant 1). The next step is the data engineering component in which machine learning techniques are used to create an AI algorithm (Quadrants 2 and 3). Finally, integration of the algorithm software into other applications must occur to fulfill the target audience's needs (Quadrant 4). In AI health care development, this translational cycle will be fueled by foundational research at each step. The interconnections will form the basis of how AI will be delivered to the health care community. Initially, we will likely see machine intelligence combine with human intelligence in small incremental ways that improve patient outcomes under specific circumstances. [1,5]. There may be no single sentinel moment that defines the ubiquitous use of AI in health care. Most people do not recognize that they have AI in their phones; they just know that their phones have become progressively "smarter" [11]. The deployment of AI tools in health care will be analogous to the progression of AI in smartphones.

As humans look for ways to incorporate AI onto the practice of medical imaging, it is equally important to understand the radiology information cycle (Fig. 2). The cycle begins as a clinical decision to request a diagnostic procedure (Quadrant 1) followed by patient preparation, determination of protocol, and other pre-acquisition steps (Quadrant 2) before performing the examination (Quadrant 3). After the examination is performed, the radiologist's interpretation informs the clinical care team with examination results and recommendations (Quadrant 4). Although the interpretation phase is getting the most attention and media hype for AI, progress in AI research and development is occurring in all areas, and subject matter experts will be necessary to convert ideas to improve care into use cases for AI development for all phases of the information cycle. Radiologists should be ready to take the lead in identifying the most important areas for AI development.

Focusing on "Narrow AI"

The rate of computational growth over time suggests that at some point in the future general AI [12] applications—in which machines will be able to learn, recognize, generalize, and perform tasks as humans do, in essence becoming a physician—will impact the way health care is delivered. But that is well into the future, and

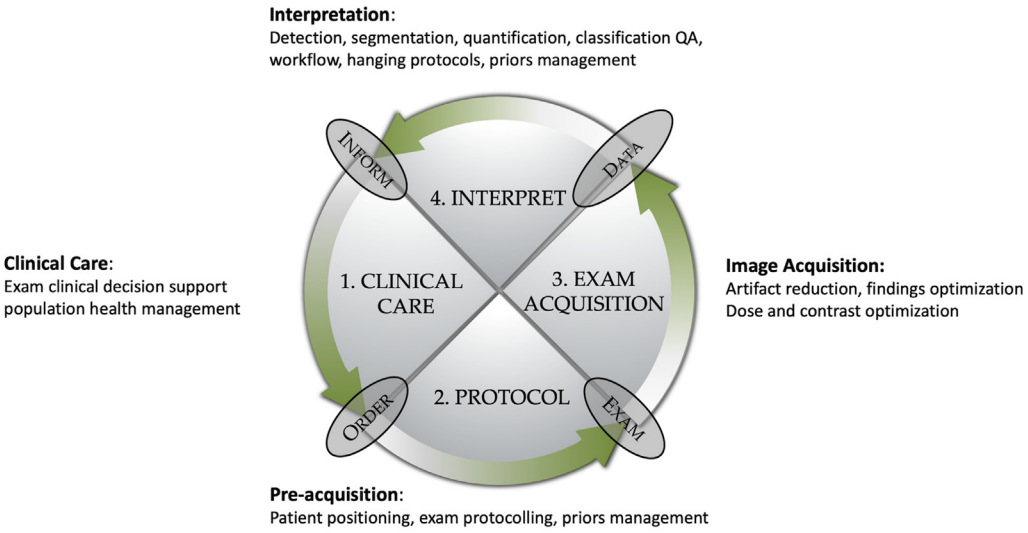

Fig 2. The radiology information cycle begins with the decision to request an imaging study (Quadrant 1). Before image acquisition, previous examinations are reviewed, protocols are determined for the current examinations, and patients are prepared and positioned (Quadrant 2). Next, images are acquired with the proper technique and radiation dose (Quadrant 3), and finally, the examination is interpreted, incorporating prior relevant information and recommendations are communicated to the rest of the care team (Quadrant 4). Opportunities exist for artificial intelligence to impact each phase of the cycle. QA = quality assurance.

inappropriately accelerating AI deployment could result in mis implementation and impede its potential benefit to health care. For now, and the foreseeable future, narrow AI applications—in which the focus is on using AI to help solve *specific challenges* in medical imaging, such as pneumothorax detection or lung nodule classification—hold considerably more promise for improving patient care than general AI applications. Additionally, algorithm performance need not be superhuman to assist radiologists in the care of their patients. If algorithm accuracy is better than the lower end of the radiologist performance scale, then patients will benefit (Fig. 3).

CHALLENGES, KNOWLEDGE GAPS, AND INFRASTRUCTURE NEEDS FOR AI IMPLEMENTATION IN CLINICAL PRACTICE

As with other new technologies that have been translated from initial research to widespread clinical practice, we need to recognize that there will be novel challenges for the clinical deployment of AI tools (Fig. 4). Understanding the nature of these new challenges, potential mitigation strategies, and a well-conceived research road map that ensure that advances in AI algorithm development are efficiently translated to clinical practice are of paramount importance. Much of the work in AI development is being done at single institutions with single institution data for training, testing, and validation of the AI algorithms. A recent review of studies that evaluated the performance of AI algorithms for the diagnostic analysis of medical images found only 6% of the 516 studies reviewed performed external validation [13], and so far, there is limited research demonstrating the generalizability of these algorithms to widespread clinical practice.

PRIORITIES FOR ADVANCING THE USE OF AI IN MEDICAL IMAGING

Developing Structured AI Use Cases for Medical Imaging

To date, AI use cases for medical imaging have been poorly defined, and there is a lack of standardization of inputs and outputs among comparable algorithms cleared by the FDA for marketing specific AI applications [4]. Because algorithms may run on the modality itself, on a local server, or in the cloud (systems that may eventually host thousands of algorithms), a standard way of accepting inputs for the algorithm to process will be required. Without standardized inputs and outputs for AI use cases, it becomes challenging to develop standard data sets for training and testing, and in the end,

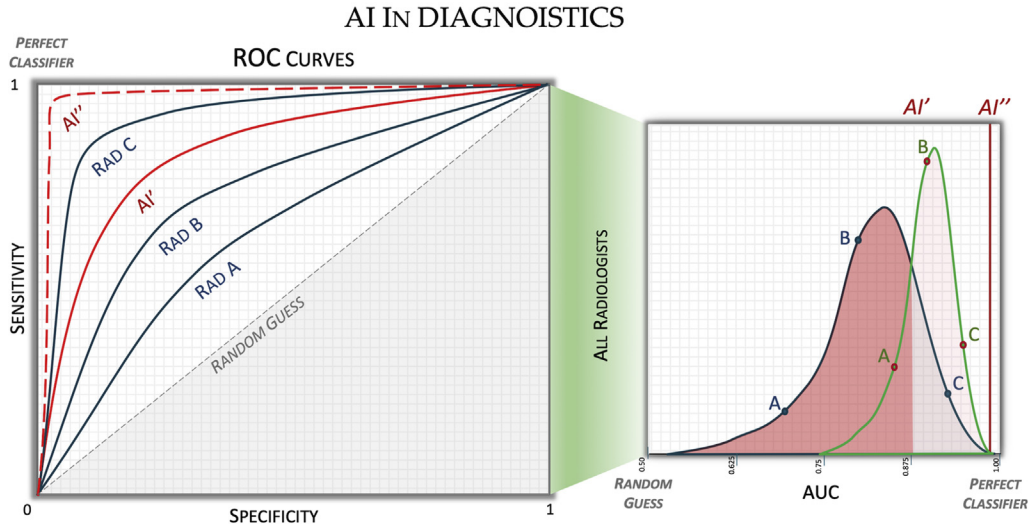

Fig 3. The performance of AI in clinical practice need not be superhuman (AI[II]) to improve patient care. If AI can perform at a level better than the lower end of the radiologist performance scale (RAD A and RAD B), then patients will benefit, and the overall performance curve for radiologist interpretations will be shifted to the right (shown in green), improving on the baseline performance (shaded in red). AI = artificial intelligence; AUC = area under the curve; ROC = receiver operating characteristic.

the resulting algorithms may show different results for the same finding. Ideally, AI use cases should be developed using a format that converts human narrative descriptions of what the algorithm should do to machine readable language such as Extensible Markup Language or JavaScript Object Notation using clearly defined data elements. Structured AI use cases can create standards for algorithm validation before marketing for clinical use and standards for monitoring algorithm performance in clinical practice [12]. The medical imaging community, including medical specialties, academic institutions, and individual radiologists, can positively influence AI development by participation in the development of these structured use cases as well as creating the general standards and structure for AI specific use cases so that AI algorithms can be built with the same definitions and deployed in clinical practice in a consistent way.

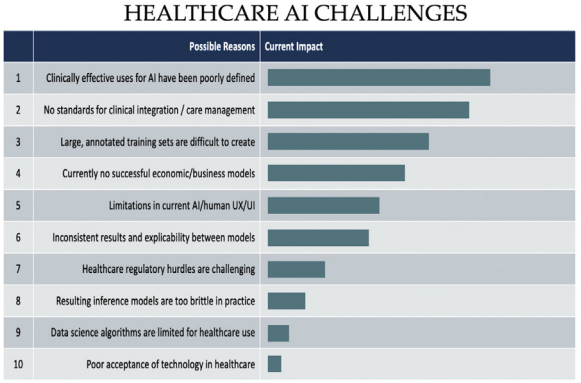

Fig 4. Although initial applications of user experience AI in health care are promising, there is currently limited use of AI in routine clinical practice. Of the challenges faced by researchers and developers for adoption of AI in clinical practice, workshop panelists, whose collective opinions are represented by the bars on the right, considered the lack of availability of structured data for training, testing, and validation of AI algorithms; the development of standards for clinical integration of AI algorithms; and the development of structured AI use cases for health care AI most important. AI = user experience; UI = user interface; UX = user experience.

Common Data Elements

The need for structured data in the form of common data elements (CDEs) is critical for AI development and its translation to clinical practice. Using CDEs in AI use case development ensures AI algorithms perform as expected across a variety of settings and stems from the changing role of the radiologist in the modern patient care. Clinical care has become more data oriented and data driven, and management of patients after imaging examinations has become more algorithmic with many being implemented in health care computer systems such as standard order sets for certain conditions. In addition, it is also imperative that radiologists are able to ingest structured information into their reporting environment. A registry of CDEs is being created [14] for use in a variety of radiology reporting tools [15,16] and in AI use case specifications so that algorithms designed for a similar purpose will be able to integrate the appropriate

examination data, create outputs that are standardized, and present final inferences in similar form and structure so as to optimally populate structured imaging reports.

Developing Standards and Methods for Data Curation, Distribution, Sharing, and Management

There is a significant need for high-quality data sets, containing appropriate annotations or rich metadata, for developing high quality AI algorithms that can meaningfully reflect desired goals of *reliability, reproducibility,* and *explainability*. Developers are building costly tools for data extraction, de-identification, labeling, and workflow integration with each developer potentially creating proprietary assets. The FAIR initiative (making data findable, accessible, interoperable, and reusable) [17] and the democratization of *narrow AI* through available libraries, relatively small data sets, and cheap access to high-performance hardware are providing opportunities for a growing number of researchers and developers to participate in AI development. However, much of the innovation continues to be concentrated at "data-rich" organizations, limiting widespread availability of data. Additionally, privacy concerns often limit the ability of institutions to allow protected health information to be shared externally, and the lack of publicly available data can slow AI development [18–20]. Accelerating the release of publicly available data sets and AI techniques such as transfer learning that allow patient data to remain behind an institution's firewall while exposing algorithm training to more diverse data may be able to help accelerate translation of AI into clinical practice [21].

Optimizing the User Interface and User Experience and Advancing Standards for Clinical Integration and Care Management

The lack of user interfaces (UIs) for bringing in the results of AI models into the clinical workflow as well as a lack of efficient user experience (UXs) are limitations to deployment of AI models for widespread clinical use. An efficient UI and UX for integration into existing clinical workflow tools such as PACS, Radiology Information System, and electronic health records will be necessary for clinical use of AI, and although individual developers will likely distinguish themselves from one another based on these interfaces, their development cannot occur without vendor-neutral interoperability standards for electronic communication between health IT (HIT) resources. Understanding the infrastructure needs, including both qualitative and quantitative analyses [17,22,23] for AI deployment in clinical practice—either local or cloud-based—will be critical in allowing use of thousands of AI algorithms in actual clinical practice. The medical imaging community must be involved in assessing the clinical and infrastructure needs and work with existing standards bodies such as the National Science Foundation and the NIH Connected Health Initiative [24] to find solutions that facilitate adoption of AI in clinical practice.

Ensuring Patient Safety and Health Equity and Developing Efficient Pathways for Safely Bringing Software Tools to Market

The medical community must work with developers, government agencies, and the public to ensure AI is deployed in medical practice, such that the end users of the software and the public can be confident that the algorithm output is accurate, free of unintended bias, and safe for patients, including protections from cybersecurity lapses. This means ensuring that the claims of the AI developers are validated using novel data sets created using technical and demographic diversity before marketing for clinical practice. The US FDA is charged with protecting the public health by "ensuring the safety, efficacy, and security" of a wide range of health care products [24,25]. The FDA regulates a broad array of medical imaging devices as well as computer-aided diagnosis software and other algorithms that provide decision-making support to medical practitioners [24]. The agency recognizes the rapid increase in digitization across the health care continuum and the importance of regulating computer software that is able to detect and classify disease processes and has been issuing regulatory guidance for software computer-aided detection and computer-aided diagnosis since 2012 [26]. The FDA is working with the International Medical Device Regulators Forum [27] to develop harmonized and convergent guidance for software as a medical device (SaMD) [28]. One of the outputs of this workgroup is a recommendation that SaMD should have a clinical assessment, analytical validation, and clinical validation as part of the development process. The FDA has adopted this principle as an important consideration in developing its upcoming guidance for SaMD [28]. Additionally, international regulatory collaborations, consistent international data standards, and methods for data sharing would also be useful in streamlining the regulatory process for considering marketing algorithms in international markets.

The FDA is committed to using real-world evidence (RWE) to support device pre- and postmarket decisions and has developed the National Evaluation System for

Healthcare Technologies (NEST) program to accelerate the development and translation of new and safe health technology to the clinic by leveraging RWE and other innovative data into the approval and postmarket surveillance processes (Fig. 5) [26]. The NEST is selecting teams for RWE assessment and value analysis initiatives including at least one demonstration for SaMD and AI [29-31]. The FDA is also planning to leverage RWE is its Pilot Software Recertification Program [32].

There are opportunities for developing partnerships with the NEST program to help provide the FDA with RWE. The Medical Device Epidemiology Network is a global public-private partnership, and it incorporates over 130 partners and over 100 national and regional registries from 37 countries [33]. Additional public-private partnerships are anticipated as well, focused on developing data registries for monitoring the performance AI algorithms in clinical practice will be developed. Data registries for collecting clinical practice and quality information have been used in health care since 1989 [34]. CMS has also advocated for the use of registry reporting for payment policy decisions (Coverage with Evidence Development) and the Quality Payment Program [35]. In radiology, the ACR National Radiology Data Registry has approximately 4,500 sites with infrastructure for allowing automated data transfer using web-based application programming interfaces (APIs) and data transfer using standards such as Health Level Seven (HL7) and Fast Healthcare Interoperability Resources (FHIR) [36]. Registry reporting is facilitated by structured reporting and the use of CDEs, and to be effective for monitoring the performance of AI algorithms, data collection will need to be as seamless as possible and will require standardized APIs. Additionally, an ability to collect appropriate metadata about the examination, such as examination manufacturer, examination parameters, and patent demographics, will be necessary so that the registry can inform stakeholders about parameters where algorithm performance did not meet expectations. Reports from these registries could be useful in developing the real-world data the FDA is seeking for postmarket surveillance of SaMD (Fig. 6). Creating models for validation and monitoring of AI algorithms and minimizing unintended bias will require collaborations between researchers, industry developers, and government agencies. The medical imaging community should play a leading role in facilitating these collaborations.

OPTIMIZING THE MACHINE-HUMAN INTERFACE: A VISION FOR THE FUTURE PRACTICE OF DIAGNOSTIC IMAGING

The plethora of information potentially available to radiologists, including pathology data, "omics" data, and digital information from the electronic health record, at the time of interpretation is adding increasingly more complexity to interpretation of diagnostic imaging. Integration of digital diagnostic information from multiple sources including medical imaging, pathology, laboratory, genomics, and radiomics is critical for generalizing advanced informatics tools into all facets of health care. This will require breaking down data silos, so

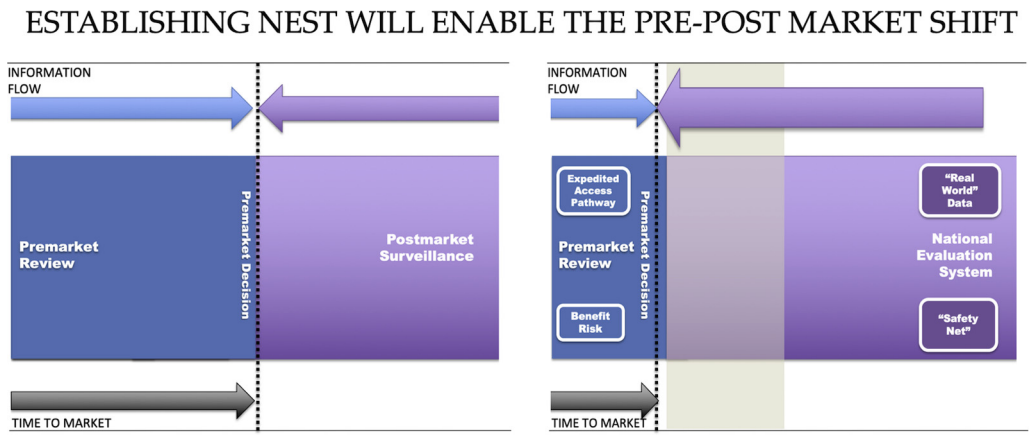

Fig 5. The FDA is creating pathways to make the premarket review process for medical devices more efficient by establishing the Software Precertification Program for premarket review and using real-world data through the National Evaluation System for Healthcare Technologies (NEST) program to make the postmarket surveillance process more robust. (Credit FDA Greg Pappas, MD.)

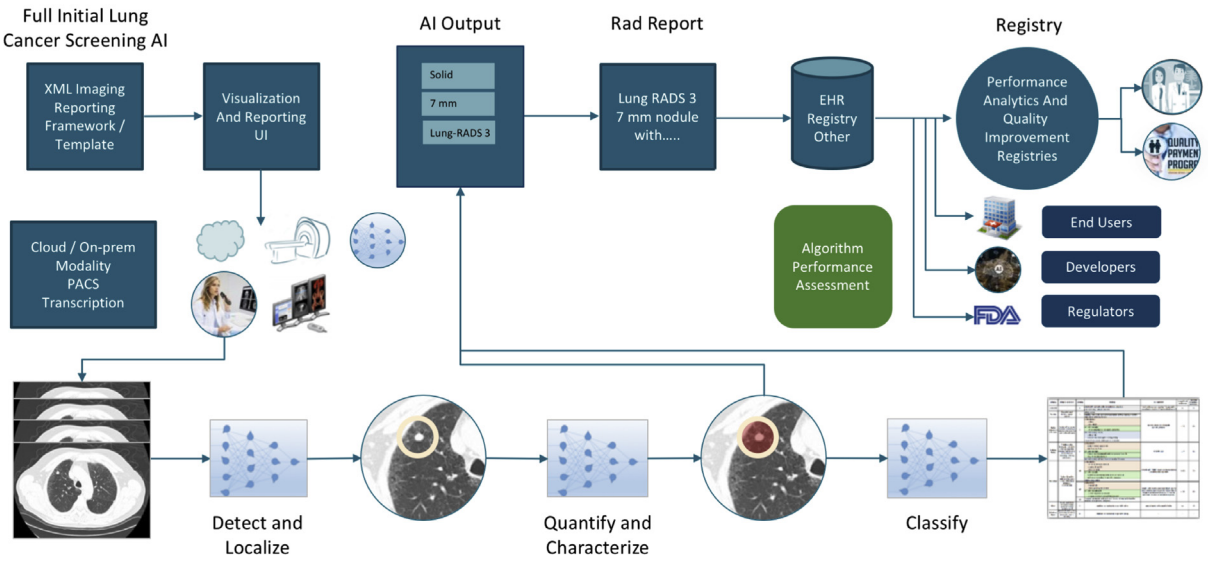

Fig 6. AI in clinical practice with registry reporting for monitoring with real-world data. Data about algorithm performance are collected through a structured reporting system, and metadata about the examination parameters are collected in the background and transferred to various registries including an AI performance monitoring registry. Reports can be provided to all stakeholders including the radiologist, the facility, the algorithm developer, and regulatory agencies. AI = artificial intelligence; EHR = electronic health record; UI = user interface; XML =Extensible Markup Language.

that a coherent integrated resource ecosystem is created, and developing a work environment that brings multiple physician specialties together and combines human intellect and judgment with AI solutions to enhance our ability to bring precision and personalized diagnosis into routine practice. Making data silos available could be accomplished developing APIs tools that allow complete integration or more elegantly by AI algorithms that efficiently data mine the disparate resources for applicable patient-specific information. Catalyzed by the Academy, this human-machine hybrid system has been conceptualized as the "Diagnostic Cockpit of the Future" [37], which is a metaphor for a "future-state digital platform that will aggregate, organize, and simplify medical imaging results and patient-centric clinical data, helping clinicians become the 'diagnostic pilots' of the future in detecting disease early, making accurate diagnoses, driving image-guided interventions, and ultimately improving the downstream clinical management of patients" [37]. This fully integrated system will receive standardized digital inputs from imaging, pathology, "omics," and other clinical data extracted from the electronic health record and house this information in a data storage center available to diagnosticians throughout the patients' clinical care. Quantitative and imaging data from the storage center must be processed and presented to the human observer in a clinically useful way for the production of a diagnostic report. AI solutions are poised to be the tools that will provide quantitative outputs to the radiologists, pathologists, and clinical domain experts of the future through advanced human-machine interfaces.

LIMITATIONS

Although the diversity and expertise of the data science and clinical and regulatory participants involved in the workshop was robust, the workshop participants do not account for the entirety of thought leaders in the field, and others may have conflicting or additional opinions. An additional limitation of the workgroup is the speed that AI in medical imaging is evolving, and as such, the current state is constantly evolving; our recommendations could rapidly become obsolete. What is presented in this road map represents a snapshot of a constantly evolving field, and additional workshops will be needed to assess progress and reassess needs.

Table 1. Research opportunities and infrastructure development requirements for translating AI for medical imaging from research to routine clinical practice

Area	Current State	Knowledge and Infrastructure Gap	Approach and Methods Needed	Comments and Limitations
Software use cases for AI	AI algorithms are being created based on use cases developed at single institutions working with single developers, limiting diversity and generalizability to widespread clinical practice.	Few structured AI use cases are available for AI development; algorithm inputs and outputs are not standardized between similar use cases; there is a lack of a registry of Common Data Elements to inform development of structured AI use cases.	Leverage the value of radiologists and radiology specialty societies to develop and promote widely available structured AI use cases for AI in healthcare; develop and promote the use of common data elements in reporting software and AI development.	No widespread use of structured data in current radiologist reporting practices; incentives are missing for culture change in clinical practice.
Data availability	Algorithms are being developed most often with single-institution data with no evidence that algorithm outputs will be generalizable to routine clinical practice and free of unintended bias.	Infrastructure changes are needed to allow availability of distributed data sets for use in training, testing, and validating AI algorithms.	Create data sets for AI training and testing based on structured use cases so that similar data sets can be created at multiple institutions and used centralized or on premises; ensure geographic, technical, and patient demographic diversity; develop directories of annotated data sets for use by AI developers; create standards for data use agreements to facilitate data sharing among organizations.	Establish incentives that promote sharing, and discourage exclusivity for partnerships between developers and organizations with data and researchers and developers that need data.
Ensuring algorithm safety and efficacy before and after deployment in clinical practice	Algorithm validation using data sets beyond those withheld from the training data are limited; pathways for performance monitoring in clinical practice are limited or nonexistent.	Embargoed data sets with ground truth for algorithm validation need to be developed and used to evaluate AI algorithms before FDA clearance; registries for algorithm performance monitoring are needed.	Develop efficient pathways for AI algorithm validation within the FDA premarket review process; establish robust registries to monitor algorithm performance in clinical practice with the ability to collect technical and demographic metadata related to the examination.	All stakeholders in the AI ecosystem will play a role, including government regulators for establishing efficient means for premarket clearance with reliance on real-world evidence to monitor the performance of AI in routine clinical practice.
Standards for clinical integration of AI algorithms	A number of standards for data transfer and archiving exist, and additional "standards" may not be needed; however, determining which standards will be best for establishing interoperability for AI algorithm integration has not occurred.	Identification and use of the proper standards for AI interoperability have not occurred; no standard APIs exist for integration of AI output data into existing HIT resources such as the PACS and EHR.	Develop recommendations for identifying proper standards for AI interoperability; develop APIs for AI integration that can be used by all researchers and developers to integrate AI algorithm outputs into the clinical environment.	The AI ecosystem community should be involved in promoting efforts to standardize interoperability and work with government agencies, including FDA, NIST, and CMS, to develop pathways and incentives to ensure interoperability of AI.

AI = artificial intelligence; API = application programming interfaces; EHR = electronic health record; HIT = health information technology; NIST = National Institute of Standards and Technology.

CONCLUSION

A summary of challenges and recommendations from the workshop are summarized in Table 1. At this time, fundamental questions regarding translating AI research to routine clinical practice remain unanswered. The road map begins with a needs assessment for what should be built. Developers are currently tasked with having to build costly tools for data extraction, de-identification, labeling, and workflow integration with each developer, potentially creating proprietary assets, which one workshop presenter indicated was akin to "the challenge of mining for gold when you have to start by building a mining pick" [8]. Open-source interoperability standards and resources should be defined and developed for how AI algorithms should be built so that these algorithms can be integrated into HIT resources, and methods should be developed for validating AI algorithms and monitoring their performance in clinical practice. Determining the most pressing clinical needs and then determining which of those needs are amenable to AI solutions will be essential to advancing AI practice. Prioritizing use cases for AI is not just determining whether an algorithm can be built but also determining whether it should be built, which should ideally include a cost-benefit analysis. The medical imaging community should describe exactly what is important to radiology and what we think data scientists, including researchers and developers, could do to improve patient care. Those descriptions should go beyond narratives and flowcharts. Human language should be converted to machine readable language using standardized data elements with specific instructions for standard inputs, relevant clinical guidelines that should be applied, and standard outputs so that inferences can be ingested by downstream HIT resources.

Structured use cases allow training and validation data sets to be built using the same standards, and data sets from multiple institutions can be aggregated for training, either centralized or distributed, to reduce unintended bias and optimize generalizability of algorithms to clinical practice as well as create robust data sets for algorithm validation. Because there are standardized inputs, algorithms may run on the modality itself, on a local server, or in the cloud, and systems that may eventually host thousands of algorithms will need a standard way of accepting inputs for the algorithm to process. Because the outputs of the algorithm are presented in a standardized way, APIs can be developed that will allow AI integration into any system or electronic resource. Finally, structured use cases should have specifications for data that should be collected to inform the developer about the performance of the algorithm in actual clinical use. Understanding performance variances that occur in different patient populations, across different equipment manufacturers, or using different acquisition protocols can then be used to refine the algorithm, modify the use case specifications, or inform regulatory agencies. Implementation strategies should also promote payment models that promote health equity so that improvements to care afforded by AI applications will be available to all patients regardless of their socioeconomic status or the resources of their health care facilities. The future for AI applications for improved diagnosis in general and for image-based diagnosis is enormous. The opportunities and challenges summarized here can serve as a guidepost and road map for future development.

TAKE-HOME POINTS

- Translating foundational research in AI for medical imaging to routine clinical practice is dependent on defining the clinical need for AI tools, establishing methods for data sharing while protecting personal health information, ensuring AI algorithms will be safe and effective in routine clinical practice, and ensuring standards for implementing AI tools into routine clinical workflows.
- An active AI ecosystem in which radiologists, their professional societies, researchers, developers, and government regulatory bodies can collaborate, contribute, and promote AI in clinical practice will be key to translating foundational AI research to clinical practice.
- AI tools will be a critical driver of a future practice state in which AI tools have provided radiologists and other diagnosticians access to a wealth of information that will inform more accurate diagnoses and identify patients at risk for significant illness.

ACKNOWLEDGMENTS

The authors thank all the presenters and participants at the NIH Workshop on Artificial Intelligence in Medical Imaging. We also are indebted to Maryellen Giger, Paul Kinahan, Greg Pappas, and Cynthia McCullough for their comments on earlier versions of the manuscript.

REFERENCES

1. Chockley K, Emanuel E. The end of radiology? Three threats to the future practice of radiology. J Am Coll Radiol 2016;13:1415-20.

2. Remnick D. Obama reckons with a Trump presidency. The New Yorker 2016. Nov 28;28.
3. Hinton G. Geoff Hinton on radiology. Machine Learning and Market for Intelligence Conference, Creative Disruption Lab Toronto, Canada. Available at: https://www.youtube.com/watch?v=2HMPRXstSvQ. Published November 24, 2016. Accessed May 1, 2019.
4. Allen B, Dreyer K. The artificial intelligence ecosystem for the radiological sciences: ideas to clinical practice. J Am Coll Radiol 2018;15:1455-7.
5. Thrall JH, Li X, Li Q, Cruz C, Do S, Dreyer K, Brink J. Artificial intelligence and machine learning in radiology: opportunities, challenges, pitfalls, and criteria for success. J Am Coll Radiol 2018;15:504-8.
6. Perspectives on Research in Artificial Intelligence and Artificial General Intelligence Relevant to DoD (Unclassified); JASON: JSR-16-Task-003; January 2017. Available at: https://fas.org/irp/agency/dod/jason/ai-dod.pdf. Accessed May 1, 2019.
7. Herper M. MD Anderson benches IBM Watson in setback for artificial intelligence in medicine. Forbes. Zugriff im Juli; 2017. Feb.
8. Langlotz CP, Allen B, Erickson BJ, Kalpathy-Cramer J, Bigelow K, Cook TS, Flanders AE, Lungren MP, Mendelson DS, Rudie JD, Wang G. A roadmap for foundational research on artificial intelligence in medical imaging: From the 2018 NIH/RSNA/ACR/The Academy Workshop. Radiology. 2019 Apr 16:190613.
9. US Department of Health and Human Services. NIH videocasting and podcasting. Artificial intelligence in medical imaging (day 1). Available at: https://videocast.nih.gov/summary.asp?live=28131&bhcp=1. Accessed April 14, 2019.
10. US Department of Health and Human Services. NIH videocasting and podcasting. Artificial intelligence in medical imaging (day 2). Available at: https://videocast.nih.gov/summary.asp?Live=28180&bhcp=1. Accessed April 14, 2019.
11. Krauth O. How 68% of consumers are already using AI without knowing it. TechRepublic. Available at: https://www.techrepublic.com/article/how-68-of-consumers-are-already-using-ai-without-knowing-it/. Accessed May 1, 2019.
12. Ranschaert ER, Morozov S, Algra PR, eds. Artificial Intelligence in Medical Imaging: Opportunities, Applications and Risks. Berlin, Germany: Springer Nature; 2019.
13. Kim DW, Jang HY, Kim KW, Shin Y, Park SH. Design characteristics of studies reporting the performance of artificial intelligence algorithms for diagnostic analysis of medical images: results from recently published papers. Korean J Radiol 2019;20:405-10.
14. RadElement.org. Common Data Elements (CDEs) for Radiology. Available at: http://radelement.org. Accessed April 14, 2019.
15. RadElement.org. Sets. Available at: http://www.radelement.org/set/all.php. Accessed April 14, 2019.
16. ACR Informatics. ACRassist. Available at: http://www.acrinformatics.org/acr-assist. Accessed April 14, 2019.
17. Wilkinson MD, Dumontier M, Aalbersberg IJ, et al. The FAIR Guiding Principles for scientific data management and stewardship. Scientific Data 2016;3:160018. https://doi.org/10.1038/sdata.2016.18.
18. HealthIT.gov. Artificial intelligence for health and health care. Available at: https://www.healthit.gov/sites/default/files/jsr-17-task-002_aiforhealthandhealthcare12122017.pdf. Accessed May 1, 2019.
19. Kobie N. AI has no place in the NHS if patient privacy isn't assured. Wired. Available at: http://www.wired.co.uk/article/ai-healthcare-gp-deepmind-privacy-problems. Accessed May 1, 2019.
20. Iyengar A, Kundu A, Pallis G. Healthcare Informatics and Privacy. IEEE Internet Comput 2018;22:29-31.
21. Kohli M, Prevedello LM, Filice RW, Geis JR. Implementing machine learning in radiology practice and research. AJR Am J Roentgenol 2017;208:754-60.
22. NIST. Standards & measurements. Available at: https://www.nist.gov/services-resources/standards-and-measurements. Accessed April 14, 2019.
23. Tang Y, Harrison AP, Bagheri M, Xiao J, Summers RM. Semi-automatic recist labeling on ct scans with cascaded convolutional neural networks. In: International Conference on Medical Image Computing and Computer-Assisted Intervention. Springer, Cham, Switzerland: Springer; 2018:405-13.
24. US Department of health and Human Services. Medical devices. Available at: https://www.fda.gov/MedicalDevices/default.htm. Accessed May 1, 2019.
25. US FDA. What we do. Available at: https://www.fda.gov/AboutFDA/WhatWeDo. Accessed April 14, 2019.
26. US FDA. Computer-Assisted Detection Devices Applied to Radiology Images and Radiology Device Data - Premarket Notification [510(k)] Submissions - Guidance for Industry and Food and Drug Administration Staff. Available at: https://www.fda.gov/MedicalDevices/DeviceRegulationandGuidance/GuidanceDocuments/ucm187249.htm. Accessed April 14, 2019.
27. US FDA. International medical device regulators forum (IMDRF). Available at: https://www.fda.gov/MedicalDevices/InternationalPrograms/IMDRF/default.htm. Accessed May 1, 2019.
28. US FDA. Software as a medical device (SaMD). Available at: https://www.fda.gov/MedicalDevices/DigitalHealth/SoftwareasaMedicalDevice/default.htm. Accessed April 14, 2019.
29. US FDA. National evaluation system for health technology (NEST). Available at: https://www.fda.gov/aboutfda/centersoffices/officeofmedicalproductsandtobacco/cdrh/cdrhreports/ucm301912.htm. Accessed April 14, 2019.
30. NEST Coordinating Center. Demonstration projects. Available at: https://nestcc.org/demonstration-projects/. Accessed April 14, 2019.
31. ACR. FDA NEST Program Names ACR DSI Use Case as Demo Project. Available at: https://www.acr.org/Media-Center/ACR-News-Releases/2018/FDA-NEST-Program-Names-ACR-DSI-Use-Case-as-Demo-Project. Accessed May 1, 2019.
32. US FDA. Digital health software precertification (pre-cert) program. Available at: https://www.fda.gov/MedicalDevices/DigitalHealth/DigitalHealthPreCertProgram/default.htm. Accessed May 1, 2019.
33. Medical Device Epidemiology Network. Available at: http://mdepinet.org. Accessed April 14, 2019.
34. National Radiology Data Registry. Available at: https://nrdr.acr.org/Portal/Nrdr/Main/page.aspx. Accessed April 14, 2019.
35. CMS Quality Payment Program. Available at: https://qpp-cm-prod-content.s3.amazonaws.com/uploads/258/2019%20QPP%20Final%20Rule%20Fact%20Sheet_Update_2019%2001%2003.pdf. Accessed April 14, 2019.
36. HL7 FHIR. FHIR overview. Available at: https://www.hl7.org/fhir/overview.html. Accessed April 14, 2019.
37. Krupinski E, Bronkalla M, Folio L, Keller B, Mather R, Seltzer S, Schnall M, Cruea R. Advancing the diagnostic cockpit of the future: An opportunity to improve diagnostic accuracy and efficiency. Academic Radiology 2019;26:579-81.