Integrating AI Algorithms into the Clinical Workflow

Krishna Juluru, MD • *Hao-Hsin Shih, MS* • *Krishna Nand Keshava Murthy, MS* • *Pierre Elnajjar, MS* •
Amin El-Rowmeim, MS • *Christopher Roth, MD* • *Brad Genereaux, BS* • *Josef Fox, MD* • *Eliot Siegel, MD* •
Daniel L. Rubin, MD

From the Department of Radiology, Memorial Sloan-Kettering Cancer Center, 1275 York Ave, Box 29, New York, NY 10065 (K.J., H.H.S., K.N.K.M., P.E., A.E.R., J.F.); Department of Radiology, Duke University Medical Center, Durham, NC (C.R.); NVIDIA, Santa Clara, Calif (B.G.); Department of Diagnostic Radiology and Nuclear Medicine, University of Maryland School of Medicine, Baltimore, Md (E.S.); and Department of Radiology, Stanford University, Stanford, Calif (D.L.R.). Received January 12, 2021; revision requested March 3; revision received June 16; accepted July 14. **Address correspondence to** K.J. (e-mail: *juluruk@mskcc.org*).

Conflicts of interest are listed at the end of this article.

Integration of artificial intelligence (AI) applications within clinical workflows is an important step for leveraging developed AI algorithms. In this report, generalizable components for deploying AI systems into clinical practice are described that were implemented in a clinical pilot study using lymphoscintigraphy examinations as a prospective use case (July 1, 2019–October 31, 2020). Deployment of the AI algorithm consisted of seven software components, as follows: *(a)* image delivery, *(b)* quality control, *(c)* a results database, *(d)* results processing, *(e)* results presentation and delivery, *(f)* error correction, and *(g)* a dashboard for performance monitoring. A total of 14 users used the system (faculty radiologists and trainees) to assess the degree of satisfaction with the components and overall workflow. Analyses included the assessment of the number of examinations processed, error rates, and corrections. The AI system processed 1748 lymphoscintigraphy examinations. The system enabled radiologists to correct 146 AI results, generating real-time corrections to the radiology report. All AI results and corrections were successfully stored in a database for downstream use by the various integration components. A dashboard allowed monitoring of the AI system performance in real time. All 14 survey respondents "somewhat agreed" or "strongly agreed" that the AI system was well integrated into the clinical workflow. In all, a framework of processes and components for integrating AI algorithms into clinical workflows was developed. The implementation described could be helpful for assessing and monitoring AI performance in clinical practice.

© RSNA, 2021

rtificial intelligence (AI) applications are increasingly being developed for diagnostic imaging (1). These AI applications can be divided broadly into two categories: first, those pertaining to logistic workflows, including order scheduling, patient screening, radiologist reporting, and other operational analytics (also termed *upstream* AI); and second, those pertaining to the acquired imaging data themselves, such as automated detection and segmentation of findings or features, automated interpretation of findings, and image post-processing (2) (also termed *downstream* AI). Numerous downstream AI applications have been developed in recent years. More than 120 AI applications in medical imaging are currently cleared by the U.S. Food and Drug Administration (3).

Although there are a variety of applications available, a major unaddressed issue is the difficulty of adopting AI algorithms into the workflow of clinical practice. AI algorithms are generally siloed systems that are not easily incorporated into existing information systems in a radiology department. Additionally, tools to measure and monitor the performance of AI systems within clinical workflows are lacking.

We sought to define the requirements for effective AI deployment in the clinical workflow by considering an exemplar downstream AI application—automated interpretation and reporting of lymphoscintigraphy examinations—and to use that exemplar to develop generalizable components to meet the defined requirements.

Materials and Methods

The institutional review board approved this retrospective study for development of the AI algorithm, which was compliant with the Health Insurance Portability and Accountability Act, and waived requirements for written informed consent.

Understanding the General Workflow and Particular Use Case

Our use case for deploying AI within the clinical workflow was an AI algorithm for evaluating lymphoscintigraphy examinations. These examinations are performed to identify sentinel lymph nodes (SLNs) in patients with invasive breast cancer, which potentially increases the accuracy of staging (4). The examination comprises images of the breasts and axillae (Fig 1), and the radiology report describes the location and positivity of SLN. The AI algorithm we developed for this use case takes the images as inputs and outputs the following data for reporting: *(a)* observed sites of injection (right breast only, left breast only, bilateral breasts), *(b)* probability of radiotracer accumulation in the axillae (probability scores for none, right, left, or bilateral axillae), *(c)* number of right axillary lymph nodes (integer), and *(d)* number of left axillary lymph nodes (integer).

Assessment of Needs for Integrating AI Algorithms

We identified the following needs for integrating an AI algorithm into the clinical workflow. The system should:

Summary

This report details requirements and an architecture for deploying artificial intelligence algorithms into the clinical workflow; the implementation of the software components described can be used to inform development of standards-based solutions.

Key Points

- Integrating artificial intelligence (AI) algorithms into clinical workflows requires a variety of components for operationalization, performance monitoring, and mechanisms for continual improvement.
- The key elements of the described AI workflow include software components for *(a)* image delivery, *(b)* quality control, *(c)* a results database, *(d)* results processing, *(e)* results presentation and delivery, *(f)* results error correction, and *(g)* a dashboard for performance monitoring.
- The software components were implemented to deploy an AI algorithm to assist in reporting lymphoscintigraphy examinations as a specific use case to demonstrate the potential value of the approach.

1. Commence the image analysis immediately after completion of image acquisition.
2. Identify and report examination quality problems.
3. Generate and send preliminary reports to the dictation system for radiologist review and signature.
4. Allow users to correct AI results.
5. Provide a dashboard to allow users to monitor system performance.

Implementation Components

Using the needs assessment as a foundation, we developed seven generalizable software components for the integration of an AI algorithm into the clinical workflow, using lymphoscintigraphy AI as a use case (Fig 2). The components include the following: *(a)* an examination orchestrator, *(b)* quality control, *(c)* a results database, *(d)* results processing, *(e)* presentation and delivery, *(f)* error correction, and *(g)* a dashboard.

All databases needed to support these components were implemented on SQL Server 2012 SP4 (Microsoft). The components were implemented as operating system services that ran continually in the background, monitoring for changes. The AI algorithm itself was implemented as a web service, a common approach for most commercial AI algorithms. All software ran on Microsoft Windows Server 2012 R2. All programming was done using C# (version 5.0; Microsoft) and Python (version 3.5; *python.org*) languages.

Although the implementation of these components within our institution is specific to our particular commercial clinical systems, we describe the design of each of them, which we believe will enable similar implementations at other institutions.

Examination orchestrator.— The examination orchestrator component monitors the radiology information system (RIS) (RIS-IC version 7.0; GE Healthcare) for lymphoscintigraphy examination orderables and their statuses. The Windows service orchestrator was designed to run on a continuous basis on a Windows 2012R2 VM server residing in a data center. A status of "complete" in RIS triggers the examination orchestrator to send a copy of the images in the examination from the picture archiving and communication system (PACS) (Centricity; GE Healthcare) to the lymphoscintigraphy AI, thereby meeting needs assessment 1. The service extracts the images from the PACS by leveraging the Digital Imaging and Communications in Medicine (DICOM) query and retrieve process and then sends the related image(s) to the AI via an application programming interface (API) call. The component also sends specified examination information to the results database, copying the necessary data from the RIS database.

Quality control.— The quality control component is a means for notifying the clinical team of issues arising in the workflow or in the quality of imaging, by e-mail or paging system, thereby meeting needs assessment 2. To implement this component, we used a free, open-source ticketing system (osTicket; Enhancesoft) that was previously implemented in a quality improvement effort (5). When a quality control event is detected, the component creates a new ticket through an API call to the ticketing system. The rules engine within the ticketing system can be configured to send an e-mail to a quality control team member assigned to the particular examination with the issue.

The quality issues that this component is meant to identify are different from AI algorithm performance issues and can occur when the AI is performing normally. We configured this component to detect the event that the wrong breast was injected (Table 1). The name of the examination order is where the requesting surgical oncologist can specify the site of injection as right breast, left breast, or bilateral breasts, for example, "Lymphoscintigraphy right breast." Meanwhile, the lymphoscintigraphy AI algorithm detects the laterality of the injection on the image (Fig 1) and passes this information to the results database. The quality control component compares the AI laterality result with the site requested in the examination order, and if there is a mismatch, a rule in the quality control component will activate, which then sends an API call to the ticketing system, resulting in an e-mail notification to a quality control team member. This all occurs within seconds of detection of an issue.

Results database.— The results database is a component that stores information about each radiology examination identified by the examination orchestrator for AI processing, the results generated by the AI algorithm, and the corrections from the error correction component. It serves as the source of data for the quality control component, the results processing component, and the system performance dashboard. Our database was implemented on SQL Server 2012 SP4, which was available
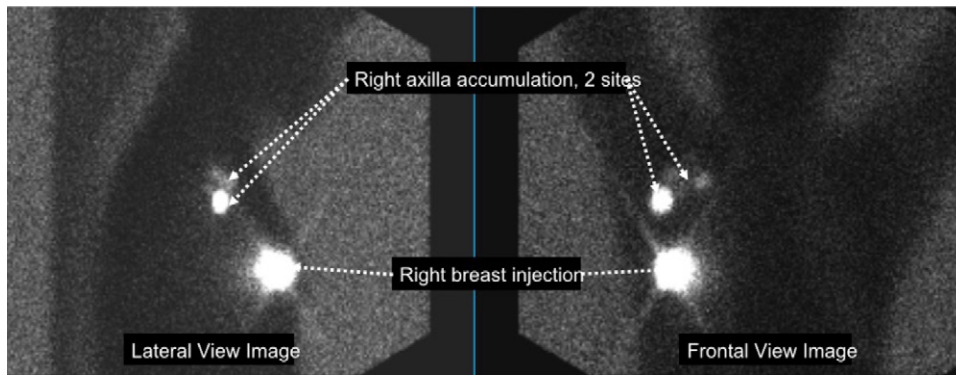
Figure 1: Images obtained from a lymphoscintigraphy examination. Following the injection of a radioactive tracer into one or both breasts, images are obtained with a gamma camera in anterior and lateral projections. Bright areas are sites of injection or radiotracer accumulation in lymph nodes where breast cancer may have spread. In the current example, there was an injection to the right breast, with two sites of accumulation in the right axilla.

Integrating AI algorithms into Clincal Workflow
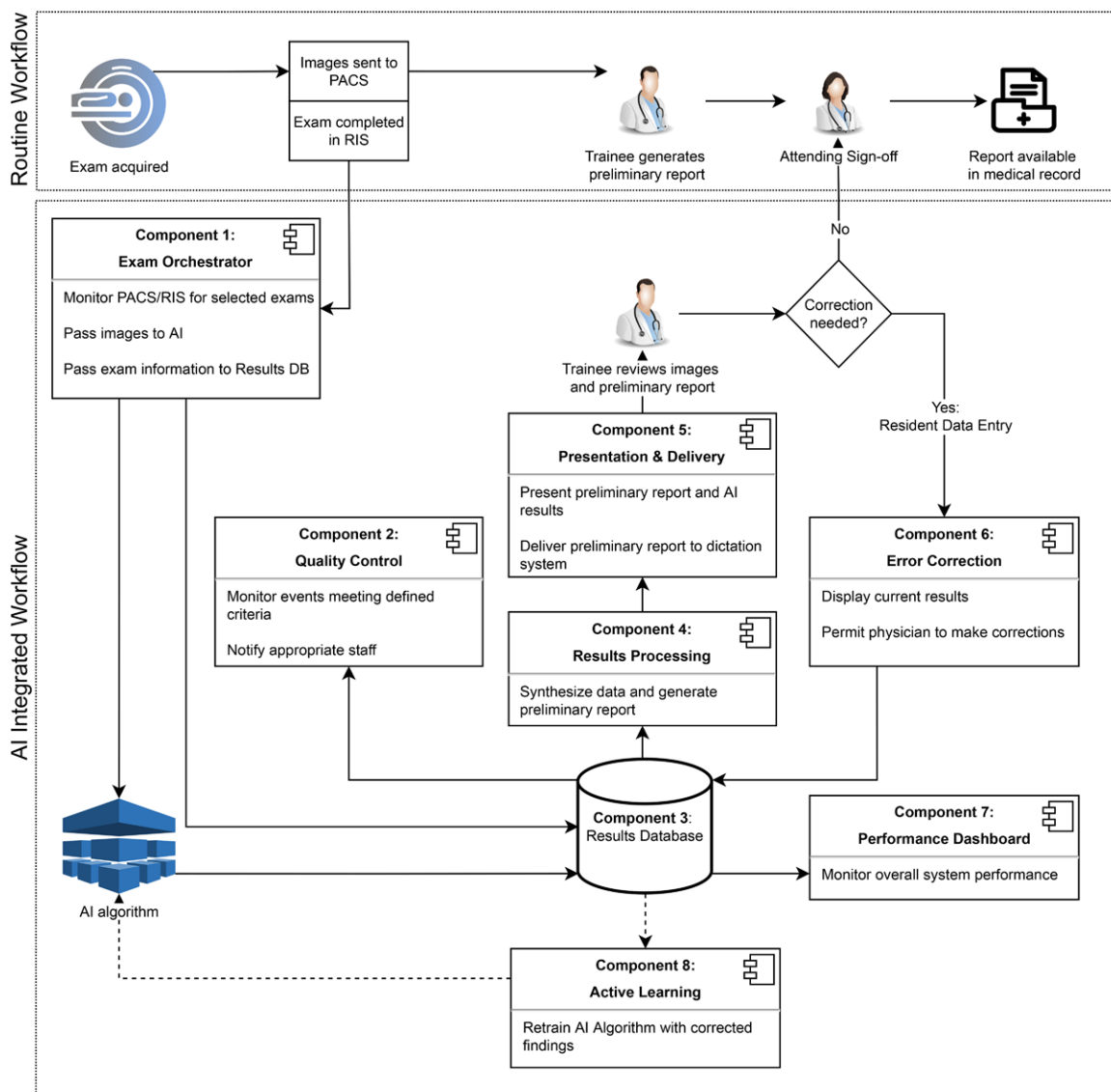
Figure 2: Clinical workflow diagram. Routine clinical workflow is compared with a workflow that includes an image analysis algorithm that performs a preliminary image analysis. Seven software components are described that are necessary for integration into a clinical environment. Component 8 was not built in this study but is included for future integration reference. AI = artificial intelligence, DB = database, PACS = picture archiving and communication system, RIS = radiology information system.

Table 1: Rules of Quality Control Component Established for Lymphoscintigraphy Use Case

Ordered Examination Name	Condition That Will Launch Quality Control Component
Lymphoscintigraphy right breast	AI {injection site} < > right
Lymphoscintigraphy left breast	AI {injection site} < > left
Lymphoscintigraphy bilateral breasts	AI (injection site) < > bilateral

Table 2: Fields Storing Lymphoscintigraphy Data for Results Database Component

Field	Source	Possible Value
Ordered examination accession number	Examination orchestrator	Integer
Ordered examination name	Examination orchestrator	Text
Injection site	AI	Right, left, bilateral
Probability of radiotracer accumulation in no axilla*	AI	Value between 0 and 1
Probability of radiotracer accumulation in right axilla*	AI	Value between 0 and 1
Probability of radiotracer accumulation in left axilla*	AI	Value between 0 and 1
Probability of radiotracer accumulation in bilateral axillae*	AI	Value between 0 and 1
No. of right axillary lymph nodes	AI	Integer
No. of left axillary lymph nodes	AI	Integer
Axillary sites	Reporting component (calculated result from probability scores)	None, right, left, bilateral
Corrected injection site	Error correction component	Right, left, bilateral
Corrected axillary sites	Error correction component	None, right, left, bilateral
Corrected no. of right axillary lymph nodes	Error correction component	Integer
Corrected no. of left axillary lymph nodes	Error correction component	Integer

Note.—AI = artificial intelligence.
*Sum of all probabilities = 1.

through an institutional site license. The fields, information sources, and possible values of items in the results database are summarized in Table 2.

Results processing.— The results processing component generates a human-friendly preliminary report by compiling and processing AI results recorded in the results database. In our use case, the results database compiles the AI-detected sites of injection, sites of axillary radiotracer accumulation, and the number of lymph nodes in each axilla. The results database also performs a processing step that is necessary to translate a probability provided by the AI to a correct accumulation site to be reported. This was done by configuring this component to report the site of highest probability. In the example provided in Table 3, the output of the lymphoscintigraphy AI included probabilities of radiotracer accumulation in no axilla, right axilla, left axilla, and bilateral axillae of .08, .88, .02, and .02, respectively. The results processing component uses the result of highest probability (right axilla) as the result to send to downstream components.

Presentation and delivery.— The presentation and delivery component delivers the compiled report generated by the

results processing component to the reporting system for review by the radiologist, thereby meeting needs assessment 3. An API available in our dictation system (PowerScribe 360, Reporting v4.0 SP1; Nuance Communications) delivers the compiled report text into the custom fields. The custom fields are then inserted into reporting templates known as *autotexts* that are set to launch by default for lymphoscintigraphy examinations. The radiologist identifies lymphoscintigraphy examinations in the PACS worklist per routine. With the launch of the dictation system in context with the examination in PACS, he or she sees the preliminary report ready for review and signature (Table 3) within the very same dictation system. The radiologist's next step is to ensure agreement between the imaging examination in PACS and the report. If agreement exists, the radiologist only needs to sign the report. If, however, corrections are needed, the radiologist uses the error correction component.

Error correction.— The error correction component tracks corrections that the radiologist makes to the preliminary report produced by the AI, thereby satisfying needs assessment 4. Although the radiologist can correct errors in the preliminary

Table 3: Sample Lymphoscintigraphy Report

Final Report Text Sent to Medical Record	Comment
CLINICAL STATEMENT: Breast cancer; referred for sentinel lymph node mapping TECHNIQUE: Radiopharmaceutical: Right breast 0.102 mCi Tc-99m sulfur colloid (unfiltered), 6 o'clock subareolar Field of view: Anterior and lateral images of the chest with transmission images for anatomic localization COMPARISON: None CORRELATION: None	Standard text, part of dictation system default template
FINDINGS: Radiotracer accumulation is seen in the right breast Sites of radiotracer accumulation outside of the breast: -Right axilla: 2 IMPRESSION: Sentinel lymph nodes in the right axilla	All text in this section (highlighted) is generated by the results processing component; in this example, the AI results were: Probability of accumulation in no axilla: 0.08 Probability of accumulation in right axilla: 0.88 Probability of accumulation in left axilla: 0.02 Probability of accumulation in bilateral axillae: 0.02 No. of right axillary lymph nodes: 2 No. of left axillary lymph nodes: 0 The results processing component selected the highest probability result (right axilla) to include in the report; it also generated the full impression The presentation and delivery component delivered all text in this section into the dictation system custom field

Note.—The "findings" and "impression" texts are automatically generated by the results processing component and delivered to a custom field in the dictation system. The custom field is part of a template that is set to automatically launch for the lymphoscintigraphy examinations. The net effect is one where at the moment the radiologist opens a lymphoscintigraphy examination for dictation, he or she will see the entire report ready for review and signature. The radiologist's next step is to ensure agreement between the imaging examination in the picture archiving and communication system and the report. If agreement exists, the radiologist only needs to sign the report. Tc-99m = technetium 99-m.

report by directly editing the text of the report in the dictation system, this does not capture disagreements with the AI output in a structured manner. Hence, we developed a web application to display current AI results (read in structured manner from the results database) with options to correct these results (Fig 3) and to also send the corrected data back to the results database as a new entry. The error correction component captures not only the existence of the correction, but also the specific correction made (for instance, if the radiologist records corrections to the number of SLNs identified by the AI). When corrections are completed, the results processing and presentation and delivery components are automatically reinitiated, generating a new report in the dictation system for the radiologist to rereview.

Dashboard to monitor AI performance.— The results database component contains all AI processing events, results, and corrections over time. Dashboards enable us to view the data in near real time and answer questions such as, "How often is a radiologist correcting the results of the AI algorithm?" There are a variety of dashboarding and business intelligence tools on the market that enable data visualization (6). Our institution has an enterprise-level subscription to one such tool known as Tableau (Tableau Server Version: 2019.1.1; Tableau Software) that has been applied to health care data visualiza-

tion challenges (7), thereby meeting needs assessment 5. We used Tableau to generate real-time metrics of data available in the results database, including quantifying discrepancies between AI results and corrected results (Fig 4). Dashboards were made available to our departmental radiology quality assurance teams for regular review.

Deployment and Data Analysis

Twenty attending faculty radiologists and radiology trainees were educated on using the system. On June 26, 2019, the system was turned on to process examinations and deliver reports prospectively. The analysis for this study was performed in November 2020 using data from a study period defined to be July 1, 2019, to October 31, 2020.

The performance of the examination orchestrator component was assessed by comparing the number of examinations processed by the AI system during the study period (available in the results database) to the number of examinations truly performed (available through an RIS query).

Our institution maintained a patient risk–reporting system that all staff were required to use to report events that may have caused patient risk. A laterality error, one in which a procedure is supposed to be performed on one side of the body but occurs on the other side, constitutes such a reportable event. The performance of the quality control component was assessed by

comparing the number of times it was activated to the number of these events reported in the institutional risk-reporting system during the study period.

To assess satisfaction with all AI system components and overall workflow, a survey was distributed in November 2020 to all faculty and trainees who had interacted with the AI system during the study period. This survey asked respondents to rate satisfaction on a five-point Likert scale (strongly agree, somewhat agree, neutral, somewhat disagree, strongly disagree, or not available). The survey included the following statements:

1. AI-generated reports were ready for review in the dictation system by the time I opened the exam.

2. AI-generated reports were well-formatted and consistent.

3. Making changes to AI-generated images was easy using the error correction component.

4. The AI system saved me time.

5. The AI system helped reduce errors.

6. Overall, the AI system was well integrated into clinical workflow.

7. Any AI system in radiology needs to be well integrated into clinical workflow for me to use it.

Although this study is about components needed for AI integration into clinical workflow and not specifically about an AI algorithm, we nonetheless also assessed algorithm performance metrics to demonstrate accessibility to this data from the integration components. Algorithm and system performance were assessed using the following metrics obtained from the results database:

• Total accuracy is true if no corrections were made to any field in the results database, otherwise, false.

• Injection accuracy is true if no correction was made to the "injection site" field, otherwise, false.

• Axillary site accuracy is true if no correction was made to the "axillary site" field, otherwise, false.

• Right axillary lymph node count error = (corrected number of right axillary lymph nodes following radiologist correction) − (number of right axillary lymph nodes originally identified by the AI).

• Left axillary lymph node count error = (corrected number of left axillary lymph nodes following radiologist correction) − (number of left axillary lymph nodes originally identified by the AI).

Results

In the 15-month study period, our nuclear medicine service performed 1782 lymphoscintigraphy examinations, of which 1748 were successfully processed by system components (processing rate, 98%). There were 26 examinations that the examination orchestrator did not send to the AI because the images available in the PACS were "secondary" screen captures, not the "primary" images needed by the AI that the examination orchestrator was configured to find and send. Eight examinations failed to process because of unexpected image artifacts. There were zero laterality mismatches reported in the institutional risk-reporting system. Likewise, the quality control component was activated zero times.


Figure 3: Error correction component–generated form that allows radiologists to provide corrections to report information (component 6 in Fig 2). The radiologist launches the form from the picture archiving and communication system. After clicking "submit," the data are recorded in the database, and the preliminary report is re-created and delivered to the dictation system by another component in the artificial intelligence system.

From an RIS query, we determined that a total of 20 faculty and trainees had generated lymphoscintigraphy reports during the study period, and all received requests to complete the anonymous survey. Five faculty and nine trainees responded. Survey results are shown in Figure 5, demonstrating an overall favorable opinion of the system components and overall workflow. Five respondents additionally provided free-text anonymous comments. The four comments from trainees were: "Great asset to the radiology department," "Great improvement to before," "Great Job," and "It provides a tremendous help with workload." One faculty member commented, "Let's work on more such solutions."

System performance metrics were readily available on the Tableau dashboard. Of the 1748 examinations, overall report accuracy was true for 1602 examinations (accuracy rate, 92%). In the remaining 146 examinations, corrections were successfully completed using the error correction component. Injection accuracy was true for 1748 of 1748 examinations (accuracy rate, 100%). Axillary accumulation site accuracy was true for 1724 of 1748 examinations (accuracy rate, 99%).

Discussion

The purpose of this study was to understand, develop, and test software components needed to incorporate the output of an AI algorithm into clinical workflow in an academic medical center. While our efforts were meant to address downstream AI needs, a similar process and methods can be applied to meet the differing needs of upstream AI solutions. We identified seven key components for a solution that would deliver AI results to our radiologists quickly, automate certain quality control processes, allow radiologists to document errors in
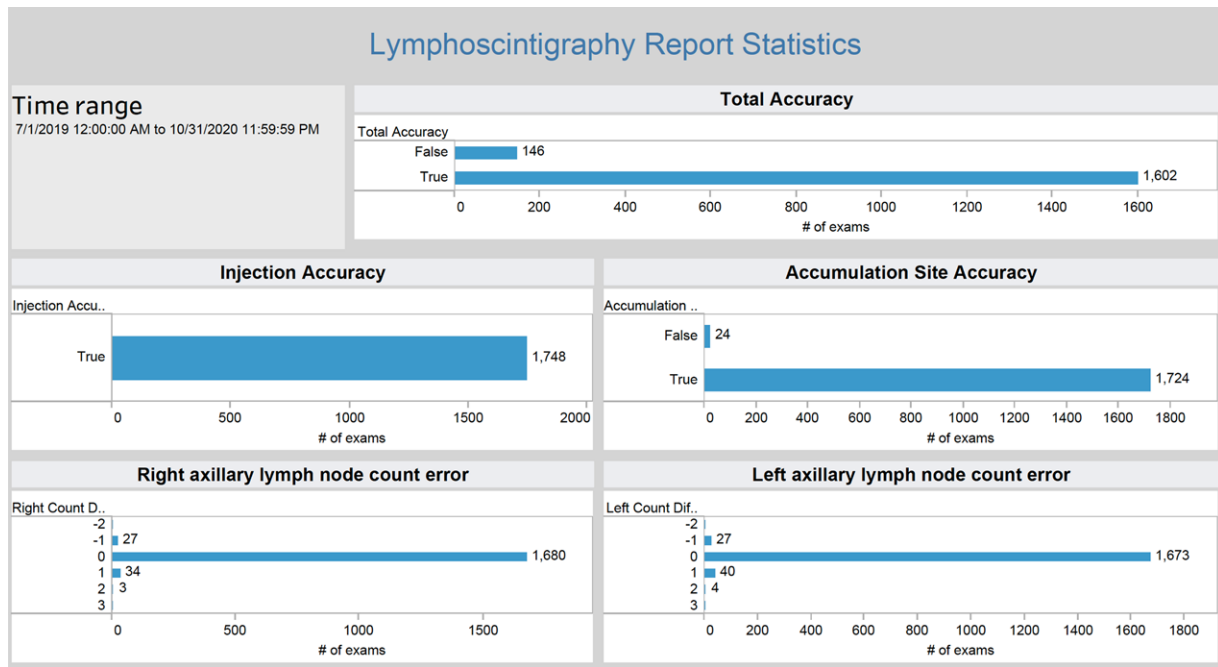
Figure 4: The dashboard specific for the lymphoscintigraphy image analysis algorithm allows close monitoring of algorithm performance (component 7 in Fig 2). The dashboard draws data in near real time from the system database and is accessible from a web browser.
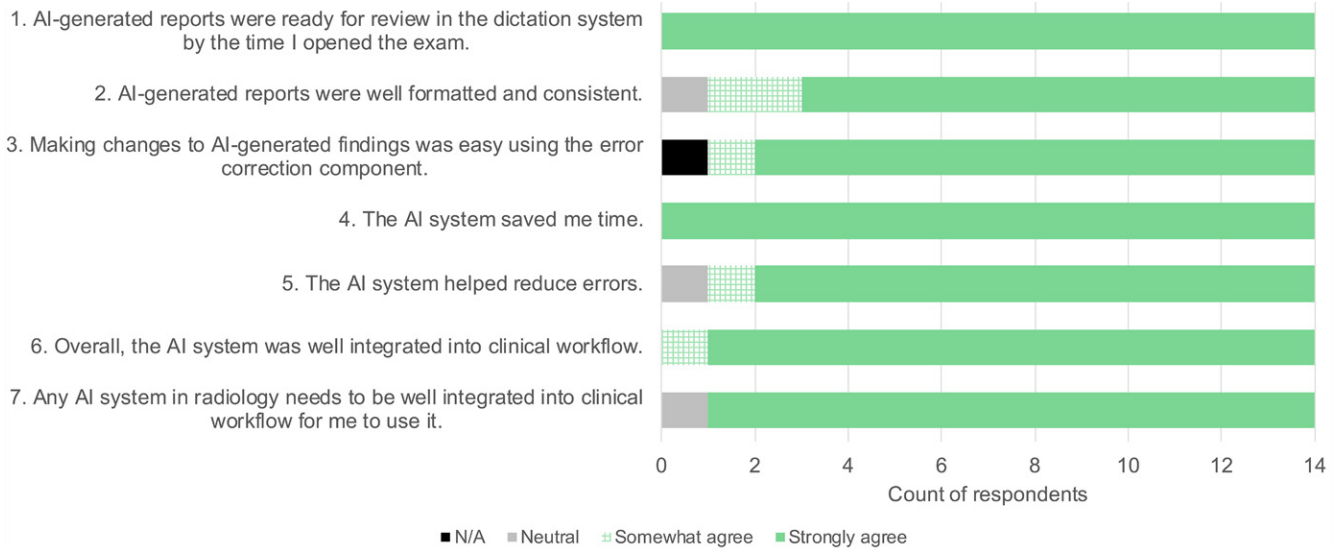
Figure 5: Survey results. Twenty faculty members and trainees received a request to complete a survey assessing satisfaction with artificial intelligence (AI) components and overall workflow. Fourteen responded. N/A = not available.

AI results, and permit ongoing monitoring of AI performance. These components allowed us to create a system that provides an end-to-end solution that met the requirements. In a successful pilot deployment, the system processed 1748 examinations over a 15-month period.

A structured process of performing a needs assessment, designing a workflow, implementing a solution, testing that solution, monitoring results in real-time, and gathering user feedback is supported as a means of building products that meet user needs (8). Several radiology AI marketplaces have been introduced in recent years that attempt to integrate a variety of disparate AI solutions into integrated frameworks (9). In

following a structured process, we identified necessary components that appear to be deficient in these commercial solutions. For example, tools that identify mismatches between the examination ordered and AI interpretation of the examination performed (part of a broader quality control component) are lacking in current commercial solutions, along with mechanisms to notify appropriate radiology staff when these issues arise. The functionalities of our reporting, presentation and delivery, and error correction components are similarly deficient in many commercial solutions. This implementation study can provide guidance to commercial vendors on specific needs and a use case on how they were implemented.

Although the implementation of the software components is customized to enable integration with the commercial systems adopted by our institution, we have described the design of each component, and we believe similar components can be implemented within other institutions. Moreover, although we built these components to support a single specific AI use case, we believe our approach is extensible to support other AI algorithms, including commercial applications. The examination orchestrator component, for instance, performs the critical role of identifying and delivering completed examinations to the appropriate AI. This component could be readily modified to support environments that use multiple AI solutions that may need to be run in a predefined sequence (eg, images delivered to one AI application and then to another). An orchestrator may also have value in managing AI usage, triggering algorithms on specified examinations at times when subspecialist expertise is not available.

To adapt our component to other workflows, changes would be needed in the data fields that capture the outputs of the AI algorithms and mappings to the results database (Table 1). The results processing and quality control components would also need to be appropriately configured to the algorithm being used and the desired needs of the department and interpreting radiologists. In the current study, for instance, the quality control component was configured to detect only a laterality error. A modification of the quality control component can be used with other algorithms that are able to detect, for instance, the presence of image artifacts (such as motion or MRI susceptibility), giving the opportunity to reimage a patient before they leave the department.

Our system failed to process 2% of examinations. The predominant reason for the failure was that technologists did not send the primary images. Instead, they sent secondary images (images that were the result of processing the primary images by another application). While the secondary images can still be interpreted by radiologists, only the primary images can be interpreted by the AI. A second reason for the failure was unexpected image artifacts. For all 34 examinations that could not be processed, the radiologists generated the reports manually with no further issue to patient care. However, as AI becomes more widely deployed, solutions are needed to detect and correct these failures as early as possible, and policies are needed to address how departmental operations should continue when an AI system malfunctions. A system performance dashboard similar to the one we implemented should play a key role in the overall strategy to monitor AI solutions. We made our dashboard available to our departmental quality assurance team, but policies for regular review of these data are still being established. Other institutions that are preparing an AI deployment strategy should be cognizant of the need for their own quality assurance policies and procedures.

A survey of users who interacted with the system during the study period showed an overall favorable response. While the survey was meant to focus on system performance and not AI performance, we recognize that these two concepts are likely intertwined in the users' experience. Therefore, similar assessments of user satisfaction should be performed with different AI algorithms matched with the same clinical integration components.

The results processing component could be extended in future versions to make calculations and measurements and integrate clinical data. For example, breast cancer risk has been shown to be correlated with age, family history, race, genetic markers, body mass index, several risk evaluation calculations, and imaging parameters of breast density at mammography (10,11). AI algorithms that generate automated breast density values have been reported in the literature (10,11). The results processing component could be extended to combine clinical and imaging parameters to generate a combined risk score to be inserted into a radiology report.

The presentation and delivery and error correction components deliver preliminarily processed results to the radiologist for review and, if needed, the ability to systematically record corrections. The practice of storing corrected results distinctly from original AI results in the database allows monitoring of the system performance through the dashboard component. Moreover, these data have value for the algorithm developers who desire to improve an algorithm over time and for regulatory review. Our current implementation injects the AI results into the radiology report using the API that is part of the commercial reporting product. Future solutions could take advantage of standards-based approaches, leveraging technologies such as Health Level Seven's Fast Healthcare Interoperability Resources DiagnosticReport (12). Moreover, presentation of AI results may additionally include overlays of annotation on images. Several standards do currently exist that support results storage, such as Annotation and Image Markup, DICOM Structured Reporting, DICOM Segmentation Object, and DICOM Parametric Maps (13), and other standardized annotation formats (14). Many PACS do not support visualization of these data, necessitating simpler solutions to enable the workflow, such as what we have developed.

Standards are emerging for facilitating adoption of AI into clinical workflows to ease solution deployment and provide long-term flexibility and interoperability. Two standards-based guidelines for AI integration that are currently in trial implementation are the Integrating the Healthcare Enterprise (IHE) Radiology AI Workflow for Imaging and AI Results profiles (15,16). Table 4 demonstrates alignment of the components in this study with the IHE profiles. The table also reveals several gaps and opportunities. Namely, while the structure of result objects and coordination of AI workflow are addressed, other necessary components are needed to transform this profile into a functional clinical solution.

For example, IHE profiles today do not adequately address the quality control workflows as described in this study. Our work demonstrated the benefit of interoperability for a ticketing system which, when implemented at scale, would benefit from having different actors being able to meaningfully contribute. Furthermore, supporting error correction workflows is meaningful regardless of which system is generating and consuming those messages. Extending the IHE profiles of Standardized Operational Log of Events and Audit Trail and Node Authentication (operation events and auditing, respectively) would be beneficial to many actors in the ecosystem. As interoperable workflows are developed and harmonized, building dashboards to monitor AI performance can be further enriched with more information. At

Table 4: Associations between AI Integration Components with Two Integrating Healthcare Enterprise Components Currently under Development

AI Integration Component	AIW-I Actor Correlate	AIR Actor Correlate
AI	Task performer	Evidence creator
Examination orchestrator	Task manager and/or watcher	Image manager
Quality control	No equivalent	No equivalent
Results database	Image manager	Image manager
Results processing	Task performer	Evidence creator
Presentation and delivery	Image manager, procedure update reporter, and/or report manager	Image display or imaging document consumer
Error correction	No equivalent	Image display or imaging document consumer
Active learning	No equivalent	No equivalent

Note.—Associations between artificial intelligence (AI) integration components in the lymphoscintigraphy use case with two Integrating the Healthcare Enterprise profile components currently under development: AI Workflow for Imaging (AIW-I) and AI Results (AIR). There are gaps and opportunities to transform the profiles into a functional clinical solution. The quality control component in our use case provides for specific validation of very rare but important conditions that can be addressed before analysis is performed. The error correction component deployed in our use case allows AI users to provide corrections to AI results as they are recognized. Moreover, data from the error correction module can be applied to retraining algorithms through a process called *active learning*. Active learning was not incorporated in our study but is included in the table for future reference, to be formalized in future profiles.

the time of writing, IHE is studying these extensions (17), and further alignment between our study and the broader community will continue to be reviewed and explored.

AI algorithms can enable radiologists to perform their jobs more accurately and efficiently, but architectures for deploying them in the clinical workflow are in the very early stages of development. The implementation of the software components we describe can ultimately be used to inform development of standards-based solutions.

References

1. Chartrand G, Cheng PM, Vorontsov E, et al. Deep learning: A primer for radiologists. RadioGraphics 2017;37(7):2113–2131.
2. Choy G, Khalilzadeh O, Michalski M, et al. Current applications and future impact of machine learning in radiology. Radiology 2018;288(2):318–328.
3. FDA Cleared AI Algorithms Cascade. https://models.acrdsi.org. Accessed May 8, 2021.
4. Albertini JJ, Lyman GH, Cox C, et al. Lymphatic mapping and sentinel node biopsy in the patient with breast cancer. JAMA 1996;276(22):1818–1822.
5. Ong L, Elnajjar P, Nyman CG, Mair T, Juluru K. Implementation of a point-of-care radiologist-technologist communication tool in a quality assurance program. AJR Am J Roentgenol 2017;209(1):W18–W25.
6. Miller JD. Big data visualization. Birmingham, England: Packt Publishing, 2017.
7. Ko I, Chang H. Interactive visualization of healthcare data using tableau. Healthc Inform Res 2017;23(4):349–354.
8. Kinzie MB, Cohn WF, Julian MF, Knaus WA. A user-centered model for web site design: needs assessment, user interface design, and rapid prototyping. J Am Med Inform Assoc 2002;9(4):320–330.
9. Tadavarthi Y, Vey B, Krupinski E, et al. The state of radiology AI: considerations for purchase decisions and current market offerings. Radiol Artif Intell 2020;2(6):e200004.
10. Kerlikowske K, Scott CG, Mahmoudzadeh AP, et al. Automated and clinical breast imaging reporting and data system density measures predict risk for screen-detected and interval cancers: A case-control study. Ann Intern Med 2018;168(11):757–765.
11. Wanders JOP, Holland K, Karssemeijer N, et al. The effect of volumetric breast density on the risk of screen-detected and interval breast cancers: a cohort study. Breast Cancer Res 2017;19(1):67.
12. Kamel PI, Nagy PG. Patient-centered radiology with FHIR: An introduction to the use of FHIR to offer radiology a clinically integrated platform. J Digit Imaging 2018;31(3):327–333.
13. Fedorov A, Clunie D, Ulrich E, et al. DICOM for quantitative imaging biomarker development: a standards based approach to sharing clinical data and structured PET/CT analysis results in head and neck cancer research. PeerJ 2016;4:e2057.
14. Clunie DA. DICOM structured reporting and cancer clinical trials results. Cancer Inform 2007;4:33–56.
15. IHE Radiology Technical Framework Supplement. AI Results. https://www.ihe.net/uploadedFiles/Documents/Radiology/IHE_RAD_Suppl_AIR.pdf. Accessed May 16, 2020.
16. IHE Radiology Technical Framework Supplement. AI Workflow for Imaging. [cited 2020 May 16]. https://www.ihe.net/uploadedFiles/Documents/Radiology/IHE_RAD_Suppl_AIW-I.pdf. Accessed May 16, 2020.
17. AI Interoperability in Imaging White Paper. IHE Wiki. https://wiki.ihe.net/index.php/AI_Interoperability_in_Imaging_White_Paper. Accessed March 20, 2021.