Radiology: Artificial Intelligence

ORIGINAL RESEARCH

An Artificial Intelligence Training Workshop for Diagnostic Radiology Residents

Ricky Hu, MASc • *Arsalan Rizwan, MD, PhD* • *Zoe Hu, MD* • *Anthony Li* • *Andrew David Chung, MD* • *Benjamin Y.M. Kwan, MD*

From the School of Medicine (R.H., A.L.) and Department of Diagnostic Radiology (R.H., A.R., Z.H., A.D.C., B.Y.M.K.), Queen's University, 76 Stuart St, Kingston, ON, Canada K7L 2V7. Received August 14, 2022; revision requested September 30; revision received December 6; accepted January 23, 2023. **Address correspondence to** R.H. (email: *rhu@qmed.ca*).

Supported by the Department of Radiology Research Grant at Queen's University.

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

Radiology: Artificial Intelligence 2023; 5(2):e220170 • https://doi.org/10.1148/ryai.220170 • Content code: AI

Purpose: To develop, implement, and evaluate feedback for an artificial intelligence (AI) workshop for radiology residents that has been designed as a condensed introduction of AI fundamentals suitable for integration into an existing residency curriculum.

Materials and Methods: A 3-week AI workshop was designed by radiology faculty, residents, and AI engineers. The workshop was integrated into curricular academic half-days of a competency-based medical education radiology training program. The workshop consisted of live didactic lectures, literature case studies, and programming examples for consolidation. Learning objectives and content were developed for foundational literacy rather than technical proficiency. Identical prospective surveys were conducted before and after the workshop to gauge the participants' confidence in understanding AI concepts on a five-point Likert scale. Results were analyzed with descriptive statistics and Wilcoxon rank sum tests to evaluate differences.

Results: Twelve residents participated in the workshop, with 11 completing the survey. An average score of 4.0 ± 0.7 (SD), indicating agreement, was observed when asking residents if the workshop improved AI knowledge. Confidence in understanding AI concepts increased following the workshop for 16 of 18 (89%) comprehension questions (*P* value range: .001 to .04 for questions with increased confidence).

Conclusion: An introductory AI workshop was developed and delivered to radiology residents. The workshop provided a condensed introduction to foundational AI concepts, developed positive perception, and improved confidence in AI topics.

Supplemental material is available for this article.

© RSNA, 2023

The field of artificial intelligence (AI) in medicine has experienced rapid growth in the past decade with an increasing number of applications being introduced into radiology practice (1). Radiologists are entrusted with integrating medical imaging AI tools with decision-making and therefore require sufficient AI literacy to evaluate its strengths and limitations. Hence, there has been frequent advocacy in the literature for radiology trainees to receive AI training (2–4). Collado-Mesa et al (5) reported that 97.1% of surveyed radiology trainees plan to pursue further AI training, but only 2.9% received any formal instruction. While workshops have been established to provide AI training for practicing radiologists within the framework of continuing medical education, few structured programs exist that are integrated into medical school or residency training.

There is currently no standardized AI curriculum for radiology residents or medical trainees, causing heterogeneity in the scope of existing AI training initiatives. For instance, the University of Toronto offers a 14-month Computing for Medicine course focused on a wide breadth of foundational computer science knowledge (6), and Harvard Medical School offers extracurricular clinical informatics courses for medical trainees (7). Specific to radiology, the Radiological Society of North America (RSNA) offers multiple extracurricular opportunities such as a 5-day National Imaging Informatics course that provides fundamental informatics not necessarily related to AI (8). RSNA also provides less didactic demonstrations such as the self-paced AI Deep Learning Lab, showcasing programming examples online, and shorter workshops such as the Imaging AI in Practice Demonstration. An institution-specific initiative was proposed by Wiggins et al (9), who implemented an 8-month data science pathway elective for three senior radiology residents paired with radiology faculty to learn from fundamental curricula and consolidate knowledge through AI research projects. Lindqwister et al (10) provided a series of lectures and journal clubs over 7 months, focusing on a different machine learning model type each month. Although trainees may participate in such extracurricular initiatives and continuing medical education workshops, AI training as an extracurricular option may be cumbersome as residents are occupied with clinical duties and mandatory academic learning obligations (11). This approach may be suitable for select radiology trainees who are interested in pursuing technical AI development, who may favor programs such as the data science pathway by Wiggins et al (9).

A general AI training program to provide all residents with a working knowledge of AI may require reduction

This copy is for personal use only. To order printed copies, contact reprints@rsna.org

Abbreviations

AI = artificial intelligence, CBME = competency-based medical education, CNN = convolutional neural network

Summary

A condensed artificial intelligence (AI) workshop implemented within curricular time increased confidence of residents in understanding AI concepts and may support future initiatives in integrating AI training within residency programs.

Key Points

- A 3-week artificial intelligence (AI) workshop for residents, containing didactic lectures, case studies, and programming examples to introduce foundational concepts, has been developed and implemented within curricular time of a radiology residency program.
- Eleven resident participants responded to a pre- and postworkshop survey assessing their confidence in understanding 18 AI topics and showed improved confidence in 16 of 18 (89%) topics in the postworkshop survey.
- The pilot workshop can be implemented in existing radiology residency curricula, as it aims to address challenges in selecting sufficiently thorough AI content without being too burdensome on residents.

Keywords

Medical Education, Machine Learning, Postgraduate Training, Competency-based Medical Education, Medical Informatics

in scope. To reduce the need to invest extracurricular hours, it may be advantageous to integrate AI into protected academic time within the core curriculum and to select only concepts for foundational literacy. Richardson and Ojeda (12) proposed a condensed 6-hour "no-code" curriculum focused on tutorials on using deep learning image classification examples in an online preprogrammed interface. The curriculum was well received, with 91% of residents reporting its helpfulness in understanding AI. While the examples provided are highly relevant and curriculum length is less burdensome than full AI courses, it may be beneficial to also introduce fundamental data science concepts and non–deep learning models to develop a general aptitude in AI.

Designing an AI curriculum for medical trainees has specific challenges related to heterogeneity of background knowledge, selection of content, and instruction method. Traditional AI training builds upon foundational knowledge in linear algebra, multivariate calculus, algorithms, and software construction (13), which are subjects not commonly taught in medical curricula. However, most medical trainees do receive basic biostatistics training targeted toward the critical appraisal of medical literature as part of their undergraduate medical education (14). Therefore, the integration of AI training at the postgraduate level ensures that participants have at least foundational knowledge in statistics.

Last, AI instruction for radiology trainees requires a curricular team with expertise in both computer science and radiology. This may require a multidisciplinary team to select topics that are clinically relevant, to convey the mathematical foundations of AI, and to highlight the strengths and limitations of AI in a particular application. Separate radiologists and computer scientists may not be necessary, particularly if the instructing team is from a clinical research facility with expertise in AI in radiology, which may have the additional strength of increasing relevance for trainees.

To address the need for AI training and its challenges in implementation, we created and delivered a pilot AI training workshop for diagnostic radiology resident physicians. The workshop was specifically designed by radiology faculty and AI engineers to condense content and develop foundational knowledge for evaluating AI applications in health care. The workshop was integrated into dedicated academic training hours which is, to our knowledge, the first such initiative in Canada. A survey was conducted before and after the workshop to assess the participants' baseline demographics, perception of AI, and confidence in understanding course concepts. Ultimately, this work aims to inform and motivate future initiatives to integrate AI teaching for medical trainees into existing teaching curricula.

Materials and Methods

Workshop Structure and Delivery Method

The workshop delivery and prospective survey data collection were approved by the Queen's University Health Sciences & Affiliated Teaching Hospitals Research Ethics Board (code: RADO-107-21) with written informed consent obtained from all participants for the anonymized survey. All participants were briefed on the data collected in the study, risks, and process of disposal of data.

The workshop was delivered to radiology residents for 3 weeks as part of weekly academic half-days, where residents were provided with a weekly time slot free of clinical duties for academic learning (15). Although participation was encouraged, attendance was not mandatory nor was it recorded for an individual's record. Delivery was virtual as the academic half-days were transitioned to virtual due to the COVID-19 pandemic, although the format is compatible with in-person instruction as well. The workshop was executed in a residency program with a competency-based medical education (CBME) curriculum (the first CBME diagnostic radiology residency program in Canada), which aims to measure educational progression primarily through the achievement of specific entrustable professional activities, or EPAs, rather than time (16). This was done to integrate AI training into the existing curriculum rather than create an entirely new extracurricular workshop, with the goal of promoting attendance and reducing disruption to individual schedules. In each week, the live workshop session length was 1 hour and consisted of equal parts of didactic lecture, case studies in literature, case studies in media, and programming demonstrations. After the first 2 weeks, optional self-study material with an approximate time commitment of 1 hour each was provided, including journal articles and annotated software code. The course structure and an example of a programming demonstration are visualized in Figure 1. The curricular content and programming examples are available at a public online repository at *https://github.com/Queens-Radiology-Intro-To-AI/Intro-to-AI*.

The goal of reducing didactic content in favor of problem-solving through examples was to promote consolidation of knowledge similar to that of the flipped classroom model (17). In total, 5 hours of curricular content were provided. The content

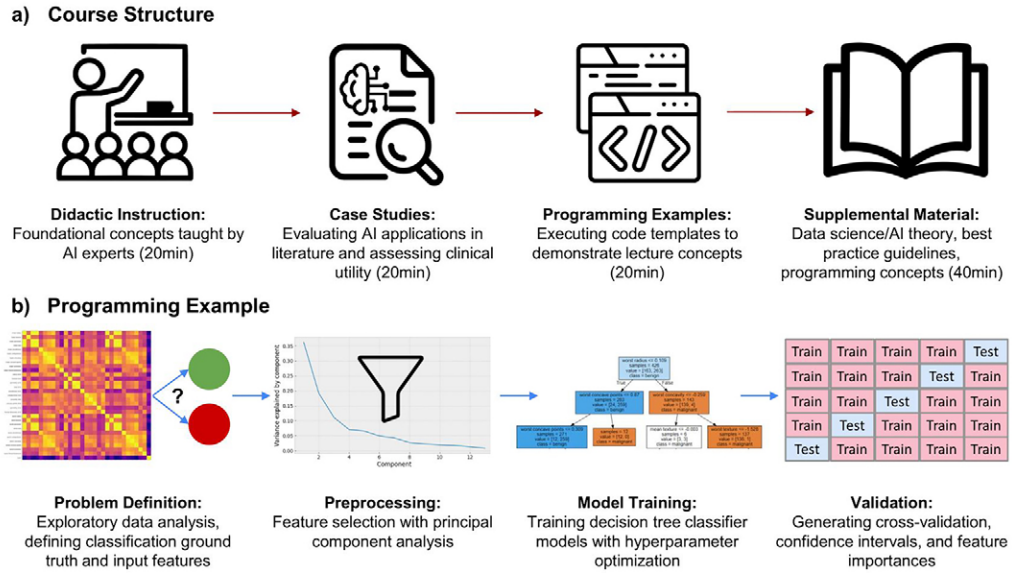


Figure 1: **(A)** Overview of the workshop structure. Each week consists of didactic lectures, consolidated by case studies and programming examples relevant to the topics from the lecture. Supplemental material was provided for participants who wished to further pursue topics beyond the scope of the workshop. **(B)** Visualization of a programming example. In this specific example, a decision tree classifier is trained to predict malignancy from tumor images. The example highlights data science best practices in organizing input features, feature selection, model optimization, and validation. AI = artificial intelligence.

was designed by a team of staff radiologists (A.D.C., B.Y.M.K.), radiology residents (A.R., Z.H.), and AI engineer (R.H.). The live content was delivered by an instructor (R.H.) with a graduate degree in engineering and 5 years of experience in AI.

Learning Objectives

The learning objectives, summarized in Table 1, followed Bloom taxonomy (18), where content delivery was designed to follow the highest taxonomic level possible to promote mastery. However, due to challenges of time constraints and varying levels of prerequisite knowledge, the learning objectives target AI literacy rather than AI proficiency. As a result, foundational principles were emphasized, such as concepts of problem definition, validation, and assessing clinical use. Advanced concepts were delivered with less rigorous objectives. For instance, convolutional neural networks (CNNs) were a topic of interest due to popularity in the literature. However, as interpreting CNN processing involves matrix multiplication and optimization with gradient descent algorithms (19), participants were only expected to recognize when a CNN was presented and understand how to evaluate performance of the model output.

Content

Detailed lecture, case study, and programming topics in the curriculum are summarized in Table 2, with corresponding learning objectives from Table 1 in parentheses. Briefly, fundamental AI concepts were introduced in lectures, particularly topics required to analyze the clinical use of an AI application, such as understanding model validation. During case studies, a particular journal article on AI technology was introduced, and the strengths and limitations of the application were evaluated. For instance, one case study computed radiomic features and used the features as an input to a logistic regression classifier (20). The goal of the study was to identify best practices in data science such as the documentation of preprocessing, feature selection to prevent overfitting, and resampling methods for validation. Finally, programming demonstrations were provided as template code in the Google Colaboratory environment (Google), where participants can execute code within their own browser. This provided the advantages of allowing participants to visualize data processing of an AI model without requiring training to install software packages, manage dependencies, or troubleshoot programming runtime errors. The programming examples would reflect topics discussed during the particular session. For instance, one topic taught was resampling methods to generate CIs when validating an AI model. The programming example demonstrated cross-validation and how to plot the range of cross-validated accuracies when retraining an AI model with different subsets of data.

Data Collection

To assess the effectiveness of the workshops, participants were asked to complete two surveys: a preworkshop survey and a postworkshop survey. The surveys were anonymous. The preworkshop survey consisted of two demographic questions on level of training, 18 comprehension questions assessing the participant's perceived confidence in a certain concept, and three perception questions surveying general opinions on AI in their careers. The postworkshop survey was nearly identical, containing two extra questions on perception of the workshop. The comprehension questions were rated on a five-point Likert scale, where five represents strong agreement and one represents strong disagreement. A list of the survey questions and data format is presented in Table S1.

Table 1: A Summary of Learning Objectives for the Artificial Intelligence Training Workshop with Bloom Taxonomic Levels in Parentheses

Session	Learning Objectives (Bloom Taxonomic Level)
1	1.1: To compare and contrast the differences between AI, ML, and deep learning (level 2)
	1.2: To define a feature, sample, ground truth, model, and task in the context of AI (level 2)
	1.3: To understand the strengths and limitations of classic statistics in predictive analysis (level 2)
	1.4: To distinguish supervised, unsupervised, and semisupervised learning (level 2)
	1.5: To assess the performance of AI classifiers when compared with radiologists in diagnostic tasks (level 5)
	1.6: To apply exploratory data analysis in software (level 3)
2	2.1: To understand the process of ML (problem definition, data preprocessing, model selection, optimizing, training, validation) (level 2)
	2.2: To present a clinical problem as an AI task with defined input and output (level 3)
	2.3: To understand how dimensionality reduction reduces overfitting (level 2)
	2.4: To recognize preprocessing techniques of first-order statistic filters, principal component analysis, and collinearity filtering (level 1)
	2.5: To recognize ML models of classification rules, logistic regression, decision trees, support vector machines, k-nearest neighbors, random forests, and neural networks (level 1)
	2.6: To select a compatible validation metric for classification, regression, and segmentation problems (level 3)
	2.7: To assess risk of bias for ML model performance (level 3)
	2.8: To contrast learning parameters and hyperparameters (level 2)
	2.9: To apply preprocessing to train an ML model in software (level 3)
3	3.1: To outline the input and output data types to convolutional neural networks and generative adversarial networks (level 1)
	3.2: To recognize increase of complexity in neural networks (level 1)
	3.3: To understand the sensitivity of neural networks to perturbations of training data (level 2)
	3.4: To recognize when a modern neural network is present in literature (level 1)
	3.5: To recognize interpretability techniques for neural networks (level 1)
	3.6: To evaluate performance of neural networks in literature (level 4)
	3.7: To understand the role of interpretability for clinical utility (level 2)
	3.8: To apply cross-validation and generate CIs for ML models (level 3)
General	4.1: To assess the potential role of AI in medicine (level 5)
	4.2: To evaluate if AI can address a clinical problem (level 5)
	4.3: To locate resources for standard guidelines for predictive analysis and assessing risk of bias (level 1)

Note.—The learning objectives focus on developing foundational literacy; hence, more emphasis and higher taxonomic levels were assigned to the initial two sessions, and more advanced topics, like neural networks, were presented in the third session to develop simple familiarity. Bloom taxonomic levels are defined as 1 = remembering, 2 = understanding, 3 = applying, 4 = analyzing, 5 = evaluating, 6 = creating. AI = artificial intelligence, ML = machine learning.

Statistical Analysis

The preworkshop and postworkshop surveys were analyzed to generate first-order summary statistics. Wilcoxon rank sum tests were conducted to observe differences between the preworkshop and postworkshop survey responses as a nonparametric significance test for small sample sizes. A *P* value less than .05 was used to indicate statistically significant differences. All analyses were performed in MATLAB R2020b (MathWorks).

Results

Overview of Participant Responses

A total of 12 residents participated in the workshop, with 11 providing survey responses. Figure 2 displays the distribution of demographic survey questions and responses regarding the perception of general knowledge improvement during the workshop and of interest in AI training. Table 3 summarizes the mean scores of the survey responses, and Figure 3 shows the distribution of Likert scores comparing preworkshop and postworkshop answers.

Comparison of Pre- and Postworkshop Scores

No participants had previous undergraduate or professional training in AI. Six of 11 (55%) respondents felt their knowledge somewhat improved, and four of 11 (36%) felt their knowledge strongly improved after the course. There was an increase in confidence among participants in 16 of 18 of the fundamental concepts after the completion of the workshop. Specifically, two

Table 2: A Summary of the Curriculum Delivered to the Resident Physicians with Associated Learning Objectives from Table 1 in Parentheses

Session/Lecture Topic (Learning Objective)	Case Studies (Learning Objective)	Programming Example (Learning Objective)
Session 1		
Defining AI, ML, and DL (1.1)		
Defining features, samples, ground truths, models, and tasks (1.2)	Traumatic brain injury decision tree [Kuppermann et al (24)] (1.3, 4.1, 4.2)	Importing and formatting features, outcome and sample (1.3, 1.6, 2.1)
Descriptive versus inferential versus predictive analysis (1.3)	Comparison of predictions of 101 radiologists versus ML models [Rodriguez-Ruiz et al (25)] (1.5, 4.1)	Displaying input data as a data frame (1.6)
Classic clinical prediction models versus ML models (1.3)	FDA-approved AI devices in health care and level of invasiveness [van Leeuwen et al (26)] (1.5, 4.1)	Generating descriptive statistics for the dataset (1.3, 1.6)
Supervised versus unsupervised learning (1.4)		Visualizing Pearson correlation ratios (1.3, 1.6)
Session 2		
The ML pipeline (problem definition, preprocessing, model selection, training, optimizing, validation) (2.1, 2.2)		
Classification versus regression versus segmentation (2.1)	Survival prediction with radiomic feature extraction and a logistic regression classifier [Bogowicz et al (20)] (2.2, 2.3, 2.4, 2.5, 2.6, 4.2)	Dimensionality reduction with PCA (2.3, 2.9)
Dimensionality reduction (2.3)	COVID-19 diagnostic prediction from acoustic analysis of cough [Laguarta et al (27)] (2.2, 2.6, 2.7, 4.2)	Training a logistic regression classifier and a decision tree classifier on the Wisconsin Breast Cancer Dataset (2.5, 2.9)
ML models: classification rules, logistic regression, SVM, kNN, decision trees, random forests, neural networks (2.5)		Visualizing training and testing classification accuracies (2.6, 2.9)
Resampling techniques: cross-validation and bootstrapping		Hyperparameter optimization of decision trees (2.8, 2.9)
Evaluation metrics for different AI tasks (2.6, 2.7)		Cross-validation and CIs for model performance (2.6, 2.7, 2.9)
Complexity versus interpretability of ML models (2.6)		
Optimization: parameters versus hyperparameters (2.8)		
Session 3		
Input and output data types to neural networks and CNNs (3.1, 3.4)		
Complexity of neural networks and sensitivity to input data (3.2, 3.3)	Predicting COVID-19 diagnosis from CT with a CNN [Li et al (28)] (2.2, 3.1, 3.2, 3.4, 3.5, 3.6, 3.7, 4.2)	Training neural network classifier on the Wisconsin Breast Cancer Dataset (3.1, 3.4)
Data augmentation for CNNs (3.3)	Predicting thoracic abnormalities from chest radiographs with a CNN [Rajpurkar et al (29)] (2.2, 3.1, 3.2, 3.4, 3.5, 3.6, 3.7, 4.2)	Evaluating classification performance of neural networks (3.6, 3.8)
Saliency maps, Grad-CAMs, and LIME for interpretability (3.5, 3.7)		Adjust hyperparameters and kernels of neural network (3.2, 3.3)
Evaluating performance of CNNs (3.6)		CNN for classification of MedNIST images (3.2, 3.3, 3.4)
Risk of bias for ML models (3.7)		
CNN versus point-data models (2.2, 2.5, 3.7)		

(Table 2 continues)

concepts with no confidence increase were "I understand what steps are required for data cleaning" and the general question of "I understand what artificial intelligence is." Most participants expressed interest in continuing AI education, with five of 11 (45%) somewhat interested and four of 11 (36%) strongly interested. We found no evidence of a difference in perception of concern over

Table 2 *(continued)*: A Summary of the Curriculum Delivered to the Resident Physicians with Associated Learning Objectives from Table 1 in Parentheses

Session/ Lecture Topic (Learning Objective)	Case Studies (Learning Objective)	Programming Example (Learning Objective)
Supplemental reference material		
List of common AI models and trade-offs between complexity and interpretability	TRIPOD checklist for predictive analysis models [Collins et al (30)] (2.6, 2.7, 4.2, 4.3)	Tutorial on running software in the Jupyter environment
Neural network kernels and activation functions	PROBAST checklist for assessing risk of bias [Wolff et al (31)] (2.6, 2.7, 4.2, 4.3)	Computation requirements for DL
Optimization of neural networks with backpropagation	Review of performance difference in external validation of DL models [Yu et al (32)] (1.5, 4.1, 4.2)	

Note.—The workshop consisted of three sessions with live lectures, case studies, and programming examples. The content was selected to satisfy learning objectives that aimed to promote literacy of artificial intelligence fundamentals. AI = artificial intelligence, CNN = convolutional neural network, DL = deep learning, FDA = U.S. Food and Drug Administration, Grad-CAM = gradient class activation map, kNN = k-nearest neighbors, LIME = local interpretable model-agnostic explanations, MedNIST = medical National Institute of Standards and Technology, ML = machine learning, PCA = principal component analysis, PROBAST = prediction model risk of bias assessment tool, SVM = support vector machine, TRIPOD = transparent reporting of a multivariable prediction model for individual prognosis or diagnosis.

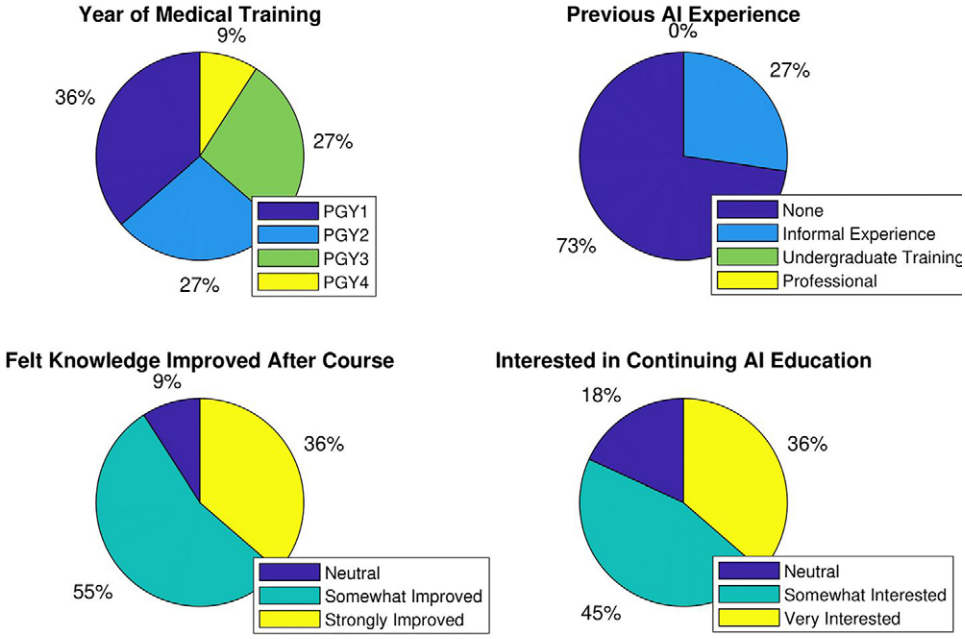


Figure 2: Baseline characteristics of all the participants and postworkshop survey responses indicate resident confidence and interest. The participants included residents from postgraduate year (PGY) 1 to 4. None of the participants in the workshop had previous undergraduate training or professional experience in artificial intelligence (AI). Most participants reported that knowledge somewhat or strongly improved and were somewhat or very interested in continuing AI training in the future.

impact of AI on career, importance of AI education, or interest in continuing education between pre- and postworkshop surveys.

Discussion

In this study, we describe the development and implementation of an AI workshop for radiology residents. Overall, the workshop improved confidence in all but two AI concepts surveyed. Every other concept resulted in a statistically significant increase in confidence rating. The workshop was delivered by a curricular team of staff radiologists and engineers as part of dedicated academic time for radiology residents, providing a condensed introduction to AI.

Although there was no statistically significant change in gauging general opinions of AI, this was likely due to baseline enthusiasm in AI and perceived importance of AI reported in the preworkshop survey. Generally, feedback from residents was positive with most agreeing that the workshop improved their AI knowledge. Particular praise was given to

Table 3: A Summary of the Mean Scores of Preworkshop and Postworkshop Survey Responses

Questions	Mean Preworkshop Score	Mean Postworkshop Score	Wilcoxon Rank Sum *P* Value
(Postworkshop survey only) The workshop improved my knowledge of AI	NA	4.0 ± 0.7	NA
(Postworkshop survey only) I am interested in similar courses in the future	NA	4.1 ± 0.7	NA
I can describe the difference between AI, machine learning, and deep learning	2.1 ± 1.0	3.7 ± 0.7	.008
I can describe the difference between descriptive and observational statistics	2.5 ± 1.1	3.7 ± 0.7	.03
I understand what features, samples, and labels are in the context of AI	2.1 ± 1.2	3.2 ± 0.8	.02
I understand what steps are required for data cleaning	2.1 ± 1.1	3.4 ± 1.8	.16
I understand when feature selection is required	2.0 ± 1.0	3.3 ± 0.7	.006
I can confidently organize a dataset to training and testing subsets	2.0 ± 1.0	3.2 ± 0.9	.04
I can recognize different point-data machine learning models	1.8 ± 0.9	3.1 ± 0.3	.003
I understand the differences between complexity and interpretability between different machine learning models	1.8 ± 1.2	3.2 ± 0.4	.02
I can identify when a model is overfitting	1.9 ± 1.0	3.6 ± 0.5	.002
I understand what cross-validation and bootstrapping are and when it is necessary	2.0 ± 1.0	3.6 ± 0.7	.02
I can describe validation processes required to evaluate machine learning models	2.0 ± 0.8	3.3 ± 0.7	.004
I can recognize when deep learning architecture is used	2.1 ± 1.4	3.3 ± 0.7	.01
I understand testing requirements for machine learning to be deployed into clinical practice	1.8 ± 0.8	3.3 ± 0.5	.001
I understand the ethical implications of AI in health care	2.8 ± 0.8	3.7 ± 0.8	.02
I feel confident in my ability to explain how AI may impact health care	2.3 ± 0.9	3.5 ± 0.5	.02
I understand what artificial intelligence is	3.4 ± 0.9	3.9 ± 0.6	.23
I am comfortable to discuss literature on artificial intelligence in medicine	2.3 ± 0.9	3.2 ± 0.6	.04
I can recognize when an AI application (eg, publication or device) does not follow best data science practices	2.1 ± 0.9	3.3 ± 0.7	.008
I am interested in continuing my training in AI literacy in the future	3.8 ± 0.8	4.0 ± 0.7	.72
I am concerned on the impact of AI on my future career	2.9 ± 0.8	3.3 ± 0.8	.33
I believe AI education is important for my future career	4.1 ± 0.5	4.1 ± 0.3	>.99

Note.—Unless otherwise indicated, data are means ± SDs. Responses were rated on a five-point Likert scale: 1 = strongly disagree, 2 = somewhat disagree, 3 = neutral, 4 = somewhat agree, and 5 = strongly agree. AI = artificial intelligence, NA = not applicable.

the use of relevant medical literature, which described how one may analyze appropriateness of a new AI model for a specific clinical problem. This provides support that training in this format can provide added value in AI literacy for future residents.

To the best of our knowledge, this workshop is the first implementation of AI teaching within a CBME residency curriculum. This is different from existing AI courses, which are external to institutional curricula and require extracurricular time commitment to attend. With future iterations and success, this work can provide support and lay the foundation for AI training to be part of residency curriculum. For instance, such training may be implemented as an accredited competency recognized by postgraduate regulatory bodies.

A particular challenge was designing content with sufficient coverage of topics without a cumbersome time commitment for the diagnostic radiology residents. A thorough course on AI, similar to programs for engineering students with consolidation tasks such as building a custom software, would likely improve understanding but is not feasible with time constraints. In addition, the involvement with AI widely varies across residents, where some may be actively pursuing AI research and others pursuing a user role. To compromise, the goal of our workshop was to develop foundational literacy of select topics. Concepts used in assessing the performance of an AI application, such as problem definition, data types, and validation, were emphasized. Selection of specific applications, however, was a greater challenge. For instance, CNNs are increasing in popularity, particularly in radiology literature, as they comprise the top-performing models on public image classification challenges (21). While it was not feasible to train participants at a level to understand how a CNN learns through convolutional kernels, we desired to improve familiarity on this popular topic. The compromise was to present a general intuition of CNN input and output data types and to review journal articles implementing CNNs so that participants can recognize if a CNN is being used in a health care tool. The goal was that participants

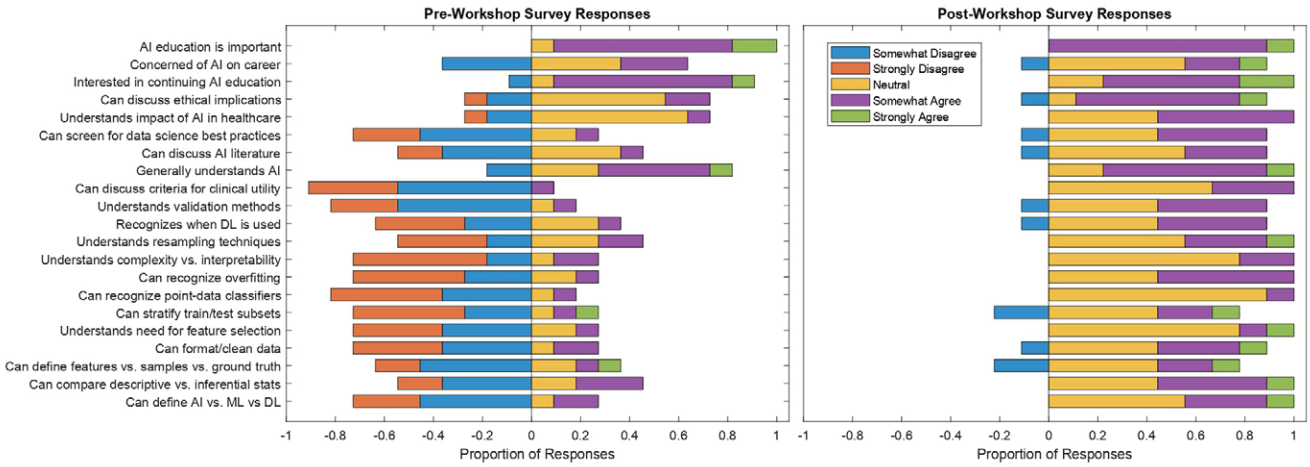


Figure 3: Survey responses before and after the workshop. The questions gauge the confidence of participants in a particular topic, with three questions gauging the perception of the impact, importance, and interest of artificial intelligence. Responses were rated on a five-point Likert scale: 1 = strongly disagree, 2 = somewhat disagree, 3 = neutral, 4 = somewhat agree, and 5 = strongly agree. AI = artificial intelligence, DL = deep learning, ML = machine learning.

can then apply foundational principles, such as assessing if proper validation practices were used given limitations of CNNs relating to sensitivity and interpretability.

Our study had several limitations. First, participants were limited to a single postgraduate program in a single university. This allowed for easier scheduling to coordinate protected academic time but resulted in a smaller sample size. To address this limitation, repeating the workshop in the future at different institutions with additional residents can provide more statistical power in assessing feasibility of the workshop. This will require multi-institutional coordination to select content of appropriate breadth, depth, and delivery. Second, knowledge retainment is unknown given the lack of immediate applications of AI concepts for most of the participants. Medical education employs strategies of reintroducing concepts in clinical rotations to consolidate knowledge (22), and existing training programs for radiology residents retrieved qualitative feedback to assess its receptions. Wiggins et al (9) reported "very positive" feedback, and Richardson and Ojeda (12) reported that most participants felt "the course taught them useful things about deep learning." The results of our workshop were similarly positive, and more quantitative assessment of confidence in concepts was measured in addition to general feedback. Lindqwister et al (10) provided a five-item Likert scale assessing confidence before and after each lecture, resulting in a statistically significant increase in confidence in every lecture. Our study was different in that we assessed specific concepts relating to a detailed list of learning objectives but produced similar results to this survey. Last, as attendance was not mandatory, there is susceptibility to selection bias such that more enthusiastic residents may have attended. This limits the applicability of the results to the general resident population. However, it should be noted that there are a total of 15 residents in the program and most attended the workshop. There remains a lack of assessment data in our study and in literature for evaluating changes in understanding. Although assessment data can be obtained through traditional methods of examinations or supervised demonstrations, it may be burdensome to add these assessments onto an already rigorous residency curriculum. For a topic such as AI, assessment of confidence was chosen as a preliminary method. Confidence assessment has been employed in the past as an analog to assessing key skills (23) by providing relative measurements on a learner's changes of skill over time but has limited use in assessing an absolute measurement of skill. One solution may be to assess resident knowledge through consolidating activities, such as by resident-led journal clubs or having a resident analyze an AI tool that a hospital deploys to analyze details of the model such as data dimensionality, training, testing, validation, and limitations. Practical demonstrations can be utilized similar to Richardson and Ojeda (12), where a graphic user interface is provided for residents to select and visualize training different models to stimulate discussion of limitations when applying the model to clinical practice.

In summary, we have implemented an introductory AI training workshop for radiology residents integrated into the teaching curriculum at a CBME institution. The workshop aimed to provide foundational knowledge for assessing AI in medicine in the future and improved self-reported confidence on fundamental topics in AI. Similar workshops may be integrated into standardized curricula in the future to provide institutional support and accreditation to motivate AI literacy.

Acknowledgment: The authors acknowledge the department of diagnostic radiology at Queen's University for institutional support.

Author contributions: Guarantors of integrity of entire study, **R.H., A.R., B.Y.M.K.**; study concepts/study design or data acquisition or data analysis/interpretation, all authors; manuscript drafting or manuscript revision for important intellectual content, all authors; approval of final version of submitted manuscript, all authors; agrees to ensure any questions related to the work are appropriately resolved, all authors; literature research, **R.H., A.R., Z.H., A.L., B.Y.M.K.**; clinical studies, **R.H.**; experimental studies, **R.H., A.L.**; statistical analysis, **R.H., Z.H.**; and manuscript editing, all authors

Disclosures of conflicts of interest: **R.H.** No relevant relationships. **A.R.** *Radiology: Artificial Intelligence* trainee editorial board member. **Z.H.** No relevant relationships. **A.L.** Unpaid volunteer director for LEADs Employment Services Board of Directors. **A.D.C.** No relevant relationships. **B.Y.M.K.** Queen's University Department of Radiology Research Grant.

References

1. Kulkarni S, Seneviratne N, Baig MS, Khan AHA. Artificial Intelligence in Medicine: Where Are We Now? Acad Radiol 2020;27(1):62–70.
2. Wood MJ, Tenenholtz NA, Geis JR, Michalski MH, Andriole KP. The Need for a Machine Learning Curriculum for Radiologists. J Am Coll Radiol 2019;16(5):740–742.
3. Grunhut J, Wyatt AT, Marques O. Educating Future Physicians in Artificial Intelligence (AI): An Integrative Review and Proposed Changes. J Med Educ Curric Dev 2021;8:23821205211036836.
4. Huisman M, Ranschaert E, Parker W, et al. An international survey on AI in radiology in 1,041 radiologists and radiology residents part 1: fear of replacement, knowledge, and attitude. Eur Radiol 2021;31(9):7058–7066.
5. Collado-Mesa F, Alvarez E, Arheart K. The Role of Artificial Intelligence in Diagnostic Radiology: A Survey at a Single Radiology Residency Training Program. J Am Coll Radiol 2018;15(12):1753–1757.
6. Windish DM, Huot SJ, Green ML. Medicine residents' understanding of the biostatistics and results in the medical literature. JAMA 2007;298(9):1010–1022.
7. McCoy LG, Nagaraj S, Morgado F, Harish V, Das S, Celi LA. What do medical students actually need to know about artificial intelligence? NPJ Digit Med 2020;3(1):86.
8. National Imaging Informatics Curriculum and Course. Radiological Society of North America. https://www.rsna.org/education/trainee-resources/national-imaging-informatics-curriculum-and-course. Published 2022. Accessed October 15, 2022.
9. Wiggins WF, Caton MT, Magudia K, et al. Preparing Radiologists to Lead in the Era of Artificial Intelligence: Designing and Implementing a Focused Data Science Pathway for Senior Radiology Residents. Radiol Artif Intell 2020;2(6):e200057.
10. Lindqwister AL, Hassanpour S, Lewis PJ, Sin JM. AI-RADS: An Artificial Intelligence Curriculum for Residents. Acad Radiol 2021;28(12):1810–1816.
11. Kijima S, Tomihara K, Tagawa M. Effect of stress coping ability and working hours on burnout among residents. BMC Med Educ 2020;20(1):219.
12. Richardson ML, Ojeda PIA. A "Bumper-Car" Curriculum for Teaching Deep Learning to Radiology Residents☆. Acad Radiol 2022;29(5):763–770.
13. Kolachalama VB, Garg PS. Machine learning and medical education. NPJ Digit Med 2018;1(1):54.
14. Wiens J, Saria S, Sendak M, et al. Do no harm: a roadmap for responsible machine learning for health care. Nat Med 2019;25(9):1337–1340. [Published correction appears in Nat Med 2019;25(10):1627.]
15. Hames K, Patlas M, Duszak R. Barriers to Resident Research in Radiology: A Canadian Perspective. Can Assoc Radiol J 2018;69(3):260–265.
16. Kwan BYM, Mbanwi A, Cofie N, et al. Creating a Competency-Based Medical Education Curriculum for Canadian Diagnostic Radiology Residency (Queen's Fundamental Innovations in Residency Education)-Part 1: Transition to Discipline and Foundation of Discipline Stages. Can Assoc Radiol J 2021;72(3):372–380.
17. Prober CG, Khan S. Medical education reimagined: a call to action. Acad Med 2013;88(10):1407–1410.
18. Adams NE. Bloom's taxonomy of cognitive learning objectives. J Med Libr Assoc 2015;103(3):152–153.
19. Mutasa S, Sun S, Ha R. Understanding artificial intelligence based radiology studies: CNN architecture. Clin Imaging 2021;80:72–76.
20. Bogowicz M, Jochems A, Deist TM, et al. Privacy-preserving distributed learning of radiomics to predict overall survival and HPV status in head and neck cancer. Sci Rep 2020;10(1):4542.
21. Morid MA, Borjali A, Del Fiol G. A scoping review of transfer learning research on medical image analysis using ImageNet. Comput Biol Med 2021;128:104115.
22. Kumaravel B, Stewart C, Ilic D. Development and evaluation of a spiral model of assessing EBM competency using OSCEs in undergraduate medical education. BMC Med Educ 2021;21(1):204.
23. Bray A, Byrne P, O'Kelly M. A Short Instrument for Measuring Students' Confidence with 'Key Skills' (SICKS): Development, Validation and Initial Results. Think Skills Creat 2020;37:100700.
24. Kuppermann N, Holmes JF, Dayan PS, et al. Identification of children at very low risk of clinically-important brain injuries after head trauma: a prospective cohort study. Lancet 2009;374(9696):1160–1170.
25. Rodriguez-Ruiz A, Lång K, Gubern-Merida A, et al. Stand-alone artificial intelligence for breast cancer detection in mammography: comparison with 101 radiologists. J Natl Cancer Inst 2019;111(9):916–922.
26. van Leeuwen KG, Schalekamp S, Rutten MJCM, van Ginneken B, de Rooij M. Artificial intelligence in radiology: 100 commercially available products and their scientific evidence. Eur Radiol 2021;31(6):3797–3804.
27. Laguarta J, Hueto F, Subirana B. COVID-19 Artificial Intelligence Diagnosis Using Only Cough Recordings. IEEE Open J Eng Med Biol 2020;1:275–281.
28. Li L, Qin L, Xu Z, et al. Using Artificial Intelligence to Detect COVID-19 and Community-acquired Pneumonia Based on Pulmonary CT: Evaluation of the Diagnostic Accuracy. Radiology 2020;296(2):E65–E71.
29. Rajpurkar P, Irvin J, Ball RL, et al. Deep learning for chest radiograph diagnosis: a retrospective comparison of the CheXNeXt algorithm to practicing radiologists. PLoS Med 2018;15(2):e1002686.
30. Collins GS, Reitsma JB, Altman DG, et al. Transparent reporting of a multivariable prediction model for individual prognosis or diagnosis (TRIPOD): the TRIPOD Statement. Ann Intern Med 2015;162(1):55–63.
31. Wolff RF, Moons KGM, Riley RD, et al. PROBAST: A Tool to Assess the Risk of Bias and Applicability of Prediction Model Studies. Ann Intern Med 2019;170(1):51–58.
32. Yu AC, Mohajer B, Eng J. External Validation of Deep Learning Algorithms for Radiologic Diagnosis: A Systematic Review. Radiol Artif Intell 2022;4(3):e210064.