Journal of Digital Imaging
https://doi.org/10.1007/s10278-018-0066-y

Efficiency Improvement in a Busy Radiology Practice: Determination of Musculoskeletal Magnetic Resonance Imaging Protocol Using Deep-Learning Convolutional Neural Networks

Young Han Lee[1]

Abstract

The purposes of this study are to evaluate the feasibility of protocol determination with a convolutional neural networks (CNN) classifier based on short-text classification and to evaluate the agreements by comparing protocols determined by CNN with those determined by musculoskeletal radiologists. Following institutional review board approval, the database of a hospital information system (HIS) was queried for lists of MRI examinations, referring department, patient age, and patient gender. These were exported to a local workstation for analyses: 5258 and 1018 consecutive musculoskeletal MRI examinations were used for the training and test datasets, respectively. The subjects for pre-processing were routine or tumor protocols and the contents were word combinations of the referring department, region, contrast media (or not), gender, and age. The CNN Embedded vector classifier was used with Word2Vec Google news vectors. The test set was tested with each classification model and results were output as routine or tumor protocols. The CNN determinations were evaluated using the receiver operating characteristic (ROC) curves. The accuracies were evaluated by a radiologist-confirmed protocol as the reference protocols. The optimal cut-off values for protocol determination between routine protocols and tumor protocols was 0.5067 with a sensitivity of 92.10%, a specificity of 95.76%, and an area under curve (AUC) of 0.977. The overall accuracy was 94.2% for the ConvNet model. All MRI protocols were correct in the pelvic bone, upper arm, wrist, and lower leg MRIs. Deep-learning-based convolutional neural networks were clinically utilized to determine musculoskeletal MRI protocols. CNN-based text learning and applications could be extended to other radiologic tasks besides image interpretations, improving the work performance of the radiologist.

Keywords Artificial neural networks · Machine learning · Magnetic resonance imaging protocol · Image protocols

Introduction

Musculoskeletal magnetic resonance imaging (MRI) is a powerful tool for the evaluation, assessment of severity, and follow-up of diseases of the musculoskeletal organs. Clinically, the MRI protocols are determined by imaging purpose, clinical indications, patient information, and clinical history based on the referring physician's clinical diagnosis [1]. Radiologists who determine the MRI pulse sequences and the imaging planes to be scanned, as well as interpret MRI must understand the details and limitations of the various imaging pulse sequences. The MRI protocols are a combination of various MRI sequences, imaging planes, slice thicknesses/gaps, imaging ranges, and various imaging parameters designed for successful clinical imaging.

General principles of protocol design include diagnostic performance, image quality and radiologic efficiency, MRI hardware and software, radiologist's preference, patient factors, and scan time constraints [2]. Scan time reduction of MRI is becoming increasingly important due to an increased demand for cost-effectiveness. There is an inherent trade-off between MRI scan time, image quality, and signal-to-noise ratio (SNR) [3]. The musculoskeletal MRI protocol can be generally divided into routine purpose and tumor/infection. The former is applicable to general evaluation and joint/intervertebral disc evaluation, focusing on the general joint problems, such as ligaments and tendons. The latter is used

✉ Young Han Lee
sando@yuhs.ac

[1] Department of Radiology, Research Institute of Radiological Science, YUHS-KRIBB Medical Convergence Research Institute and Center for Clinical Imaging Data Science, Yonsei University College of Medicine, 50-1 Yonsei-ro, Seodaemun-gu, Seoul 03722, South Korea

Published online: 04 April 2018

Table 1 Summary of training and test sets of the patients

	Training set	Test set
Numbers	5720	1018
Age	53.89 ± 17.26	54.68 ± 17.67
Sex Gender (M:F)	2449:3271	464:554

to evaluate the target region with decreased or increased field of view (FOV) and axial/coronal/sagittal contrast-enhanced images. The determination of MRI protocols is needed to acquire optimal image quality and is essential for a radiologist to read the MRI. This can be time consuming depending on the radiologist's expertise and experiences. Providing a best-practices recommendation for an MRI protocol could improve efficiency and decrease suboptimal or erroneous studies. However, the task of MRI protocol determinations may place additional burden on the radiologic environment. With a rapidly growing volume and diversity of MRI examinations, traditional MRI protocol determinations often need to be changed for efficiency and accuracy.

Deep learning is a form of machine learning whereby neural networks with multiple hidden layers are trained to perform a certain task [4] and has been applied to health and biomedical research with large data sets for predictive models [5, 6]. The flexibility and prowess of machine learning models has also been applied in medical imaging and radiology [4, 7]. Recently, the artificial intelligence in musculoskeletal MRI was applied to contrast determination using IBM Watson's natural language processing [8]. However, to our knowledge, no previous studies have been performed of MRI protocol determination in the radiologic environment. The purposes of this study are to (1) evaluate the feasibility of protocol determination with a convolutional neural network (CNN) classifier based on short-text classification and (2) evaluate the agreements by comparing the protocols determined by CNN with those determined by musculoskeletal radiologists.

Table 2 Details of the training set

Region	Tumor protocol	Routine protocol	Total
Cervical spine	52	281	333
Cervicothoracic spine	7	–	7
Thoracic spine	77	12	89
Thoracolumbar spine	34	15	49
Lumbar spine	163	918	1081
Whole spine	469	53	522
Pelvic bone	289	4	293
Sacroiliac joint	2	106	108
Brachial plexus	1	36	37
Chest wall	20	–	20
Sternum	26	–	26
Rib	44	–	44
Back	12	–	12
Shoulder	116	527	643
Upper arm	50	–	50
Elbow	33	74	107
Forearm	24	1	25
Wrist	15	109	124
Hand	46	5	51
Finger	26	8	34
Hip	64	75	139
Thigh	176	2	178
Knee	107	699	806
Lower leg	97	1	98
Ankle	6	252	258
Foot	83	24	107
Toe	4	–	4
Extremities	–	13	13
Total number	2043	3215	5258

Note: – represents no exam in the training set

Materials and Methods

Data Preparation

The database of hospital information system (HIS) was queried for lists of MRI examinations as well as referring

Table 3 Details of the test set

Region	Tumor protocol	Routine protocol	Total
Cervical spine	7	40	47
Thoracic spine	17	1	18
Thoracolumbar spine	3	9	12
Lumbar spine	32	171	203
Whole spine	107	8	115
Pelvic bone	75	0	75
Shoulder	37	172	209
Upper arm	11	–	11
Elbow	0	13	13
Wrist	0	26	26
Hand	12	0	12
Hip	21	7	28
Thigh	37	0	37
Knee	4	123	127
Lower leg	16	0	16
Ankle	0	47	47
Foot	21	1	22
Total number	400	618	1018

Note: – represents no exam in the test set. For no exam, but a training set, it is represented by 0

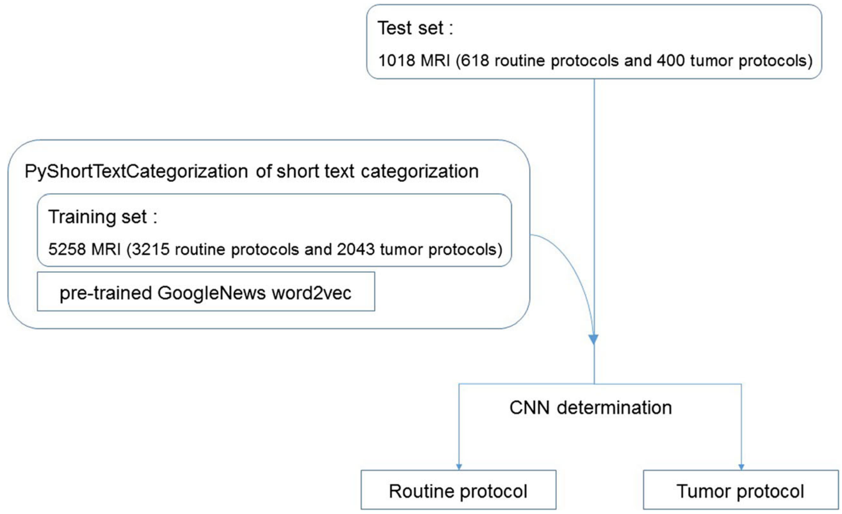

Fig. 1 Flowchart of the study. The study consisted of 5258 musculoskeletal MRI examinations for the training dataset and 1018 consecutive musculoskeletal MRI examinations for the test dataset

department, patient age, and patient gender. These were exported to a local workstation for analysis. The summary of the training and test sets are given in Table 1.

For the training dataset, 5258 musculoskeletal MRI examinations were collected from the institution's electronic medical records (EMR) between January and December 2016. The patient ages, genders, referring departments, examination names, and use of contrast agent that matched each test were collected. The training set of the MRI examinations consisted of various regions in each protocol (Table 2). Among the 5258 MRI examinations, 2043 and 3215 examinations were performed with the tumor/infection and routine protocols, respectively. We simplified the protocol into two protocols: routine protocol and tumor/infection protocol. In the spine, the disc level oriented imaging protocol (i.e., axial images for the intervertebral disc space) was recorded as a routine protocol, and the vertebral body (or continuous axial sections) was recorded as a tumor protocol. For the joint MRI excluding the spine, the exams were classified as a routine joint imaging for ligaments or tendons and tumor/infection imaging focusing on soft tissue infections, septic/pyogenic arthritis, and various bone and soft tissue tumors. In the test dataset, 1050 consecutive musculoskeletal MRI examinations were collected from the institution's EMR between January and February 2017, and less than 10 exams were excluded. Finally, 1018 exams were included (Table 3). The study protocol was reviewed by an institutional review board, and the informed consents were waived for this type of study.

Pre-processing of Text Data

The data format was prepared using the guidelines of PyShortTextCategorization (available at http://shorttext.readthedocs.io/en/latest/tutorial_dataprep.html#user-provided-training-data). The subject was the routine or tumor protocols. The contents were the word combinations of referring department, region, contrast media (or not), gender, and age (e.g., OS patient 56-year-old male hip MRI contrast) (Fig. 1). The pre-processing was performed using Microsoft Excel (Microsoft, Redmond, WA) by one musculoskeletal radiologist. The protocols of all examinations were divided into routine and tumor protocols.

Convolutional Neural Network for Sentence Classification

In this study, the short-text categorization was performed using PyShortTextCategorization of short-text categorization written in Python (available at https://github.com/stephenhky/PyShortTextCategorization) as an implementation of the CNN

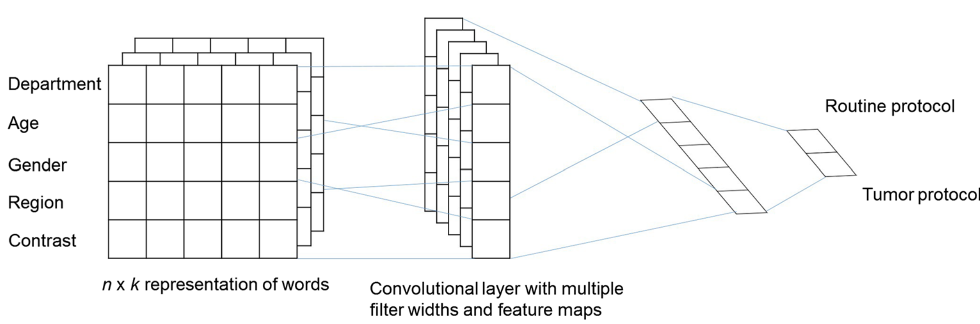

Fig. 2 Model architecture with two channels for the routine or tumor protocols of the musculoskeletal MRI

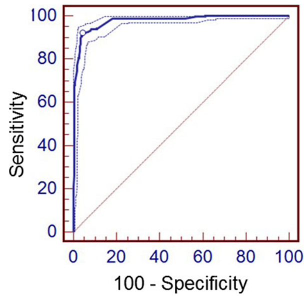

Fig. 3 Receiver operating characteristic (ROC) curve (solid line) with 95% confidence interval (dashed line) of CNN determinations

(Fig. 2). This module is a collection of algorithms for the multi-class classification to short texts using Python, Keras, and Tensorflow as the backend.

The deep neural networks with Word-Embedding were performed with neural network algorithms of supervised short text using PyShortTextCategorization of short-text categorization. Each class label contained short sentences and each token was converted to an embedded vector given by a pre-trained word-embedding model (Word2Vec model of Google news vectors available at https://github.com/mmihaltz/word2vec-GoogleNews-vectors/blob/master/GoogleNews-vectors-negative300.bin.gz). The input short sentences in the training set were processed by a matrix for data training. The predictive input short sentences in the test set were converted to unit vectors in the same way. The scores were calculated according to the trained neural network model and the higher scores of MRI protocols (routine or tumor protocols) were reported. The neural network for supervised learning was used as the convolutional neural network (ConvNet), as demonstrated in Ref. [9].

Training for Protocol Determination Model

The prepared 5258 MRI examinations were included with an epoch number of 100 during training. The CNN Embedded vector classifier was used [9] with the Word2Vec of Google news vectors. The Word2Vec model is a pre-trained Google News model of 3 billion running words, consisting of 3 million 300-dimension English word vectors. A trained model of ConvNet was made that was a supervised learning of the convolutional neural network, as demonstrated in Ref. [9]: CNNWordEmbed(nb_labels, wvmodel = pre-trained GoogleNews Word2Vec model, nb_filters=1200, n_gram=2, maxlen=15, vecsize=100, cnn_dropout=0.0, final_activation='softmax', dense_wl2reg=0.0, dense_bl2reg=0.0, optimizer='adam'', with_gensim=false).

Test of Protocol Determination Model

To validate the trained model, the consecutive 1018 musculoskeletal MRI examinations were tested with the trained model, and the results were output as routine or tumor protocols. The

Table 4 Accuracies of the test set

Region	Exam number	Tumor protocol		Routine protocol		Accuracies (%)
		Correct	Incorrect	Correct	Incorrect	
Cervical spine	47	6	1	38	2	93.62
Thoracic spine	18	16	1	1	0	94.44
Thoracolumbar spine	12	2	1	8	1	83.3
Lumbar spine	203	27	5	169	2	96.56
Whole spine	115	104	3	8	0	97.39
Pelvic bone	75	75	0	0	0	100
Shoulder	209	32	5	172	0	97.61
Upper arm	11	11	0	0	0	100
Elbow	13	0	0	10	3	76.92
Wrist	26	0	0	26	0	100
Hand	12	11	1	0	0	91.67
Hip	28	19	2	7	0	92.86
Thigh	37	34	3	0	0	91.89
Knee	127	3	1	105	18	85.04
Lower leg	16	16	0	0	0	100
Ankle	47	0	0	42	5	89.36
Foot	22	17	4	0	1	77.27
Total number	1018	373	27	582	36	Overall, 94.2

Note: * indicates that the test set consisted of > 10 cases

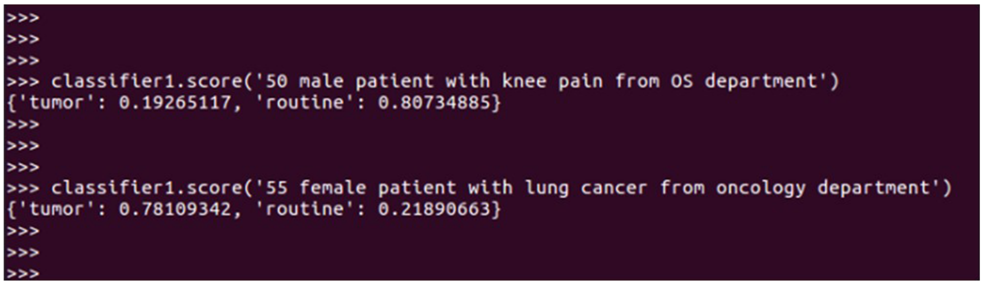

Fig. 4 Representative screenshot of the python shell. In the case of the '50 male patient with knee pain from the OS department' as input text, the recommended protocol was the routine protocol with a score of 0.80734885. In other case of the '55 female patient with lung cancer from the oncology department,' the recommended protocol was the tumor protocol with a score of 0.78109342. OS, orthopedic surgery

proposed protocol was determined by the optimal cut-off score from the ROC curve. The accuracies were evaluated with a radiologist-confirmed protocol as the reference protocol. The confusion matrix and Kappa value were calculated for the statistical analysis.

The resulted protocols were recorded according to the highest probability scores. All codes were implemented with python 2.7, based on the Keras (version 2.0.8, available at https://keras.io) and Google TensorFlow library (version 1.3. 0, available at https://www.tensorflow.org) with an Ubuntu Linux operating system (version 16.04.3). All codes were run on a quad-core CPU (i5-2500 3.3 GHz, Intel, Santa Clara, CA) with one NVIDIA GeForce GTX 1060 GPU (NVIDIA Corp., Santa Clara, CA).

Statistical Analyses

The determination performance of the CNN model was evaluated by using the receiver operating characteristic (ROC) curve. Statistical analyses were performed using MedCalc Statistical Software version 11.2 (MedCalc Software BVBA, Ostend, Belgium).

Results

The training time was approximately 10 min for the 5258 examinations. The results of the protocol determinations were output as probability scores (Figs. 3 and 4). The agreement between the CNN and radiologist's determinations was 0.88 of the Kappa value (Table 4).

ROC analysis revealed an optimal cut-off score of 0.5067 for protocol determinations between routine protocols and tumor protocols. This cut-off showed a sensitivity of 95.60% and specificity of 92.10%. The area under curve (AUC) was 0.977 (9% confidence interval, 0.966–0.985). The overall accuracy was 94.20% for the ConvNet model ($n = 915/1018$). All MRI protocols were correct in the pelvic bone, upper arm, wrist, and lower leg MRIs. The details are provided in Table 5.

Discussion

The application of deep learning to the radiologic field is a new area of radiology [10]. Although there have been some unfounded controversies in the applications of machine learning due to their complexity of various imaging data [2, 11], the machine learning technique can be applied to non-imaging radiological tasks in the radiologic workflow. Besides imaging data, there are many phrases and words in the workflow of the radiologic reports and radiology information system. The deep learning in radiology could be applied to the text data of radiology, such as radiologic protocol determination. The determination of the MRI protocol is an essential process in radiologic workflow. The most appropriate protocol determination is essential for accurate radiologic interpretations and definitive radiologic decisions. However, this is time consuming and can result in a work burden for radiologists. Machine learning have potential uses in radiologic tasks; deep learning is useful for image analytics [12, 13] and somewhat simple tasks, such as contrast study determination [8], that are more suited to computer-aided tasks.

In our study, the text classifier module written by Python was utilized; the package shorttext facilitates supervised learning for short-text categorization. This module can be

Table 5 Confusion matrix to demonstrate the protocol of CNNs' suggestion compared to that of the radiologists' decision

	Routine [reference]	Tumor [reference]	
Routine [CNN]	586	32	PPV, 94.82%
Tumor [CNN]	27	373	NPV, 93.25%
	Sensitivity, 95.60%	Specificity, 92.10%	Accuracy, 94.20%

PPV positive predict value, *NPV* negative predict value downloaded from the GitHub website (available at https://github.com/stephenhky/PyShortTextCategorization). The application of deep learning to text classification in radiologic protocol determination was feasible by using the resultant scores. These results showed excellent agreement between the CNN and the radiologist's determinations with an optimal cut-off score of 0.5067 for protocol determinations. It will be improved further, providing higher accuracy in the results with a number of training sets. This module could be integrated into radiologic services, such as Radiologic Chatbot, with the following functionalities; (1) a busy radiologist could confirm the pre-determined radiologic protocol and (2) radiotechnologists could use chatbot to recheck the MRI protocols before patient scanning.

In the field of radiology, machine learning can be applied not only for image analysis but also for patient safety, improving work efficiency, and optimization of radiology workflow. One of the possible applications is to minimize human errors in radiology, and this includes laterality errors of radiologic reports [14–16]. These error-minimizing warning tools could be powered by applying machine learning or deep learning [17]. In terms of improving efficiency in the field of radiology, there are potential applications. To maximize efficiency of radiologic imaging workflow, radiologists can utilize CNN protocol determination as a template before doctor's protocol confirmation. Junior radiologic residents can use it when the senior radiologist on duty is temporarily not available. In the viewpoint of radiologic imaging technologists, MRI radiotechnologists can start MRI imaging by using CNN protocol determination in order to improve efficiency. Regarding patient imaging recalls, approximately 20% of the causes for patient recall is protocol error in the outpatient department [18]. The application of CNN protocol determination can be utilized as double check, which leads to a reduction in the number of patient recalls for MRI, resulting in reduced cost and time loss as well as enhanced patient satisfaction.

There are several limitations to this study. Firstly, the simplified protocols were trained as routine versus tumor protocols (the disc protocol versus continuous body in spine MRI). There are customized protocols for certain clinical settings in radiology. In this feasibility study, the customized protocol was not included. However, an advanced machine learning algorithm is expected to learn various radiologic protocols that could be applied in more complex radiologic protocols. The clinical application should be used with caution until it is verified with more data collected in the future. When the imaging is started with this protocol determination module alone, the radiotechnologist should get a physician's confirmation before ending the imaging. Secondly, the patient EMR data was not utilized because of difficulties in the indexing and categorization of the EMR data due to its heterogeneity nature. Advancement of the EMR data application could be possible with advancements in natural language processing.

In conclusion, deep-learning-based convolutional neural networks can be clinically utilized to determine musculoskeletal MRI protocols. These results support using deep learning to assist radiologists in their work by providing timely and highly accurate protocol determinations that only require rapid confirmation. Furthermore, CNN-based text learning and applications could be extended to other radiologic task besides image interpretations, facilitating an improved work performance for radiologists.

Acknowledgements The authors would like to thank Kwan-Yuet Ho, PhD (Developer of ShortText) for his help with the installation and application.

Funding This work was supported by a National Research Foundation (NRF) grant funded by the Korea Government, Ministry of Science and ICT (MSIP, 2018R1A2B6009076).

Compliance with Ethical Standards

Conflict of Interest The authors declare that they have no conflict of interest.

References

1. ACR Practice parameter for performing and interpreting magnetic resonance imaging (MRI). Available at: https://www.acr.org/~/media/EB54F56780AC4C6994B77078AA1D6612.pdf. Accessed Nov 1, 2017
2. Edelstein WA, Mahesh M, Carrino JA: MRI: time is dose—and money and versatility. J Am Coll Radiol 7:650–652, 2010
3. Mekle R, Wu EX, Meckel S, Wetzel SG, Scheffler K: Combo acquisitions: balancing scan time reduction and image quality. Magn Reson Med 55:1093–1105, 2006
4. LeCun Y, Bengio Y, Hinton G: Deep learning. Nature 521:436–444, 2015
5. Ayaru L, Ypsilantis PP, Nanapragasam A, Choi RCH, Thillanathan A, Min-Ho L, Montana G: Prediction of outcome in acute lower gastrointestinal bleeding using gradient boosting. PLoS One 10: e0132485, 2015
6. Ogutu JO, Schulz-Streeck T, Piepho HP: Genomic selection using regularized linear regression models: ridge regression, lasso, elastic net and their extensions. BMC Proc 6(Suppl 2):S10, 2012
7. Forsberg D, Sjoblom E, Sunshine JL: Detection and labeling of vertebrae in MR images using deep learning with clinical annotations as training data. J Digit Imaging 30:406–412, 2017
8. Trivedi H, Mesterhazy J, Laguna B, Vu T, Sohn JH: Automatic determination of the need for intravenous contrast in musculoskeletal MRI examinations using IBM Watson's natural language processing algorithm. J Digit Imaging. https://doi.org/10.1007/s10278-017-0021-3, 2017
9. Kim Y: Convolutional neural networks for sentence classification. ArXiv e-prints abs/1408.5882, 2014
10. 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 15:504–508, 2018
11. Kim Y: Convolutional neural networks for sentence classification. CoRR abs/1408.5882 %U http://arxiv.org/abs/1408.5882, 2014

12. Pereira S, Pinto A, Alves V, Silva CA: Brain tumor segmentation using convolutional neural networks in MRI images. IEEE Trans Med Imaging 35:1240–1251, 2016
13. Setio AAA, Ciompi F, Litjens G, Gerke P, Jacobs C, van Riel SJ, Wille MMW, Naqibullah M, Sanchez CI, van Ginneken B: Pulmonary nodule detection in CT images: false positive reduction using multi-view convolutional networks. IEEE Trans Med Imaging 35:1160–1169, 2016
14. Sangwaiya MJ, Saini S, Blake MA, Dreyer KJ, Kalra MK: Errare humanum est: frequency of laterality errors in radiology reports. AJR Am J Roentgenol 192:W239–W244, 2009
15. Luetmer MT, Hunt CH, McDonald RJ, Bartholmai BJ, Kallmes DF: Laterality errors in radiology reports generated with and without voice recognition software: frequency and clinical significance. J Am Coll Radiol 10:538–543, 2013
16. Lee YH, Yang J, Suh JS: Detection and correction of laterality errors in radiology reports. J Digit Imaging 28:412–416, 2015
17. Siegel E. It's the effect size, stupid. Will Computers Replace Radiologists? Debunking the hype of AI. **Available at:** http://www.carestream.com/blog/2016/11/01/why-computers-cant-replace-radiologists. Accessed Apr 1, 2017
18. Gyftopoulos S, Kim D, Aaltonen E, Horwitz LI: Patient recall imaging in the ambulatory setting. Am J Roentgenol 206:787–791, 2016