Utilizing Deep Learning to Develop a Conversational Support Tool for Clinicians: Creating a Radiology Chatbot

Kevin F. Seals, MD, Ronald Reagan Medical Center, UCLA; Johnathan Hedge, MD; Robert Suh, MD; Edward Zaragoza, MD; Dieter Enzmann, MD

Background

Radiologists provide clinicians with significant value not directly related to the interpretation of medical images. This includes guiding the optimal imaging study for a particular clinical scenario, managing contrast administration in renal failure, determining which metallic implants are MRI-appropriate, and various other skills. Although some of these tasks are relatively formulaic, with established national or institutional guidelines, few suitable informational tools exist for clinicians beyond direct conversation with human radiologists. While discourse between clinicians and radiologists is critical to optimal patient care, repeat phone conversations regarding redundant, easily automated issues is both distracting for radiologists and unnecessarily time-consuming for clinicians.

We therefore used cutting-edge machine learning to create a conversational virtual assistant—a chatbot—that can assist clinicians with needs related to both diagnostic and interventional radiology. In creating this application, we aimed to encapsulate much of the value of speaking with a human radiologist in the context of an intelligent, conversational virtual radiology assistant accessible via mobile devices. Using a deep learning approach, we are able to create a smart, powerful tool that both understands complex user inputs and continually grows smarter and more powerful with continued use (1, 2).

Evaluation

We began the project with an online survey of clinicians at our institution regarding the non-interpretive radiology skills they find most valuable—that is, the skills that can be implemented in a chatbot that can intelligently converse and present data but not interpret medical images. We asked clinicians to list the three non-interpretive radiology skills they most commonly utilize, receiving responses from 64 clinicians, including both internists and subspecialists. The responses allowed us to identify 32 categories of non-interpretive radiology skill for implementation in the chatbot. Common responses included determining the optimal imaging study for a particular clinical scenario (72% of respondents), assessing if intravenous contrast is necessary and/or feasible at a given creatinine level (53%), and identifying if a metallic implant is MRI-compatible (20%).

We developed the RadChat iOS application to provide clinicians with rapid, automated information related to these non-interpretive skill categories. The application was programmed using the XCode Software Development Kit (SDK) and the Swift programming language for use on mobile devices. An interface similar to text messaging on a cellular phone or messaging via a chat application (WhatsApp, Facebook Messenger, etc.) was created in XCode. To understand user inputs, a neural network was implemented using the Natural Language Classifier of the Watson Application Programming Interface (API). The neural network was trained with more than 2000 data points created by a team of physicians, including both radiologists and clinicians.
Users interact with the application by entering a line of text regarding their medical need in a conversational interface that is similar to text-based communication with a human radiologist (Figure 1). The Natural Language Classifier then processes the user input and generates the probability that it relates to a given information category. Each information category corresponds to a particular non-interpretive radiology skill, for example the MRI safety of metallic objects, contrast administration in renal failure, or the optimal imaging protocol for pulmonary embolism. Categories found to be above a certain threshold probability of matching the user query are presented, and the user either selects a category of interest or specifies that none of the provided categories are relevant. All user queries and their corresponding category selections are stored in an online Firebase database to facilitate further training of the neural network. Using a Python algorithm, a JavaScript Object Notation (JSON) representation of this database can be converted to Comma Separated Values (CSV) format and used to further enhance the neural network of the Natural Language Classifier.

When the user selects a particular information category, he or she is taken to the corresponding information module. Information modules represent custom programs providing the user with data related to the topic of interest. These custom programs range from a screen presenting a particular website or image to an interactive Well’s score calculator to a searchable database of phone numbers in the hospital. The relationship between user queries, neural network processing, and information modules is summarized in Figure 2. With the tap of a button, the user can leave a particular information module and resume chatting to enter additional queries.
Discussion

A wide variety of useful information modules are currently implemented in the application, which is available to a number of test users at our institution. Guidance regarding the optimal imaging study for a particular clinical scenario has been of particular interest to clinicians. There is a separate information module for each American College of Radiology (ACR) Appropriateness Criteria category. If the neural network determines that the user is interested in the optimal imaging study for a clinical issue that has a corresponding ACR Appropriateness Criteria category, the user is provided with a link to the relevant section of the ACR website. The website conveniently opens in the application for rapid review (Figure 3). Of note, the neural network is trained using numerous keywords that are associated with each information module. The trained neural network is thus smart enough to understand a wide variety of user inputs that may suggest a particular clinical issue. For example, in the case of pulmonary embolism, terms such as “deep venous thrombosis,” “positive D-Dimer,” and “chest pain in the setting of malignancy,” are all appropriate inputs associated with the pulmonary embolism category. The neural network search algorithm thus represents a powerful approach that can integrate large amounts of medical knowledge to understand user inputs.

Figure 3

Welcome to RadChat! I am your virtual radiology assistant.

How can I help you today? To learn more about what I can do, just ask!

rq pain

Here are some information categories that may be relevant to you. If a category seems relevant, tap to learn more.

- Right Lower Quadrant Pain, Suspected Appendicitis
- Acute (Nonlocalized) Abdominal Pain and Fever or Suspected Abdominal Abscess
- None of These Categories Are Relevant

There are many additional useful information modules. If the user has a question regarding the optimal follow-up protocol for an incidental finding, he or she is taken to an algorithmic flowchart derived from the literature, complete with pinch-to-zoom functionality. This flowchart can assist the clinician in formulating an appropriate, evidence-based follow-up imaging protocol. Should a user need the phone number for a particular radiology-related area in the hospital, such as a reading room or CT scanner, he or she is provided with a searchable database of hospital phone numbers (Figure 4). If the user is attempting to determine the appropriate imaging protocol for suspected pulmonary embolism, the application offers a Well’s score calculator to help guide clinical management (Figure 5).
By providing these information categories and many others, each offering radiologist curated and validated information that is rapidly accessible, we have created a powerful tool helping clinicians interface with radiology. Even a simple website link has significant value to clinicians, as that link is curated and validated by a team of academic radiologists through careful analysis of the available radiology literature. The alternative, performing a web search for this information, would inevitably yield innumerable results of variable quality and scientific legitimacy. Careful radiologist curation is thus a fundamental component of the RadChat application and its value to clinicians.

We continue to develop concepts for new information modules through the analysis of queries from test users, which are automatically stored in an online database and regularly reviewed by our team. RadChat will grow progressively more powerful as we expand to a larger number of clinician users, with user data continually refining the abilities of our neural network and generating new ideas for application improvement.

Conclusion

We have successfully created a virtual radiology assistant that uses deep learning to assist clinicians with a wide variety of radiology needs. We believe this clinical tool will be valuable to both clinicians, who will be able to more rapidly access useful information, and radiologists, who will have more uninterrupted time to focus on caring for their patients. The application will naturally grow more powerful with continued clinical use, as the analysis of user inputs identifies useful new information categories and user inputs are used as training data to make the neural network progressively more powerful and intelligent.
References


Keywords

deep learning, machine learning, natural language processing, decision support, chatbot