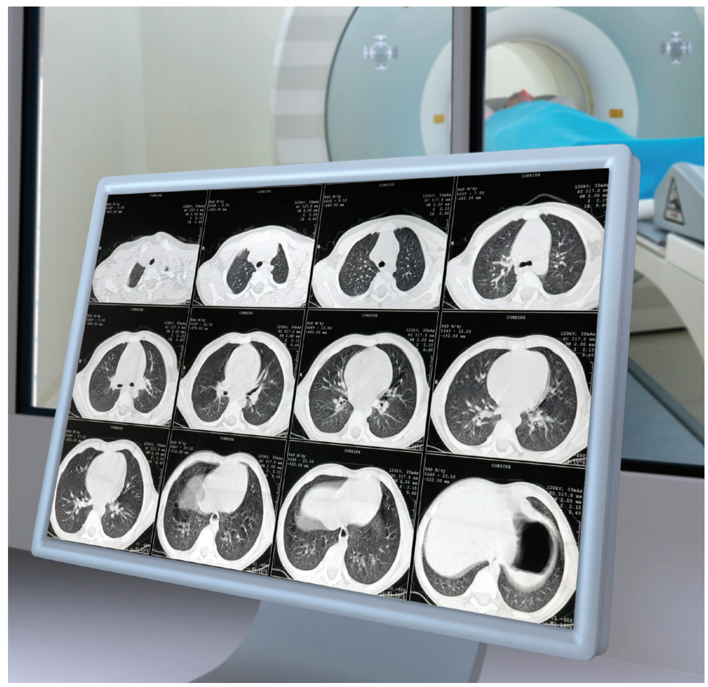
Shedding Light on Neural Networks ... as they Shed Light on Indeterminate Lung Nodules

Denise R. Aberle, MD
Professor of Radiology and Bioengineering
Vice Chair for Research
Department of Radiological Sciences
David Geffen School of Medicine at UCLA

"Among the most important questions that radiologists are looking to answer with artificial intelligence is whether a particular lesion — in any organ in the body — is a cancer or is not a cancer," says Denise Aberle, MD, professor of radiology and bioengineering and vice chair for radiological sciences research at David Geffen School of Medicine at UCLA. "Secondly, if we know that we have a cancer, we're looking to artificial intelligence (AI) to help distinguish between indolent, slow-growing cancers and those that are aggressive, enabling more individualized management of patients. Various forms of AI, including deep learning with neural networks, are making inroads into these challenges. For example, researchers recently used a deep learning algorithm to distinguish lung cancer in nodules seen on low-dose computed tomography (CT) scans. For those cases where prior CT imaging was not available, the algorithm outperformed a panel of six radiologists, reducing false negative results by 5 percent and false positives by 11 percent.

Within AI, there are a number of machine learning approaches, including random forests, support vector machines and deep learning with convolutional neural networks (CNN). "The excitement — actually the frenzy — in imaging today is the use of convolutional neural networks to answer questions for us based on the extraction of information beyond what humans can see in the image, says Dr. Aberle."



Challenges of developing convolutional neural networks

While convolutional neural networks can be very powerful tools, there are challenges in developing these systems. First, they typically require large data sets in order to be able to perform well. Second, in order for the neural network to learn from the training data, the images must be well annotated to contain all variables relevant to each individual scan. Of critical importance, the training data must be representative of the population to which the model will be applied.

Training the CNN can require data sets sufficiently large that they exceed what is available at a single institution. "In the case of training a neural network to learn what is a lung cancer and what is not, this may involve thousands of images," says Dr. Aberle.

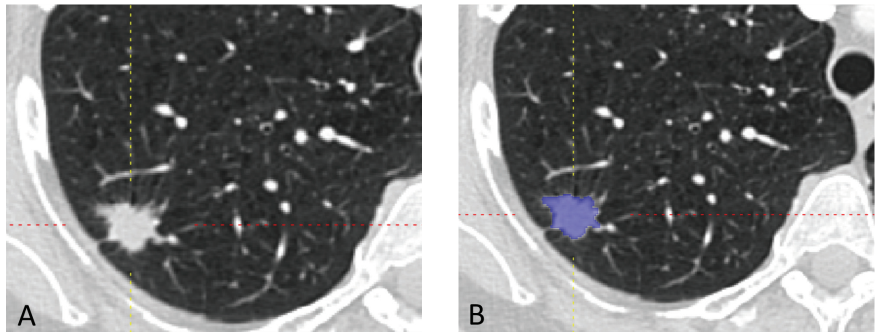
In addition, annotation of the data set is very labor intensive. A vast amount of human effort is required to annotate the very large data sets required for training convolutional neural networks. The largest annotated data set currently available for lung cancer is from the National Lung Screening Trial (NLST), for which Dr. Aberle was principal investigator. The NLST was a large, multi-center clinical trial that enrolled 53,000 patients, half of whom had low-dose chest CT scans annually for three years. The NLST data has already been used to train various models approved for clinical use in the European Union.

Another challenge of training neural networks is that models perform their best when data acquisition and reconstructions parameters for the training images match those of the clinical images to which the model will be applied. Current CT imaging practices are highly variable in terms of parameters like slice thickness, beam energy, dose and image convolution kernel.

"The excitement — actually the frenzy — in imaging today is the use of convolutional neural networks to answer questions for us based on the extraction of information beyond what humans can see in the image"
— Denise Aberle, MD

A current concern in using convolutional neural networks in patient care is that the reasoning of the network classification is often not known; the analytical queues that inform the classification task remain hidden within many layers of the computer model.

Example of a lung nodule on computed tomography (A) with automated segmentation of the nodule (B). This segmented region of interest can then be analyzed for features associated with a lung cancer.



Computer models are most reliable when asked to interpret data that is similar to the data used in their training.

Artificial intelligence algorithms developed for clinical practice must be carefully tested and validated to ensure their soundness as decision-support tools. Testing and validation should each employ an independent data set that was not previously fed through the model. “These are the types of vigorous validation and performance assessments we need before we begin to use these AI tools in clinical practice,” states Dr. Aberle. “We’re not far off, but we’re currently still a ways away.”

Augmenting the performance of neural networks

While convolutional neural networks rely on very large sets of well annotated data to learn to interpret image data, some of the other artificial intelligence approaches, such as random forests and support vector machines, can develop good functionality with smaller training data sets. Moreover, we are increasingly seeing ensemble approaches in which analysis pipelines combine multiple strategies to exploit the efficiencies, or the unique advantages, of different approaches in a single system.

Similarly, a model’s performance can be improved when more than one kind of data is used to train the system. “If I’m training a CNN using image data from CT scans, it may reach a certain level of performance and begin to plateau unless I’m able to feed it many more hundreds or thousands of annotated images,” explains Dr. Aberle. “An alternative approach is to introduce other variables — non-imaging variables — that also influence the classification task.” These other variables, if they represent orthogonal data (meaning that there is little or no overlap with imaging data), provide a different form of training that is complementary to the imaging data. One example might be demographic variables — such as race, sex, age or region of the country — that have been shown to be associated with the clinical condition being investigated.

Artificial intelligence and lung nodules

Under multiple NIH grants, Dr. Aberle is applying CNN to the diagnosis of indeterminate lung nodules. While the majority of lung nodules detected on CT scans are benign — the result of scarring, inflammation or infection — nodules can also be early cancers or cancers that have metastasized from elsewhere in the body. In collaboration with colleagues in the fields of computer science, medical informatics and molecular biology, Dr. Aberle is applying CNN to CT images to help diagnose when indeterminate lung nodules are cancer and to help distinguish indolent growths from those that are more aggressive.

The team is hoping to enhance their ability to diagnose these indeterminate nodules by adding molecular biomarker data to the CT image data. By collecting readily accessible biospecimens — either blood specimens, saliva, or cells easily collected from inside the nose or mouth — they are working to identify useful biomarkers. “Our hope is that the combination of molecular and imaging data will be able to tell us that an individual has lung cancer much earlier so we can intercept the cancer early on and maximize the likelihood of long-term cure,” explains Dr. Aberle.

In other research related to neural networks, Dr. Aberle hopes to shed light on how deep learning algorithms derive their conclusions. A current concern in using convolutional neural networks in patient care is that the reasoning of the network classification is often not known; the analytical queues that inform the classification task remain hidden within many layers of the computer model. “We think being able to follow the process will be important to the ultimate implementation of these machine learning approaches,” says Dr. Aberle. “If the system is not understandable to humans, we are less comfortable implementing them into direct patient management.” she says. R