Going deep: Artificial intelligence improves accuracy of breast ultrasound diagnoses
5 April 2021

Ultrasound is an invaluable diagnostic tool for the early detection of breast cancer, but the classification of lesions is sometimes challenging and time consuming. Could artificial intelligence hold the answer to solving these problems? Credit: Chinese Medical Journal

In 2020, the International Agency for Research on Cancer of the World Health Organization stated that breast cancer accounts for most cancer morbidities and mortalities in women worldwide. This alarming statistic not only necessitates newer methods for the early diagnosis of breast cancer, but also brings to light the importance of risk prediction of the occurrence and development of this disease. Ultrasound is an effective and noninvasive diagnostic procedure that truly saves lives; however, it is sometimes difficult for ultrasonologists to distinguish between malignant tumors and other types of benign growths. In particular, in China, breast masses are classified into four categories: benign tumors, malignant tumors, inflammatory masses, and adenosis (enlargement of milk-producing glands). When a benign breast mass is misdiagnosed as a malignant tumor, a biopsy usually follows, which puts the patient at unnecessary risk. The correct interpretation of ultrasound images is made even harder when factoring in the large workload of medical specialists.

Could deep learning algorithms be the solution to this conundrum? Professor Wen He (Beijing Tian Tan Hospital, Capital Medical University, China) thinks so. "Artificial intelligence is good at identifying complex patterns in images and quantifying information that humans have difficulty detecting, thereby complementing clinical decision making," he states. Although much progress has been made in the integration of deep learning algorithms into medical image analysis, most studies in breast ultrasound deal exclusively with the differentiation of malignant and benign diagnoses. In other words, existing approaches do not try to categorize breast masses into the four abovementioned categories.

To tackle this limitation, Dr. He, in collaboration with scientists from 13 hospitals in China, conducted the largest multicenter study on breast ultrasound yet in an attempt to train convolutional neural networks (CNNs) to classify ultrasound images. As detailed in their paper published in Chinese Medical Journal, the scientists collected 15,648 images from 3,623 patients and used half of them to train and the other half to test three different CNN models. The first model only used 2D ultrasound intensity images as input, whereas the second model also included color flow Doppler images, which provide information on blood flow surrounding breast lesions. The third model further added pulsed wave Doppler images, which provide spectral information over a specific area within the lesions.

Each CNN consisted of two modules. The first one, the detection module, contained two main submodules whose overall task was to determine the position and size of the breast lesion in the original 2D ultrasound image. The second module, the classification module, received only the extracted portion from the ultrasound images containing the detected lesion. The output layer contained four categories corresponding to each of the four classifications of breast masses commonly
used in China.

First, the scientists checked which of the three models performed better. The accuracies were similar and around 88%, but the second model including 2D images and color flow Doppler data performed slightly better than the other two. The reason the pulsed wave Doppler data did not contribute positively to performance may be that few pulsed wave images were available in the overall dataset. Then, researchers checked if differences in tumor size caused differences in performance. While larger lesions resulted in increased accuracy in benign tumors, size did not appear to have an effect on accuracy when detecting malignancies. Finally, the scientists put one of their CNN models to the test by comparing its performance to that of 37 experienced ultrasonologists using a set of 50 randomly selected images. The results were vastly in favor of the CNN in all regards, as Dr. He remarks: "The accuracy of the CNN model was 89.2%, with a processing time of less than two seconds. In contrast, the average accuracy of the ultrasonologists was 30%, with an average time of 314 seconds."

This study clearly showcases the capabilities of deep learning algorithms as complementary tools for the diagnosis of breast lesions through ultrasound. Moreover, unlike previous studies, the researchers included data obtained using ultrasound equipment from different manufacturers, which hints at the remarkable applicability of the trained CNN models regardless of the ultrasound devices present at each hospital. In the future, the integration of artificial intelligence into diagnostic procedures with ultrasound could speed up the early detection of cancer. It would also bring about other benefits, as Dr. He explains: "Because CNN models do not require any type of special equipment, their diagnostic recommendations could reduce predetermined biopsies, simplify the workload of ultrasonologists, and enable targeted and refined treatment."

Let us hope artificial intelligence soon finds a home in ultrasound image diagnostics so doctors can work smarter, not harder.


Provided by Chinese Medical Journal