THYROID CANCER

Sonographic diagnosis of thyroid cancer with support of AI

Frederik Verburg and Christoph Reiners

Thyroid ultrasonography is an important element of clinical thyroid diagnostics. Unfortunately, the results of this technique can vary based on the skill and experience of the operator. A new study suggests that assessment of ultrasound images using artificial intelligence has similar sensitivity and improved specificity compared with the judgement of experienced radiologists.

Refers to Li, X. et al. Diagnosis of thyroid cancer using deep convolutional neural network models applied to sonographic images: a retrospective, multicohort, diagnostic study. *Lancet Oncol.* **20**, 193–201 (2019).

Xiangchun Li and colleagues recently published the results of a promising retrospective, multicohort diagnostic study that used artificial intelligence (AI) with deep convolutional neural network models applied to sonographic images for the diagnosis of thyroid cancer[1]. The cohorts recruited for training ($n = 42{,}952$) and validation ($n = 2{,}692$) of the neural network models comprised a large number of patients from three Chinese hospitals and included a training cohort and a validation cohort of patients with thyroid cancer. According to the authors, the AI approach, which was validated with the data from the three hospitals, resulted in slightly inferior sensitivity and substantially improved specificity compared with the judgement of three experienced radiologists using the Thyroid Imaging, Reporting and Data System (TI-RADS) classification, in the version as defined by the American College of Radiology, at each participating hospital (TABLE 1).

> the AI approach … improved specificity compared with the judgement of three experienced radiologists

Presently there are various ultrasonography techniques available. The basis of thyroid examination is still conventional greyscale imaging, amended by more advanced technologies such as colour Doppler ultrasonography, elastography (ultrasonography elasticity imaging) and 3D imaging. These techniques improve — independently from AI technologies — the detection of thyroid nodules and help physicians differentiate benign nodules from malignant nodules on the basis of certain features (mostly a combination of these features). The implementation of high-resolution colour Doppler microvascular imaging results in receiver operating characteristics for the diagnosis of malignant nodules that are markedly superior to greyscale imaging alone[2]. Elastography has the potential to improve the identification of thyroid malignancies as well[2]. 3D image acquisition enables the capture and storage of data from a complete sweep with the ultrasonography probe over the whole neck[3]. Such 3D volume scans are ideal for the training and validation of AI technologies for thyroid imaging.

According to a meta-analysis of the available literature, the Thyroid Imaging Reporting and Data System (TI-RADS) reporting scheme for thyroid ultrasonography[4] used for validation by Li and colleagues[1] discriminates moderately well between benign and malignant nodules with a pooled sensitivity of 75% and a pooled specificity of 69%[5]. The inclusion of additional parameters can improve the accuracy of these existing reporting schemes. For example, the specificity of TI-RADS for selecting patients who require further diagnostics in the form of fine-needle aspiration biopsy can be increased from 20% to 47% by the inclusion of elastography parameters without a decrease in sensitivity[6]. When these data are compared with the aforementioned meta-analysis[5], the low specificity of TI-RADS[6] seems to be confusing. The reason for this confusion is that it is often very difficult to properly classify results of such studies as the cohorts included are usually small and selective, meaning that the reference standards to ascertain thyroid cancer (cytology or, better, histology) are missing. Furthermore, the applied ultrasonographic techniques are not always described sufficiently and 2D ultrasonography is an operator-dependent technique, meaning that the results will depend on the skills of ultrasonographers, which are very difficult to qualify or quantify objectively. For the purpose of AI-based assessment of the malignant potential of a thyroid nodule, the operator-dependent recording of relevant areas of a nodule using 2D transverse or axial images can be replaced by the 3D recording of an entire thyroid lobe in a single volume, thus also completely encompassing any relevant nodule.

AI and machine learning approaches, which include deep learning or neural networks, have the potential to improve image segmentation and, in the case of thyroid ultrasonography, volume definition of the whole gland and of focal lesions[7]. More importantly, the data derived with these technologies can achieve a high accuracy of up to 98% in discriminating between malignant and benign lesions[8]. According to the first results derived with these technologies, the sensitivity for discrimination of malignant thyroid nodules using AI might be equal to that of experienced radiologists, whereas the specificity of the AI approach might not yet be satisfactory[9] or — on the contrary — be surprisingly better than the results achievable with conventional visual image analysis[1,10].

A weakness of AI studies thus far is that non-ultrasound clinical information is largely neglected. However, it is possible that information, such as lab results, could allow for further improvement of diagnostic AI assessment. The current confusing situation demands for well-controlled studies using

Table 1 | **Sensitivity and specificity for US classification of thyroid cancer**

Hospital	Artificial intelligence[a]	Radiologists (TI-RADS)[b]	Sensitivity of artificial intelligence (lower than radiologists; significance)	Specificity of artificial intelligence (higher than radiologists; significance)
Tianjin Hospital (cases with cancer $n = 555$; controls $n = 563$)	Sensitivity = 93.4%; specificity = 86.1%	Sensitivity = 96.9%; specificity = 59.4%	$P < 0.003$	$P < 0.0001$
Jilin Hospital (cases with cancer $n = 70$; controls $n = 84$)	Sensitivity = 84.3%; specificity = 86.9%	Sensitivity = 92.9%; specificity = 57.1%	$P = 0.048$	$P < 0.001$
Weihai Hospital (cases with cancer $n = 912$; controls $n = 848$)	Sensitivity = 84.7%; specificity = 87.8%	Sensitivity = 89.0%; specificity = 68.6%	NS	$P < 0.0001$

Sensitivity and specificity for US classification of thyroid cancer with the artificial intelligence deep learning approach[a] compared with Thyroid Imaging Reporting and Data System (TI-RADS) classification by radiologists[b] from three different Chinese Hospitals (assembled with data from REF.[1]). NS, not significant.

AI for thyroid imaging assessment, including clinical parameters in large cohorts of cases and controls. Therefore, we suggest a possible outline for such a (multicentric international) study, which uses a large training pool of anonymized patient data.

> A weakness of AI studies thus far is that non-ultrasound clinical information is largely neglected

Studies should include patients with thyroid nodules who have an indication for thyroid surgery. Cases should include those with histologically proven cancer and controls should have had cancer excluded (by cytology or, preferably, histology); biomaterial samples should be stored for later examination. Indication for surgery should be recommended because of risk factors such as growth of nodule, suspicious palpation and/or elastography or TI-RADS classification. Imaging should be undertaken with conventional ultrasonography (TI-RADS), supplemented by colour Doppler and elastography. The data sets should also include additional US data acquisition in 3D and/or 4D mode and storage of complete data sets that encompass each thyroid lobe entirely, together with patients' histories, clinical examinations and thyroid lab results.

In order to optimize training of the AI model, the training data set has to be checked very carefully for quality and completeness. Self-evidently, mandatory patient informed consent and data safety should be assured according to procedures of established common practice in biomaterial and data banks. In parallel, a second sample of strictly controlled, blinded data following the same prerequisites has to be collected prospectively for validation of the AI procedures resulting from training with the first data set.

*Frederik Verburg[1] and Christoph Reiners[2]**

[1]Nuclear Medicine, University Hospital Marburg, Marburg, Germany.

[2]Nuclear Medicine, University Hospital Würzburg, Würzburg, Germany.

**e-mail: reiners_c@ukw.de*

https://doi.org/10.1038/s41574-019-0204-8

1. Li, X. et al. Diagnosis of thyroid cancer using deep convolutional neural network models applied to sonographic images: a retrospective, multicohort, diagnostic study. *Lancet Oncol.* **20**, 193–201 (2019).
2. Hu, X., Liu, Y. & Qian, L. Diagnostic potential of real-time elastography (RTE) and shear wave elastography (SWE) to differentiate benign and malignant thyroid nodules: a systematic review and meta-analysis. *Medicine (Baltimore)* **96**, e8282 (2017).
3. Lyshchik, A., Drozd, V. & Reiners, C. Accuracy of three-dimensional ultrasound for thyroid volume measurement in children and adolescents. *Thyroid* **14**, 113–120 (2004).
4. Tessler, F. N., Middleton, W. D. & Grant, E. G. Thyroid Imaging Reporting and Data System (TI-RADS): a user's guide. *Radiology* **287**, 29–36 (2018).
5. Wei, X. et al. Meta-analysis of thyroid imaging reporting and data system in the ultrasonographic diagnosis of 10,437 thyroid nodules. *Head Neck* **38**, 309–315 (2016).
6. Mao, F. et al. Assessment of virtual touch tissue imaging quantification and the ultrasound Thyroid Imaging Reporting and Data System in patients with thyroid nodules referred for biopsy. *J. Ultrasound Med.* **37**, 725–736 (2018).
7. Poudel, P. et al. Evaluation of commonly used algorithms for thyroid ultrasound images segmentation and improvement using machine learning approaches. *J. Healthc. Eng.* **2018**, 8087624 (2018).
8. Song, W. et al. Multi-task cascade convolution neural networks for automatic thyroid nodule detection and recognition. *IEEE J. Biomed. Health Inform.* https://doi.org/10.1109/JBHI.2018.2852718 (2018).
9. Wu, H. et al. Classifier model based on machine learning algorithms: application to differential diagnosis of suspicious thyroid nodules via sonography. *AJR Am. J. Roentgenol.* **207**, 859–864 (2016).
10. Chi, J. et al. Thyroid nodule classification in ultrasound images by fine-tuning deep convolutional neural network. *J. Digit. Imaging* **30**, 477–486 (2017).

Competing interests
The authors declare no competing interests.