SAMSUNG

White paper

A Comparison as to the Advocacy and Inter-Observer Agreement of Using S-Detect™ against Sonographers Classifying Thyroid Lesions Using the British Thyroid Association (BTA) Guidelines

Matthew Southam, Warren Foster

AECC University College School of Medical Ultrasound, UK

Introduction

Nodules are a common finding within the thyroid with up to 67% of adults found to have them. However, only 4-7% of nodules are found to be malignant on cytology [1~2].

Given the large number of patients presenting with thyroid nodules it poses a diagnostic problem determining which nodules require intervention and which do not.

Artificial intelligence (AI) is becoming increasingly utilised in a healthcare environment, with the Care Quality Commission (CQC) stating in 2018 that "diagnostic imaging will be 'revolutionised' by machine learning".

AI aims to mitigate the issue of inter-operator or inter-observer reliability, which is a well-documented limitation of ultrasound as a modality. There is growing support of AI in a radiology setting with Buda et al (2019) demonstrating AI may match performance of radiologists using an American grading system for nodules.

This study evaluates a novel technology created by SAMSUNG MEDISON, CO. LTD., Korea termed S-Detect™. This software aims to classify nodules discovered with ultrasound to give an indicator of malignancy, and as such guide the further management of the patient.

S-Detect™ has been recently utilised in the application of breast and thyroid imaging, with studies finding that there is good agreement and high sensitivity [3~5]. However, these studies have not utilised BTA guidelines in the classification of nodules which is the current standard of practice within most healthcare providers in the UK.

Objectives:

The purpose of this study is to assess the S-Detect™ software for its agreement in applying the BTA guidelines to thyroid nodules, with an experienced sonographer. The S-Detect™ software will be assessed for its sensitivity, specificity, positive predictive value and negative predictive value against the sonographer, with agreement assessed using the Kappa Coefficient. The main purpose of this initial study is to assess whether there is potential clinical relevance to the S-Detect™ software in a clinical setting within the UK.

Materials and Methods

A total of 51 patients with thyroid nodules from a single NHS foundation trust were prospectively evaluated using a RS80A with prestige (SAMSUNG MEDISON, CO. LTD., Korea) ultrasound machine equipped with S-Detect™, utilising BTA guidelines. The researcher applied the S-Detect™ software to determine a classification of the nodule, with the result being compared with the classification given by the sonographer, using the same single image. The results of each observer was not available to the other at the time of classification, to avoid bias. The BTA classifications U1 and U2 were termed as negative for the disease, as in practice the classifications would not require further intervention. The BTA classifications U3, U4 and U5 were termed as positive for the disease, as in practice the classifications would require intervention.

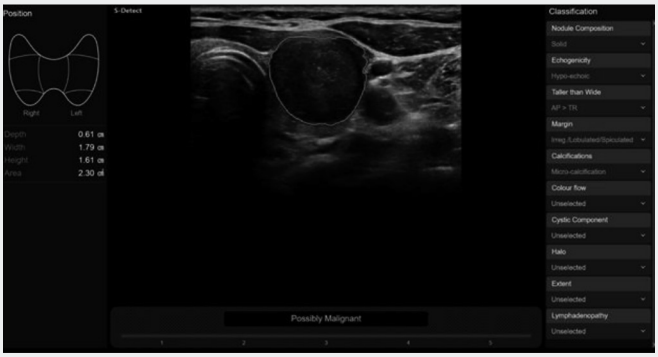

Figure 1. S-Detect™ image showing possibly malignant nodule

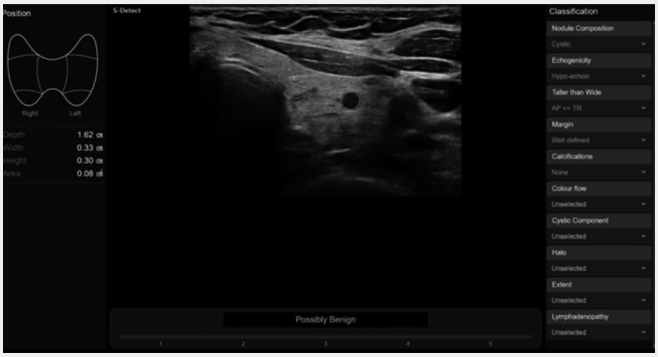

Figure 2. S-Detect™ image possibly benign nodule.

Results

Table 1 summarizes the BTA Positive/Negative Classification results as determined by S-Detect™ or the sonographer.

Of the 51 cases there were 2 cases classified as positive by the S-Detect™ software, which were also characterised as positive by the sonographer, meaning they are true positive results.

There were no cases classified by the S-Detect™ software as negative that were classified as positive by the Sonographer, meaning there were no false negatives.

There were 41 cases that were classified as negative by both the S-Detect™ software and by the sonographer, meaning there were 41 true negatives.

There were 8 cases that were classified as positive by the S-Detect™ software that were classified as negative by the sonographer, meaning that there was a false positive rate of 16.3% (8/51).

The Kappa rating for agreement between the sonographer and the S-Detect™ software in determining whether disease is present and as such determining the need for further investigation is demonstrated as 0.287 with an approximate significance of 0.003. This result determines that there is fair agreement between the S-Detect™ software and the sonographer (Table 2).

Table 1A. S-Detect™ BTA Positive/Negative Classification* Sonographer BTA Positive/Negative:

S-Detect™ Classification* Sonographer Classification Cross tabulation					
			Sonographer Classification		Total
			Positive U3, U4, and U5	Negative U1 and U2	
S-Detect™ Classification	Positive U3, U4 and U5	Count	2	8	10
		% within S-Detect™ Classification	20.0%	80.0%	
		% within Sonographer Classification	100.0%	16.3%	
	Negative U1 and U2	Count	0	41	41
		% within S-Detect™ Classification	0.0%	100.0%	
		% within Sonographer Classification	0.0%	83.7%	
Total		Count	2	49	51

Table 1B. Sonographer/S-Detect™ BTA Positive/Negative Classification:

		Sonographer Classification		S-Detect™ Classification	
		Frequency	Percent (%)	Frequency	Percent (%)
Valid	Positive U3, U4, and U5	2	3.9	10	19.6
	Negative U1 and U2	49	96.1	41	80.4
	Total	51	100.0	51	100.0

Table 1C. S-Detect™ classification* sonographer classification cross tabulation

	Disease Present (Sonographer Classification) : U3, U4 and U5	Disease absent (Sonographer Classification) : U1 and U2	Total
S–Detect™ Classification positive (U3, U4 and U5)	2 (True Positive)	8 (False Positive)	10 (Test Positive)
S-Detect™ Classification Negative (U1 and U2)	0 (False Negative)	41 (True Negative)	41 (Test Negative)

The positive predictive value of the S-Detect™ software is found to be 20%, which is the likelihood of a nodule being positive for disease if classified by the software.

The negative predictive value of the S-Detect™ software is 100%, which is the probability of a nodule being negative for disease if classified by the software.

The sensitivity of the S-Detect™ software is 100%, which is the ability of the software to detect the disease if present.

The specificity of the S-Detect™ software is 83.7%, which is the ability of the software to correctly classify the disease as not being present when it truly is not present.

Table 2. Kappa Coefficient Results:

Symmetric Measures					
		Value	Asymptotic Standard Error[a]	Approximate T[b]	Approximate Significance
Measure of Agreement	Kappa	0.287	0.162	2.921	0.003
No. of Valid Cases		51			

Discussion

The results obtained from this study show some promising results as to the potential effectiveness of S-Detect™. The S-Detect™ software correctly identified the disease as present, i.e. classified the nodules graded as U3 and above, for all the cases that the sonographer classified as present.

Of the 51 sampled cases, there were 2 cases that were identified as positive by the sonographer and 49 cases identified as negative. Whilst this demonstrates a clear significant difference in

classification of the samples, this is representative of thyroid nodules clinically. Of this sample 3.9% were identified as being potentially positive for thyroid malignancy, similar to literature review which reports that 4-7% of nodules are malignant [1~2]. Although the sample size is relatively small, it suggests that it is a representative sample of the overall population.

The main aim of this study was to evaluate whether there was sufficient agreement between the S-Detect™ software and the sonographer in order to establish whether there is potential for the S-Detect™ software to be clinically viable. The Kappa rating for agreement between the two reporters at identifying positive and negative presence of the disease was 287 which indicates "fair agreement". This shows that there is some agreement between the S-Detect™ software and the sonographer, considered as the clinical gold standard.

When the data is assessed it becomes apparent that the S-Detect™ software did identify all positive cases, giving the S-Detect™ a sensitivity of 100%.

The positive predictive value was low at 20%, essentially meaning that the S-Detect™ software, if incorrect, is likely to over classify a lesion and as such result in a greater number of interventions.

More importantly, the S-Detect™ software did not under-classify any nodules that the sonographer graded as positive, and as such implies that the software is unlikely to miss a malignant nodule.

Conclusion

It can be concluded that there is fair agreement between the S-Detect™ software and the sonographer at determining whether there was positive or negative thyroid disease, which in practice is used to determine whether the need for further investigation of a nodule is warranted. There are some disagreements within the results with the S-Detect™ software more likely to assign a positive classification than the sonographer. This could lead to fewer misses of malignant nodules, but may also increase the amount of further investigation. Further research with a greater sample size, using cytology or histology data and two images or 3-dimensional data is recommended for more accurate evaluation. This study shows that the S-Detect™ software utilising BTA guidelines can be considered as having a high potential to be used as an adjunct to current clinical practice.

References

[1] Hoang, J., Raduazo, P., Yousem, D. and Eastwood, J. (2012). What to Do with Incidental Thyroid Nodules on Imaging? An Approach for the Radiologist. Seminars in Ultrasound, CT and MRI, 33(2), pp.150-157.

[2] Sabih, D. and Rahim, K. (2015). Thyroid Nodule Imaging, Status and Limitations. Asia Oceania Journal of Nuclear Medicine & Biology, 3(1).

[3] Choi, Y., Baek, J., Park, H., Shim, W., Kim, T., Shong, Y. and Lee, J. (2017). A Computer-Aided Diagnosis System Using Artificial Intelligence for the Diagnosis and Characterization of Thyroid Nodules on Ultrasound: Initial Clinical Assessment. Thyroid, 27(4), pp.546-552.

[4] Yoo, Y., Ha, E., Cho, Y., Kim, H., Han, M. and Kang, S. (2018). Computer-Aided Diagnosis of Thyroid Nodules via Ultrasonography: Initial Clinical Experience. Korean Journal of Radiology, 19(4), p.665.

[5] Jeong, E., Kim, H., Ha, E., Park, S., Cho, Y. and Han, M. (2019). Computer-aided diagnosis system for thyroid nodules on ultrasonography: diagnostic performance and reproducibility based on the experience level of operators. European Radiology, 29(4), pp.1978-1985.

[6] British Thyroid Association (2014). British Thyroid Association Guidelines for the Management of Thyroid Cancer. Clinical Endocrinology, 81(Supplement 1).

[7] Buda, M., Wildman-Tobriner, B., Hoang, J., Thayer, D., Tessler, F., Middleton, W. and Mazurowski, M. (2019). Management of Thyroid Nodules Seen on US Images: Deep Learning May Match Performance of Radiologists. Radiology, p.181343.

[8] Cqc.org.uk. (2018). [online] Available at: https://www.cqc.org.uk/sites/default/files/20180718-radiology-reporting-review-report-final-for-web.pdf [Accessed 18 Nov. 2018].

SAMSUNG MEDISON CO., LTD.

Disclaimer

[Do not distribute this document to customers unless relevant regulatory and legal affairs officers approve such distribution.]

* The features mentioned in this document may not be commercially available in all countries. Due to regulatory reasons, their future availability cannot be guaranteed.
* Images may have been cropped to better visualize its pathology.
* This clinical practice review is not an official clinical study or paper presented at a conference. It is a result of a personal study conducted by collaboration between Samsung Medison and Prof. Warren Foster and Matthew Southam. This review is to aid customer in their understanding, but the objectivity is not secured.

*** 본자료는 공식 임상시험 결과물이나 학회에 발표된 논문이 아니며 삼성메디슨이 Warren Foster 교수님 및 Matthew Southam와 협업하여 산출된 개인 연구의 결과물입니다.**
고객의 요청에 따라 이해를 돕기 위해 제공하는 자료일 뿐 객관성은 확보되지 않았습니다.

Please visit http://www.samsung.com/global/business/healthcare