

# Evaluation of Diagnostic Usefulness of Thyroid Lesions of Deep Learning-based CAD System

Chae Won Kang<sup>1,2</sup>, Hyo Yeong Lee<sup>1,3,\*</sup>

<sup>1</sup>Department of Biomedical Health Science, Graduate School of Dong-Eui University

<sup>2</sup>Department of Radiology, Numberone Medical Clinic

<sup>3</sup>Department of Radiological Science, Dongeui University

Received: August 23, 2024. Revised: October 24, 2024. Accepted: October 31, 2024.

## ABSTRACT

This study aims to evaluate the diagnostic concordance and accuracy by comparing thyroid lesions diagnosed with the artificial intelligence-based computer-aided diagnosis (CAD) system, S-Detect™, to the results of fine-needle aspiration biopsy(FNAB). A retrospective study was conducted involving 60 patients at N Hospital in Gyeongnam from May 2023 to September 2023. The study used S-Detect™ to analyze ultrasound findings and malignancy risk of thyroid nodules and compared these findings with FNAB results to determine accuracy. The study assessed the sensitivity, specificity, accuracy, positive predictive value (PPV), and negative predictive value (NPV) of S-Detect™ and evaluated the diagnostic concordance between the two methods using Kappa analysis. S-Detect™ demonstrated a sensitivity of 90.5%, specificity of 83.2%, accuracy of 88.3%, PPV of 80.7%, and NPV of 92.7%. The Kappa value for diagnostic agreement between S-Detect™ and FNAB was 0.719 (p<0.05), indicating a high level of agreement between the methods. Therefore, the CAD system S-Detect™ proves valuable in distinguishing between malignant and benign thyroid lesions and could reduce unnecessary tissue examinations when used appropriately before thyroid fine-needle aspiration.

Keywords: Computer-Aided Diagnosis (CAD) System, Fine-Needle Aspiration Biopsy (FNAB), Predictive Value (PPV), Negative Predictive Value (NPV)

## I. INTRODUCTION

갑상샘 결절은 매우 흔한 질환으로 우리나라 성인의 30~40%의 유병률을 보이며, 갑상샘 결절 중 갑상샘암의 유병률은 2-6%로 알려져 있다<sup>[1]</sup>. 2023년에 발표된 중앙암등록본부 자료에 의하면 2021년에 우리나라에서는 277,523건의 암이 새로이 발생했는데, 그중 갑상샘암이 남녀를 합쳐서 35,303건, 전체 암 발생의 12.7%로 1위를 차지했다<sup>[2]</sup>.

영상기술의 지속적인 발전으로 인해 갑상샘 질환과 갑상샘암의 발견율은 국내외적으로 급속한 증가 추세를 보이고 있다<sup>[3]</sup>. 갑상샘 질환을 평가하

기 위해 초음파, 컴퓨터 단층촬영, 방사성동위원소 스캔 등 다양한 보조 검사 방법이 사용되고 있다. 그중 갑상샘 초음파는 비침습적으로 시각화하고 평가하는 데 사용되며, 결절의 진단부터 위험도 분류, 세침흡인 검사를 결정하는 일차적인 도구로써 이용되고 있다<sup>[4]</sup>. 그러나 초음파 영상의 정확한 식별은 의사의 경험과 밀접한 관련이 있으며 관찰자 간의 차이가 클 수 있다<sup>[5]</sup>. 따라서 경험이 부족한 의사는 오진, 상태를 과소평가하거나 불필요하게 세침흡인 검사를 수행할 위험이 더 크다<sup>[6]</sup>. 최근 인공지능기반의 컴퓨터 진단 보조(CAD, Computer-Aided Diagnosis) 시스템이 초음파의 단점을 보완하고, 결

\* Corresponding Author: Hyo Yeong Lee E-mail: lhy250@deu.ac.kr Tel: +82-51-890-2679 Address: Dong-Eui University of Busan, 176 Eomgwang-ro, Busan-jingu, 614-714, Republic of Korea

질의 특징을 정확하고 일관되게 해석하며, 작업의 흐름을 반자동화하는 방식으로 도입되었다. 이를 통해 불필요한 세침흡인 검사 및 진단적 수술을 줄이는데 기여가 되고 있다<sup>[7]</sup>. 선행연구에서도 초음파 이미지에서 CAD시스템 성능연구와 실제 임상 결과에 대한 결절을 구분하는데 있어 두결과가 비슷한 진단성능을 보여준 것으로 보고되고 있다<sup>[8]</sup>. 또한 CAD시스템이 Bethesda II 및 VI 범주의 결절에 대해서는 신뢰할 수 있는 검사 결과를 가지고 있는 것으로 나타났다<sup>[9]</sup>. 따라서 갑상샘 암 진단에서 인공지능 기반의 CAD 시스템 결과와 세침흡인 검사 결과를 비교분석하여 CAD 시스템과 FNA(Fine Needle Aspiration) 진단 결과의 일치도를 분석하고 정확도를 비교분석해 보고자 한다.

## II. MATERIAL AND METHODS

### 1. 위험도 분류방법

획득한 영상을 갑상샘 병리학 보고 시스템 (Bethesda System for Reporting Thyroid catopathology)로 결절의 위험 단계를 6가지로 분류하고 예측 악성도<sup>[10]</sup>를 Table 1에 나타내었다.

Table 1. TI-RADS Category (2023)

분류	위험도
TR I	비진단적 (Non diagnostic)
TR II	양성 (Benign) Clinical follow-up
TR III	비정형 (Atypia of undetermined significance) Repeat FNA
TR IV	여포종양 (Follicular nodule/suspicious follicular nodule) Surgical lobectomy
TR V	악성의심 (Suspicious for malignancy) Surgical lobectomy or near-total thyroidectomy
TR VI	악성 (Malignant) Near-total thyroidectomy

### 2. Cohen's Kappa 계수값 분류

Cohen's Kappa 계수는 두 명의 평가자 간의 일치도를 측정하는 통계적 방법으로, 0부터 1 사이의

값을 가진다. 값이 클수록 높은 일치도를 의미하며, Table 2와 같이 일반적으로 분류하였다.

Table 2. Agreement level according to Cohen's Kappa coefficient.

K	Strength Agreement
> 0.000	Poor
0.000 - 0.200	Sght
0.201 - 0.400	Fair
0.401 - 0.600	Moderate
0.601 - 0.800	Substantial
0.801 - 1.000	Almost Perfect

### 3. 대상 환자

경남 소재 내과병원에서 2023년 5월부터 9월 말까지 갑상샘에 세침흡인 검사를 받은 환자군 60명을 대상으로 후향적 연구를 수행하였다. 본 연구에 사용된 영상은 S-Detect<sup>TM</sup>의 자동화된 기법을 바탕으로 갑상샘 결절의 초음파 소견과 악성 위험도를 나타냈다.

### 4. 검사장비

본 연구에 사용한 장비는 삼성메디슨(주) Samsung Medison Co.(성남, 대한민국)에서 개발한 DL-CAD 소프트웨어(S-Detect; 삼성메디슨(주), 대한민국)를 활용하여 병변의 데이터를 추출하였고 본 연구에서 제시된 모든 증례는 3-12MHz 선형 다중 주파수 트랜스듀서(RS80A with Prestige, Samsung Medison, Co. Ltd.)를 사용하였다.

### 5. CAD system

S-Detect<sup>TM</sup> (Samsung Ultrasound RS80A, Samsung Medison Co. Ltd., Seoul, South Korea)는 딥러닝 기반의 컴퓨터 보조 진단(Deep learning based computer-aided diagnosis, DL-CAD) 시스템으로 초음파 이미지상에서 선택된 병변을 TI-RADS (Thyroid Imaging Reporting and Data System)을 기반으로 형태학적 분석을 통해 악성 또는 양성의 가능성을 분류하여 지원하는 소프트웨어 프로그램으로<sup>[11]</sup> Fig. 1에 나타내었다.



Fig. 1. Deep learning-based computer-aided diagnostic system for ultrasound images."

## 6. 통계분석

통계분석은 PASW(PASW statistics, ver. 26.0, SPSS, Chicago, USA)와 MedCalc(MedCalc Software, Ostend, Belgium)를 이용하였다.

연구 대상의 일반적 특성을 파악하기 위해 기술 통계와 빈도 분석을 하였으며, FNA와 S-Detect™에서의 DL-CAD 시스템의 진단의 일치도를 알아보기 위한 Kappa 분석을 실시하였다. 그리고 두 진단 결과를 비교하고 이진 분류 모델의 성능을 시각적으로 평가하기 위하여 ROC curve를 사용하였고, 유의 수준은 0.05로 설정하였다.

## III. RESULT

### 1. FNA 결과

환자 60명중에서 FNA 검사 결과 TR Category I 0명, TR Category II 3명, TR Category III 18명, TR Category IV 10명, TR Category V 16명, TR Category VI 13명으로 Table 3에 나타났다.

Table 3. Results of FNA

Category	N(60)
TR I	0
TR II	3
TR III	18
TR IV	10
TR V	16
TR VI	13

### 2. Cohen's Kappa 분석 결과

S-detect의 DL-CAD 방법과 FNA방법 간의 진단이 얼마만큼 일치하는지 알아보기 위해 kappa 분석을 실시하였고 결과는 Table 4에 나타내었다.

S-detect의 DL-CAD 방법과 FNA 방법 간의 일치도 분석 결과는 Kappa 계수값 0.765(Cohen's Kappa 계수(k) 0.601-0.800은 일치 강도 Substantial로 상당한 일치도로 높게 나타났으며, 두 방법 간에 상당한 일치도를 보였다.

Table 4. Results of Cohen's Kappa analysis

		FNA		Total
		Benign	Malignant	
DL-CAD	Benign	30	6	36
	Malignant	1	23	24
Total		31	29	60

K=0.765(p<0.001)

### 3. ROC(Receiver Operating Characteristics) 곡선

FNA와 DL-CAD의 ROC 곡선 분석결과 AUC는 0.880(95% 신뢰구간 0.771-0.950)으로 나타났고, FNA는 0.865(95% 신뢰구간 0.753-0.940)이고 p값은 0.689로 두 검사 방법 간의 유의한 차이가 없었다. (p>0.05), 이는 Fig. 2와 Table 5, 6에 나타내었다.

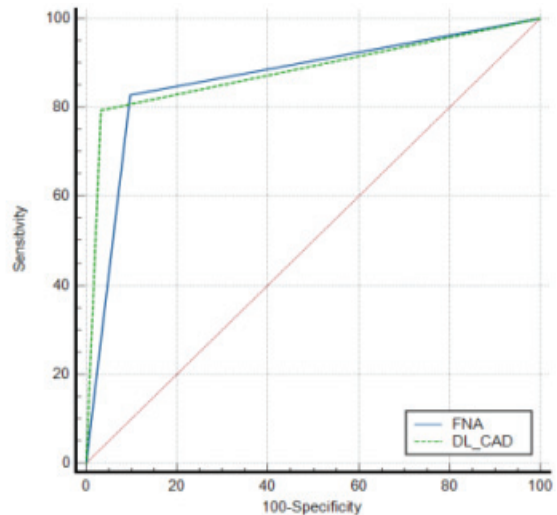


Fig. 2. Comparison of ROC curves between DL-CAD and FNA.

Table 5. DL-CAD to FNA

difference between areas	0.0150
Standard Error	0.0376
95% CI	-0.0587 to 0.0888
Z statistic	0.399
Significance level	P=0.6898

Table 6. AUC and 95% confidence interval

Variable	AUC	SE	95% CI
DL-CAD	0.880	0.0415	0.771-0.950
FNA	0.865	0.447	0.753-0.940

인공지능 기반 CAD 시스템의 민감도(sensitivity)는 95%, 특이도(specificity) 80%, 정확도(accuracy) 85%로 나타났다.

#### IV. DISCUSSION

갑상샘 질환 진단에 초음파 검사가 널리 사용되면서 갑상샘 결절의 발견률이 크게 높아졌고<sup>[12]</sup>, 현행 지침에서는 갑상샘암 진단에 초음파검사를 일차적으로 권고하고 있다<sup>[13]</sup>. 1989년에 Hirning et al.은 초음파 질감 분석을 기반으로 갑상샘 결절을 구별하는 최초의 연구에서 분류 시스템의 정확도는 85%였고<sup>[14]</sup> Chang et al.의 연구에서 민감도 90%, 특이도 50.9%로 나타났다<sup>[15]</sup>. 본 연구에서 민감도 95%, 특이도 80% 그리고, 정확도 85%로 분석되었다. 그리고 갑상샘 병리학 보고 시스템(Bethesda System for Reporting Thyroid catopathology)에 따른 갑상샘 진단분류에서 category I에 해당되는 환자는 표본에서 제외하였다. Acharya et al.은 3차원 고해상도 조영증강 초음파 영상을 이용하여 악성 갑상샘 결절과 양성 갑상샘 결절을 자동으로 분류하는 CAD 시스템을 제안하였다. 이 시스템은 분류 정확도가 98.1%에서 100% 범위였으며, 이 CAD 시스템은 갑상샘 결절의 특성을 식별하는 유용한 결과를 보였다<sup>[16]</sup>. 인공지능 기반 초음파 S-Detect 기술은 대부분 유방 종양 진단에 활용되었으며, 갑상샘 결절 진단에는 상대적으로 사용이 적었다<sup>[17]</sup>. 본 연구에서는 초음파 S-Detect 기술을 적용하여 갑상샘 결절 진단의 특이성, 민감도, 정확도를 향상시켰으며, TI-RADS 분류 및 진단 방법을 통해 갑상샘 결절에

진단 능력을 개선할 수 있었다. 그러나, Sollini 등은 체계적 고찰 연구를 통해 기존의 많은 머신 러닝 기반의 연구 결과들이 서로 다른 영상 프로토콜, 세분화 방법 등의 다양한 교란 변수로 인해 연구 결과들의 비교 가능성과 재현성이 제한될 수 있다고 보고하였다<sup>[18]</sup>.

또한 적은 수의 환자를 대상으로 진행하고 특정 지역, 특정병원에 내원한 환자를 대상으로 연구하였기 때문에 연구결과가 전체 환자를 대변할 수가 없다는 제한점이 있다. 그러나 향후 더 많은 환자 다양한 환경에서 추가적인 검사를 실시하면 연구 결과 정확도와 활용도를 더 높일 수 있을 것이다.

#### V. CONCLUSION

갑상샘 초음파에서 인공지능 기술을 접목하면 의료 진단 및 관리 분야에서 중요한 발전을 가져올 수 있다. AI 기술을 활용하면 초음파 검사의 해석에 대한 차이를 줄일 수 있고, 갑상샘 결절 환자의 불필요한 세침흡인 검사나 진단적 수술 등을 줄이는 데 도움이 될 수 있을 것으로 생각한다.

#### Acknowledgement

이 논문은 2024년 동의대학교 교내 연구비에 의해 연구되었음.(202401240001)

#### Reference

- [1] J. H. Moon, M. K. Hyun, J. Y. Lee, et al., "Prevalence of thyroid nodules and their associated clinical parameters : a large-scale, multicenter-based health checkup study", Korean journal of internal medicine, Vol. 33, No. 4, pp. 753-762, 2018. <http://dx.doi.org/10.3904/kjim.2015.273>
- [2] National Cancer Information Center, From URL; <https://www.cancer.go.kr/lay1/S1T639C641/contents.do>
- [3] F. Bray, J. Ferlay, I. Soerjomataram, et al., "Global cancer statistics 2018: GLOBOCAN estimates of incidence and mortality worldwide for 36 cancers in 185 countries", CA: A Cancer Journal for Clinicians, Vol. 68, No. 6 pp. 394-424, 2018. <http://dx.doi.org/10.3322/caac.21492>

- [4] B. R. Haugen, "2015 American Thyroid Association Management Guidelines for Adult Patients with Thyroid Nodules and Differentiated Thyroid Cancer: What is new and what has changed?", *Cancer*, Vol. 123, No. 3, pp. 372-381, 2017. <http://dx.doi.org/10.1002/cncr.30360>
- [5] M. Dighe, R. Barr, J. Bojunga, et al., "Thyroid ultrasound: State of the art. part 2 - focal thyroid lesions", *Medical Ultrasonography*, Vol. 19, No. 2, pp. 195-210, 2017. <http://dx.doi.org/10.11152/mu-999>
- [6] J. K. Hoang, W. D. Middleton, A. E. Farjat, et al., "Interobserver Variability of Sonographic Features Used in the American College of Radiology Thyroid Imaging Reporting and Data System", *American Journal of Roentgenology*, Vol. 211, No. 1, pp. 162-167, 2018. <http://dx.doi.org/10.2214/AJR.17.19192>
- [7] H. J. Lee, D. Y. Yoon, Y. L. Seo, et al., "Intraobserver and Interobserver Variability in Ultrasound Measurements of Thyroid Nodules", *Journal of Ultrasound Medicine*, Vol. 37, No. 1, pp. 173-178, 2018. <http://dx.doi.org/10.1002/jum.14316>
- [8] Y J Kim, Y S Choi, et al., "Deep convolutional neural network for classification of thyroid nodules on ultrasound: Comparison of the diagnostic performance with that of radiologists" *European Journal of Radiology* ,Vol. 152, pp.110335, 2022. <https://doi.org/10.1016/j.ejrad.2022.110335>
- [9] J. H. Lee, Y. K. Seong, C. H. Chang, et al., "Computer- aided lesion diagnosis in B-mode ultrasound by border irregularity and multiple sonographic features", *Computer Science Medicine*, Vol. 8670, 2013. <https://doi.org/10.1117/12.2007452>
- [10] S. Z. Ali, Z. W. Baloch, B. Cochand-Priollet, et al., "The 2023 Bethesda System for Reporting Thyroid Cytopathology", *Thyroid*, Vol. 12, No. 5, 2023. <https://doi.org/10.1089/thy.2023.0141>
- [11] Y. Chang, A. K. Paul, N. Kim, et al., "Computer-aided diagnosis for classifying benign versus malignant thyroid nodules based on ultrasound images: A comparison with radiologist-based assessments", *Medical Physics*, Vol. 43, No. 1 pp. 554-567, 2016. <http://dx.doi.org/10.1118/1.4939060>
- [12] E. J. Ha, D. G. Na, Y. H. Lee, et al., "A Multicenter Prospective Validation Study for the Korean Thyroid Imaging Reporting and Data System in Patients with Thyroid Nodules", *Korean Journal Radiology*, Vol. 17, No. 5, pp. 811-821, 2016.
- [13] J. H. Shin, J. H. Baek, J. Chung, et al., "Ultrasonography Diagnosis and Imaging-Based Management of Thyroid Nodules: Revised Korean Society of Thyroid Radiology Consensus Statement and Recommendations", *Korean Journal Radiology*, Vol. 17 No. 3, pp. 370-395, 2016. <http://dx.doi.org/10.3348/kjr.2016.17.3.370>
- [14] T. Hirning, I. Zuna, D. Schlaps, et al., "Quantification and classification of echographic findings in the thyroid gland by computerized B-mode texture analysis", *European Journal Radiology*, Vol. 9, No. 4, pp. 244-247, 1989.
- [15] C. L. Chng, H. C. Tan, C. W. Too, et al., "Diagnostic performance of ATA, BTA and TIRADS sonographic patterns in the prediction of malignancy in histologically proven thyroid nodules", *Singapore Medical Journal*, Vol. 59, No. 11, pp. 578-583, 2018. <http://dx.doi.org/10.11622/smedj.2018062>
- [16] U. R. Acharya, O. Faust, S. V. Sree, et al., "Cost-effective and non-invasive automated benign and malignant thyroid lesion classification in 3D contrast-enhanced ultrasound using combination of wavelets and textures: a class of ThyroScan™ algorithms", *Technology in Cancer Research & Treatment*. Vol. 10, No. 4, pp. 371-380, 2011. <http://dx.doi.org/10.7785/tcr.2012.500214>
- [17] E. Szczepanek-Parulska, K. Wolinski, K. Dobruch-Sobczak, et al., "S-Detect Software vs. EU-TIRADS Classification: A Dual-Center Validation of Diagnostic Performance in Differentiation of Thyroid Nodules", *Journal of Clinical Medicine*, Vol. 9, No. 8, pp. 2495-2576, 2020. <http://dx.doi.org/10.3390/jcm9082495>
- [18] S. Martina, C. Luca, C. Arturo, et al., "Texture analysis and machine learning to characterize suspected thyroid nodules and differentiated thyroid cancer: Where do we stand?", *European Journal of Radiology*, Vol. 99, pp.1-8, 2018. <http://dx.doi.org/10.1016/j.ejrad.2017.12.004>

## 딥러닝을 기반으로 한 CAD 시스템의 갑상샘 질환의 진단 유용성

강채원<sup>1,2</sup>, 이효영<sup>1,3,\*</sup>

<sup>1</sup>동의대학교 대학원 보건의과학과

<sup>2</sup>넘버원메디컬클리닉 영상의학과

<sup>3</sup>동의대학교 방사선학과

### 요 약

본 연구는 인공지능 기반 컴퓨터 진단 보조 시스템(CAD)인 S-Detect™를 통해 진단된 갑상샘 병변과 세침흡인 검사(FNAB) 결과를 비교하여 분석하고, 진단의 일치도와 정확도를 평가하고자 한다. 2023년 5월부터 2023년 9월까지 경남 소재 N 병원 내과에서 60명의 환자를 대상으로 후향적 연구를 수행하였다. S-Detect™를 사용하여 갑상샘 결절의 초음파 소견과 악성 위험도를 분석하고, 이를 세침흡인 검사 결과와 비교하여 정확도를 확인하였다. S-Detect 방법과 세침흡인 검사 방법 간의 민감도, 특이도, 정확도, 양성예측도 및 음성 예측도를 분석하였으며, 두 방법 간의 진단 일치도를 Kappa 분석을 통해 확인하였다. S-Detect 분석 결과는 민감도 90.5%, 특이도 83.2%, 정확도 88.3%, 양성 예측도 80.7%, 음성 예측도 92.7%로 나타났다. 또한, S-Detect 방법과 세침흡인검사 방법 간의 진단 일치도 분석 결과, Kappa 값이 0.719(p<0.05)로 높게 나타났으며, 두 방법 간에 유사한 일치도를 보였다. 따라서, 인공지능 기반 컴퓨터 진단 보조 시스템(CAD)인 S-Detect는 갑상샘 병변에서 악성 결절과 양성 결절을 구별하는데 유용하며, 갑상선 세침흡인 검사 전에 적절히 활용하면 불필요한 조직 검사를 줄일 수 있을 것으로 생각한다.

중심단어: 컴퓨터 진단보조시스템, 세침흡인검사, 민감도, 특이도, 정확도, 양성 예측도, 음성 예측도

### 연구자 정보 이력

	성명	소속	직위
(제1저자)	강채원	동의대학교 대학원 보건의과학과	대학원생
(교신저자)	이효영	동의대학교 방사선학과	교수