Implications of US radiomics signature for predicting malignancy in thyroid nodules with indeterminate cytology

Jiyoung Yoon

Department of Medicine

The Graduate School, Yonsei University

Implications of US radiomics signature for predicting malignancy in thyroid nodules with indeterminate cytology

Jiyoung Yoon

Department of Medicine

The Graduate School, Yonsei University

Implications of US radiomics signature for predicting malignancy in thyroid nodules with indeterminate cytology

Directed by Professor Jin Young Kwak

The Doctoral Dissertation
submitted to the Department of Medicine,
the Graduate School of Yonsei University
in partial fulfillment of the requirements for the degree
of Doctor of Philosophy in Medical Science

Jiyoung Yoon

June 2021

This certifies that the Doctoral Dissertation of Jiyoung Yoon is approved.

Thesis Supervisor: Jin Young Kwak

Thesis Committee Member#1: Sang-Wook Kang

Thesis Committee Member#2: Eunjung Lee

Thesis Committee Member#3: Vivian Youngjean Park

Thesis Committee Member#4: Kyunghwa Han

The Graduate School
Yonsei University

June 2021

ACKNOWLEDGEMENTS

I acknowledge my deep gratitude to professor Jin Young Kwak for supporting my efforts with total commitment and facilitating every step in the process of completing this study. My appreciation for her guidance and encouragement is tremendous.

In addition, I would like to take this opportunity to express my gratitude to the professors of the committee who gave generous advice and words of encouragement to this study even during the busy days.

Through this study, not only thyroid radiology but also radiomics could be studied in-depth, and it is expected to be the basis for an active research process in the future.

<TABLE OF CONTENTS>

LIST OF FIGURES

LIST OF TABLES

ABSTRACT

Implications of US radiomics signature for predicting malignancy in thyroid nodules with indeterminate cytology

Jiyoung Yoon

Department of Medicine
The Graduate School, Yonsei University

(Directed by Professor Jin Young Kwak)

Objectives: The purpose of this study was to evaluate the role of the radiomics score using US images to predict malignancy in AUS/FLUS and FN/SFN nodules.

Methods: 155 indeterminate thyroid nodules in 154 patients who received initial US-guided FNA for diagnostic purposes were included in this retrospective study. A representative US image of each tumor was acquired, and square ROIs covering the whole nodule were drawn using the Paint program of Window 7. Texture features were extracted by in-house texture analysis algorithms implemented in MATLAB 2019b. The LASSO logistic regression model was used to choose the most useful predictive features, and ten-fold cross-validation was performed. Two prediction models were constructed using multivariable logistic regression analysis: one based on clinical variables, and the other based on clinical variables with the radiomics score. Predictability of the two models was assessed with the AUC of the ROC curves.

Results: Clinical characteristics did not significantly differ between malignant and benign nodules, except for mean nodule size. Among 730 candidate texture features generated from a single US image, 15 features were selected. Radiomics signatures were constructed with a radiomics

score, using selected features. In multivariable logistic regression analysis, higher radiomics score was associated with malignancy (OR = 10.923; P <0.001). The AUC of the malignancy prediction model composed of clinical variables with the radiomics score was significantly higher than the model composed of clinical variables alone (0.839 vs 0.583).

Conclusions: Quantitative US radiomics features can help predict malignancy in thyroid nodules with indeterminate cytology.

Key words: ultrasonography, thyroid nodule, computer assisted radiographic image interpretation, computer-assisted diagnosis, thyroid neoplasms

Implications of US radiomics signature for predicting malignancy in thyroid nodules with indeterminate cytology

Jiyoung Yoon

Department of Medicine
The Graduate School, Yonsei University

(Directed by Professor Jin Young Kwak)

I. INTRODUCTION

Thyroid nodules are a common disease entity accounting for 1-5% of palpable nodules in iodine-sufficient areas [1,2], and 19-67% of patients who undergo ultrasonography (US) for screening despite not having symptoms or palpable nodules [3]. Up to now, US-guided fine-needle aspiration (US-FNA) has been regarded as the first-line diagnostic method for nodular thyroid disease due to its cost-effectiveness and accuracy [4,5]. However, an important limitation of US-FNA is that clinicians cannot definitely determine whether a nodule is benign or malignant on some specimens. These are considered to have indeterminate cytology and are classified into category III of the Bethesda system, atypia of undetermined significance/follicular lesions of undetermined significance (AUS/FLUS) and category IV of the Bethesda system, follicular neoplasm or suspicious for a follicular neoplasm (FN/SFN) [5]. Indeterminate cytology can result in a delayed diagnosis of thyroid malignancy or unnecessary operations for a definitive diagnosis [6,7].

Many studies have tried to identify US features that can be used to select patients at high risk for malignancies or discover molecular markers (such as RAS mutation or BRAF mutation) which help differentiate malignancy from these indeterminate nodules [8-12]. Taller-than-wide shape, marked

hypoechogenicity, irregular or spiculated margin, and microcalcifications were US features that were significantly associated with malignancy [7,13]. But as gray-scale US is highly subjective and its results depend on the experience and skill of the operator, with both inter-observer and intra-observer variations [14-16]. In contrast, molecular markers are objective but the testing needed to evaluate relevant genetic mutations is still expensive [17,18]. When diagnosing indeterminate cytology nodules, 18F-FDG-PET-CT has shown a high negative predictive value (NPV) for malignancy [19]. However, since PET-CT is not routinely performed in patients with thyroid nodules of indeterminate cytology, using PET-CT to predict malignancy at this point in time results in increased medical costs and radiation exposure.

Recently, radiomics features extracted from tomographic images have been introduced as a potential tool to overcome intra- and inter-observer variability in subjective visual analysis [20,21]. Radiomics can potentially be applied to various conditions to aid cancer detection, diagnosis, assessment of prognosis, prediction of treatment response, and monitoring of disease status [20,21]. A few studies have applied this approach to the analysis of thyroid nodules, and a majority of them focus on predicting the prognosis of already diagnosed papillary thyroid carcinomas [23-25].

Recent studies suggest that radiomics features extracted from US can be used to predict malignancy in thyroid nodules [26,27]. But to our knowledge, no study has yet determined whether radiomics using US images can predict malignancy in nodules with indeterminate cytology. Therefore, the purpose of this study was to evaluate the role of the radiomics score using US images to predict malignancy in AUS/FLUS and FN/SFN nodules.

II. MATERIALS AND METHODS

This study was approved by the institutional review board (IRB) of Severance Hospital (Seoul, South Korea) and as the study was of retrospective

design, requirement of informed consent was waived by the IRB committee.

1. Study population

From February 2016 through July 2017, initial US-guided FNA was performed for diagnostic purposes at our institution (a referral center) in 2,382 nodules measuring 1 cm or larger at the maximal dimension in 2,252 consecutive patients. Among them, 353 nodules in 331 patients had AUS/FLUS cytology and 16 nodules in 16 patients had SFN/FN cytology. Thyroid nodules fulfilling one of the following inclusion criteria were included in this study: (a) nodules pathologically confirmed with surgery, (b) nodules that were diagnosed as benign or malignant on follow-up US-FNA, (c) nodules that were diagnosed as benign or malignant on US-core biopsy, and (d) nodules that decreased in size on follow-up US performed 6 months after the initial US-guided FNA. Finally, 155 nodules in 154 patients were included in this study (Figure 1).

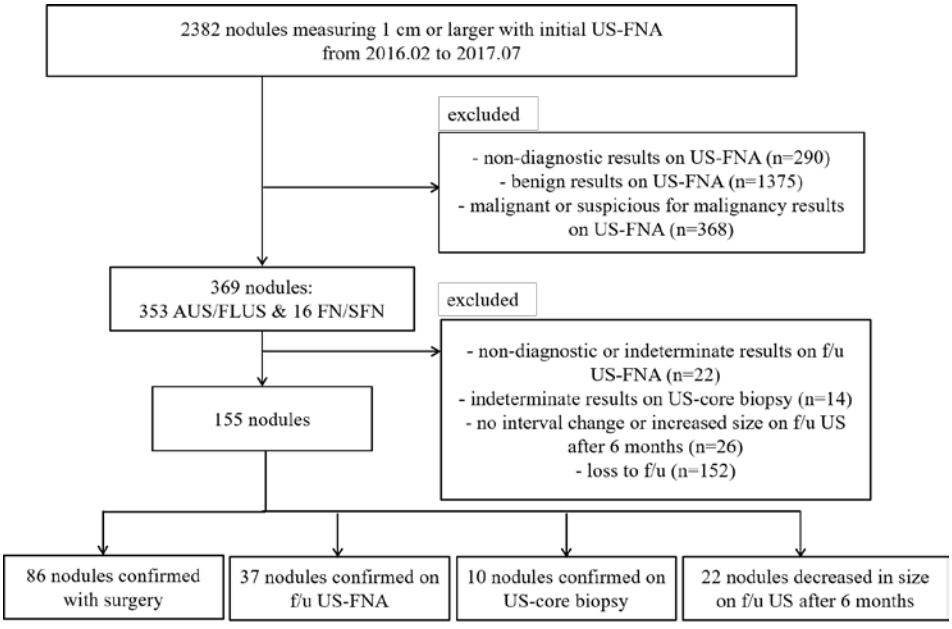

Figure 1. Patient selection diagram

2. Image Acquisition

All gray-scale USs were performed using a 7−12-MHz linear transducer (HDI 3000 or 5000; Philips Medical Systems, Bothell, WA), a 5-to 12-MHz linear transducer (iU22; Philips Healthcare, Bothell, WA) or a 5-to 12-MHz linear transducer (EPIQ5; Philips Healthcare, Bothell, WA). All US scans were done by one of 13 radiologists with 1 to 22 years of experience in thyroid imaging. US features of each nodule that underwent US-FNA were prospectively recorded in our institution database. The radiologist who performed the US-FNA recorded the following US feature categories: margin, shape, echogenicity, internal component, calcifications, and echogenicity [28]. The prospectively recorded US features and US images were both retrospectively reviewed by one radiologist with 18 years of experience in thyroid imaging (K.J.Y.) and reclassified according to the 2015 American Thyroid Association management guidelines. A representative US image of each nodule was selected by the thyroid radiologist (K.J.Y.) from previously captured images with prospectively recorded US features being taken into consideration. Representative US images were retrieved from the picture archiving and communication system (PACS) of our institution. Using the Paint program of Window 7, square regions-of-interest (ROIs) were drawn by a radiologist (K.J.Y.) to cover the whole nodule.

3. Extraction of Radiomics feature

Texture features were extracted by in-house texture analysis algorithms implemented in MATLAB 2019b (The MathWorks, Inc., Natick, Massachusetts, United States). ROIs drawn on US images were firstly saved as JPG images and then converted into grayscale intensity images by eliminating the hue and saturation information while retaining luminance. 730 candidate radiomics features such as energy, entropy, kurtosis, features for the GLCM and GLRLM texture matrices, and features for the single-level

discrete 2D wavelet transform were generated from a single US image. Each extracted ROI image was normalized for direct comparison between patients when textural features were calculated. More details on the methods used to extract radiomics features are described in the APPENDICES.

4. Cytological analysis

Slides were interpreted by one of 8 cytopathologists who specialize in thyroid pathology. Cytologic reports of the US-guided FNA were based on the BSRTC (Bethesda System For Reporting Thyroid Cytopathology) used in our institution since 2009 [29]. The BSRTC categories are the following: non-diagnostic or unsatisfactory (Bethesda System I), benign (Bethesda System II), AUS/FLUS (Bethesda System III), FN/SFN (Bethesda System IV), suspicious for malignancy (Bethesda System V), and malignant (Bethesda System VI). The final cytologic reports were obtained from electronic medical records.

5. Data and Statistical analysis

To diminish the high dimension of the texture features to the number of events, the least absolute shrinkage and selection operator (LASSO) logistic regression model was used to choose the most useful predictive features. Ten-fold cross-validation was performed to avoid overfitting. The calculated area under the curve (AUC) values of the receiver operating characteristic (ROC) curves were plotted versus log (Lambda); a tuning parameter (Lambda) was selected when the mean AUC value was maximized under 10-fold cross-validation. Features with non-zero coefficients were selected and a radiomics score was calculated by a linear combination of selected features weighted by their respective coefficients.

Categorical variables are shown as numbers with percentages. Continuous variables are presented as mean values with standard deviations.

Univariable and multivariable logistic regression analyses were performed to build malignancy prediction models using clinical variables and radiomics scores. We constructed two models for multivariable logistic regression analysis: one based on clinical variables, and another based on clinical variables with the radiomics score. Selected clinical variables to construct malignancy prediction model are nodule size, patient's gender, patient's age, and the Bethesda category of nodule (AUS/FLUS vs FN/SFN). To evaluate the incremental predictive value of the radiomics score when added to clinical variables, the predictability of the two models was evaluated with the AUC of the ROC curves. Bootstrapping methods were used for internal validation and comparison of predictability between the two models. We calculated the difference between bootstrapped AUCs of the two ROC curves, and considered the predictability difference of the two models was statistically significant if the 95% confidence interval (CI) for the AUC difference does not straddle zero. A subgroup analysis was also performed on nodules that were pathologically diagnosed with surgical resection.

All statistical analyses were performed with R software (version 3.4.3.; R Foundation for Statistical Computing, Vienna, Austria). A P value of 0.05 or less was considered statistically significant.

III. RESULTS

1. Patient Characteristics.

Of 155 enrolled nodules in 154 patients, 42 nodules (27.1%) were confirmed as malignancy and the remaining 113 nodules (72.9%) were confirmed as benign. Table 1 compares the demographic features of patients between the malignant and benign thyroid nodules. The mean size of malignant nodules was significantly larger than that of benign nodules (25.7±16.1 vs 23.8±13.5, P = 0.048). The mean age of patients with malignant nodules was 47.2±15.1 years and the mean age of patients with benign nodules

was 50.1±13.0 years, so there was no significant difference in age between the two groups (P = 0.223). There was no significant difference in gender distribution between patients with malignant nodules and patients with benign nodules (P = 0.640). Also, visual analysis results of the US images based on the classifications of the ATA guidelines did not differ between the two groups (P = 0.053) (Table 1).

Table 1. Demographic features of enrolled patients

	Malignant (n=42)	Benign (n=113)	*P* value
Mean age (years)	47.2±15.1	50.1±13.0	0.223
Gender			
Men	10 (23.8%)	23 (20.4%)	0.640
Women	32 (76.2%)	90 (79.6%)	
Mean size of tumor (mm)	25.7±16.1	23.8±13.5	0.048
US features*			0.053
High suspicion	15 (35.7%)	20 (17.7%)	
Intermediate suspicion	12 (28.6%)	29 (25.7%)	
Low suspicion	13 (31.0%)	49 (43.4%)	
Very low suspicion	2 (4.8%)	15 (13.3%)	
Benign	0	0	

*Based on the 2015 American Thyroid Association management guidelines.

Among 155 enrolled nodules, 86 nodules (55.5%) were confirmed by surgery. All of the 8 FN/SFN (Bethesda System IV) nodules (100.0%) were confirmed by surgery (Figure 2). Of 86 surgically resected nodules, 47 (54.7%) were benign and 39 (45.3%) were malignant. Table 2 summarizes the final pathologic diagnosis of the 86 surgically confirmed nodules.

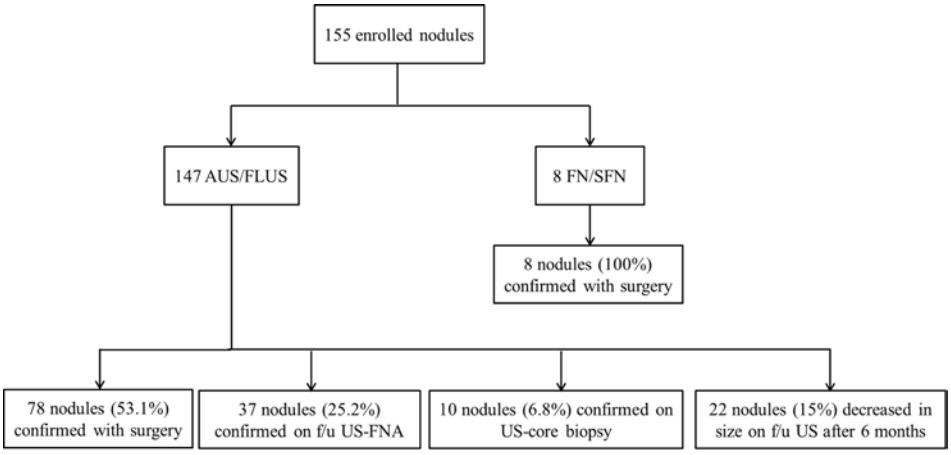

Figure 2. Pathological confirm methods of enrolled nodules

Table 2. Histopathological results of surgically resected nodules

	Histopathological Results	Number (%)
Benign (n=47)	Follicular adenoma	19 (40.4%)
	Adenomatous hyperplasia	15 (31.9%)
	Hurthle cell adenoma	10 (21.3%)
	Follicular proliferative lesion	1 (2.1%)
	Reactive hyperplasia	1 (2.1%)
	Cyst with degeneration	1 (2.1%)
Malignant (n=39)	Papillary carcinoma, conventional	13 (33.3%)
	Follicular carcinoma, Minimally invasive	11 (28.2%)
	Papillary carcinoma, follicular variant	10 (25.6%)
	Medullary carcinoma	3 (7.7%)
	Papillary carcinoma, solid variant	1 (2.6%)
	Anaplastic carcinoma, arising from follicular carcinoma	1 (2.6%)

2. Selection of texture features and construction of radiomics signature

Among 730 texture features, 15 features were selected using lambda

(the effective degrees of freedom) as tuning parameters in the LASSO logistic regression model with ten-fold cross-validation (Figure 3).

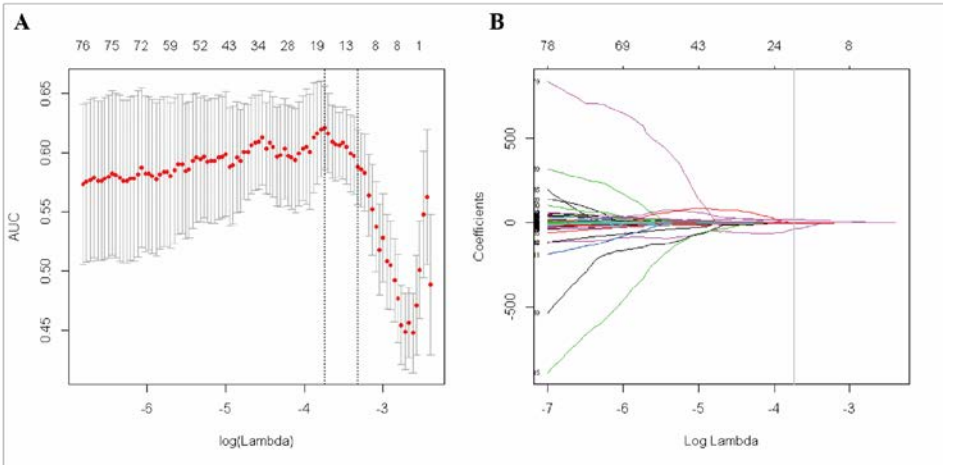

Figure 3. Radiomics feature selection using the LASSO logistic regression model.

A. Selection of tuning parameters (lambda) in the LASSO logistic regression model with ten-fold cross-validation. The area under the curve (AUC) was plotted versus log (lambda). Vertical lines were drawn at the optimal values using the minimum criteria and the 1-standard error (1-SE) criteria. 15 features were selected with the maximal AUC.

B. LASSO coefficient profile plot of the 15 selected features. A vertical line was plotted at the selected value using ten-fold cross-validation, where optimal lambda resulted in 15 features with nonzero coefficients

A radiomics signature was constructed with a radiomics score calculated using the following formula:

Radiomics score = 649.9212647 - 0.0000645 X cp_24_0 + 19.1741088 X rp_49_0
- 0.000031 X lrhgle_55_0 - 0.0000425 X cor_28_45
- 5.4475087 X lglre_50_90 - 0.0002809 X cor_28_135

- 41.0713995 X HH_rms_10_0 - 0.0044085 X HH_lrlgle_54_0
- 2.9513879 X HH_imc1_35_135 - 3.1804719 X HL_imc1_35_0
- 5.4042076 X HL_imc1_35_90 + 0.1033506 X LH_iv_39_0
+ 0.0000004 X LH_rln_48_0 - 0.5962091 X LH_mp_40_90
- 0.0087816 X LL_srhgle_53_90

The mean radiomics score of the malignant nodule group is -0.514 (range from -1.528 to 1.432) and the mean radiomics score of the benign nodule group is -1.382 (range from -7.769 to -0.036).

3. Prediction of thyroid malignancy

Univariable logistic regression analysis revealed that the Bethesda category (AUS/FLUS (Bethesda System III) vs FN/SFN (Bethesda System IV)) (odds ratio [OR] = 4.910; 95% CI = 1.149-24.884; *P* = 0.035) and radiomics scores (OR = 9.126; 95% CI = 4.218-22.473; *P* < 0.001) were predictive factors of malignancy. ATA classification was not a predictive factor of malignancy (P = 0.064). Multivariable logistic regression analysis in our cohort found that the variables independently associated with malignancy were age (OR = 0.963; 95% CI = 0.930-0.995; *P* = 0.029) and radiomics score (OR = 10.923; 95% CI = 4.735-29.415; *P* <0.001) (Table 3).

After internal validation, the AUC value of model 2 (clinical variables + radiomics score) was higher than the AUC value of model 1 (clinical variables alone) (0.839 vs 0.583). The predictability of the two models was statistically different (difference between bootstrapped AUCs of the two ROC curves = 0.256; 95% CI = 0.138-0.423) (Table 3).

Table 3. Multivariable logistic regression model for predicting malignancy

	Model 1 (Clinical variables)		Model 2 (Clinical variables + Radiomics score)	
AUC (95% CI)	0.593 (0.487, 0.698)		0.841 (0.777, 0.905)	
Bootstrapped AUC (95% CI)	0.583 (0.435, 0.693)		0.839 (0.775, 0.897)	
Variables	Odds Ratio (95% CI)	*P*-value	Odds Ratio (95% CI)	*P*-value
Size	1.010 (0.985, 1.035)	0.414	1.003 (0.974, 1.033)	0.817
Men	1.169 (0.493, 2.719)	0.722	1.495 (0.526, 4.135)	0.440
Age	0.987 (0.960, 1.014)	0.333	0.963 (0.93, 0.995)	0.029
FN/SFN cytology	4.492 (1.031, 23.096)	0.145	2.327 (0.34, 19.535)	0.405
Radiomics score			10.923 (4.735, 29.415)	<0.001

In the subgroup analysis of surgically confirmed nodules (86/155, 55.5%), a multivariable logistic regression analysis found that the radiomics score was the only factor independently associated with malignancy (OR = 11.503; 95% CI = 4.144-41.444; *P* <0.001). After internal validation, the AUC value of model 2sub (clinical variables + radiomics score) was higher than the AUC value of model 1sub (clinical variables alone) (0.831 vs 0.626). The predictability of the two models was statistically different (difference between bootstrapped AUCs of the two ROC curves = 0.206; 95% CI =

0.076-0.337) (Table 4).

Table 4. Multivariable logistic regression model for predicting malignancy in surgically proven nodules

	Model 1 (Clinical variables)		Model 2 (Clinical variables + Radiomics score)	
AUC (95% CI)	0.627 (0.505, 0.750)		0.841 (0.777, 0.905)	
Bootstrapped AUC (95% CI)	0.626 (0.486, 0.748)		0.831 (0.737, 0.904)	
Variables	Odds Ratio (95% CI)	*P*-value	Odds Ratio (95% CI)	*P*-value
Size	0.983 (0.954, 1.011)	0.243	0.972 (0.936, 1.005)	0.108
Men	1.141 (0.393, 3.296)	0.806	2.579 (0.689, 10.572)	0.168
Age	0.997 (0.963, 1.032)	0.871	0.953 (0.903, 0.998)	0.054
FN/SFN cytology	1.955 (0.437, 10.276)	0.388	1.061 (0.143, 8.608)	0.953
Radiomics score			11.503 (4.144, 41.444)	<0.001

IV. DISCUSSION

Radiomics is a series of processes that extract a large number of quantitative features from digital images, then mine the data for hypothesis generation/testing, and finally develop decision support tools. Since digital radiologic images are obtained for almost all cancer patients, and all of these

images are potential sources for the radiomics database, we anticipate radiomics to play a new role in oncology such as cancer diagnosis and prediction of prognosis or treatment response. Although radiomics is a relatively new field and practical issues have to be resolved before its implementation in clinical settings, visualization of tumor heterogeneity is expected to play an important role in tumor evaluation [20].

Our study demonstrates that quantitative radiomics features can help predict malignancy in thyroid nodules with indeterminate cytology. Adding a radiomics score to the prediction model increases the predictability of malignancy among indeterminate cytology nodules compared to a model constructed only with clinical variables.

Indeterminate cytology readings including Bethesda category III and IV comprise approximately 20-30% of all FNA results and malignancy risk is reported as 10-30% in Bethesda category III, and 25-40% in Bethesda category IV[5]. To reduce the diagnostic uncertainty of cytologically indeterminate nodules prior to surgery, many studies have been attempted, using additional molecular tests or US features [30-33]. Currently, there are two main molecular tests available: the Afirma Gene Expression Classifier (GEC) (Veracyte Inc, South San Francisco, California) which uses gene expression from mRNA, and the ThyroSeq gene mutation panel (GMP) developed by a team at the University of Pittsburgh Medical Center (UPMC).

The GEC showed high sensitivity (92%) and high negative predictive value (NPV) (93%) for predicting malignancy among nodules of indeterminate cytology [30]. Therefore, GEC-negative nodules with indeterminate cytology are considered as candidates for clinical observation instead of performing diagnostic surgery [30,34]. However, the GEC has fairly low specificity (52%), and thus, limited accuracy for identifying malignant nodules [30,35,36]. Especially, GEC analysis has limited value in Hurthle cell neoplasms [37,38]. Advances have been made to the GMP over the last few years, and the latest version has shown high

sensitivity (98%) and specificity (81.8%) [31,39-41]. The main value of the GMP is also strong NPV, up to 97-100%, that helps some patients avoid unnecessary surgery [42,43]. However, binary classification of nodules as positive or negative on the GMP does not add additional value to the risk stratification of indeterminate nodules, and the sensitivity and specificity of GMP decreases in Hurthle cell neoplasms (92.9% and 69.3% respectively) [41,43].

Molecular-based tests are still expensive and comprehensive tests are restricted to only a few, highly specialized laboratories [18]. Instead, a review of patients' US images does not add costs. High-risk US features have been previously associated with malignancy in thyroid nodules with indeterminate cytology [32,33]. Other studies have demonstrated that various criteria for malignant US features, including the Thyroid Imaging Reporting And Data System (TI-RADS) and American Thyroid Association (ATA) classification, have similar sensitivity and NPV for diagnosing malignancy among indeterminate thyroid nodules compared to molecular tests [44,45]. However, relatively low specificity has been a common problem and there is considerable intra-observer and inter-observer variability in the visual interpretation of US images, especially according to experience level of the performing radiologists [14,45-47].

Although interest in quantitative imaging biomarkers is increasing, the application of radiomics in thyroid oncology has been limited to predicting mutations or nodal metastasis status [48-51]. A recent study reported that radiomics can be used to predict malignancy in thyroid nodules and may even outperform American College of Radiology (ACR) TI-RADS scoring by junior radiologists, but follow-up studies have not been conducted [26]. In this study, visual analysis of US images which were reported with the ATA classifications did not reveal statistically significant differences between benign and malignant nodules. However, the radiomics score which was composed of 15 radiomics features was independent variable to predict malignancy in indeterminate thyroid

nodules. Furthermore, adding the radiomics score to the malignancy prediction model which was originally composed of only clinical features enhanced its predictive ability. Our results show that US image-derived radiomics features may have the potential to act as a non-invasive objective biomarker to predict malignancy among cytologically indeterminate thyroid nodules.

There are several limitations to this study. First, our study was of retrospective design and patients were collected from a single tertiary referral center, which can lead to a selection bias. Second, thyroid nodules in this study were classified according to the ATA guidelines by using prospectively collected data analyzed by radiologists with widely varying levels of experience (1 to 22 years) in thyroid imaging. Therefore, interobserver variability may exist. Third, sample size was relatively small, especially for the Bethesda category IV (FN/SFN) nodules. Further research based on larger data is needed to confirm our results. Fourth, radiomics features can be affected by the type of US machine used, and this in turn may have affected our results. Fifth, when saving exported representative US images from the PACS, we used the Paint program and JPEG format. This process may have caused some data loss, and been the reason for artifacts seen in later texture analysis. Sixth, the prevalence of each subtype of thyroid cancer differs from country to country. The Korean population is known for its high papillary thyroid cancer prevalence [52]. Consequently, results would differ because of the different prevalence even when the same study protocols are applied to populations from other countries. Last, the prediction model that we proposed was not validated externally with other cohorts. Therefore, our findings might be difficult to generalize.

V. CONCLUSION

In conclusion, quantitative US radiomics features can help predict malignancy in thyroid nodules with indeterminate cytology. Furthermore, a prediction model composed of clinical variables and US radiomics features may

aid in determining the extent of resection before surgical treatment; if a nodule with indeterminate cytology is at high risk for malignancy according to our prediction model, the surgeon should be informed that the likelihood of total thyroidectomy needing to be performed after frozen biopsy is high. Further multi-center studies with more thyroid nodules are needed to verify our study results.

REFERENCES

1. Vander JB, Gaston EA, Dawber TR. The significance of nontoxic thyroid nodules. Final report of a 15-year study of the incidence of thyroid malignancy. Ann Intern Med 1968;69:537-40.

2. Tunbridge WM, Evered DC, Hall R, Appleton D, Brewis M, Clark F, et al. The spectrum of thyroid disease in a community: the Whickham survey. Clin Endocrinol (Oxf) 1977;7:481-93.

3. Tan GH, Gharib H. Thyroid incidentalomas: management approaches to nonpalpable nodules discovered incidentally on thyroid imaging. Ann Intern Med 1997;126:226-31.

4. Haugen BR, Alexander EK, Bible KC, Doherty GM, Mandel SJ, Nikiforov YE, et al. 2015 American Thyroid Association Management Guidelines for Adult Patients with Thyroid Nodules and Differentiated Thyroid Cancer: The American Thyroid Association Guidelines Task Force on Thyroid Nodules and Differentiated Thyroid Cancer. Thyroid 2016;26:1-133.

5. Cibas ES, Ali SZ. The 2017 Bethesda System for Reporting Thyroid Cytopathology. Thyroid 2017;27:1341-6.

6. Yoon JH, Moon HJ, Kim EK, Kwak JY. Inadequate cytology in thyroid nodules: should we repeat aspiration or follow-up? Ann Surg Oncol 2011;18:1282-9.

7. Yoon JH, Kwak JY, Kim EK, Moon HJ, Kim MJ, Kim JY, et al. How to approach thyroid nodules with indeterminate cytology. Ann Surg Oncol 2010;17:2147-55.

8. Kim DW, Lee EJ, Jung SJ, Ryu JH, Kim YM. Role of sonographic diagnosis in managing Bethesda class III nodules. AJNR Am J Neuroradiol 2011;32:2136-41.

9. Yoon JH, Lee HS, Kim EK, Moon HJ, Kwak JY. A nomogram for predicting malignancy in thyroid nodules diagnosed as atypia of undetermined significance/follicular lesions of undetermined significance on fine needle

aspiration. Surgery 2014;155:1006-13.

10. Grani G, Lamartina L, Ascoli V, Bosco D, Nardi F, D'Ambrosio F, et al. Ultrasonography scoring systems can rule out malignancy in cytologically indeterminate thyroid nodules. 2017;57:256-61.

11. Xing M, Tufano RP, Tufaro AP, Basaria S, Ewertz M, Rosenbaum E, et al. Detection of BRAF mutation on fine needle aspiration biopsy specimens: a new diagnostic tool for papillary thyroid cancer. J Clin Endocrinol Metab 2004;89:2867-72.

12. Yoon JH, Kwon HJ, Lee HS, Kim EK, Moon HJ, Kwak JY. RAS Mutations in AUS/FLUS Cytology: Does it Have an Additional Role in BRAFV600E Mutation-Negative Nodules? Medicine (Baltimore) 2015;94:e1084.

13. Mendez W, Rodgers SE, Lew JI, Montano R, Solorzano CC. Role of surgeon-performed ultrasound in predicting malignancy in patients with indeterminate thyroid nodules. Ann Surg Oncol 2008;15:2487-92.

14. Choi SH, Kim EK, Kwak JY, Kim MJ, Son EJ. Interobserver and intraobserver variations in ultrasound assessment of thyroid nodules. Thyroid 2010;20:167-72.

15. Park CS, Kim SH, Jung SL, Kang BJ, Kim JY, Choi JJ, et al. Observer variability in the sonographic evaluation of thyroid nodules. J Clin Ultrasound 2010;38:287-93.

16. Park SJ, Park SH, Choi YJ, Kim DW, Son EJ, Lee HS, et al. Interobserver variability and diagnostic performance in US assessment of thyroid nodule according to size. Ultraschall Med 2012;33:E186-90.

17. Koh J, Choi JR, Han KH, Kim EK, Yoon JH, Moon HJ, et al. Proper indication of BRAF(V600E) mutation testing in fine-needle aspirates of thyroid nodules. PLoS One 2013;8:e64505.

18. Sciacchitano S, Lavra L, Ulivieri A, Magi F, De Francesco GP, Bellotti C, et al. Comparative analysis of diagnostic performance, feasibility and cost of different test-methods for thyroid nodules with indeterminate cytology.

Oncotarget 2017;8:49421-42.

19. Piccardo A, Puntoni M, Dezzana M, Bottoni G, Foppiani L, Marugo A, et al. Indeterminate thyroid nodules. The role of (18)F-FDG PET/CT in the "era" of ultrasonography risk stratification systems and new thyroid cytology classifications. Endocrine 2020;69:553-61.

20. Gillies RJ, Kinahan PE, Hricak H. Radiomics: Images Are More than Pictures, They Are Data. Radiology 2016;278:563-77.

21. Lambin P, Rios-Velazquez E, Leijenaar R, Carvalho S, van Stiphout RG, Granton P, et al. Radiomics: extracting more information from medical images using advanced feature analysis. Eur J Cancer 2012;48:441-6.

22. Nie K, Al-Hallaq H, Li XA, Benedict SH, Sohn JW, Moran JM, et al. NCTN Assessment on Current Applications of Radiomics in Oncology. Int J Radiat Oncol Biol Phys 2019;104:302-15.

23. Liu T, Zhou S, Yu J, Guo Y, Wang Y, Zhou J, et al. Prediction of Lymph Node Metastasis in Patients With Papillary Thyroid Carcinoma: A Radiomics Method Based on Preoperative Ultrasound Images. Technol Cancer Res Treat 2019;18:1533033819831713.

24. Liu T, Ge X, Yu J, Guo Y, Wang Y, Wang W, et al. Comparison of the application of B-mode and strain elastography ultrasound in the estimation of lymph node metastasis of papillary thyroid carcinoma based on a radiomics approach. Int J Comput Assist Radiol Surg 2018;13:1617-27.

25. Park VY, Han K, Lee E, Kim EK, Moon HJ, Yoon JH, et al. Association Between Radiomics Signature and Disease-Free Survival in Conventional Papillary Thyroid Carcinoma. Sci Rep 2019;9:4501.

26. Liang J, Huang X, Hu H, Liu Y, Zhou Q, Cao Q, et al. Predicting Malignancy in Thyroid Nodules: Radiomics Score Versus 2017 American College of Radiology Thyroid Imaging, Reporting and Data System. Thyroid 2018;28:1024-33.

27. Zhao CK, Ren TT, Yin YF, Shi H, Wang HX, Zhou BY, et al. A

Comparative Analysis of Two Machine Learning-Based Diagnostic Patterns with Thyroid Imaging Reporting and Data System for Thyroid Nodules: Diagnostic Performance and Unnecessary Biopsy Rate. Thyroid 2020.

28. Kim EK, Park CS, Chung WY, Oh KK, Kim DI, Lee JT, et al. New sonographic criteria for recommending fine-needle aspiration biopsy of nonpalpable solid nodules of the thyroid. AJR Am J Roentgenol 2002;178:687-91.

29. Cibas ES, Ali SZ. The Bethesda System For Reporting Thyroid Cytopathology. Am J Clin Pathol 2009;132:658-65.

30. Alexander EK, Kennedy GC, Baloch ZW, Cibas ES, Chudova D, Diggans J, et al. Preoperative diagnosis of benign thyroid nodules with indeterminate cytology. N Engl J Med 2012;367:705-15.

31. Nikiforova MN, Wald AI, Roy S, Durso MB, Nikiforov YE. Targeted next-generation sequencing panel (ThyroSeq) for detection of mutations in thyroid cancer. J Clin Endocrinol Metab 2013;98:E1852-60.

32. Barbosa TLM, Junior COM, Graf H, Cavalvanti T, Trippia MA, da Silveira Ugino RT, et al. ACR TI-RADS and ATA US scores are helpful for the management of thyroid nodules with indeterminate cytology. BMC Endocr Disord 2019;19:112.

33. Ahmadi S, Herbst R, Oyekunle T, Jiang X, Strickland K, Roman S, et al. USING THE ATA AND ACR TI-RADS SONOGRAPHIC CLASSIFICATIONS AS ADJUNCTIVE PREDICTORS OF MALIGNANCY FOR INDETERMINATE THYROID NODULES. Endocr Pract 2019;25:908-17.

34. Kloos RT. Molecular Profiling of Thyroid Nodules: Current Role for the Afirma Gene Expression Classifier on Clinical Decision Making. Mol Imaging Radionucl Ther 2017;26:36-49.

35. Deaver KE, Haugen BR, Pozdeyev N, Marshall CB. Outcomes of Bethesda categories III and IV thyroid nodules over 5 years and performance of the

Afirma gene expression classifier: A single-institution study. Clin Endocrinol (Oxf) 2018;89:226-32.

36. Lastra RR, Pramick MR, Crammer CJ, LiVolsi VA, Baloch ZW. Implications of a suspicious afirma test result in thyroid fine-needle aspiration cytology: an institutional experience. Cancer Cytopathol 2014;122:737-44.

37. Wu JX, Young S, Hung ML, Li N, Yang SE, Cheung DS, et al. Clinical Factors Influencing the Performance of Gene Expression Classifier Testing in Indeterminate Thyroid Nodules. Thyroid 2016;26:916-22.

38. Brauner E, Holmes BJ, Krane JF, Nishino M, Zurakowski D, Hennessey JV, et al. Performance of the Afirma Gene Expression Classifier in Hürthle Cell Thyroid Nodules Differs from Other Indeterminate Thyroid Nodules. Thyroid 2015;25:789-96.

39. Nikiforov YE, Ohori NP, Hodak SP, Carty SE, LeBeau SO, Ferris RL, et al. Impact of mutational testing on the diagnosis and management of patients with cytologically indeterminate thyroid nodules: a prospective analysis of 1056 FNA samples. J Clin Endocrinol Metab 2011;96:3390-7.

40. Nikiforov YE, Carty SE, Chiosea SI, Coyne C, Duvvuri U, Ferris RL, et al. Highly accurate diagnosis of cancer in thyroid nodules with follicular neoplasm/suspicious for a follicular neoplasm cytology by ThyroSeq v2 next-generation sequencing assay. Cancer 2014;120:3627-34.

41. Nikiforova MN, Mercurio S, Wald AI, Barbi de Moura M, Callenberg K, Santana-Santos L, et al. Analytical performance of the ThyroSeq v3 genomic classifier for cancer diagnosis in thyroid nodules. Cancer 2018;124:1682-90.

42. Steward DL, Carty SE, Sippel RS, Yang SP, Sosa JA, Sipos JA, et al. Performance of a Multigene Genomic Classifier in Thyroid Nodules With Indeterminate Cytology: A Prospective Blinded Multicenter Study. JAMA Oncol 2019;5:204-12.

43. Jug R, Foo WC, Jones C, Ahmadi S, Jiang XS. High-risk and intermediate-high-risk results from the ThyroSeq v2 and v3 thyroid genomic

classifier are associated with neoplasia: Independent performance assessment at an academic institution. Cancer Cytopathol 2020.

44. Gao LY, Wang Y, Jiang YX, Yang X, Liu RY, Xi XH, et al. Ultrasound is helpful to differentiate Bethesda class III thyroid nodules: A PRISMA-compliant systematic review and meta-analysis. Medicine (Baltimore) 2017;96:e6564.

45. He YP, Xu HX, Zhao CK, Sun LP, Li XL, Yue WW, et al. Cytologically indeterminate thyroid nodules: increased diagnostic performance with combination of US TI-RADS and a new scoring system. Sci Rep 2017;7:6906.

46. Trimboli P, Fulciniti F, Zilioli V, Ceriani L, Giovanella L. Accuracy of international ultrasound risk stratification systems in thyroid lesions cytologically classified as indeterminate. Diagn Cytopathol 2017;45:113-7.

47. Kim SH, Park CS, Jung SL, Kang BJ, Kim JY, Choi JJ, et al. Observer variability and the performance between faculties and residents: US criteria for benign and malignant thyroid nodules. Korean J Radiol 2010;11:149-55.

48. Yoon JH, Han K, Lee E, Lee J, Kim EK, Moon HJ, et al. Radiomics in predicting mutation status for thyroid cancer: A preliminary study using radiomics features for predicting BRAFV600E mutations in papillary thyroid carcinoma. PLoS One 2020;15:e0228968.

49. Park VY, Han K, Kim HJ, Lee E, Youk JH, Kim EK, et al. Radiomics signature for prediction of lateral lymph node metastasis in conventional papillary thyroid carcinoma. PLoS One 2020;15:e0227315.

50. Kwon MR, Shin JH, Park H, Cho H, Hahn SY, Park KW. Radiomics Study of Thyroid Ultrasound for Predicting BRAF Mutation in Papillary Thyroid Carcinoma: Preliminary Results. AJNR Am J Neuroradiol 2020;41:700-5.

51. Lu W, Zhong L, Dong D, Fang M, Dai Q, Leng S, et al. Radiomic analysis for preoperative prediction of cervical lymph node metastasis in patients with papillary thyroid carcinoma. Eur J Radiol 2019;118:231-8.

52. Cho BY, Choi HS, Park YJ, Lim JA, Ahn HY, Lee EK, et al. Changes in the

clinicopathological characteristics and outcomes of thyroid cancer in Korea over the past four decades. Thyroid 2013;23:797-804.

APPENDICES

To properly analyze ultrasound images, it is necessary to extract a series of distinct features from each image. In this study, texture-based features[1] and wavelet transformation[2] were used. Features that characterize textures are based on first- and second-order statistics. Features from first-order statistics lead to information on the gray-level distribution of pixels in images using a histogram computed by the number of gray-level pixels over the total number of pixels in the ROI. Thus, they provide characteristics that are rotation and scale invariant. The features from first-order statistics include energy, entropy, kurtosis, maximum, minimum, mean, mean absolute deviation, median, range, root mean square, skewness, standard deviation, uniformity, and variance[3]. However, first-order features do not give any information regarding the relative locations of the various gray levels in the image. For instance, first-order statistics cannot distinguish textures in different spatial arrangements but with the same gray-level value distribution. This can be achieved by taking features from second-order statistics. Here, we used the gray level co-occurrence matrix (GLCM) and gray level run-length matrix (GLRLM) that describe the occurrences of gray-level configuration, or the pattern. These matrices can be generated differently depending on distance and direction parameters. We set the distance to 1 and chose 4 different directions which corresponded to orientations of the 0, 45, 90, and 135 degree angle. As introduced in (3), the following features were collected; autocorrelation, cluster prominence, cluster shade, cluster tendency, contrast, correlation, difference entropy, dissimilarity, energy, entropy (H), homogeneity1, homogeneity 2, informational measure of correlation 1, informational measure of correlation 2, inverse difference moment normalized, inverse difference normalized, inverse variance, maximum probability, sum average, sum entropy, sum variance, variance, short run emphasis, long run emphasis, gray level non-uniformity, run length non-uniformity, run percentage, low gray level run emphasis, high gray level

run emphasis, short run low gray level emphasis, short run high gray level emphasis, long run low gray level emphasis, and long run high gray level emphasis.

Wavelets are functions that break up data into frequency components and then explore each component with a resolution that matches to its scale. Wavelet offers a more pliable way to analyze spatial content and frequency content (unlike Fourier transformation that only analyzes frequency content) using a variable-sized window, so it provides an effective representation of an image for feature extraction in pattern recognition. Here, a single-level discrete two-dimensional wavelet (Coiflet 1) transform in 4 decompositions, LL, LH, HL, HH, were established and the aforementioned first- and second-order statistics were calculated again to generate more features, with L and H standing for the respective low- and high-pass filters in the x- and y-directions.

APPENDICES REFERENCES

1. Haralick RM, Shanmugam K, Dinstein IH. Textural features for image classification. IEEE Transactions on systems, man, and cybernetics 1973:610-621.

2. El-Dahshan E, Salem A-BM, Younis TH. A hybrid technique for automatic MRI brain images classification. Studia Univ. Babes-Bolyai, Informatica 2009:54:55-67.

3. Aerts HJ, Velazquez ER, Leijenaar RT, Parmar C, Grossmann P, Carvalho S, Bussink J, Monshouwer R, Haibe-Kains B, Rietveld D. Decoding tumour phenotype by noninvasive imaging using a quantitative radiomics approach. Nature communications 2014:5:1-9.

ABSTRACT (IN KOREAN)

세포 검사에서 불확정 결과를 얻은 갑상선 결절의 악성 가능성을 예측하기 위한 초음파 radiomics signature의 적용

<지도교수 곽진영>

연세대학교 대학원 의학과

윤 지 영

목적: 이 연구의 목적은 세포 검사에서 AUS/FLUS 및 FN/SFN의 결과를 얻은 갑상선 결절이 악성 종양일 가능성을 예측하기 위해 초음파 이미지를 사용하여 얻은 radiomics 점수의 역할을 평가하는 것이다.

방법: 본원에서 진단 목적으로 최초의 초음파 유도하 세침 흡인술을 받은 154 명의 환자의 총 155 개의 불확정 갑상선 결절이 이 후향적 연구에 포함되었다. 각 종양에서 한 장의 대표 초음파 영상을 획득하고, Window 7의 그림판 프로그램을 사용하여 전체 결절을 덮는 정사각형 모양의 ROI를 그렸다. 텍스처 특성을 MATLAB 2019b에서 구현 된 사내 텍스처 분석 알고리즘으로 추출되었다. LASSO 로지스틱 회귀 모델을 사용하여 가장 유용한 예측 특성을 선택하고, 10배 교차 검증을 수행했다. 다변수 로지스틱 회귀 분석을 사용하여 두 가지 예측 모델을 구성했는데, 하나는 임상 변수만을 기반으로 하였고, 다른 하나는 임상 변수와 radiomics 점수를 기반으로 하였다. 두 모델의 악성 예측 능력은 ROC 곡선의 AUC로 평가 및 비교되었다.

결과: 평균 결절 크기를 제외하고는 악성 결절과 양성 결절 간에 유의한 차이를 보이는 임상적 특징이 없었다. 단일 초음파 이미지에서 생성 된 730개의 후보 텍스처 특성 중 15 개의 특성이 선택되었다. Radiomics signature는 선택된 특성들을

사용하여 각 결절 별 하나의 radiomics 점수로 구성되었다. 다변수 로지스틱 회귀 분석에서 더 높은 radiomics 점수가 악성 종양과 관련이 있었다 (OR = 10.923; P <0.001). Radiomics 점수와 임상 변수로 구성된 악성 예측 모델의 AUC가 임상 변수만으로 구성된 예측 모델과 비교하여 통계학적으로 유의하게 높았다 (0.839 vs 0.583).

결론: 정량적인 초음파 radiomics 특성은 세포 검사에서 불확실한 결과를 얻은 갑상선 결절에서 악성 종양을 예측하는 데 도움이 될 수 있다.

핵심 단어: 초음파 영상, 갑상선 결절, 컴퓨터의 도움을 받은 영상 해석, 컴퓨터의 도움을 받은 진단, 갑상선 신생물

PUBLICATION LIST

1. Yoon J, Lee E, Kang SW, Han K, Park VY, Kwak JY. Implications of US radiomics signature for predicting malignancy in thyroid nodules with indeterminate cytology. Eur Radiol 2021.

2. Yoon J, Lee E, Koo JS, Yoon JH, Nam KH, Lee J, et al. Artificial intelligence to predict the BRAFV600E mutation in patients with thyroid cancer. PLoS One 2020;15:e0242806.

3. Yoon J, Yoon JH, Han K, Lee J, Kim EK, Moon HJ, et al. Ultrasonography surveillance in papillary thyroid carcinoma patients after total thyroidectomy according to dynamic risk stratification. Endocrine 2020;69:347-57.