Nuclear Medicine and Molecular Imaging (2020) 54:241–248
https://doi.org/10.1007/s13139-020-00659-2

ORIGINAL ARTICLE

ISSN (print) 1869-3482
ISSN (online) 1869-3474

Evaluating Focal ^{18}F-FDG Uptake in Thyroid Gland with Radiomics

Ayşegül Aksu[1] · Nazlı Pınar Karahan Şen[1] · Emine Acar[2,3] · Gamze Çapa Kaya[1]

Received: 28 March 2020 / Revised: 17 July 2020 / Accepted: 23 July 2020 / Published online: 28 July 2020

Abstract

Purpose The aim of this study was to evaluate the ability of ^{18}F-FDG PET/CT texture analysis to predict the exact pathological outcome of thyroid incidentalomas.

Methods ^{18}F-FDG PET/CT images between March 2010 and September 2018 were retrospectively reviewed in patients with focal ^{18}F-FDG uptake in the thyroid gland and who underwent fine needle aspiration biopsy from this area. The focal uptake in the thyroid gland was drawn in 3D with 40% SUVmax threshold. Features were extracted from volume of interest (VOI) using the LIFEx package. The features obtained were compared in benign and malignant groups, and statistically significant variables were evaluated by receiver operating curve (ROC) analysis. The correlation between the variables with area under curve (AUC) value over 0.7 was examined; variables with correlation coefficient less than 0.6 were evaluated with machine learning algorithms.

Results Sixty patients (70% train set, 30% test set) were included in the study. In univariate analysis, a statistically significant difference was observed in 6 conventional parameters, 5 first-, and 16 second-order features between benign and malignant groups in train set ($p < 0.05$). The feature with the highest benign-malignant discriminating power was $GLRLM_{RLNU}$ (AUC:0.827). AUC value of SUVmax was calculated as 0.758. $GLRLM_{RLNU}$ and SUVmax were evaluated to build a model to predict the exact pathology outcome. Random forest algorithm showed the best accuracy and AUC (78.6% and 0.849, respectively).

Conclusion In the differentiation of benign-malignant thyroid incidentalomas, $GLRLM_{RLNU}$ and SUVmax combination may be more useful than SUVmax to predict the outcome.

Keywords Thyroid incidentaloma · FDG · PET · Radiomics · Texture

Introduction

Especially in the field of oncology, with the increase in the use of ^{18}F-fluorodeoxyglucose (FDG) positron emission tomography/computed tomography (PET/CT), an increase in the detection of some unexpected lesions is observed in whole body imaging [1]. Thyroid incidentaloma is a thyroid lesion detected in non-thyroid imaging. Benign-malignant differentiation of focal ^{18}F-FDG uptake in thyroid gland in PET/CT is important for both clinicians and nuclear medicine physicians. For this purpose, the most studied parameter is the standard uptake value (SUV). In many studies, there was a significant difference in SUV between benign and malignant thyroid lesions. However, it is still difficult to distinguish between benign and malignant thyroid incidentalomas by SUV alone due to significant overlap between benign and malignant lesions. Also, in some studies, a certain SUVmax cut-off value could not be determined which could provide a distinction between benign and malignant lesions [2–10].

✉ Ayşegül Aksu
aaysegulgedikli@gmail.com

Nazlı Pınar Karahan Şen
drpinarkarahan@hotmail.com

Emine Acar
emineacar87@gmail.com

Gamze Çapa Kaya
gamze.capa@hotmail.com

[1] Department of Nuclear Medicine, School of Medicine, Dokuz Eylul University, İzmir, Turkey

[2] Department of Nuclear Medicine, Kent Hospital, İzmir, Turkey

[3] Department of Translational Oncology, Institute of Health Sciences, Dokuz Eylul University, İzmir, Turkey

In some studies, the use of parameters other than SUVmax in the differentiation of benign and malignant lesions was investigated. Volume-based parameters such as total lesion glycolysis (TLG) and metabolic tumor volume (MTV) were considered to have potential clinical value as well as SUVmax in the differential diagnosis of thyroid incidentaloma [11, 12].

Radiomics is a rapidly developing field of research that expresses quantitative data extraction and analysis from medical images such as CT or PET [13]. Radiomic approach has recently increased attention because radiomic data can help in the diagnosis, prognosis, and predicting the response of the disease [14–18]. The main idea underlying radiomic is that there is hidden information that cannot be seen in medical images. This information can be accessed by advanced texture and shape analysis. Texture analysis refers to various mathematical models to measure the relationships between the signal intensity of pixels and their relative position in the image.

Few data are available in the literature regarding ^{18}F-FDG PET/CT radiomics in thyroid nodules. Sollini et al. evaluated thyroid incidentalomas with texture analysis and predicted that some parameters could provide benign and malignant differentiation [19]. However, machine learning modeling was not performed in this study. The aim of our study was to evaluate focal ^{18}F-FDG uptake detected by PET/CT with SUV and texture parameters, to be able to predict the definitive diagnosis with radiomic model.

Materials and Methods

Patients

PET/CT images performed in our center between March 2010 and September 2018 were retrospectively analyzed. Patients who had focal ^{18}F-FDG uptake in the thyroid gland and who underwent fine needle aspiration biopsy (FNAB) from this area were selected. In patients included in our study, PET/CT imaging was performed for oncological purposes (diagnosis, staging, restaging, treatment response evaluation) in 57 patients and non-oncological (infectious/inflammatory diseases) for 3 patients. Focal ^{18}F-FDG involvement in the thyroid gland was detected incidentally in these patients, and PET/CT imaging was not performed to evaluate the thyroid nodule specifically. These patients underwent ultrasound-guided FNAB. Cytology and/or histopathology results were used as reference to identify the exact diagnosis: benign or malignant. In case of uncertain significance of atypia or suspicious results, thyroidectomy results were accepted as gold standard to determine the definitive diagnosis.

Patients with inadequate cytology result were excluded. Also, diffuse ^{18}F-FDG uptake in the thyroid gland was not included in this study because diffuse involvement is often associated with benign diseases [4]. In addition, patients with atypia of undetermined significance or suspicious results who had no definitive pathology were excluded from the study.

PET/CT Imaging

Patients with appropriate patient preparation (fasting for at least 4 h) and adequate blood glucose levels were enrolled in PET/CT. Approximately 1 h after the average injection of 4.07 Megabecquerels (MBq)/kg (0.11 miliCurie/kg) ^{18}F-FDG, first, a contrast-free CT scan from the vertex to the middle of the thigh was obtained using the following parameter in Philips Gemini TOF PET/CT (Eindhoven, Netherlands): 120 kVp; 50 mAs; 5.0-mm re-structured section thickness. After the CT imaging was completed, the PET scan in the same area was performed with a position of 10–12 beds per patient and 1.5 min/bed position. Images were reconstructed using a row action maximum likelihood algorithm.

Texture Analysis

PET/CT images of the patients included in the study were analyzed by two nuclear medical physicians by the program Local Image Features Extraction (LIFEx) package (http://www.lifexsoft.org) [20]. The focal ^{18}F-FDG uptake, which can be distinguished from the peripheral thyroid tissue, was drawn with a 40% SUVmax threshold in 3D from PET images [21, 22]. Texture matrices were computed after resampling to a 4 mm × 4 mm × 4 mm grid, with 64 bins and absolute scale bounds between 0 and 35. Conventional parameters (SUVmin, SUVmean, SUVstd, SUVmax, SUVpeak, TLG), first-order (histogram and shape), and second-order features (gray-level co-occurrence matrix: GLCM, neighborhood gray-level different matrix: NGLDM, gray-level run-length matrix: GLRLM, gray-level zone-length matrix: GLZLM), in total, 46 features were extracted from the obtained VOI. The evaluated parameters are shown in Table 1.

Statistical Analysis

Statistical analysis was performed with SPSS v24.0 (IBM, USA). The patients randomly split into the train (70%) and test (30%) sets, as in the literature [23–25]. Normally distributed data were presented as mean ± standard deviation, and non-normal distributed data were given as median (range). The relationship between gender and pathology was evaluated by chi-square test. In univariate analysis, the comparison of texture analysis features in benign and malignant groups was performed by Mann-Whitney *U* test. Significant parameters were evaluated by ROC analysis. The correlation between the variables with AUC value over 0.7 was examined by spearman correlation test. The features with a correlation coefficient of above 0.6 were not included in the model [26]. The features with a correlation coefficient of less than 0.6 were

Table 1 Extracted features by LIFEx

Conventional	SUVmin, SUVmax, SUVmean, SUVstd, SUVpeak, TLG
Histogram based	Skewness, kurtosis, entropy (log 2&10), energy
Shape based	Volume, sphericity, compacity
GLCM	Homogeneity, energy, contrast, correlation, entropy (log 2&10), dissimilarity
GLRLM	SRE, LRE, LGRE, HGRE, SRLGE, SRHGE, LRLGE, LRHGE, GLNU, RLNU, RP
NGLDM	Coarseness, contrast, busyness
GLZLM	SZE, LZE, LGZE, HGZE, SLZLGE, SZHGE, LZLGE, LZHGE, GLNU, ZLNU, ZP

SUV, standard uptake value; *TLG*, total lesion glycolysis; *GLCM*, gray-level co-occurrence matrix; *GLRLM*, gray-level run length matrix; *NGLDM*, neighborhood gray-level different matrix; *GLZLM*, gray-level zone length matrix; *SRE*, short-run emphasis; *LRE*, long-run emphasis; *LGRE*, low gray-level run emphasis; *HGRE*, high gray-level run emphasis; *SRLGE*, short-run low gray-level emphasis; *SRGHE*, short-run high gray-level emphasis; *LRLGE*, long-run low gray-level emphasis; *LRHGE*. long-run high gray-level emphasis; *GLNU*, gray-level non-uniformity; *RLNU*, run length non-uniformity; *RP*, run percentage; *SZE*, short-zone emphasis; *LZE*, long-zone emphasis; *LGZE*, low gray-level zone emphasis; *HGZE*, high gray-level zone emphasis; *SZLGE*, short-zone low gray-level emphasis; *SZHGE*, short-zone high gray-level emphasis; *LZLGE*, long-zone low gray-level emphasis; *LZHGE*, long-zone high gray-level emphasis; *ZLNU*, zone length non-uniformity; *ZP*, zone percentage

evaluated with Weka data mining program to form a model. Five machine learning algorithms (random forest, naive bayes, *k* nearest neighbor, decision tree, and support vector machine) and logistic regression for binary risk classification were compared. To compare their predictive performance for differentiating focal ^{18}F-FDG uptake in thyroid gland between models, the receiver operating characteristic (ROC) curve analysis was used. Tenfold cross-validation was used as internal validation method. A $p < 0.05$ was considered statistically significant.

Results

Train Set

Forty-two patients with definitive pathology were included in train group. Fifty-two percent of these patients were women. The mean age of all patients was 63.5 ± 14.3 (32–88) years. Mean fasting blood glucose values of patients were 98 ± 2 g/dl (73–125 g/dl). Between PET/CT and FNAB, there was an average of 99.1 days (1–294). FNAB results were benign in 24 patients (57%), atypia of undetermined significance in two patients (5%), suspicion of malignancy in 10 patients (24%), and malignancy in six patients (14%). Sixteen patients had total thyroidectomy (Table 2). One of the patients with atypia of undetermined significance was diagnosed as papillary carcinoma due to thyroid carcinoma metastasis in right femur. Another patient who suspicious FNAB result was diagnosed with papillary carcinoma metastasis as a result of wedge biopsy from the lung. Total thyroidectomy was not performed in these patients. And also, total thyroidectomy was not performed in two patients who had malignant FNAB result due to their oncological diseases.

In total, twenty of the patients (47.6%) had malignant pathology. Fourteen patients were diagnosed with papillary thyroid carcinoma, 1 patient with follicular thyroid carcinoma, 2 with medullary thyroid carcinoma, 2 with thyroid anaplastic carcinoma, and 1 with differential squamous cell and papillary carcinoma.

Although the FNAB result obtained from two patients was benign, it was diagnosed as papillary thyroid carcinoma after thyroidectomy.

The rates of malignancy were 65% and 31.8% in males and females, respectively; there was a statistically significant difference (*p*: 0.032).

Test Set

Eighteen patients with definitive pathology were included in test group. Sixty-one percent of these patients were women.

Table 2 The number of patients undergoing biopsy and thyroidectomy

FNAB groups	Number of patients (train set)	Number of patient who had total thyroidectomy (train set)	Number of patients (test set)	Number of patient who had total thyroidectomy (test set)
Benign	24	2	10	1
Atypia of indeterminate significance	2	1	-	-
Suspicion of malignancy	10	9	4	4
Malignant	6	4	4	-
Total	42	16	18	5

The mean age of all patients was 65.7 ± 12.1 (38–84) years. Mean fasting blood glucose values of patients were 101 ± 3 g/dl (87–129 g/dl). FNAB results were benign in 10 patients (55.6%), suspicion of malignancy in four patients (22.2%), and malignancy in four patients (22.2%). Five patients had total thyroidectomy (Table 2). Four patients with suspected biopsy were diagnosed as malignant after thyroidectomy.

In total, eight of the patients (44.4%) had malignant pathology.

Radiomic Analysis

In the univariate analysis, all conventional parameters, 5 first-order features, and 16 second-order features were significantly different between benign and malignant groups in train set (Table 3). The feature with the highest benign-malignant discriminating power was $GLRLM_{RLNU}$. When the cutoff was 68.6, the sensitivity, specificity, PPV, NPV, and accuracy were 75%, 90.9%, 88.2%, 100%, and 83.3%, respectively (AUC: 0.827, 0.693–0.962, 95% CI). The median value of $GLRLM_{RLNU}$ was 60.7 (20.0–645.7).

The median value of SUVmax was 4.75 (2.2–30.4). When the cutoff was 3.9, the sensitivity, specificity, and accuracy were 90%, 55.5%, and 77.8% respectively (AUC: 0.758, 0.612–0.904, 95% CI).

AUC values of six conventional parameters, two first-order features, and 10 second-order features were greater than 0.7 (Table 4).

Table 3 The features with significantly different median (range) values between benign and malignant groups (train set)

Features	Benign group	Malignant group	*p* value
SUVmin	1.8 (1.0–2.7)	2.3 (1.1–12.1)	0.008
SUVmean	2.5 (1.4–7.4)	3.3 (2.0–19.5)	0.009
SUVstd	0.5 (0.2–2.9)	0.8 (0.3–4.0)	0.012
SUVmax	3.7 (2.2–14.7)	5.6 (2.6–30.4)	0.004
SUVpeak	2.9 (1.7–10.3)	4.3 (2.5–25.4)	0.01
TLG	14.8 (7.0–39.7)	48.1 (6.7–491.2)	< 0.001
HISTO_entropylog10	0.53 (0.23–1.22)	0.67 (0.40–1.42)	0.039
HISTO_entropylog2	1.76 (0.78–4.06)	2.21 (1.32–4.73)	0.044
SHAPEvolume	5.1 (3.3–27.7)	12.0 (3.1–61.1)	0.008
SHAPE_sphericity	1.055 (0.990–1.160)	1.025 (0.910–1.160)	0.036
SHAPE_compacity	0.90 (0.67–1.62)	1.26 (0.68–3.02)	0.004
GLCM_entropylog10	1.06 (0.45–1.73)	1.29 (0.72–2.41)	0.026
GLCM_entropylog2	3.53 (1.49–5.76)	4.28 (2.39–8.00)	0.024
GLCM_correlation	0.196 (0.051–0.506)	0.316 (0.084–0.521)	0.014
GLRLM_LGRE	0.046 (0.009–0.107)	0.026 (0.001–0.067)	0.008
GLRLM_HGRE	25.8 (10.4–226.3)	43.9 (16.7–1355.9)	0.011
GLRLM_SRLGE	0.033 (0.009–0.064)	0.023 (0.001–0.054)	0.008
GLRLM_SRHGE	20.15 (5.80–219.40)	38.00 (12.70–1307.50)	0.012
GLRLM_LRLGE	0.090 (0.010–0.805)	0.046 (0.001–0.155)	0.022
GLRLM_LRHGE	62.6 (28.5–253.9)	85.1 (40.8–1565.1)	0.006
GLRLM_RLNU	43.2 (20.0–84.5)	105.2 (24.7–645.7)	< 0.001
NGLDMcoarseness	0.074 (0.022–0.131)	0.030 (0.006–0.118)	0.001
GLZLM_LGZE	0.049 (0.010–0.161)	0.027 (0.001–0.082)	0.007
GLZLM_HGZE	25.35 (9.30–230.50)	47.30 (16.10–1284.60)	0.008
GLZLM_SZHGE	5.65 (0.01–151.50)	15.75 (0.10–932.40)	0.01
GLZLM_GLNU	2.3 (1.0–4.2)	3.7 (1.0–21.4)	0.001
GLZLM_ZLNU	1.8 (1.0–15.0)	3.2 (1.0–92.7)	0.012

SUV, standard uptake value; *TLG*, total lesion glycolysis; *GLCM*, gray-level co-occurrence matrix; *GLRLM*, gray-level run length matrix; *NGLDM*, neighborhood gray-level different matrix; *GLZLM*, gray-level zone length matrix; *LGRE*, low gray-level run emphasis; *HGRE*, high gray-level run emphasis; *SRLGE*, short-run low gray-level emphasis; *SRGHE*, short-run high gray-level emphasis; *LRLGE*, long-run low gray-level emphasis; *LRHGE*, long-run high gray-level emphasis; *GLNU*, gray-level non-uniformity; *RLNU*, run length non-uniformity; *RP*, run percentage; *LGZE*, low gray-level zone emphasis; *HGZE*, high gray-level zone emphasis; *SZHGE*, short-zone high gray-level emphasis; *ZLNU*, zone length non-uniformity

Table 4 Features with AUC greater than 0.7

Features	AUC	95% CI
SUVmin	0.739	0.586–0.891
SUVmean	0.734	0.585–0.884
SUVstd	0.725	0.572–0.878
SUVmax	0.758	0.612–0.904
SUVpeak	0.733	0.577–0.889
TLG	0.822	0.687–0.956
SHAPEvolume	0.741	0.583–0.899
SHAPEcompacity	0.759	0.607–0.911
GLCM_entropylog10	0.701	0.543–0.860
GLCM_entropylog2	0.703	0.545–0.861
GLRLM_HGRE	0.730	0.579–0.880
GLRLM_SRHGE	0.727	0.576–0.879
GLRLM_LRHGE	0.750	0.601–0.899
GLRLM_RLNU	0.827	0.700–0.964
GLZLM_HGZE	0.740	0.590–0.890
GLZLM_SZHGE	0.732	0.579–0.884
GLZLM_GLNU	0.788	0.641–0.934
GLZLM_ZLNU	0.725	0.569–0.881

SUV, standard uptake value; *TLG*, total lesion glycolysis; *GLCM*, gray-level co-occurrence matrix; *GLRLM*, gray-level run length matrix; *GLZLM*, gray-level zone length matrix; *HGRE*, high gray-level run emphasis; *SRGHE*, short-run high gray-level emphasis; *LRHGE*, long-run high gray-level emphasis; *RLNU*, run length non-uniformity; *RP*, run percentage; *HGZE*, high gray-level zone emphasis; *SZHGE*, short-zone high gray-level emphasis; *GLNU*, gray-level non-uniformity; *ZLNU*, zone length non-uniformity; *AUC*, area under curve; *CI*, confidence interval

The correlation between 18 features with AUC values above 0.7 was investigated (Fig. 1). The features with a correlation coefficient of less than 0.6 were $GLRLM_{RLNU}$ and SUVmax. Other parameters were not included in the model because the correlations between these features were high. $GLRLM_{RLNU}$ and SUVmax were evaluated to build a model to predict the exact pathology outcome. Random forest algorithm showed the best model accuracy and the highest AUC (Table 5). Sensitivity, specificity, and accuracy of the model in benign-malignant differentiation were calculated as 75%, 81.8%, and 78.6%, respectively (AUC: 0.849). The ROC curve of the model is shown in Fig. 2.

When the random forest model was tested with a previously unused test set, sensitivity, specificity, and accuracy were calculated as 75%, 80%, and 77.8%, respectively (AUC: 0.731).

Discussion

In the literature, there are limited studies on the heterogeneity of thyroid incidentalomas in ^{18}F-FDG PET/CT. Sollini et al. [19] evaluated the benign-malignant thyroid incidentaloma differentiation using conventional parameters, first-, and second-order features. In this study, $HISTO_{skewness}$ was found to be the best feature for benign-malignant differentiation (AUC = 0.66). A significant difference was observed between SUVstd, SUVmax, TLG, MTV, $HISTO_{kurtosis}$, and $GLCM_{correlation}$ features between benign and malign groups. In our study, conventional parameters, first-, and second-order features could be obtained from all patients. A univariate analysis revealed a significant difference in 27 parameters between benign and malignant groups. In contrast to the study by Sollini et al., the $HISTO_{kurtosis\ and\ skewness}$ features did not differ significantly between the benign and malign groups in our study. In the ROC analysis, the variable with the highest AUC value was $GLRLM_{RLNU}$ (AUC = 0.827). $GLRLM_{RLNU}$ is a parameter that measures the similarity of run lengths in the image, with a high value showing the heterogeneity in the image.

There are many studies in the literature about the use of SUV in benign and malignant differentiation. In some of these studies, a cutoff could be determined for SUVmax [2, 3, 6, 7], while in some studies, a suitable cutoff could not be detected [8–10]. And because SUVmax show the highest metabolic activity in a single pixel, volume-based metabolic parameters were used in literature. TLG is a semiquantitative parameter obtained by SUVmean and tumor volume and used for evaluating metabolic activity. In a study, Shi et al. [12] reported that TLG indices were useful in the differentiation of benign and malignant thyroid incidentalomas. Similarly, in our study, TLG values were higher in malignant group and there was a statistically significant difference between benign and malignant groups. Also, in our patient group, the discriminative power of TLG was higher than SUVmax. But TLG was not included in our model training because it showed a high correlation with $GLRLM_{RLNU}$.

Machine learning algorithms provide powerful modeling tools to mine the huge amount of image data available, reveal underlying complex biological mechanisms, and make personalized precision cancer diagnosis and treatment planning possible. With machine learning algorithms, we evaluated the features of AUC above 0.7 and the correlations with each other in order to prevent overfitting. SUVmax and $GLRLM_{RLNU}$ significantly contributed to model. The sensitivity, specificity, and accuracy of the model for detecting malignant lesions were 75%, 81.8%, and 78.6%. Our model's discriminating power was calculated higher than SUVmax and TLG. This has led us to conclude that the model we obtained can be used before the FNAB in the differentiation of benign-malignant thyroid incidentalomas and may be more useful than SUVmax to predict the outcome.

One of the problems in texture analysis is that the results may vary with different patient groups. Before using a model, it is important to evaluate model performance in data sets that

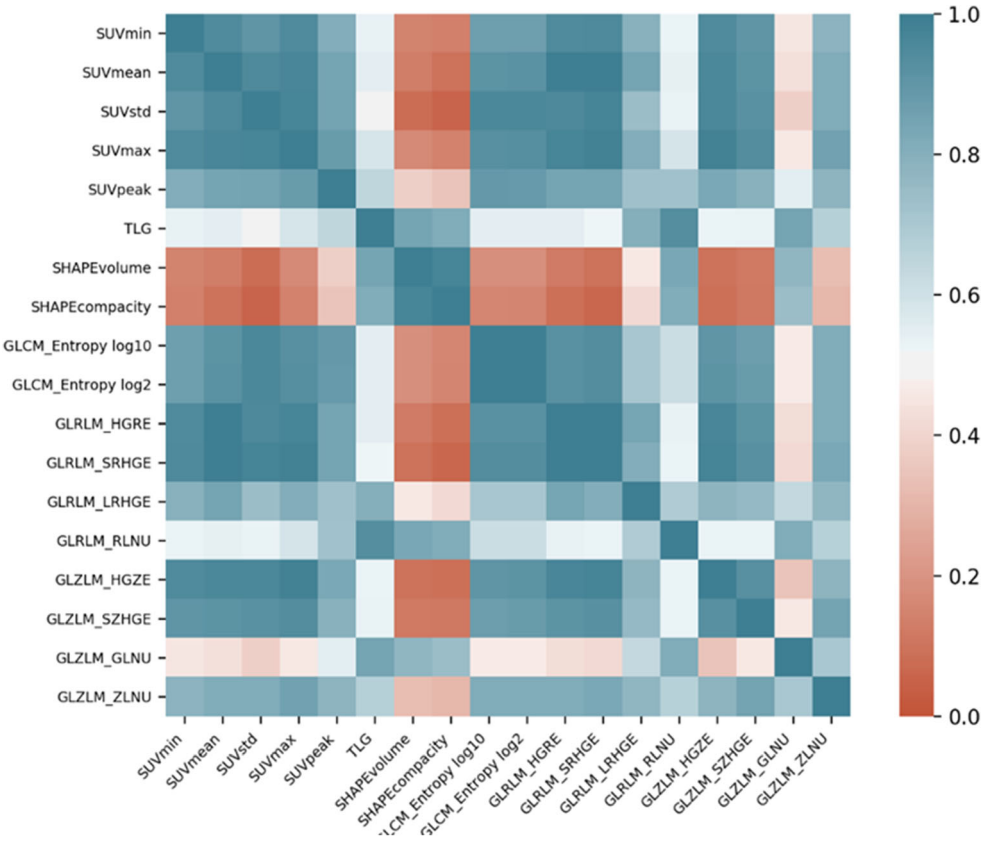

Fig. 1 The correlation matrix of features with AUC greater than 0.7

are not used to develop the model. External validation uses an independent dataset to evaluate the accuracy of the predictive model and to assess the generalizability of a predictive model. Therefore, we evaluated the performance of the model we created in our study using external validation.

In the literature, false negative rates in FNAB in thyroid nodules ranged from 1 to 39.72% [27, 28]. Cytomorphological overlap or sampling error between benign and low-grade malignant lesions can cause false negativity [29]. In two patients, FNAB results were benign, but the histopathology was malignant. In these patients, ultrasound was suspicious for malignancy and SUV values were at the level of malignancy. The model identified these patient's pathology results as malignant. It was a limitation that the definitive pathology results could not be obtained in other patients with benign results. Another limitation was that this method could not be applied to non-FDG avid nodules that can be observed in ^{18}F-FDG PET/CT.

In the LIFEx program, for technical reasons, second-order features of VOIs below 64 voxels cannot be obtained. This can lead to limitations on the use of the second-order features. Therefore, this method may not work for nodules smaller than 64 voxels. It is also known that blood glucose values, the time between injection and imaging, and the reconstruction methods may have effect on the radiomic analysis [30]. A standardization on this issue is not yet available. Therefore,

Table 5 AUC and accuracy values of models obtained by different algorithms in train set

Algorithms	Area under curve	Accuracy (%)
Naive bayes	0.714	64.3
Logistic regression	0.770	76.2
Support vector machine	0.577	59.5
k nearest neighbor	0.650	64.3
Decision tree	0.638	61.9
Random forest	0.849	78.6

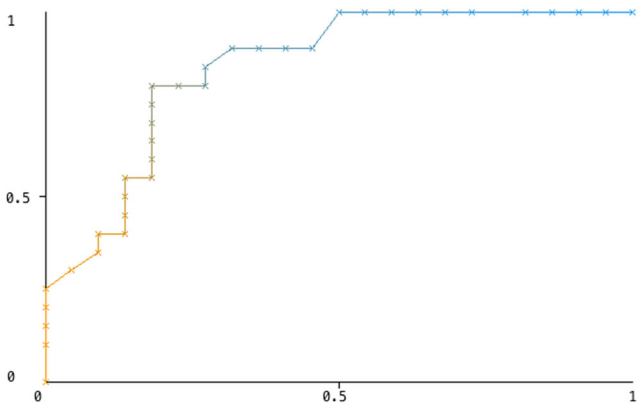

Fig. 2 The ROC curve of the model

prospective studies with larger patient groups are needed in this regard.

In conclusion, it is thought that, ^{18}F-FDG PET/CT texture analysis may be more useful than SUVmax in predicting the benign-malignant distinction of focal involvement in the thyroid gland.

Acknowledgments The authors would like to thank the members of the Dokuz Eylül University, Department of Nuclear Medicine for their technical assistance and support.

Compliance with Ethical Standards

Conflict of Interest Ayşegül Aksu, Nazlı Pınar Karahan Şen, Emine Acar, and Gamze Çapa Kaya declare that they have no conflict of interest.

Ethics Approval All procedures performed in studies involving human participants were in accordance with the ethical standards of the institutional and national research committee and with the 1964 Helsinki declaration and its later amendments or comparable ethical standards.

Informed Consent The institutional review board of our institute approved this retrospective study, and the requirement to obtain informed consent was waived.

References

1. Thuillier P, Roudaut N, Crouzeix G, Cavarec M, Robin P, Abgral R, et al. Malignancy rate of focal thyroid incidentaloma detected by FDG PET–CT: results of a prospective cohort study. Endocr Connect. 2017;6:413–21.
2. Choi JY, Lee KS, Kim HJ, Shim YM, Kwon OJ, Park K, et al. Focal thyroid lesions incidentally identified by integrated 18F-FDG PET/CT: clinical significance and improved characterization. J Nucl Med. 2006;47:609–15.
3. Pagano L, Sama MT, Morani F, Prodam F, Rudoni M, Boldorini R, et al. Thyroid incidentaloma identified by ^{18}F-fluorodeoxyglucose positron emission tomography with CT (FDG-PET/CT): clinical and pathological relevance. Clin Endocrinol. 2011;75:528–34.
4. Bertagna F, Treglia G, Piccardo A, Giubbini R. Diagnostic and clinical significance of F-18-FDG-PET/CT thyroid incidentalomas. J Clin Endocrinol Metab. 2012;97:3866–75.
5. Shie P, Cardarelli R, Sprawls K, Fulda KG, Taur A. Systematic review: prevalence of malignant incidental thyroid nodules identified on fluorine-18 fluorodeoxyglucose positron emission tomography. Nucl Med Commun. 2009;30:742–8.
6. Soelberg KK, Bonnema SJ, Brix TH, Hegedüs L. Risk of malignancy in thyroid incidentalomas detected by 18F-fluorodeoxyglucose positron emission tomography: a systematic review. Thyroid. 2012;22:918–25.
7. Cohen MS, Arslan N, Dehdashti F, Doherty GM, Lairmore TC, Brunt LM, et al. Risk of malignancy in thyroid incidentalomas identified by fluorodeoxyglucose-positron emission tomography. Surgery. 2001;130:941–6.
8. Hagenimana N, Dallaire J, Vallée É, Belzile M. Thyroid incidentalomas on 18FDG-PET/CT: a metabolico-pathological correlation. J Otolaryngol Head Neck Surg. 2017:46–22.
9. Kim JM, Ryu JS, Kim TY, Kim WB, Kwon GY, Gong G, et al. 18F-Fluorodeoxyglucose positron emission tomography does not predict malignancy in thyroid nodules cytologically diagnosed as follicular neoplasm. J Clin Endocrinol Metab. 2007;92:1630–4.
10. Are C, Hsu JF, Schoder H, Shah JP, Larson SM, Shaha AR. FDG-PET detected thyroid incidentalomas: need for further investigation? Ann Surg Oncol. 2007;14:239–47.
11. Kim BH, Kim SJ, Kim H, Jeon YK, Kim SS, Kim IJ, et al. Diagnostic value of metabolic tumor volume assessed by 18F-FDG PET/CT added to SUVmax for characterization of thyroid 18F-FDG incidentaloma. Nucl Med Commun. 2013;34:868–76.
12. Shi H, Yuan Z, Yuan Z, Yang C, Zhang J, Shou Y, et al. Diagnostic value of volume-based fluorine-18-fluorodeoxyglucose PET/CT parameters for characterizing thyroid incidentaloma. Korean J Radiol. 2018;19:342–51.
13. Parvez A, Tau N, Hussey D, Maganti M, Metser U, et al. 18F-FDG PET/CT metabolic tumor parameters and radiomics features in aggressive non-Hodgkin's lymphoma as predictors of treatment outcome and survival. Ann Nucl Med. 2018;32:410–6.
14. Zhou L, Zhang Z, Chen YC, Zhao ZY, Yin XD, Jiang HB. A deep learning-based radiomics model for differentiating benign and malignant renal tumors. Transl Oncol. 2019;12:292–300.
15. Mao N, Yin P, Wang Q, Liu M, Dong J, Zhang X, et al. Added value of radiomics on mammography for breast cancer diagnosis: a feasibility study. J Am Coll Radiol. 2019;16:485–91.
16. Abdollahi H, Mofid B, Shiri I, Razzaghdoust A, Saadipoor A, Mahdavi A, et al. Machine-learning-based radiomic models to predict intensity-modulated radiation therapy response, Gleason score and stage in prostate cancer. Radiol Med. 2019;124:555–67.
17. Yang L, Yang J, Zhou X, Huang L, Zhao W, Wang T, et al. Development of a radiomics nomogram on the 2D and 3D CT features to predict the survival of non-small cancer patients. Eur Radiol. 2019;29:2196–206.
18. Zheng BH, Liu LZ, Zhang ZZ, Shi JY, Dong LQ, Tian LY, et al. Radiomics score: a potential prognostic imaging feature for postoperative survival of solitary HCC patient. BMC Cancer. 2018;18:1148.
19. Sollini M, Cozzi L, Pepe G, Antunovic L, Lania A, Di Tommaso L, et al. [18F]FDG-PET/CT texture analysis in thyroid incidentalomas: preliminary results. Eur J Hybrid Imaging. 2017;1:3.
20. Nioche C, Orlhac F, Boughdad S, Reuzé S, Goya-Outi J, Robert C, et al. LIFEx: a freeware for radiomic feature calculation in multimodality imaging to accelerate advances in the characterization of tumor heterogeneity. Cancer Res. 2018;78:4786–9.
21. Ouyang FS, Guo BL, Zhang B, Dong YH, Zhang L, Mo XK, et al. Exploration and validation of radiomics signature as an independent prognostic biomarker in stage III-IVb nasopharyngeal carcinoma. Oncotarget. 2017;24:74869–79.
22. Huang Z, Zhang W, He D, Cui X, Tian S, Yin H, et al. Development and validation of a radiomics model based on T2WI images for preoperative prediction of microsatellite instability status in rectal cancer: study protocol clinical trial (SPIRIT Compliant). Medicine (Baltimore). 2020;99:e19428.
23. Schernberg A, Reuze S, Orlhac F, et al. A score combining baseline neutrophilia and primary tumor SUV peak measured from FDG PET is associated with outcome in locally advanced cervical cancer. Eur J Nucl Med Mol Imaging. 2018;45:187–95.
24. Boughdad S, Nioche C, Orlhac F, et al. Influence of age on radiomic features in 18 F-FDG PET in normal breast tissue and in breast cancer tumors. Oncotarget. 2018;20(9):30855–68.
25. Zhou H, Jiang J, Lu J, Wang M, Zhang H, Zuo C. Dual-model radiomic biomarkers predict development of mild cognitive impairment progression to Alzheimer's disease. Front Neurosci. 2019;12:1045.
26. Chan YH. Bioistatistics 104: correlation analysis. Singap Med J. 2003;44:614–9.
27. Sharma C. Diagnostic accuracy of fine needle aspiration cytology of thyroid and evaluation of discordant cases. J Egypt Natl Canc Inst. 2015;27:147–53.

28. Machała E, Sopiński J, Iavorska I, Kołomecki K. Correlation of fine needle aspiration cytology of thyroid gland with histopathological results. Pol Przegl Chir. 2018;21(90):1–5.
29. Sukumaran R, Kattoor J, Pillai KR, Ramadas PT, Nayak N, Somanathan T, et al. Fine needle aspiration cytology of thyroid lesions and its correlation with histopathology in a series of 248 patients. Indian J Surg Oncol. 2014;5:237–41.
30. Cook GJR, Azad G, Owczarczyk K. Challenges and promises of PET radiomic. Int J Radiat Oncol Biol Phys. 2018;15(102):1083–9.