

Received:
11 September 2021

Revised:
28 February 2022

Accepted:
14 March 2022

Published online:
28 March 2022

<https://doi.org/10.1259/bjr.20211050>

Cite this article as:

Nakajo M, Takeda A, Katsuki A, Jinguji M, Ohmura K, Tani A, et al. The efficacy of ^{18}F -FDG-PET-based radiomic and deep-learning features using a machine-learning approach to predict the pathological risk subtypes of thymic epithelial tumors. *Br J Radiol* (2022) 10.1259/bjr.20211050.

FULL PAPER

The efficacy of ^{18}F -FDG-PET-based radiomic and deep-learning features using a machine-learning approach to predict the pathological risk subtypes of thymic epithelial tumors

¹MASATOYO NAKAJO, MD, PhD, ²AYA TAKEDA, MD, ³AKIE KATSUKI, BST, ¹MEGUMI JINGUJI, MD, PhD, ³KAZUYUKI OHMURA, MEng, ¹ATSUSHI TANI, MD, PhD, ²MASAMI SATO, MD, PhD and ¹TAKASHI YOSHIURA, MD, PhD

¹Department of Radiology, Kagoshima University, Graduate School of Medical and Dental Sciences, Kagoshima, Japan

²Department of General Thoracic Surgery, Kagoshima University, Graduate School of Medical and Dental Sciences, Kagoshima, Japan

³Research and Development Department, GE Healthcare Japan, Tokyo, Japan

Address correspondence to: Radiologist Masatoyo Nakajo
E-mail: toyo.nakajo@dolphin.ocn.ne.jp

Objective: To examine whether the machine-learning approach using 18-fluorodeoxyglucose positron emission tomography (^{18}F -FDG-PET)-based radiomic and deep-learning features is useful for predicting the pathological risk subtypes of thymic epithelial tumors (TETs).

Methods: This retrospective study included 79 TET [27 low-risk thymomas (types A, AB and B1), 31 high-risk thymomas (types B2 and B3) and 21 thymic carcinomas] patients who underwent pre-therapeutic ^{18}F -FDG-PET/CT. High-risk TETs (high-risk thymomas and thymic carcinomas) were 52 patients. The 107 PET-based radiomic features, including SUV-related parameters [maximum SUV (SUV_{max}), metabolic tumor volume (MTV), and total lesion glycolysis (TLG)] and 1024 deep-learning features extracted from the convolutional neural network were used to predict the pathological risk subtypes of TETs using six different machine-learning algorithms. The area under the curves (AUCs) were calculated to compare the predictive performances.

Results: SUV-related parameters yielded the following AUCs for predicting thymic carcinomas: SUV_{max} 0.713, MTV 0.442, and TLG 0.479 or high-risk TETs: SUV_{max} 0.673, MTV 0.533, and TLG 0.539. The best-performing algorithm was the logistic regression model for predicting thymic carcinomas (AUC 0.900, accuracy 81.0%), and the random forest (RF) model for high-risk TETs (AUC 0.744, accuracy 72.2%). The AUC was significantly higher in the logistic regression model than three SUV-related parameters for predicting thymic carcinomas, and in the RF model than MTV and TLG for predicting high-risk TETs (each; $p < 0.05$).

Conclusion: ^{18}F -FDG-PET-based radiomic analysis using a machine-learning approach may be useful for predicting the pathological risk subtypes of TETs.

Advances in knowledge: Machine-learning approach using ^{18}F -FDG-PET-based radiomic features has the potential to predict the pathological risk subtypes of TETs.

INTRODUCTION

Thymic epithelial tumors (TETs) are the most common mediastinal tumors in adults, and thymomas and thymic carcinomas are the most common histologic subtypes.¹ The incidence of TETs is 0.15–0.32 cases per million.² Their prognoses mainly depend on their Masaoka staging and World Health Organization (WHO) histological classification as well as resectability.^{1–5} In the WHO classification,⁶ TETs are classified into the following categories: five types of thymomas (type A, AB, B1, B2 and B3) and thymic carcinomas. It has been reported that the overall survival rates are higher in patients with type A, AB, or B1 tumors

(low-risk thymomas) than those with type B2 or B3 tumors (high-risk thymomas),⁷ and thymic carcinomas have a poor prognosis, with a 5 year survival rate of 30–50%.^{3,8} Therefore, a non-invasive diagnostic method that can predict the pathological risk subtypes (low-risk thymomas, high-risk thymomas and thymic carcinoma) may be useful for pre-treatment risk stratification and prognostication in TETs patients.^{1,5}

Glucose analog 2-deoxy-2- ^{18}F -fluorodeoxyglucose (^{18}F -FDG) uptake represents the glucose metabolic activity and is widely used as a tracer of positron emission tomography

(PET) in the field of oncology.⁸ Several studies have investigated the ability of conventional SUV-related parameters [*e.g.* maximum SUV (SUV_{max}), metabolic tumor volume (MTV), and total lesion glycolysis (TLG)] to predict the malignant nature of TETs.^{9–14} Treglia *et al*¹⁵ performed a meta-analysis of SUV_{max} and demonstrated significant differences in the SUV_{max} between low- and high-risk thymomas and between low-risk (type A, AB, or B1) or high-risk (type B2 or B3) thymomas and thymic carcinomas, although a SUV_{max} cut-off for discriminating between the groups could not be defined because of the large overlap of SUV_{max} among the different TETs. There has been controversy regarding the ability of MTV and TLG to predict TET grades; one report showed that MTV was significantly higher in type B3 thymomas than in other types of thymomas,¹⁶ while the other reported that MTV and TLG were not related to the tumor grade.¹⁷

Radiomics refers to various mathematical methods that extract a large number of quantitative features that describe imaging phenotypes (*e.g.* pixel intensity, shape, and texture) and provide useful biologic information.¹⁸ A few studies have examined the ¹⁸F-FDG PET-based texture features for predicting the malignant nature of TETs: Lee *et al*¹⁹ reported that some of the GLSZM indices were independent of SUV_{max} when discriminating between the TET grades. Nakajo *et al*²⁰ investigated the SUV-related parameters and six texture parameters (entropy, homogeneity, dissimilarity, intensity variability, size-zone variability, and zone percentage) individually and in combination for discriminating between TET grades and reported that although the diagnostic performances of individual SUV_{max} and texture parameters were relatively low, a combination of these parameters can increase the diagnostic performance when differentiating low-risk thymomas from high-risk TETs. However, their predictive performances and representative features for outcome prediction were inconsistent.

Machine learning relies on computer algorithms to learn and identify complex interactions among all variables by minimizing the error between the predicted and observed outcomes.²¹ Compared to conventional statistical methods, machine learning can detect interactions among variables at a deep level and can learn from the data and update algorithms.²²

Recently, researchers in the field of nuclear medicine have proposed classification methods based on a machine-learning approach or a deep-learning model.^{23–26} However, to our knowledge, no study has investigated the usefulness of a machine-learning approach or a deep-learning model for assessing ¹⁸F-FDG-PET-based images to predict the pathological risk subtypes of TETs.

The present study was performed to examine whether the machine-learning approach using ¹⁸F-FDG PET-based radiomic and deep-learning features is useful for predicting the pathological risk subtypes of TETs.

METHODS AND MATERIALS

Patients

The institutional review board approved this retrospective study and waived the requirement for written informed patient consent. Pre-treatment ¹⁸F-FDG-PET/CT was

performed for 86 consecutive patients with suspected or known TETs between January 2011 and January 2018, and their clinical records were reviewed to identify patients who were eligible for analysis.

In the previous study,²⁰ the diagnostic performance of ¹⁸F-FDG-PET/CT for discriminating between TET grades was examined using the SUV-related parameters and six texture parameters (entropy, homogeneity, dissimilarity, intensity variability, size-zone variability and zone percentage) in 34 TETs patients who were enrolled between January 2011 and June 2016. However, analyses by the machine-learning approach for predicting the risk subtypes of TETs using other ¹⁸F-FDG PET-based radiomic or deep-learning features were not performed. Thus, these 34 patients were included in the total 86 patients. Patients were included in the current study if they met the following inclusion and exclusion criteria; the inclusion criteria were: (1) pathologically proven TETs; (2) no pre-operative history of radiotherapy, chemotherapy, or chemoradiotherapy; and (3) primary tumor with visible ¹⁸F-FDG uptake on the PET/CT reports. The exclusion criterion was incomplete clinical or follow-up data.

Of the 86 patients, 5 patients (3 type AB, 1 type B2 and 1 type B3 thymomas) were excluded because of lack of focal ¹⁸F-FDG uptake in the TETs. Two patients (one type AB and one type B1 thymomas) were excluded because the volumes of interest (VOIs) could not be created due to low ¹⁸F-FDG uptake (SUV_{max} : 1.5 and 1.4) with small tumor size (1.4 cm and 1.8 cm). Finally, 79 patients [41 men and 38 women; mean (\pm standard deviation) age, 62 ± 15 y; range, 22–87 y] were eligible for the analyses (69 TETs were diagnosed by surgical excision, 10 by percutaneous biopsy).

The previous study was only analyzed for the ¹⁸F-FDG-avid TETs with MTV of >10.0 cm³ and $SUV \geq 2.5$, because texture analysis has been reported to be influenced by tumor volume.²⁷ Thus, we also performed the same analyses for the ¹⁸F-FDG-avid TETs with MTV of >10.0 cm³ and $SUV \geq 2.5$ as the additional supplemental analyses. When we applied the above criteria, 59 patients [31 men and 28 women; mean (\pm standard deviation) age, 63 ± 15 y; range, 22–87 y] were eligible for the analyses. Thus, in this group, 20 patients were excluded from the original 79 patients.

The data collection was completed on December 31, 2020.

Imaging protocols

The patients were instructed to fast for at least 5 h before the examinations, and the scans were performed using a Discovery 600M PET/CT (GE Medical Systems, Milwaukee, WI). The mean plasma glucose level was 107 mg dl⁻¹ (range, 82–167 mg dl⁻¹) immediately before administering intravenous injection of ¹⁸F-FDG (FDG Scan; Nihon Medi-Physics, Tokyo, Japan). The emission scan started 1 h after ¹⁸F-FDG [206 MBq \pm 31 (range, 146–278 MBq)] injection following CT data acquisition (slice thickness, 3.75 mm; pitch, 1.75 mm; 120 keV; auto mA [40–100 mA depending on patient body mass]). Acquisition time was

2.5 min per bed position, with eight bed positions. The CT attenuation-corrected acquired data were reconstructed with a three-dimensional ordered subset expectation–maximization algorithm (image matrix size, 192×192 ; 1 PET voxel size, $3.125 \times 3.125 \times 3.27 \text{ mm}^3$; 16 subsets, 2 iterations; VUE Point Plus).

Image and radiomic feature analyses

Two radiologists, one with 11 years and another with 18 years of ^{18}F -FDG-PET/CT experience, who were aware of the study purpose but were blinded to the clinical and pathological information confirmed, in consensus, whether the primary lesion had abnormal FDG uptake (greater than background activity in the surrounding tissue).²⁰ A third radiologist who had 16 years of ^{18}F -FDG-PET/CT experience performed quantitative analyses of the visible primary lesions. He generated the VOI by manually placing a region of interest on a suitable reference fused axial image and defined the craniocaudal and mediolateral extent encompassing the entire visible primary lesion, excluding any adjacent physiological ^{18}F -FDG-avid structure.

A 40% threshold of SUV_{max} was used to define the VOI boundaries.²⁸ The Pyradiomic software (v. 2.2.0)²⁹ was used to extract 107 PET-radiomic features, including the shape and first-order features, gray-level co-occurrence matrix, gray-level dependence matrix, gray-level run-length matrix (GLRLM), gray-level size-zone matrix (GLSZM), and neighborhood gray-level different matrix on the PET images (Supplementary Table 1).

Histological analyses

All the clinical records, including the pathological reports, were reviewed and the TETs were classified as per the 2004 WHO histological classification [thymoma types A, AB, B1, B2, B3 and thymic carcinoma (C)].⁶ Types A, AB and B1 are less aggressive and exhibit better prognoses than B2, B3 and thymic carcinomas.^{3–8} Thus, all TETs were grouped into low-risk thymomas

(type A, AB and B1), high-risk thymomas (type B2 and B3) and thymic carcinomas.⁵ Moreover, high-risk thymomas and thymic carcinomas were defined as high-risk TETs.³

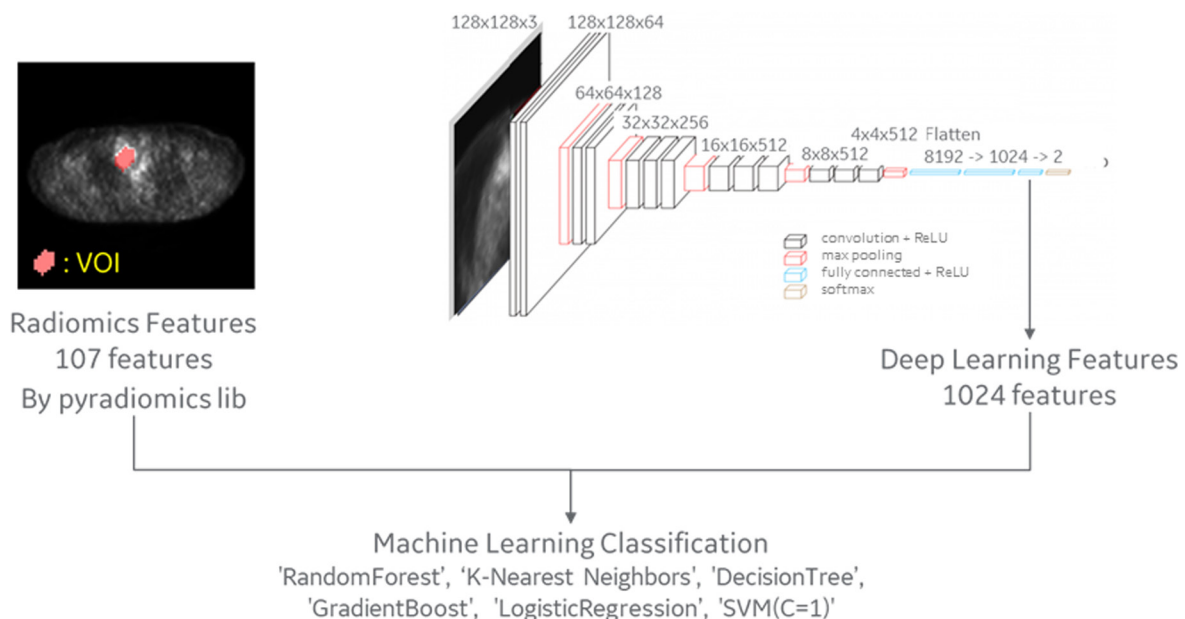
Machine-learning approach

The 107 PET-based radiomic features and 1024 deep-learning features were used to predict the risk subtypes of TETs, using machine-learning approaches (Figure 1). Deep-learning features were extracted at the second layer from the end. The model was trained as transfer learning on VGG16 convolutional neural network (CNN). As CNN input data, the volume data were cropped out with an $180 \times 180 \times 90$ -mm boundary box around the VOI, resized to $128 \times 128 \times 32$ pixels, and picked the center and ± 1 slices in the Z direction up. $\text{SUV}_{\text{max}} = 10$ and $\text{min} = 0$ normalization was applied. Training was performed with data augmentation (zoom, shift, rotate, flip) and 50 epochs. The model was saved at the best during the training.

The following six machine-learning algorithms were evaluated for binary classification: random forest (RF), k-nearest neighbors, logistic regression, decision tree, gradient boost, and a support vector machine (SVM). To overcome the limitation caused by imbalanced data, bagging was used.³⁰ To minimize the negative influence of overfitting, five-fold cross-validation was performed in this study^{31,32} and repeated 10 times. The fivefold cross-validation randomly split the data set into five subsets. For each repeated time, four subsets were used as the training groups, and the remaining subset was used as the testing data.

Each machine-learning algorithm calculated the probability score (range, 0–1) of thymic carcinoma for each tumor, and the score was averaged for all the repeated assessments. As per the averaged probability score, each tumor was classified as thymic carcinoma (probability score ≥ 0.50) or thymoma

Figure 1. Diagram of the machine-learning approach.



(probability score <0.50), and the areas under the receiver operating characteristic curve (AUC), sensitivity, specificity, and accuracy for predicting thymic carcinoma were calculated. Moreover, each machine-learning algorithm calculated the probability score for high-risk TET, and each tumor was classified as high-risk TET (probability score ≥ 0.50) or low-risk thymomas (probability score <0.50); further, the AUC, sensitivity, specificity, and accuracy for predicting high-risk TETs were obtained. Additionally, among the patients with thymomas, the same analyses were also performed for predicting high-risk thymomas. Receiver operating characteristic curve analysis was performed to compare the predictive performances of the models.

In order to identify the most important features, Spearman's rank correlation coefficient and recursive feature elimination (RFE) were used to select the most informative and valuable radiomic features.³³ First, the correlation coefficient of each radiomic feature was selected using Spearman's rank correlation, and a strong correlation ($\rho \geq 0.95$) was selected. Then, the importance of each feature using RFE with linear SVM was calculated, and the features were reduced to the required number of features recursively.

The machine-learning approach was used with scikit-learn (v. 0.21.3), a library of python package.³⁴

Statistical analyses

The Mann-Whitney *U* test was used to appropriately assess the difference between two quantitative variables. Receiver operating characteristic curve analysis was performed to examine the diagnostic performance of conventional SUV-related parameters (SUVmax, MTV and TLG) or each machine-learning algorithm for predicting thymic carcinoma, high-risk TET or high-risk thymomas. The DeLong method was used to analyze the statistical significance of the differences between the AUCs.³⁵

As supplemental analyses, the above machine-learning and statistical analyses were also performed for 59 patients who were met the inclusion and exclusion criteria of the previous study.

Data were presented as medians and interquartile ranges (IQRs). All the *p* values were two-sided, and a *p*-value <0.05 was considered to indicate statistical significance. The statistical analyses were performed using MedCalc Statistical Software (MedCalc Software, Mariakerke, Belgium) and R software (v. 4.0.3).

RESULTS

Patient characteristics

Among 79 patients with TET, there were 58 (7 type A, 10 type AB, 10 type B1, 10 type B2 and 21 type B3) thymomas and 21 thymic carcinomas. Thus, there were 27 low-risk thymomas and 31 high-risk thymomas. Among these 79 TETs, 52 TETs (31 high-risk thymomas and 21 thymic carcinomas) were high-risk TETs.

On the additional supplemental analyses of 59 TETs with MTV > 10.0 cm³ and SUV ≥ 2.5 , there were 43 (5 type A, 7 type AB, 7 type B1, 6 type B2, and 18 type B3) thymomas and 16 thymic carcinomas. Thus, there were 19 low-risk thymomas and 24 high-risk thymomas. Among these 59 TETs, 40 TETs (24 high-risk thymomas and 16 thymic carcinomas) were high-risk TETs.

Conventional SUV-related parameters and machine-learning methods for predicting thymic carcinomas

Thymic carcinomas showed significantly higher SUVmax than thymomas ($p = 0.004$); however, there was no significant difference in the MTV and TLG between thymic carcinomas and thymomas (MTV: $p = 0.43$; TLG: $p = 0.78$) (Table 1). On the additional supplemental analyses of 59 TETs with MTV > 10.0 cm³ and SUV ≥ 2.5 , thymic carcinomas also showed significantly higher SUVmax than thymomas ($p = 0.006$); neither MTV nor TLG was significantly different between these two groups (MTV: $p = 0.36$; TLG: $p = 0.89$) (Supplementary Table 2).

Table 2 shows the following AUCs of three conventional SUV-related parameters to predict thymic carcinomas: SUVmax, 0.713 ($p = 0.002$); MTV, 0.442 ($p = 0.44$); and TLG, 0.479 ($p = 0.79$). The parameters yielded sensitivity of 42.9% (TLG) to 71.4% (SUVmax), specificity of 31.0% (MTV) to 69.0% (SUVmax), and accuracy of 36.7% (MTV) to 69.6% (SUVmax) for predicting thymic carcinomas. The additional supplemental analyses of 59 TETs with MTV > 10.0 cm³ and SUV ≥ 2.5 (Supplementary Table 3) showed the following AUCs of three conventional SUV-related parameters to predict thymic carcinomas: SUVmax, 0.733 ($p = 0.002$); MTV, 0.578 ($p = 0.39$); and TLG, 0.512 ($p = 0.90$).

The overall classification performance of six machine-learning methods for predicting thymic carcinomas was compared using AUCs (Table 3). The logistic regression model was the best-performing classifier for predicting thymic carcinomas (AUC = 0.900, sensitivity 81.0%, specificity 81.0%, accuracy 81.0%), and the three most important features for predicting thymic carcinomas were skewness, sphericity, and GLSZM-gray level non-uniformity. The AUC and diagnostic accuracy of these three features were 0.580–0.828 and 40.5–78.5%, respectively (Table 4). The features yielded sensitivity of 76.2% (sphericity) to 95.2% (GLSZM-gray level non-uniformity) and specificity of 20.7% (GLSZM-gray level non-uniformity) to 79.3% (sphericity) for predicting thymic carcinomas.

No significant difference was observed in the AUC between the logistic regression model and skewness ($p = 0.16$). However, the AUC of the logistic regression model was significantly higher than those of the other two important features (sphericity, $p = 0.023$ and GLSZM-gray level non-uniformity, $p < 0.001$) or three conventional SUV-related parameters (SUVmax, $p = 0.007$; MTV and TLG, each, $p < 0.001$).

Table 1. Comparison of the SUV-related parameters between thymomas and thymic carcinomas, between low-risk thymomas and high-risk TETs or between low-risk thymomas and high-risk thymomas

Feature	Thymoma (n = 58)			Thymic carcinoma (n = 21)			p-value
	Median	IQR	Range	Median	IQR	Range	
SUVmax	4.2	3.4–5.7	1.7–14.2	6.3	4.7–9.3	2.0–46.9	0.004
MTV (cm ³)	46.5	12.0–96.3	0.6–350.9	19.7	10.6–73.9	1.0–724.0	0.43
TLG	108.3	25.2–244.9	0.9–1061.8	44.0	17.4–294.2	0.7–250.1	0.78
	Low-risk thymoma (n = 27)			High-risk TET (n = 52)			
	Median	IQR	Range	Median	IQR	Range	p-value
SUVmax	4.0	3.4–4.4	2.2–14.2	5.3	3.8–7.6	1.7–46.9	0.012
MTV (cm ³)	33.2	12.2–75.1	1.8–209.2	46.0	14.4–68.7	0.6–724.0	0.64
TLG	68.0	25.5–172.9	3.1–1061.8	120.1	19.2–266.5	0.7–3092.2	0.58
	Low-risk thymoma (n = 27)			High-risk thymoma (n = 31)			
	Median	IQR	Range	Median	IQR	Range	p-value
SUVmax	4.0	3.4–4.4	2.2–14.2	4.8	3.5–6.4	1.7–10.6	0.16
MTV (cm ³)	33.2	12.2–75.1	1.8–209.2	55.8	10.9–109.1	0.6–350.9	0.32
TLG	68.0	25.5–172.9	3.1–1061.8	140.4	20.5–261.8	0.9–720.2	0.40

IQR, interquartile range; MTV, metabolic tumor volume; TET, thymic epithelial tumor; TLG, total lesion glycolysis.

On the additional supplemental analyses of 59 TETs with MTV > 10.0 cm³ and SUV ≥ 2.5, the RF model was the best-performing classifier for predicting thymic carcinomas (AUC = 0.935, sensitivity 87.5%, specificity 88.4%, accuracy 88.1%), and followed by the logistic regression model (AUC = 0.908, sensitivity 75.0%, specificity 86.1%, accuracy 83.1%) (Supplementary Table 4). The three most important features for predicting thymic carcinomas were skewness, maximum 2D diameter row, and GLSZM-gray level non-uniformity. The AUC and diagnostic accuracy of these three most important features were 0.563–0.788 and 39.0–76.3%, respectively. On the other hand, the AUC of the RF model was significantly higher than those of three important features (Skewness, $p = 0.028$, Maximum 2D diameter row, $p < 0.001$, and GLSZM-gray level non-uniformity, $p < 0.001$) or three conventional SUV-related parameters (SUVmax, $p = 0.008$; MTV and TLG, each, $p < 0.001$).

Conventional SUV-related parameters and machine-learning methods for predicting high-risk TETs

High-risk TETs showed significantly higher SUVmax than low-risk thymomas ($p = 0.012$); neither MTV nor TLG was significantly different between these two groups (MTV: $p = 0.64$; TLG: $p = 0.58$) (Table 1). On the additional supplemental analyses of 59 TETs with MTV > 10.0 cm³ and SUV ≥ 2.5, high-risk TETs also showed significantly higher SUVmax than low-risk thymomas ($p = 0.013$); neither MTV nor TLG was significantly different between these two groups (MTV: $p = 0.52$; TLG: $p = 0.46$) (Supplementary Table 2).

Three conventional SUV-related parameters yielded the following AUCs for the ability to predict high-risk TETs: SUVmax, 0.673 ($p = 0.006$); MTV, 0.533 ($p = 0.62$); and TLG, 0.539 ($p = 0.57$) (Table 2). The parameters yielded sensitivity

from 32.7% (MTV) to 57.7% (SUVmax), specificity from 74.1% (TLG) to 85.2% (SUVmax), and accuracy from 50.6% (MTV) to 67.1% (SUVmax) for predicting high-risk TETs. The additional supplemental analyses of 59 TETs with MTV > 10.0 cm³ and SUV ≥ 2.5 (Supplementary Table 3) showed the following AUCs of three conventional SUV-related parameters to predict high-risk TETs: SUVmax, 0.702 ($p = 0.007$); MTV, 0.553 ($p = 0.50$); and TLG, 0.561 ($p = 0.44$).

The overall classification performance of six machine-learning methods for predicting high-risk TETs was compared using AUCs (Table 3). The RF model was the best-performing classifier for predicting high-risk TETs (AUC = 0.744, sensitivity 71.2%, specificity 74.1%, accuracy 72.2%). The three most important features for the prediction of high-risk TETs were skewness, kurtosis, and GLSZM-small area emphasis, their AUC and diagnostic accuracies were 0.506–0.741 and 54.4–75.9%, and yielded sensitivity of 44.2% (kurtosis) to 90.4% (skewness) and specificity of 48.1% (skewness) to 77.8% (kurtosis), respectively (Table 4).

No significant difference was observed in the AUC between the RF model and skewness ($p = 0.97$), kurtosis ($p = 0.12$), and SUVmax ($p = 0.32$). The AUC of the above RF model was significantly higher than those of the other one important feature (GLSZM-small area emphasis, $p = 0.006$) and two other SUV-related parameters (MTV, $p = 0.022$ and TLG, $p = 0.021$).

On the additional supplemental analyses of 59 TETs with MTV > 10.0 cm³ and SUV ≥ 2.5, the SVM model was the best-performing classifier for predicting high-risk TETs (AUC = 0.844, sensitivity 80.0%, specificity 78.9%, accuracy 79.7%), followed by the logistic regression model (AUC = 0.838, sensitivity 75.0%,

Table 2. SUV-related parameters for predicting thymic carcinomas, high-risk TETs or high-risk thymomas

Feature	Thymic carcinoma				
	Sensitivity (%)	Specificity (%)	Accuracy (%)	AUC	p-value
SUVmax	71.4 (15/21)	69.0 (40/58)	69.6 (55/79)	0.713	0.002
	47.8-88.7 ^a	55.5-80.5 ^a	58.2-79.5 ^a	0.600-0.809 ^a	
MTV (cm ³)	52.4 (11/21)	31.0 (18/58)	36.7 (29/79)	0.442	0.44
	29.8-74.3 ^a	19.5-44.5 ^a	26.1-48.3 ^a	0.301-0.583 ^a	
TLG	42.9 (9/21)	43.1 (25/58)	43.0 (34/79)	0.479	0.79
	21.8-66.0 ^a	30.2-56.8 ^a	31.9-54.7 ^a	0.356-0.623 ^a	
Feature	High-risk TET				
	Sensitivity (%)	Specificity (%)	Accuracy (%)	AUC	p-value
SUVmax	57.7 (30/52)	85.2 (23/27)	67.1 (53/79)	0.673	0.006
95% CI	43.2-71.3 ^a	66.2-95.8 ^a	55.6-77.3 ^a	0.558-0.774 ^a	
MTV (cm ³)	32.7 (17/52)	85.2 (23/27)	50.6 (40/79)	0.533	0.62
	20.3-47.1 ^a	66.2-95.8 ^a	39.1-62.1 ^a	0.417-0.646 ^a	
TLG	44.2 (23/52)	74.1 (20/27)	54.4 (43/79)	0.539	0.57
	30.5-58.7 ^a	53.7-88.9 ^a	42.8-65.7 ^a	0.422-0.651 ^a	
Feature	High-risk thymoma				
	Sensitivity (%)	Specificity (%)	Accuracy (%)	AUC	p-value
SUVmax	48.4 (15/31)	85.2 (23/27)	65.5 (38/58)	0.609	0.15
	30.2-66.9 ^a	66.3-95.8 ^a	51.9-77.5 ^a	0.472-0.734 ^a	
MTV (cm ³)	38.7 (12/31)	85.2 (23/27)	60.3 (35/58)	0.576	0.32
	21.8-57.8 ^a	66.3-95.8 ^a	46.6-73.0 ^a	0.439-0.705 ^a	
TLG	58.1 (18/31)	66.7 (18/27)	62.1 (36/58)	0.565	0.40
	39.1-75.5 ^a	46.0-83.5 ^a	48.4-74.5 ^a	0.428-0.695 ^a	

AUC, area under the receiver operating characteristic curve; MTV, metabolic tumor volume; TET, thymic epithelial tumor; TLG, total lesion glycolysis. ^a95% confidence interval.

specificity 89.5%, accuracy 79.7%) (Supplementary Table 4). The three most important features of predicting high-risk TETs were skewness, kurtosis, and GLRLM-short run emphasis and their AUC and diagnostic accuracy were 0.599–0.750 and 59.3–67.8%, respectively (Supplementary Table 5). No significant difference was observed in the AUC between the SVM model and skewness ($p = 0.24$), and SUVmax ($p = 0.069$). The AUC of the above SVM model was significantly higher than those of two other important features (kurtosis, $p = 0.027$, GLRLM-short run emphasis, $p = 0.004$) and two other SUV-related parameters (MTV, $p = 0.002$ and TLG, $p = 0.002$).

The representative ¹⁸F-FDG-PET/CT images of thymoma and thymic carcinoma are shown in Figures 2 and 3, respectively.

Conventional SUV-related parameters and machine-learning methods for predicting high-risk thymomas

There were no significant differences in three conventional SUV-related parameters between 31 high-risk thymomas and 27 low-risk thymomas (each, $p > 0.05$) (Table 1). On the additional

supplemental analyses of 43 thymomas with MTV > 10.0 cm³ and SUV ≥ 2.5, there were also no significant differences in three conventional SUV-related parameters between 24 high-risk thymomas and 19 low-risk thymomas (each, $p > 0.05$) (Supplementary Table 2).

Three conventional SUV-related parameters yielded the following AUCs for the ability to predict high-risk thymomas: SUVmax, 0.609 ($p = 0.15$); MTV, 0.576 ($p = 0.32$); and TLG, 0.565 ($p = 0.40$) and yielded sensitivity from 38.7% (MTV) to 58.1% (TLG), specificity from 66.7% (TLG) to 85.2% (SUVmax, MTV), and accuracy from 60.3% (MTV) to 65.5% (SUVmax), respectively (Table 2). The additional supplemental analyses of 43 thymomas with MTV > 10.0 cm³ and SUV ≥ 2.5 (Supplementary Table 3) showed the following AUCs of three conventional SUV-related parameters to predict high-risk thymomas: SUVmax, 0.639 ($p = 0.12$); MTV, 0.610 ($p = 0.22$); and TLG, 0.594 ($p = 0.30$).

The decision tree model was the best-performing classifier for predicting high-risk thymomas (AUC = 0.680, sensitivity 71.0%, specificity 63.0%, accuracy 67.2%) (Table 3). The three most

Table 3. Classification performances of machine-learning models for the prediction of thymic carcinomas, high-risk TEIs or high-risk thymomas

Feature	Thymic carcinoma					High-risk TEI					High-risk thymoma				
	Sensitivity(%)	Specificity(%)	Accuracy(%)	AUC		Sensitivity(%)	Specificity(%)	Accuracy(%)	AUC		Sensitivity(%)	Specificity(%)	Accuracy(%)	AUC	
Random forest	85.7 (18/21)	81.0 (47/58)	82.3 (65/79)	0.885		71.2 (37/52)	74.1 (20/27)	72.2 (57/79)	0.744		77.4 (24/31)	44.4 (12/27)	62.1 (36/58)	0.600	
	63.7-97.0 ^d	68.6-90.1 ^d	72.1-90.0 ^d	0.793-0.946 ^d		56.9-82.9 ^d	53.7-88.9 ^d	60.9-81.7 ^d	0.633-0.835 ^d		58.9-90.4 ^d	25.5-64.7 ^d	48.4-74.5 ^d	0.450-0.750 ^d	
k-nearest neighbors	71.4 (15/21)	75.9 (44/58)	74.7 (59/79)	0.691		50.0 (26/52)	77.8 (21/27)	59.5 (47/79)	0.565		54.8 (17/31)	59.3 (16/27)	56.9 (33/58)	0.632	
	47.8-88.7 ^d	62.8-86.1 ^d	63.6-83.8 ^d	0.576-0.790 ^d		35.8-64.2 ^d	57.7-91.4 ^d	47.9-70.4 ^d	0.449-0.676 ^d		36.0-72.7 ^d	38.8-77.6 ^d	43.2-69.8 ^d	0.486-0.778 ^d	
Logistic regression	81.0 (17/21)	81.0 (47/58)	81.0 (64/79)	0.900		63.5 (33/52)	74.1 (20/27)	67.1 (53/79)	0.696		64.5 (20/31)	51.9 (14/27)	58.6 (34/58)	0.582	
	58.1-94.6 ^d	68.6-90.1 ^d	70.6-89.0 ^d	0.812-0.956 ^d		49.0-76.4 ^d	53.7-88.9 ^d	55.6-77.3 ^d	0.582-0.794 ^d		45.4-80.8 ^d	31.9-71.3 ^d	44.9-71.4 ^d	0.431-0.734 ^d	
Decision tree	76.2 (16/21)	77.6 (45/58)	77.2 (61/79)	0.716		51.9 (27/52)	92.6 (25/27)	65.8 (52/79)	0.598		71.0 (22/31)	63.0 (17/27)	67.2 (39/58)	0.680	
	52.8-91.8 ^d	64.7-87.5 ^d	66.4-85.9 ^d	0.603-0.811 ^d		37.6-66.0 ^d	75.7-99.1 ^d	54.3-76.1 ^d	0.482-0.707 ^d		52.0-85.8 ^d	42.4-80.6 ^d	53.7-79.0 ^d	0.541-0.819 ^d	
Gradient boost	47.6 (10/21)	81.0 (47/58)	72.2 (57/79)	0.805		55.8 (29/52)	81.5 (22/27)	64.6 (51/79)	0.627		71.0 (22/31)	51.9 (14/27)	72.4 (42/58)	0.679	
	25.7-70.2 ^d	68.6-90.1 ^d	60.9-81.7 ^d	0.700-0.885 ^d		41.3-69.5 ^d	61.9-93.7 ^d	53.0-75.0 ^d	0.511-0.733 ^d		52.0-85.8 ^d	31.9-71.3 ^d	59.1-83.3 ^d	0.540-0.818 ^d	
SVM	71.4 (15/21)	75.9 (44/58)	74.7 (59/79)	0.851		55.8 (29/52)	74.1 (20/27)	62.0 (49/79)	0.719		80.6 (25/31)	14.8 (4/27)	50.0 (29/58)	0.566	
	47.8-88.7 ^d	62.8-86.1 ^d	63.6-83.8 ^d	0.754-0.921 ^d		41.3-69.5 ^{aa}	53.7-88.9 ^{aa}	50.4-72.7 ^d	0.606-0.814 ^d		62.5-92.5 ^d	4.2-33.7 ^d	36.6-63.4 ^d	0.415-0.717 ^d	

AUC, area under the receiver operating characteristic curve; SVM, support vector machine; TEI, thymic epithelial tumor.
^d95% confidence interval.

Table 4. Three most important features for predicting thymic carcinomas, high-risk TETs or high-risk thymomas

Thymic carcinoma				
Top three features	Sensitivity (%)	Specificity (%)	Accuracy (%)	AUC
Skewness	81.0 (17/21)	75.9 (44/58)	77.2 (61/79)	0.828
	58.1-94.6 ^a	62.8-86.3 ^a	66.4-85.9 ^a	0.727-0.904 ^a
Sphericity	76.2 (16/21)	79.3 (46/58)	78.5 (62/79)	0.739
	52.8-91.8 ^a	66.6-88.8 ^a	67.8-86.9 ^a	0.628-0.831 ^a
GLSZM-gray level non-uniformity	95.2 (20/21)	20.7 (12/58)	40.5 (32/79)	0.580
	76.2-99.9 ^a	11.1-33.6 ^a	29.6-52.1 ^a	0.464-0.691 ^a
High-risk TET				
Top three features	Sensitivity (%)	Specificity (%)	Accuracy (%)	AUC
Skewness	90.4 (47/52)	48.1 (13/27)	75.9 (60/79)	0.741
	79.0-96.8 ^a	28.7-68.1 ^a	65.0-84.9 ^a	0.631-0.833 ^a
Kurtosis	44.2 (23/52)	77.8 (21/27)	55.7 (44/79)	0.596
	30.5-58.7 ^a	57.7-91.4 ^a	44.1-66.9 ^a	0.480-0.705 ^a
GLSZM-small area emphasis	51.9 (27/52)	59.3 (16/27)	54.4 (43/79)	0.506
	37.6-66.0 ^a	38.8-77.6 ^a	42.8-65.7 ^a	0.391-0.620 ^a
High-risk thymoma				
Top three features	Sensitivity (%)	Specificity (%)	Accuracy (%)	AUC
Skewness	83.9 (26/31)	48.1 (13/27)	67.2 (39/58)	0.649
	66.3-94.5 ^a	28.7-68.1 ^a	53.7-79.0 ^a	0.512-0.770 ^a
Surface volume ratio	35.5 (11/31)	81.5 (22/27)	56.9 (33/58)	0.520
	19.2-54.6 ^a	61.9-93.7 ^a	43.2-69.8 ^a	0.385-0.653 ^a
Kurtosis	48.4 (15/31)	77.8 (21/27)	62.1 (36/58)	0.616
	30.2-66.9 ^a	57.7-91.4 ^a	48.4-74.5 ^a	0.479-0.741 ^a

AUC, area under the receiver operating characteristic curve; GLSZM, gray-level size-zone matrix; TET, thymic epithelial tumor.
^a95% confidence interval.

important features for predicting high-risk thymomas were skewness, surface volume ratio and kurtosis, and the AUC, sensitivity, specificity, and accuracy of skewness for predicting high-risk thymomas were 0.649, 83.9%, 48.1% and 67.2%, respectively (Table 4). No significant difference was observed in the AUC between the decision tree model and skewness ($p = 0.76$), surface volume ratio ($p = 0.10$), kurtosis ($p = 0.52$) and three conventional SUV-related parameters (SUVmax, $p = 0.47$; MTV, $p = 0.28$; and TLG, $p = 0.24$).

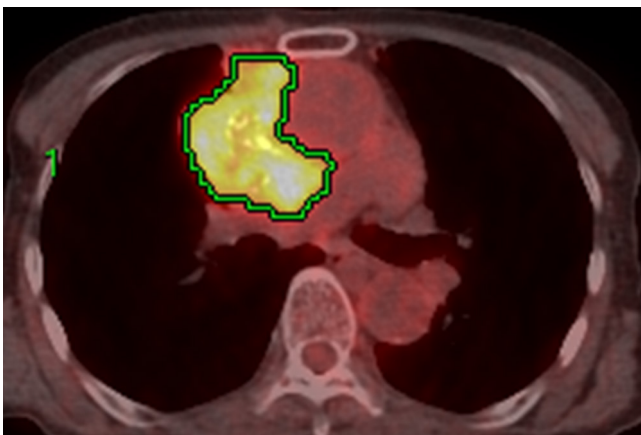
On the additional supplemental analyses of 43 thymomas with MTV > 10.0 cm³ and SUV ≥ 2.5, the SVM model was the best-performing classifier for predicting high-risk thymomas (AUC = 0.670, sensitivity 83.3%, specificity 36.8%, accuracy 62.8%) followed by the logistic regression model (AUC = 0.666, sensitivity 75.0%, specificity 52.6%, accuracy 65.1%) (Supplementary Table 4). The three most important features for predicting high-risk thymomas were skewness, kurtosis, and SUV_{Interquartile range} and their AUC and diagnostic accuracy were 0.524–0.675 and 65.1%, respectively (Supplementary Table 5). No significant difference was observed in the AUC between the SVM model and skewness ($p = 0.96$),

kurtosis ($p = 0.89$), SUV_{Interquartile range} ($p = 0.11$) or three conventional SUV-related parameters (SUVmax, $p = 0.77$; MTV, $p = 0.61$; and TLG, $p = 0.52$).

DISCUSSION

In our study, SUVmax was significantly higher in patients with thymic carcinomas than in those with thymomas or in those with high-risk TETs than in those with low-risk thymomas, while neither MTV nor TLG was significantly different between these respective groups. Moreover, no significant differences were observed in these conventional SUV-related parameters between high- and low-risk thymomas. These conventional SUV-related parameters yielded an accuracy of 36.7% (MTV) to 69.6% (SUVmax) for predicting thymic carcinomas, from 50.6% (MTV) to 67.1% (SUVmax) for predicting high-risk TETs, and from 60.3% (MTV) to 65.5% (SUVmax) for predicting high-risk thymomas. On the additional supplemental analyses of TETs with MTV > 10.0 cm³ and SUV ≥ 2.5, the significant differences in conventional SUV-related parameters among the risk subtypes of TETs were almost the same as those of the whole population analyses (Supplementary Tables 2 and 3).

Figure 2. Image of a 75-year-old female with type B3 thymoma. The transaxial ^{18}F -FDG-PET/CT image shows ^{18}F -FDG uptake in the primary lesion. The green lines represent the borders of the VOI. The SUVmax (cut-off, 5.1), MTV (cut-off, 14.8), and TLG (cut-off, 63.8) were 7.8, 210.8 and 720.2, respectively. All the SUV-related parameters were positive for thymic carcinoma as per each threshold criterion. The calculated probability score as per the logistic regression model for the diagnosis of thymic carcinoma was 0.178. As per this probability score, this tumor was classified as a thymoma. Moreover, the calculated probability score based on the RF model for the diagnosis of high-risk TET was 0.60. As per this probability score, this tumor was classified as a high-risk TET. ^{18}F -FDG-PET, 18-fluodeoxyglucose positron emission tomography; MTV, metabolic tumor volume; RF, random forest; SUVmax, maximum standardized uptake value; TET, thymic epithelial tumor; TLG, total lesion glycolysis; VOI, volume of interest.



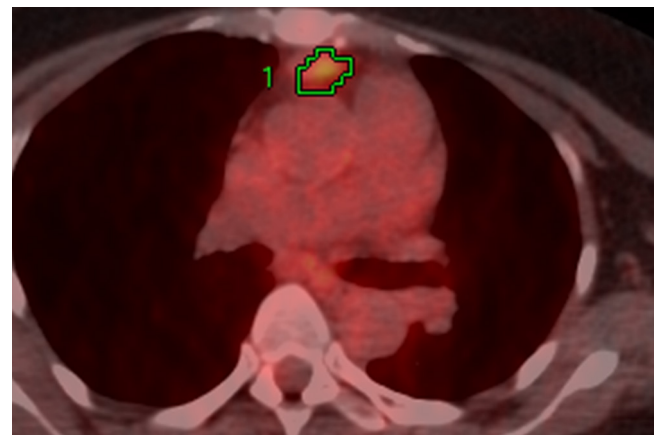
As mentioned in the introduction, SUV-related parameters have been reported to be difficult to correctly predict the TET risk subtypes, and our current findings also suggest that it might be difficult to correctly predict the TET risk subtypes using conventional SUV-related parameters due to the large overlap of these quantitative values among the different TET risk subtypes.

The term radiomics reflects a process of converting digital medical images into high-dimensional data by extracting a high number of handcrafted quantitative imaging features based on a wide range of mathematical and statistical methods.¹⁸ Most of the quantitative features extracted through computerized algorithms are beyond visual interpretation,^{36,37} and these radiomic features reflect the texture features of tumors, which are important biomarkers of tumor heterogeneity.³⁸ However, the certain association between biological behaviors and radiomic features has not still been clarified.

A few studies have examined characteristics of radiomic features with ^{18}F -FDG-PET/CT in TETs^{19,20} as mentioned in the Introduction.

In our study, one first-order feature, skewness (first-ranked radiomic feature); one shape feature, sphericity (second-ranked radiomic

Figure 3. Image of a 52-year-old female with thymic carcinoma. The transaxial ^{18}F -FDG-PET/CT image shows ^{18}F -FDG uptake in the primary lesion. The green lines represent the borders of the VOI. The SUVmax (cut-off, 5.1), MTV (cut-off, 14.8), and TLG (cut-off, 63.8) were 4.2, 2.3 and 4.1, respectively. All the SUV-related parameters were negative for thymic carcinoma as per each threshold criterion. The calculated probability score as per the logistic regression model for the diagnosis of thymic carcinoma was 0.890. As per this probability score, this tumor was classified as thymic carcinoma. Moreover, the calculated probability score based on the RF model for the diagnosis of high-risk TET was 0.50. As per this probability score, this tumor was classified as a high-risk TET. ^{18}F -FDG-PET, 18-fluodeoxyglucose positron emission tomography; MTV, metabolic tumor volume; RF, random forest; SUVmax, maximum standardized uptake value; TET, thymic epithelial tumor; TLG, total lesion glycolysis; VOI, volume of interest.



feature); and one higher-order feature, GLSZM-gray level non-uniformity (third-ranked radiomic feature) were highly associated with the prediction of thymic carcinomas, while two first-order features, including skewness (first-ranked radiomic feature) and kurtosis (second-ranked radiomic feature) and one higher-order feature, GLSZM-small area emphasis (third-ranked radiomic feature), were highly associated with the prediction of high-risk TETs. About the prediction of high-risk thymomas, three most important features including skewness (first-ranked radiomic feature), surface volume ratio (second-ranked radiomic feature) and kurtosis (third-ranked radiomic feature) were highly associated. Moreover, skewness was the first-ranked radiomic feature for the prediction of thymic carcinomas, high-risk TETs and high-risk thymomas, respectively. On the additional supplemental analyses of TETs with $\text{MTV} > 10.0\text{ cm}^3$ and $\text{SUV} \geq 2.5$, although the three most important features for the prediction of thymic carcinomas, high-risk TETs or high-risk thymomas were different from those of the whole population analyses, skewness was also the first-ranked radiomic feature for the prediction of thymic carcinomas, high-risk TETs and high-risk thymomas, respectively (Supplementary Table 5). These findings suggest that skewness may be the most useful ^{18}F -FDG PET-based radiomic feature for predicting the high risk TETs.

Some recent studies have proposed classification methods based on machine-learning approaches.^{23-25,39} Hyun et al³⁹ examined the usefulness of a machine-learning approach by using the ^{18}F -FDG

PET-based radiomics to predict the histological subtypes of lung cancers. In their study, the adoption of the logistic regression model as a classifier yielded the highest classification performance. Ahn et al²⁵ used a machine-learning approach to examine the prognostic value of ¹⁸F-FDG PET-based radiomics in patients with non-small cell lung cancer. The RF model best predicted disease recurrence in their study.

In our study, 107 PET-based radiomic and 1024 deep-learning features were examined to predict thymic carcinomas, high-risk TETs or high-risk thymomas with machine-learning approaches; the logistic regression model was the best-performing classifier for thymic carcinomas (AUC = 0.900, accuracy 81.0%), while the RF model was the best-performing classifier for high-risk TETs (AUC = 0.744, accuracy 72.2%). The AUC for predicting thymic carcinomas was significantly higher in the logistic regression model than the three conventional SUV-related parameters (SUVmax, MTV, and TLG, each; $p < 0.01$). Moreover, the AUC for predicting high-risk TETs was significantly higher in the RF model than the conventional SUV-related parameters except for SUVmax (MTV and TLG, each; $p < 0.05$). Although the decision tree model was the best-performing classifier for high-risk thymomas (AUC = 0.680, accuracy 67.2%), no significant difference was observed in the AUC between the decision tree model and three conventional SUV-related parameters (SUVmax, MTV, and TLG, each; $p > 0.05$). These findings indicate that a machine-learning approach using ¹⁸F-FDG PET radiomic features might be useful for predicting thymic carcinomas or high-risk TETs. However, it might be difficult for differentiating between low- and high-risk thymomas using ¹⁸F-FDG PET radiomic features, even if the machine-learning approach is applied for the analyses.

On the additional supplemental analyses for TETs with MTV > 10.0 cm³ and SUV ≥ 2.5 , the best-performing classifier for thymic carcinomas, high-risk TETs and high-risk thymomas were RF model (AUC = 0.935, accuracy 86.4%), SVM model (AUC = 0.844, accuracy 79.7%) and SVM model (AUC = 0.670, accuracy 62.8%), respectively (Supplementary Table 4). The AUC for predicting thymic carcinomas was significantly higher in the RF model than the three conventional SUV-related parameters (SUVmax, MTV, and TLG, each; $p < 0.01$), and the AUC for predicting high-risk TETs was significantly higher in the SVM model than the conventional SUV-related parameters except for SUVmax (MTV and TLG, each; $p < 0.05$), and no significant difference was observed in the AUC between the SVM model and three conventional SUV-related parameters (SUVmax, MTV, and TLG, each; $p > 0.05$). Although the type of best-performing model was different from that of the whole population analyses, the diagnostic performance including AUC and accuracy of each best-performing model for predicting thymic carcinomas, high-risk TETs or high-risk thymomas was almost the same diagnostic performance as the whole population analyses.

This study has certain limitations. First, this was a retrospective study with a relatively small sample; therefore, case selection bias was unavoidable. A prospective study

involving a much larger population is needed to validate and confirm our current findings. Second, texture analysis has been reported to be influenced by tumor volume, especially small tumor volume.²⁷ Thus, the influence of tumor volume might not be ignored for the obtained results of the whole population analyses. However, in the supplemental analyses of 59 TETs with MTV > 10.0 cm³ and SUV ≥ 2.5 (MTV range; 10.4–724 cm³), the obtained results were almost the same as the whole population analyses (MTV range; 0.6–724 cm³). Moreover, the influence of noise or low resolution in PET images might not be ignored for PET-based radiomic analysis.⁴⁰ However, the presented machine-learning methods produced the good results. In this connection, the MTV (0.6–46.0 cm³) was relatively small in the 20 TETs (5 thymic carcinomas, 7 high-risk thymomas and 8 low-risk thymomas) which were excluded by MTV > 10.0 cm³ and SUV ≥ 2.5 criteria from 79 whole patients. However, in this group, the diagnostic accuracy by the respective best model was 85.0% (17/20) for predicting thymic carcinomas by the logistic regression model (5 thymic carcinomas vs 15 thymomas), 85.0% (17/20) for predicting the high-risk TETs by the RF model (12 high-risk TETs vs 8 low-risk thymomas), and 73.3% (11/15) for predicting high-risk thymomas by the decision tree model (7 high-risk thymomas vs 8 low-risk thymomas), respectively. These results are also almost the same as those of 79 and 59 patient groups.

Machine-learning algorithms have the ability to analyze various data types and focus on making predictions as accurate as possible,⁴¹ thus, a machine-learning approach might have the potential to overcome these limitations of PET-based radiomic analyses. Third, lesions with slight ¹⁸F-FDG uptake with small tumor size were excluded from the analyses, because VOIs could not be created. Thus, it is problematic how to predict the risk of TETs in the tumors which could not be performed the ¹⁸F-FDG PET-based radiomic analyses using a machine-learning approach. Further researches are required how to predict the risk of TETs in these tumors. Fourth, although the internal validation showed high classification performance by the machine-learning models, the external validation was not performed due to the difficulty to collect additional patients, and the lack of external validation limits the generalizability of our results. Therefore, a training–test scheme, requiring a large sample, is preferred for the validation of the classifiers.

CONCLUSIONS

The ¹⁸F-FDG PET-based radiomic analysis using a machine-learning approach may be useful for predicting the pathological risk subtypes of the TETs.

CONFLICT OF INTEREST

The authors declare that they have no conflict of interest except employment by GE Healthcare Japan (Akie Katsuki and Kazuyuki Ohmura). The Kagoshima University Department of Radiology received the technical support from GE Healthcare Japan.

REFERENCES

- Venuta F, Anile M, Diso D, Vitolo D, Rendina EA, De Giacomo T, et al. Thymoma and thymic carcinoma. *Eur J Cardiothorac Surg* 2010; **37**: 13–25. <https://doi.org/10.1016/j.ejcts.2009.05.038>
- Girard N, Ruffini E, Marx A, Faivre-Finn C, Peters S, ESMO Guidelines Committee. Thymic epithelial tumours: ESMO Clinical Practice Guidelines for diagnosis, treatment and follow-up. *Ann Oncol* 2015; **26** Suppl 5: v40–55. <https://doi.org/10.1093/annonc/mdv277>
- Okumura M, Ohta M, Tateyama H, Nakagawa K, Matsumura A, Maeda H, et al. The World Health Organization histologic classification system reflects the oncologic behavior of thymoma: a clinical study of 273 patients. *Cancer* 2002; **94**: 624–32. <https://doi.org/10.1002/cncr.10226>
- Okumura M, Miyoshi S, Fujii Y, Takeuchi Y, Shiono H, Inoue M, et al. Clinical and functional significance of WHO classification on human thymic epithelial neoplasms: a study of 146 consecutive tumors. *Am J Surg Pathol* 2001; **25**: 103–10. <https://doi.org/10.1097/00000478-200101000-00012>
- Jeong YJ, Lee KS, Kim J, Shim YM, Han J, Kwon OJ. Does CT of thymic epithelial tumors enable us to differentiate histologic subtypes and predict prognosis. *AJR Am J Roentgenol* 2004; **183**: 283–89. <https://doi.org/10.2214/ajr.183.2.1830283>
- Müller-Hermelink HK, Engel P, Kuo TT, et al. Tumors of the thymus. In: *Pathology & Genetics, Tumours of the Lung, Pleura, Thymus and Heart*. Lyon: IARC Press; 2004, pp. 145–247.
- Chen G, Marx A, Chen W-H, Yong J, Puppe B, Stroebel P, et al. New WHO histologic classification predicts prognosis of thymic epithelial tumors: a clinicopathologic study of 200 thymoma cases from China. *Cancer* 2002; **95**: 420–29. <https://doi.org/10.1002/cncr.10665>
- Kondo K, Yoshizawa K, Tsuyuguchi M, Kimura S, Sumitomo M, Morita J, et al. WHO histologic classification is a prognostic indicator in thymoma. *Ann Thorac Surg* 2004; **77**: 1183–88. <https://doi.org/10.1016/j.athoracsur.2003.07.042>
- Sung YM, Lee KS, Kim BT, Choi JY, Shim YM, Yi CA. 18F-FDG PET/CT of thymic epithelial tumors: usefulness for distinguishing and staging tumor subgroups. *J Nucl Med* 2006; **47**: 1628–34.
- Terzi A, Bertolaccini L, Rizzardi G, Luzzi L, Bianchi A, Campione A, et al. Usefulness of 18-F FDG PET/CT in the pre-treatment evaluation of thymic epithelial neoplasms. *Lung Cancer* 2011; **74**: 239–43. <https://doi.org/10.1016/j.lungcan.2011.02.018>
- Nakajo M, Kajiya Y, Tani A, Yoneda S, Shirahama H, Higashi M, et al. 18F-FDG PET for grading malignancy in thymic epithelial tumors: significant differences in 18F-FDG uptake and expression of glucose transporter-1 and hexokinase II between low and high-risk tumors: preliminary study. *Eur J Radiol* 2012; **81**: 146–51. <https://doi.org/10.1016/j.ejrad.2010.08.010>
- Endo M, Nakagawa K, Ohde Y, Okumura T, Kondo H, Igawa S, et al. Utility of 18F-FDG PET for differentiating the grade of malignancy in thymic epithelial tumors. *Lung Cancer* 2008; **61**: 350–55. <https://doi.org/10.1016/j.lungcan.2008.01.003>
- Viti A, Bertolaccini L, Cavallo A, Fortunato M, Bianchi A, Terzi A. 18-Fluorine fluorodeoxyglucose positron emission tomography in the pretreatment evaluation of thymic epithelial neoplasms: a ‘metabolic biopsy’ confirmed by Ki-67 expression. *Eur J Cardiothorac Surg* 2014; **46**: 369–74. <https://doi.org/10.1093/ejcts/ezu030>
- Matsumoto I, Oda M, Takizawa M, Waseda R, Nakajima K, Kawano M, et al. Usefulness of fluorine-18 fluorodeoxyglucose-positron emission tomography in management strategy for thymic epithelial tumors. *Ann Thorac Surg* 2013; **95**: 305–10. <https://doi.org/10.1016/j.athoracsur.2012.09.052>
- Treglia G, Sadeghi R, Giovanella L, Cafarotti S, Filosso P, Lococo F. Is 18F-FDG PET useful in predicting the WHO grade of malignancy in thymic epithelial tumors? A meta-analysis. *Lung Cancer* 2014; **86**: 5–13. <https://doi.org/10.1016/j.lungcan.2014.08.008>
- Benveniste MFK, Moran CA, Mawlawi O, Fox PS, Swisher SG, Munden RF, et al. FDG PET-CT aids in the preoperative assessment of patients with newly diagnosed thymic epithelial malignancies. *J Thorac Oncol* 2013; **8**: 502–10. <https://doi.org/10.1097/JTO.0b013e3182835549>
- Park SY, Cho A, Bae MK, Lee CY, Kim DJ, Chung KY. Value of 18F-FDG PET/CT for predicting the world health organization malignant grade of thymic epithelial tumors. *Clin Nucl Med* 2016; **41**: 15–20. <https://doi.org/10.1097/RLU.0000000000001032>
- Gillies RJ, Kinahan PE, Hricak H. Radiomics: images are more than pictures, they are data. *Radiology* 2016; **278**: 563–77. <https://doi.org/10.1148/radiol.2015151169>
- Lee HS, Oh JS, Park YS, Jang SJ, Choi IS, Ryu JS. Differentiating the grades of thymic epithelial tumor malignancy using textural features of intratumoral heterogeneity via (18)F-FDG PET/CT. *Ann Nucl Med* 2016; **30**: 309–19. <https://doi.org/10.1007/s12149-016-1062-2>
- Nakajo M, Jinguji M, Shinaji T, Nakajo M, Aoki M, Tani A, et al. Texture analysis of 18F-FDG PET/CT for grading thymic epithelial tumours: usefulness of combining SUV and texture parameters. *Br J Radiol* 2018; **91**: 20170546: 1083. <https://doi.org/10.1259/bjr.20170546>
- Erickson BJ, Korfiatis P, Akkuz Z, Kline TL. Machine learning for medical imaging. *Radiographics* 2017; **37**: 505–15. <https://doi.org/10.1148/rg.2017160130>
- Waljee AK, Higgins PDR. Machine learning in medicine: A primer for physicians. *Am J Gastroenterol* 2010; **105**: 1224–26. <https://doi.org/10.1038/ajg.2010.173>
- Gao X, Chu C, Li Y, Lu P, Wang W, Liu W, et al. The method and efficacy of support vector machine classifiers based on texture features and multi-resolution histogram from (18)F-FDG PET-CT images for the evaluation of mediastinal lymph nodes in patients with lung cancer. *Eur J Radiol* 2015; **84**: 312–17. <https://doi.org/10.1016/j.ejrad.2014.11.006>
- Ypsilantis P-P, Siddique M, Sohn H-M, Davies A, Cook G, Goh V, et al. Predicting response to neoadjuvant chemotherapy with PET imaging using convolutional neural networks. *PLoS One* 2015; **10**(9): e0137036. <https://doi.org/10.1371/journal.pone.0137036>
- Ahn HK, Lee H, Kim SG, Hyun SH. Pre-treatment 18F-FDG PET-based radiomics predict survival in resected non-small cell lung cancer. *Clin Radiol* 2019; **74**: 467–73. <https://doi.org/10.1016/j.crad.2019.02.008>
- Shen W-C, Chen S-W, Wu K-C, Hsieh T-C, Liang J-A, Hung Y-C, et al. Prediction of local relapse and distant metastasis in patients with definitive chemoradiotherapy-treated cervical cancer by deep learning from [18F]-fluorodeoxyglucose positron emission tomography/computed tomography. *Eur Radiol* 2019; **29**: 6741–49. <https://doi.org/10.1007/s00330-019-06265-x>
- Yip SSE, Aerts HJWL. Applications and limitations of radiomics. *Phys Med Biol* 2016; **61**: R150–66. <https://doi.org/10.1088/0031-9155/61/13/R150>
- Nakajo M, Jinguji M, Tani A, Kikuno H, Hirahara D, Togami S, et al. Application of a

- machine learning approach for the analysis of clinical and radiomic features of pretreatment [¹⁸F]-fdg pet/ct to predict prognosis of patients with endometrial cancer. *Mol Imaging Biol* 2021; **23**: 756–65. <https://doi.org/10.1007/s11307-021-01599-9>
29. van Griethuysen JJM, Fedorov A, Parmar C, Hosny A, Aucoin N, Narayan V, et al. Computational radiomics system to decode the radiographic phenotype. *Cancer Res* 2017; **77**: e104–7. <https://doi.org/10.1158/0008-5472.CAN-17-0339>
30. Petersen ML, Molinaro AM, Sinisi SE, van der Laan MJ. Cross-validated bagged learning. *J Multivar Anal* 2008; **25**: 260–66. <https://doi.org/10.1016/j.jmva.2007.07.004>
31. Jung Y. Multiple predicting K-fold cross-validation for model selection. *Journal of Nonparametric Statistics* 2017; **30**: 197–215. <https://doi.org/10.1080/10485252.2017.1404598>
32. Cook JA, Ranstam J. Overfitting. *Br J Surg* 2016; **103**: 1814. <https://doi.org/10.1002/bjs.10244>
33. Guyon I, Weston J, Barnhill S, Vapnik V. Gene selection for cancer classification using support vector machines. *Mach Learn* 2002; **46**: 389–422. <https://doi.org/10.1023/A:1012487302797>
34. Pedregosa F, Varoquaux G, Gramfort A, Michel V, Thirion B, Grisel O, et al. Scikit-learn: machine learning in Python. *J Mach Learn Res* 2011; **12**: 2825–30.
35. DeLong ER, DeLong DM, Clarke-Pearson DL. Comparing the areas under two or more correlated receiver operating characteristic curves: A nonparametric approach. *Biometrics* 1988; **44**: 837–45. <https://doi.org/10.2307/2531595>
36. Braman NM, Etesami M, Prasanna P, Dubchuk C, Gilmore H, Tiwari P, et al. Intratumoral and peritumoral radiomics for the pretreatment prediction of pathological complete response to neoadjuvant chemotherapy based on breast DCE-MRI. *Breast Cancer Res* 2017; **19**: 57. <https://doi.org/10.1186/s13058-017-0846-1>
37. Li Y, Liu X, Xu K, Qian Z, Wang K, Fan X, et al. MRI features can predict EGFR expression in lower grade gliomas: A voxel-based radiomic analysis. *Eur Radiol* 2018; **28**: 356–62. <https://doi.org/10.1007/s00330-017-4964-z>
38. Tran B, Dancey JE, Kamel-Reid S, McPherson JD, Bedard PL, Brown AMK, et al. Cancer genomics: technology, discovery, and translation. *J Clin Oncol* 2012; **30**: 647–60. <https://doi.org/10.1200/JCO.2011.39.2316>
39. Hyun SH, Ahn MS, Koh YW, Lee SJ. A machine-learning approach using PET-based radiomics to predict the histological subtypes of lung cancer. *Clin Nucl Med* 2019; **44**: 956–60. <https://doi.org/10.1097/RLU.0000000000002810>
40. Brooks FJ, Grigsby PW. The effect of small tumor volumes on studies of intratumoral heterogeneity of tracer uptake. *J Nucl Med* 2014; **55**: 37–42. <https://doi.org/10.2967/jnumed.112.116715>
41. Ngiam KY, Khor IW. Big data and machine learning algorithms for health-care delivery. *Lancet Oncol* 2019; **20**: e262–73. [https://doi.org/10.1016/S1470-2045\(19\)30149-4](https://doi.org/10.1016/S1470-2045(19)30149-4)