Original Article

Machine learning models from computed tomography to diagnose thymic epithelial tumors requiring combined resection

Yuki Onozato[1]^, Hidemi Suzuki[1], Hiroki Matsumoto[2], Takamasa Ito[1], Takayoshi Yamamoto[3], Kazuhisa Tanaka[1], Yuichi Sakairi[1], Yukiko Matsui[1], Takekazu Iwata[3], Tomohiko Iida[2], Toshihiko Iizasa[3], Ichiro Yoshino[1]

[1]Department of General Thoracic Surgery, Chiba University Graduate School of Medicine, Inohana, Chuo-ku, Chiba-shi, Chiba, Japan; [2]Department of Thoracic Surgery, Kimitsu Chuo Hospital, Sakurai, Kisarazu-shi, Chiba, Japan; [3]Division of Thoracic Surgery, Chiba Cancer Centre, Nitona-cho, Chuo-ku, Chiba-shi, Chiba, Japan

Contributions: (I) Conception and design: Y Onozato, H Suzuki; (II) Administrative support: Y Sakairi, T Iwata, T Iida, T Iizasa; (III) Provision of study materials or patients: H Matsumoto, T Ito, T Yamamoto, K Tanaka; (IV) Collection and assembly of data: Y Onozato; (V) Data analysis and interpretation: Y Onozato, Y Matsui; (VI) Manuscript writing: All authors; (VII) Final approval of manuscript: All authors.

Correspondence to: Hidemi Suzuki, MD, PhD. Department of General Thoracic Surgery, Chiba University Graduate School of Medicine, 1-8-1, Inohana, Chuo-ku, Chiba 260-8670, Japan. Email: hidemisuzukidesu@yahoo.co.jp.

Background: Minimally invasive approaches have been a standard choice of surgery for noninvasive thymic epithelial tumors (TETs), but we sometimes experience cases requiring combined resection of adjacent structures. We develop and validate machine learning models to predict combined resection based on preoperative contrast-enhanced computed tomography (CT).

Methods: This study included 212 patients with TETs (140 in the training cohort and 72 in the validation cohort) who underwent radical surgery. Radiomics features were extracted from contrast-enhanced CT and predicted with five feature selection methods and seven machine learning models in nested cross validation. The clinical utility of the models was analyzed by a decision curve analysis (DCA).

Results: Fifty-five patients in the training cohort and 28 in the validation cohort required combined resection. The classifiers random forest (RF), gradient boosting (GB), and eXtreme Gradient Boosting (XGB) indicated high predictive performance, with the XGB classifier based on features selected by GB performing the best, with an area under the curve (AUC) of 0.797. In the validation cohort, the classifier had an AUC of 0.817. The DCA showed the validity of the model with a threshold range of 15–72%. When restricted to combined pulmonary and pericardial resection, the respective AUCs were 0.736 and 0.674 for the training cohort and 0.806 and 0.924 for the validation cohort.

Conclusions: The machine learning model based on preoperative CT images was able to diagnose TETs requiring combined resection with high accuracy. The DCA demonstrated a wide range of model validity and may aid in surgical approach selection.

Keywords: Thymic epithelial tumor (TET); surgical procedure; radiomics; machine learning; computed tomography (CT)

---

^ ORCID: 0000-0002-6298-707X.

Introduction

Thymic epithelial tumors (TETs) account for about half of resected mediastinal tumors, with thymoma being the most common at approximately 85% (1). Thymic carcinomas and neuroendocrine tumors have a poor prognosis, with a 5-year survival rate of approximately 60% (2,3). In contrast, thymomas have an oncologically low to intermediate grade and are histopathologically classified as types A, AB, B1, B2, or B3 with varying clinical courses (4). Surgery is the first choice for treatment of TETs, and complete resection is required for improving the overall survival and reducing the recurrence rate (5,6). Since complete resection may require combined resection or reconstruction, depending on the extent of tumor invasion and adhesion to adjacent structures, detailed preoperative surgical planning is essential.

In terms of the surgical approach, the conventional standard approach for TETs has been sternotomy or lateral open thoracotomy. In recent years, several new approaches have been used, such as video-assisted thoracic surgery (VATS) including a uniport (7,8), and robot-assisted thoracic surgery (RATS) (9,10). Minimally invasive surgery (MIS) for locally invasive thymoma has been performed in some centers (8,9), but no oncologic consensus has yet been reached concerning these approaches, and it is not always easy to convert to thoracotomy during an operation.

Previous reports have identified a large tumor size, lobulated tumor contour, presence of calcifications and pulmonary changes adjacent to the tumor on computed tomography (CT) (11,12) as factors associated with tumor invasiveness. We previously reported that fluorine-18-fluorodeoxyglucose positron emission tomography coupled with CT (^{18}F-FDG-PET/CT) is useful for predicting the malignancy grade, staging, and invasiveness of TETs (13). However, PET/CT, although useful, is not routinely performed for TETs. Traditionally, the combination of these risk factors has been used to assess tumor invasiveness, and the surgical approach has been determined but they do not have satisfactory predictive performance.

The field of radiomics, which quantitatively captures imaging findings, has been developed in recent years, and several studies have reported the usefulness of radiomics for evaluating TETs. Most predicted pathological low- and high-risk thymomas using radiomics and machine learning models with CT imaging (14-16). A small number of models have also been developed to predict the residual-factor, which is indicative of complete resection (17). However, models predicting the surgical T-factor, which is directly related to the surgical procedure, have not been constructed.

Machine learning for radiomics is usually performed in two steps. First, radiomics features are selected to improve the accuracy of the classifier and to reduce the amount of calculation. Next, a machine learning model is built based on the selected features. Feature selections are performed selecting features based on the relationship between individual variables and the objective variable, learning and selecting features using a subset of features such as recursive feature elimination (RFE) (18), and feature selection simultaneously with model learning such as least absolute shrinkage and selection operator (LASSO) and machine learning models (19). While all have proven effective, the model that fits each task needs to be validated individually, and multiple methods of analysis were tested. Various machine learning models have been devised, classically logistic regression (LR), support vector machine (SVM) (20,21), k-nearest neighbor (KNN) (22), Naïve Bayes (NB) (23), etc., and more recently random forest (RF) (24), gradient boosting (GB) (25), and eXtreme Gradient Boosting (XGB) (26). LR is an analytical method often used in statistics, with an emphasis on prediction in machine learning. SVM maps data into a high-dimensional space and divides it by finding optimal boundaries between different classes. KNN uses majority voting to determine which group the unknown data fall into. NB is an algorithm based on Bayes' theorem, which uses a probabilistic approach to learn. LR, SVM, KNN, and NB have been developed decades ago and have been validated in many studies. RF builds multiple decision trees based on the data extracted by bootstrapping and

Highlight box

Key findings
• The machine learning model based on preoperative computed tomography (CT) images was able to diagnose thymic epithelial tumors (TETs) requiring combined resection.

What is known and what is new?
• Several studies have shown validity in predicting pathologic grade of TETs from CT images.
• However, machine learning models have not yet been constructed to directly predict combined resections for TETs, which are critical to surgery.

What is the implication, and what should change now?
• Predicting combined resection with a high degree of accuracy might be useful in selecting procedures and surgical approaches.

outputs predictions by ensembling the results. GB and XGB are ensemble learning algorithms, but they differ from RF in that they train a new decision tree based on the gradient of the loss function for the weak model created. XGB is an efficient implementation of GB. A simplified figure is shown in Figure S1. In the present study, radiomic features were extracted and selected from tumors in contrast-enhanced CT scans, and we investigated whether machine learning models could diagnose tumors requiring combined resection of adjacent structures (TCRs). We present this article in accordance with the TRIPOD reporting checklist (available at https://jtd.amegroups.com/article/view/10.21037/jtd-23-1840/rc).

Methods

Patients

This was a three-center retrospective study. The training cohort consisted of patients who underwent thymectomy at Chiba University Graduate School of Medicine from January 2009 to October 2022 and were diagnosed with TET by pathology. The validation cohort included patients with the same conditions at other two hospitals (Chiba Cancer Center and Kimitsu Chuo Hospital) from January 2006 to October 2022 and January 2011 to October 2022. Exclusion criteria were as follows: (I) contrast-enhanced CT with slice thickness ≤5 mm not taken within 90 days prior to surgery; (II) contrast-enhanced CT obtained other than 60 seconds after contrast injection; (III) preoperative chemotherapy conducted. Ultimately, 140 patients in the training cohort and 72 patients in the validation cohort were included in the study (*Figure 1*). TCR was defined as cases requiring combined resection equivalent to T2, T3, or T4 based on the tumor, node, metastasis (TNM) classification. Tumors other than TCR were defined as the control group. The study was conducted in accordance with the Declaration of Helsinki (as revised in 2013). The study was approved by institutional ethics board of Chiba University Graduate School of Medicine (No. M10536), and Chiba Cancer Center and Kimitsu Chuo Hospital were informed and agreed with this study. Individual consent for this retrospective analysis was waived.

Feature extraction and selection

Contrast-enhanced CT images were loaded with an open-source 3D slicer software program (version 4.11) and used for segmentation of TETs. For segmentation, we used the grow from seeds algorithm implemented in the 3D slicer program and manually checked and adjusted. The Pyradiomics software package (version 3.0.1; https://github.com/Radiomics/pyradiomics) was used to extract radiomics features for the segmented tumors. The voxels were resampled to a uniform voxel size of 1.0×1.0×1.0 and discretization of CT images was set to a bin width of 25. The filters used were square, square root, logarithm, exponential, gradient, wavelet, and Laplacian of Gaussian.

Ultimately, a total of 1,795 features were extracted and were standardized in the training set. The feature selection methods used were LASSO, RFE, RF, GB, and Boruta. LASSO adds a regularization term to the least squares method, and many coefficients are compressed to 0 (19). RFE is a method of selecting features up to an arbitrary number of features by repeatedly constructing the model and deleting features (18). RF and GB are commonly used machine learning models, and the variable importance can be used to determine which variables contributed to the model construction. Boruta selects features based on variable importance by comparing shadow variables with actual variables. The parameters of the feature selection methods are shown in Table S1.

Machine learning model construction and validation

Seven machine learning models were used to predict TCR. The models used were LR, SVM, KNN, NB, RF, GB, and XGB. The hyperparameters tuned are listed in Table S2. The machine learning model was built using nested cross validation. The training cohort was divided into five parts, four of which were assigned to the training set and the remaining one to the test set. In the training set, features were selected, and the model was constructed by five-fold cross validation. All parts were assigned once to the test set. The prediction probability in the validation cohort was calculated as the average of the five patterns of feature selection and the model constructed in the training cohort. A decision curve analysis (DCA) was also performed to measure the clinical utility of the model (27). A DCA calculates the net benefit of the model based on the relationship between the risk of false positives and false negatives of the model.

Statistical analyses

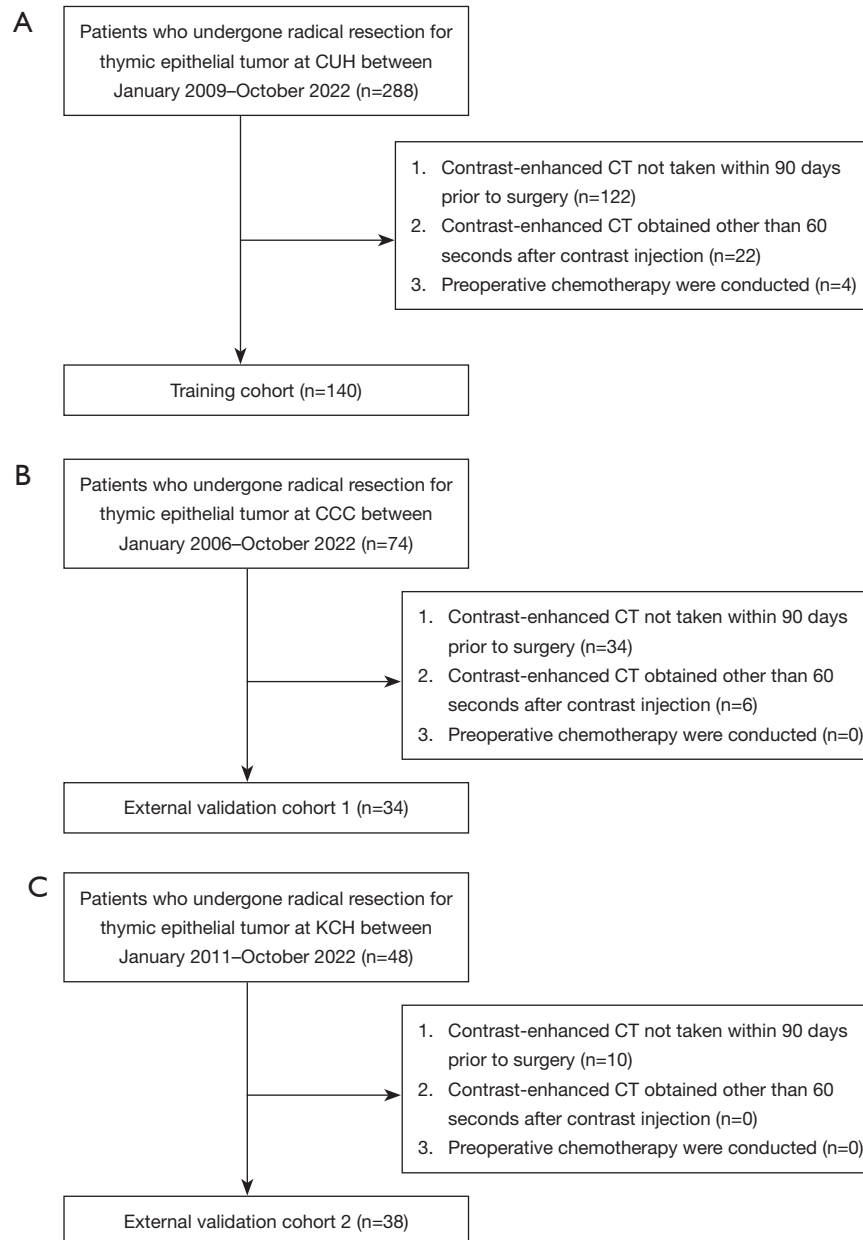Statistical analyses were performed using Fisher's exact test

Figure 1 The flow chart of patient selection. (A) For the training cohort, a total of 288 patients underwent thymectomy in CUH, and 140 met the criteria. (B,C) For the validation cohort, a total of 122 patients underwent thymectomy in CCC and KCH, and 72 met the criteria. CUH, Chiba University Hospital; CT, computed tomography; CCC, Chiba Cancer Center; KCH, Kimitsu Chuo Hospital.

and the Mann-Whitney U test appropriately. All analyses were two-tailed, and P<0.05 was considered significantly different. Statistical analyses and the DCA were performed using the R software program (version 3.6.3, http://www.R-project.org). Machine learning was performed and evaluated using Python (version 3.7) and the scikit-learn package (version 1.0.2).

Results

Patient characteristics

One hundred and forty patients were ultimately included in the training cohort and 72 patients in the validation cohort, and the numbers of TCRs were 55 and 28, respectively.

Patient characteristics are shown in *Table 1*. Gender

tended to be male in the training cohort (P=0.02). Tumor size was significantly different in the training cohort (P=0.02) and validation cohort (P=0.006), with a trend toward larger tumors in the TCR group. Regarding tumor characteristics, there was a significant difference in edge irregularity (P<0.001, P<0.001), tumor heterogeneity (P=0.003, P=0.01), respectively. Tumor calcification was significantly different in the training cohort, with a trend toward more calcification in the TCR group (P=0.007). The most common approach was sternotomy in both cohorts. Blood loss was higher in the TCR group in both cohorts (P<0.001, P<0.001), and the operative time was longer in the TCR group as well (P<0.001, P<0.001). The number of years of experience of the surgeons tended to be longer in the TCR group in the training cohort (P=0.02) than in control group. The clinical T-factor was higher in the TCR group in both cohorts (P<0.001 and P<0.001, respectively) than control group. Of the TCRs, 32 were cT1, of which 9 (28.1%) were pT1. Conversely, 51 patients had cT2 or higher, of which 40 (78.4%) had pT2 and T3. There was no difference in the prevalence of myasthenia gravis in both cohorts. Postoperative complications were more common in the TCR group in the training cohort (P<0.01) than control group, and there were no complications of grade 3 or higher at Clavien-Dindo in the control group in the validation cohort. There was no difference in the follow-up period between the two cohorts. Histologic types were more common with high-grade in the TCR group in both cohorts (P<0.001, P=0.005, respectively). In the training

cohort, the lung was the most frequently resected structure (44 cases, 80.0%) and the pericardium was resected in 33 cases (60.0%); in the validation cohort, 21 cases (75.0%) and 14 cases (50.0%), respectively, were resected.

Machine learning construction and prediction

The results for the training cohort are shown in *Table 2*. The best performing combination was the method with GB feature selection and XGB model construction, with an area under the curve (AUC) of 0.797 [95% confidence interval (CI): 0.721–0.873]. The method using GB for feature selection had a slightly higher AUC than that using Boruta for feature selection, so GB-XGB was used for further analyses. RF and GB with Boruta had AUCs of 0.794 (95% CI: 0.720–0.868) and 0.785 (95% CI: 0.707–0.862), respectively, followed by LR and KNN with GB for feature selection with AUCs of 0.768 (95% CI: 0.688–0.849) and 0.776 (95% CI: 0.698–0.854), respectively. SVM and NB had AUCs of 0.748 (95% CI: 0.665–0.832) and 0.705 (95% CI: 0.613–0.797), respectively, when RFE was used for feature selection. The receiver operating characteristic (ROC) curves for the best-performing feature selection and machine learning models are shown in *Figure 2A*. SVM, RF, and XGB performed well in the validation cohort with AUCs of 0.829, 0.822, and 0.817, respectively (*Figure 2B*). Individual hospital results in the validation cohort are shown in Figure S2.

The number of machine learning models that predicted

Table 1 Patients' characteristics

Characteristics	Training cohort (n=140)			Validation cohort (n=72)		
	TCR (n=55)	Control group (n=85)	P value	TCR (n=28)	Control group (n=44)	P value
Age (years)	58.6±15.3	58.0±14.5	0.44	63.2±10.8	58.7±13.8	0.14
Gender, male/female	35/20	37/48	0.02	17/11	24/20	0.63
CT findings						
Tumor size (mm)	58.0 [44.4–71.2]	46.8 [35.2–60.9]	0.02	60.1 [42.9.5–75.1]	45.5 [36.0–52.4]	0.006
Edge irregular/regular	30/25	24/61	<0.001	16/12	7/37	<0.001
Heterogeneity	35/20	32/53	0.003	15/13	10/34	0.01
Calcification	16/39	9/76	0.007	8/20	6/38	0.14

Table 1 (*continued*)

Table 1 (*continued*)

Characteristics	Training cohort (n=140)			Validation cohort (n=72)		
	TCR (n=55)	Control group (n=85)	P value	TCR (n=28)	Control group (n=44)	P value
Surgical procedure			0.22			0.41
Thymectomy	37	48		23	32	
Extended thymectomy	18	37		5	12	
Surgical approach			0.002			–
Sternum	42	43		24	22	
Thoracotomy	6	7		3	7	
VATS	6	32		1	14	
RATS	1	3		0	1	
Bleeding (mL)	200 [85–370]	70 [5–140]	<0.001	294 [194–450]	80 [15–149]	<0.001
Surgery time (min)	208 [169–255]	151 [123–200]	<0.001	256 [207–297]	158 [131–221]	<0.001
Years of surgeon (years)	15 [12–18]	13 [10–15]	0.02	12 [9–15]	10 [6–16]	0.29
Clinical classification						
cT1/2/3/4	21/3/31/0	71/3/11/0	<0.001	11/10/7/0	38/4/2/0	<0.001
cN0/1/2	55/0/0	85/0/0	–	28/0/0	44/0/0	–
cM0/1	54/1	85/0	–	26/2	44/0	–
MG +/–	14/41	21/64	>0.99	2/26	4/40	>0.99
Clavien-Dindo ≥3	13	7	0.01	4	0	–
Follow-up period (days)	2,181 [955–3,066]	1,842 [1,141–3,269]	0.91	1,879 [1,376–2,612]	2,307 [1,296–3,417]	0.44
Tumor histology			<0.001			0.005
A/AB/B1	2/7/7	1/36/13		5/3/2	3/18/9	
B2/B3/carcinoma	13/12/14	26/5/4		3/4/11	6/2/6	
Pathological classification						
pT1/2/3/4	8/13/34/0	85/0/0/0	–	12/3/13/0	44/0/0/0	–
pN0/1/2	52/2/1	85/0/0	–	25/2/1	44/0/0	–
pM0/1	51/4	85/0	–	22/6	44/0	–
Residual tumor	10	2	0.002	7	4	0.10
Combined resection			–			–
Lung	44 (80.0)	0		21 (75.0)	0	
Pericardium	33 (60.0)	0		14 (50.0)	0	
SVC or BV	11 (20.0)	0		9 (32.1)	0	
Phrenic nerve	13 (23.6)	0		7 (25.0)	0	

Data are presented as mean ± SD, number, median [IQR], or n (%). TCR, tumors requiring combined resection of adjacent structure; CT, computed tomography; VATS, video-assisted thoracic surgery; RATS, robot-assisted thoracic surgery; MG, myasthenia gravis; SVC, superior vena cava; BV, brachiocephalic vein; IQR, interquartile range; SD, standard deviation.

Table 2 The results of machine learning

Model	Feature selector				
	LASSO	RFE	RF	GB	Boruta
LR	0.710 (0.622–0.799)	0.757 (0.673–0.840)	0.729 (0.642–0.816)	0.768 (0.688–0.849)	0.759 (0.677–0.840)
SVM	0.670 (0.579–0.761)	0.748 (0.665–0.832)	0.702 (0.614–0.789)	0.729 (0.646–0.811))	0.693 (0.604–0.783)
KNN	0.704 (0.615–0.792)	0.733 (0.648–0.818)	0.724 (0.638–0.811)	0.776 (0.698–0.854)	0.715 (0.629–0.801)
NB	0.666 (0.573–0.758)	0.705 (0.613–0.797)	0.696 (0.604–0.788)	0.666 (0.574–0.758)	0.668 (0.573–0.763)
RF	0.778 (0.699–0.856)	0.772 (0.694–0.850)	0.785 (0.710–0.861)	0.780 (0.701–0.859)	0.794 (0.720–0.868)
GB	0.732 (0.641–0.822)	0.733 (0.648–0.817)	0.747 (0.664–0.830)	0.772 (0.689–0.855)	0.785 (0.707–0.862)
XGB	0.771 (0.691–0.852)	0.751 (0.669–0.834)	0.778 (0.700–0.856)	0.797 (0.721–0.873)	0.797 (0.723–0.871)

Data are presented as AUC (95% CI). LASSO, least absolute shrinkage and selection operator; RFE, recursive feature elimination; RF, random forest; GB, gradient boosting; LR, logistic regression; SVM, support vector machine; KNN, k-nearest neighbor; NB, Naïve Bayes; XGB, eXtreme Gradient Boosting; AUC, area under the curve; CI, confidence interval.
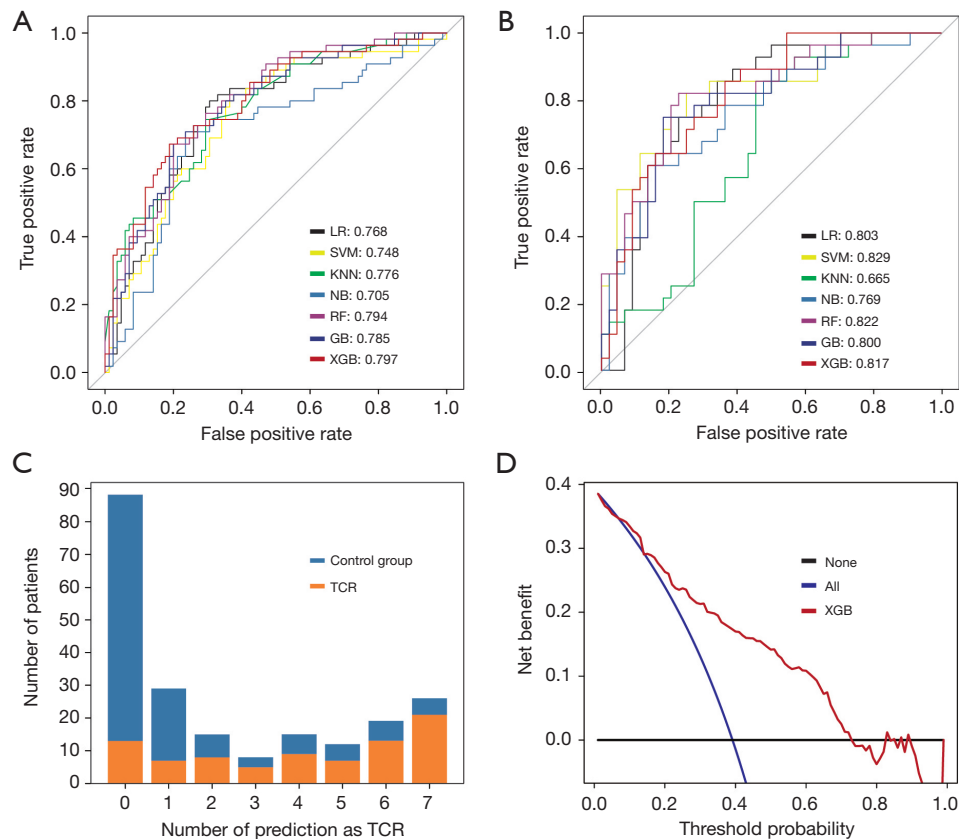

Figure 2 The prediction results for TCR. (A) ROC curves show the results of each machine learning model with the best performing feature selection. (B) The results of predicting TCR in the validation cohort. (C) Combining the training and validation cohorts and showing how many models predicted as TCR when the cut-off of the machine learning model was set to 0.5. (D) Results of a DCA based on the predicted probability of TCR for all patients. LR, logistic regression; SVM, support vector machine; KNN, k-nearest neighbor; NB, Naïve Bayes; RF, random forest; GB, gradient boosting; XGB, eXtreme Gradient Boosting; TCR, tumors requiring combined resection of adjacent structure; ROC, receiver operating characteristic; DCA, decision curve analysis.
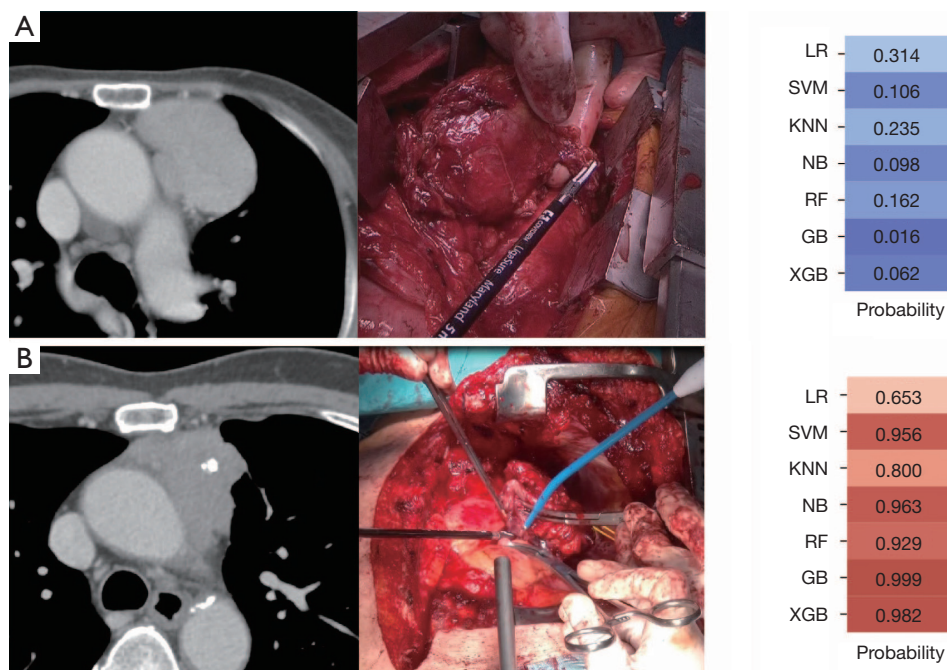
4942

Onozato et al. Predicting combined resection for TETs

Figure 3 Representative cases. (A) Case 1 is an 84-year-old woman. The tumor was 7.5 cm in diameter with a smooth surface and uniform interior. Thymectomy was performed through a median sternotomy approach, but no adhesion or invasion was observed. The predicted probability of TCR was 6.2% for XGB. (B) Case 2 is a 72-year-old man. The tumor was 7.0 cm with irregular margins and internal calcification. A median sternotomy approach was initiated, but the tumor was firmly adhered to or had infiltrated the aorta, brachiocephalic vein, and pulmonary artery. Therefore, additional left fourth intercostal thoracotomy was performed. The brachiocephalic vein and transverse nerves had to be reconstructed. The predicted probability of TCR was 98.2% by XGB. LR, logistic regression; SVM, support vector machine; KNN, k-nearest neighbor; NB, Naïve Bayes; RF, random forest; GB, gradient boosting; XGB, eXtreme Gradient Boosting; TCR, tumors requiring combined resection of adjacent structure.

as TCRs in all cases and the actual number of TCRs are shown in *Figure 2C*. With a cut-off of 0.5, the percentage of correct responses was higher when all were predicted to be negative or positive than the others. The overall accuracy was 75.0% but was 84.2% for all model-matched cases. The predictive probability of the model was analyzed by a DCA with GB-XGB and showed the usefulness of the model in the threshold range of 15–72% (*Figure 2D*). Representative cases were shown in *Figure 3*.

Prediction of the surgical invasiveness of lung and pericardium

A new model was built using a combination of GB-XGB classifier, to predict the need for combined pulmonary and pericardial resection, respectively. The performance of the respective constructed models was AUC 0.736 (95% CI: 0.639–0.833) for lung resection and AUC 0.674 (95% CI: 0.556–0.793) for pericardial resection. The same model was fitted to the validation cohort and predicted AUCs of 0.806 for lung resection and 0.924 for pericardial resection (*Figure 4A,4B*). The DCA analysis of the combined training and validation cohort patients demonstrated the usefulness of the model, with a threshold range of 14–61% for the lung and 10–48% for the pericardium (*Figure 4C,4D*).

Discussion

In the present study, TETs were segmented from preoperative contrast-enhanced CT images, and radiomic features were extracted. Based on the extracted features, TCRs were predicted by combining multiple feature selection methods and machine learning models.

For predicting TCRs, the model using GB as the feature selection and XGB as the classifier performed best with an AUC of 0.797 (95% CI: 0.721–0.873) for the training
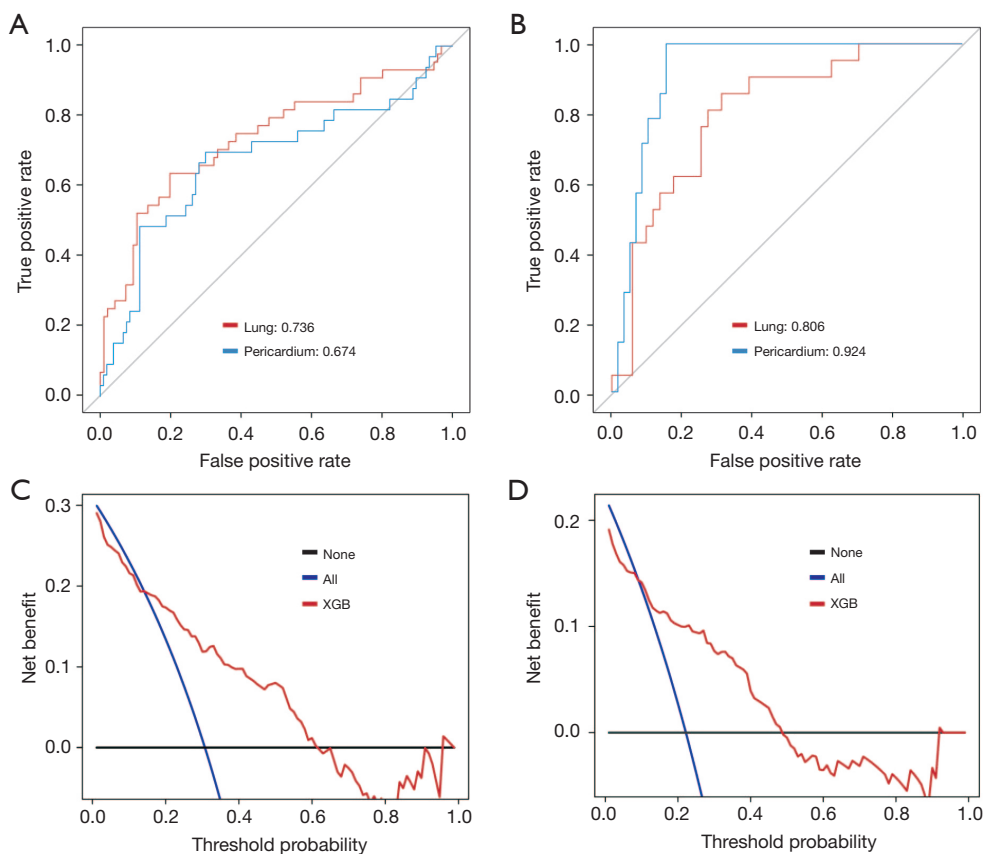
Figure 4 The prediction results for combined resection of lung and pericardium. (A) ROC curves show the results of the GB feature selector and XGB classifier for lung and pericardium. (B) The results of prediction in the validation cohort. (C) Results of a DCA based on the predicted probability of lung. (D) Results of a DCA based on the predicted probability of pericardium. XGB, eXtreme Gradient Boosting; ROC, receiver operating characteristic; GB, gradient boosting; DCA, decision curve analysis.

cohort and an AUC of 0.817 for the validation cohort. It is generally acknowledged that gradient boost performs well on tabular data, and in this study, GB and XGB performed well, regardless of the feature selection method used. In contrast, LR, KNN, SVM, and NB performed inadequately on the average. For the task of predicting TCRs in TETs, RF, GB, and XGB are considered as preferrable. Several studies have constructed multiple machine learning models based on radiomics to make predictions. Shang *et al.* predicted histology in TETs with five feature selection methods and seven classifiers, with the model using SVM-GB achieving an AUC of 0.876 (15). For a similar purpose, Dong *et al.* compared five classifiers with LASSO and obtained the highest AUC of 0.819 for LR (16). Since different machine learning models are suitable for different tasks, it is considered important to build multiple models.

Thymectomy through median sternotomy has

traditionally been the gold-standard approach, but in recent years, MIS have been devised. The International Thymic Malignancy Interest Group states that MIS should be converted to thoracotomy if oncologic principles, such as capsule destruction, incomplete resection, risk of discontinuous resection, are not followed (28). For TETs, the CT findings such as the tumor diameter and edge irregularity, are used to determine the surgical procedure. Conventional methods of determining procedure are reasonably reliable, as MIS have achieved low conversion rates. Burt *et al.* reported that of 943 patients undergoing surgery for stage I or II thymoma, 295 (31.3%) were resected with MIS, of which only 2.6% were converted to open thoracic surgery (29). Conversely, 68.7% of patients underwent thoracotomy, which may include some cases where a MIS could have been selected. Thus, although conventional image evaluations have been shown to be

　　　　J Thorac Dis 2024;16(8):4935-4946 | https://dx.doi.org/10.21037/jtd-23-1840

4944

Onozato et al. Predicting combined resection for TETs

statistically useful to some extent (11,12), to our knowledge, no machine learning model has been constructed that directly predicts the surgical T factor.

The efficacy of the model in identifying tumor invasiveness in relation to surgery has been demonstrated in multiple organs. A SVM classifier using LASSO feature selection to predict local invasiveness of craniopharyngioma was reported to have an AUC of 0.79 (30). Zheng *et al.* also constructed a nomogram with LASSO to discriminate muscular invasiveness of bladder cancer from non-invasive cancer achieving an AUC of 0.876 (31). Given these present and previous findings, the analysis of radiomic features appears highly effective in predicting its relationship with adjacent structures. By combining machine learning models that predict the malignancy of TETs that have been accumulated to date and using machine learning models, it may be possible to construct a prediction model that is more valuable for actual clinical settings.

The present study also included a DCA, which showed that the model was valid over a wide threshold range for diagnosing TCR, but the threshold range of validity was narrower for predicting lung and pericardium alone. This may be due to the low value of the AUC in the training cohort, resulting in insufficient predictive accuracy. The reason for the reduced prediction accuracy may be that only the tumor was segmented and predicted, and the relationship with the adjacent structures was not quantified. Even for similar tumors, the likelihood of invasion may vary depending on the area in contact with adjacent structures.

Several limitations associated with the present study warrant mention. First, the conditions under which contrast CT was performed were not strictly defined to collect as many cases as possible. Our model was constructed and evaluated based on contrast-enhanced CT images acquired with a variety of models, so differences in individual models and imaging conditions may have introduced bias. Second, tumors were segmented using semi-automatic methods. To ensure reproducibility, it is desirable to eliminate manual operations as much as possible and perform automatic segmentation. Third, this study predicted the surgical invasiveness of tumors with surgery in mind. The ability to dissect adhesions depends largely on the skill of the surgeon, which is a surgeon-dependent factor. In the training cohort, more skilled surgeons operated on TCR cases. The influence of the surgeon was not included in the analysis and may be a bias.

Conclusions

In conclusion, the radiomics machine learning model based on preoperatively obtained contrast CT images of TETs was able to predict TCRs and control cases with high accuracy. The model using XGB as the classifier with GB feature selection showed the best performance with an AUC of 0.797 in the training cohort and an AUC of 0.817 in the validation cohort. Machine learning models using contrast-enhanced CT scans can provide accurate information for predicting combined resection of adjacent structures in surgery and may be useful for determining the surgical procedure and approach.

Footnote

Reporting Checklist: The authors have completed the TRIPOD reporting checklist. Available at https://jtd.amegroups.com/article/view/10.21037/jtd-23-1840/rc

Data Sharing Statement: Available at https://jtd.amegroups.com/article/view/10.21037/jtd-23-1840/dss

Peer Review File: Available at https://jtd.amegroups.com/article/view/10.21037/jtd-23-1840/prf

Conflicts of Interest: All authors have completed the ICMJE uniform disclosure form (available at https://jtd.amegroups.com/article/view/10.21037/jtd-23-1840/coif). I.Y. receives grants from Taiho Pharmaceutical, Chugai Pharmaceutical, Shionogi Pharmaceutical, Daiichi-sankyo Chemial Pharma, Eli Lily and Pfizer; consulting fees from Astra Zeneca, Chugai Pharmaceutical, Johnson and Johnson, Medicaroid, Covidien and Intuitive Surgical; honoraria from Astra Zeneca, Chugai Pharmaceutical, Johnson and Johnson, Covidien Japan, Daiichi-sankyo Chemical Pharma, Taho, Eli Lily, Intuitive Surgical, MSD and Bristol-Myers Squib. The other authors have no conflicts of interest to declare.

Ethical Statement: The authors are accountable for all aspects of the work in ensuring that questions related to the accuracy or integrity of any part of the work are

appropriately investigated and resolved. The study was conducted in accordance with the Declaration of Helsinki (as revised in 2013). The study was approved by institutional ethics board of Chiba University Graduate School of Medicine (No. M10536), and Chiba Cancer Center and Kimitsu Chuo Hospital were informed and agreed with this study. Individual consent for this retrospective analysis was waived.

Open Access Statement: This is an Open Access article distributed in accordance with the Creative Commons Attribution-NonCommercial-NoDerivs 4.0 International License (CC BY-NC-ND 4.0), which permits the non-commercial replication and distribution of the article with the strict proviso that no changes or edits are made and the original work is properly cited (including links to both the formal publication through the relevant DOI and the license). See: https://creativecommons.org/licenses/by-nc-nd/4.0/.

References

1. Committee for Scientific Affairs, The Japanese Association for Thoracic Surgery, Shimizu H, et al. Thoracic and cardiovascular surgeries in Japan during 2018 : Annual report by the Japanese Association for Thoracic Surgery. Gen Thorac Cardiovasc Surg 2021;69:179-212.

2. Huang J, Ahmad U, Antonicelli A, et al. Development of the international thymic malignancy interest group international database: an unprecedented resource for the study of a rare group of tumors. J Thorac Oncol 2014;9:1573-8.

3. Tagawa T, Suzuki H, Nakajima T, et al. Long-term outcomes of surgery for thymic carcinoma: experience of 25 cases at a single institution. Thorac Cardiovasc Surg 2015;63:212-6.

4. Weis CA, Yao X, Deng Y, et al. The impact of thymoma histotype on prognosis in a worldwide database. J Thorac Oncol 2015;10:367-72.

5. Girard N, Ruffini E, Marx A, et al. Thymic epithelial tumours: ESMO Clinical Practice Guidelines for diagnosis, treatment and follow-up. Ann Oncol 2015;26 Suppl 5:v40-55.

6. Detterbeck F, Youssef S, Ruffini E, et al. A review of prognostic factors in thymic malignancies. J Thorac Oncol 2011;6:S1698-704.

7. Wu CF, Gonzalez-Rivas D, Wen CT, et al. Single-port video-assisted thoracoscopic mediastinal tumour resection. Interact Cardiovasc Thorac Surg 2015;21:644-9.

8. Jiang L, Chen H, Hou Z, et al. Subxiphoid Versus Unilateral Video-assisted Thoracoscopic Surgery Thymectomy for Thymomas: A Propensity Score Matching Analysis. Ann Thorac Surg 2022;113:1656-62.

9. Kneuertz PJ, Kamel MK, Stiles BM, et al. Robotic Thymectomy Is Feasible for Large Thymomas: A Propensity-Matched Comparison. Ann Thorac Surg 2017;104:1673-8.

10. Geraci TC, Ferrari-Light D, Pozzi N, et al. Midterm Results for Robotic Thymectomy for Malignant Disease. Ann Thorac Surg 2021;111:1675-81.

11. Marom EM, Milito MA, Moran CA, et al. Computed tomography findings predicting invasiveness of thymoma. J Thorac Oncol 2011;6:1274-81.

12. Zhao Y, Chen H, Shi J, et al. The correlation of morphological features of chest computed tomographic scans with clinical characteristics of thymoma. Eur J Cardiothorac Surg 2015;48:698-704.

13. Ito T, Suzuki H, Sakairi Y, et al. 18F-FDG-PET/CT predicts grade of malignancy and invasive potential of thymic epithelial tumors. Gen Thorac Cardiovasc Surg 2021;69:274-81.

14. Feng XL, Wang SZ, Chen HH, et al. Optimizing the radiomics-machine-learning model based on non-contrast enhanced CT for the simplified risk categorization of thymic epithelial tumors: A large cohort retrospective study. Lung Cancer 2022;166:150-60.

15. Shang L, Wang F, Gao Y, et al. Machine-learning classifiers based on non-enhanced computed tomography radiomics to differentiate anterior mediastinal cysts from thymomas and low-risk from high-risk thymomas: A multi-center study. Front Oncol 2022;12:1043163.

16. Dong W, Xiong S, Lei P, et al. Application of a combined radiomics nomogram based on CE-CT in the preoperative prediction of thymomas risk categorization. Front Oncol 2022;12:944005.

17. Araujo-Filho JAB, Mayoral M, Zheng J, et al. CT Radiomic Features for Predicting Resectability and TNM Staging in Thymic Epithelial Tumors. Ann Thorac Surg 2022;113:957-65.

18. Karami G, Giuseppe Orlando M, Delli Pizzi A, et al. Predicting Overall Survival Time in Glioblastoma Patients Using Gradient Boosting Machines Algorithm and Recursive Feature Elimination Technique. Cancers (Basel) 2021;13:4976.

19. Tibshirani R. Regression Shrinkage and Selection Via the Lasso. J Royal Statistical Soc Ser B Methodol 1996;58:267-88.

20. Guyon I, Weston J, Barnhill S, et al. Gene Selection for Cancer Classification using Support Vector Machines. Mach Learn 2002;46:389-422.

21. Xu L, Yang P, Liang W, et al. A radiomics approach based on support vector machine using MR images for preoperative lymph node status evaluation in intrahepatic cholangiocarcinoma. Theranostics 2019;9:5374-85.

22. Guo G, Wang H, Bell D, et al. On The Move to Meaningful Internet Systems 2003: CoopIS, DOA, and ODBASE, OTM Confederated International Conferences CoopIS, DOA, and ODBASE 2003 Catania, Sicily, Italy, November 3–7, 2003 Proceedings. Lecture Notes in Computer Science; 2003:986-96.

23. Wu W, Parmar C, Grossmann P, et al. Exploratory Study to Identify Radiomics Classifiers for Lung Cancer Histology. Front Oncol 2016;6:71.

24. Breiman L. Random forests. Mach Learn 2001;45:5-32.

25. Natekin A, Knoll A. Gradient boosting machines, a tutorial. Front Neurorobot 2013;7:21.

26. Chen T, Guestrin C. XGBoost: a scalable tree boosting system. In: Proceedings of the 22nd ACM SIGKDD International Conference on Knowledge Discovery and Data Mining. Association for Computing Machinery; 2016:785-94.

27. Vickers AJ, Elkin EB. Decision curve analysis: a novel method for evaluating prediction models. Med Decis Making 2006;26:565-74.

28. Toker A, Sonett J, Zielinski M, et al. Standard terms, definitions, and policies for minimally invasive resection of thymoma. J Thorac Oncol 2011;6:S1739-42.

29. Burt BM, Nguyen D, Groth SS, et al. Utilization of Minimally Invasive Thymectomy and Margin-Negative Resection for Early-Stage Thymoma. Ann Thorac Surg 2019;108:405-11.

30. Ma G, Kang J, Qiao N, et al. Non-Invasive Radiomics Approach Predict Invasiveness of Adamantinomatous Craniopharyngioma Before Surgery. Front Oncol 2020;10:599888.

31. Zheng J, Kong J, Wu S, et al. Development of a noninvasive tool to preoperatively evaluate the muscular invasiveness of bladder cancer using a radiomics approach. Cancer 2019;125:4388-98.