Acta Radiologica Open
13(10) 1–9
© The Author(s) 2024
Article reuse guidelines:
sagepub.com/journals-permissions
DOI: 10.1177/20584601241288509
journals.sagepub.com/home/arr
Sage

Differential diagnosis between low-risk and high-risk thymoma: Comparison of diagnostic performance of radiologists with and without deep learning model

Yuriko Yoshida[1], Masahiro Yanagawa[1], Yukihisa Sato[2], Tomo Miyata[3], Atsushi Kawata[4], Akinori Hata[1] and Noriyuki Tomiyama[1]

Abstract

Background: There are few CT-based deep learning (DL) studies on thymoma according to the World Health Organization classification.

Purpose: To develop a CT-based DL model to distinguish between low-risk and high-risk thymoma and to compare the diagnostic performance of radiologists with and without the DL model.

Material and Methods: 159 patients with 160 thymomas were included. A fine-tuning VGG16 network model with Adam optimizer was used, followed by k-fold cross validation. The dataset consisted of three axial slices, including the maximum tumor size from the CT volume data. The data were augmented 50 times by rotation, zoom, shear, and horizontal/vertical flip. Three independent networks for the CT dataset were considered, and the result was determined by voting. Three radiologists independently diagnosed thymomas with and without the model. The area under the curve (AUC) of the diagnostic performance was compared using receiver operating characteristic analysis.

Results: Accuracy of the DL model was 71.3%. Diagnostic performance of the radiologists was as follows: AUC and accuracy without the DL model, 0.61–0.68 and 61.9%–69.3%; and with the DL model, 0.66–0.69 and 68.1%–70.0%, respectively. AUC of the diagnostic performance showed no significant differences between radiologists with and without the DL model. The DL model tended to increase the diagnostic accuracy, but AUC was not significantly improved.

Conclusion: Diagnostic performance of the DL was comparable to that of radiologists. The DL model assistance tended to increase diagnostic accuracy.

Keywords
Deep learning, thymoma, mediastinum, risk classification, CT imaging, radiology

Received 7 December 2023; accepted 17 September 2024

[1]Department of Diagnostic and Interventional Radiology, Osaka University Graduate School of Medicine, Osaksa, Japan
[2]Department of Diagnostic Radiology, Suita Municipal Hospital, Osaka, Japan
[3]Department of Diagnostic Radiology, Sakai City Medical Center, Osaka, Japan
[4]Department of Diagnostic Radiology, Osaka International Cancer Institute, Osaka, Japan

Corresponding author:
Masahiro Yanagawa, Department of Diagnostic and Interventional Radiology, Osaka University Graduate School of Medicine, 2-2 Yamadaoka Suita, Osaksa 565-0871, Japan.
Email: m-yanagawa@radiol.med.osaka-u.ac.jp

Introduction

Thymomas are the most common anterior mediastinal tumors, which are classified into five subtypes (A, AB, B1, B2, and B3) according to the World Health Organization (WHO) histological classification. This histological classification has been reported to represent a prognostic factor for patients with thymomas.[1] Another classification defines types A, AB, and B1 thymomas as low-risk thymomas and types B2 and B3 as high-risk thymomas.[2] The prognosis for low-risk thymoma is reported to be good, whereas the prognosis for high-risk thymoma is poor.[2,3] In cases where an anterior mediastinal tumor is suspected to be a thymoma, preoperative biopsy is generally not recommended due to the potential risk of pleural dissemination associated with needle biopsy procedures. This emphasizes the critical importance of accurate preoperative diagnosis through noninvasive methods. Contrast-enhanced chest CT is an essential preoperative test for thymoma, and several studies have investigated the relationship between CT findings and the WHO classification or simplified risk classification of thymoma. However, considerable overlap between classification-based findings limits the ability to sort by CT findings.[4–6]

In recent years, many deep learning (DL) studies have been conducted on diagnostic imaging. DL has emerged as a potential means for analyzing medical images and has demonstrated a diagnostic accuracy similar to that of radiologists. Currently, various algorithms are being developed to evaluate nodules, classify interstitial lung diseases, and detect aortic dissection on non-contrast-enhanced CT.[7–10] However, there are few deep learning studies on thymoma, and none of them examined the WHO classification of thymoma on CT. Therefore, this study aimed to develop a CT-based DL model to distinguish between low- and high-risk thymoma and to compare the diagnostic performance of radiologists with and without the DL model.

Material and Methods

Patients

This study was approved by the ethical review committee of Osaka University (No. 18096-2) and was conducted in accordance with the principles of the Declaration of Helsinki. Informed consent was waived by the ethical review committee of Osaka University (No. 18096-2) as this was a retrospective review of images and records. A retrospective search of patients who underwent surgical resection between January 2005 and November 2016, had pathologically confirmed thymoma, and underwent CT before surgery in a single institution, identified a total of 185 consecutive patients. Patients were included if they underwent non-enhanced and contrast-enhanced CT with the specified protocol before surgery, while those who underwent dual-energy CT were excluded. In addition, patients who underwent only non-contrast enhanced CT (n = 8), only contrast-enhanced CT (n = 3), contrast-enhanced CT with different timing (n = 3), and dual-energy CT (n = 12) were also excluded. Of the 185 patients, 159 patients with 160 thymomas were finally included (Supplemental Figure 1). Pathologists reported according to the WHO histologic classification of thymomas in our study cohort. Thymomas were classified into two subgroups according to a simplified WHO classification system: types A, AB, and B1 as low-risk thymomas and types B2 and B3 as high-risk thymomas.[2] A radiologist with 8 years of experience (Y.Y.) measured the maximum diameter of each tumor on its largest cross-sectional area in CT images. A total of 160 tumors were evaluated. Measurements were performed on contrast-enhanced CT images with a slice thickness of 5 mm.

CT protocol

All participants were examined using one of the seven CT scanners as follows: Aquilion 4, Aquilion 64, Aquilion ONE GENESIS, Aquilion ONE (Toshiba Medical Systems, Otawa, Tochigi), Light Speed VCT, Discovery CT 750 HD, and Discovery CT 750 HD FE (GE Medical Systems, Milwaukee, WI, USA). The imaging parameters used were as follows: tube current, automatic exposure control; tube voltage, 120 kVp; field of view, 345 mm; collimation, 0.5 mm for Aquilion 4, Aquilion 64, Aquilion ONE GENESIS, and Aquilion ONE or 0.625 mm for Light Speed VCT, Discovery CT 750 HD, and Discovery CT750 HD FE; use of iterative reconstruction, none; matrix size, 512 × 512; scan direction, craniocaudal direction; slice interval of reconstructed images, 5 mm; and reconstruction slice thickness, 5 mm. Non-enhanced and contrast-enhanced CT were performed. Contrast-enhanced CT was performed 60 s after injection. For 20 patients scanned from January 2005 to May 2006, the amount of contrast material was 150 mL at 2 mL/s. For the remaining 140 patients scanned from June 2006 through 2016, the amount was 2 mL/kg and the injection time was 60 s.

Construction of the DL model

The hyperparameters were adjusted and the following models were used for accuracy and reproducibility: All CT data had thymoma lesions graded according to the WHO classification by a single radiologist. Types A, AB, and B1 of the WHO classification were labeled as low-risk thymomas, and types B2 and B3 of the WHO classification were labeled as high-risk thymomas. The center of the thymoma lesion and the boundary box were annotated by the same radiologist. A three-dimensional patch was extracted from the center of the thymoma lesion from the enhanced CT as a digital imaging and communications in medicine (DICOM) standard to cover the whole lesion and resized to a 128 × 128 × 128 cube. The center slice and pre- and post-slices in the axial direction of the cube were generated (Figure 1). These three slices were used as three channel inputs for training with a two-dimensional convolutional neural network (CNN), which is a fine-tuning model based on VGG16 (Figure 2), classifying input data into high- and low-risk thymoma classes. The number of total data was 160 examinations (low-risk: 92 and high-risk: 68); it was randomly divided into 80% training data and 20% test data. Validation data ratio was 10% of the training data. Data augmentation was also applied to the training data with rotation, zoom, shear, and flip, and the augmentation ratio was 50. Training was performed with 128 × 128 × 3 pixel input data and 30 epochs with early stopping. The weights of the original VGG16 were maintained until 15 layers and then tuned after 15 layers. After the full connection, the hidden fully connected layer had 256 nodes, the output layer was designed with two outputs, and a sigmoid activation function was used. Batch = 32 and Adam optimizer with a learning rate of 0.0001 were used as model parameters, and categorical_crossentropy was used for loss function. K-fold cross-validation (K = 5) was applied to the data, and an interim output label was predicted each time. The trial was repeated five times, and the final output label was calculated from the five predicted interim outputs by voting (threshold = 2.5) (Supplemental Figure 2). For preprocessing, the generated input data were converted to tiff images from the DICOM image with WW/WL= (350, 50), and Min = 0/Max = 255 normalization was applied.

Evaluation by radiologists with and without the DL model

Three radiologists (A.K., T.M., and Y.S. with 1, 6, and 12 years of radiology experience, hereafter referred to as R1, R2, and R3, respectively) independently diagnosed thymomas from the non-enhanced and enhanced CT images without the DL model. One month later, they re-diagnosed them with reference to the DL model. Radiologists interpreted the entire chest, including the absence or presence of pleural dissemination. If necessary, the window level or width was freely changed to review the CT images. The following information, summarized from previous reports on CT imaging features of thymoma, was presented before

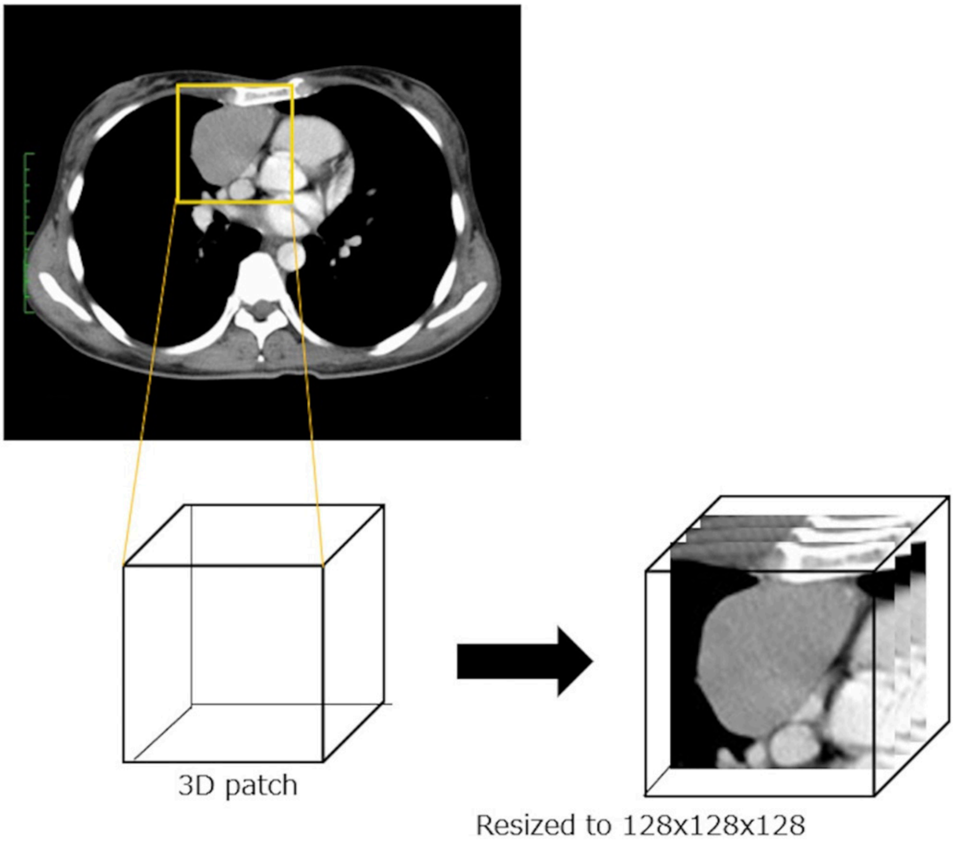

Figure 1. Input data generation. The dataset consisted of three axial slices, including the maximum tumor size from the CT volume data, and was resized to a 128 × 128 × 128 cube. CT, computed tomography.

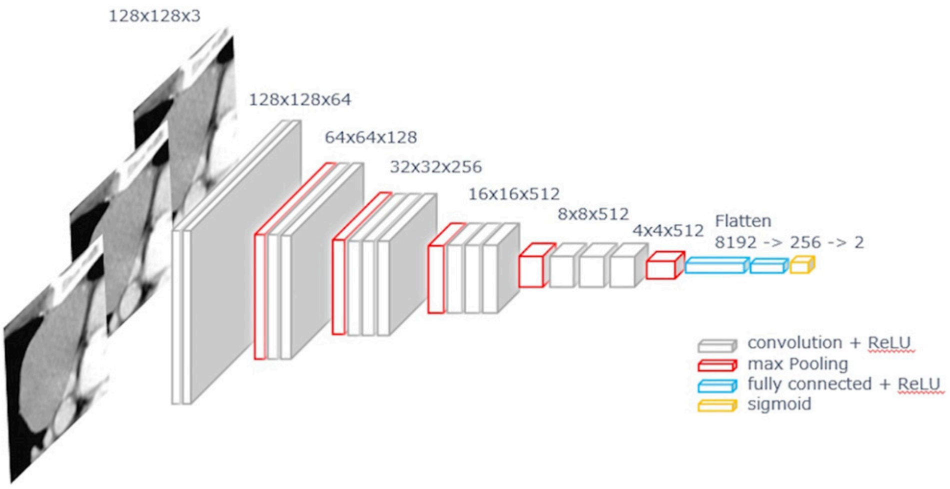

Figure 2. Convolutional neural network model. The center slice and pre- and post-slices from CT images of thymomas in the axial direction were used as three channel inputs, and training was performed using a 2-dimensional convolutional neural network, which was a fine-tuning model based on VGG16. CT, computed tomography.

interpretation: Type A is more likely to be spherical with smooth margins.[6] High-risk thymomas are more likely than low-risk thymomas to have a lobulated or irregular morphology.[11,12] Calcification is more common in types B1, B2, and B3 than in type AB.[6] The incidence of necrotic or cystic changes, capsular destruction, and pericapsular invasion increases with the increase in degree of malignancy. High-risk thymomas are larger in size than low-risk thymomas (>50% of type B3 tumors are greater than 10 cm).[13]

Statistical analysis

Data were analyzed using commercially available software, JMP (version 16; SAS Institute, Cary, NC, USA) and MedCalc (version 20; MedCalc Software Inc. Mariakerke, Belgium). The performance of each radiologist and the DL model was evaluated based on four indices: the area under the receiver operating characteristic curve (AUC), accuracy, specificity, and sensitivity. The McNemar test with Bonferroni correction was used to compare the accuracy, sensitivity, and specificity between performances of the DL model and the radiologists with and without the DL model. A p-value of <0.0167 (0.05/3) was considered statistically significant. Pairwise comparison was performed for AUC between performance of the DL model and each radiologist and between each radiologist with and without the DL model using the receiver operating characteristic (ROC) analysis. The AUC by ROC analysis was statistically analyzed using DeLong's test to compare the performance between the DL model and each radiologist and between each radiologist with and without the DL model. In this study, we defined sensitivity and specificity as follows:

- True Positive (TP): Correctly identifying a high-risk thymoma case.
- True Negative (TN): Correctly identifying a low-risk thymoma case.
- False Positive (FP): Incorrectly classifying a low-risk case as high-risk.
- False Negative (FN): Incorrectly classifying a high-risk case as low-risk.

Based on these definitions, sensitivity was calculated as TP / (TP + FN), and specificity was calculated as TN / (TN + FP).

Results

Patient characteristics

A total of 159 patients with 160 tumors (55 men and 104 women; mean age, 56.0 years [range, 27–83 years]) were included. There were 92 low-risk thymomas and 68 high-risk thymomas (type A: 5, type AB: 38, type B1: 49, type B2: 45, and type B3: 23). Only one patient had two thymomas (low-risk, type B1; and high-risk, type B2). All the others had one thymoma per patient. All the thymomas were located in the anterior mediastinum. The tumor sizes ranged from 12 mm to 125 mm, with a mean of 47.0 mm.

Performance of the DL model and radiologists without DL model assistance

The performance of the DL model and each radiologist without the DL model is listed in Table 1. In eight cases, all the radiologists answered incorrectly, and only the DL model answered correctly: low-risk thymoma in five cases

Table 1. Comparison of the diagnostic performance of each radiologist and the DL model.

(Radiology experience)	AUC	p	Accuracy (%)	p	Sensitivity (%)	p	Specificity (%)	p
R1 (1 year)	0.62	0.08	61.9	0.04	58.8	0.52	64.1	<0.001*
R2 (6 years)	0.68	0.67	69.4	0.61	57.3	0.61	78.3	0.167
R3 (12 years)	0.65	0.34	67.5	0.38	48.5	0.7	81.5	0.503
DL	0.69		71.9		53.0		85.0	

AUC was calculated using ROC (receiver operating characteristic) analysis.
*Significantly different from McNemar's test with Bonferroni correction.
Abbreviations: DL, deep learning; AUC, area under the receiver operating characteristic curve.

and high-risk thymoma in three cases (Figure 3). The specificity of the DL model was higher than that of all the radiologists; in particular, significantly higher than that of R1 ($p < .001$). The accuracy and AUC of performance of the DL model were higher than those of all the radiologists but were not significantly different. The performance of the DL model was comparable to that of the radiologists. The results of the DL model and radiologists are shown according to WHO subtype in Figure 4. The DL model performed better in types A and AB than the radiologists, giving the correct outputs in almost all cases. All the cases identified by the DL model as high-risk were type B thymomas (type B1, 12; type B2, 20; and type B3, 16), except for one case of type AB.

Performance of radiologists with the DL model assistance

The performance of each radiologist with and without the DL model is shown in Table 2. Only the specificity of R1 differed significantly with and without the DL model ($p = .002$). The DL model increased the AUCs of performances of R1 and R3, but not significantly. The DL model also increased the accuracy of two of the three radiologists, but the difference was not significant. For R1 and R2, the DL model increased the number of correct answers for types A and AB (Figure 4).

Discussion

We developed a CT-based DL model for risk classification of thymoma. Although the DL model performed risk classification based on only the primary lesions of thymomas, its diagnostic performance was comparable to that of radiologists who interpreted the entire chest CT imaging, including pulmonary dissemination, for risk classification. The AUC of performance of the DL model was higher than that of all the chest radiologists, but no significant difference. Accuracy of the DL-model was better than that of the radiologists, but not significantly different, especially in the first year. Specificity of diagnosis by R1 and the DL-model was significantly different. There was no significant difference in the diagnostic performance of the radiologists with and without the DL model. However, for two of the three radiologists, assistance of the DL model increased the accuracy and AUC. Our results suggest that DL model support may enhance radiologist's performance; in particular, the high specificity of the DL model may be useful for reducing unnecessary examinations. Further studies using a larger cohort are required to confirm our results.

Predicting the risk classification of thymoma from CT images can evaluate the prognosis of patients, especially those with poor performance status who were not fit for surgery. This may influence treatment planning in operable patients. For instance, if a low-risk thymoma with a favorable prognosis is expected preoperatively, the likelihood of invasion of adjacent structures or pleural dissemination is evaluated to be low. Therefore, robot-assisted surgery or thoracoscopic thymectomy may be an option because they are less invasive than thoracotomy. In contrast, surgeons must operate more carefully than in patients with a low-risk thymoma to avoid iatrogenic dissemination, when a high-risk thymoma is speculated. Compared to those with low-risk thymomas, high-risk thymomas are more likely to be unresectable, which may necessitate the use of neoadjuvant chemotherapy.[14]

There are several reports on CT findings associated with risk classification based on the WHO classification. For example, high-risk thymomas tend to have a more irregular morphology and invade the surrounding tissues more than low-risk thymomas.[11–13] However, when the radiologists in this study performed risk classification of thymomas based on the results of previous studies, the correct answer was only 61.9% for the first-year radiologist. In cases with findings suggestive of both low- and high-risk thymomas, different radiologists would make different decisions. Such cases can be difficult to classify, particularly for less experienced radiologists. Risk classification by radiologists based on CT findings alone has limitations in terms of reproducibility. After referring to the DL model results, the percentage of correct answers was 68.1%–70.0%, with less variability and better reproducibility. The use of the DL model allows risk classification with a certain degree of accuracy, regardless of the experience of the radiologist.

The DL model has the advantage of evaluating features that cannot be confirmed visually. For type A and AB tumors, the

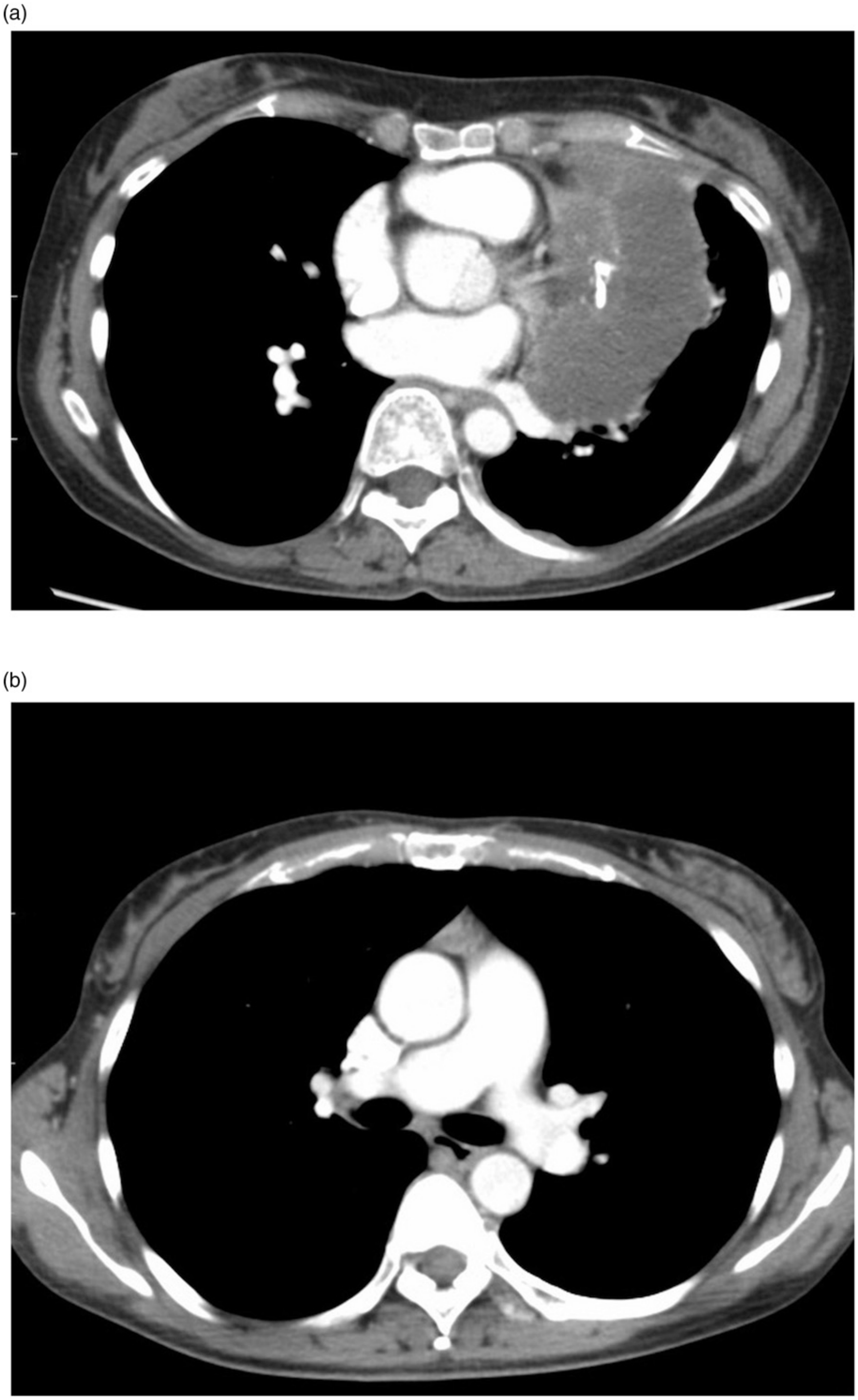

Figure 3. (a) A 47-year-old woman with a low-risk thymoma (type B1). The DL model correctly classified it as a low-risk thymoma, although all three radiologists classified it as high-risk thymoma. (b) A 46-year-old woman with a high-risk thymoma (type B2). The DL model correctly classified it as a high-risk thymoma, although all three radiologists classified it as low-risk thymoma.

DL model correctly classified the risk for almost all tumors. The reason for the accuracy of the DL model for only types A and AB is not clear because the basis for diagnosis by the DL model is unknown. We speculated that it reflected the fact that there was a 15% overlap in diagnosing types B1 and B2 by pathologists, since the constituent cells of types A and B are different, and type B is diagnosed by the number of lymphocytes.[15] A previous study reported that the maximum

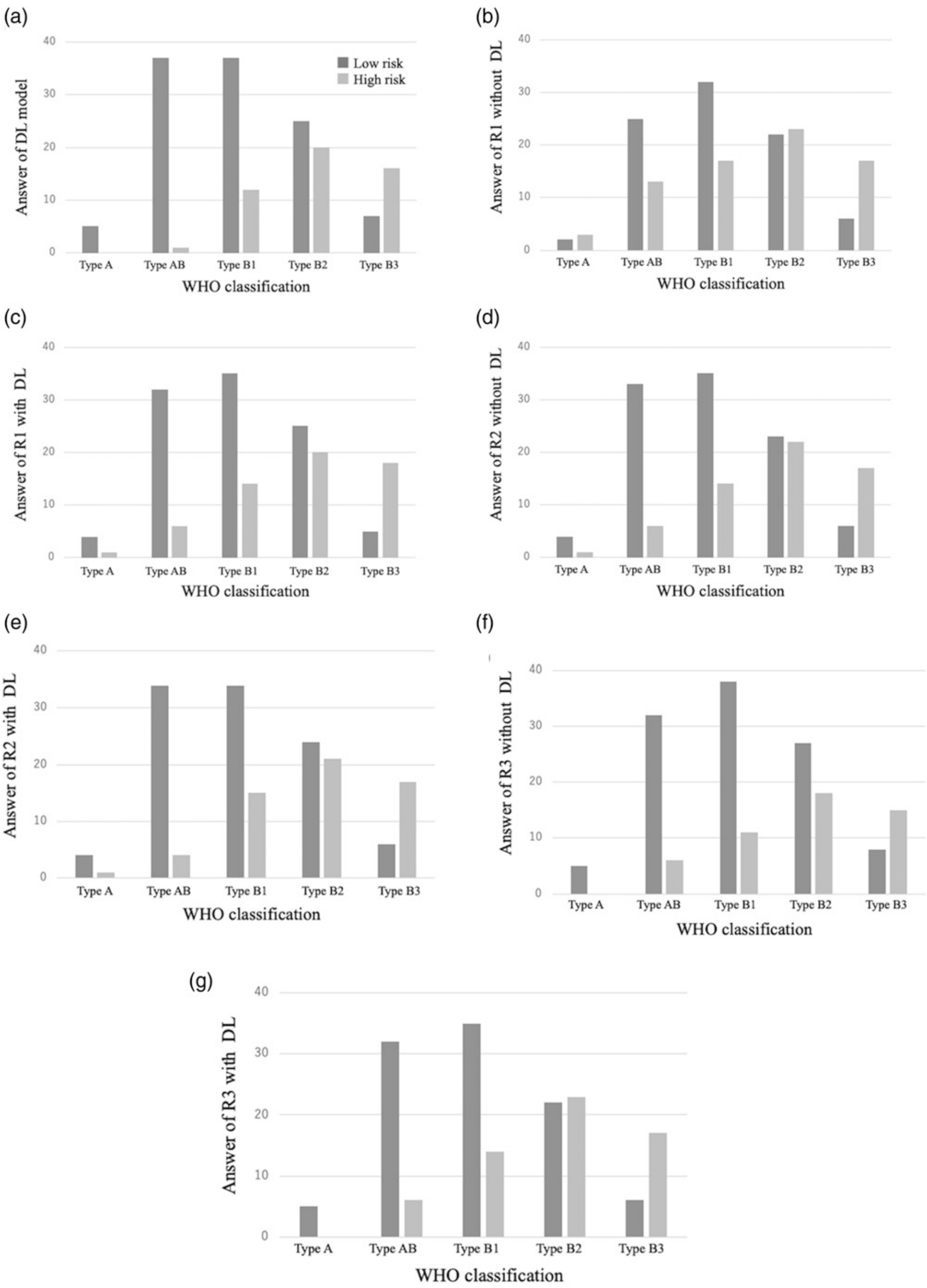

Figure 4. Graphs of the answers of radiologists and the DL model for each WHO subtype. (a) Answers of the DL model for each WHO subtype. (b) Answers of R1 without the DL model for each WHO subtype. (c) Answers of R1 with the DL model for each WHO subtype. (d) Answers of R2 without the DL model for each WHO subtype. (e) Answers of R2 with the DL model for each WHO subtype. (f) Answers of R3 without the DL model for each WHO subtype. (g) Answers of R3 with the DL model for each WHO subtype.

difference between non-enhanced and enhanced CT attenuation and the spectral/perfusion parameter values of dual-energy CT of type B1 thymoma was closer to those of types B2 and B3 than those of types A and AB, which might be due to differences in the cells that constitute the tumor.[16,17] To improve the accuracy of the DL model, it is necessary to improve the diagnostic performance for type B thymoma and accumulate more cases.

This study had several limitations. First, the number of patients was small. A cross-validation was performed to compensate for this shortcoming. Overfitting, a considerable issue in DL, was exacerbated when training instances were

Table 2. Comparison of the diagnostic performance of each radiologist with and without the DL model. AUC was calculated using ROC (receiver operating characteristic) analysis.

(Radiology experience)	DL	AUC	*p*	Accuracy (%)	*p*	Sensitivity (%)	*p*	Specificity (%)	*p*
R1 (1 year)	(−)	0.62	0.08	61.9	0.05	58.8	0.72	64.1	0.002*
	(+)	0.66		68.1		55.9		77.2	
R2 (6 years)	(−)	0.68	0.49	69.4	1.00	57.3	1.00	78.3	1.00
	(+)	0.67		68.8		55.9		78.3	
R3 (12 years)	(−)	0.65	0.31	67.5	0.57	48.5	0.14	81.5	0.55
	(+)	0.69		70.0		58.8		78.3	
DL		0.69		71.9		53.0		85.0	

*Significantly different from McNemar's test with Bonferroni correction.
Abbreviations: DL, deep learning; AUC, area under the receiver operating characteristic curve.

limited. With a larger number of cases, it may be possible to create an algorithm with even higher accuracy. Second, all the CTs were performed at the same facility. If the contrast protocol changes, the attenuation of thymomas will also change and the results may vary. Further studies are needed to determine the best protocol for the WHO classification of thymoma. Third, the slice thickness of the CT images used in this study was 5 mm. Thinner slice has a higher resolution, and it is possible that a DL model with a higher diagnostic performance could have been constructed. However, because thin-slice CT was not available for all cases, CT images with 5 mm-thick slices were used in this study. Fourth, we used only 2D data. In this study, only the three axial slices, including the maximum cleavage plane, were used for the DL model. Therefore, it was not possible to evaluate the morphology in the coronal or sagittal directions. A previous study in which 3D morphometric analysis using sphericity and ellipticity was performed reported that sphericity and ellipticity are useful and objective indices for the risk classification of thymoma.[18] DL models using 3D data are expected to further improve the accuracy of thymoma risk classification. Fifth, determining the method the DL model used to reach its conclusions is difficult. R2 changed their answer in only one case after referring to the DL results in our study. We speculate that R2 may not have given significant weight to the DL results when the readers' own impressions differed from the DL output. This observation may stem from the inherent "black box" nature of current DL systems, which often lack transparency in their decision-making processes. However, several techniques have been suggested to explain the behavior of DL algorithms, including gradient-weighted class activation mapping (Grad-CAM).[19] A detailed explanation of the algorithm's decision may help to detect new clinically useful imaging findings. Furthermore, it could provide insights into the reasoning behind DL outputs, potentially enhancing radiologists' understanding and confidence in DL-assisted diagnoses, particularly in cases of disagreement between human and DL interpretations. Further analysis is needed to elucidate the black box of diagnostic process of the DL model in the future. Sixth, our study focused exclusively on thymomas due to sample size constraints. However, we recognize that anterior mediastinal tumors include not only thymomas but also malignant lymphomas, germ cell tumors, and thymic carcinomas. Accurate differentiation among these neoplasms on preoperative CT remains challenging. In future research, we aim to expand our dataset to include a wider range of anterior mediastinal tumors. With larger and more diverse datasets, we anticipate that DL could be developed to differentiate between various anterior mediastinal neoplasms, enhancing the clinical utility of DL-assisted preoperative diagnosis. Seventh, while our model showed promising results within our single-center cohort, its generalizability to diverse populations and settings remains unestablished. This limits our ability to assess the performance of the model across varying imaging protocols and institutional practices. Eighth, the DL model was trained on only three CT slices per thymoma, including the slice with maximum tumor size, while radiologists interpreted the entire chest CT. This difference in available information may influence the comparative performance assessment. The limited slice for the DL model might not capture all the complexities of various thymomas that radiologists can interpret in a full CT series, potentially affecting its diagnostic ability. Future research should need using more comprehensive CT data for the DL model to enable a more equitable comparison with radiologist performance.

In conclusion, we developed the DL model for risk classification of thymoma from CT images based on the WHO classification, and the DL model performed as well as the radiologists.

Acknowledgments

We thank Ms. Akie Katsuki and Mr. Kazuyuki Ohmura (GE HealthCare Japan/Research and Development) for their technical supports including a construction of the deep learning model.

Declaration of conflicting interests

The author(s) declared no potential conflicts of interest with respect to the research, authorship, and/or publication of this article.

Funding

The author(s) received no financial support for the research, authorship, and/or publication of this article.

ORCID iD

Masahiro Yanagawa https://orcid.org/0000-0002-0911-6769

Data availability statement

The datasets used and/or analyzed during the current study are available from the corresponding author on reasonable request.

Supplemental Material

Supplemental material for this article is available online.

References

1. Okumura M, Ohta M, Tateyama 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: 624–632.
2. Chen G, Marx A, Chen WH, et al. New WHO histologic classification predicts prognosis of thymic epithelial tumors: a clinicopathologic study of 200 thymoma cases from China. Cancer 2002; 95: 420–429.
3. Ströbel P, Bauer A, Puppe B, et al. Tumor recurrence and survival in patients treated for thymomas and thymic squamous cell carcinomas: a retrospective analysis. J Clin Oncol 2004; 22: 1501–1509.
4. Han J, Lee KS, Yi CA, et al. Thymic epithelial tumors classified according to a newly established WHO scheme: CT and MR findings. Korean J Radiol 2003; 4: 46–53.
5. Sadohara J, Fujimoto K, Müller NL, et al. Thymic epithelial tumors: comparison of CT and MR imaging findings of low-risk thymomas, high-risk thymomas, and thymic carcinomas. Eur J Radiol 2006; 60: 70–79.
6. Blinded for anonymity’.
7. Chassagnon G, Vakalopolou M, Paragios N, et al. Deep learning: definition and perspectives for thoracic imaging. Eur Radiol 2020; 30: 2021–2030.
8. Wang S, Shi J, Ye Z, et al. Predicting EGFR mutation status in lung adenocarcinoma on computed tomography image using deep learning. Eur Respir J 2019; 53: 1800986.
9. Walsh SLF, Calandriello L, Silva M, et al. Deep learning for classifying fibrotic lung disease on high-resolution computed tomography: a case-cohort study. Lancet Respir Med 2018; 6: 837–845.
10. Blinded for anonymity’.
11. Ozawa Y, Hara M, Shimohira M, et al. Associations between computed tomography features of thymomas and their pathological classification. Acta Radiol 2016; 57: 1318–1325.
12. Jeong YJ, Lee KS, Kim J, et al. Does CT of thymic epithelial tumors enable us to differentiate histologic subtypes and predict prognosis? AJR Am J Roentgenol 2004; 183: 283–289.
13. Liu GB, Qu YJ, Liao MY, et al. Relationship between computed tomography manifestations of thymic epithelial tumors and the WHO pathological classification. Asian Pac J Cancer Prev APJCP 2012; 13: 5581–5585.
14. 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.
15. Marx A, Ströbel P, Badve SS, et al. ITMIG consensus statement on the use of the WHO histological classification of thymoma and thymic carcinoma: refined definitions, histological criteria, and reporting. J Thorac Oncol 2014; 9: 596–611.
16. Yu C, Li T, Zhang R, et al. Dual-energy CT perfusion imaging for differentiating WHO subtypes of thymic epithelial tumors. Sci Rep 2020; 10: 5511.
17. Hu YC, Wu L, Yan LF, et al. Predicting subtypes of thymic epithelial tumors using CT: new perspective based on a comprehensive analysis of 216 patients. Sci Rep 2014; 4: 6984.
18. Yamazaki M, Oyanagi K, Umezu H, et al. Quantitative 3D shape analysis of CT images of thymoma: a comparison with histological types. AJR Am J Roentgenol 2020; 214: 341–347.
19. Selvaraju RR, Cogswell M, Das A, et al. Grad-CAM: visual explanations from deep networks via gradient-based localization. Int J Comput Vis 2020; 128: 336–359.