



Development of a deep learning model for classifying thymoma as Masaoka-Koga stage I or II via preoperative CT images

Lei Yang^{1#}, Wenjia Cai^{2#}, Xiaoyu Yang^{3#}, Haoshuai Zhu¹, Zhenguo Liu¹, Xi Wu³, Yiyan Lei¹, Jianyong Zou¹, Bo Zeng¹, Xi Tian⁴, Rongguo Zhang⁴, Honghe Luo¹, Ying Zhu³

¹Department of Thoracic Surgery, the First Affiliated Hospital of Sun Yat-sen University, Guangzhou, Guangdong 510080, China; ²State Key Laboratory of Ophthalmology, Zhongshan Ophthalmic Center, Sun Yat-sen University, Guangzhou 510060, China; ³Department of Radiology, the First Affiliated Hospital of Sun Yat-sen University, Guangzhou 510080, China; ⁴Advanced Institute, Infervision, Beijing 100000, China

Contributions: (I) Conception and design: L Yang, Y Zhu, W Cai, H Luo; (II) Administrative support: None; (III) Provision of study materials or patients: Y Zhu, H Luo; (IV) Collection and assembly of data: Y Zhu, W Cai, L Yang; (V) Data analysis and interpretation: Y Zhu, W Cai, L Yang, H Luo; (VI) Manuscript writing: All authors; (VII) Final approval of manuscript: All authors.

[#]These authors contributed equally to this work as co-first authors.

Correspondence to: Ying Zhu. Department of Radiology, the First Affiliated Hospital of Sun Yat-sen University, Guangzhou 510080, China. Email: zhuying10246@hotmail.com; Honghe Luo. Department of Thoracic Surgery, the First Affiliated Hospital of Sun Yat-sen University, Guangzhou 510080, China. Email: luohhzm@163.com.

Background: Accurate thymoma staging via computed tomography (CT) images is difficult even for experienced thoracic doctors. Here we developed a preoperative staging tool differentiating Masaoka-Koga (MK) stage I patients from stage II patients using CT images.

Methods: CT images of 174 thymoma patients were retrospectively selected. Two chest radiologists independently assessed the images. Variables with statistical differences in univariate analysis were adjusted for age, sex, and smoking history in multivariate logical regression to determine independent predictors of the thymoma stage. We established a deep learning (DL) 3D-DenseNet model to distinguish the MK stage I and stage II thymomas. Furthermore, we compared two different methods to label the regions of interest (ROI) in CT images.

Results: In routine CT images, there were statistical differences ($P < 0.05$) in contour, necrosis, cystic components, and the degree of enhancement between stage I and II disease. Multivariate logical regression showed that only the degree of enhancement was an independent predictor of the thymoma stage. The area under the receiver operating characteristic curve (AUC) of routine CT images for classifying thymoma as MK stage I or II was low (AUC = 0.639). The AUC of the 3D-DenseNet model showed better performance with a higher AUC (0.773). ROIs outlined by segmentation labels performed better (AUC = 0.773) than those outlined by bounding box labels (AUC = 0.722).

Conclusions: Our DL 3D-DenseNet may aid thymoma stage classification, which may ultimately guide surgical treatment and improve outcomes. Compared with conventional methods, this approach provides improved staging accuracy. Moreover, ROIs labeled by segmentation is more recommendable when the sample size is limited.

Keywords: Thymoma; Masaoka-Koga stage (MK stage); artificial intelligence (AI) (computer vision systems); neural networks; X-ray computed tomography (X-ray CT)

Submitted Sep 22, 2019. Accepted for publication Jan 10, 2020.

doi: 10.21037/atm.2020.02.183

View this article at: <http://dx.doi.org/10.21037/atm.2020.02.183>

Introduction

Thymoma, originating from thymic epithelial cells, is a common anterior mediastinal tumor. The Masaoka-Koga (MK) staging system, based on the anatomic extent of tumor invasion, has been widely used to evaluate the thymoma stage (1-9). According to the MK classification criteria, the distinction between stage I and II thymoma is if the tumor has invaded the capsule. So stage I thymomas are noninvasive, whereas invasive for stage II tumors and above (10). Extracapsular involvement is proven to be an essential factor concerning disease recurrence and patient survival (11-13).

Complete thymectomy includes the excision of the lesion and the mediastinal fatty tissue between both phrenic nerves. This surgical method is a standard treatment for the disease at any stage (14-16). Thymomectomy, defined as the resection of the thymoma only, has been performed routinely for patients with stage I thymomas. Previous studies showed no difference in survival and recurrence between patients with stage I thymomas who received thymectomy or thymomectomy. In contrast, complete thymectomy was still the preferable treatment for stage II thymomas (17-20). However, recent reports indicated that thymectomy was unnecessary for the MK stage I tumor. Moreover, thymomectomy, the less invasive procedure, could be the more reasonable option for these patients. Because it was associated with shorter operative time, less blood loss, and shorter hospital stay (21-23). Thus, it is crucial to distinguish the MK stage I from stage II thymoma preoperatively. Enhanced computed tomography (CT) is a standard examination for preoperative staging by The International Thymic Malignancy Interest Group (ITMIG) (24). Marom *et al.* suggested that CT imaging features could distinguish between MK stage I/II and MK stage III/IV thymomas (25). However, it is challenging to identify extracapsular invasion on CT images by visual inspection. In most cases, postoperative histopathological examination of the surgical specimen was essential for staging (25,26). Most thoracic surgeons make these surgery decisions based on clinical experience, which are sometimes inaccurate and might lead to overtreatment.

Artificial intelligence (AI) has the potential to assist in making medical diagnoses and clinical decisions (27-29). Deep learning (DL), an AI method, is an approach that utilizes a convolutional neural network (CNN) to enable feature extraction from accurately labeled images and to generate classifications as output (30-32). DL extracts image information with high data throughput (33). The

densely connected convolutional network (DenseNet) (34,35) required less computational power, less model complexity, and yielded significant improvement over a conventional CNN. However, like any other CNN model, DenseNet had to learn each parameter from scratch, and a large volume of data was required. So for diseases with low incidences, transfer learning (TL) enabled the addressing of less information by using a feed-forward approach (36). TL transferred knowledge from a previous model to another when the latter one had only a limited amount of training data (37,38). TL proved to be a highly effective DL technique for model generalization. Nevertheless, the usefulness of DenseNet with TL for thymoma staging remained unclear.

In this study, we developed and validated a 3-dimensional (3D)-DenseNet DL model based on preoperative enhanced CT images for classifying thymoma as MK stage I or MK stage II. The technique may provide a more accurate assessment of tumor stage, which may ultimately facilitate the guidance for surgical treatment and improved clinical outcomes.

Methods

Patients

The records of 174 patients diagnosed with MK stage I or MK stage II thymoma by preoperative CT imaging were retrospectively reviewed and analyzed in this study. All patients received complete thymectomy or thymomectomy at the Department of Thoracic Surgery, the First Affiliated Hospital of Sun Yat-sen University from January 1, 2011, to June 31, 2018. The disease stage was based on the postoperative pathological examination of the surgical specimen. Patients with MK stage III or IV were excluded from this study. The Institutional Ethics committee of the First Affiliated Hospital of Sun Yat-sen University approved the study protocol.

A standardized data form was created to retrieve the clinical characteristics of the patients, including age, sex, smoking history, tumor size, World Health Organization (WHO) histological classification, surgical approach, and if the patient also had myasthenia gravis (MG).

Enhanced CT examination

Two radiologists with more than ten years of experience in chest imaging assessed the enhanced CT image

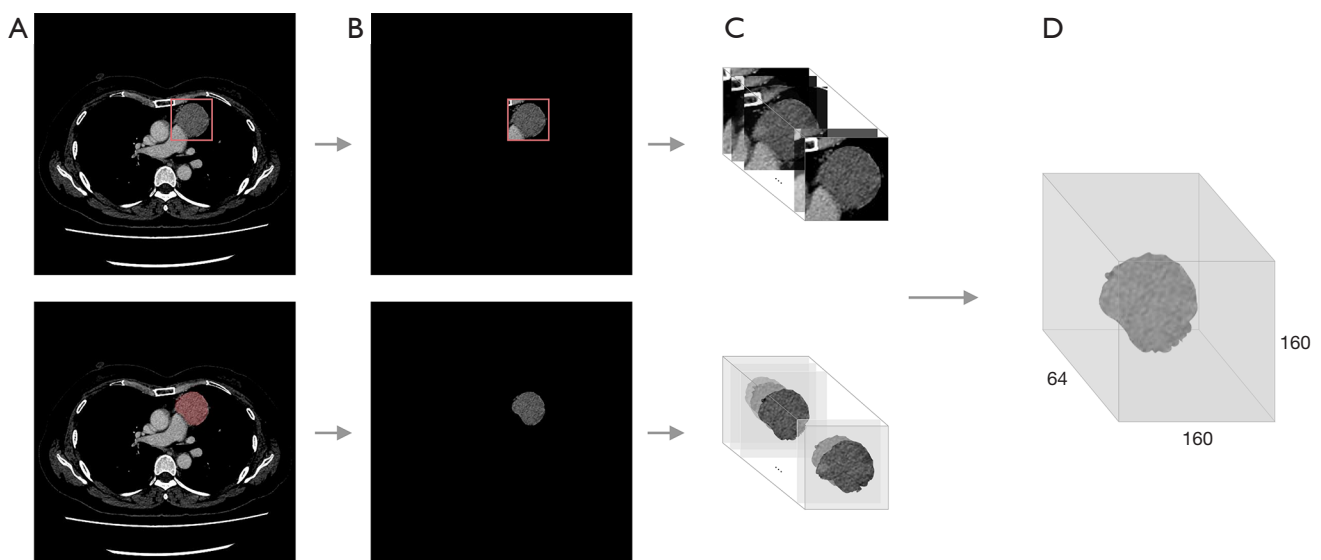


Figure 1 Preprocessing of raw CT images. (A) Datasets labeled with bounding box (first row) and segmentation (second row). (B) Extracted thymoma region. (C) 3D-reconstruction of all the CT slices contained the thymoma region. (D) The final input for the deep learning—a cube with a fixed size.

characteristics separately. Both of whom were blinded to the pathological examination results. Seven imaging characteristics were evaluated: shape (round, oval, lobulated or irregular), contour (smooth or irregular), necrosis/cystic component (indicating components without enhancement, and classified as 0–25%, 26–50%, 51–75%, 75–100% according to the volume percentage), degree of enhancement in Hounsfield units (HU) (indicating degree of enhancement of the solid components excluding necrotic/cystic components within the lesion), enhancement (homogeneous or heterogeneous enhancement of the lesion as a whole), the presence of calcifications, and the presence of an effusion (pleural/pericardial).

CT scanning was performed with a Toshiba Aquilion 64 spiral CT volume scanner. The tube voltage was 120 kV, the tube current was 200 mAs, and the slice thickness and the slice spacing were 1 mm. Iopromide (300 mgI/mL, Schering Pharmaceutical Ltd.) was used as the contrast agent, and 80–100 mL was injected at a flow rate of 3 mL/s.

Image preprocessing with two different methods

Raw CT images required preprocessing before being input into the deep neural network, and thus regions of the thymomas were extracted. CT images were labeled by two methods using labeling software ITK-SNAP 3.4.0.

As shown in *Figure 1*, the first row included a bounding box (the red rectangle) that located the thymoma region with surrounding tissue, while the second row labeled only the thymoma itself (the red-colored region). Both the red rectangle and red-colored region in *Figure 1A* indicate the desired region of interest (ROI). *Figure 1B* shows the final extracted thymoma region used for analysis.

The entire CT image data were analyzed by the two labeling methods described above. Then a 3D-reconstruction of the extracted thymoma region was established (*Figure 1C*; first and second row, respectively). Next, the extracted 3D image of the thymoma was randomly placed in a cube of a fixed size and used as input for the neural network. The size of the cube was 160×160×64 pixels, based on statistical data of the size of a common thymoma (*Figure 1D*). These steps allowed the neural network to process the data directly. None of the imaging technicians responsible for labeling was aware of the final pathological diagnosis.

Development of the DL 3D-DenseNet model

Since our dataset was relatively small, data augmentation was applied to expand the size of the training dataset and avoid overfitting. By introducing random factors, such as random cropping, more data with different directions were

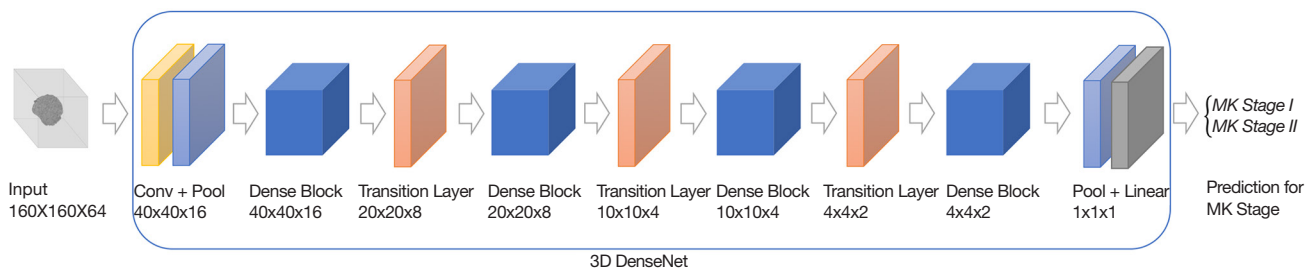


Figure 2 The development of the 3D-DenseNet model. The input is the 3D image of the extracted thymoma region, and the output is the prediction for MK stages. MK, Masaoka-Koga.

created. TL was then applied to speed up the learning process by making full use of the pre-trained parameters. In the 3D-DenseNet model, there were direct connections between any two layers in a feed-forward fashion. That is to say, each input layer of the network was the union of all previous output layers (Figure 2).

Cross-validation of the DL 3D-DenseNet model

We evaluated our model by 5-fold cross-validation because of the computational efficiency yielded by the 3D-DenseNet. Rather than randomly located, the tumor was placed in the center of the specified cube in the validation dataset. Batch normalization, rectified linear unit (ReLU), and softmax functions were applied to activate layers, and computed the probability of each sample. The loss function of our model was binary cross-entropy, which was optimized by stochastic gradient descent (SGD) with a mini-batch size of 16. Cosine annealing was used to schedule our learning rate for a total of 720 steps by setting the initial learning rate to $1.2e^{-5}$, and the minimum learning rate to $1e^{-7}$.

The dataset was randomly divided into five groups without overlapping, and each group had the same proportion of the MK stage I and II data. During the process of 1-fold, we split the data into a validation cohort and a training cohort. Finally, the evaluation result was the average of the 5-fold validations.

Implementation and evaluation metrics

The DL model was implemented under the MXNet (version 1.2.0, Apache Software Foundation, Forest Hill, MD USA) framework, using Python programming language (version 2.7.12). We trained our model on four NVIDIA GeForce GTX 1080 GPUs (NVIDIA, Beijing, China). The performance of the DL model was evaluated based on four

metrics: the area under the receiver operating characteristic curve (AUC), accuracy (ACC), specificity (SP), and sensitivity (SN).

Statistical analyses

Statistical analyses were performed using SPSS version 22.0 software (IBM, USA). Variables were grouped based on the MK stage I or II. Categorical variables were compared using the chi-square test. Continuous variables were compared using the *t*-test or Mann-Whitney U test for variables with a non-normal distribution. Variables with statistical differences ($P < 0.05$) in univariate analysis were adjusted for age, sex, and smoking history, and input into multivariate logical regression to determine independent predictors of the thymoma stage. Values of $P < 0.05$ were considered statistically significant. The AUC was calculated for evaluating the accuracy of models, and we regarded an $AUC \geq 0.7$ as a good predictive performance.

Results

Clinical characteristics of patients

Of the 174 patients included in the study, 48.3% (84/174) were MK stage I, and 51.7% (90/174) were MK stage II. There was an apparent correlation between the MK stage and WHO histological classification. There were no significant differences between the two groups concerning age, sex, tumor size, smoking history, the presence of MG, operation time, and blood loss. Patient characteristics are summarized in Table 1.

Imaging characteristics

Imaging characteristics used in this study are listed in Table 1.

Table 1 General clinical and image characteristics of patients with thymoma (n=174)

Variables	MK stage		P
	I	II	
General clinical characteristics			
Age [†] (year)	50.35±12.53	48.82±12.70	0.427
Gender, n (%)			0.537
Male	52 (61.90)	58 (64.44)	
Female	32 (38.10)	32 (35.56)	
Tumor size (cm) [†]	5.07±2.58	5.42±2.72	0.376
Smoking history, n (%)			0.735
No	75 (89.29)	78 (86.67)	
Yes	9 (10.71)	12 (13.33)	
WHO histologic classification, n (%)			0.001*
A	15 (17.85)	7 (7.78)	
AB	22 (26.19)	14 (15.56)	
B1	11 (13.10)	10 (11.11)	
B2	32 (38.10)	37 (41.11)	
B3	4 (4.76)	16 (17.78)	
C	0 (0)	6 (6.67)	
Combine with MG, n (%)			0.611
No	43 (51.19)	49 (54.44)	
Yes	41 (48.81)	41 (45.56)	
Surgical approach, n (%)			0.000*
Thymoma resection	20 (23.81)	11 (12.22)	
Thymectomy	13 (15.48)	18 (20.00)	
Extended thymectomy	51 (60.71)	61 (67.78)	
Surgery under VATS, n (%)			0.000*
No	57 (67.86)	64 (71.11)	
Yes	27 (32.14)	26 (28.89)	
Operation time (min) [†]	118.45±48.53	121.14±48.18	0.714
Blood loss (mL) [†]	75.73±64.87	78.30±47.97	0.769
Image characteristics			
Shape, n (%)			0.209
Round or oval	45 (54.57)	42 (46.67)	
Lobulated	21 (25.00)	18 (20.00)	
Irregular	18 (21.43)	30 (33.33)	

Table 1 (continued)

Table 1 (continued)

Variables	MK stage		P
	I	II	
Contour, n (%)			0.045*
Smooth	79 (94.05)	77 (85.56)	
Irregular	5 (5.95)	13 (14.44)	
Necrosis/cystic component, n (%)			0.000*
0–25%	25 (29.76)	38 (42.22)	
26–50%	50 (59.52)	30 (33.33)	
51–75%	6 (7.14)	7 (7.78)	
75–100%	3 (3.57)	15 (16.67)	
Degree of enhancement (HU) [†]	37.12±22.49	27.88±19.66	0.004*
Enhancement, n (%)			0.383
Homogeneous	43 (51.19)	42 (46.67)	
Heterogeneous	41 (48.81)	48 (53.33)	
Calcification, n (%)			0.455
No	70 (83.33)	71 (78.89)	
Yes	14 (16.67)	19 (21.11)	
Effusion (pleural/pericardial), n (%)			0.407
No	83 (98.80)	86 (95.56)	
Yes	1 (1.2)	4 (4.44)	

[†], data are mean ± standard deviation; *, P<0.05 was considered as statistically significant. MK, Masaoka-Koga.

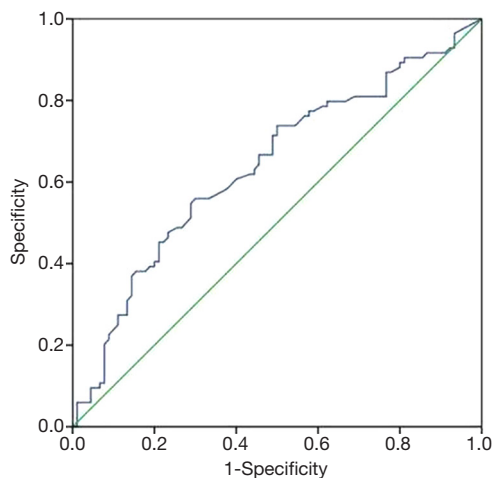


Figure 3 The ROC results of the Logistic regression of routine CT images in the prediction of Masaoka-Koga stage I or II (AUC =0.639, 95% CI: 0.556–0.721). AUC, area under the receiver operating characteristic curve.

There were statistical differences between the two groups in contour (smooth/ irregular), necrosis or cystic component, and degree of enhancement. There were no significant differences between the tumors from MK stage I and MK stage II patients in shape, enhancement (homogeneous/ heterogeneous), the presence of calcifications, and the presence of effusion (pleural/pericardial). Multivariate logical regression showed that only degree of enhancement was an independent predictor of the thymoma stage, with an AUC =0.639 (Figure 3).

MK stage classification by the 3D-DenseNet model

The 3D-DenseNet model was applied to predict the MK stage I or II for each image. The AUC for the training dataset using segmentation labels and bounding box labels was 0.966 and 0.951, respectively. The average AUC for the

Table 2 Mean cross-validation results of both training dataset and validation dataset in Masaoka-Koga staging across the two data forms

Label	Training dataset				Validation dataset			
	AUC	ACC	SN	SP	AUC	ACC	SN	SP
Segmentation	0.966	0.914	0.917	0.924	0.773	0.771	0.766	0.776
Bounding box	0.951	0.905	0.886	0.923	0.722	0.690	0.703	0.682

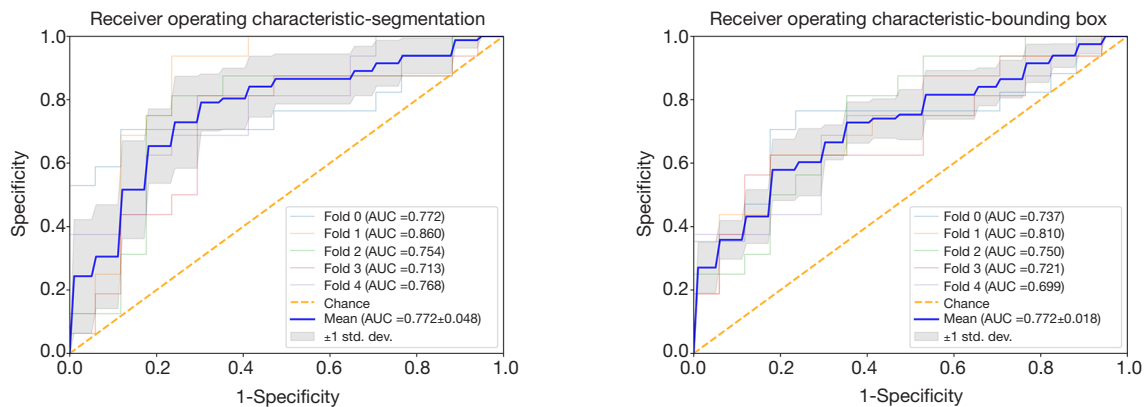


Figure 4 AUC for the 5-fold cross-validation dataset from segmentation (A) labels and bounding box (B) labels were 0.773 and 0.722, respectively. The ROC results showed that the proposed 3D-DenseNet model had achieved good performance in the prediction of Masaoka-Koga stage I or II. AUC, area under the receiver operating characteristic curve.

5-fold cross-validation dataset from the two labels was 0.773 and 0.722, respectively. The training dataset and cross-validation results are shown in *Table 2* and *Figure 4*.

Comparison of model performance with the two different data labeling methods

To further investigate whether the data labeling form affected the model performance, we compared the results of the two training datasets. As *Figure 1A,B* showed, ROI was extracted from segmentation labels (the second line) and bounding box labels (the first line). In the training dataset, both forms of data can fit well using the proposed DL method. In the validation dataset, the averaged results show that the segmentation labeled data outperforms the bounding box labeled data. The average AUC for the final validation dataset using segmentation labels and bounding box labels was 0.773 and 0.722, respectively ($P=0.017$). Also, the model performance was higher when using segmentation as compared with using the bounding box in ACC ($P=0.00141$) and SP ($P=0.0026$) (*Table 2* and *Figure 5*). These results indicated that segmentation yield higher accuracy than the bounding box in thymoma stage

classification.

Discussion

Currently, enhanced chest CT is the preferred preoperative examination for evaluating thymomas. We aimed to distinguish between the MK stage I and II thymomas via preoperative CT images, which might influence the selection of surgical procedures and clinical outcomes.

For routine CT image features, we found that the degree of enhancement and the presence of a necrosis/cystic component were significantly different between the two groups, demonstrating that CT image characteristics play a pivotal role in MK staging. Similar findings were reported in other studies (39,40). A recent study found correlations between preoperative CT imaging features and the biologic behavior of thymomas (41). However, our multivariate logical regression analysis showed that only the degree of enhancement was an independent predictor of the stage, with a relatively low AUC (AUC = 0.639). A retrospective study (42) indicated that routine CT imaging was not adequate for determining the MK stage. In the study, 437 patients with thymic epithelial tumors were included, in which 51% of stage III thymomas were

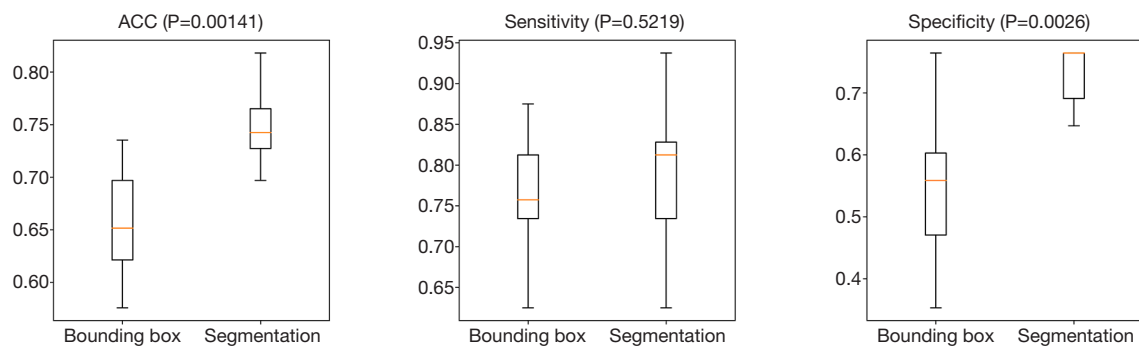


Figure 5 This figure illustrates the differences between the two groups on average accuracy, sensitivity, and specificity across a range of thresholds (we chose 0.48–0.51) of the five-fold cross-validation results. The result shows that the model trained with data in segmentation form outperforms model trained with data labeled with the bounding box. There were significant differences in ACC ($P=0.00141$) and SP ($P=0.0026$). ACC, accuracy; SP, specificity.

misclassified as stage I or II, and 37% stage I or II thymomas as stage III.

In order to improve the identification of stage I and stage II thymomas, we used a DenseNet model, which achieved a significant advantage over the state-of-the-art CNN (34). However, DenseNet was used mostly for 2-dimensional (2D) images; thus, we sought to improve the algorithm by developing a 3D-DenseNet model for 3D CT images of thymoma. Besides, we adopted a 5-fold cross-validation procedure to reduce variance caused by splitting data, preventing overfitting, and maximizing data utilization. The 3D-DenseNet model predicted MK stage I thymomas with a higher AUC (AUC =0.773) in the validation set, which indicated that DL algorithms greatly enhanced the ability to classify thymomas as MK stage I or II. The emerging technique of DL enables automated feature extraction, which has the advantage of evaluating features that cannot be observed by visual observation, and is not limited to evaluate the “interesting image features” only (30). This technique holds great promise for more accurate preoperative thymoma staging.

We also compared the results based on two different types of image labeling methods, which are most common, apart from mere classification. Bounding box labels could spatially constrain the objects in a fixed form for analysis (43,44). However, segmentation labels describe the contour of tumors more precisely and are more recommended when the training sets are limited. Our results showed that ROIs outlined by segmentation labels displayed more accurate performance than bounding box labels for predicting the MK stage. Our study is the first to explore model performance for predicting disease stage based on different data extraction forms. Based

on our findings, segmentation labels are preferable when the sample size is relatively small, with low disease incidence and will lead to a more reliable result.

Our study had its limitations. First, our study is a retrospective study from a single-center, which might cause selection bias. A large sample size and multicenter study is required to validate these results. Second, regarding the labeling methods, when the training set is relatively large, the segmentation method might take more time and effort. Thus bounding box labels might be more applicable for saving time on image processing. Third, patients’ long-term prognosis and clinical trials (45) need to be supplemented before applying this model to clinical practice in the future.

In summary, this is the first study to examine the DL approach for thymoma staging, and the results suggest that the method holds promise for more accurate preoperative staging of thymomas.

Conclusions

DL has a great potential for the preoperative staging of thymomas. Compared with visual observation, it dramatically improves the identification between the MK stage I and stage II thymomas. When the sample size of the training set is small, using segmentation labels for the ROIs results in better performance. The results of this study suggest that further studies of DL models for thymoma staging are warranted.

Acknowledgments

Funding: None.

Footnote

Conflicts of Interest: The 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 Institutional Ethics Committee of the First Affiliated Hospital of Sun Yat-sen University approved the study protocol.

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. Detterbeck FC, Nicholson AG, Kondo K, et al. The Masaoka-Koga stage classification for thymic malignancies: clarification and definition of terms. *J Thorac Oncol* 2011;6:S1710-6.
2. Filosso PL, Venuta F, Oliaro A, et al. Thymoma and inter-relationships between clinical variables: A multi-centre study in 537 patients. *Eur J Cardiothorac Surg* 2014;45:1020-7.
3. Ruffini E, Detterbeck F, Van Raemdonck D, et al. Tumours of the thymus: A cohort study of prognostic factors from the European Society of Thoracic Surgeons database. *Eur J Cardiothorac Surg* 2014;46:361-8.
4. Ruffini E, Filosso PL, Mossetti C, et al. Thymoma: inter-relationships among World Health Organization histology, Masaoka staging and myasthenia gravis and their independent prognostic significance: a single-centre experience. *Eur J Cardiothorac Surg* 2011;40:146-53.
5. Moon JW, Lee KS, Shin MH, et al. Thymic epithelial tumors: prognostic determinants among clinical, histopathologic, and computed tomography findings. *Ann Thorac Surg* 2015;99:462-70.
6. Liang G, Gu Z, Li Y, et al. Comparison of the Masaoka-Koga staging and the International Association for the Study of Lung Cancer/the International Thymic Malignancies Interest Group proposal for the TNM staging systems based on the Chinese Alliance for Research in Thymomas retrospective database. *J Thorac Dis* 2016;8:727-37.
7. Kim DJ, Yang WI, Choi SS, et al. Prognostic and clinical relevance of the World Health Organization schema for the classification of thymic epithelial tumors: A clinicopathologic study of 108 patients and literature review. *Chest* 2005;127:755-61.
8. Lee GD, Kim HR, Choi SH, et al. Prognostic stratification of thymic epithelial tumors based on both Masaoka-Koga stage and WHO classification systems. *J Thorac Dis* 2016;8:901-10.
9. Zhao Y, Shi J, Fan L, et al. Surgical treatment of thymoma: an 11-year experience with 761 patients. *Eur J Cardiothorac Surg* 2016;49:1144-9.
10. Jackson MW, Palma DA, Camidge DR, et al. The Impact of Postoperative Radiotherapy for Thymoma and Thymic Carcinoma. *J Thorac Oncol* 2017;12:734-44.
11. Roden AC, Yi ES, Jenkins SM, et al. Modified Masaoka Stage and Size Are Independent Prognostic Predictors in Thymoma and Modified Masaoka Stage Is Superior to Histopathologic Classifications. *J Thorac Oncol* 2015;10:691-700.
12. Detterbeck FC. Evaluation and treatment of stage I and II thymoma. *J Thorac Oncol* 2010;5:S318-22.
13. Bae MK, Lee SK, Kim HY, et al. Recurrence after thymoma resection according to the extent of the resection. *J Cardiothorac Surg* 2014;9:51-8.
14. National Comprehensive Cancer Network (NCCN) Guidelines. Available online: http://www.nccn.org/professionals/physician_gls/f_guidelines.asp
15. Burt BM, Yao X, Shrager J, et al. Determinants of complete resection of thymoma by minimally invasive and open thymectomy: analysis of an international registry. *J Thorac Oncol* 2017;12:129-36.
16. Mori T, Nomori H, Ikeda K, et al. Three cases of multiple thymoma with a review of the literature. *Jpn J Clin Oncol* 2007;37:146-9.
17. Onuki T, Ishikawa S, Iguchi K, et al. Limited thymectomy for stage I or II thymomas. *Lung Cancer* 2010;68:460-5.
18. Carillo C, Diso D, Mantovani S, et al. Multimodality treatment of stage II thymic tumours. *J Thorac Dis* 2017;9:2369-74.
19. Sakamoto M, Murakawa T, Konoeda C, et al. Survival after extended thymectomy for thymoma. *Eur J Cardiothorac Surg* 2012;41:623-7.
20. Gu Z, Fu J, Shen Y, et al. Thymectomy versus tumor

- resection for early-stage thymic malignancies: a Chinese Alliance for Research in Thymomas retrospective database analysis. *J Thorac Dis* 2016;8:680-6.
21. Nakagawa K, Yokoi K, Nakajima J, et al. Is Thymomectomy Alone Appropriate for Stage I (T1N0M0) Thymoma? Results of a propensity-score analysis. *Ann Thorac Surg* 2016;101:520-6.
 22. Tseng YC, Hsieh CC, Huang HY, et al. Is thymectomy necessary in non-myasthenic patients with early thymoma? *J Thorac Oncol* 2013;8:952-8.
 23. Sakamaki Y, Kido T, Yasukawa M. Alternative choices of total and partial thymectomy in video-assisted resection of noninvasive thymomas. *Surg Endosc* 2008;22:1272-7.
 24. Marom EM, Rosado-de-Christenson ML, Bruzzi JF, et al. Standard report terms for chest computed tomography reports of anterior mediastinal masses suspicious for thymoma. *J Thorac Oncol* 2011;6:S1717-23.
 25. Marom EM, Milito MA, Moran CA, et al. Computed tomography findings predicting invasiveness of thymoma. *J Thorac Oncol* 2011;6:1274-81.
 26. Priola AM, Priola SM, Di Franco M, et al. Computed tomography and thymoma: distinctive findings in invasive and noninvasive thymoma and predictive features of recurrence. *Radiol Med* 2010;115:1-21.
 27. Kermany DS, Goldbaum M, Cai W, et al. Identifying medical diagnoses and treatable diseases by image-based deep learning. *Cell* 2018;172:1122-31.e9.
 28. He J, Baxter SL, Xu J, et al. The practical implementation of artificial intelligence technologies in medicine. *Nat Med* 2019;25:30-6.
 29. Long EP, Lin HT, Liu ZZ, et al. An artificial intelligence platform for the multihospital collaborative management of congenital cataracts. *Nat Biomed Engineering* 2017;1:24.
 30. McBee MP, Awan OA, Colucci AT, et al. Deep learning in radiology. *Acad Radiol* 2018;25:1472-1480.
 31. Litjens G, Kooi T, Bejnordi BE, et al. A survey on deep learning in medical image analysis. *Med Image Anal* 2017;42:60-88.
 32. Halevy A, Norvig P, Pereira F. The unreasonable effectiveness of data. *IEEE Intelligent Systems* 2009;24:8-12.
 33. LeCun Y, Bengio Y, Hinton G. Deep learning. *Nature* 2015;521:436-44.
 34. Densely Connected Convolutional Networks. In *Proceedings of the 2017 IEEE Conference on Computer Vision and Pattern Recognition (CVPR)*, Honolulu, HI, USA, 21-26 July 2017;2261-9.
 35. Liang S, Zhang R, Liang D, et al. Multimodal 3D DenseNet for IDH Genotype Prediction in Gliomas. *Genes (Basel)* 2018. doi: 10.3390/genes9080382.
 36. LeCun Y, Bengio Y, Hinton G. Deep learning. *Nature* 2015;521:436-44.
 37. Pan SJ, Yang Q. A Survey on Transfer Learning. *IEEE Trans Knowl Data Eng* 2010;22:1345-59.
 38. Yosinski, J, Clune, J, Bengio Y, et al. How transferable are features in deep neural networks? *NIPS'14 Proceedings of the 27th International Conference on Neural Information Processing Systems* 2014;2:3320-8.
 39. 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.
 40. Ried M, Marx A, Götz A, et al. State of the art: diagnostic tools and innovative therapies for treatment of advanced thymoma and thymic carcinoma. *Eur J Cardiothorac Surg* 2016;49:1545-52.
 41. Ozawa Y, Hara M, Shimohira M, et al. Associations between computed tomography features of thymomas and their pathological classification. *Acta Radiol* 2016;57:1318-25.
 42. Moon JW, Lee KS, Shin MH, et al. Thymic epithelial tumors: prognostic determinants among clinical, histopathologic, and computed tomography findings. *Ann Thorac Surg* 2015;99:462-70.
 43. Rother C, Kolmogorov V, Blake A. Grabcut: Interactive foreground extraction using iterated graph cuts. *ACM Transactions on Graphics (TOG)* 2004;23:309-14.
 44. Lempitsky V, Kohli P, Rother C, et al. Image segmentation with a bounding box prior. in *Computer Vision, 2009 IEEE 12th International Conference on*. IEEE 2009;277-84.
 45. Lin H, Li R, Liu Z, et al. Diagnostic Efficacy and Therapeutic Decision-making Capacity of an Artificial Intelligence Platform for Childhood Cataracts in Eye Clinics: A Multicentre Randomized Controlled Trial. *EClinicalMedicine* 2019;9:52-9.

Cite this article as: Yang L, Cai W, Yang X, Zhu H, Liu Z, Wu X, Lei Y, Zou J, Zeng B, Tian X, Zhang R, Luo H, Zhu Y. Development of a deep learning model for classifying thymoma as Masaoka-Koga stage I or II via preoperative CT images. *Ann Transl Med* 2020;8(6):287. doi: 10.21037/atm.2020.02.183