



# Machine learning for predicting the prognosis of patients with thymoma and thymic carcinoma

Haijie Xu<sup>1,2#</sup>, Xirui Lin<sup>1,2#</sup>, Junhan Wu<sup>2,3#</sup>, Jianrong Chen<sup>1,2</sup>, Jiaying Wu<sup>1,2</sup>, Zheng Lin<sup>2</sup>, Xiaoming Cai<sup>1,2</sup>, Jiong Lin<sup>1,2</sup>, Peishen Li<sup>1,2</sup>, Chaoquan He<sup>1,2</sup>, Zefeng Xie<sup>1</sup>, Hansheng Wu<sup>1</sup>

<sup>1</sup>Department of Thoracic Surgery, The First Affiliated Hospital of Shantou University Medical College, Shantou, China; <sup>2</sup>Shantou University Medical College, Shantou, China; <sup>3</sup>Department of Thoracic Surgery, Guangdong Provincial People's Hospital (Guangdong Academy of Medical Sciences), Southern Medical University, Guangzhou, China

**Contributions:** (I) Conception and design: H Xu, H Wu; (II) Administrative support: Z Xie, H Wu; (III) Provision of study materials or patients: Z Xie, H Wu; (IV) Collection and assembly of data: H Xu, X Lin, Junhan Wu, J Chen, Jiaying Wu; (V) Data analysis and interpretation: H Xu, X Lin, Junhan Wu; (VI) Manuscript writing: All authors; (VII) Final approval of manuscript: All authors.

<sup>#</sup>These authors contributed equally to this work.

**Correspondence to:** Hansheng Wu, MD. Department of Thoracic Surgery, The First Affiliated Hospital of Shantou University Medical College, No. 57 Changping Road, Shantou 515041, China. Email: wu-han-sheng@163.com.

**Background:** Thymoma and thymic carcinoma are the most common tumors of the anterior mediastinum. However, there are little research on applying machine learning (ML) approaches to the prognostic prediction of thymoma and thymic carcinoma. The study aims to develop predictive models utilizing ML techniques to accurately forecast the 5-year survival of patients with thymoma and thymic carcinoma.

**Methods:** Patients with malignant thymic neoplasms were identified in the Surveillance, Epidemiology, and End Results (SEER) 17 database, and their demographic and clinicopathological characteristics were collected. ML classifiers, including elastic net regularized logistic regression, random forest (RF), non-linear support vector machine (SVM), extreme gradient boosting (XGBoost) machine, and categorical boosting (CatBoost) were trained. The hyper-parameter of the algorithms was optimized by a grid search with five repeats of 10-fold cross-validation. Ensemble models were built based on the three algorithms with the highest area under the receiver operator characteristic (ROC) curve (AUC) in the validation set. The best model among the single models and ensemble model was selected as the final model. Calibration curve and decision curve were adopted to evaluate the calibration performance and clinical utility. For comparison, we constructed a baseline model consisting of age and Masaoka stages using logistic regression.

**Results:** After data cleaning, 1,363 patients and 841 patients were included in the overall survival (OS) dataset and disease-specific survival (DSS) dataset, respectively. CatBoost [AUC: 0.755; 95% confidence interval (CI): 0.698–0.811] had the best performance in the OS prediction for the original dataset. The ensemble model achieved the highest prognostic efficiency for the original dataset, with an AUC of 0.833 (95% CI: 0.765–0.901). Calibration showed favorable goodness of fit and was further verified with the Hosmer-Lemeshow test (CatBoost:  $\chi^2=12.63$ ,  $P=0.13$ ; ensemble model:  $\chi^2=7.61$ ,  $P=0.47$ ). The decision curve showed that the final model provided a high net benefit. The model could significantly distinguish the prognosis of patients (all  $P$  values  $<0.001$ ). Finally, World Health Organization (WHO) histological classification, Masaoka stage, and age were the variables that significantly contributed to the models' prediction of OS and DSS.

**Conclusions:** We trained ML-based predictive models that could accurately predict the 5-year OS and DSS of patients with thymoma and thymic carcinoma.

<sup>^</sup> ORCID: 0000-0002-2692-1575.

**Keywords:** Thymoma; thymic carcinoma; machine learning (ML); prognostic prediction

Submitted Aug 05, 2024. Accepted for publication Dec 20, 2024. Published online Feb 20, 2025.

doi: 10.21037/jtd-24-1263

View this article at: <https://dx.doi.org/10.21037/jtd-24-1263>

## Introduction

Thymoma and thymic carcinoma are the most common tumors of the anterior mediastinum and account for 20% of mediastinal neoplasms (1,2), but their prognoses vary considerably (3,4). Previous research has identified several factors that can influence the prognosis of patients with thymoma and thymic carcinoma (4-9).

The Masaoka staging system has been widely accepted in clinical practice (9-11), with which thymic epithelial tumors can be categorized into stages I to IV according to the extent of local tumor invasion and the involvement of adjacent organs (11). The prognostic significance of the Masaoka staging system has been confirmed in several studies (10-12). Histological type is another significant prognostic factor for thymic epithelial tumors and is strongly correlated with Masaoka stage (4,13). Patients diagnosed with type A, AB, or B1 disease tend to present with stage I or II disease and are associated with a more favorable prognosis (13). In contrast, individuals with B2 or B3 thymoma or thymic carcinoma are more likely to

have advanced stage disease and are associated with higher rates of recurrence and mortality (13). In addition to these two factors, neoadjuvant therapy, surgical extent, lymph nodes metastasis and distant metastasis have also been demonstrated to be associated with the prognosis of patients with thymoma and thymic carcinoma (5,14-17). Despite so many valuable factors are found, it is challenging for clinical practitioners to comprehensively consider all the relevant factors without the help of certain tools. However, due to the rarity of thymic epithelial tumors, this area has garnered relatively little research attention, and the models that are currently available have deficiencies in diagnostic efficacy and goodness of fit (8).

Machine learning (ML) has become widely adopted in the analyses of medical data (18-20). Research has proven that these ML approaches exhibit superior predictive performance as compared to conventional statistical methods (21,22). This can be ascribed to ML's ability to model the high-dimensional and nonlinear relationships that are common in clinical information (18). Moreover, prior features selection is usually unnecessary in ML algorithms, allowing greater flexibility in model fitting. Although the interpretability of deep learning models reminds one of main drawbacks to this approach, significant advancements in recent years have been made in enhancing the interpretability of ML (23-25). In the context of thymoma and thymic carcinoma, ML methods may presumably offer robust performance in predicting outcomes. However, little research has been conducted to confirm this speculation.

Therefore, in this study, we assessed the prognostic performance of ML methods in thymoma and thymic carcinoma using the Surveillance, Epidemiology, and End Results (SEER) database. We included patients who had undergone surgery and trained single and ensemble models base on these data. We also compared the prognostic significance of complex ML models and the baseline models comprising only age and Masaoka classification. We present this article in accordance with the TRIPOD reporting checklist (available at <https://jtd.amegroups.com/article/>

### Highlight box

#### Key findings

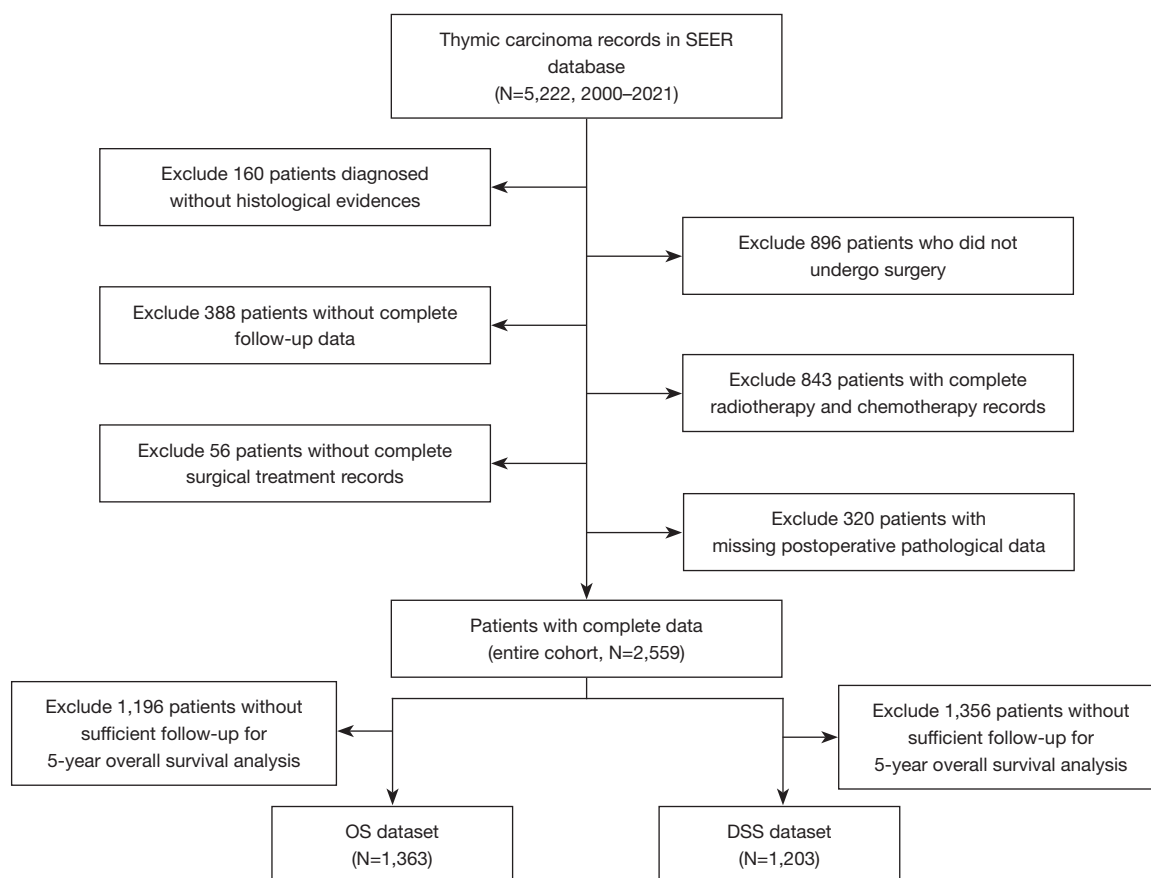
- We have developed predictive models for the 5-year overall survival and disease-specific survival of patients with thymoma and thymic carcinoma.

#### What is known and what is new?

- Previous research has identified several factors that influenced the prognosis of patients with thymoma and thymic carcinoma. However, it is challenging for clinical practitioners to comprehensively consider all the relevant factors without the help of tools.
- Our models demonstrate robust performance in predicting patient outcomes.

#### What is the implication, and what should change now?

- Machine learning tools have the potential to make the clinical judgement more swift, more precise, and more objective.
- Further study should pay attention to the exploration of multimodal data.



**Figure 1** Flow diagram of samples inclusion and exclusion. SEER, Surveillance, Epidemiology, and End Results; OS, overall survival; DSS, disease-specific survival.

view/10.21037/jtd-24-1263/rc).

## Methods

### Study population

Patients with malignant thymic neoplasms were identified in the SEER 17 database, which contains data collected from 2000 to 2021. International Classification of Oncology, Third Edition (ICO-O-3) histology codes of thymoma [8580-8585] and thymic carcinoma [8586], along with the malignant behavior code [3] were used to screen patients with malignant thymic neoplasms. A total of 5,222 patients were identified on the preliminary screen. Other inclusion criteria included patients (I) with definite pathological diagnosis; (II) with a history of completed surgery; (III) with complete follow-data; (IV) with complete radiotherapy and chemotherapy records; (V) with complete surgical treatments records; and (VI) with complete postoperative

pathological data. A flowchart of the patient screening process is shown in *Figure 1*. The demographic and clinicopathological characteristics of patients were collected, which included age, sex, race, chemotherapy, radiotherapy, tumor size, number of harvested lymph nodes, number of positive lymph nodes, histological type, distant metastasis, pulmonary metastasis, and previous tumor history. The Masaoka stage of patients were determined based on the primary tumor extension records in the SEER database. The staging scheme included the following categories: stage I–IIA (localized; confined to the gland of origin, not otherwise specified), stage IIB (regional; invasion to the adjacent connective tissue), and stage III–IV (distant; invasion to the adjacent organs/structures or pleural/pericardial implants and metastases). These classification criteria were used in a previous research as neither tumor-node-metastasis (TNM) nor Masaoka staging is directly available in the SEER database (26). The study was conducted in accordance with

the Declaration of Helsinki (as revised in 2013).

### **Data preprocessing**

Patients with missing data were excluded during the screening process. Therefore, only samples with intact data were incorporated into the model building and validation stages. To meet the requirements of ML algorithms, all the categorical data were converted to numeric format. One-hot-encoding was adopted to process all the categorical data. However, for improved interpretation of the models and comparison of data preprocessing, the original data were also preserved and used to build the model. Dataset were divided into a training set and validating set in a ratio of 7:3 ratio. To reduce the influence of different quantitative scales and alleviate data leakage, data were centered and scaled after division. Finally, the synthetic minority over-sampling technique, an improved algorithm based on oversampling, was performed to optimize the imbalanced distribution of negative and positive samples.

### **Model building and validation**

Five ML classifiers were trained in the study, including elastic net regularized logistic regression (ELR) (27), random forest (RF) (28), nonlinear support vector machine (SVM), extreme gradient boosting (XGBoost) (29), and categorical boosting (CatBoost) (30). A brief description of these algorithms can be found in Table S1. The hyperparameter of the algorithms was optimized via a grid search with five repeats of 10-fold cross-validation, and cross-entropy loss was set as the metric for optimization. The final hyper-parameters are shown in Table S2.

The discrimination ability of the models was evaluated via the area under the receiver operator characteristic (ROC) curve (AUC). The final AUC in the training set was obtained with 1,000 times of 0.632 bootstrapping. Ensemble models were built according to the prediction of the three algorithms that demonstrated the highest AUC in the internal validating set. The best model within the single models and ensemble model was selected as the final model. Calibration curve and decision curve were adopted to evaluate calibration performance and clinical utility.

To further validate the advantages of ML, we constructed a baseline model incorporating age and Masaoka stage via logistic regression. The building and evaluation process of the baseline model was consistent with that of the ML models.

### **Variable significance**

In the concise liner regression model, variable significance can be obtained easily via coefficients. However, the interpretability of ML models is a persistent issue in ML. In this study, we used the VarImp function in the “caret” package of R (The R Foundation for Statistical Computing, Vienna, Austria), which can calculate an AUC for each variable and access its contribution to the entire ROC curve (31).

### **Risk score**

The surv\_cutpoint function in the “survminer” package (<https://github.com/kassambara/survminer>) was used to determine the best classification thresholds for the predictive value in the training sets. This threshold was used to separate the patients into a high-risk group (predictive value over the threshold) and a low-risk group (predictive value under the threshold). Survival analysis was performed according to this grouping in the validation set and the entire cohort.

### **Statistical analysis**

Continuous variables are expressed as the median and interquartile range and were compared between groups using the Wilcoxon test. Categorical variables are expressed as the frequency (percentage) and were compared with Chi-squared test or Fisher exact test. The outcomes examined in this study included both 5-year overall survival (OS) and 5-year disease-specific survival (DSS). OS was defined as the interval between surgery and either the occurrence of death or the date of the last follow-up. DSS was defined as the interval between surgery procedure and the occurrence of death caused by malignant thymic tumors. Survival analysis was conducted with Kaplan-Meier curves and the log-rank test.

All tests were two-sided, with a P value <0.05 indicating statistical significance. All statistical analyzes were performed using R version 4.2.1.

## **Results**

### **Patient characteristics**

A total of 2,559 patients with complete data were screened from the raw data (Figure 1). Patients with a follow-up duration of less than 5 years and lacking survival outcomes were excluded from the study in order to meet the requirements for conducting 5-year OS and 5-year DSS analyzes. Finally, 1,363 patients and 1,203 patients were included in the OS dataset (Table 1) and DSS dataset

**Table 1** Demographic and clinical characters for patients in the overall survival dataset

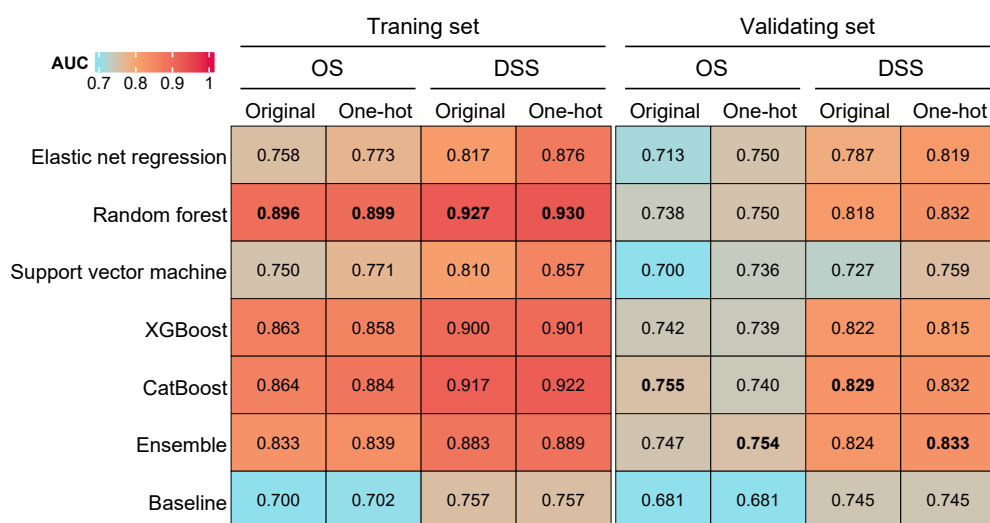
Variables	Overall (N=1,363)	Training set (n=954)	Validating set (n=409)	P
Age (years)	61 [11, 89]	61 [11, 89]	60 [17, 86]	0.20
Sex				0.92
Female	671 (49.2)	471 (49.4)	200 (48.9)	
Male	692 (50.8)	483 (50.6)	209 (51.1)	
History of other tumors				0.97
Yes	414 (30.4)	289 (30.3)	125 (30.6)	
No	949 (69.6)	665 (69.7)	284 (69.4)	
Race				0.90
White	938 (68.8)	659 (69.1)	279 (68.2)	
Black	185 (13.6)	130 (13.6)	55 (13.4)	
Other	240 (17.6)	165 (17.3)	75 (18.3)	
Tumor size, mm				0.31
≤6	969 (71.1)	670 (70.2)	299 (73.1)	
>6	394 (28.9)	284 (29.8)	110 (26.9)	
Masaoka stage				0.91
I–IIA	528 (38.7)	373 (39.1)	155 (37.9)	
IIB	646 (47.4)	449 (47.1)	197 (48.2)	
III–IV	189 (13.9)	132 (13.8)	57 (13.9)	
Chemotherapy				0.61
Yes	324 (23.8)	231 (24.2)	93 (22.7)	
No	1,039 (76.2)	723 (75.8)	316 (77.3)	
Radiotherapy				0.61
Yes	647 (47.5)	448 (47.0)	199 (48.7)	
No	716 (52.5)	506 (53.0)	210 (51.3)	
Surgery type				0.36
Radical/total resection	807 (59.2)	558 (58.5)	249 (60.9)	
Local/partial excision	517 (37.9)	365 (38.3)	152 (37.2)	
Debulking	39 (2.9)	31 (3.2)	8 (2.0)	
WHO classification				0.64
Type A	96 (7.0)	70 (7.3)	26 (6.4)	
Type AB	249 (18.3)	182 (19.1)	67 (16.4)	
Type B1	147 (10.8)	105 (11.0)	42 (10.3)	
Type B2	220 (16.1)	147 (15.4)	73 (17.8)	
Type B3	213 (15.6)	141 (14.8)	72 (17.6)	
Thymic carcinoma	222 (16.3)	157 (16.5)	65 (15.9)	
NOS	216 (15.8)	152 (15.9)	64 (15.6)	

**Table 1** (continued)

Table 1 (continued)

Variables	Overall (N=1,363)	Training set (n=954)	Validating set (n=409)	P
Number of harvested lymph nodes				0.49
≤5	419 (30.7)	284 (29.8)	135 (33.0)	
>5	166 (12.2)	118 (12.4)	48 (11.7)	
No node dissection performed	778 (57.1)	552 (57.9)	226 (55.3)	
Lymph node invasion				0.60
Negative	526 (38.6)	363 (38.1)	163 (39.9)	
Positive	59 (4.3)	39 (4.1)	20 (4.9)	
No node dissection performed	778 (57.1)	552 (57.9)	226 (55.3)	
Lung metastasis				0.54
Yes	54 (4.0)	41 (4.3)	13 (3.2)	
No	919 (67.4)	637 (66.8)	282 (68.9)	
Unknown	390 (28.6)	276 (28.9)	114 (27.9)	

Data were presented as n (%) or median [IQR]. WHO, World Health Organization; NOS, not otherwise specified; IQR, interquartile range.



**Figure 2** Heatmap for the model performance of each machine learning algorithm for 5-year OS and 5-year DSS prediction. AUC, area under the curve; OS, overall survival; DSS, disease-specific survival; XGBoost, extreme gradient boosting machine; CatBoost, categorical boosting.

(Table S3) respectively. All characteristics were well balanced between the training set and the validation set (all P values >0.05).

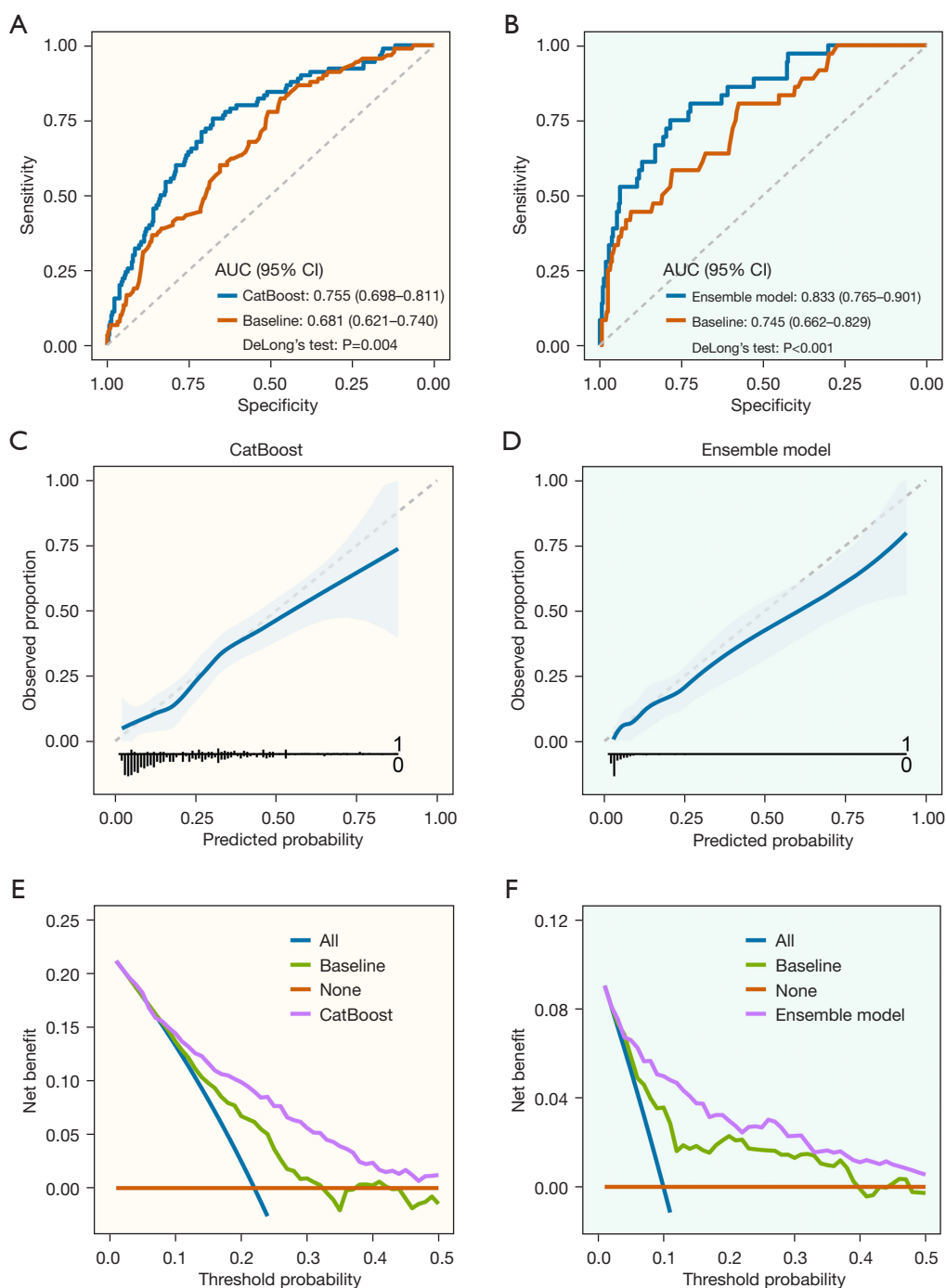
### Model performance

The performances of the models are presented in Figure 2. All the models performed excellently in the training set, with RF demonstrating the best performance in all the situations. However, this superiority did not persist into the validation

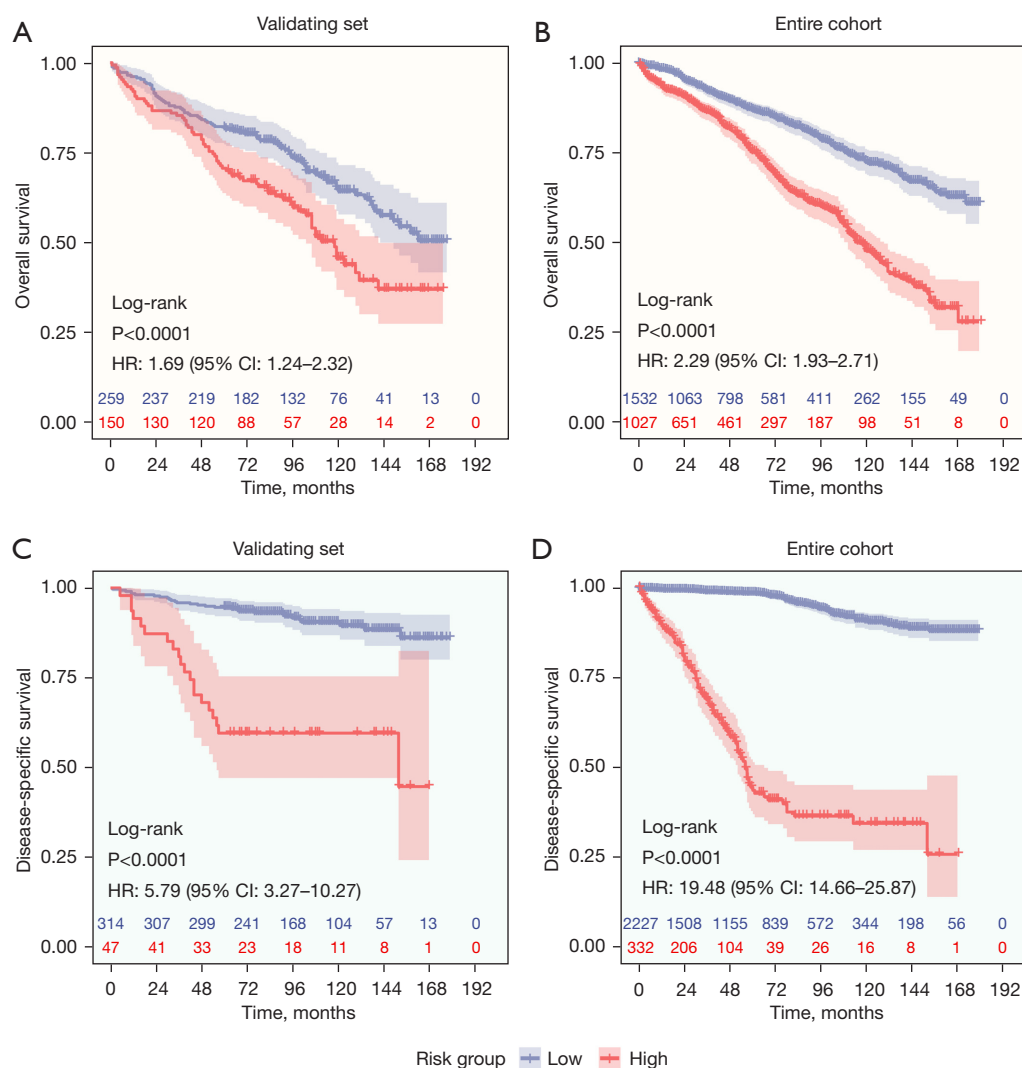
sets. CatBoost [AUC: 0.755; 95% confidence interval (CI): 0.698–0.811] had the best performance in the OS prediction for the original dataset (Figure 2 and Figure 3A). Additionally, the ensemble model consisting of ELR, RF and CatBoost achieved the highest prognostic efficiency in DSS prediction over the original dataset, with an AUC of 0.833 (95% CI: 0.765–0.901) (Figure 2 and Figure 3B). Thus, these two models were selected as the final models.

The transverse contrast indicated that all the models performed generally better in the DSS prediction task than





**Figure 3** Performance of the final classification model. (A,C,E) ROC, calibration and DCA for CatBoost in 5-year OS prediction. (B,D,F) ROC, calibration and DCA for ensemble model in 5-year DSS prediction. The ensemble model consisted of elastic net regression, random forest and CatBoost. Baseline model was logistic regression model based only on age and Masaoka stage. AUC, area under the curve; ROC, receiver operator characteristic; DCA, decision curve analysis; CatBoost, categorical boosting; OS, overall survival; DSS, disease-specific survival.



**Figure 4** Kaplan Meier survival curves for low and high-risk groups in both validating set and entire cohort for the final models. The efficacy of the CatBoost (A,B) predictive value and the ensemble model (C,D) predictive value in distinguish the prognosis of patients. HR, hazard ratio; CI, confidence interval; CatBoost, categorical boosting.

in the OS task (Figure 2). Moreover, the more complex ML models had a higher prognostic efficiency compared to the baseline model comprised only of age and Masaoka stage in the longitudinal contrast (Figure 2 and Figure 3A,3B).

The calibrations from the validation set are presented in Figure 3C,3D. Both the CatBoost model and the ensemble model showed favorable goodness of fit and can be further verified via Hosmer-Lemeshow test (CatBoost:  $\chi^2=12.63$ ,  $P=0.13$ ; ensemble model:  $\chi^2=7.61$ ,  $P=0.47$ ) (Figure 3C,3D). Additionally, the results from the decision curve analysis

confirmed that the final model provided high net benefit (Figure 3E,3F).

### Survival analysis

To further verify the prognostic value of our models, we conducted a survival analysis with the optimal classification thresholds identified via the “surv\_cutpoint” function. The optimal cut-off values for OS and DSS prediction were 0.62 and 0.75, respectively. Kaplan-Meier curves for



Table 2 Variable importance

Variables	Contribution to the ROC curves	
	OS (%)	DSS (%)
Age	23.43	14.42
Sex	0.00	1.37
History of other tumors	1.53	1.62
Race	3.77	0.00
Tumor size	6.15	1.28
Masaoka stage	12.50	14.45
Chemotherapy	6.43	10.60
Radiotherapy	6.11	2.23
Surgery type	8.14	7.75
WHO classification	16.81	31.00
Number of harvested lymph nodes	6.12	6.65
Lymph node invasion	7.02	2.63
Lung metastasis	1.99	6.00

ROC, receiver operation characteristic; OS, overall survival; DSS, disease-specific survival; WHO, World Health Organization.

the validation set and the entire cohort for each outcome are shown in *Figure 4*. Patients in the low-risk group had significantly better survival than did those in the high-risk group (all P values <0.001).

Variable significance

The significance of variables from the optimal model in the original dataset are shown in *Table 2*. The five most significant variables for OS prediction were age, World Health Organization (WHO) histological classification, Masaoka stage, surgery type, and chemotherapy; meanwhile, the most significant variables for DSS prediction were WHO histological classification, Masaoka stage, age, chemotherapy, and surgery type. *Table S4* presents the significant variables of the best model in the one-hot encoding dataset.

Discussion

In this study, we examined the prognostic value of ML algorithms in patients who had undergone surgery for thymoma and thymic carcinoma. Our models receive

routinely available clinical data and categorize patients into low- and high-risk groups in terms of 5-year OS and DSS. Our models demonstrated robust performance in the prognostic prediction of thymoma and thymic carcinoma, especially in the prediction of 5-year DSS.

In general, all of the models exhibited superior discrimination ability and consistency as compared to the basic prognostic model consisting of only age and Masaoka stage. All the variables contributed to the final models to some degree, but the significance of variables varied across the different models and different prediction tasks. Accounting for the myriad array of risk predictors can be challenging for clinicians, and thus developing these complex models is sensible and productive. Several factors contributed significantly to the prediction of both OS and DSS. There is a general consensus regarding the relevance of WHO histological classification, Masaoka stage, surgery type and lymph node invasion (4-10); meanwhile, age, which has also been established as a relevant factor, surpassed our expectation in OS prediction (8,32,33). This may be explained by the following: first, older adults have a higher rate of thymic carcinoma as compared to younger individuals (34). The proportion of postoperative myasthenia gravis is lower in younger patients, but that of complete stable remission is higher (35,36). Myasthenia gravis is a significant paraneoplastic disorder of thymoma, and it is often fatal (34-36). Other contributing risk factors, such as remaining life expectancy and resilience to concurrent diseases, may further explain the significance of age in OS prediction.

Chemotherapy was also found to be a significant factor in this study. However, the role of chemotherapy in the thymoma and thymic carcinoma remains controversial (4,16). Chemotherapy can be adopted as neoadjuvant therapy for patients with potentially resectable thymoma or thymic carcinoma and may serve as primary therapy for patients with unresectable thymoma or thymic carcinoma (37). For patients with a positive margin or residual disease after resection, chemotherapy is also practicable (37). However, the majority of (61.3%) the patients in the presenting study do not belong to the above-mentioned condition. Which supported the potential effect of chemotherapy on patients with thymoma or thymic carcinoma. It should be noted, however, that ML algorithms are not particularly suited to characterizing the relationship between variables and outcomes. Thus, further work is needed to determine the prognostic effect of chemotherapy on patients with thymoma and thymic carcinoma.

As expected, we found that all the models examined in our study generally performed better in DSS prediction than in OS prediction. The features included in the presenting study are more likely to have a stronger connection with disease-related prognosis, rather than the general prognosis of patients. Features from the data directly determines the performance ceiling of the models, while more powerful algorithms can merely determine how close the model gets to that ceiling.

Although several methods were adopted to avoid overfitting, our model demonstrated overfitting in the training set to some degree. This could potentially be attributed to a relatively lower number of positive cases, possibly arising from a comparatively favorable prognosis. However, we consider this overfitting to be acceptable for the following reasons: first, our models had a fairly good performance in the validation set. Second, more stringent control of overfitting during the training process did not significantly improve the performance of model. Regardless, our final model yielded a better performance in OS prediction than that reported in a previous study based on Cox regression (AUC in the validation set: 0.746) (8). Although the differences were not substantial, this is significant, as the improvement was completely “free”. The Cox regression model has limitations in that the proportional hazard assumption and linearity of each variable are difficult to be satisfied in real-world data (19). Second, in tied samples, approximations are used for better computational efficiency, but this can lead to significantly different outcomes (38). Furthermore, issues related to multiple-testing may arise when a comparatively substantial number of variables are processed. Potential false positives from significant features may reduce the reliability of models (39). Introducing one-hot encoding did not significantly improve the performance of models in the presenting study; rather, one-hot encoding may entail risks, including dimensionality increase, information loss, and reduction in the interpretability of the clinical data. Thus, the application of one-hot encoding in the clinical data should not be considered a necessity but should be carefully examined in this context.

Several limitations to this study should be acknowledged. First, due to the morbidity of thymoma and thymic carcinoma, external validation was lacking in this study. Therefore, larger-scale studies are required to confirm our results. Second, imaging data and biological data are not available from the SEER database and were thus not included for the further boosting of model performance.

Using a multimodal dataset to construct more robust models is not merely a prevalent direction in current mainstream research but is also an effective approach for more comprehensively exploiting clinical information and thus improving model performance.

## Conclusions

We trained ML-based predictive models that demonstrated good accuracy in predicting the 5-year OS and DSS of patients with thymoma and thymic carcinoma. The application of ML algorithms has not yet been widely applied in the prognostic prediction of cancers, especially in patients with thymoma and thymic carcinoma. ML tools have the potential to render clinical judgement more swift, precise, and objective, and the inclusion of multimodal data may confer additional benefit and should be investigated.

## Acknowledgments

None.

## Footnote

*Reporting Checklist:* The authors have completed the TRIPOD reporting checklist. Available at <https://jtd.amegroups.com/article/view/10.21037/jtd-24-1263/rc>

*Peer Review File:* Available at <https://jtd.amegroups.com/article/view/10.21037/jtd-24-1263/prf>

*Funding:* This work was supported by the Science and Technology Program of Guangdong, China (grant number: 210716126901104) and the “Talent Support” Program of The First Affiliated Hospital of Shantou University Medical College. The sponsors were not involved in the study design, data collection, analysis, and interpretation, writing or the decision to submit for publication.

*Conflicts of Interest:* All authors have completed the ICMJE uniform disclosure form (available at <https://jtd.amegroups.com/article/view/10.21037/jtd-24-1263/coif>). 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 study was

conducted in accordance with the Declaration of Helsinki (as revised in 2013). All data were derived from publicly available open-access databases and so ethical approval was not required.

**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. Xin Z, Lin M, Hao Z, et al. The immune landscape of human thymic epithelial tumors. *Nat Commun* 2022;13:5463.
2. Ao YQ, Gao J, Wang S, et al. Immunotherapy of thymic epithelial tumors: molecular understandings and clinical perspectives. *Mol Cancer* 2023;22:70.
3. Li Q, Pu Y, Gong Z, et al. Preoperative systemic immune-inflammation index for predicting the prognosis of thymoma with radical resection. *Thorac Cancer* 2023;14:1192-200.
4. Scorsetti M, Leo F, Trama A, et al. Thymoma and thymic carcinomas. *Crit Rev Oncol Hematol* 2016;99:332-50.
5. Alkaaki A, Abo Al-Saud A, Di Lena É, et al. Factors predicting recurrence in thymic epithelial neoplasms. *Eur J Cardiothorac Surg* 2022;62:ezac274.
6. Wu J, Fang W, Chen G. The enlightenments from ITMIG Consensus on WHO histological classification of thymoma and thymic carcinoma: refined definitions, histological criteria, and reporting. *J Thorac Dis* 2016;8:738-43.
7. Wang ZM, Li F, Sarigül L, et al. A predictive model of lymph node metastasis for thymic epithelial tumours. *Eur J Cardiothorac Surg* 2022;62:ezac210.
8. Zhao M, Yin J, Yang X, et al. Nomogram to predict thymoma prognosis: A population-based study of 1312 cases. *Thorac Cancer* 2019;10:1167-75.
9. 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.
10. Tassi V, Vannucci J, Ceccarelli S, et al. Stage-related outcome for thymic epithelial tumours. *BMC Surg* 2019;18:114.
11. Masaoka A, Monden Y, Nakahara K, et al. Follow-up study of thymomas with special reference to their clinical stages. *Cancer* 1981;48:2485-92.
12. 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.
13. 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;22:1501-9.
14. Du X, Cui J, Yu XT, et al. Risk factor analysis of thymoma resection and its value in guiding clinical treatment. *Cancer Med* 2023;12:13408-14.
15. Dai H, Lan B, Li S, et al. Prognostic CT features in patients with untreated thymic epithelial tumors. *Sci Rep* 2023;13:2910.
16. Falkson CB, Vella ET, Ellis PM, et al. Surgical, Radiation, and Systemic Treatments of Patients With Thymic Epithelial Tumors: A Systematic Review. *J Thorac Oncol* 2023;18:299-312.
17. Jiang YG, Ma MY, Wu JJ, et al. Prognostic factors in patients with thymoma who underwent surgery. *World J Surg Oncol* 2023;21:203.
18. Rahman SA, Walker RC, Lloyd MA, et al. Machine learning to predict early recurrence after oesophageal cancer surgery. *Br J Surg* 2020;107:1042-52.
19. Hindocha S, Charlton TG, Linton-Reid K, et al. A comparison of machine learning methods for predicting recurrence and death after curative-intent radiotherapy for non-small cell lung cancer: Development and validation of multivariable clinical prediction models. *EBioMedicine* 2022;77:103911.
20. Zhang Y, Zhang L, Li B, et al. Machine learning to predict occult metastatic lymph nodes along the recurrent laryngeal nerves in thoracic esophageal squamous cell carcinoma. *BMC Cancer* 2023;23:197.
21. Ye W, Yan T, Zhang C, et al. Detection of Pesticide Residue Level in Grape Using Hyperspectral Imaging with Machine Learning. *Foods* 2022;11:1609.
22. Lu MY, Liu TW, Liang PC, et al. Decision tree algorithm predicts hepatocellular carcinoma among chronic hepatitis C patients following viral eradication. *Am J Cancer Res* 2023;13:190-203.
23. Kong G, Lin K, Hu Y. Using machine learning methods to predict in-hospital mortality of sepsis patients in the ICU. *BMC Med Inform Decis Mak* 2020;20:251.

24. Rosenberg MA. Trusting Magic: Interpretability of Predictions From Machine Learning Algorithms. *Circulation* 2021;143:1299-301.
25. Sealfon RSG, Mariani LH, Kretzler M, et al. Machine learning, the kidney, and genotype-phenotype analysis. *Kidney Int* 2020;97:1141-9.
26. Zhang C, Wang Q, Hu L, et al. The Prognostic Value of Postoperative Radiotherapy for Thymoma and Thymic Carcinoma: A Propensity-Matched Study Based on SEER Database. *Cancers (Basel)* 2022;14:4938.
27. Zou H, Hastie T. Regularization and Variable Selection Via the Elastic Net. *J R Stat Soc Ser B Stat Methodol* 2005;67:301-20.
28. Breiman L. Random Forests. *Mach Learn* 2001;45:5-32.
29. 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 (KDD '16)*. New York, NY, USA: Association for Computing Machinery; 2016:785-94.
30. Dorogush AV, Ershov V, Gulin A. CatBoost: gradient boosting with categorical features support. *arXiv* 2018. *arXiv:1810.11363*.
31. Kuhn M. caret: Classification and Regression Training. *Astrophysics Source Code Library*. 2015;ascl:1505.003.
32. Nam JG, Goo JM, Park CM, et al. Age- and gender-specific disease distribution and the diagnostic accuracy of CT for resected anterior mediastinal lesions. *Thorac Cancer* 2019;10:1378-87.
33. Giorgetti OB, Nusser A, Boehm T. Human thymoma-associated mutation of the GTF2I transcription factor impairs thymic epithelial progenitor differentiation in mice. *Commun Biol* 2022;5:1037.
34. Rich AL. Epidemiology of thymoma. *J Thorac Dis* 2020;12:7531-5.
35. Zhang J, Zhang Z, Zhang H, et al. Thymectomy in ocular myasthenia gravis-prognosis and risk factors analysis. *Orphanet J Rare Dis* 2022;17:309.
36. Zhang J, Zhang P, Zhang H, et al. Thymectomy in thymomatous generalized myasthenia gravis: An analysis of the prognosis and risk factors. *Eur J Neurol* 2023;30:2012-21.
37. NCCN [Internet]. [cited 2024 Jul 25]. National Comprehensive Cancer Network guidelines. Available online: [https://www.nccn.org/guidelines/category\\_1](https://www.nccn.org/guidelines/category_1)
38. Lee B, Chun SH, Hong JH, et al. DeepBTS: Prediction of Recurrence-free Survival of Non-small Cell Lung Cancer Using a Time-binned Deep Neural Network. *Sci Rep* 2020;10:1952.
39. Zhang Y, Oikonomou A, Wong A, et al. Radiomics-based Prognosis Analysis for Non-Small Cell Lung Cancer. *Sci Rep* 2017;7:46349.

**Cite this article as:** Xu H, Lin X, Wu J, Chen J, Wu J, Lin Z, Cai X, Lin J, Li P, He C, Xie Z, Wu H. Machine learning for predicting the prognosis of patients with thymoma and thymic carcinoma. *J Thorac Dis* 2025;17(2):824-835. doi: 10.21037/jtd-24-1263