**Peer Review File**

**Article Information:** https://dx.doi.org/10.21037/jtd-21-1948

**Reviewer A**

In this article, the authors study the usefulness of a CT-based radiomics analysis to differentiate thymomas (low and high-grade) from thymic carcinomas.
The topic is timely, the methodology sound, and the presentation and English language are appropriate.
The authors elegantly show that two texture features, namely GLCM-energy and solidity, are good predictors of thymic malignancy, rendering it a potentially useful tool in clinical practice.
I have no specific revision to suggest, whether on the methodology/results or the way the results are displayed.
Ans.) Thank you.

**Reviewer B**

Interesting analysis. I suggest clarifying a couple of things briefly:
- Why atypical type A variant thymoma was not included in the present study?
Ans.) Thank you for your comment.
The present study included data of thymectomy (n = 51) or biopsy (n = 12) performed between January 2010 to December 2013. We performed this study according to the pathological results diagnosed during this period. The atypical type A variant is not included in this study because it is a classification proposed in 2015.

- Considering radiomics analysis will be particularly useful in cases for which surgery or biopsy is not feasible, which advantages will it have over PET/CT?
Ans.) Thank you for your comment.
We are soryy that we cannot clearly answer which advantages the radiomics analysis will have over PET/CT because we did not compare results between the radiomics analysis and PET/CT. Our present study demonstrated that the CT radiomics features GLCM_energy and Solidity were useful and had high specificity for predicting thymic carcinoma. The high specificity of our model suggests its suitability for use in tertiary hospitals to reduce unnecessary examinations. Radiomics analysis might be an effective tool in differentiating between thymoma and thymic carcinoma in clinical treatment planning, particularly in cases for which surgery or biopsy is not feasible. In addition, if thymic carcinoma is diagnosed using radiomics features, metastasis can be detected by PET/CT, histological type can be determined by biopsy, and staging and an appropriate treatment plan can be determined.

- Necrotic component, very common in these tumors, can affect tumor size and volumen and be confusing. How will this affects radiomics analysis?
Ans.) Thank you for your comment.
As you indicated, tumor includes heterogeneous area because of necrosis and/or new

blood vessels. In the present study, semi-automatic segmentation was performed in the maximum cross-sectional image of each tumor, which was as large as possible to minimize inter-tumor variability. Therefore, necrotic component was included in the segmented area in this study. Evaluation of these areas is important for diagnosing. The radiomics features such as GLCM tend to increase due to hetertogenity casued by necrosis.

**Reviewer C**

The authors tested radiomics features on CT imaging to differentiate between thymoma and thymic carcinomas. The idea is interesting and has potential clinical value, but crucial methodological issues undermine the power of the results. The differentiation of thymic carcinoma from high-risk thymomas is usually challenging, even in the pathology, what was marginally addressed by the authors. A subanalysis comparing high-risk thymomas and thymic carcinomas is imperative.

Ans.) Thank you for your suggestion.

We added a subanalysis comparing high-risk thymomas and thymic carcinomas. Please see the 'Predictive performance for thymic carcinoma using radiomics features' in the result section of the revised text:

In a subanalysis comparing high-risk thymoma and thymic carcinoma, multiple logistic regression analysis revealed GLCM-energy as an independent indicator associated with thymic carcinoma (odds ratio, 69.3; 95% confidence interval, 6.4–748.1; P = 0.0005). AUC for diagnosing thymic carcinoma was 0.877 (95% confidence interval, 0.72–0.97): sensitivity, 81.3% and specificity, 94.1%.

We also added the following sentence in the first paragraph of the discussion section (clean main body, Page17 Line4-5):

Particularly, in comparing high-risk thymoma and thymic carcinoma, GLCM-energy was as an independent indicator associated with thymic carcinoma.

The sample is very small (16 thymic carcinomas), even for rare tumors, and class imbalance (45 thymomas) potentially affects the investigation and induces some bias, what was not alleviated by proper strategies.

Ans.) Thank you for your suggestion.

We agree with you. However, thymic carcinoma is rare and imbalances may not be avoided. Therefore, we added the following sentence to the limitation section (Page21 Line4-5):

Class imbalance (45 thymomas and 16 thymic carcinomas) may potentially affect the investigation and induces some bias, what was not alleviated by proper strategies.

The feature selection criteria is very simplistic. A solid radiomics model recquires more advanced strategies. Validation is a critical step in the worklow radiomics process and the presented validation strategy (cross-validation is subjected to criticisms. Differences in the training and testing cohorts were not properly presented.

Ans.) Thank you for your suggestion.

We agree with you. We revised the statistical section for the feature selection criteria. However, because of the small number cases, it was not possible to create a sufficient set of training, test, and validation. Therefore, 10-fold cross validation was performed in this study. If the number of cases is small, cross-validation may be permissible: e.g., Medicine (Baltimore). 2019 Jun; 98 (25): e16119.

We added the following explanation into the statistical analysis section (Page12 Line14 - Page13 Line1):
The hyper-parameter of LASSO regression (weight parameter for the regularization term) was determined by the internal cross validation within each fold. Features with non-zero regression coefficients were considered important and importance of a feature was determined by counting how many times the feature was considered important through the repeated cross validation.

We are sorry not to validate our results by using other cohort this time.
We also added the following sentence to the limitation section (Page21 Line6-8):
Moreover, only 10-fold cross validation was performed because of the small number of cases in the present study. Validation using other cohort might have been a critical step in the worklow radiomics process.

Comparison with visual assessment may be of interest. Radiomics workflow is time consuming and requires expertise, therefore a comparison analysis would support the real need for advanced image analysis. Lesion size may affect radiomic feature calculation.
Ans.) Thank you for your suggestion.
Comparison with visual assessment may be interesting. Although there are previous reports of specific findings among low-risk, high-risk, and thymic carcinoma, it is very difficult to visually and accuratly differenciate among them because of overlap CT findings. For example, the presence or absence of lymph node metastasis has been reported to be a significant finding for differentiating thymic carcinoma. However, only tumor itself was evaluated using radiomics analysis, so all lung field was not evaluated in the present study. As you suggested, we would like to compare between visual assessment and radiomics/artificiall intelligence assessment using all lung field data in the future.
With regard to lesion size, tumor volume and ratio (major axis length/minor axis length) have been already used as one of the explanatory factors for logistic regression model: please see table 2 and 3.

Some papers have been reported about the relationship between size/volume and histological types: JTD 2018:10;5822-5832, Eur Radiol 2017;27:1992-2001, Asian Pacific J Cancer Prev 2012;13:5581-5585, etc. On these papars, tumor size has also been reported to be one of the useful factors for histology. However, tumor volume and ratio (major axis length / minor axis length) were not useful factors in multivariate analysis including other radiomic features of the present study. Multiple

logistic regression analysis using a stepwise method revealed two features (Solidity and GLCM-energy) as independent indicators associated with thymic carcinoma (odds ratio, 14.7 and 14.3; 95% confidence interval, 1.6–139.0 and 3.0–68.7; P = 0.045 and 0.002, respectively) in this study.

**Reviewer D**

General comments:

I have reviewed the manuscript with great interest because radiomics features are evolving and proven helpful in tumor diagnosis and predicting patient's prognosis. The authors applied radiomics features to classify thymic epithelial tumors, particularly in diagnosing thymic carcinoma. The manuscript is well written, with a clear hypothesis and conclusion. However, as of the nature of radiomics features ---- some of the features are not correlate to visual assessment--- it may be helpful for readers to follow the context with a little more explanation about the parameters (GLCM-energy and Solidity) revealed to predict thymic carcinoma. Also, it needs more discussion regarding how other parameters (i.e., volume and homogeneity), expected to be valid from visual features, turned out to be less valuable. In addition, it is a little unclear in the comparison groups and statistical analysis (in particular, tumor volume). Also, I would suggest that the model could include other clinical features such as age and visual features to be more clinically practical.

Ans.) Thank you for your thoughtful comments and suggestions. Please see our answers below.

Specific comments:

Page8 Line9-10: slice thickness of <1mm

Is there any data showing if slice thickness affects the radiomics parameters? I assume some of the features may be more accurate and vice versa.

Ans.) Thank you for your comment.

For example, the reproducibility of radiomics features in lung cancer is significantly influenced by CT slice thickness, which can be improved by the convolutional neural network-based super-resolution algorithms ( Korean J Radiol. 2019;20:1431-1440.)

The other paper (Sci Rep. 2016;6:23428) demonstreated that increased noise levels associated with thinner slice images may disturb texture features as many texture features are quite sensitive to fluctuation of image densities. On the other hand, although thicker slices decrease noise levels, they can blur the images (diminish texture details) due to poor spatial resolution along the axial direction and larger partial volume effects. Moreover, at the same slice thickness, a smoother reconstruction algorithm can reduce more noise from images than a sharper one, but the smoother algorithm may hold back useful texture details from images.

Moreover, in the present study, because volumetry was also performed, volume can be measured more accurately on thinner images than on thicker images.

Therefore, CT images that were obtained with slice thickness of <1 mm were

included.

Page8-9: Histopathological data
Please explain if there are any features useful for low and high-risk thymomas. And it is unclear how the comparison was made, thymoma vs. thymic carcinoma or high-risk thymoma vs. carcinoma. I am not sure a detailed discussion of WHO classification is necessary, which was repeated in the Background and here again, although the authors used three categories. Mentioning atypical variant is probably irrelevant in this study.
Ans.) Thank you for your suggestion.
According to your suggestion, we deleted the term 'atypical type A' in the text, and deleted a detailed discussion of WHO classification in this section.
Survival rates for patients with type A, AB, and B1 tumors were higher than for those with type B2, B3, and thymic carcinoma (references 14, 15). Therefore, in our study, we classified the WHO histologic classification of thymic epithelial tumors into three subgroups—that is, low-risk thymoma (A, AB, and B1), high-risk thymoma (B2 and B3), and thymic carcinoma.

We revised this section as follows (Page10 Line1-3):
we classified all tumors into low-risk thymoma, high-risk thymoma, and thymic carcinoma according to the prognostic value of WHO histologic classification, as described previously[4,14,15].

Page10 Line15: Did the authors include tumor volume in the analysis? I do see major and minor diameters in Figure 3, but volume.
Ans.) Thank you for your suggestion.
Tumor volume and ratio (major axis length/minor axis length) have been already used as one of the explanatory factors for logistic regression model: please see table 2 and 3.
We added the following sentence in the text (Page11 Line16-17): The volume was calculated in the same way as in the previous study[15].

Figure 2 showed CT radiomics features extracted using our developed software.
Figure 3 showed important radiomics features associated with the three groups (low-risk thymoma; high-risk thymoma; and thymic carcinoma).
We revised the legend of figure 2 and inseted figure 2 into the correct place of the text (Page11 Line2).

Some papers have been reported about the relationship between size/volume and histological types: JTD 2018:10;5822-5832, Eur Radiol 2017;27:1992-2001, Asian Pacific J Cancer Prev 2012;13:5581-5585, etc. On these papars, tumor size has also been reported to be one of the useful factors for histology. However, tumor volume and ratio (major axis length / minor axis length) were not useful factors in multivariate analysis including other radiomic features of the present study. Multiple

logistic regression analysis using a stepwise method revealed two features (Solidity and GLCM-energy) as independent indicators associated with thymic carcinoma (odds ratio, 14.7 and 14.3; 95% confidence interval, 1.6–139.0 and 3.0–68.7; P = 0.045 and 0.002, respectively) in this study.

We added the following sentence and new references in the statistical analysis section (Page13 Line4-7):
Moreove, we added volume as one of the explanatory factors because tumor size has been reported to be one of the characteristic factors for histological subtypes of thymic epithelial tumors[16,17,18].

Because new references were added, the other document numbers in the text were newly assigned.
New reference 16: Chang S, et al. Volume-based quantification using dual-energy computed tomography in the differentiation of thymic epithelial tumours: an initial experience. Eur Radiol. 2017;27:1992-2001.

New reference 17: Liu GB, Qu YJ, Liao MY, Hu HJ, Yang GF, Zhou SJ. Relationship between computed tomography manifestations of thymic epithelial tumors and the WHO pathological classification. Asian Pac J Cancer Prev. 2012;13:5581-5585.

New reference 18: Blumberg D, et al. Thymoma: a multivariate analysis of factors predicting survival. Ann Thorac Surg. 1995 ;60:908-913.

Discussion:
Page15 Line 11: I would like to see more explanation about GLCM-energy even though it does not correlate with the visual feature. Any speculation in relating to pathologic features?
Ans.) Thank you for your suggestion.
It is just a speculation, but GLCM_energy is to measure the degree of fluctuation in the space in the tumor. The feature GLCM-energy extracted in this study is difficult to observe subjectively on CT images. Generally, squamous cell carcinoma is the most frequent subtype of thymic carcinoma, which is composed of large polyhedral cells arranged in nests and cords, showing evidence of keratinization and/or intercelluar bridges. Foci of spontaneous necrosis are frequently seen, as is the invasion of intratumoral blood vessels, resulting in the heterogeneity of the tumor. GLCM-energy might correlate to the heterogeneity of thymic carcinoma.
Please see the discussion section (Page17 Line13 – Page18 Line2).

In the same context, please add more explanation of Solidity rather than just a definition.
Ans.) Thank you for your suggestion.
A value of 1 signifies a solid object, and a value less than 1 will signify an object having an irregular boundary, or containing holes. Thymic carcinoma might indicate lower

solidity than thymoma because of irregular margin and internal necrosis.
Please see the discussion section (Page18 Line4-7).

Page17 Line3-11: This explanation may be inconsistent with the previous publication. Generally, tumor volume is a significant prognostic factor in most tumors, and therefore, the previous conclusion may be more convincing. Need more explanation how the authors handled tumor volume data and concluded.
Ans.) Thank you for your suggestion.
Some papers have been reported about the relationship between size/volume and histological types: JTD 2018:10;5822-5832, Eur Radiol 2017;27:1992-2001, Asian Pacific J Cancer Prev 2012;13:5581-5585, etc. On these papars, tumor size has also been reported to be one of the useful factors for histology.
(In the paper [Eur Radiol 2017;27:1992-2001], there were no significant difference in volume among low-risk thymoma, high-risk thymoma, and thymic carcinoma.)

In the previous studies[7, 15] also, univariate analysis revealed that tumor volume was useful for distinguishing between thymoma and thymic carcinoma, which was in accordance with the present study. (Please see table 2 and 3.)
However, tumor volume and ratio (major axis length / minor axis length) were not useful factors in multivariate analysis including other radiomic features of the present study. Multiple logistic regression analysis using a stepwise method revealed two features (Solidity and GLCM-energy) as independent indicators associated with thymic carcinoma (odds ratio, 14.7 and 14.3; 95% confidence interval, 1.6–139.0 and 3.0–68.7; P = 0.045 and 0.002, respectively) in this study.

We revised the discussion section a little. Please see the revised text (Page19 Line14 – Page20 Line2).

I would suggest combining these radiomics features with non-binary volume data and demographic data (especially age) so that the authors could build a better predictive model.
Ans.) Thank you for your suggestion.
As in many other papers, in the present study, for each feature including volume, the cutoff value that yielded the largest difference in numbers of patients with and without thymic carcinoma was determined using the ROC method. Optimal thresholds were determined for each variable separately using the Youden index (the highest sum of sensitivity and specificity). Associations between thymic carcinoma and each binary group designated by the cutoff value for the seven radiomics features were evaluated by univariate logistic regression analysis. Significant parameters identified by univariate analysis were included in multiple logistic regression (stepwise method; P value of 0.05 or less was used for entry into the model and P value greater than 0.1 was selected for removal).
Therefore, we would like to keep the present model.

As you indicated in your below comment for Table 1, we added the analysis comparing low-risk, high-risk, and carcinoma with p-values for sex and age.
In the present study, there were no significant differences in the distribution of age between thymoma and thymic carcinoma. However, there was a significant difference in the distribution of sex between thymoma and thymic carcinoma. (Please see a new table 1.) Therefore, we also added new results by a new multivariate analysis adjusted for sex. Please see the result section (Patient data & Predictive performance for thymic carcinoma using radiomics features). (Page14 Line1 – Page16 Line8)

Moreover, we added some revision in the discussion section as follows (Page20 Line5-8):
In general, thymic carcinoma has little gender difference[4], but in the present study, multivariate analysis adjusted for sex was also performed because of the predominant distribution in men. The result was almost the same: two texture features (GLCM-energy and Solidity) were significant predictors of thymic carcinoma.

Page18 Line 12-14: This is not very clear to me. Did the results derive from thymoma vs. thymic carcinoma? Please be consistent with analysis methods and results.
Ans.) We apologize for the confusion.
We revised the conclusion (Page21 Line 14 - Page 22 Line2).

Figure3:
Volume data included here?
What are the x-axis parameters?
Please explain why GLCM Homogeneity resulted not useful in diagnosing thymic carcinoma because the feature ranked top in the "importance." It is counterintuitive to say that "Homogeneity" is less helpful.
Ans.) Thank you for your suggestion.
As we have already answered, tumor volume and ratio (major axis length/minor axis length) have been already used as one of the explanatory factors for logistic regression model: please see table 2 and 3.
Figure 3 showed important radiomics features associated with the three groups (low-risk thymoma; high-risk thymoma; and thymic carcinoma).
Therefore, volume is not incuded in this figure.
Apart from the radiomics features, volume was used as one of the explanatory factors because tumor size has been reported to be one of the characteristic factors for histological subtypes of thymic epithelial tumors[16,17,18].

X-axis shows the number of times each radiomics feature was considered important (regression factor was not 0) in the cross validation. Feature importance was defined as the number of times for a feature to have non-zero LASSO regression coefficient over the repeated cross validation. For our 100-times repeated 10-fold cross validation, maximum possible importance was 1000.

We consulted with a statistician of our institution. The degree of importance does not necessarily correlate with the statistical significance of logistic analysis including odds ratio, etc.
In the present study, univariate analysis revealed that GLCM-homogeneity was the significant indicator associated with thymic carcinoma (odds ratio, 15.5; 95% confidence interval, 3.1-77.6; P = 0.001). However, GLCM-homogeneity was statistically removed by multiple logistic regression analysis using a stepwise method, resulting in less significance than Solidity and GLCM-energy.

Table1: Please provide data regarding the study comparing low-risk, high-risk, and carcinoma with p-values for sex and age. I assume age may be significantly different in the categories, and if so, it should be adjusted or included in the model.
Ans.) Thank you for your suggestion. We added a new analysis.
Please see the revised result section (Page14 Line1 – Page16 Line8).

Table3: It seems the authors compared thymoma vs. carcinoma based on the data in this table. Please be consistent throughout the manuscript. In addition, please provide the P values of multivariate analysis of other features even though those are insignificant. The univariate analysis results seem promising in volume, compactness, and homogeneity. Therefore, I would see precise data to reject those features.
Ans.) Thank you for your comments.
In this study, significant parameters identified by univariate analysis were included in multiple logistic regression. Multiple logistic regression was performed using stepwise method. In the stepwise method, P value of 0.05 or less was used for entry into the model and P value greater than 0.1 was selected for removal. Therefore, no p-value is given for items that are not significantly different (MedCalc, version20.015-64bit). This statistical method has been used in many papers.
For example, please see our previous paper: Yanagawa M, et al. Lung Adenocarcinoma at CT with 0.25-mm Section Thickness and a 2048 Matrix: High-Spatial-Resolution Imaging for Predicting Invasiveness. Radiology. 2020 Nov;297(2):462-471. doi: 10.1148/radiol.2020201911.