Application of Chest CT Imaging Feature Model in Distinguishing Squamous Cell Carcinoma and Adenocarcinoma of the Lung

Chunmei Liu,1 Yuzheng He,2 Jianmin Luo3

1Department of Radiation Oncology, The Second Hospital of Hebei Medical University, Shijiazhuang, Hebei Province, People’s Republic of China; 2Department of Thoracic Surgery, The Second Hospital of Hebei Medical University, Shijiazhuang, Hebei Province, People’s Republic of China; 3Department of Hematology, The Second Hospital of Hebei Medical University, Shijiazhuang, Hebei Province, People’s Republic of China

Correspondence: Jianmin Luo, Department of Hematology, The Second Hospital of Hebei Medical University, 215 West Heping Road, Shijiazhuang, Hebei Province, People’s Republic of China, Tel +86 1348368527, Email luojm31555@163.com

Purpose: In situations where pathological acquisition is difficult, there is a lack of consensus on distinguishing between adenocarcinoma and squamous cell carcinoma from imaging images, and each doctor can only make judgments based on their own experience. This study aims to extract imaging features of chest CT, extract sensitive factors through logistic univariate and multivariate analysis, and model to distinguish between lung squamous cell carcinoma and lung adenocarcinoma.

Methods: We downloaded chest CT scans with clear diagnosis of adenocarcinoma and squamous cell carcinoma from The Cancer Imaging Archive (TCIA), extracted 19 imaging features by a radiologist and a thoracic surgeon, including location, spicule, lobulation, cavity, vacuolar sign, necrosis, pleural traction sign, vascular bundle sign, air bronchogram sign, calcification, enhancement degree, distance from pulmonary hilum, atelectasis, pulmonary hilum and bronchial lymph nodes, mediastinal lymph nodes, interlobular septal thickening, pulmonary metastasis, adjacent structures invasion, pleural effusion. Firstly, we apply the glm function of R language to perform logistic univariate analysis on all variables to select variables with P < 0.1. Then, perform logistic multivariate analysis on the selected variables to obtain a predictive model. Next, use the roc function in R language to calculate the AUC value and draw the ROC curve, use the val.prob function in R language to draw the Calibrat curve, and use the rmda package in R language to draw the DCA curve and clinical impact curve. At the same time, 45 patients diagnosed with lung squamous cell carcinoma and lung adenocarcinoma through surgery or biopsy in the Radiotherapy Department and Thoracic Surgery Department of our hospital from 2023 to 2024 were included in the validation group. The chest CT features were jointly determined and recorded by the two doctors mentioned above and included in the validation group. The included image feature data are complete and do not require preprocessing, so directly entering statistical calculations. Perform ROC curves, calibration curves, DCA, and clinical impact curves in the validation group to further validate the predictive model. If the predictive model performs well in the validation group, further draw a nomogram to demonstrate.

Results: This study extracted 19 imaging features from the chest CT scans of 75 patients downloaded from TCIA and finally selected 18 complete data for analysis. First, univariate analysis and multivariate analysis were performed, and a total of 5 variables were obtained: spicule, necrosis, air bronchogram Sign, atelectasis, pulmonary hilum and bronchial lymph nodes. After conducting modeling analysis with AUC = 0.887, a validation group was established using clinical cases from our hospital, Draw ROC curve with AUC = 0.865 in the validation group, evaluate the accuracy of the model through Calibrate calibration curve, evaluate the reliability of the model in clinical practice through DCA curve, and further evaluate the practicality of the model in clinical practice through clinical impact curve.

Conclusion: It is possible to extract influential features from ordinary chest CT scans to determine lung adenocarcinoma and squamous cell carcinoma. The model we have set up performs well in terms of discrimination, accuracy, reliability, and practicality.

Keywords: lung cancer, LUAD, LSCC, image features, predict

Introduction

The pathology of lung cancer is very important in personalized and precise treatment,1,2 but in clinical practice, we inevitably encounter a challenge, which is that sometimes only a small number of tumor cells can be found after biopsy, and even necrotic cells can only be seen. Even after immunohistochemistry, the specific pathological type cannot be
determined. Based on tumor markers, we can sometimes further rule out the diagnosis of small cell lung cancer, but further distinguishing between squamous cell carcinoma and adenocarcinoma is basically impossible. Pathological patients who undergo biopsy again often refuse due to risk and pain. But there are significant differences in treatment options between squamous cell carcinoma and adenocarcinoma, according to the National Comprehensive Cancer Network (NCCN) guidelines. At this point, doctors have no other choice but to make further judgments based on imaging features. However, this judgment requires rich clinical experience and a high error rate. This leads to a lack of precise treatment for patients with unclear pathology, resulting in a decrease in treatment effectiveness and a shortened survival time. Therefore, how to distinguish lung squamous cell carcinoma and adenocarcinoma based on common chest CT imaging features has become a challenge in treating patients with unclear pathology.

Some studies have also attempted to model by extracting features from various image data, some based on CT texture features, some based on the iodine concentration of enhanced CT, some rely on the attenuation rate of enhanced CT, some models are established based on the perfusion parameters of brain metastases, some use PET-CT, and some combine machine learning. However, these examinations increase the economic burden on patients or are difficult to operate, and they have not been proven to be clinically applicable. Because chest CT scans are more common and cost-effective, it is more meaningful and beneficial for patients to distinguish between adenocarcinoma and squamous cell carcinoma based on chest CT scans. In the past, only one study based on chest CT features studied the pathological differences of lung cancer, but he did not further analyze the differences between squamous cell carcinoma and adenocarcinoma. This study aims to extract traditional imaging features from chest CT plain scans and analyze them, aiming to obtain a reliable predictive model to help clinical doctors distinguish between lung squamous cell carcinoma and lung adenocarcinoma.

Method
Download images with squamous cell carcinoma or adenocarcinoma from The Cancer Imaging Archive (TCIA). Inclusion criteria: 1. Pathologically confirmed as lung squamous cell carcinoma or lung adenocarcinoma; 2. The contour of lung tumors is relatively clear; 3. Complete lung tumor scanning; 4. CT is a plain or enhanced image. Exclusion criteria: 1. Unclear pathology; 2. A large amount of pleural effusion or pneumonia obstructs the contour of the tumor; 3. Incomplete lung tumor scanning; 4. Identify cases with significant tumor changes after radiotherapy or chemotherapy. A radiologist and a thoracic surgeon jointly determined 19 chest CT features, including location, spicule, lobulation, cavity, vacuolar sign, necrosis, pleural traction sign, vascular bundle sign, air bronchogram sign, calcification, enhancement degree, distance from pulmonary hilum, atelectasis, pulmonary hilum and bronchial lymph nodes, mediastinal lymph nodes, interlobular septal thickening, pulmonary metastasis, adjacent structures invasion, pleural effusion, pathology and record, enter the training group. If there is a lot of missing data in the feature, it will be removed. Firstly, we apply the glm function of R language to perform logistic univariate analysis on all variables to select variables with \( P < 0.1 \). Then, perform logistic multivariate analysis on the selected variables to obtain a predictive model. Next, use the roc function in R language to calculate the AUC value and draw the ROC curve, use the val.prob function in R language to draw the Calibrat curve, use the rmda package in R language to draw the DCA curve and clinical impact curve. At the same time, 45 patients diagnosed with lung squamous cell carcinoma and lung adenocarcinoma through surgery or biopsy in the Radiotherapy Department and Thoracic Surgery Department of our hospital from 2023 to 2024 were included. The chest CT features were jointly determined and recorded by the two doctors mentioned above and included in the validation group. The included image feature data is complete and does not require preprocessing, so directly entering statistical calculations. Perform ROC curves, calibration curves, DCA, and clinical impact curves in the validation group to further validate the predictive model. If the predictive model performs well in the validation group, further draw a nomogram to demonstrate.

Result
Univariate Analysis of Training Group
Chest CT scans of 75 patients diagnosed with adenocarcinoma and squamous cell carcinoma were downloaded from The Cancer Imaging Archive (TCIA). They will be included in the training group (CMB-LCA, CPTAC-LUAD, CPTAC-LSCC). After sorting out its 19 image features, it was found that the degree of enhancement was missing by 30%, so
it was removed. Therefore, the actual image features that entered the analysis were 18, which including location, spicule, lobulation, cavity, vacuolar sign, necrosis, pleural traction sign, vascular bundle sign, air bronchogram sign, calcification, distance from pulmonary hilum, atelectasis, pulmonary hilum and bronchial lymph nodes, mediastinal lymph nodes, interlobular septal thickening, pulmonary metastasis, adjacent structures invasion, pleural effusion. The glm package of R language was applied for logistic univariate analysis, and the results are shown in Table 1. Further extract features with P < 0.05 and P < 0.1 as follows: spicule, necrosis, air bronchogram sign, atelectasis, pulmonary hilum and bronchial lymph nodes (Table 2) and spicule, necrosis, pleural traction sign, air bronchogram sign, atelectasis, pulmonary hilum and bronchial lymph nodes (Table 3). Select variables with P < 0.1 for the next step of multi factor analysis.

Multivariate Analysis of Training Group
Selecting variables with univariate analysis result P < 0.1, and conducting multiple factor regression analysis using enter, forward, back, and both methods, it was found that the AIC values for forward, back, and both were the same, with AIC = 73.66533, which was lower than the AIC value of the enter method. Therefore, both stepwise regression method was selected for multivariate analysis, and the results are shown in Table 4. The final image features that enter modeling are spicule, necrosis, air bronchogram sign, atelectasis, pulmonary hilum and bronchial lymph nodes.

ROC Curves for Training Groups
We use ROC curves for model discrimination evaluation in training groups. First, draw the ROC curve in the training group with AUC = 0.887 and 95% confidence interval of 0.812–0.963, as shown in Figure 1.

### Table 1 Results of Logistic Univariate Analysis

<table>
<thead>
<tr>
<th>Characteristics</th>
<th>B</th>
<th>SE</th>
<th>OR</th>
<th>CI</th>
<th>Z</th>
<th>P</th>
</tr>
</thead>
<tbody>
<tr>
<td>Location</td>
<td>-0.065</td>
<td>0.18</td>
<td>0.94</td>
<td>0.66–1.33</td>
<td>-0.361</td>
<td>0.718</td>
</tr>
<tr>
<td>Spicule</td>
<td>-1.158</td>
<td>0.333</td>
<td>0.31</td>
<td>0.16–0.6</td>
<td>-3.478</td>
<td>0.001</td>
</tr>
<tr>
<td>Lobulation</td>
<td>-0.584</td>
<td>0.802</td>
<td>0.56</td>
<td>0.12–2.69</td>
<td>-0.728</td>
<td>0.467</td>
</tr>
<tr>
<td>Cavity</td>
<td>17.972</td>
<td>1769.258</td>
<td>63,817,219</td>
<td>0-Inf</td>
<td>0.01</td>
<td>0.992</td>
</tr>
<tr>
<td>Vesicles</td>
<td>-1.003</td>
<td>0.713</td>
<td>0.37</td>
<td>0.09–1.48</td>
<td>-1.408</td>
<td>0.159</td>
</tr>
<tr>
<td>Necrosis</td>
<td>0.999</td>
<td>0.481</td>
<td>2.71</td>
<td>1.06–6.97</td>
<td>2.077</td>
<td>0.038</td>
</tr>
<tr>
<td>Pleuraltraction</td>
<td>-0.875</td>
<td>0.479</td>
<td>0.42</td>
<td>0.16–1.07</td>
<td>-1.828</td>
<td>0.068</td>
</tr>
<tr>
<td>Vascular</td>
<td>-0.468</td>
<td>0.572</td>
<td>0.63</td>
<td>0.2–1.92</td>
<td>-0.819</td>
<td>0.413</td>
</tr>
<tr>
<td>Airbron</td>
<td>-1.121</td>
<td>0.491</td>
<td>0.33</td>
<td>0.12–0.85</td>
<td>-2.284</td>
<td>0.022</td>
</tr>
<tr>
<td>Calcification</td>
<td>-0.241</td>
<td>0.233</td>
<td>NA</td>
<td>NA–NA</td>
<td>-1.037</td>
<td>0.3</td>
</tr>
<tr>
<td>Distance</td>
<td>-0.096</td>
<td>0.093</td>
<td>0.91</td>
<td>0.76–1.09</td>
<td>-1.029</td>
<td>0.304</td>
</tr>
<tr>
<td>Atelectasis</td>
<td>1.872</td>
<td>0.704</td>
<td>6.5</td>
<td>1.64–25.83</td>
<td>2.66</td>
<td>0.008</td>
</tr>
<tr>
<td>Hilarlym</td>
<td>1.821</td>
<td>0.635</td>
<td>6.17</td>
<td>1.78–21.44</td>
<td>2.867</td>
<td>0.004</td>
</tr>
<tr>
<td>Mediastinallym</td>
<td>0.949</td>
<td>0.852</td>
<td>2.58</td>
<td>0.49–13.72</td>
<td>1.113</td>
<td>0.266</td>
</tr>
<tr>
<td>Interlobular</td>
<td>-0.464</td>
<td>0.476</td>
<td>0.63</td>
<td>0.25–1.6</td>
<td>-0.976</td>
<td>0.329</td>
</tr>
<tr>
<td>Metastasis</td>
<td>0.27</td>
<td>0.749</td>
<td>1.31</td>
<td>0.3–5.69</td>
<td>0.361</td>
<td>0.718</td>
</tr>
<tr>
<td>Invasion</td>
<td>-0.276</td>
<td>0.625</td>
<td>0.76</td>
<td>0.22–2.58</td>
<td>-0.442</td>
<td>0.659</td>
</tr>
<tr>
<td>Hydrothorax</td>
<td>-0.47</td>
<td>1.247</td>
<td>0.63</td>
<td>0.05–7.2</td>
<td>-0.377</td>
<td>0.706</td>
</tr>
</tbody>
</table>

### Table 2 Image Characteristics with P<0.05

<table>
<thead>
<tr>
<th>Characteristics</th>
<th>B</th>
<th>SE</th>
<th>OR</th>
<th>CI</th>
<th>Z</th>
<th>P</th>
</tr>
</thead>
<tbody>
<tr>
<td>Spicule</td>
<td>-1.158</td>
<td>0.333</td>
<td>0.31</td>
<td>0.16–0.6</td>
<td>-3.478</td>
<td>0.001</td>
</tr>
<tr>
<td>Necrosis</td>
<td>0.999</td>
<td>0.481</td>
<td>0.37</td>
<td>1.06–6.97</td>
<td>2.077</td>
<td>0.038</td>
</tr>
<tr>
<td>Airbron</td>
<td>-1.121</td>
<td>0.491</td>
<td>0.33</td>
<td>0.12–0.85</td>
<td>-2.284</td>
<td>0.022</td>
</tr>
<tr>
<td>Atelectasis</td>
<td>1.872</td>
<td>0.704</td>
<td>6.5</td>
<td>1.64–25.83</td>
<td>2.66</td>
<td>0.008</td>
</tr>
<tr>
<td>Hilarlym</td>
<td>1.821</td>
<td>0.635</td>
<td>6.17</td>
<td>1.78–21.44</td>
<td>2.867</td>
<td>0.004</td>
</tr>
</tbody>
</table>

Cancer Management and Research 2024:16
https://doi.org/10.2147/CMAR.S462951

DovePress

Liu et al

Powered by TCPDF (www.tcpdf.org)
Calibrate Validation Curves for Training Groups

We use the Calibrate curve to evaluate the accuracy of the model. First, draw the Calibrate curve (Figure 2) in the training group, which shows $p = 0.956$, showing that the accuracy of the model is extremely high.

DCA curves and confidence intervals for training groups

We use DCA curves to evaluate the reliability of the model in clinical practice. Draw the DCA curve and confidence interval (Figure 3A and B) in the training group using the RMDA package of R language, and search for literature\textsuperscript{21,22} to

### Table 3 Image Characteristics with $P<0.1$

<table>
<thead>
<tr>
<th>Characteristics</th>
<th>B</th>
<th>SE</th>
<th>OR</th>
<th>CI</th>
<th>Z</th>
<th>P</th>
</tr>
</thead>
<tbody>
<tr>
<td>Spicule</td>
<td>-1.158</td>
<td>0.333</td>
<td>0.31</td>
<td>0.16–0.6</td>
<td>-3.478</td>
<td>0.001</td>
</tr>
<tr>
<td>Necrosis</td>
<td>0.999</td>
<td>0.481</td>
<td>2.71</td>
<td>1.06–6.97</td>
<td>2.077</td>
<td>0.038</td>
</tr>
<tr>
<td>Pleuraltraction</td>
<td>-0.875</td>
<td>0.479</td>
<td>0.42</td>
<td>0.16–1.07</td>
<td>-1.828</td>
<td>0.068</td>
</tr>
<tr>
<td>Airbron</td>
<td>-1.121</td>
<td>0.491</td>
<td>0.33</td>
<td>0.12–0.85</td>
<td>-2.284</td>
<td>0.022</td>
</tr>
<tr>
<td>Atelectasis</td>
<td>1.872</td>
<td>0.704</td>
<td>6.5</td>
<td>1.64–25.83</td>
<td>2.66</td>
<td>0.008</td>
</tr>
<tr>
<td>Hilarlyn</td>
<td>1.821</td>
<td>0.635</td>
<td>6.17</td>
<td>1.78–21.44</td>
<td>2.867</td>
<td>0.004</td>
</tr>
</tbody>
</table>

### Table 4 Results of Logistic Multivariate Analysis

<table>
<thead>
<tr>
<th>Characteristics</th>
<th>B</th>
<th>SE</th>
<th>OR</th>
<th>CI</th>
<th>Z</th>
<th>P</th>
</tr>
</thead>
<tbody>
<tr>
<td>Spicule</td>
<td>-1.368</td>
<td>0.446</td>
<td>0.25</td>
<td>0.11–0.61</td>
<td>-3.065</td>
<td>0.002</td>
</tr>
<tr>
<td>Necrosis</td>
<td>1.209</td>
<td>0.661</td>
<td>3.35</td>
<td>0.92–12.24</td>
<td>1.829</td>
<td>0.067</td>
</tr>
<tr>
<td>Airbron</td>
<td>-1.312</td>
<td>0.695</td>
<td>0.27</td>
<td>0.07–1.05</td>
<td>-1.888</td>
<td>0.059</td>
</tr>
<tr>
<td>Atelectasis</td>
<td>2.42</td>
<td>0.91</td>
<td>11.25</td>
<td>1.89–66.93</td>
<td>2.66</td>
<td>0.008</td>
</tr>
<tr>
<td>Hilarlyn</td>
<td>2.865</td>
<td>0.898</td>
<td>17.56</td>
<td>3.02–102.05</td>
<td>3.19</td>
<td>0.001</td>
</tr>
</tbody>
</table>

Figure 1 Training group ROC curve.
determine a 20% incidence rate for squamous cell carcinoma and an 80% incidence rate for adenocarcinoma. Therefore, setting the parameter population validity = 0.2 shows good reliability.

**ROC Curves for Validation Groups**

We use ROC curves for model discrimination evaluation in validation groups. We use R language to draw the ROC curve in the validation group with AUC = 0.865 and 95% confidence interval of 0.756–0.975, as shown in Figure 4.

Figure 2 Training group Calibrate curve.

Figure 3 (A). Training group DCA curve. (B). confidence interval.
Calibrate Validation Curves for Validation Groups
We use the Calibrate curve to evaluate the accuracy of the model in validation groups. Using R language to draw a Calibrate curve (Figure 5) in the validation group, with \( P = 0.179 \). The accuracy of the model was not as good as in the training group, but the \( P \) value was greater than 0.05, showing good accuracy.

DCA Curves and Confidence Intervals for Validation Groups
We use DCA curves to evaluate the reliability of the model in clinical practice in validation groups. First, set the parameter population validity = 0.2. Then, in the validation group, a DCA curve and confidence interval (Figure 6A and B) were plotted, and the results showed that reliability was not as good as in the training group.

Clinical Impact Curve
We further evaluate the practicality of the model in clinical practice using clinical impact curves. By using the RMDA package in R language to draw a clinical impact curve (Figure 7), it can be seen that the solid red line is very close to the solid blue line, indicating good reliability of the model.

Nomogram Column Chart
Finally, we visualize the established model and use the rms package in R language to draw a Nomogram column chart (Figure 8).

Discussion
For patients who are difficult to undergo surgery or biopsy, or whose pathology cannot be decided after biopsy, distinguishing between lung squamous cell carcinoma and adenocarcinoma is a widespread problem that troubles clinical
doctors, and there is currently no good method. This study extracted image features from chest CT downloaded from TCIA and conducted modeling analysis with AUC = 0.887. A validation group was set up using clinical cases from our hospital, and an ROC curve with AUC = 0.865 was drawn in the validation group. The accuracy of the model was evaluated through the Calibration curve, the reliability of the model in clinical practice was evaluated through the DCA curve, and the clinical impact curve was further evaluated to evaluate the practicality of the model in clinical practice. Prove that the model has certain clinical application value.

Figure 5 Validation group Calibrate curve.

Figure 6 (A) Verification group DCA curve. (B) confidence interval.
There have been many literature reports on the application of imaging features for pathological prediction in the past.\(^5\)\(^–\)\(^13\),\(^17\)\(^–\)\(^33\) Han et al found that a model for distinguishing lung squamous cell carcinoma and lung adenocarcinoma was constructed using CT texture features, with an AUC of 0.803,\(^5\) Zhang et al established a model based on the variation of iodine concentration in enhanced CT, with an AUC of 0.871,\(^7\) Fukuma et al established a model based on the attenuation rate of enhanced CT, with an AUC of only 0.625,\(^8\) Jiang et al established a model based on the perfusion parameters of brain metastases, with an AUC of 0.845,\(^9\) Tomori et al established a model using a combination of PET-CT

Figure 7 Clinical impact curve.

Figure 8 Nomo column chart.
and CT imaging features, with an AUC of 0.92.\textsuperscript{11} Lin et al combined machine learning to establish a model with a maximum AUC 0.8.\textsuperscript{12} Tang et al combined machine learning to establish a model with a maximum AUC 0.87.\textsuperscript{13} But these examinations will increase the financial burden on patients, and some patients with allergic constitution are not suitable for enhanced examinations, while ordinary chest CT plain scans are more common and economical. Yue et al\textsuperscript{24} found that the imaging features of chest CT are a fast and widely available way to distinguish between lung squamous cell carcinoma and adenocarcinoma. However, these articles usually require the use of radiomics to extract data features from images in order to make judgments, and this process requires complex computer technologies such as image segmentation and Python software operation, which is still very difficult for clinical doctors. Therefore, I conducted a literature search based on how to apply chest CT to distinguish adenocarcinoma and squamous cell carcinoma. I did not find any research on using the imaging features of chest CT for modeling to help distinguish between lung squamous cell carcinoma and adenocarcinoma. This study downloaded chest CT from the TCIA database, reducing human intervention in cases and setting up a model for predicting lung adenocarcinoma and squamous cell carcinoma with AUC = 0.887. The AUC of the models developed in the above studies is 0.625 for the lowest, 0.92 for the highest, and mostly 0.8. The AUC of this study is 0.887, which belongs to the upper moderate level. Moreover, the data source of this study is the most ordinary chest CT, so the results of this study are reliable and practical.

This study has some limitations. Firstly, the number of cases is relatively small, which to some extent reduces the applicability of the model. Secondly, due to the high number of cases with enhanced CT, there is a lack of enhancement factors. Also, due to the selection of multiple influencing features in this study, no identical or similar articles were found, so it was not possible to compare the ROC with other articles. Finally, using chest CT alone to determine lung adenocarcinoma and squamous cell carcinoma itself has significant limitations. This study intends to analyze this only because it has found that many medical colleagues are interested in using chest CT to determine pathological subtypes and have engaged in heated discussions. However, there is no consensus. The image features adopted in this study are the basis mentioned by doctors on the Chinese Internet during their discussion. It is hoped that this study can also provide some judgmental basis for doctors. However, due to the small sample size of this study, it may result in biased results. Therefore, the five features proposed in this article require further clinical validation and future research improvements with larger sample sizes. Clinicians can strengthen the weights of five features, including spicule, necrosis, air bronchogram sign, atelectasis, pulmonary hilum and bronchial lymph nodes when using their own experience to make judgments, with a basis to preliminarily distinguish between lung squamous cell carcinoma and lung adenocarcinoma, and make personalized treatment choices with more confidence in treatment. In the future, we will accumulate more cases to enrich the research sample of this experiment, so as to make the results closer to the real world and improve the accuracy of evaluation.

**Conclusion**

It is currently very difficult to distinguish between lung adenocarcinoma and lung squamous cell carcinoma in clinical practice when pathology is difficult to obtain. Our study found that extracting specific features from regular chest CT scans is feasible for identifying lung adenocarcinoma and squamous cell carcinoma. The model we have established performs well in terms of discrimination, accuracy, reliability, and practicality. This provides more evidence for ordinary clinical doctors to distinguish between squamous cell carcinoma and adenocarcinoma, which is more reliable than relying solely on personal experience. In the future, we need more cases to improve the model.

**Data Sharing Statement**

The datasets generated during and/or analyzed during the current study are not publicly available due to privacy of patient information but are definitely available from the corresponding author if requested. All requests are greeted.

**Ethics Declarations and Consent to Participate**

This retrospective study was carried out using the opt-out method for the case series of our hospital. The study was approved by the Research Ethics Committee of the second hospital of Hebei Medical University (approval no 2024-R019) and was conducted in accordance with the 1964 Helsinki Declaration and its later amendments or comparable
ethical standards. Informed consent was waived by the Research Ethics Committee of the second hospital of Hebei Medical University due to the retrospective nature of the study, and all clinical data was anonymized during analysis to maintain the patient privacy. And all methods were performed in accordance with the relevant guidelines and regulations of the committee of The Second Hospital of Hebei Medical University.

Author Contributions
Chunmei Liu wrote the manuscript and all the figures, downloaded imaging, and read image features, Yuzheng He jointly analyze image data and extract image features, Jianmin Luo design the study. All authors made a significant contribution to the work reported, whether that is in the conception, study design, execution, acquisition of data, analysis and interpretation, or in all these areas; took part in drafting, revising or critically reviewing the article; gave final approval of the version to be published; have agreed on the journal to which the article has been submitted; and agree to be accountable for all aspects of the work.

Funding
This work was supported by the Health of Hebei Province Commission for medical scientific research funding of China (No. 20180396).

Disclosure
The authors declare that there is no conflict of interests.

References


