A CT-based integrated model for preoperative prediction of occult lymph node metastasis in early tongue cancer

Wei Han [Equal first author, 1, 2], Yingshu Wang [Equal first author, 3], Tao Li [2], Yuke Dong [2], Yanwei Dang [2], Liang He [1], Lianfang Xu [2], Yuhao Zhou [2], Yujie Li [2], Xudong Wang [Corresp. 1]

[1] Department of Maxillofacial and Otorhinolaryngological Oncology, Tianjin Medical University Cancer Institute and Hospital, National Clinical Research Center for Cancer, Tianjin Clinical Research Center for Cancer, Key Laboratory of Cancer Prevention and Therapy, Tianjin, China

[2] Department of Otolaryngology, Head and Neck Surgery, Zhengzhou Central Hospital Affiliated to Zhengzhou University, Zhengzhou, China

[3] Department of Radiology, the Affiliated Cancer Hospital of Zhengzhou University, Zhengzhou, China

Corresponding Author: Xudong Wang
Email address: wxd.1133@163.com

Background: Occult lymph node metastasis (OLNM) is an essential prognostic factor for early-stage tongue cancer (cT1-2N0M0) and a determinant of treatment decisions. This, accurate prediction of OLNM can significantly impact the clinical management and outcomes of tongue cancer patients. The aim of this study was to develop and validate a multiomics-based model to predict OLNM in patients with early-stage tongue cancer.

Methods: The data of 125 patients diagnosed with early-stage tongue cancer (cT1-2N0M0) who underwent primary surgical treatment and elective neck dissection was retrospectively analyzed. A total of 100 patients were randomly assigned to the training set and 25 to the testing set. The preoperative contrast-enhanced computed tomography (CT) and clinical data of these patients were collected. Radiomics features were extracted from the primary tumor region of interest (ROI) on CT images, and correlation analysis and LASSO were used to identify the most relevant features. A SVM classifier was constructed and compared with other machine learning algorithms. Using the same method, a clinical model was built and the peri-tumoral and intra-tumoral images were selected as the input for the deep learning model. Stacking ensemble technique was used to combine the multiple models. The predictive performance of the integrated model was evaluated on the basis of the accuracy, sensitivity, specificity, and the area under the receiver operating characteristic curve (AUC-ROC), and compared to expert assessment. Internal validation was performed using a stratified 5-fold cross-validation approach. **Results:** Of the 125 patients, 41 (32.8%) showed OLNM on postoperative pathological examination. The integrated model achieved higher predictive performance compared to the individual models, with accuracy of 84%, sensitivity 100%, specificity 76.5%, and AUC-ROC 0.949 (95% CI 0.870-1.000). In addition, the performance of the integrated model surpassed that of younger doctors and was comparable to the evaluation of experienced doctors.

Conclusions: The multiomics-based model can accurately predict OLNM in early-stage tongue cancer patients, and may serve as a valuable decision-making tool to determine the appropriate treatment and avoid unnecessary neck surgery in patients without OLNM.

A CT-based Integrated Model for Preoperative Prediction of Occult Lymph Node Metastasis in Early Tongue Cancer

Running title: Prediction of OLNM in Tongue cancer

Wei Han[1,2#], Yingshu Wang[3#], Tao Li[2], Yuke Dong[2], Yanwei Dang[2], Liang He[1], Lianfang Xu[2], Yuhao Zhou[2], Yujie Li[2], Xudong Wang[1*]

[1]Department of Maxillofacial and Otorhinolaryngological Oncology, Tianjin Medical University Cancer Institute and Hospital, National Clinical Research Center for Cancer, Tianjin Clinical Research Center for Cancer, Key Laboratory of Cancer Prevention and Therapy, Tianjin 300060, China

[2]Department of Otolaryngology, Head and Neck Surgery, Zhengzhou Central Hospital Affiliated to Zhengzhou University, Zhengzhou 450000, China

[3]Department of Radiology, the Affiliated Cancer Hospital of Zhengzhou University, Zhengzhou 450000, China

# Wei Han and Yingshu Wang contributed equally to this work.

* Corresponding Authors:
Xudong Wang,
Tianjin Medical University Cancer Institute and Hospital, National Clinical Research Center for Cancer, Tianjin Clinical Research Center for Cancer, Key Laboratory of Cancer Prevention and Therapy, Hexi District, Tiyuan North Ring, Huxi Road, Tianjin 300060, China
E-mail: wxd.1133@163.com

Abstract

Background Occult lymph node metastasis (OLNM) is an essential prognostic factor for early-stage tongue cancer (cT1-2N0M0) and a determinant of treatment decisions. This, accurate prediction of OLNM can significantly impact the clinical management and outcomes of tongue cancer patients. The aim of this study was to develop and validate a multiomics-based model to predict OLNM in patients with early-stage tongue cancer.

Methods: The data of 125 patients diagnosed with early-stage tongue cancer (cT1-2N0M0) who underwent primary surgical treatment and elective neck dissection was retrospectively analyzed. A total of 100 patients were randomly assigned to the training set and 25 to the testing set. The preoperative contrast-enhanced computed tomography (CT) and clinical data of these patients were collected. Radiomics features were extracted from the primary tumor region of interest (ROI) on CT images, and correlation analysis and LASSO were used to identify the most relevant features. A SVM classifier was constructed and compared with other machine learning

38   algorithms. Using the same method, a clinical model was built and the peri-tumoral and intra-
39   tumoral images were selected as the input for the deep learning model. Stacking ensemble
40   technique was used to combine the multiple models. The predictive performance of the
41   integrated model was evaluated on the basis of the accuracy, sensitivity, specificity, and the area
42   under the receiver operating characteristic curve (AUC-ROC), and compared to expert
43   assessment. Internal validation was performed using a stratified 5-fold cross-validation approach.
44   **Results:** Of the 125 patients, 41 (32.8%) showed OLNM on postoperative pathological
45   examination. The integrated model achieved higher predictive performance compared to the
46   individual models, with accuracy of 84%, sensitivity 100%, specificity 76.5%, and AUC-ROC
47   0.949 (95% CI 0.870-1.000). In addition, the performance of the integrated model surpassed that
48   of younger doctors and was comparable to the evaluation of experienced doctors.
49   **Conclusions:** The multiomics-based model can accurately predict OLNM in early-stage tongue
50   cancer patients, and may serve as a valuable decision-making tool to determine the appropriate
51   treatment and avoid unnecessary neck surgery in patients without OLNM.
52   **Keywords:** tongue cancer, deep learning, radiomics, computed tomography, lymph node
53   metastasis
54

55   ## Introduction

56   Tongue cancer affects the oral and maxillofacial area, and is more common in men due to higher
57   frequency of tobacco and alcohol consumption. Treatment for early-stage tongue cancer usually
58   involves surgery, radiation therapy, or a combination of both*(Coletta et al., 2020, Ansarin et al.,*
59   *2019, Pfister et al., 2020)*. Cervical lymph node metastasis is a major concern in the early stage
60   of tongue cancer, and can reduce patient survival by almost half. It is currently the primary
61   prognostic factor for tongue cancer*(Ren et al., 2022, Imai et al., 2017, Li et al., 2023)*. Occult
62   lymph node metastasis (OLNM) is diagnosed when pathological examination after neck
63   dissection indicates micrometastasis in the lymph nodes but the results of palpation, ultrasound,
64   CT, MRI, and other imaging tests are negative*(Baba et al., 2020)*. About 20%-40% of patients in
65   the early stages of oral cancer have OLNM*(Almangush et al., 2014, Greenberg et al., 2003)*. In
66   addition, patients without pathological lymph node metastasis (pN0) who undergo neck
67   dissection must bear unnecessary intraoperative risks and postoperative complications such as
68   shoulder dysfunction. On the other hand, lack of effective treatment in patients negative for
69   clinical lymph node metastasis (cN0) but with OLNM may lead to cervical lymph node
70   metastasis or even extracapsular spread during follow-up, resulting in worse prognosis*(Wang et*
71   *al., 2022)*. Therefore, early detection and accurate prediction of OLNM are critical for planning
72   the optimal treatment and improving patient outcomes.
73   In recent years, radiomics and deep learning techniques have helped identify early signs of
74   metastasis, and predict the likelihood of cancer recurrence*(Gillies et al., 2016, Leger et al., 2017,*
75   *Shen et al., 2017)*. Radiomics involves extraction of quantitative features from medical images to
76   identify subtle patterns that are not visible to the naked eye*(Aerts et al., 2014, van Griethuysen et*

77   *al., 2017, Lambin et al., 2012).* Deep learning refers to training deep neural networks, such as
78   conventional convolutional neural network (CNN), on large datasets and fine-tuning them on
79   smaller datasets to improve predictive ability*(Shen et al., 2017, Hosny et al., 2018).* It uses an
80   optimization algorithm called back-propagation to adjust its internal parameters for the best
81   image output prediction.
82   Combining radiomics with deep learning techniques can help improve the accuracy of the
83   prediction models for lymph node metastasis in early tongue cancer*(Gillies et al., 2016, Leger et*
84   *al., 2017, Shen et al., 2017).* Machine learning algorithms can identify patterns indicative of
85   metastasis by analyzing radiomics features, such as texture, shape and intensity*(Aerts et al., 2014,*
86   *van Griethuysen et al., 2017, Lambin et al., 2012).* Furthermore, deep transfer learning can
87   automatically extract more meaningful features from medical images to avoid massive
88   engineering of radiomics, leading to more accurate predictions in shorter duration, and achieve
89   good performance*(Shen et al., 2017, Hosny et al., 2018).*
90   Previous studies have utilized this approach on the intra-tumoral region without considering the
91   peritumoral area. However, there is evidence that peritumoral features can improve diagnostic
92   efficiency*(Huang et al., 2016, Sun et al., 2018, Zhou et al., 2018).* Huang et al demonstrated that
93   a radiomics signature including peritumoral features can predict disease-free survival in early-
94   stage non-small cell lung cancer patients with high accuracy*(Huang et al., 2016).* Sun et al used
95   peritumoral radiomics features to assess tumor-infiltrating CD8 cells and the response to
96   immunotherapy*(Sun et al., 2018).* In addition, Zhou et al generated a radiogenomics map for
97   non-small cell lung cancer to illustrate the relationship between molecular and imaging
98   phenotypes*(Zhou et al., 2018).*
99   The aim of this study was to combine radiomics and deep transfer learning to predict OLNM in
100 patients with early-stage tongue cancer on the basis of contrast-enhanced CT images. The
101 peritumoral region was incorporated into the radiomics-based model to improve the predictive
102 accuracy. In a clinical setting, this model can minimize decision-making errors by surgeons and
103 reduce the risk of unnecessary surgery. The workflow of the study is shown in Figure 1.
104

Materials and Methods

Patients

107 A retrospective study was conducted using data of patients diagnosed with tongue cancer
108 between 2015 and 2022. 125 patients were randomly divided into the training set (100 cases) and
109 test set (25 cases). The study was approved by the ethics committee of Zhengzhou Central
110 Hospital (202336) and conducted in line with the Declaration of Helsinki. Due to the
111 retrospective nature of the study, the need for informed consent was waived by the committee.
112 The inclusion criteria were as follows: (1) pathological diagnosis of tongue squamous cell
113 carcinoma, (2) planned surgery and elective neck dissection, (3) clinical diagnosis of cT1-
114 2N0M0 based on palpation and imaging tests, (4) contrast-enhanced CT scans of the head and
115 neck area performed within 30 days prior to surgery, and (5) availability of complete clinical
116 information. Patients with a history of previous malignancy, or with unidentifiable lesions or

117 artifacts-induced radiological interference on CT scans were excluded. The baseline
118 characteristics including gender, sex, T stage, tumor size, smoking history, and alcohol
119 consumption were retrieved from the hospital database.
120
121 **Region of interest segmentation and preprocessing**
122 Two radiologists blinded to the clinicopathological information of the patients independently
123 conducted tumor segmentation on the CT dataset using ITK-SNAP version 3.8.0
124 (http://www.itksnap.org). A senior radiologist verified and corrected all delineation tags. Since
125 the images came from two CT scanners with different parameters, resampling was done before
126 the features could be extracted from the ROI of the CT. The image space was normalized to
127 1x1x1. Inter-observer reliability and intra-observer repeatability of the extraction of radiomic
128 characteristics are generally evaluated using inter- and intra-class correlation coefficients (ICCs).
129 Features with ICC > 0.75 were retained for good consistency. The extracted image features were
130 then normalized by adjusting the distribution of all features to the same range. The data was
131 standardized according to the formula: $\mathbf{Z = (X-\mu) / \sigma}$, where $\boldsymbol{\mu}$ is the feature mean and $\boldsymbol{\sigma}$ is the
132 feature standard deviation.
133
134 **Feature extraction and selection**
135 The radiomics features were classified into three groups to describe different aspects of the
136 tumors: (I) geometry - three-dimensional form of the tumor; (II) intensity - first-order statistical
137 distribution of the voxel intensities within the tumor; (III) texture - the patterns, or second- and
138 high-order spatial distributions of the intensities. The neighborhood gray-tone difference matrix,
139 the gray-level co-occurrence matrix, gray-level run length matrix, gray level size zone matrix,
140 and other techniques were used to extract the texture features (Figure 2A).
141 The radiomic features with p values less than 0.05 in the Mann-Whitney U test were screened.
142 The correlation between these features was analyzed by Spearman's method, and those with a
143 correlation coefficient > 0.9 were selected. A greedy recursive deletion technique was used for
144 feature filtering, wherein the feature with the highest level of duplication in the current set is
145 removed each time in order to preserve the ability to display features as fully as possible. Finally,
146 115 features were retained and subjected to least absolute shrinkage and selection operator
147 (LASSO) regression with the scikit-learn package in Python using 10-fold cross validation with
148 minimum criteria. The features with non-zero coefficients in the regression model were used to
149 construct a radiomics signature, and a linear combination of retained features weighted by their
150 model coefficients was used to calculate the radiomics score for each patient (Figure 2B, C, D).
151 To construct deep transfer learning models, the most representative images were identified. The
152 images were processed through three steps: M1: Assuming that the image background would
153 create interference, the background pixels were removed and the primary lesion area was
154 extracted along with the area surrounding the tumor. M2: The largest ROI was selected from the
155 CT images and enlarged outward by 5 pixels. M3: In the 3D image, the tumor area was expanded

156 evenly by 2 voxels to obtain the peritumoral region, and the image with the largest ROI was
157 selected (Figure 2E).

158

159 **Development of different models**
160 Radiomics model: The final features were input into various machine learning models such as
161 SVM, KNN, RandomForest, ExtraTrees, XGBoost, LightGBM, MLP, and LR to construct a
162 radiomics model*(Cortes and Vapnik, 1995, Geurts et al., 2006, Chen and Guestrin, 2016, Ke et*
163 *al., 2017)*.
164 Clinical model: The preoperative clinical data included age, gender, T stage, tumor size, drinking
165 and smoking history. The same methodology as above was used to construct the clinical model.
166 Deep transfer learning model: Based on literature review and preliminary experimental testing,
167 ResNet50*(He et al., 2016)* was selected as the deep transfer learning model, and three datasets
168 were incorporated for training. The SGD optimizer was utilized with a learning rate of 0.001 and
169 100 epochs. A fully connected layer, followed by a dropout layer with a 0.5 drop rate, was
170 employed to prevent overfitting. The optimal dataset was selected for training, and the structured
171 data was extracted from the average pooling layer (i.e., the penultimate layer). The deep learning
172 model was then constructed using the same methodology as mentioned above. Finally, the three
173 meta-models were fused using stacking strategy to obtain the integrated model.

174

175 **Statistical analysis**
176 Student's t-test or the Mann-Whitney U test were used to compare continuous variables, and the
177 chi-squared test was used for categorical variables. All statistical tests were two-tailed and P <
178 0.05 was considered statistically significant. Python 3.7 was used for statistical analysis.

179

180 # Results

181 **Clinical data**
182 A total of 125 patients met the inclusion criteria, of which 100 were randomly assigned to the
183 training set and 25 to the test set. The clinical features of the patients are summarized in Table 1.

184

185 **Performance of the different models**
186 The LightGBM model demonstrated the best performance and was selected to construct the
187 radiomics signature, which achieved AUC values of 0.896 (95%CI 0.825 – 0.967) and 0.824
188 (95%CI 0.658 – 0.989) in the training and testing datasets respectively (Figure 3A). The
189 confusion matrix for LightGBM is shown in Figure 3B. The AUC of the clinical signature was
190 0.728. In the deep learning models, the performance of M3 was significantly higher than that of
191 the other datasets, with AUC of 0.826 in the training cohort and 0.674 in the test cohort (Figure
192 4). Subsequently, the deep transfer learning model trained on the M3 dataset was selected as the
193 deep signature.
194 Once the base models were trained, we used stacking to combine their predictions*(Džeroski and*
195 *Ženko, 2004)*. The dataset was split into training and testing sets, and the predictions generated

from each base model on the testing set were used as input features to train a meta-model for optimally combining the predictions of the three base models. Logistic regression was used as the meta-model due to its simplicity and interpretability. Our integrated model demonstrated improved performance compared to individual models, achieving an accuracy of 95%, sensitivity 100% and specificity 92.5% in the training dataset, and 84% accuracy, 100% sensitivity and 76.5% specificity in the testing dataset (Table 2). The AUC of the integrated model was 0.949 (95% CI: 0.870 – 1.000), which was higher than the AUCs of the radiomics (0.893), clinical (0.728), and deep learning models (0.798)(Figure 5A, B).

The performance of the integrated model was also evaluated by the DeLong test and decision curve analysis (DCA). The DeLong test indicated that the integrated model had a higher AUC compared to the individual models (nomogram vs clinical: 0.002; nomogram vs radiomics: 0.073; nomogram vs deep learning: 0.035), demonstrating its superior discriminative ability. The DCA further showed that our integrated model yielded a greater net benefit compared to other models across the majority of practical threshold probabilities (Figure 6). The nomogram of the integrated model is shown in Figure 7.

Comparison of the integrated model with clinical assessment

The performance of the integrated model was compared to that of clinical experts. Three clinicians with varying levels of experience in oral diseases, i.e., less than 5 years (D1), less than 10 years but longer than D1 (D2), and more than 15 years (D3), were asked to interpret imaging and clinical data. The decisions of the three clinicians were compared to that of the integrated model by plotting ROC curves. As shown in Figure 8, the integrated model achieved an AUC of 0.838 compared to 0.654, 0.768 and 0.893 of clinicians D1, D2 and D3 respectively. This suggests that there are differences between the model and expert evaluations.

Discussion

OLNM is an essential prognostic factor in early-stage tongue cancer (cT1-2N0M0) and influences the treatment decisions*(Kubo et al., 2022, Kwon, 2022)*. Thus, accurate prediction of OLNM can significantly impact clinical management and outcomes of tongue cancer patients*(Baba et al., 2019)*. In this study, we developed a decision-level fusion system for classifying CT images by integrating radiomics, clinical, and deep learning models using the stacking ensemble technique. A nomogram was employed to visualize the integrated prediction results and allow clinical interpretation.

The integrated model exhibited superior performance compared to the individual radiomics, clinical, and deep learning models in terms of the AUC values, which can be attributed to the combination of the complementary strengths of the three models that allowed diverse aspects of the data to be captured. For instance, the radiomics model extracted quantitative features from CT images and captured subtle patterns that might be overlooked by human observers, while the clinical model provided additional demographic and clinical information. We also applied deep learning to predict the presence of OLNM in cT1-2N0 patients*(Jin et al., 2021)*. The deep

236 learning model based on ResNet50 can directly analyze raw medical images, and learn complex
237 and high-level representations of the data*(Yu et al., 2020)*. However, the format of the raw
238 images is a matter of debate. While some researchers advocate for directly incorporating 3D
239 images into the model, others believe that the images should first undergo a flattening process
240 before being fed into the model*(Harmon et al., 2020, Calabrese et al., 2022, Tian et al., 2020,*
241 *Shi et al., 2022, Cho et al., 2022, Charles et al., 2017)*. However, direct use of 3D data for deep
242 learning training often requires significant computational resources, such as high-performance
243 GPUs and large memory capacities, which make the training and deployment processes
244 expensive and time-consuming. Furthermore, medical image datasets are often small and
245 imbalanced, which may lead to overfitting when 3D data is used directly for deep learning
246 training. Overfitting can cause the model to perform well with training data but poorly in real-
247 world applications due to reduced generalization. Therefore, we converted the 3D CT images to
248 2D images, which required determination of the feature extraction range due to relatively fewer
249 features. In order to extract more representative features, we adopted three cropping methods to
250 obtain suitable images as input for the deep learning model. The M1 dataset was first obtained by
251 setting pixel values of the background to zero. The largest cross-section from the 3D ROI was
252 selected and expanded by 5 pixels, resulting in dataset M2. Finally, the 3D ROI was evenly
253 expanded by 2 voxels around the tumor, and the largest cross-section was cropped to generate
254 the M3 dataset. Incorporating peri-tumoral information can prevent overfitting when applying
255 deep learning models to most small medical imaging datasets by increasing the number of
256 features. Among the two peri-tumoral processing methods, M3 uniformly expanded a fixed
257 length in 3D images, while M2 involved rectangular expansion in 2D images. Consequently, M3
258 encompassed more critical features than M2, eventually yielding superior results.
259 By combining the outputs of multiple base models, stacking leverages the strengths of each to
260 achieve better predictive accuracy*(Mostafaei et al., 2020)* and reduces overfitting by diversifying
261 the model ensemble, leading to more robust and generalizable results. We compared the
262 performance of our integrated model to the decision-making abilities of clinicians, and found that
263 the integrated model had superior accuracy. As already explained, this can be attributed to the
264 ability of the machine learning algorithms to detect subtle patterns and relationships in the data
265 that may not be easily discernible by human experts. Our findings indicate that a targeted multi-
266 omics integration model may help clinical practitioners, especially those with less experience,
267 attain a performance level similar to that of experts. Clinicians often need to analyze vast
268 amounts of information, including medical images, clinical records, and laboratory results,
269 within limited time, which can lead to cognitive overload and increase the likelihood of errors.
270 The integrated model can rapidly process and analyze large amount of data, thus reducing the
271 time needed for diagnosis and treatment planning. Another benefit of the integrated model is its
272 ability to continuously learn and adapt as new data becomes available. As medical knowledge
273 and clinical practice evolve, clinicians need to update their knowledge and decision-making
274 processes accordingly. However, this process can be time-consuming and challenging. The
275 integrated model can be retrained with new data, enabling it to stay up-to-date with the latest

276   advancements in the field and provide more accurate predictions. While the expertise of
277   clinicians is invaluable, personal biases and preferences can sometimes influence their decisions,
278   potentially leading to sub-optimal patient outcomes. By providing an objective, data-driven
279   approach, the integrated model can minimize the impact of subjectivity in decision-making,
280   leading to more consistent and optimal patient care. The nomogram incorporated the predictions
281   from the radiomics, clinical, and deep learning models, as well as the output from the meta-
282   model, to provide a comprehensive and intuitive representation of the decision-level fusion
283   process. This can enable more effective communication of the results to both clinical
284   practitioners and researchers, facilitating better understanding and application of the model in
285   real-world settings.
286   Nevertheless, there are several limitations to our study that should be acknowledged. First, the
287   sample size used for training and validation was relatively small, which may limit the
288   generalizability of our findings. Future studies should incorporate larger and more diverse patient
289   cohorts to further validate the performance of the integrated model. Second, the choice of the
290   meta-model in the stacking ensemble technique could potentially impact the results. In our study,
291   we employed logistic regression as the meta-model due to its simplicity and interpretability.
292   However, other meta-models, such as support vector machines or random forests, could be
293   explored in future to assess their impact on the performance of the integrated model.
294

Conclusion

296   In conclusion, integration of radiomics, clinical, and deep learning models using the stacking
297   ensemble technique can enable clinical decision-making based on imaging data. Furthermore, the
298   nomogram can visualize the integrated prediction results and thus enhance the interpretability
299   and applicability of the model. Despite the noted limitations, our integrated model could serve as
300   a valuable tool for improving diagnosis and treatment planning, eventually leading to better
301   patient outcomes. To the best of our knowledge, this is also the first application of an ensemble
302   model for predicting OLNM in early-stage tongue cance.
303

Ethics approval

305   The study was approved by the ethics committee of Zhengzhou Central Hospital (202336) and
306   conducted in line with the Declaration of Helsinki. Due to the retrospective nature of the study,
307   the need for informed consent was waived by the committee.

Competing Interests

309   The authors declare there are no competing interests.
310

Author Contributions

312 **Wei Han**: analyzed the data,conceived and designed, performed the experiments, drafted the
313 work, prepared figures and/or tables; **Yingshu Wang**: analyzed the data,conceived and
314 designed,performed the experiments, drafted the work, prepared figures and/or tables; **Tao Li:**
315 performed the experiments, analyzed the data, prepared figures and/or tables**;Yuke Dong:**
316 performed the experiments, analyzed the data, prepared figures and/or tables**;Yanwei Dang**:
317 performed the experiments, analyzed the data, prepared figures and/or tables; **Liang He**:
318 performed the experiments, prepared figures and/or tables; **Lianfang Xu**: performed the
319 experiments, prepared figures and/or tables; **Yuhao Zhou**: performed the experiments, prepared
320 figures and/or tables; **Yujie Li**: performed the experiments, analyzed the data, prepared figures
321 and/or tables;**Xudong Wang**: conceived and designed the experiments, authored or reviewed
322 drafts of the article.
323

Data Availability

325 The original contributions presented in the study are included in the article/Supplementary
326 Material.

327

Funding

329 There was no project specific funding for this work.

330

Acknowledgements

332 Not applicable.

333

References

335 **Aerts HJ, Velazquez ER, Leijenaar RT, Parmar C, Grossmann P, Carvalho S, Bussink J,**
336 **Monshouwer R, Haibe-Kains B, Rietveld D, Hoebers F, Rietbergen MM, Leemans CR,**
337 **Dekker A, Quackenbush J, Gillies RJ & Lambin P. 2014.** Decoding tumour phenotype by
338 noninvasive imaging using a quantitative radiomics approach. *Nat Commun* **5**(4006.
339 **Almangush A, Bello IO, Keski-Säntti H, Mäkinen LK, Kauppila JH, Pukkila M, Hagström**
340 **J, Laranne J, Tommola S, Nieminen O, Soini Y, Kosma VM, Koivunen P, Grénman R,**
341 **Leivo I & Salo T. 2014.** Depth of invasion, tumor budding, and worst pattern of invasion:
342 prognostic indicators in early-stage oral tongue cancer. *Head Neck* **36**(6): 811-8.
343 **Ansarin M, Bruschini R, Navach V, Giugliano G, Calabrese L, Chiesa F, Medina JE,**
344 **Kowalski LP & Shah JP. 2019.** Classification of GLOSSECTOMIES: Proposal for tongue
345 cancer resections. *Head Neck* **41**(3): 821-827.

346 **Baba A, Hashimoto K, Kayama R, Yamauchi H, Ikeda K & Ojiri H. 2020.** Radiological
347   approach for the newly incorporated T staging factor, depth of invasion (DOI), of the oral
348   tongue cancer in the 8th edition of American Joint Committee on Cancer (AJCC) staging
349   manual: assessment of the necessity for elective neck dissection. *Jpn J Radiol* **38**(9): 821-
350   832.
351 **Baba A, Okuyama Y, Ikeda K, Kozakai A, Suzuki T, Saito H, Ogane S, Yamazoe S,**
352   **Yamauchi H, Ogino N, Seto Y, Kobashi Y, Mogami T & Ojiri H. 2019.** Undetectability
353   of oral tongue cancer on magnetic resonance imaging; clinical significance as a predictor to
354   avoid unnecessary elective neck dissection in node negative patients. *Dentomaxillofac*
355   *Radiol* **48**(3): 20180272.
356 **Calabrese E, Rudie JD, Rauschecker AM, Villanueva-Meyer JE, Clarke JL, Solomon DA**
357   **& Cha S. 2022.** Combining radiomics and deep convolutional neural network features from
358   preoperative MRI for predicting clinically relevant genetic biomarkers in glioblastoma.
359   *Neurooncol Adv* **4**(1): vdac060.
360 Charles RQ, Su H, Kaichun M & Guibas LJ 2017. PointNet: Deep Learning on Point Sets for 3D
361   Classification and Segmentation. *2017 IEEE Conference on Computer Vision and Pattern*
362   *Recognition (CVPR)*. IEEE Computer Society.
363 Chen T & Guestrin C 2016. XGBoost: A Scalable Tree Boosting System. *Proceedings of the*
364   *22nd ACM SIGKDD International Conference on Knowledge Discovery and Data Mining*.
365   San Francisco, California, USA: Association for Computing Machinery.
366 **Cho HH, Kim H, Nam SY, Lee JE, Han BK, Ko EY, Choi JS, Park H & Ko ES. 2022.**
367   Measurement of Perfusion Heterogeneity within Tumor Habitats on Magnetic Resonance
368   Imaging and Its Association with Prognosis in Breast Cancer Patients. *Cancers (Basel)* **14**(8).
369 **Coletta RD, Yeudall WA & Salo T. 2020.** Grand Challenges in Oral Cancers. *Front Oral*
370   *Health* **1**(3.
371 **Cortes C & Vapnik V. 1995.** Support-vector networks. *Machine Learning* **20**(3): 273-297.
372 **Džeroski S & Ženko B. 2004.** Is Combining Classifiers with Stacking Better than Selecting the
373   Best One? *Machine Learning* **54**(3): 255-273.
374 **Geurts P, Ernst D & Wehenkel L. 2006.** Extremely randomized trees. *Machine Learning* **63**(1):
375   3-42.
376 **Gillies RJ, Kinahan PE & Hricak H. 2016.** Radiomics: Images Are More than Pictures, They
377   Are Data. *Radiology* **278**(2): 563-77.
378 **Greenberg JS, El Naggar AK, Mo V, Roberts D & Myers JN. 2003.** Disparity in pathologic
379   and clinical lymph node staging in oral tongue carcinoma. Implication for therapeutic
380   decision making. *Cancer* **98**(3): 508-15.
381 **Harmon SA, Sanford TH, Xu S, Turkbey EB, Roth H, Xu Z, Yang D, Myronenko A,**
382   **Anderson V, Amalou A, Blain M, Kassin M, Long D, Varble N, Walker SM, Bagci U,**
383   **Ierardi AM, Stellato E, Plensich GG, Franceschelli G, Girlando C, Irmici G, Labella D,**
384   **Hammoud D, Malayeri A, Jones E, Summers RM, Choyke PL, Xu D, Flores M,**
385   **Tamura K, Obinata H, Mori H, Patella F, Cariati M, Carrafiello G, An P, Wood BJ &**

386 **Turkbey B. 2020.** Artificial intelligence for the detection of COVID-19 pneumonia on chest
387 CT using multinational datasets. *Nat Commun* **11**(1): 4080.

388 He K, Zhang X, Ren S & Sun J. Deep Residual Learning for Image Recognition. 2016 IEEE
389 Conference on Computer Vision and Pattern Recognition (CVPR), 27-30 June 2016 2016.
390 770-778.

391 **Hosny A, Parmar C, Quackenbush J, Schwartz LH & Aerts H. 2018.** Artificial intelligence
392 in radiology. *Nat Rev Cancer* **18**(8): 500-510.

393 **Huang Y, Liu Z, He L, Chen X, Pan D, Ma Z, Liang C, Tian J & Liang C. 2016.** Radiomics
394 Signature: A Potential Biomarker for the Prediction of Disease-Free Survival in Early-Stage
395 (I or II) Non-Small Cell Lung Cancer. *Radiology* **281**(3): 947-957.

396 **Imai T, Satoh I, Matsumoto K, Asada Y, Yamazaki T, Morita S, Saijo S, Okubo JI,**
397 **Wakamori S, Saijo S & Matsuura K. 2017.** Retrospective observational study of occult
398 cervical lymph-node metastasis in T1N0 tongue cancer. *Jpn J Clin Oncol* **47**(2): 130-136.

399 **Jin Y, Li M, Zhao Y, Huang C, Liu S, Liu S, Wu M & Song B. 2021.** Computed
400 Tomography-Based Radiomics for Preoperative Prediction of Tumor Deposits in Rectal
401 Cancer. *Front Oncol* **11**(710248.

402 Ke G, Meng Q, Finley T, Wang T, Chen W, Ma W, Ye Q & Liu T-Y 2017. LightGBM: a highly
403 efficient gradient boosting decision tree. *Proceedings of the 31st International Conference*
404 *on Neural Information Processing Systems.* Long Beach, California, USA: Curran
405 Associates Inc.

406 **Kubo K, Kawahara D, Murakami Y, Takeuchi Y, Katsuta T, Imano N, Nishibuchi I, Saito**
407 **A, Konishi M, Kakimoto N, Yoshioka Y, Toratani S, Ono S, Ueda T, Takeno S &**
408 **Nagata Y. 2022.** Development of a radiomics and machine learning model for predicting
409 occult cervical lymph node metastasis in patients with tongue cancer. *Oral Surg Oral Med*
410 *Oral Pathol Oral Radiol* **134**(1): 93-101.

411 **Kwon M. 2022.** Prediction of Occult Lymph Node Metastasis in Early Tongue Cancer. *Clin Exp*
412 *Otorhinolaryngol* **15**(4): 297-298.

413 **Lambin P, Rios-Velazquez E, Leijenaar R, Carvalho S, Van Stiphout RG, Granton P,**
414 **Zegers CM, Gillies R, Boellard R, Dekker A & Aerts HJ. 2012.** Radiomics: extracting
415 more information from medical images using advanced feature analysis. *Eur J Cancer* **48**(4):
416 441-6.

417 **Leger S, Zwanenburg A, Pilz K, Lohaus F, Linge A, Zöphel K, Kotzerke J, Schreiber A,**
418 **Tinhofer I, Budach V, Sak A, Stuschke M, Balermpas P, Rödel C, Ganswindt U, Belka**
419 **C, Pigorsch S, Combs SE, Mönnich D, Zips D, Krause M, Baumann M, Troost EGC,**
420 **Löck S & Richter C. 2017.** A comparative study of machine learning methods for time-to-
421 event survival data for radiomics risk modelling. *Sci Rep* **7**(1): 13206.

422 **Li C, Zhu L, Guo Y, Ji T & Ren Z. 2023.** Three-dimensional assessment of tongue cancer
423 prognosis by preoperative MRI. *Oral Dis* **29**(5): 2006-2011.

424 **Mostafaei S, Abdollahi H, Kazempour Dehkordi S, Shiri I, Razzaghdoust A, Zoljalali**
425 **Moghaddam SH, Saadipoor A, Koosha F, Cheraghi S & Mahdavi SR. 2020.** CT imaging

426   markers to improve radiation toxicity prediction in prostate cancer radiotherapy by stacking
427   regression algorithm. *Radiol Med* **125**(1): 87-97.

428   **Pfister DG, Spencer S, Adelstein D, Adkins D, Anzai Y, Brizel DM, Bruce JY, Busse PM,**
429   **Caudell JJ, Cmelak AJ, Colevas AD, Eisele DW, Fenton M, Foote RL, Galloway T,**
430   **Gillison ML, Haddad RI, Hicks WL, Hitchcock YJ, Jimeno A, Leizman D, Maghami E,**
431   **Mell LK, Mittal BB, Pinto HA, Ridge JA, Rocco JW, Rodriguez CP, Shah JP, Weber**
432   **RS, Weinstein G, Witek M, Worden F, Yom SS, Zhen W, Burns JL & Darlow SD. 2020.**
433   Head and Neck Cancers, Version 2.2020, NCCN Clinical Practice Guidelines in Oncology. *J*
434   *Natl Compr Canc Netw* **18**(7): 873-898.

435   **Ren J, Yuan Y & Tao X. 2022.** Histogram analysis of diffusion-weighted imaging and dynamic
436   contrast-enhanced MRI for predicting occult lymph node metastasis in early-stage oral
437   tongue squamous cell carcinoma. *Eur Radiol* **32**(4): 2739-2747.

438   **Shen D, Wu G & Suk HI. 2017.** Deep Learning in Medical Image Analysis. *Annu Rev Biomed*
439   *Eng* **19**(221-248.

440   **Shi J, Dong Y, Jiang W, Qin F, Wang X, Cui L, Liu Y, Jin Y, Luo Y & Jiang X. 2022.** MRI-
441   based peritumoral radiomics analysis for preoperative prediction of lymph node metastasis in
442   early-stage cervical cancer: A multi-center study. *Magn Reson Imaging* **88**(1-8.

443   **Sun R, Limkin EJ, Vakalopoulou M, Dercle L, Champiat S, Han SR, Verlingue L, Brandao**
444   **D, Lancia A, Ammari S, Hollebecque A, Scoazec JY, Marabelle A, Massard C, Soria**
445   **JC, Robert C, Paragios N, Deutsch E & Ferté C. 2018.** A radiomics approach to assess
446   tumour-infiltrating CD8 cells and response to anti-PD-1 or anti-PD-L1 immunotherapy: an
447   imaging biomarker, retrospective multicohort study. *Lancet Oncol* **19**(9): 1180-1191.

448   Tian Z, Jia Y, Men X & Sun Z 2020. 3DCNN for Pulmonary Nodule Segmentation and
449   Classification.

450   **Van Griethuysen JJM, Fedorov A, Parmar C, Hosny A, Aucoin N, Narayan V, Beets-Tan**
451   **RGH, Fillion-Robin JC, Pieper S & Aerts H. 2017.** Computational Radiomics System to
452   Decode the Radiographic Phenotype. *Cancer Res* **77**(21): e104-e107.

453   **Wang F, Tan R, Feng K, Hu J, Zhuang Z, Wang C, Hou J & Liu X. 2022.** Magnetic
454   Resonance Imaging-Based Radiomics Features Associated with Depth of Invasion Predicted
455   Lymph Node Metastasis and Prognosis in Tongue Cancer. *J Magn Reson Imaging* **56**(1):
456   196-209.

457   **Yu J, Deng Y, Liu T, Zhou J, Jia X, Xiao T, Zhou S, Li J, Guo Y, Wang Y, Zhou J &**
458   **Chang C. 2020.** Lymph node metastasis prediction of papillary thyroid carcinoma based on
459   transfer learning radiomics. *Nat Commun* **11**(1): 4807.

460   **Zhou M, Leung A, Echegaray S, Gentles A, Shrager JB, Jensen KC, Berry GJ, Plevritis SK,**
461   **Rubin DL, Napel S & Gevaert O. 2018.** Non-Small Cell Lung Cancer Radiogenomics Map
462   Identifies Relationships between Molecular and Imaging Phenotypes with Prognostic
463   Implications. *Radiology* **286**(1): 307-315.

464

465   **Figure legends**

466 **Figure 1.** The workflow of this study.

467 **Figure 2.** Features selection process. (A). Number and ratio of handcrafted features; Nonzero

468 coefficients were selected to establish the Rad-score with a LASSO logistic regression model.

469 (B). Coefficients and (D). MSE (mean standard error) of 10 folds validation; (C). The histogram

470 of the Rad-score based on the selected features. (E). Deep learning datasets based on different

471 methods (M1, M2, M3)

472 **Figure 3.** Development of radiomics model. (A). The ROC curves of the radiomics model; (B).

473 The confusion matrix of predicting lymph node metastasis

474 **Figure 4.** The ROC curves of M3.

475 **Figure 5.** The ROC curves of the clinical model, radiomics model, Deep learning model and

476 combined model in the training cohort(A) and validation cohort (B). The integrated model

477 demonstrated significantly higher AUCs in the training and validation cohorts (AUCs of 0.989

478 and 0.949) than the clinical model (AUCs of 0.815 and 0.728), the radiomics model (AUCs of

479 0.896 and 0.893) and the deep learning model (AUCs of 0.980 and 0.798)

480 **Figure 6.** The DCA of the clinical model, radiomics model, Deep learning model and combined

481 model.

482 **Figure 7.** The development of the nomogram for clinical use.

483 **Figure 8.** The ROC curves of the integrated model and diagnostic results of each radiologist in

484 the test cohort. The integrated model reached a high AUCs (AUC: 0.838), doctor1 acquired

485 0.654 of AUC, the AUC of doctor2 was 0.768 and doctor3 demonstrated the highest with 0.893.

486

487 **Table Legends**

488 **Table 1.** The detailed clinical characteristics of the patients.

489 **Table 2.** The performance of different models.

490

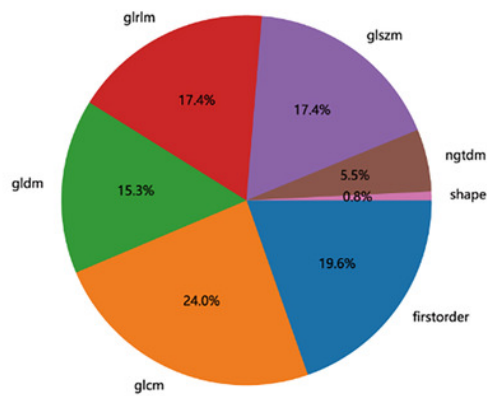
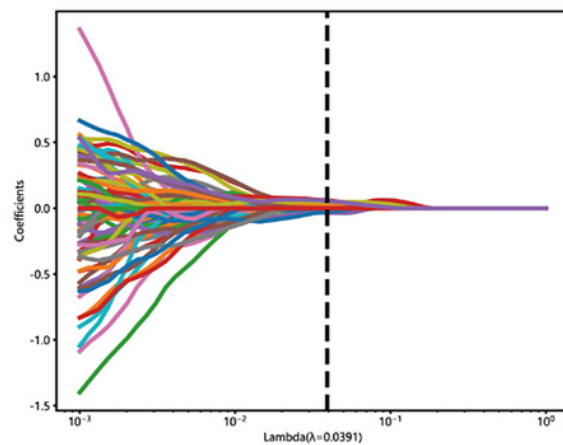
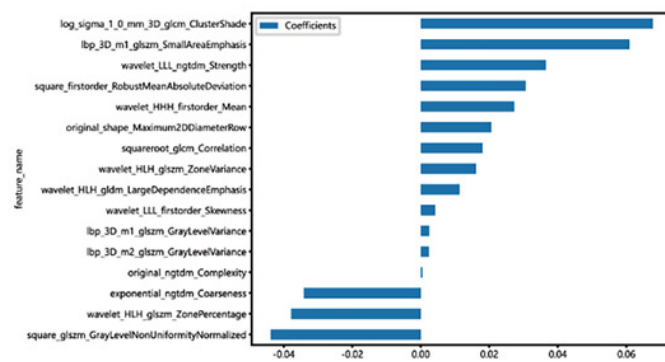
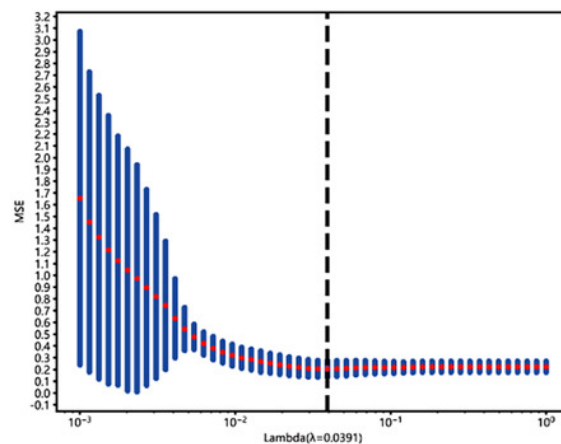
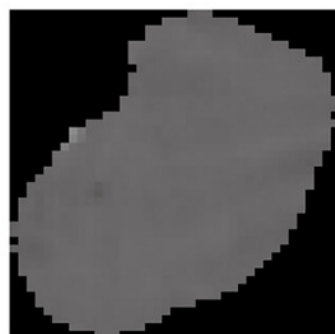
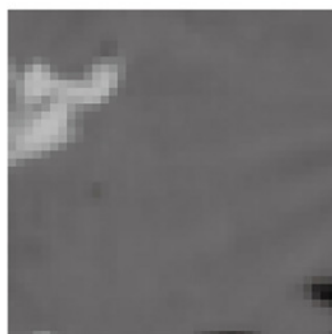
Figure 1

The workflow of this study.

Figure 2

Features selection process.

Number and ratio of handcrafted features; Nonzero coefficients were selected to establish the Rad-score with a LASSO logistic regression model. (B). Coefficients and (D). MSE (mean standard error) of 10 folds validation; (C). The histogram of the Rad-score based on the selected features. (E). Deep learning datasets based on different methods (M1, M2, M3)

A

B

C

D

E

M1

M2

M3

Figure 3

Development of radiomics model.

(A). The ROC curves of the radiomics model; (B). The confusion matrix of predicting lymph node metastasis

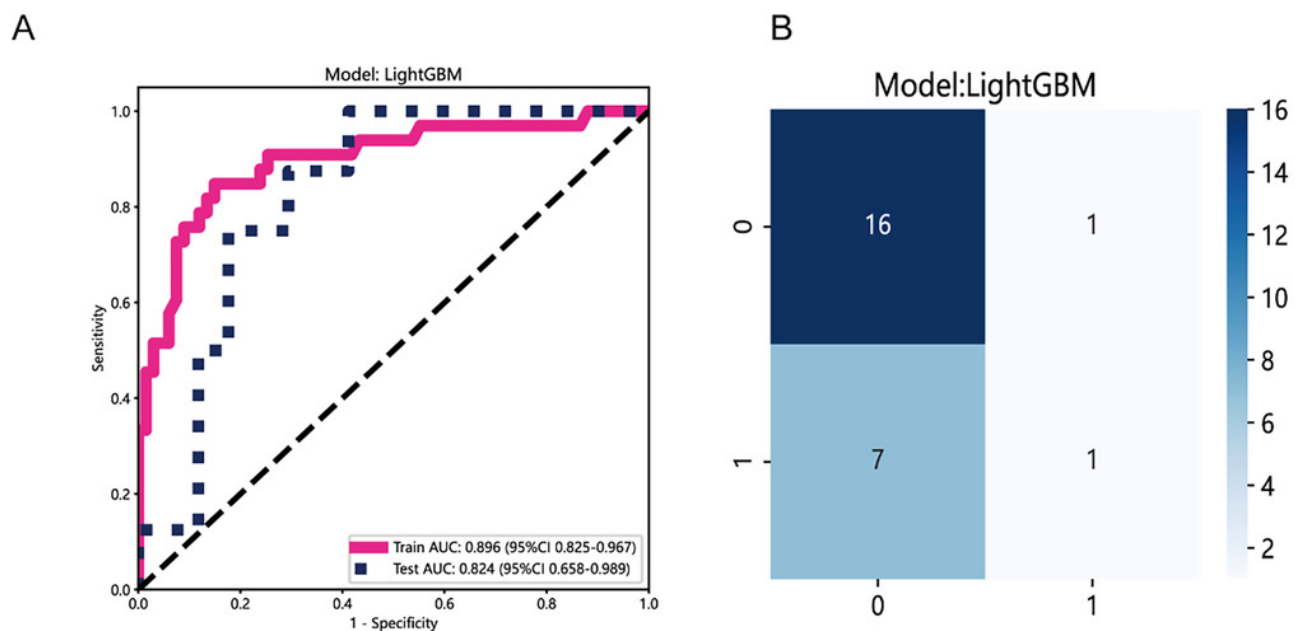



Figure 4

The ROC curves of M3.

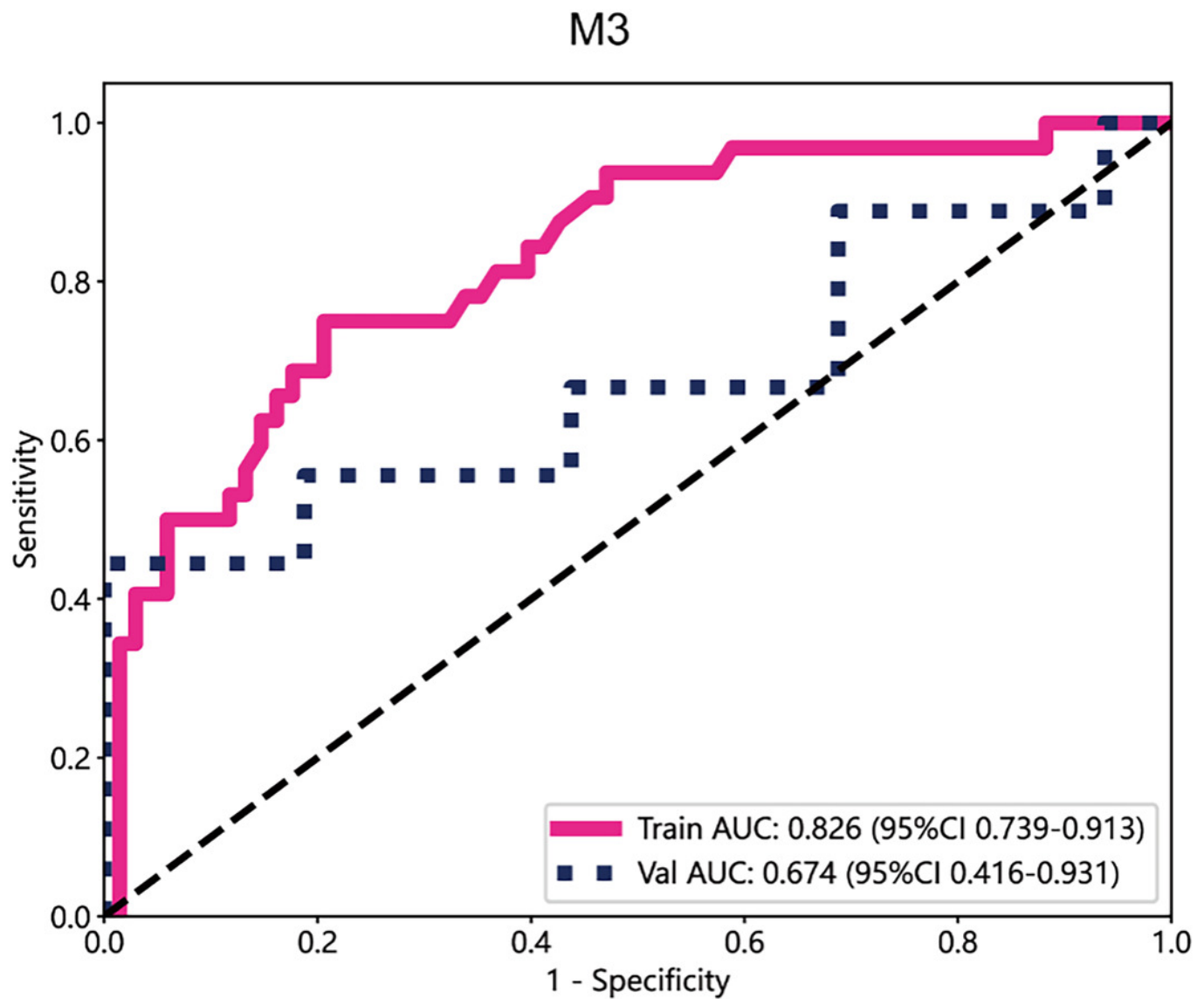



Figure 5

The ROC curves of the clinical model, radiomics model, Deep learning model and combined model in the training cohort

(A) and validation cohort (B). The integrated model demonstrated significantly higher AUCs in the training and validation cohorts (AUCs of 0.989 and 0.949) than the clinical model (AUCs of 0.815 and 0.728), the radiomics model (AUCs of 0.896 and 0.893) and the deep learning model (AUCs of 0.980 and 0.798)

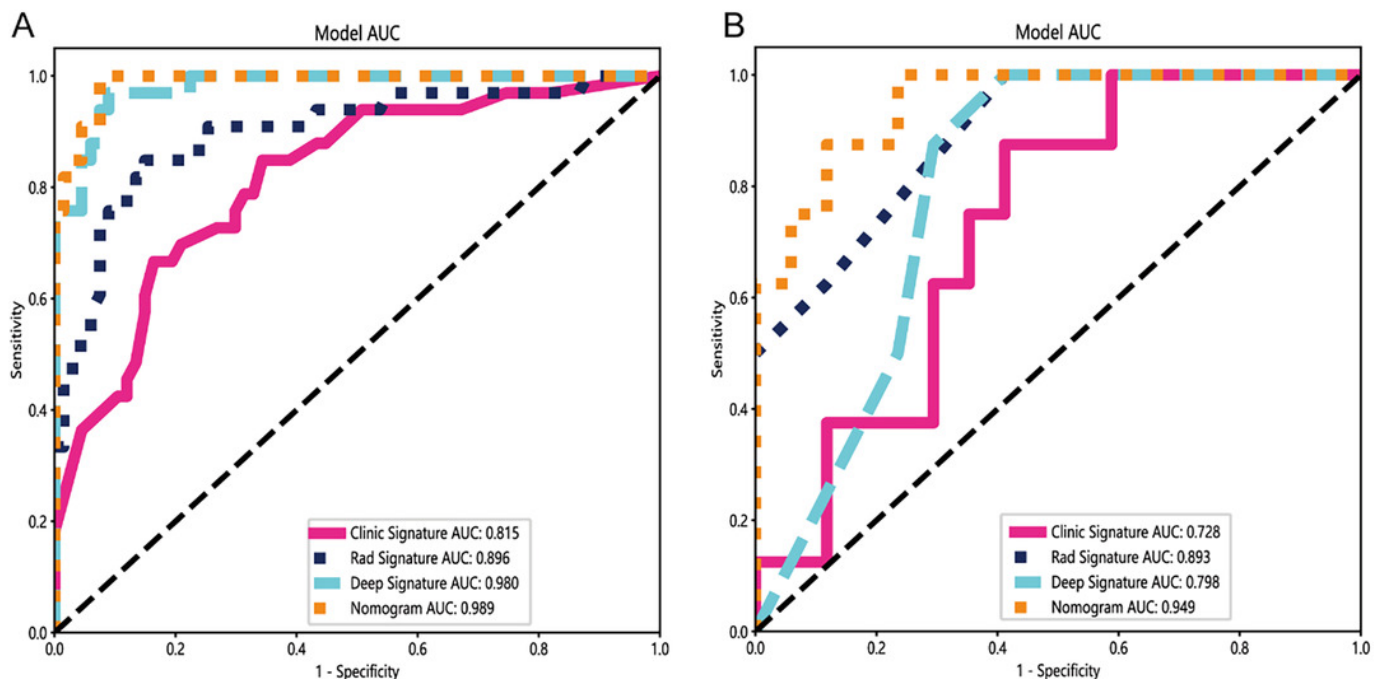



Figure 6

The DCA of the clinical model, radiomics model, Deep learning model and combined model.

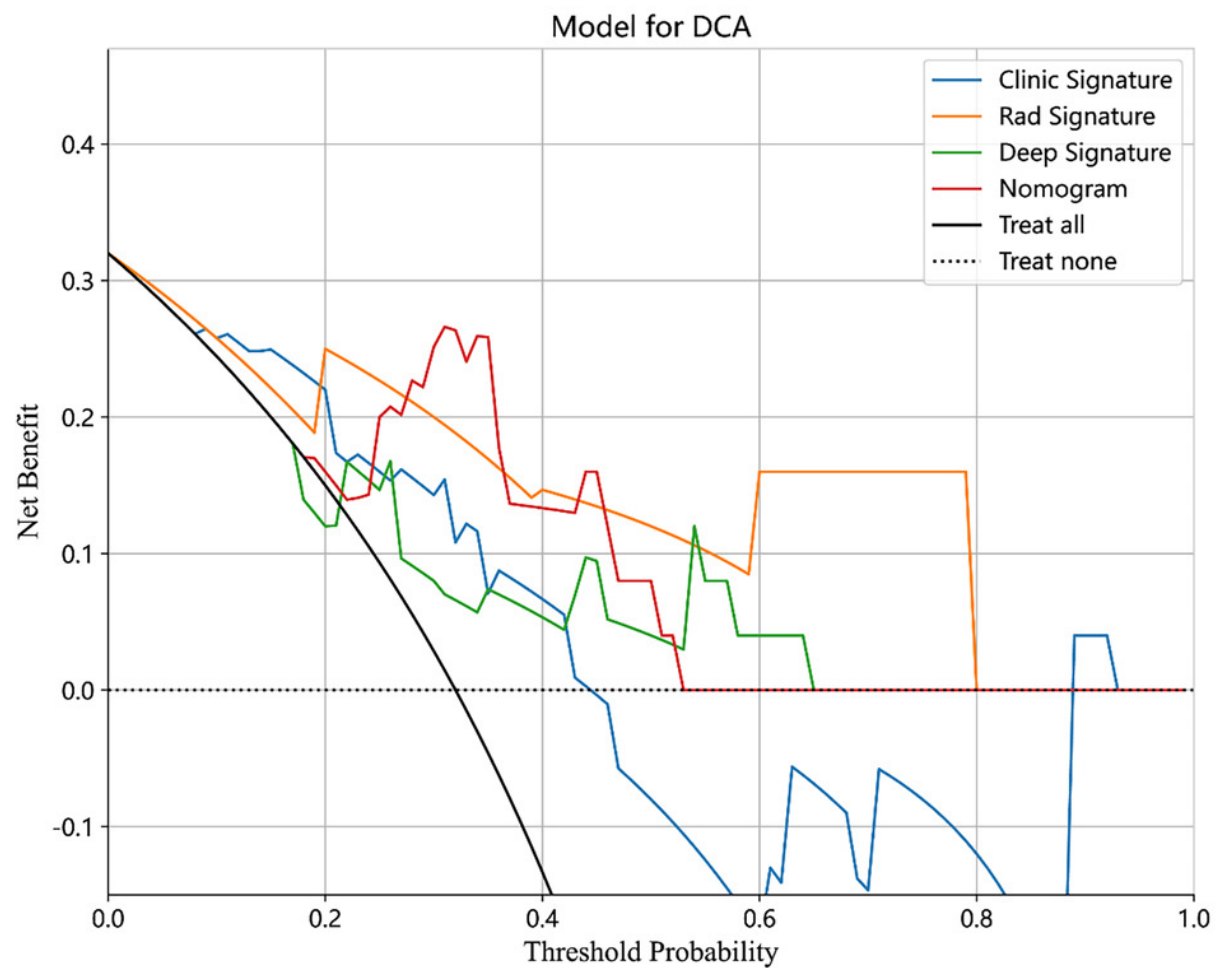



Figure 7

The development of the nomogram for clinical use.

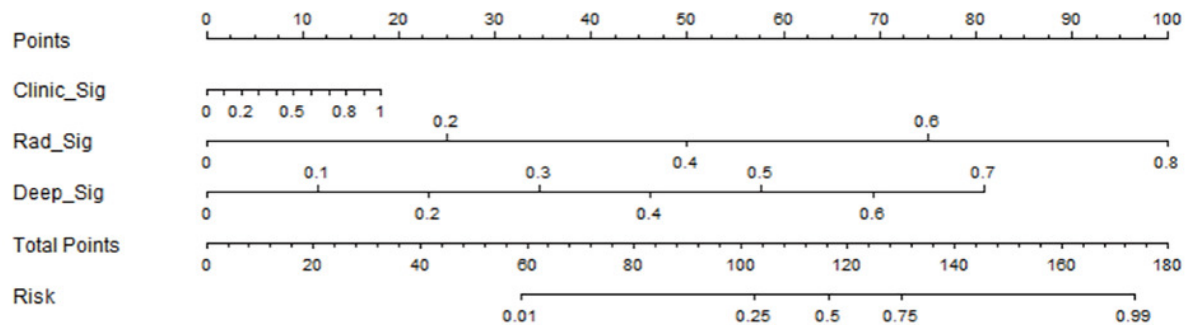
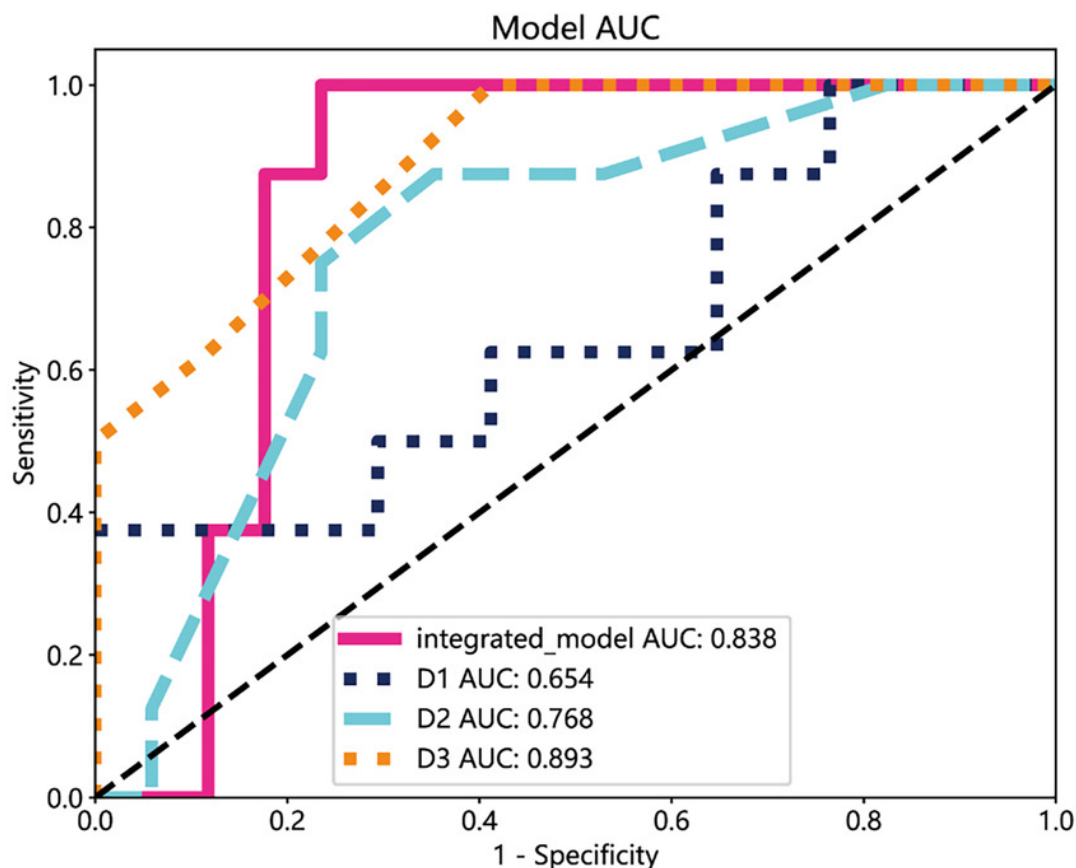



Figure 8

The ROC curves of the integrated model and diagnostic results of each radiologist in the test cohort.

The integrated model reached a high AUCs (AUC: 0.838), doctor1 acquired 0.654 of AUC, the AUC of doctor2 was 0.768 and doctor3 demonstrated the highest with 0.893.

The detailed clinical characteristics of the patients.

1 **Table 1.** The detailed clinical characteristics of the patients.

	train-label=ALL	train-label=0	train-label=1	pvalue	test-label=ALL	test-label=0	test-label=1	pvalue
age	58.54±12.65	59.61±10.54	56.36±16.09	0.229041	54.92±15.39	56.18±16.05	52.25±14.54	0.563049
tumorsize	2.10±0.91	2.03±1.01	2.24±0.68	0.286474	1.99±0.73	1.74±0.68	2.53±0.53	0.008355
ender				0.876726				0.479802
male	58(58.00)	38(56.72)	20(60.61)		18(72.00)	11(64.71)	7(87.50)	
female	42(42.00)	29(43.28)	13(39.39)		7(28.00)	6(35.29)	1(12.50)	
Tstage				0.056452				0.479802
1	29(29.00)	24(35.82)	5(15.15)		7(28.00)	6(35.29)	1(12.50)	
2	71(71.00)	43(64.18)	28(84.85)		18(72.00)	11(64.71)	7(87.50)	
drink				0.859987				1
no	67(67.00)	44(65.67)	23(69.70)		14(56.00)	10(58.82)	4(50.00)	
yes	33(33.00)	23(34.33)	10(30.30)		11(44.00)	7(41.18)	4(50.00)	
smoke				0.993163				0.378313
no	56(56.00)	37(55.22)	19(57.58)		11(44.00)	9(52.94)	2(25.00)	
yes	44(44.00)	30(44.78)	14(42.42)		14(56.00)	8(47.06)	6(75.00)	

The performance of different models.

1 **Table 2.** The performance of different models.

	Accuracy	AUC	95% CI	Sensitivity	Specificity	Recall	F1	Cohort
Clinic Sig	0.720	0.815	0.725 - 0.904	0.848	0.657	0.848	0.667	Train
Rad Sig	0.850	0.896	0.825 - 0.967	0.848	0.851	0.848	0.789	Train
Deep Sig	0.930	0.980	0.960 - 1.000	0.970	0.910	0.970	0.901	Train
Nomogram	0.950	0.989	0.976 - 1.000	1.000	0.925	1.000	0.930	Train
Clinic Sig	0.680	0.728	0.526 - 0.930	0.875	0.588	0.875	0.636	Test
Rad Sig	0.720	0.893	0.777 - 1.000	1.000	0.588	1.000	0.696	Test
Deep Sig	0.720	0.798	0.625 - 0.971	1.000	0.625	1.000	0.696	Test
Nomogram	0.840	0.949	0.870 - 1.000	1.000	0.765	1.000	0.800	Test