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. Therefore, accurate prediction of OLNM can significantly impact the clinical management and outcomes of patients with tongue cancer. 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 were retrospectively analyzed. A total of 100 patients were randomly assigned to the training set and 25 to the test set. The preoperative contrast-enhanced computed tomography (CT) and clinical data on these patients were collected. Radiomics features were extracted from the primary tumor as the region of interest (ROI) on CT images, and correlation analysis and the least absolute shrinkage and selection operator (LASSO) method were used to identify the most relevant features. A support vector machine (SVM) classifier was constructed and compared with other machine learning algorithms. With 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. The stacking ensemble technique was used to combine the multiple models. The predictive performance of the integrated model was evaluated for accuracy, sensitivity, specificity, and the area under the receiver operating characteristic curve (AUC-ROC), and compared with 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 with the individual models, with an accuracy of 84%, a sensitivity of 100%, a specificity of 76.5%, and an AUC-ROC of 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 patients with early-stage tongue cancer, 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 Author:

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. Therefore, accurate prediction of OLNM can significantly impact the clinical management and outcomes of patients

29   with tongue cancer. The aim of this study was to develop and validate a multiomics-based model
30   to predict OLNM in patients with early-stage tongue cancer.

31   **Methods**: The data of 125 patients diagnosed with early-stage tongue cancer (cT1-2N0M0) who
32   underwent primary surgical treatment and elective neck dissection were retrospectively analyzed.
33   A total of 100 patients were randomly assigned to the training set and 25 to the test set. The
34   preoperative contrast-enhanced computed tomography (CT) and clinical data on these patients
35   were collected. Radiomics features were extracted from the primary tumor as the region of
36   interest (ROI) on CT images, and correlation analysis and the least absolute shrinkage and
37   selection operator (LASSO) method were used to identify the most relevant features. A support
38   vector machine (SVM) classifier was constructed and compared with other machine learning
39   algorithms. With the same method, a clinical model was built and the peri-tumoral and intra-
40   tumoral images were selected as the input for the deep learning model. The stacking ensemble
41   technique was used to combine the multiple models. The predictive performance of the
42   integrated model was evaluated for accuracy, sensitivity, specificity, and the area under the
43   receiver operating characteristic curve (AUC-ROC), and compared with expert assessment.
44   Internal validation was performed using a stratified 5-fold cross-validation approach.

45   **Results:** Of the 125 patients, 41 (32.8%) showed OLNM on postoperative pathological
46   examination. The integrated model achieved higher predictive performance compared with the
47   individual models, with an accuracy of 84%, a sensitivity of 100%, a specificity of 76.5%, and
48   an AUC-ROC of 0.949 (95% CI 0.870-1.000). In addition, the performance of the integrated
49   model surpassed that of younger doctors and was comparable to the evaluation of experienced
50   doctors.

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

54

55   **Keywords:** tongue cancer, deep learning, radiomics, computed tomography, lymph node
56   metastasis

57

Introduction

59   Tongue cancer affects the oral and maxillofacial area and is more common in men due to the
60   higher frequency of tobacco and alcohol consumption. Treatment for early-stage tongue cancer
61   usually involves surgery, radiation therapy, or a combination of both*(Coletta et al., 2020,*
62   *Ansarin et al., 2019, Pfister et al., 2020).* Cervical lymph node metastasis is a major concern in

63   the early stage of tongue cancer and can reduce patient survival by almost half. It is currently the
64   primary prognostic factor for tongue cancer*(Ren et al., 2022, Imai et al., 2017, Li et al., 2023)*.
65   Occult lymph node metastasis (OLNM) is diagnosed with the presence of micrometastasis in the
66   lymph nodes through pathological examination after neck dissection, before which the results of
67   palpation, ultrasound, CT, MRI, and other imaging tests are negative *(Baba et al., 2020)*. About
68   40%-60% of the patients in the early stage of oral cancer have OLNM*(Yuan, Ren & Tao, 2021,*
69   *Ren, Yuan & Tao, 2022, Doll et al., 2022)*. In addition, patients without pathological lymph node
70   metastasis (pN0) who undergo neck dissection have to bear unnecessary intraoperative risks and
71   postoperative complications such as shoulder dysfunction. On the other hand, a lack of effective
72   treatment in patients without clinical lymph node metastasis (cN0) but with OLNM may lead to
73   cervical lymph node metastasis or even extracapsular spread during follow-up, resulting in a
74   worse prognosis*(Wang et al., 2022)*. Therefore, early detection and accurate prediction of OLNM
75   are critical for determining the optimal treatment and improving patient outcomes.

76   In recent years, radiomics and deep learning have helped identify early signs of metastasis and
77   predict the likelihood of cancer recurrence*(Ferreira-Junior et al., 2020, Jalalifar et al., 2022,*
78   *Bang et al., 2023)*. Radiomics involves the extraction of quantitative features from medical
79   images to identify subtle patterns that are not visible to the naked eye*(Guiot et al., 2022, Currie*
80   *& Rohren, 2022, Currie, Hawk & Rohren, 2023)*. Deep learning refers to training deep neural
81   networks, such as conventional convolutional neural network (CNN) on large datasets, and fine-
82   tuning them on smaller datasets to improve the predictive ability*(Hosny et al., 2018, Liu et al.,*
83   *2020)*. It uses an optimization algorithm called back-propagation to adjust its internal parameters
84   for the best image output prediction.

85   Combining radiomics with deep learning can help improve the accuracy of the prediction models
86   for lymph node metastasis in early tongue cancer*(Gillies et al., 2016, Leger et al., 2017)*.
87   Machine learning algorithms can identify patterns indicative of metastasis by analyzing
88   radiomics features, such as texture, shape, and intensity*(Tang et al., 2022)*. Furthermore, deep
89   transfer learning can automatically extract more meaningful features from medical images to
90   avoid massive engineering of radiomics for more accurate predictions in a shorter duration,
91   achieving good performance*(Hosny et al., 2018)*.

92   Previous studies have used this approach on the intratumoral region without considering the
93   peritumoral area. However, there is evidence that peritumoral features can improve diagnostic
94   efficiency*(Zhou et al., 2018, Zhang et al., 2022, Zhao et al., 2023)*. Zhang et al demonstrated that
95   time-serial CT-based radiomics signature integrating intra- and peri-tumoral features offered the
96   potential to predict progression-free survival for lung adenocarcinoma patients treated with
97   epidermal growth factor receptor-tyrosine kinase inhibitors*(Zhang et al., 2022)*. Zhao et al.
98   developed and validated a nomogram based on intratumoral and peritumoral radiomics
99   signatures to predict distant metastasis-free survival in patients with locally advanced rectal
100 cancer before the treatment with neoadjuvant chemoradiotherapy*(Zhao et al., 2023)*. In addition,

101 Zhou et al. generated a radiogenomics map for non-small cell lung cancer to illustrate the
102 relationship between molecular and imaging phenotypes*(Zhou et al., 2018)*.

103 This study aimed to combine radiomics and deep transfer learning to predict OLNM in patients
104 with early-stage tongue cancer based on contrast-enhanced CT images. The peritumoral region
105 was incorporated into the radiomics-based model to improve the predictive accuracy. In clinical
106 settings, this model can minimize decision-making errors by surgeons and reduce the risk of
107 unnecessary surgery. The workflow of the study is shown in Figure 1.

108

Materials and Methods

Participants

111 A retrospective study was conducted using data on patients diagnosed with tongue cancer
112 between 2015 and 2022 at the Tianjin Medical University Cancer Institute and Hospital and
113 Zhengzhou Central Hospital Affiliated to Zhengzhou University. A total of 125 patients were
114 randomly divided into the training set (100 cases) and the test set (25 cases). The study was
115 approved by the ethics committee of Zhengzhou Central Hospital (202336) and conducted in line
116 with the Declaration of Helsinki. Due to the retrospective nature of the study, the need for
117 informed consent was waived by the committee.

118 The inclusion criteria were as follows: (1) The participants must have a pathological diagnosis of
119 tongue squamous cell carcinoma, (2) and were treated with planned surgery and elective neck
120 dissection, (3) with a clinical diagnosis of cT1-2N0M0 based on palpation and imaging tests. (4)
121 The contrast-enhanced CT scans of the head and neck area were performed within 30 days prior
122 to surgery, and (5) the complete clinical information was available. Patients with a history of
123 previous malignancy, or with unidentifiable lesions or artifacts-induced radiological interference
124 on CT scans were excluded. The baseline characteristics including gender, sex, T stage, tumor
125 size, smoking history, and alcohol consumption were retrieved from the databases of the
126 hospitals.

127

Region of interest (ROI) segmentation and preprocessing

129 Two radiologists blinded to the clinicopathological information of the patients independently
130 conducted tumor segmentation on the CT dataset using ITK-SNAP version 3.8.0
131 (http://www.itksnap.org). A senior radiologist verified all delineation tags and corrected any
132 errors. Since the images came from two CT scanners with different parameters, resampling was
133 done before the features could be extracted from the ROI of the CT. The image space was
134 normalized to 1x1x1. The inter-observer reliability and intra-observer repeatability of the

135 extraction of radiomic characteristics were generally evaluated using inter- and intra-class
136 correlation coefficients (ICCs). Features with ICC > 0.75 were retained for good consistency.
137 The extracted image features were then normalized by adjusting the distribution of all features to
138 the same range. The data were standardized according to the formula: $\mathbf{Z = (X-\mu) / \sigma}$, where $\boldsymbol{\mu}$ is
139 the feature's mean and $\boldsymbol{\sigma}$ is the feature's standard deviation.

140

141 **Feature extraction and selection**

142 The radiomics features were classified based on geometry, intensity, and texture to describe
143 different aspects of the tumors, specifically from the following aspects: (I) geometry - a three-
144 dimensional form of the tumor; (II) intensity - first-order statistical distribution of the voxel
145 intensities within the tumor; (III) texture - the patterns, or second- and high-order spatial
146 distributions of the intensities. The neighborhood gray-tone difference matrix, gray-level co-
147 occurrence matrix, gray-level run length matrix, gray level size zone matrix, and other
148 techniques were used to extract the texture features (Figure 2A).

149 The radiomics features with p values less than 0.05 in the Mann-Whitney U test were screened.
150 The correlation between these features was analyzed by Spearman's method, and those with a
151 correlation coefficient > 0.9 were selected. A greedy recursive deletion strategy was used for
152 feature filtering, and the feature with the highest level of duplication in the current set was
153 removed each time to preserve the ability to display features as fully as possible. Finally, 115
154 features were retained and subjected to least absolute shrinkage and selection operator (LASSO)
155 regression with the scikit-learn package in Python using 10-fold cross validation with minimum
156 criteria. The features with non-zero coefficients in the regression model were used to construct
157 the radiomics signature, and a linear combination of retained features weighted by their model
158 coefficients was used to calculate the radiomics score for each patient (Figures 2B, C, D).

159 To construct deep transfer learning models, the most representative images were identified. The
160 images were processed through three steps: M1: Assuming that the image background would
161 create interference, the background pixels were removed and the primary lesion area was
162 extracted along with the area surrounding the tumor. M2: The largest ROI was selected from the
163 CT images and enlarged by 5 pixels. M3: In the 3D image, the tumor area was expanded evenly
164 by 2 voxels to obtain the peritumoral region, and the image with the largest ROI was selected
165 (Figure 2E).

166

167 **Development of different models**

168 Radiomics model: The final features were input into various machine learning models such as
169 support vector machine (SVM), k-nearest neighbor (kNN), random forest, extra trees, extreme

170   gradient boosting (XGBoost), light gradient boosting machine (LightGBM), multilayer
171   perceptron (MLP), and logistic regression (LR) to construct the radiomics model (Cortes and
172   Vapnik, 1995, Geurts et al., 2006, Chen and Guestrin, 2016, Ke et al., 2017).

173   Clinical model: The preoperative clinical data included age, gender, T stage, tumor size, drinking
174   and smoking history. The same methodology described above was used to construct the clinical
175   model.

176   Deep transfer learning model: Based on literature review and preliminary experimental testing,
177   ResNet50*(He et al., 2016)* was selected as the deep transfer learning model, and three datasets
178   were incorporated for training. The stochastic gradient descent (SGD) optimizer was adopted
179   with a learning rate of 0.001 and 100 epochs. A fully connected layer, followed by a dropout
180   layer with a 0.5 drop rate, was employed to prevent overfitting. The optimal dataset was selected
181   for training, and the structured data were extracted from the average pooling layer (i.e., the
182   penultimate layer). The deep learning model was then constructed using the same methodology
183   mentioned above. Finally, the three base models were integrated using the stacking strategy to
184   obtain the integrated model.

185

186   **Statistical analysis**

187   Student's t-test or the Mann-Whitney U test was used to compare continuous variables, and the
188   chi-squared test was used for categorical variables. All statistical tests were two-tailed and $P <$
189   0.05 was considered statistically significant. Python 3.7 was used for statistical analysis.

190

191   # Results

192   **Clinical data**

193   A total of 125 patients met the inclusion criteria, of which 100 were randomly assigned to the
194   training set and 25 to the test set. The clinical features of the patients are summarized in Table 1.

195

196   **Performance of the radiomics model**

197   The LightGBM model demonstrated the best performance and was selected to construct the
198   radiomics signature, with the area under curve (AUC) values of 0.896 (95%CI 0.825 – 0.967)
199   and 0.824 (95%CI 0.658 – 0.989) in the training and testing datasets, respectively (Figure 3A).
200   The confusion matrix for LightGBM is shown in Figure 3B. The AUC of the clinical signature
201   was 0.728.

202

Performance of the deep learning model

204 In the deep learning models, the performance of M3 was significantly higher than that of other
205 datasets, with an AUC value of 0.826 in the training cohort and 0.674 in the test cohort (Figure
206 4). Subsequently, the deep transfer learning model trained on the M3 dataset was selected as the
207 deep signature.

208

The integrated model performed better than a single model

210 Once the base models were trained, we used stacking to combine their predictions (Džeroski and
211 Ženko, 2004). The dataset was split into the training and test sets, and the predictions generated
212 from each base model on the test set were used as input features to train a meta-model for
213 optimally combining the predictions of the three base models. Logistic regression was used as
214 the meta-model for its simplicity and interpretability. Our integrated model demonstrated
215 improved performance compared with individual models, achieving an accuracy of 95%, a
216 sensitivity of 100%, and a specificity of 92.5% in the training dataset, and 84% accuracy, 100%
217 sensitivity, and 76.5% specificity in the test dataset (Table 2). The AUC of the integrated model
218 was 0.949 (95% CI: 0.870 – 1.000), which was higher than the AUCs of the radiomics (0.893),
219 clinical (0.728), and deep learning models (0.798)(Figures 5A, B).

220

**Compared with a single model, the integrated model showed superior discriminative ability
and produced larger net benefits**

223 The performance of the integrated model was also evaluated by the DeLong test and decision
224 curve analysis (DCA). The DeLong test indicated that the integrated model had a higher AUC
225 compared with the individual models (nomogram vs. clinical: 0.002; nomogram vs. radiomics:
226 0.073; nomogram vs. deep learning: 0.035), demonstrating the superior discriminative ability of
227 the integrated model. The DCA further showed that our integrated model yielded a greater net
228 benefit compared with other models across the majority of practical threshold probabilities
229 (Figure 6). The nomogram of the integrated model is shown in Figure 7.

230

231 Comparison of the integrated model with clinical assessment

232 The performance of the integrated model was compared with that of clinical experts. Three
233 clinicians with varying levels of experience in oral diseases, i.e., less than 5 years (D1), less than
234 10 years but longer than D1 (D2), and more than 15 years (D3), were asked to interpret the
235 imaging and clinical data. The decision of the three clinicians was compared with that of the

236   integrated model by plotting ROC curves. As shown in Figure 8, the integrated model achieved
237   an AUC of 0.838 compared with 0.654, 0.768, and 0.893 for clinicians D1, D2, and D3,
238   respectively, suggesting differences between the model and the expert evaluation.

239

Discussion

241   OLNM is an essential prognostic factor in early-stage tongue cancer (cT1-2N0M0) and
242   influences treatment decisions*(Kubo et al., 2022, Kwon, 2022)*. Thus, accurate prediction of
243   OLNM can significantly impact clinical management and outcomes of tongue cancer
244   patients*(Baba et al., 2019)*. In this study, we developed a decision-level fusion system for
245   classifying CT images by integrating radiomics, clinical, and deep learning models with the
246   stacking ensemble technique. A nomogram was employed to visualize the integrated prediction
247   results and allow clinical interpretation. To our knowledge, this is also the first time that an
248   integrated model capturing the features of CT images around the tumor has been used to predict
249   OLNM in early-stage tongue cancer.

250   The integrated model exhibited superior performance compared with the individual radiomics,
251   clinical, and deep learning models in terms of the AUC values, and this can be attributed to the
252   combination of the complementary strengths of the three models that allowed diverse aspects of
253   the data to be captured. For instance, the radiomics model allowed quantitative features to be
254   extracted from CT images and subtle patterns that might be overlooked by human observers to
255   be captured, while the clinical model provided additional demographic and clinical information.
256   We also applied deep learning to predict the presence of OLNM in cT1-2N0 patients*(Jin et al.,
257   2021)*. The deep learning model based on ResNet50 can directly analyze raw medical images and
258   learn complex and high-level representations of the data*(Yu et al., 2020)*. However, the format of
259   the raw images is a matter of debate. While some researchers advocate for directly incorporating
260   3D images into the model, others believe that the images should first undergo a flattening
261   process before being fed into the model*(Harmon et al., 2020, Calabrese et al., 2022, Tian et al.,
262   2020, Shi et al., 2022, Cho et al., 2022, Charles et al., 2017)*. However, direct use of 3D data for
263   deep learning training often requires significant computational resources, such as high-
264   performance Graphic Processing Units (GPUs) and large memory capacities, making the training
265   and deployment processes expensive and time-consuming. Furthermore, medical image datasets
266   are often small and imbalanced, possibly leading to overfitting when 3D data are directly used
267   for deep learning training. Overfitting can cause the model to perform well with the training data
268   but underperform in real-world applications due to reduced generalization. Therefore, we
269   converted the 3D CT images to 2D images, during which determination of the feature extraction
270   range was required due to relatively fewer features. To extract more representative features, we
271   adopted three cropping methods to obtain suitable images as the input for the deep learning
272   model. The M1 dataset was first obtained by setting pixel values of the background to zero. Then,

273 the largest cross-section from the 3D ROI was selected and expanded by 5 pixels to obtain
274 dataset M2. Finally, the 3D ROI was evenly expanded by 2 voxels around the tumor, and the
275 largest cross-section was cropped to generate the M3 dataset. Incorporating peri-tumoral
276 information can prevent overfitting when deep learning models are applied to most small
277 medical imaging datasets by increasing the number of features. Among the two peri-tumoral
278 processing methods, M3 uniformly expanded a fixed length in 3D images, while M2 involved
279 rectangular expansion in 2D images. Consequently, M3 encompassed more critical features than
280 M2, eventually yielding superior results.

281 By combining the outputs of multiple base models, stacking leverages the strengths of each to
282 achieve better predictive accuracy*(Mostafaei et al., 2020)* and reduces overfitting by diversifying
283 the model ensemble, leading to more robust and generalizable results. We compared the
284 performance of our integrated model with the decision-making abilities of clinicians and found
285 that the integrated model had superior accuracy. As already explained, this can be attributed to
286 the ability of machine learning algorithms to detect subtle patterns and relationships in the data
287 that may not be easily discernible by human experts. Our findings indicate that a targeted
288 multiomics integration model may help clinical practitioners, especially those with less
289 experience, attain a performance level similar to that of experts. Clinicians often need to analyze
290 vast amounts of information, including medical images, clinical records, and laboratory results,
291 within a limited time, which can lead to cognitive overload and increase the likelihood of errors.
292 The integrated model can rapidly process and analyze a large number of data, thus reducing the
293 time needed for diagnosis and treatment planning. Another benefit of the integrated model is its
294 ability to continuously learn and adapt new data that become available. As medical knowledge
295 and clinical practice evolve, clinicians need to update their knowledge and decision-making
296 process accordingly. However, this process can be time-consuming and challenging. The
297 integrated model can be retrained with new data, enabling it to stay up-to-date with the latest
298 advancements in the field and provide more accurate predictions. While the expertise of
299 clinicians is invaluable, personal biases and preferences can sometimes influence their decisions,
300 potentially leading to sub-optimal patient outcomes. By providing an objective, data-driven
301 approach, the integrated model can minimize the impact of subjectivity in decision-making,
302 leading to more consistent and optimal patient care. The nomogram incorporated the predictions
303 from the radiomics, clinical, and deep learning models, as well as the output from the meta-
304 model, to provide a comprehensive and intuitive representation of the decision-level fusion
305 process. This can enable more effective communication of the results to both clinical
306 practitioners and researchers, facilitating better understanding and application of the model in
307 real-world settings.

308 In clinical practice, clinical data and CT image features of patients with early tongue cancer can
309 be collected before surgery. Through objective data analysis and processing, the integrated
310 model can be used to predict OLNM more accurately in a short period, and unnecessary neck
311 surgery can be minimized for patients with negative OLNM. During the follow-up period, new

312 data can be continuously collected for prediction, patients with missed diagnosis or OLNM are
313 detected as early as possible, and lymph node dissection can be completed in time to effectively
314 improve the prognosis. For patients with positive OLNM, cervical surgery should be completed
315 simultaneously to reduce the risk of cervical lymph node metastasis and spread during follow-up.
316 The integrated model is beneficial to the standardized and individualized diagnosis and treatment
317 of early tongue cancer and improves the curative effect and prognosis.

318 However, several limitations to our study should be acknowledged. First, the sample size used
319 for training and validation was relatively small, and this may limit the generalizability of our
320 findings. Future studies should include larger and more diverse patient cohorts to further
321 optimize the comprehensive model, and prospective studies are needed to verify its efficacy.
322 Second, the choice of the meta-model in the stacking ensemble technique could potentially
323 impact the results. In our study, we employed logistic regression as the meta-model due to its
324 simplicity and interpretability. However, other meta-models, such as SVM or random forest,
325 could be explored in the future to assess their impact on the performance of the integrated model.

326

Conclusion

328 The multiomics-based model demonstrated high accuracy, sensitivity, and specificity for
329 predicting OLNM in patients with early-stage tongue cancer. Despite the noted limitations, our
330 integrated model could serve as a valuable tool for improving diagnosis and treatment planning,
331 eventually leading to better patient outcomes.

332

Ethics approval

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

337

Competing Interests

339 The authors declare that there are no competing interests.

340

Author Contributions

Wei Han analyzed the data,conceived and designed the work, performed the experiments, drafted the work, and prepared figures and/or tables; **Yingshu Wang** analyzed the data,conceived and designed the work, performed the experiments, drafted the work, and prepared figures and/or tables; **Tao Li** performed the experiments, analyzed the data, and prepared figures and/or tables; **Yuke Dong** performed the experiments, analyzed the data, and prepared figures and/or tables; **Yanwei Dang** performed the experiments, analyzed the data, and prepared figures and/or tables; **Liang He** performed the experiments and prepared figures and/or tables; **Lianfang Xu** performed the experiments and prepared figures and/or tables; **Yuhao Zhou** performed the experiments and prepared figures and/or tables; **Yujie Li** performed the experiments, analyzed the data, and prepared figures and/or tables; **Xudong Wang** conceived and designed the experiments, and authored or reviewed drafts of the article.

Data Availability

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

Funding

There was no project specific funding for this work.

Acknowledgements

Not applicable.

References

Bang C, Bernard G, Le WT, Lalonde A, Kadoury S, Bahig H. 2023. Artificial intelligence to predict outcomes of head and neck radiotherapy. Clinical and Translational Radiation Oncology 39:100590.

368 **Currie G, Hawk KE, Rohren E. 2023.** The transformational potential of molecular radiomics.
369 Journal of Medical Radiation Sciences 70:77–88.

370 **Currie G, Rohren E. 2022.** The deep radiomic analytics pipeline. Veterinary Radiology &
371 Ultrasound 63:889–896.

372 **Doll C, Mrosk F, Wuester J, Runge A-S, Neumann F, Rubarth K, Heiland M, Kreutzer K,**
373 **Voss J, Raguse J-D, Koerdt S. 2022.** Pattern of cervical lymph node metastases in squamous
374 cell carcinoma of the upper oral cavity – How to manage the neck. Oral Oncology 130:105898.

375 **Ferreira-Junior JR, Koenigkam-Santos M, Magalhães Tenório AP, Faleiros MC, Garcia**
376 **Cipriano FE, Fabro AT, Näppi J, Yoshida H, De Azevedo-Marques PM. 2020.** CT-based
377 radiomics for prediction of histologic subtype and metastatic disease in primary malignant lung
378 neoplasms. International Journal of Computer Assisted Radiology and Surgery 15:163–172.

379 **Guiot J, Vaidyanathan A, Deprez L, Zerka F, Danthine D, Frix A, Lambin P, Bottari F,**
380 **Tsoutzidis N, Miraglio B, Walsh S, Vos W, Hustinx R, Ferreira M, Lovinfosse P, Leijenaar**
381 **RTH. 2022.** A review in radiomics: Making personalized medicine a reality via routine imaging.
382 Medicinal Research Reviews 42:426–440.

383 **Hosny A, Parmar C, Quackenbush J, Schwartz LH, Aerts HJWL. 2018.** Artificial
384 intelligence in radiology. Nature Reviews Cancer 18:500–510.

385 **Jalalifar SA, Soliman H, Sahgal A, Sadeghi-Naini A. 2022.** Predicting the outcome of
386 radiotherapy in brain metastasis by integrating the clinical and MRI-based deep learning features.
387 Medical Physics 49:7167–7178.

388 **Liu F, Kijowski R, Feng L, El Fakhri G. 2020.** High-performance rapid MR parameter
389 mapping using model-based deep adversarial learning. Magnetic Resonance Imaging 74:152–
390 160.

391 **Ren J, Yuan Y, Tao X. 2022.** Histogram analysis of diffusion-weighted imaging and dynamic
392 contrast-enhanced MRI for predicting occult lymph node metastasis in early-stage oral tongue
393 squamous cell carcinoma. European Radiology 32:2739–2747.

394 **Tang Y, Che X, Wang W, Su S, Nie Y, Yang C. 2022.** Radiomics model based on features of
395 axillary lymphatic nodes to predict axillary lymphatic node metastasis in breast cancer. Medical
396 Physics 49:7555–7566.

397 **Yuan Y, Ren J, Tao X. 2021.** Machine learning–based MRI texture analysis to predict occult
398 lymph node metastasis in early-stage oral tongue squamous cell carcinoma. European Radiology
399 31:6429–6437.

400   **Zhang X, Lu B, Yang X, Lan D, Lin S, Zhou Z, Li K, Deng D, Peng P, Zeng Z, Long L.**
401   **2022.** Prognostic analysis and risk stratification of lung adenocarcinoma undergoing EGFR-TKI
402   therapy with time-serial CT-based radiomics signature. European Radiology 33:825–835.

403   **Zhao R, Wan L, Chen S, Peng W, Liu X, Wang S, Li L, Zhang H. 2023.** MRI-based
404   Multiregional Radiomics for Pretreatment Prediction of Distant Metastasis After Neoadjuvant
405   Chemoradiotherapy in Patients with Locally Advanced Rectal Cancer. Academic
406   Radiology:S1076633223004749.

407   **Zhou M, Leung A, Echegaray S, Gentles A, Shrager JB, Jensen KC, Berry GJ, Plevritis SK,**
408   **Rubin DL, Napel S, Gevaert O. 2018.** Non–Small Cell Lung Cancer Radiogenomics Map
409   Identifies Relationships between Molecular and Imaging Phenotypes with Prognostic
410   Implications. Radiology 286:307–315.

411   **Ansarin M, Bruschini R, Navach V, Giugliano G, Calabrese L, Chiesa F, Medina JE,**
412   **Kowalski LP & Shah JP. 2019.** Classification of GLOSSECTOMIES: Proposal for tongue
413   cancer resections. Head Neck 41(3): 821-827.

414   **Baba A, Hashimoto K, Kayama R, Yamauchi H, Ikeda K & Ojiri H. 2020.** Radiological
415   approach for the newly incorporated T staging factor, depth of invasion (DOI), of the oral tongue
416   cancer in the 8th edition of American Joint Committee on Cancer (AJCC) staging manual:
417   assessment of the necessity for elective neck dissection. *Jpn J Radiol* **38**(9): 821-832.

418   **Baba A, Okuyama Y, Ikeda K, Kozakai A, Suzuki T, Saito H, Ogane S, Yamazoe S,**
419   **Yamauchi H, Ogino N, Seto Y, Kobashi Y, Mogami T & Ojiri H. 2019.** Undetectability of
420   oral tongue cancer on magnetic resonance imaging; clinical significance as a predictor to avoid
421   unnecessary elective neck dissection in node negative patients. *Dentomaxillofac Radiol* 48(3):
422   20180272.

423   **Calabrese E, Rudie JD, Rauschecker AM, Villanueva-Meyer JE, Clarke JL, Solomon DA**
424   **& Cha S. 2022.** Combining radiomics and deep convolutional neural network features from
425   preoperative MRI for predicting clinically relevant genetic biomarkers in glioblastoma.
426   *Neurooncol Adv* **4**(1): vdac060.

427   **Charles RQ, Su H, Kaichun M & Guibas LJ 2017.** PointNet: Deep Learning on Point Sets for
428   3D Classification and Segmentation. *2017 IEEE Conference on Computer Vision and Pattern*
429   *Recognition (CVPR).* IEEE Computer Society.

430   **Chen T & Guestrin C 2016.** XGBoost: A Scalable Tree Boosting System. *Proceedings of the*
431   *22nd ACM SIGKDD International Conference on Knowledge Discovery and Data Mining.* San
432   Francisco, California, USA: Association for Computing Machinery.

433 **Cho HH, Kim H, Nam SY, Lee JE, Han BK, Ko EY, Choi JS, Park H & Ko ES. 2022.**
434 Measurement of Perfusion Heterogeneity within Tumor Habitats on Magnetic Resonance
435 Imaging and Its Association with Prognosis in Breast Cancer Patients. *Cancers (Basel)* **14**(8).

436 **Coletta RD, Yeudall WA & Salo T. 2020.** Grand Challenges in Oral Cancers. *Front Oral*
437 *Health* **1**(3.

438 **Cortes C & Vapnik V. 1995.** Support-vector networks. *Machine Learning* **20**(3): 273-297.

439 **Džeroski S & Ženko B. 2004.** Is Combining Classifiers with Stacking Better than Selecting the
440 Best One? *Machine Learning* **54**(3): 255-273.

441 **Geurts P, Ernst D & Wehenkel L. 2006.** Extremely randomized trees. *Machine Learning* **63**(1):
442 3-42.

443 **Gillies RJ, Kinahan PE & Hricak H. 2016.** Radiomics: Images Are More than Pictures, They
444 Are Data. *Radiology* **278**(2): 563-77.

445 **Harmon SA, Sanford TH, Xu S, Turkbey EB, Roth H, Xu Z, Yang D, Myronenko A,**
446 **Anderson V, Amalou A, Blain M, Kassin M, Long D, Varble N, Walker SM, Bagci U,**
447 **Ierardi AM, Stellato E, Plensich GG, Franceschelli G, Girlando C, Irmici G, Labella D,**
448 **Hammoud D, Malayeri A, Jones E, Summers RM, Choyke PL, Xu D, Flores M, Tamura K,**
449 **Obinata H, Mori H, Patella F, Cariati M, Carrafiello G, An P, Wood BJ & Turkbey B.**
450 **2020.** Artificial intelligence for the detection of COVID-19 pneumonia on chest CT using
451 multinational datasets. *Nat Commun* 11(1): 4080.

452 **He K, Zhang X, Ren S & Sun J.** Deep Residual Learning for Image Recognition.   2016 IEEE
453 Conference on Computer Vision and Pattern Recognition (CVPR), 27-30 June 2016 2016. 770-
454 778.

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

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

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

464 **Kubo K, Kawahara D, Murakami Y, Takeuchi Y, Katsuta T, Imano N, Nishibuchi I, Saito**
465 **A, Konishi M, Kakimoto N, Yoshioka Y, Toratani S, Ono S, Ueda T, Takeno S & Nagata Y.**
466 **2022.** Development of a radiomics and machine learning model for predicting occult cervical

467 lymph node metastasis in patients with tongue cancer. *Oral Surg Oral Med Oral Pathol Oral*
468 *Radiol* 134(1): 93-101.

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

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

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

478 Mostafaei S, Abdollahi H, Kazempour Dehkordi S, Shiri I, Razzaghdoust A, Zoljalali
479 **Moghaddam SH, Saadipoor A, Koosha F, Cheraghi S & Mahdavi SR. 2020.** CT imaging
480 markers to improve radiation toxicity prediction in prostate cancer radiotherapy by stacking
481 regression algorithm. *Radiol Med* 125(1): 87-97.

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

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

492 **Tian Z, Jia Y, Men X & Sun Z 2020.** 3DCNN for Pulmonary Nodule Segmentation and
493 Classification.

494 **Wang F, Tan R, Feng K, Hu J, Zhuang Z, Wang C, Hou J & Liu X. 2022.** Magnetic
495 Resonance Imaging-Based Radiomics Features Associated with Depth of Invasion Predicted
496 Lymph Node Metastasis and Prognosis in Tongue Cancer. *J Magn Reson Imaging* **56**(1): 196-
497 209.

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

501

Figure legends

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

504   **Figure 2.** Feature selection process. (A). Number and ratio of handcrafted features; Nonzero
505   coefficients were selected to establish the Rad-score with a LASSO logistic regression model.
506   (B). Coefficients and (D). MSE (mean standard error) of 10-fold validation; (C). The histogram
507   of the Rad-score based on the selected features. (E). Deep learning datasets based on different
508   methods (M1, M2, M3)

509   **Figure 3.** Development of radiomics model. (A). The ROC curves of the radiomics model; (B).
510   The confusion matrix of predicting lymph node metastasis

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

512   **Figure 5.** The ROC curves of the clinical model, radiomics model, deep learning model, and
513   integrated model in the training cohort (A) and validation cohort (B). The integrated model
514   demonstrated significantly higher AUCs in the training and validation cohorts (0.989 and 0.949)
515   than the clinical model (0.815 and 0.728), the radiomics model (0.896 and 0.893), and the deep
516   learning model (0.980 and 0.798)

517   **Figure 6.** The DCA of the clinical model, radiomics model, deep learning model, and integrated
518   model.

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

520   **Figure 8.** The ROC curves of the integrated model and diagnostic results of each radiologist in
521   the test cohort. The integrated model reached a high AUC of 0.838, clinician D1 acquired 0.654
522   of AUC, the AUC of clinician D2 was 0.768, and clinician D3 demonstrated the highest with
523   0.893.

524

Table legends

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

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

528

Figure 1

The workflow of this study.
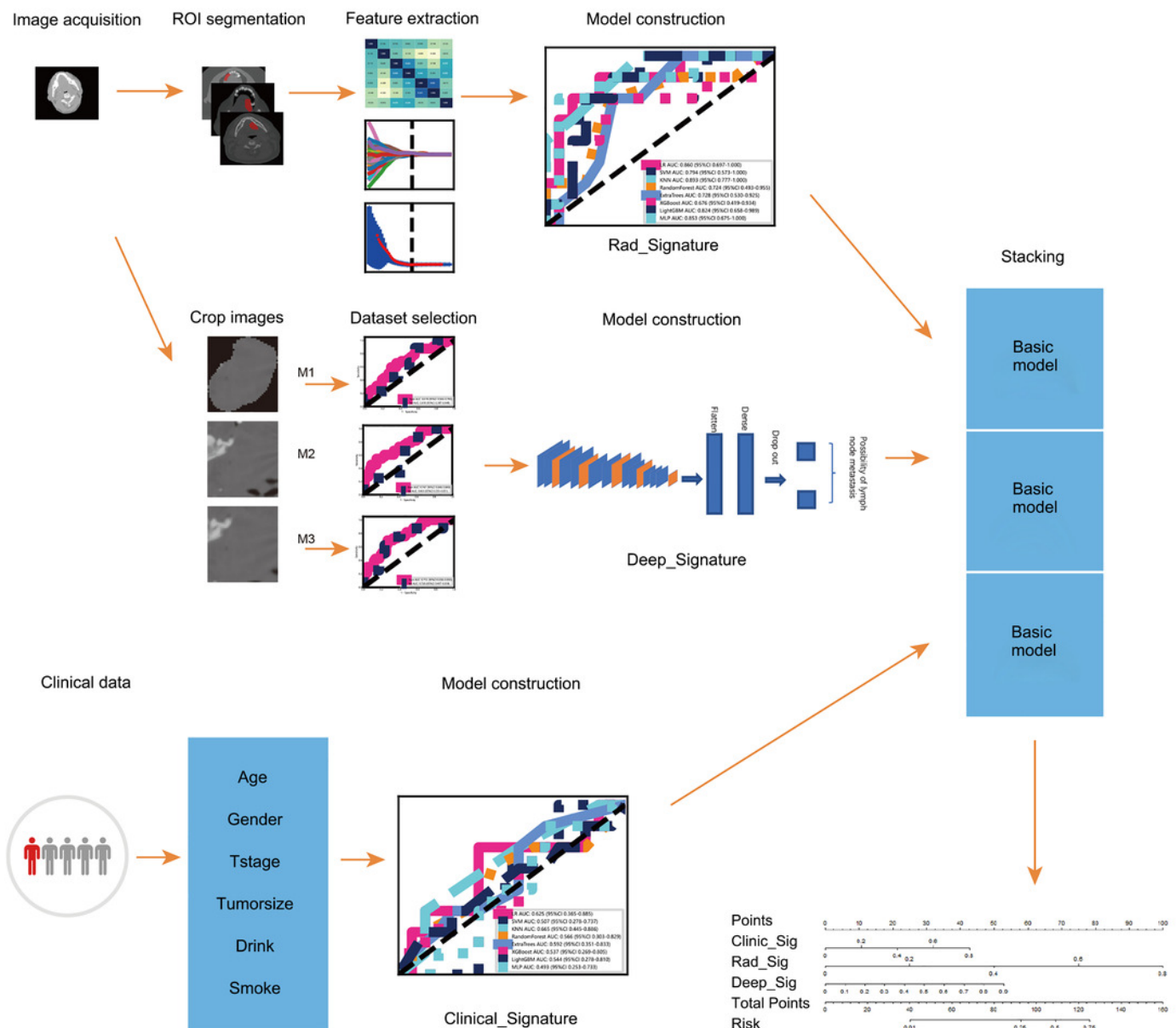
Figure 2

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

glrlm 17.4%
glszm 17.4%
ngtdm 5.5%
shape 0.8%
firstorder 19.6%
glcm 24.0%
gldm 15.3%

B

C

log_sigma_1_0_mm_3D_glcm_ClusterShade
lbp_3D_m1_glszm_SmallAreaEmphases
wavelet_LLL_ngtdm_Strength
square_firstorder_RobustMeanAbsoluteDeviation
wavelet_HHH_firstorder_Mean
original_shape_Maximum2DDiameterRow
squareroot_glcm_Correlation
wavelet_HLH_glszm_ZoneVariance
wavelet_HLH_gldm_LargeDependenceEmphasis
wavelet_LLL_firstorder_Skewness
lbp_3D_m1_glszm_GrayLevelVariance
lbp_3D_m2_glszm_GrayLevelVariance
original_ngtdm_Complexity
exponential_ngtdm_Coarseness
wavelet_HLH_glszm_ZonePercentage
square_glszm_GrayLevelNonUniformityNormalized

Coefficients

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
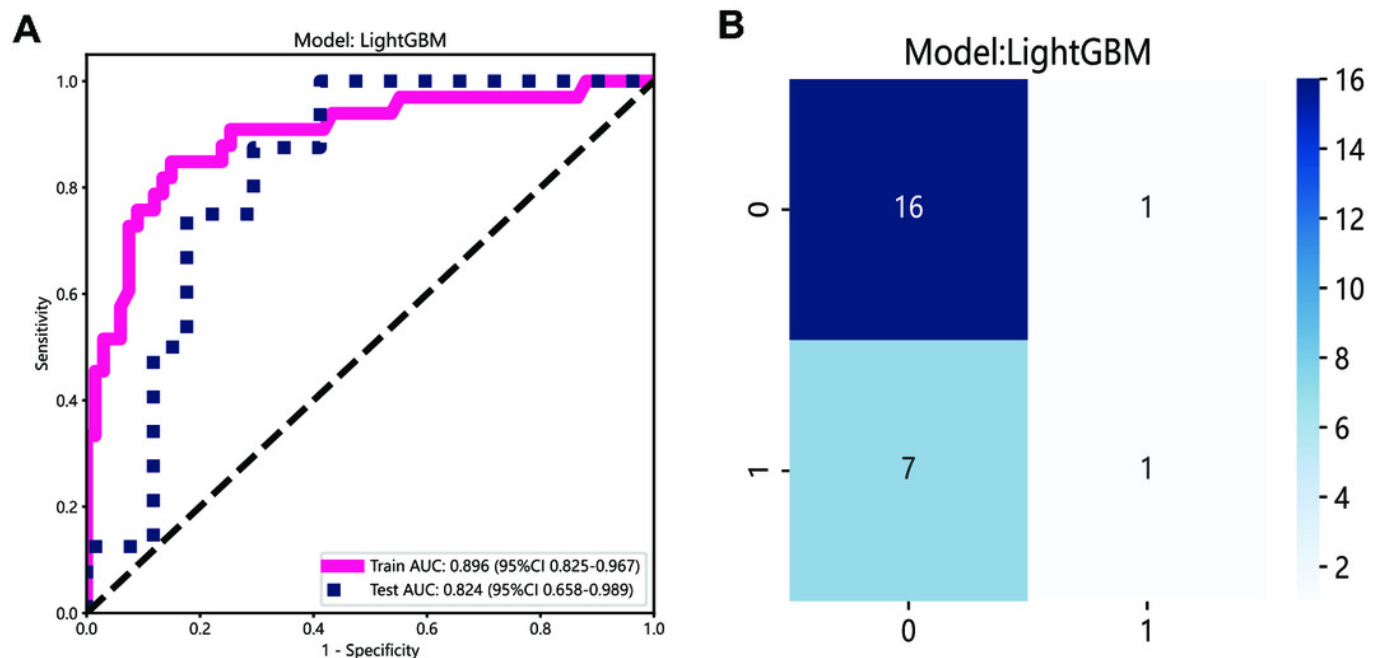
Figure 4

The ROC curves of M3.
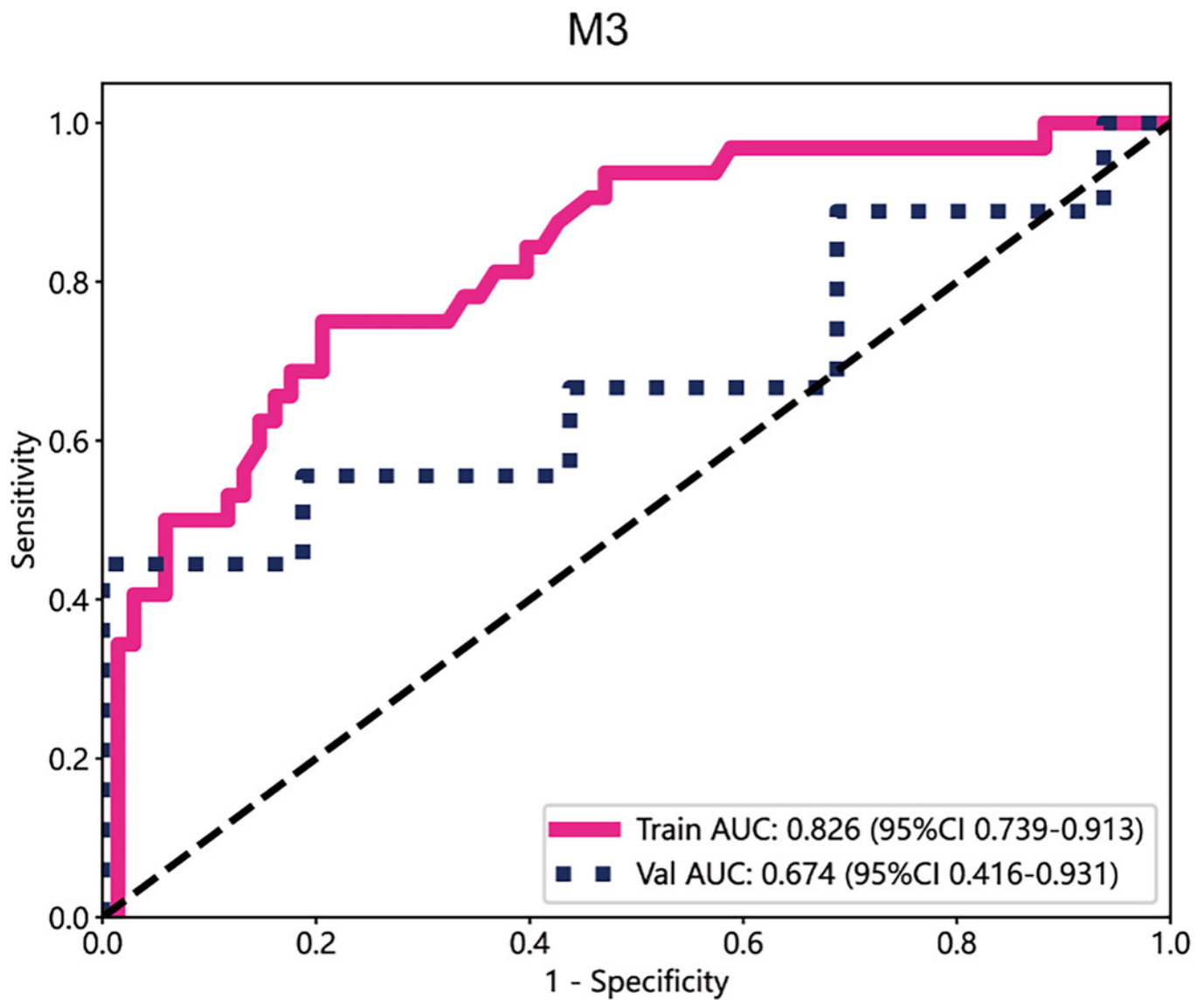
Figure 5

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)
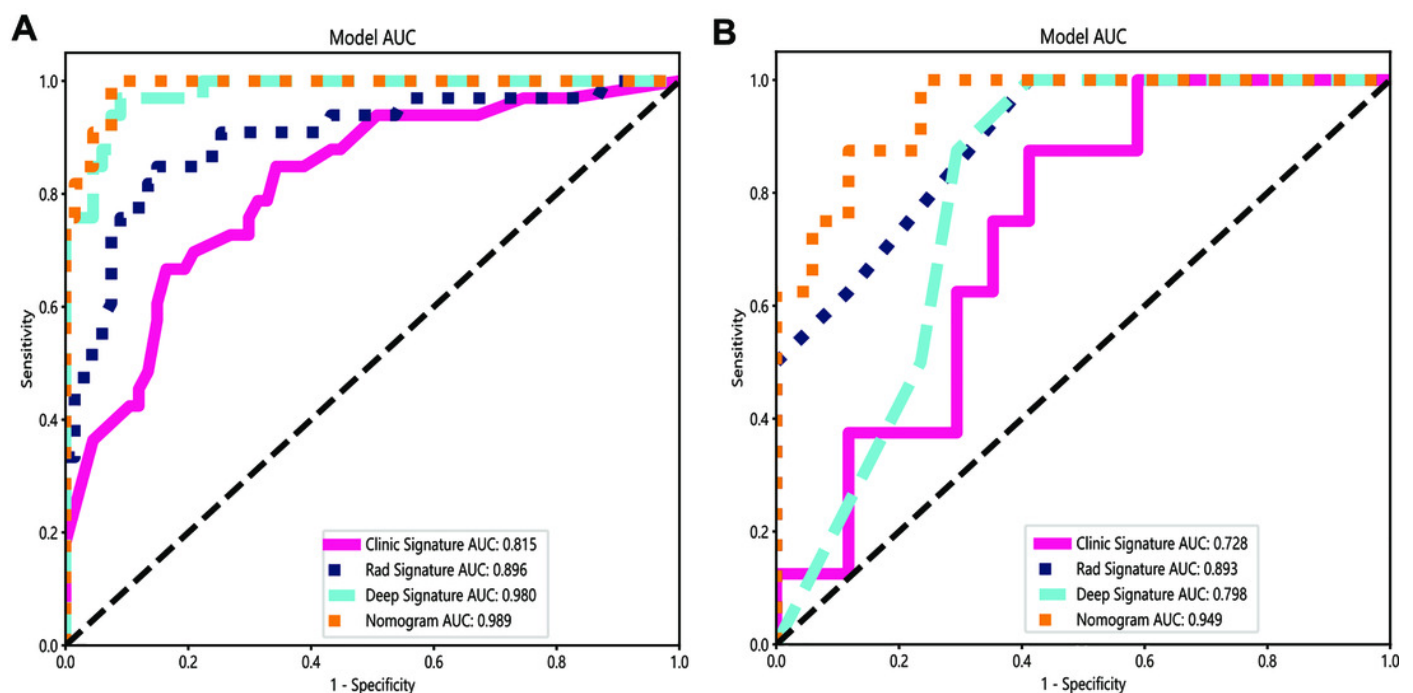
Figure 6

The DCA of the clinical model, radiomics model, Deep learning model and combined model.
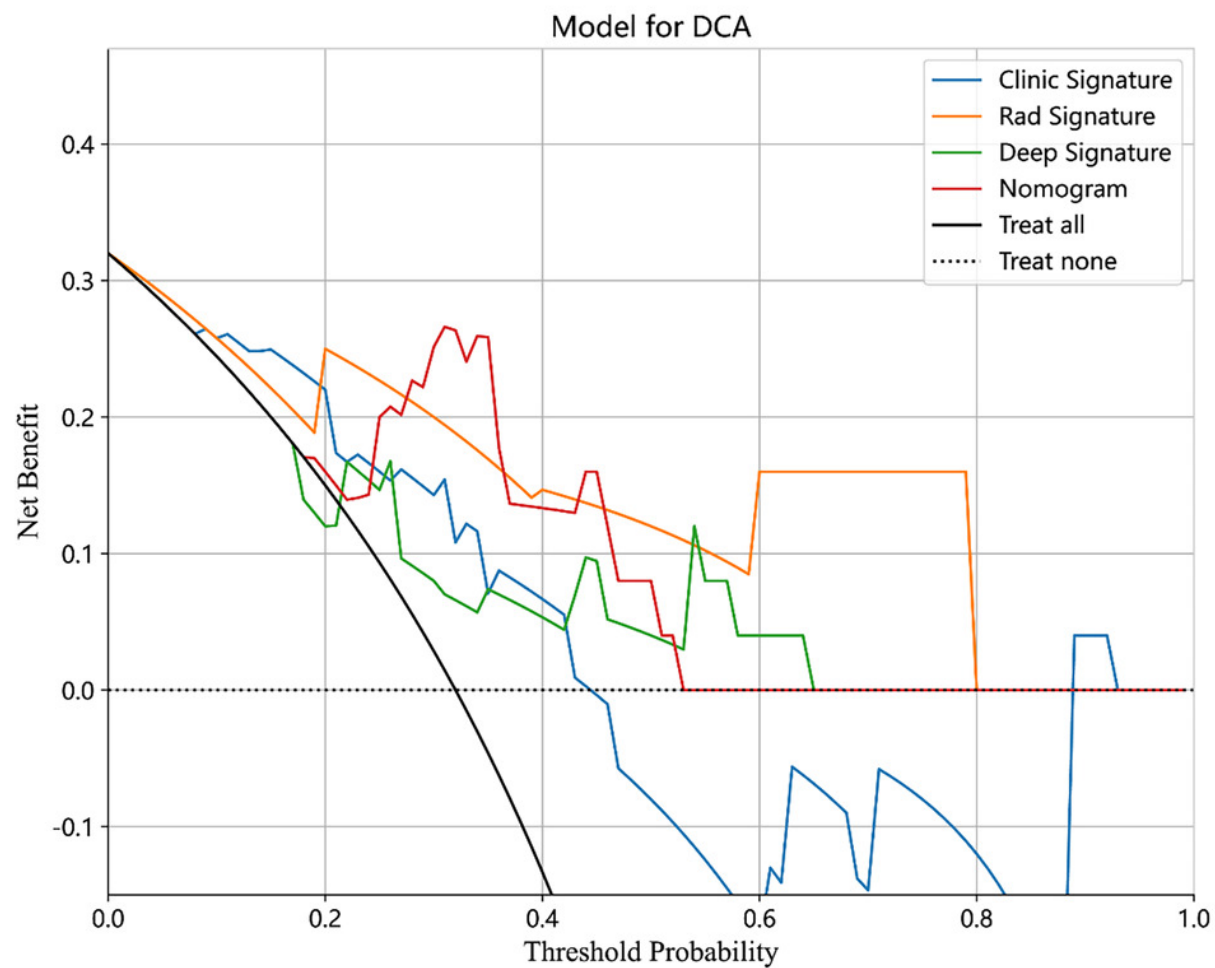
Figure 7

The development of the nomogram for clinical use.
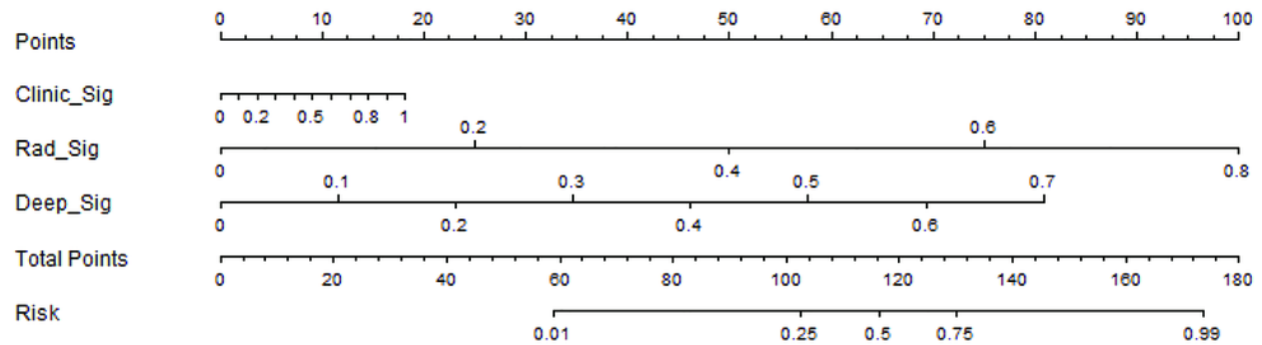
Figure 8

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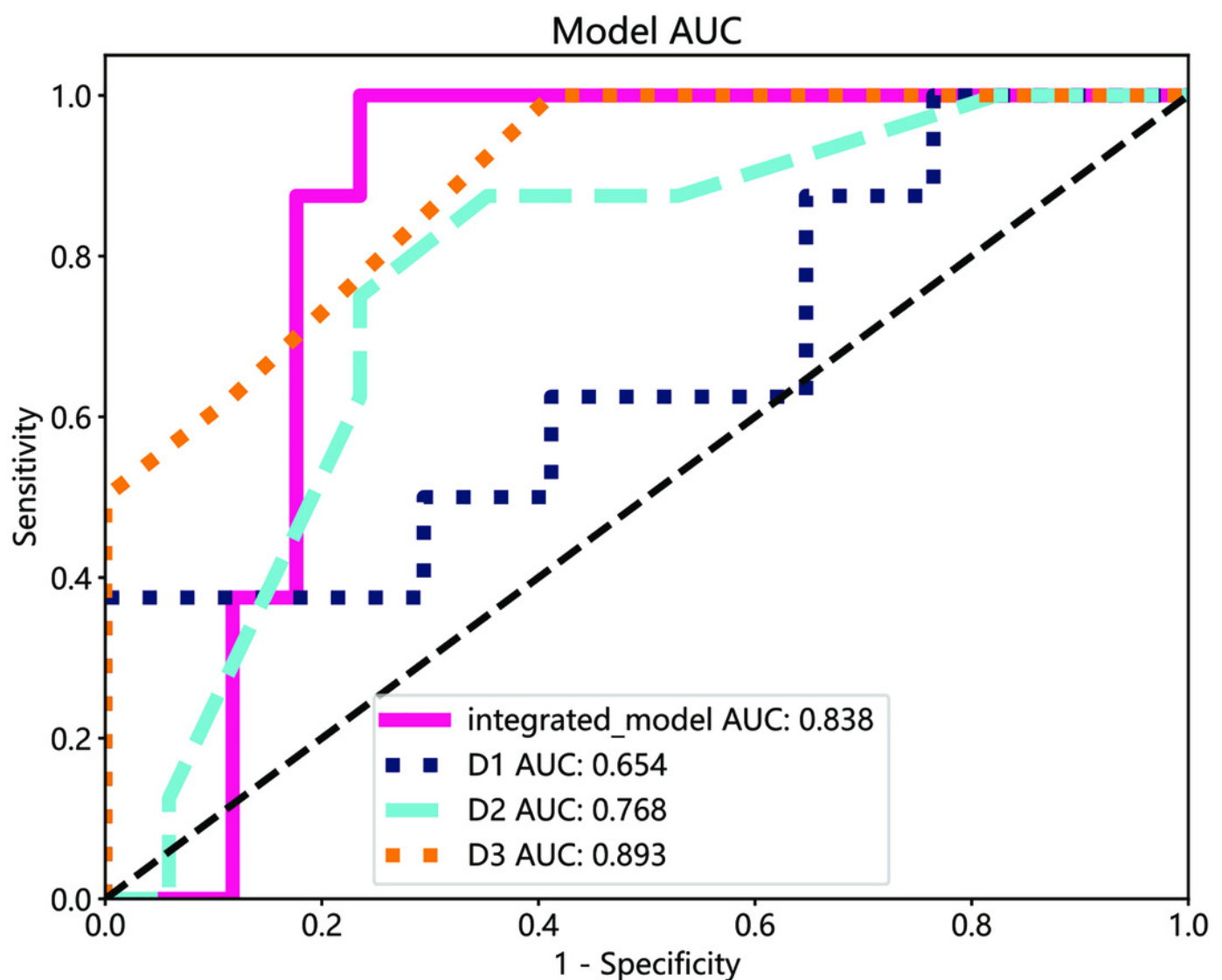
Table 1. The detailed clinical characteristics of the patients.

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
Tumor size	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
gender				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)	
T stage				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)	

Table 2. 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