



Deep learning assisted contrast-enhanced CT-based diagnosis of cervical lymph node metastasis of oral cancer: a retrospective study of 1466 cases

Xiaoshuai Xu¹ · Linlin Xi² · Lili Wei³ · Luping Wu¹ · Yuming Xu¹ · Bailve Liu² · Bo Li³ · Ke Liu^{1,4} · Gaigai Hou¹ · Hao Lin¹ · Zhe Shao^{1,4} · Kehua Su² · Zhengjun Shang^{1,4} 

Received: 21 June 2022 / Revised: 23 September 2022 / Accepted: 29 November 2022 / Published online: 28 December 2022
© The Author(s) 2022

Abstract

Objectives Lymph node (LN) metastasis is a common cause of recurrence in oral cancer; however, the accuracy of distinguishing positive and negative LNs is not ideal. Here, we aimed to develop a deep learning model that can identify, locate, and distinguish LNs in contrast-enhanced CT (CECT) images with a higher accuracy.

Methods The preoperative CECT images and corresponding postoperative pathological diagnoses of 1466 patients with oral cancer from our hospital were retrospectively collected. In stage I, full-layer images (five common anatomical structures) were labeled; in stage II, negative and positive LNs were separately labeled. The stage I model was innovatively employed for stage II training to improve accuracy with the idea of transfer learning (TL). The Mask R-CNN instance segmentation framework was selected for model construction and training. The accuracy of the model was compared with that of human observers.

Results A total of 5412 images and 5601 images were labeled in stage I and II, respectively. The stage I model achieved an excellent segmentation effect in the test set (AP_{50} -0.7249). The positive LN accuracy of the stage II TL model was similar to that of the radiologist and much higher than that of the surgeons and students (0.7042 vs. 0.7647 ($p = 0.243$), 0.4216 ($p < 0.001$), and 0.3629 ($p < 0.001$)). The clinical accuracy of the model was highest (0.8509 vs. 0.8000, 0.5500, 0.4500, and 0.6658 of the Radiology Department).

Conclusions The model was constructed using a deep neural network and had high accuracy in LN localization and metastasis discrimination, which could contribute to accurate diagnosis and customized treatment planning.

Key Points

- Lymph node metastasis is not well recognized with modern medical imaging tools.
- Transfer learning can improve the accuracy of deep learning model prediction.
- Deep learning can aid the accurate identification of lymph node metastasis.

Keywords Mouth neoplasms · Lymphatic metastasis · Tomography, X-ray computed · Deep learning · Diagnosis, computer-assisted

Xiaoshuai Xu, Linlin Xi and Lili Wei contributed equally to this work.

✉ Zhe Shao
shaozhe@whu.edu.cn

✉ Kehua Su
skh@whu.edu.cn

✉ Zhengjun Shang
shangzhengjun@whu.edu.cn

² School of Computer Science, Wuhan University, 299 Bayi Road, Wuhan 430072, Hubei, China

³ Department of Radiology, School & Hospital of Stomatology, Wuhan University, Wuhan, China

⁴ Department of Oral and Maxillofacial Head Neck Surgery, School & Hospital of Stomatology, Wuhan University, 237 Luoyu Road, Wuhan 430079, Hubei, China

¹ The State Key Laboratory Breeding Base of Basic Science of Stomatology (Hubei-MOST) & Key Laboratory of Oral Biomedicine Ministry of Education, School & Hospital of Stomatology, Wuhan University, Wuhan, China

Abbreviations

AI	Artificial intelligence
AP	Average precision
CECT	Contrast-enhanced computed tomography
DL	Deep learning
Faster R-CNN	Faster region-based convolutional neural network
FPN	Feature pyramid network
IoU	Intersection over union
IRB	Institutional review board
LN	Lymph node
LN–	Negative LN
LN+	Positive LN
Mask R-CNN	Mask region-based convolutional neural network
ML	Machine learning
PR	Precision/recall
ResNet101	Residual Network101
RoI	Region of interest
RPN	Region proposal network
SGD	Stochastic gradient descent
TL	Transfer learning

Introduction

Oral cancer is a prevalent malignancy worldwide with a high recurrence rate [1, 2]. Lymph node (LN) metastasis is a common cause of recurrence for oral cancer patients [3]. Poor identification of LN metastasis often causes undertreatment with occult metastasis (30–40%) and overtreatment at an early stage (60–70%), both of which could cause poor prognosis [4, 5]. Additionally, whether elective neck dissection should be performed and the extent of tissue removal essentially depend on the diagnosis of LN metastasis.

Currently, CT and MRI are frequently utilized to identify the size, internal heterogeneity, and contour of LNs [6, 7]. With the adjunction of [¹⁸F]FDG PET, these imaging approaches can reach higher sensitivity [8, 9]. Despite advancements in modern medical imaging technology, even experienced radiologists often miss diagnoses with LN metastasis due to limited working time and a heavy daily workload [10, 11]. Additional invasive procedures, such as fine-needle aspiration biopsy, could be performed to increase the diagnostic accuracy of suspicious LNs, which is harmful to patients [12]. Therefore, more accurate and noninvasive LN metastasis diagnosis methods are needed.

As a subdomain of machine learning (ML), deep learning (DL) is a method that uses more complex network model structures, is much better at discovering deeper features in input data, and performs well in many practical application scenes of ML [13, 14]. Many studies have shown that DL achieves excellent performance in image processing problems

[15–17]. DL-based image classification and object detection have been widely employed in the medical field to provide supporting advice for diagnoses [13, 18–24], such as skin cancer [25], breast cancer (LNs) [26], and COVID-19 [27]. Furthermore, DL can achieve or exceed the performance of human experts in several tasks of medical image analysis [28, 29]. In the field of medical image processing based on DL, the lack of annotated datasets is a major problem [18]. The introduction of transfer learning (TL) to DL can alleviate the shortage of annotated datasets. TL can improve the performance of learning on target domains by transferring knowledge from different but related source domains [30].

To help decrease the high misdiagnosis rate of LNs, it is worthwhile to introduce artificial intelligence (AI) to the field of imaging identification of LNs in oral cancer. Consequently, this research intends to develop a DL model that can identify, locate, and distinguish LNs in CECT images with a higher accuracy to replace existing inefficient manual identification methods.

Materials and methods

Patient cohort

This study was reported according to the recommendations of the STROBE guidelines and was approved by the Institutional Review Board (IRB) of the Ethics Committee of the Hospital of Stomatology, Wuhan University (IRB No. 2020-B63). As it was a retrospective study, all patients could not be identified, only imaging data and pathological reports were collected, and the extracted data did not contain patient names. Therefore, consent for participation was not obtained.

The CECT images of 2773 patients with oral cancer who were admitted to the Hospital of Stomatology, Wuhan University, between September 1, 2012, and September 22, 2020, were retrospectively collected. After screening according to our inclusion and exclusion criteria, 1307 invalid samples were removed (Table S1). The baseline information of the included patients is shown in Table 1 and Table S2. For the 1466 included samples, 5412 images were selected for full-layer data labeling (stage I), and 5601 images were selected for LN metastasis discrimination data labeling (stage II) (Fig. 1).

Data processing

Three researchers were specially trained by a radiologist with 20 years of working experience in our hospital to be responsible for this study. All data were labeled based on pathological information for consensus regarding segmentation accuracy. If a consensus was not reached, researchers consulted the radiologist to reduce personnel errors. After the labeling was

Table 1 Baseline information of 1466 oral cancer patients in this study

		Total (N = 1466)	Male (N = 1059)	Female (N = 407)
Age (median [IQR])		55.41 [48,64]	54.80 [48,63]	57.01 [49,67]
T (tumor stage)	T1	34 (2.32%)	20 (1.36%)	14 (0.96%)
	T2	434 (29.60%)	287 (19.58%)	147 (10.02%)
	T3	693 (47.27%)	532 (36.29%)	161 (10.98%)
	T4a	305 (20.81%)	220 (15.01%)	85 (5.80%)
	T4b	0	0	0
N (node stage)	N0	778 (53.07%)	559 (38.13%)	219 (14.94%)
	N1	268 (18.28%)	189 (12.89%)	79 (5.39%)
	N2a	15 (1.02%)	10 (0.68%)	5 (0.34%)
	N2b	304 (20.74%)	225 (15.35%)	79 (5.39%)
	N2c	23 (1.57%)	17 (1.16%)	6 (0.41%)
	N3a	0	0	0
M (metastasis stage)	N3b	78 (5.32%)	59 (4.02%)	19 (1.30%)
	M0	1466 (100%)	1059 (72.24%)	407 (27.76%)
	M1	0	0	0
Clinical stage	I	29 (1.98%)	18 (1.23%)	11 (0.75%)
	II	310 (21.15%)	205 (13.99%)	105 (7.16%)
	III	511 (34.85%)	385 (26.26%)	126 (8.59%)
	IVA	537 (36.63%)	392 (26.74%)	145 (9.89%)
	IVB	79 (5.39%)	59 (4.03%)	20 (1.36%)
	IVC	0	0	0

We used the Joint Commission on Cancer eighth edition TNM classification (TNM8) criteria
IQR interquartile range

completed, it was reviewed and approved (or modified as needed) by the radiologist.

This study was carried out in two stages. In stage I, CECT images were screened according to the traditional CT level guidelines for cervical LNs (seven categories) [31]. Most of the possible locations of LNs were included in the images, especially the most common levels, I to V. The target areas were outlined by forming polygons based on Label Studio (a privately deployed platform, which is based on the open-source data-labeling platform). For example, LNs were outlined by polygons consisting of green dots and lines. The remaining structures that appeared in the image were synchronously outlined, such as teeth (light blue), bone (dark blue), blood vessels (red), and other soft tissue (yellow). In stage II, positive LN (LN+) and negative LN (LN−) were outlined by polygons consisting of red (LN+) and green (LN−) dots and lines, respectively, based on Label Studio (Fig. 2). Detailed information on data processing (including CT parameters) is provided in the Supplement.

Model introduction

Mask R-CNN, a simple, flexible, and general DL framework in the field of instance segmentation, was introduced in this study. Mask R-CNN adds a branch of mask generation based

on Faster R-CNN, which can realize object detection and generate high-quality masks for different instances of an object [32].

The relevant code based on the Mask R-CNN framework using TensorFlow was implemented. The network structure consisted of three parts: the backbone network, region proposal network (RPN), and head network (Fig. S1 and S2).

Due to the similar nature of the training tasks in these two stages, both of which were multiclass instance segmentation (five classes in stage I and two classes in stage II), the overall network structure was similar, but the training process parameter settings were different.

ResNet101, which is composed of 101 layers, was selected as the backbone feature extraction network, and simultaneously constructed feature pyramids based on a feature pyramid network (FPN). In the following part, C1 to C5 were employed to define the feature map output by different ResNet convolutional layers [16]. The construction of the FPN was to realize the fusion of features at multiple scales [33]. The extracted P2 to P6 could be utilized as the effective feature layer of the RPN (Fig. S2).

The RPN was used to generate a region of interest (RoI) for the head network [34]. Considering the size of the input image, five anchors of different scales— 16×16 , 32×32 , 64×64 , 128×128 , and 256×256 —were defined. The gridlines

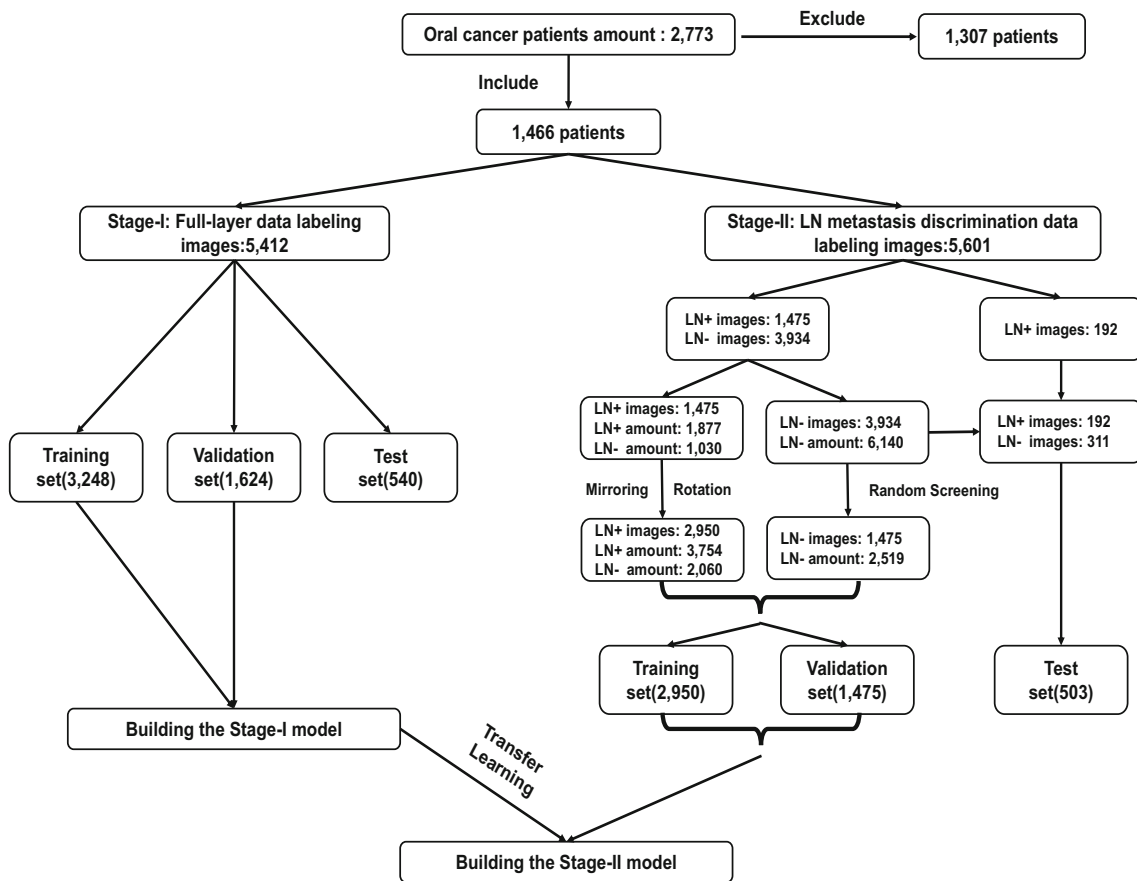
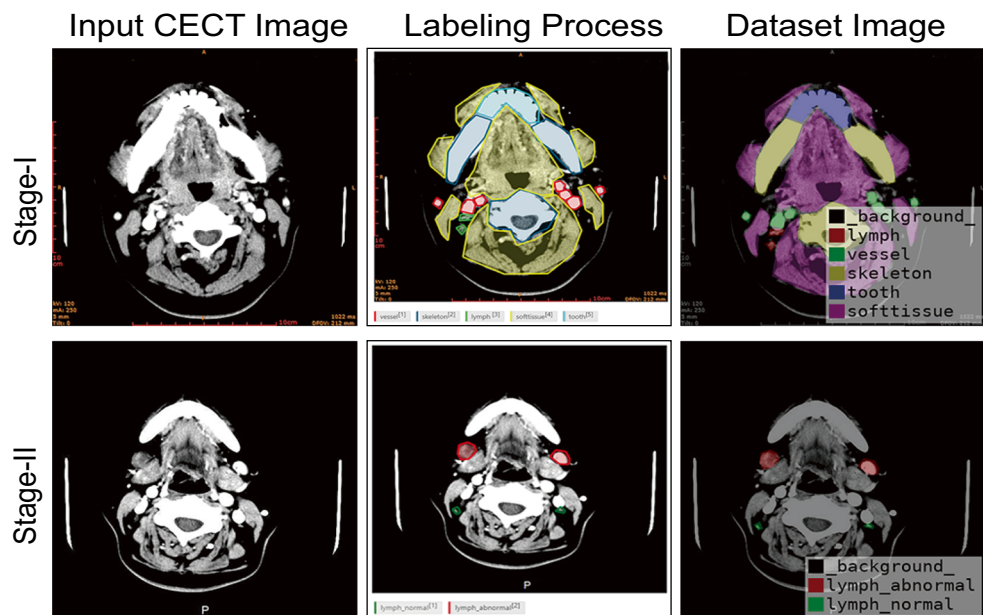


Fig. 1 Process of patient data collection, screening, classification, and data labeling. Flowchart of the study population and assignments for data labeling and machine learning for construction of prediction models. LN, lymph node; LN+, positive lymph node; LN-, negative lymph node

were generated through the RPN area, and the preselection bounding boxes that may contain objects were generated and referred to as the proposal area. A set of RoIs was

generated according to the evaluation of each proposal area, which greatly reduced the task size of subsequent classification detection.

Fig. 2 The process of data labeling. The process of data labeling for the input CECT images of stage I and stage II using our data-labeling tools. The dataset image used for training was generated from labeled data



The head network, which included a classifier model and mask generation model, was responsible for classification and mask generation. Both branches were simultaneously employed. The head network determined the object category of the input RoI through a classifier model and a generated mask for each instance of the object by classifying the pixel level through a mask generation model [32].

To minimize the influence of overfitting, the L2 regularization method was utilized for the model.

Model training

In the two stages, the loss function (L) applied was multitask loss, which was the sum of classification loss (L_{cls}), bounding-box loss (L_{box}), and mask branch loss (L_{mask}).

$$L = L_{cls} + L_{box} + L_{mask}$$

In our study, the datasets were shuffled and split into training (60%), validation (30%), and test sets (10%). For model training, Mask R-CNN COCO model weights were used as the initial weights, which were optimized by the stochastic gradient descent (SGD) strategy.

In stage I, the model was trained for 50 epochs in total on the training set and validated after each epoch on the validation set. The trends of loss during training and validation are shown in Fig. S3a, b.

In stage II, the LN metastasis discrimination labeled data (enlarged by means of rotation and mirroring) were served as the dataset. The stage II model (without TL) was trained for 100 epochs in total. The trends of loss are shown in Fig. S3c, d. Subsequently, the stage I model of the training results was used as the initial model weight for TL to improve the effect. However, the stage II-TL model (with TL) was trained for 10 epochs in total. The trends of loss are shown in Fig. S3e, f.

During training, validation loss was calculated after each epoch, and model weights were saved following each epoch that showed improvement in validation loss.

Assessment by clinicians

After model construction and optimization, another radiologist with 11 years of working experience, two surgeons with 12 and 11 years of experience, and two graduate students with 3 years of clinical experience in the Department of Oral and Maxillofacial Head Neck Surgery were invited to identify and distinguish LN_s+. The accuracy of the model was compared with clinicians' results to assess the clinical usability of the model.

Statistical analysis

The chi-square test was performed to examine the differences in the categorical variables. A two-sided $p < 0.05$ means that the corresponding estimate reaches a significant difference. The performance of the stage I model was evaluated on the independent test set using the average precision (AP) of the model to measure the segmentation effect. However, in stage II, the AP could not directly reflect the overall effect of the model. Hence, three new customized but much stricter model evaluation criteria were introduced: LN accuracy, LN+ accuracy, and clinical accuracy. Detailed information about the calculation of the AP and PR curves and these three new evaluation criteria are provided in the Supplement.

Results

Identify different anatomical structures (stage I)

Stage I aimed to automatically recognize different soft and hard tissues in the cervical CECT images, especially to distinguish and identify LN_s, hoping to achieve more accurate recognition and positioning of LN_s in the next stage of training. All anatomical structures, including teeth, bones, LN_s, blood vessels, and other soft tissues were labeled (Fig. 2). A total of 5412 CECT images were screened. The prediction result image and precision/recall (PR) curve for different intersection-over-union (IoU) thresholds of the stage I model are shown in Fig. 3a, b.

AP₅₀, AP₆₀, AP₇₀, and AP reached 72.49%, 58.72%, 35.88%, and 29.63%, respectively, for the discriminant segmentation of different anatomical structures in the test set (Table 2). The AP rapidly decreased with an increase in the IoU threshold. Considering the small size of LN_s and blood vessels in images, the number of negative samples increased with an increase in the IoU threshold. In summary, in stage I, the different anatomical structures could be effectively recognized in the cervical CECT images.

Accuracy of cervical LN metastasis discrimination was improved by TL (stage II)

Stage II was designed to accurately distinguish LN metastasis. LN₋ and LN₊ were specifically labeled (Fig. 2). A total of 5601 CECT images were screened, in which 2126 were LN_s+, with 1667 images involved, and 6140 were LN_s-, with 3934 images involved (Fig. 1).

Since AP₅₀ represented the best recognition effect in the stage I model, only AP₅₀ was shown in different stage II models (Table 2). The prediction result images and the PR curve at the default IoU (0.50) threshold of the stage II model are shown in Fig. 3c, d.

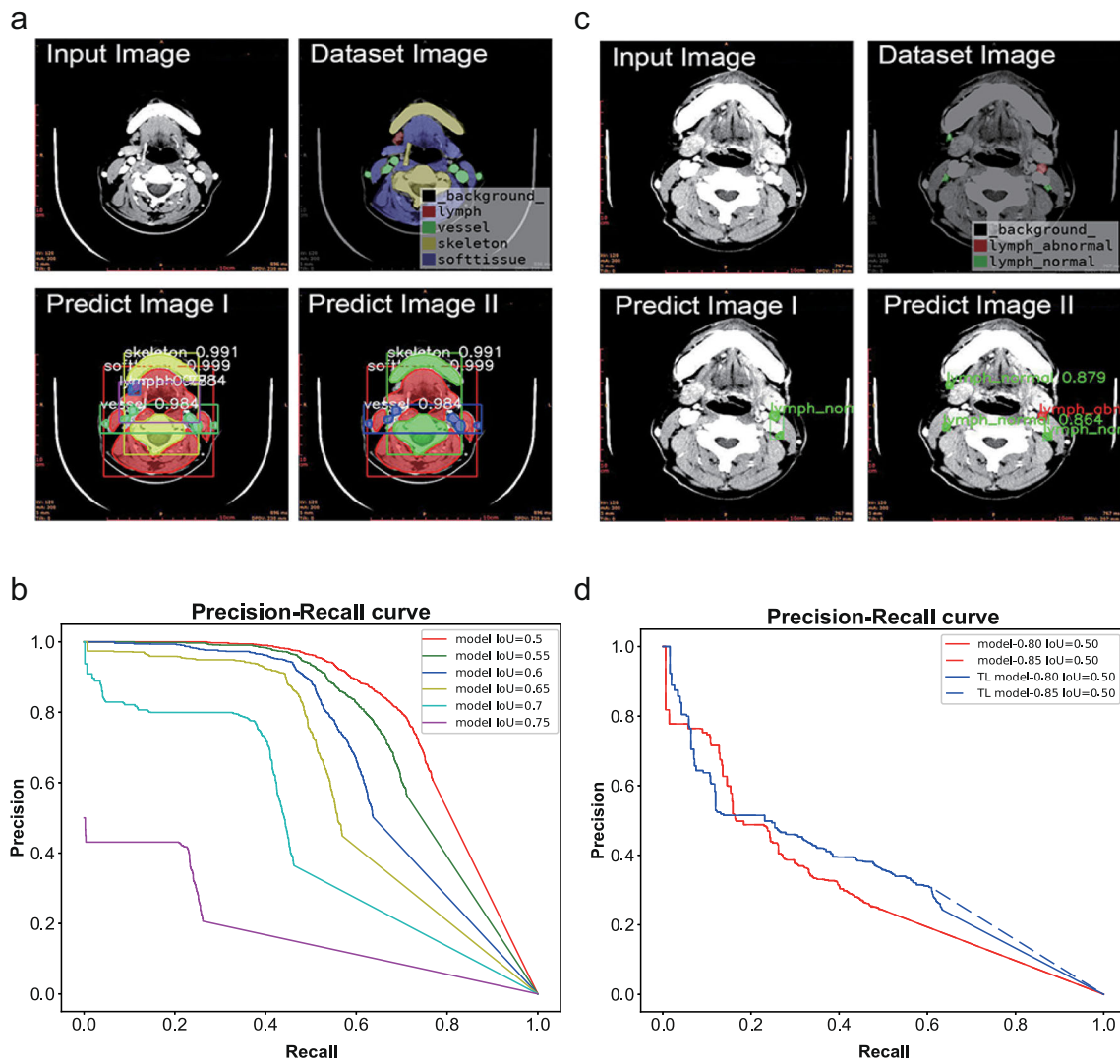


Fig. 3 Prediction process and PR curves. **a** Prediction process using the stage I model. Predicted images I and II were generated from the input image with confidence levels of 0.75 and 0.85, respectively. The dataset image was labeled for comparison. **b** PR curves at different IoU threshold values in stage I. **c** Prediction process using the stage II model. Predicted

image I was generated from the input image in the stage II model. Predicted image II was generated from the input image in the stage II-TL model. The dataset image was labeled for comparison. **d** PR curves of the model before and after the stage I model was utilized for training of transfer learning at the default IoU threshold value (0.50)

However, only the partial neck levels that underwent elective neck dissection had LNs’ pathological information for the labeling process, while the LNs’ pathological information in

the neck levels without surgery was ambiguous, and the LN status in these levels could not be labeled. But the prediction included levels without pathological information, and when these predicted instance objects appeared in the model evaluation, the corresponding instance could not be identified in the labeled samples. As a result, many negative samples were generated during the matching calculation, which affected the overall evaluation of the model, and its AP could not directly reflect the effect of the model. Hence, three new customized but much stricter model evaluation criteria were introduced: LN accuracy, LN+ accuracy, and clinical accuracy.

Table 2 AP of each model at different IoU thresholds

	AP	AP ₅₀	AP ₆₀	AP ₇₀
Stage I model	29.63	72.49	58.72	35.88
Stage II model-0.80	\	24.35	\	\
Stage II-TL model-0.80	\	30.63	\	\
Stage II model-0.85	\	24.35	\	\
Stage II-TL model-0.85	\	29.97	\	\

AP average precision, TL transfer learning, “\” means there is no value

Notably, when the stage I model was employed in stage II for TL, the stage II-TL model significantly improved the discrimination of cervical LN metastasis (Tables 3, 4 and 5) ($p < 0.001$). Each indicator reached the highest value at the 0.85

Table 3 Chi-square test between TL model-0.80 and model-0.80

	Right	Wrong	Accuracy (%)	<i>p</i> values
TL model-0.80	209	135	60.76	
Model-0.80	179	272	39.69	<i>p</i> = 3.92274E-09

confidence level (Table 5). In conclusion, the stage II-TL model achieved the highest LN accuracy (71.90%), LN+ accuracy (70.42%), and clinical accuracy (85.09%) at the 0.85 confidence level.

Comparison of predicting accuracy in different neck levels

The neck is divided into a total of seven levels based on the topographical subdivision [30, 31]. In the stage II test set, the data at levels I, II, and III were dominant, with 195, 191, and 198 images, respectively. These three levels were peculiarly prone to LN metastasis in oral cancer. Therefore, the prediction results of LN+ accuracy at these levels were compared. The results indicated that for different models and different confidence levels, the highest LN+ accuracy was obtained at level II (Fig. 4a and Table 5).

Prediction of cervical LN metastasis: model versus clinicians

To verify the accuracy of the model, the stage II-TL model-0.85 predictions were compared with those of the radiologist, surgeons, surgical students, and diagnostic reports from the Radiology Department in our hospital. Considering that the accurate identification of LN+ is the most important indicator, LN+ accuracy and clinical accuracy were selected for the comparison of the prediction results. Unfortunately, medical imaging diagnostic reports did not describe every LN, so the

Table 4 Chi-square test between TL model-0.85 and other models or clinicians

	Right	Wrong	Accuracy (%)	<i>p</i> values
TL model-0.85	200	84	70.42	
Model-0.85	175	253	40.89	<i>p</i> = 1.08402E-14
TL model-0.80	209	135	60.76	<i>p</i> = 0.011407545
TL model-0.84	202	94	68.24	<i>p</i> = 0.569445805
TL model-0.86	198	84	70.21	<i>p</i> = 0.956442618
TL model-0.90	158	108	59.40	<i>p</i> = 0.006720646
Radiologist	78	24	76.47	<i>p</i> = 0.243141608
Surgeons	85	119	42.16	<i>p</i> = 2.05967E-10
Students	90	159	36.29	<i>p</i> = 2.23635E-15

LN+ accuracy and LN accuracy of the Radiology Department could not be calculated.

Collectively, our model achieved LN+ accuracy, which was comparable to that of the radiologist (70.42% vs. 76.47%) (*p* = 0.243), and its LN+ accuracy was much higher than that of the surgeons (42.16%) (*p* < 0.001) and students (36.29%) (*p* < 0.001) (Table 4). Our model achieved the highest clinical accuracy (85.09% vs. 80%, 55%, 45%, and 66.58% of the Radiology Department) (Fig. 4b and Table 5).

Discussion

In this study, DL using a neural network combined with CECT images could identify, locate, and distinguish cervical metastatic LNs of oral cancer patients with a very high accuracy and was demonstrated to exceed that of an experienced radiologist and surgeons. Even in terms of clinical accuracy, our model exceeded the average of the Radiology Department in our hospital by nearly 20%. To the best of our knowledge, this is the largest study on the prediction of cervical LN metastasis by CECT images based on DL in oral cancer. Our model was very effective not only in assisting radiologists in accurately diagnosing LN metastasis before surgery but also in helping surgeons in designing customized and accurate surgical plans to improve prognosis.

For many years, neck dissection has been a major concern for oral cancer patients worldwide. With all fields being facilitated by AI, especially in the medical imaging field, it is regrettable that few studies have carried out AI analysis on cervical LNs [25, 27, 35]. Reza et al and Yuan et al performed texture analysis on cervical LN metastasis by dual-energy CT and MRI, respectively, based on ML with high accuracy [36, 37]. The content of our study was CECT, consistent with China's national conditions because of their strong universality and applicability. [¹⁸F]FDG PET/CT can better detect occult neck metastasis than CT/MRI imaging [9]. However, this method is expensive, and only a few large hospitals are equipped with such expensive machines. Additionally, the sample sizes adopted by previous researchers were measured in units of 10 and 100, while the present study directly collected data from 2773 patients and included 1466 samples after screening. Most significantly, the present study first divided the work into two stages and innovatively integrated the idea of TL to apply the results of the first stage of the model to the training of the second stage of the model, which significantly improved the prediction accuracy. In previous studies, the prediction required an experienced operator to outline the LNs before the model recognized them, which was laborious and time-consuming. In this study, we need only to input a complete CECT image set of the patient into the model, after which the prediction results can be obtained in seconds, which

Table 5 Stage II predictions of each model and predictions of radiologist, surgeons, students, and the Radiology Department

	Clinical accuracy	LN accuracy	Positive LN accuracy			
			Total	Level I	Level II	Level III
Model-0.80	58.45 (294/503)	45.62 (583/(1007 + 271))	39.69 (179/(249 + 202))	33.63	46.63	33.68
TL model-0.80	76.34 (384/503)	71.76 (846/(1007 + 172))	60.76 (209/(249 + 95))	56.60	64.38	53.62
Model-0.85	60.44 (304/503)	42.33 (527/(1007 + 238))	40.89 (175/(249 + 179))	34.93	46.24	36.36
TL model-0.85	85.09 (428/503)	71.90 (801/(1007 + 102))	70.42 (200/(249 + 35))	66.67	72.22	65.45
TL model-0.84	83.10 (418/503)	71.83 (811/(1007 + 122))	68.24 (202/(249 + 47))	64.18	70.00	62.07
TL model-0.86	85.09 (428/503)	68.38 (757/(1007 + 100))	70.21 (198/(249 + 33))	65.60	72.00	66.06
TL model-0.90	82.31 (414/503)	53.01 (563/(1007 + 55))	59.40 (158/(249 + 17))	51.69	60.17	55.88
Radiologist	80.00 (80/100)	\	76.47 (78/(91 + 11))	\	\	\
Surgeons	55.00 (55/100)	\	42.16 (43/(91 + 11))	\	\	\
Students	45.00 (45/100)	\	36.29 (45/(91 + 33))	\	\	\
Radiology Department	66.58 (976/1466)	\	\	\	\	\

TL transfer learning, LN lymph node, "\" means there is no value

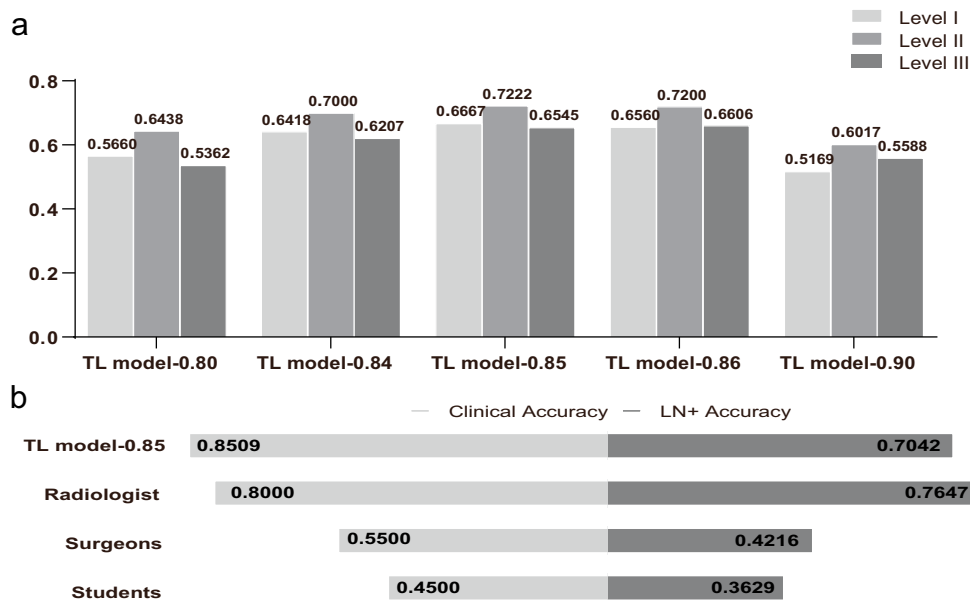
is convenient and time-efficient with strong clinical translation ability.

As the model built by combining DL and radiology was noninvasive, we could use only CECT image to perform operational analysis before surgery and to minimize harm to the patients. Although the results of our study were impressive and provide a basis for future research, more challenging tasks, such as the identification of LNs at different levels, need to be completed [31]. Nearly all patients included in the study had neck LN dissection at levels I, II, and III (images dominated the dataset), while only some patients had surgery at levels IV and V. Notably, levels I to III were peculiarly prone to LN metastasis in oral cancer. Incorrect discrimination of

LNs at levels I to III is prone to increase as LNs at these levels are often accompanied by blood vessels. In the statistical results of our study, we presented the first analysis of discrepancies in the accuracy of the prediction of LNs at different levels and discovered that level II (upper group of the internal jugular chain) had a much higher accuracy at any confidence level in any model.

Although the AI utilized in the present study could greatly facilitate the diagnosis of diseases by clinicians, the current study has limitations. First, this was a single-center retrospective experiment; in particular, the LNs in pathological reports and CECT images were not completely accurate in one-to-one correspondence. Second, persuading patients to embrace this

Fig. 4 Model evaluation and prediction results. **a** LN+ accuracy of the stage II-TL model at different levels at different confidence intervals. **b** Comparison results of LN+ accuracy and clinical accuracy among the stage II-TL model-0.85, radiologist, surgeons, and students. TL, transfer learning; LN+, positive lymph node



new diagnostic system remains challenging and requires large-scale clinical application and verification. Third, because the current ability of medical imaging (e.g., CT and MRI) to detect micrometastases in LNs with normal morphology and dimension is rather limited [6], our model, which is based on CECT imaging, still cannot reach satisfactory prediction within this aspect. Last, multicenter prospective studies and a larger sample size are needed to validate our results to popularize our AI results and to overcome the subtle discrepancies of different CECT machines. Given these issues, AI has great potential in cervical LN metastasis diagnosis; however, whether it can be successfully promoted remains to be determined.

In conclusion, despite the fact that more framework optimization and large-scale verification are still required before official clinical application, AI-DL-Mask R-CNN-assisted CECT can accurately predict the metastasis of cervical LNs, providing strong support for the efficient diagnosis of radiologists to a certain extent. Importantly, convenient manipulation, instant diagnosis, and excellent effect endow this technology with a powerful clinical translation ability.

Supplementary Information The online version contains supplementary material available at <https://doi.org/10.1007/s00330-022-09355-5>.

Acknowledgements We thank Zhiyuan Xiong from The Network Center of Hospital of Stomatology, Wuhan University, for providing data download assistance in this study.

Code availability Our networks were constructed based on the Mask R-CNN framework, and the code for the Mask R-CNN framework we referenced can be downloaded from the GitHub repository <https://github.com/facebookresearch/Detectron> and https://github.com/matterport/Mask_RCNN. Our data-labeling platform was deployed based on the privatization of the open-source data-labeling platform Label Studio at <https://github.com/heartexlabs/label-studio>. We trained on a computer with an NVIDIA V100 GPU, and the codes we used for training are available at the GitHub repository <https://github.com/whucsss/LNdiagnosis-stage1> and <https://github.com/whucsss/LNdiagnosis-stage2>.

Funding This study has received funding by the National Natural Science Foundation of China 81772897, 81672666, 61772379 and the National Key R&D Program of China 2018YFB2100500.

Declarations

Guarantor The scientific guarantor of this publication is Zhengjun Shang.

Conflict of interest The authors of this manuscript declare no relationships with any companies whose products or services may be related to the subject matter of the article.

Statistics and biometry No complex statistical methods were necessary for this paper.

Informed consent Written informed consent was waived by the Institutional Review Board.

Ethical approval Institutional Review Board approval was obtained.

Methodology

- retrospective
- diagnostic or prognostic study
- performed at one institution

Open Access This article is licensed under a Creative Commons Attribution 4.0 International License, which permits use, sharing, adaptation, distribution and reproduction in any medium or format, as long as you give appropriate credit to the original author(s) and the source, provide a link to the Creative Commons licence, and indicate if changes were made. The images or other third party material in this article are included in the article's Creative Commons licence, unless indicated otherwise in a credit line to the material. If material is not included in the article's Creative Commons licence and your intended use is not permitted by statutory regulation or exceeds the permitted use, you will need to obtain permission directly from the copyright holder. To view a copy of this licence, visit <http://creativecommons.org/licenses/by/4.0/>.

References

1. Pare A, Joly A (2017) Oral cancer: risk factors and management. *Presse Med* 46:320–330
2. Bray F, Ferlay J, Soerjomataram I, Siegel RL, Torre LA, Jemal A (2018) Global cancer statistics 2018: GLOBOCAN estimates of incidence and mortality worldwide for 36 cancers in 185 countries. *CA Cancer J Clin* 68:394–424
3. Walk EL, Weed SA (2011) Recently identified biomarkers that promote lymph node metastasis in head and neck squamous cell carcinoma. *Cancers (Basel)* 3:747–772
4. Oh LJ, Phan K, Kim SW, Low TH, Gupta R, Clark JR (2020) Elective neck dissection versus observation for early-stage oral squamous cell carcinoma: systematic review and meta-analysis. *Oral Oncol* 105:104661
5. Liao LJ, Hsu WL, Wang CT, Lo WC, Lai MS (2016) Analysis of sentinel node biopsy combined with other diagnostic tools in staging cN0 head and neck cancer: a diagnostic meta-analysis. *Head Neck* 38:628–634
6. Forghani R, Yu E, Levental M, Som PM, Curtin HD (2015) Imaging evaluation of lymphadenopathy and patterns of lymph node spread in head and neck cancer. *Expert Rev Anticancer Ther* 15:207–224
7. Jin TP, Roh JL, Kim JS et al (2016) 18 F FDG PET/CT versus CT/MR imaging and the prognostic value of contralateral neck metastases in patients with head and neck squamous cell carcinoma. *Radiology* 279:481–491
8. Heusch P, Sproll C, Buchbender C et al (2014) Diagnostic accuracy of ultrasound, 18F-FDG-PET/CT, and fused 18F-FDG-PET-MR images with DWI for the detection of cervical lymph node metastases of HNSCC. *Clin Oral Investig* 18:969–978
9. Bae MR, Roh JL, Kim JS et al (2020) (18)F-FDG PET/CT versus CT/MR imaging for detection of neck lymph node metastasis in palpably node-negative oral cavity cancer. *J Cancer Res Clin Oncol* 146:237–244
10. Kann BH, Aneja S, Loganadane GV, Kelly JR, Husain ZAJSR (2018) Pretreatment identification of head and neck cancer nodal

- metastasis and extranodal extension using deep learning neural networks. *Sci Rep* 8:14036
11. McDonald RJ, Schwartz KM, Eckel LJ et al (2015) The effects of changes in utilization and technological advancements of cross-sectional imaging on radiologist workload. *Acad Radiol* 22:1191–1198
 12. van der Linden N, Flach GB, de Bree R, Uyl-de Groot CA (2016) Cost-utility of sentinel lymph node biopsy in cT1-T2N0 oral cancer. *Oral Oncol* 53:20–26
 13. Esteva A, Robicquet A, Ramsundar B et al (2019) A guide to deep learning in healthcare. *Nat Med* 25:24–29
 14. LeCun Y, Bengio Y, Hinton G (2015) Deep learning. *Nature* 521:436–444
 15. Szegedy C, Liu W, Jia Y et al (2015) Going deeper with convolutions. 2015 IEEE Conference on Computer Vision and Pattern Recognition (CVPR), pp 1–9
 16. He K, Zhang X, Ren S, Sun J (2016) Deep residual learning for image recognition. 2016 IEEE Conference on Computer Vision and Pattern Recognition (CVPR), pp 770–778
 17. Chan TH, Jia K, Gao S, Lu J, Zeng Z, Ma Y (2015) PCANet: a simple deep learning baseline for image classification? *IEEE Trans Image Process* 24:5017–5032
 18. Litjens G, Kooi T, Bejnordi BE et al (2017) A survey on deep learning in medical image analysis. *Med Image Anal* 42:60–88
 19. Coudray N, Ocampo PS, Sakellaropoulos T et al (2018) Classification and mutation prediction from non-small cell lung cancer histopathology images using deep learning. *Nat Med* 24:1559–1567
 20. Hirasawa T, Aoyama K, Tanimoto T et al (2018) Application of artificial intelligence using a convolutional neural network for detecting gastric cancer in endoscopic images. *Gastric Cancer* 21:653–660
 21. Horie Y, Yoshio T, Aoyama K et al (2019) Diagnostic outcomes of esophageal cancer by artificial intelligence using convolutional neural networks. *Gastrointest Endosc* 89:25–32
 22. Xu Y, Hosny A, Zeleznik R et al (2019) Deep learning predicts lung cancer treatment response from serial medical imaging. *Clin Cancer Res* 25:3266–3275
 23. Bulten W, Pinckaers H, van Boven H et al (2020) Automated deep-learning system for Gleason grading of prostate cancer using biopsies: a diagnostic study. *Lancet Oncol* 21:233–241
 24. Zhou D, Tian F, Tian X, Sun L, Li XJNC (2020) Diagnostic evaluation of a deep learning model for optical diagnosis of colorectal cancer. *Nat Commun* 11:2961
 25. Esteva A, Kuprel B, Novoa RA et al (2017) Dermatologist-level classification of skin cancer with deep neural networks. *Nature* 542:115–118
 26. Golden JAJJ (2017) Deep learning algorithms for detection of lymph node metastases from breast cancer: helping artificial intelligence be seen. *JAMA* 318:2184–2186
 27. Ezzat D, Hassanien A, Ella HJAsc (2021) An optimized deep learning architecture for the diagnosis of COVID-19 disease based on gravitational search optimization. *Appl Soft Comput* 98:106742
 28. Ardila D, Kiraly AP, Bharadwaj S et al (2019) End-to-end lung cancer screening with three-dimensional deep learning on low-dose chest computed tomography. *Nat Med* 25:954–961
 29. Topol EJ (2019) High-performance medicine: the convergence of human and artificial intelligence. *Nat Med* 25:44–56
 30. Zhuang F, Qi Z, Duan K et al (2021) A comprehensive survey on transfer learning. *Proc IEEE* 109:43–76
 31. Robbins KT, Shaha AR, Medina JE et al (2008) Consensus statement on the classification and terminology of neck dissection. *Arch Otolaryngol Head Neck Surg* 134:536
 32. He K, Gkioxari G, Dollár P, Girshick R (2017) Mask R-CNN. 2017 IEEE International Conference on Computer Vision (ICCV), pp 2980–2988
 33. Lin TY, Dollár P, Girshick R, He K, Hariharan B, Belongie S (2017) Feature pyramid networks for object detection. 2017 IEEE Conference on Computer Vision and Pattern Recognition (CVPR), pp 936–944
 34. Ren S, He K, Girshick R, Sun J (2017) Faster R-CNN: towards real-time object detection with region proposal networks. *IEEE Trans Pattern Anal Mach Intell* 39:1137–1149
 35. Schwyzer M, Martini K, Benz DC et al (2020) Artificial intelligence for detecting small FDG-positive lung nodules in digital PET/CT: impact of image reconstructions on diagnostic performance. *Eur Radiol* 30:2031–2040
 36. Forghani R, Chatterjee A, Reinhold C et al (2019) Head and neck squamous cell carcinoma: prediction of cervical lymph node metastasis by dual-energy CT texture analysis with machine learning. *Eur Radiol* 29:6172–6181
 37. Yuan Y, Ren J, Tao X (2021) Machine learning-based MRI texture analysis to predict occult lymph node metastasis in early-stage oral tongue squamous cell carcinoma. *Eur Radiol* 31:6429–6437

Publisher's note Springer Nature remains neutral with regard to jurisdictional claims in published maps and institutional affiliations.