

DeepThy-Net: A Multimodal Deep Learning Method for Predicting Cervical Lymph Node Metastasis in Papillary Thyroid Cancer

Jincao Yao, Zhikai Lei, Wenwen Yue, Bojian Feng, Wei Li, Di Ou, Na Feng, Yidan Lu, Jing Xu, Wencong Chen, Chen Yang, Lijing Wang, Liping Wang, Junping Liu, Peiying Wei, Huixiong Xu,* and Dong Xu*

Papillary thyroid cancer (PTC) accounts for more than 80% of thyroid cancers, and ultrasound (US) imaging is the preferred method for the diagnosis of PTC. However, accurate prediction of different patterns of cervical lymph node metastasis (CLNM) in PTC continues to be a challenge. Herein, US images and clinical factors of PTC patients from three hospitals for more than 11 years are collected, and a multimodal deep learning model called DeepThy-Net is then developed to predict different CLNM patterns. The proposed model not only uses the convolutional features extracted by deep learning but also integrates traditional clinical factors that are highly related to lymph node metastasis. Finally, the model is tested in two independent test sets, and the experimental results show that the area under curve (AUC) is between 0.870 and 0.905, indicating clinical applicability. The proposed method provides an important reference for the treatment and management of PTC. Moreover, for PTC cases involving an active surveillance strategy, the proposed method can serve as an important CLNM early warning tool.

1. Introduction

Thyroid cancer is a common disease with an incidence rate of $\approx 6.6\%$, and its incidence is threefold higher among women compared with men.^[1,2] Papillary thyroid cancer (PTC) accounts for more than 80% of all thyroid cancers, and some of them develop cervical lymph node metastases (CLNM).^[3-6] Although ultrasound (US) imaging is the preferred method for the diagnosis of PTC,^[7-9] due to anatomical particularities, the lymph nodes in the posterior trachea, esophagus, and pharynx cannot be well detected.^[10,11] For lateral CLNM, the sensitivity of ultrasonography is higher than that in the central region. However, given the varying operating skills and experience of radiologists, the incidences of misdiagnoses and missed

J. Yao, W. Li, D. Ou, N. Feng, Y. Lu, C. Yang, L. Wang, L. Wang, J. Liu, D. Xu
 Medical Ultrasonics Department
 Cancer Hospital of the University of Chinese Academy of Sciences
 (Zhejiang Cancer Hospital)
 No.1 East Banshan Road, Hangzhou 310022, P. R. China
 E-mail: xudong@zjcc.org.cn

J. Yao, W. Li, D. Ou, N. Feng, Y. Lu, C. Yang, L. Wang, L. Wang, J. Liu, D. Xu
 Institute of Basic Medicine and Cancer (IBMC)
 Chinese Academy of Sciences
 No.150 Dongfangjie, Hangzhou 310000, P. R. China

J. Yao, D. Xu
 Key Laboratory of Head & Neck Cancer Translational Research of Zhejiang Province
 Zhejiang Provincial Research Center for Cancer Intelligent Diagnosis and Molecular Technology
 No.1 East Banshan Road, Hangzhou 310022, P. R. China

Z. Lei, P. Wei
 Zhejiang University School of Medicine
 Affiliated Hangzhou First People's Hospital
 No.261 Huansha Road, Hangzhou 310006, P. R. China

W. Yue, H. Xu
 Department of Medical Ultrasound
 Shanghai Tenth People's Hospital
 School of Medicine
 Shanghai Engineering Research Center of Ultrasound in Diagnosis and Treatment
 Tongji University
 No. 301 Middle Yanchang Road, Shanghai 200072, P. R. China
 E-mail: xuhuixiong2022@126.com

B. Feng, J. Xu
 School of Statistics and Mathematics
 Zhejiang Gongshang University
 No.18 Xuezhengjie, Hangzhou 310018, P. R. China

W. Chen
 Department of Biostatistics
 Vanderbilt University Medical Center
 2215 Garland Ave, Nashville, TN 440039, USA

H. Xu
 Department of Ultrasound
 Zhongshan Hospital
 Fudan University
 No.921 Tongxin road, Shanghai 200032, P. R. China

 The ORCID identification number(s) for the author(s) of this article can be found under <https://doi.org/10.1002/aisy.202200100>.

© 2022 The Authors. Advanced Intelligent Systems published by Wiley-VCH GmbH. This is an open access article under the terms of the Creative Commons Attribution License, which permits use, distribution and reproduction in any medium, provided the original work is properly cited.

DOI: [10.1002/aisy.202200100](https://doi.org/10.1002/aisy.202200100)

diagnoses are still high, especially in the early stages of metastasis.^[12–15] Although many studies have shown that the features of PTC lesion in US images were highly correlated with CLNM, there is still no accurate or efficient method that can predict the risk of different CLNM patterns prior to operation.^[16–19] As is known, CLNM is a fundamental indicator of the recurrence and prognosis of PTC patients and the important basis of lymph node dissection (LND).^[16,17] And exploring a noninvasive assessment method for different CLNM patterns is of great value for the treatment and disease management of PTC.

To address this problem, earlier studies attempted to use clinical factor-based statistical methods to predict the risk of different CLNM in PTC.^[20–24] For example, a univariable analysis by Wu et al. found that Hashimoto's thyroiditis was an independent factor for CLNM in PTC ($p = 0.016$).^[20] Roh et al.'s multivariate analyses showed that tumor size and extrathyroidal extension of PTC are two important factors for CLNM ($p < 0.05$).^[21] In addition, Nie et al. reported that invasion and tumor location are correlated with CLNM ($p < 0.05$).^[22] Generally, these traditional methods mainly relied on clinical experience or risk factors, and the prediction accuracy was not high. With the development of medical image processing and feature extraction technology, many studies have begun to use machine learning (ML) models and high-throughput radiomic features (HTRF) of the tumor US images to obtain better results.^[23,24] For example, Liu et al. adopted a classic support vector machine (SVM) algorithm together with 50 radiomic features from the tumor US images to predict central CLNM in PTC.^[23] Jin et al. used a logistic regression (LR) model to predict lateral CLNM in 106 PTC patients and obtained an AUC value of 0.74.^[24] And these methods improved the prediction accuracy to some extent.

Although these efforts have made progress, there are many limitations. First, the amount of data in these studies is small, and the accuracy of the model is not satisfactory. Second, when assessing the risk of CLNM in PTC, we need to face multiple patterns, such as central-only CLNM, both central and lateral CLNMs, and non-metastatic cases. However, a unified model that can accurately assess the risk of multiple CLNM patterns has not been reported yet. Moreover, earlier studies have shown that some clinical factors, such as extrathyroidal extension, capsular invasion, and Hashimoto's thyroiditis, are correlated with CLNM;^[20–22] unfortunately, a model capable of comprehensively considering both clinical factors and US image features has not been developed yet.

Recently, deep learning (DL) and cancer diagnosis methods have made important breakthroughs.^[25–31] Considering that the deep convolutional neural network (DCNN) has a strong feature extraction ability, in this study, we developed a multimodal DL model called DeepThy-Net to predict different CLNM patterns in PTC. And to effectively train the model, a total of 23 617 US images from 6,032 PTC cases were collected from three hospitals for more than 11 years.

To the best of our knowledge, this is the first study using multimodal DL method to assess the risks of multiple CLNM patterns in PTC. Meanwhile, it is also the largest dataset used to study this problem so far. The experimental results show that our method achieved encouraging accuracy. To better illustrate the design of the experiment, Figure 1 presents the overall

schematics. As is shown, PTC patients will first undergo an ultrasound examination, and their US images are collected. After a doctor's evaluation, clinical factors highly related to lymph node metastasis are recorded. Our model not only uses the convolutional image features but also integrates traditional clinical factors. Finally, all the features and factors are input into a fully connected network to infer the predictive results of CLNMs in PTC.

2. Results

2.1. Performance in Independent Test Set 1

The overall receiver operating characteristic (ROC) curves of the proposed model using both convolution features and clinical factors in the independent test set 1 are shown in Figure 2. The ROC curves of the ResNet50, DenseNet121, InceptionV4, and CSAC-Net are also shown as control methods.^[29–32] Table S1, Supporting Information, lists the aforementioned clinical factors. The training set contained 17 956 US images from 5,129 patients, and the independent test set 1 contained 3,271 US images from 487 patients. Based on the pathological results of the LND, the test samples were divided into three categories: central-only CLNM (170 patients), both central and lateral CLNM (126 patients), and non-metastatic cases (191 patients). We used the one-hot method to label the samples. Samples without CLNM, with central-only CLNM, and with both central and lateral CLNM were labeled as "001," "010," and "100," respectively.

As shown in Figure 2, the ResNet50, DenseNet121, InceptionV4, and CSAC-Net methods obtained similar experimental results, their ROC curves are concentrated in a close range, and the overall AUC is between 0.75 and 0.81. Our single-modal DeepThy-Net (the green lines) yielded slightly better results compared with the aforementioned methods, and the overall AUC was between 0.81 and 0.85. Finally, the multimodal DeepThy-Net (the red curve with CFs) achieved better performance, and the overall AUC was between 0.87 and 0.90. Figure 3a shows the US images of PTC with different types of CLNM and their convolutional feature maps. Columns 1 and 2 show the US images and corresponding convolutional feature maps of central-only CLNM cases; Columns 3 and 4 show the US images and feature maps of both central and lateral CLNM cases; Columns 5 and 6 show the US images and feature maps of non-metastatic cases. As seen in the figure, the feature energies of samples without CLNM are visibly lower than those of the other samples. Figure 3b,c shows the distribution of the identification results and the confusion matrix, respectively. The abscissa of Figure 3b represents the output value of the proposed model; the green histogram represents the cases that should be predicted as 0; the blue histogram represents the cases that should be predicted as 1. As can be seen, despite some overlap, the samples with different labels have begun clustering.

Based on the ROC curves we calculate the true positive rate (TPR), true negative rate (TNR) together with the KL-divergence (KLD), and the details of the comparison are shown in Table 1. In addition, the extra positive predictive value (PPV), negative

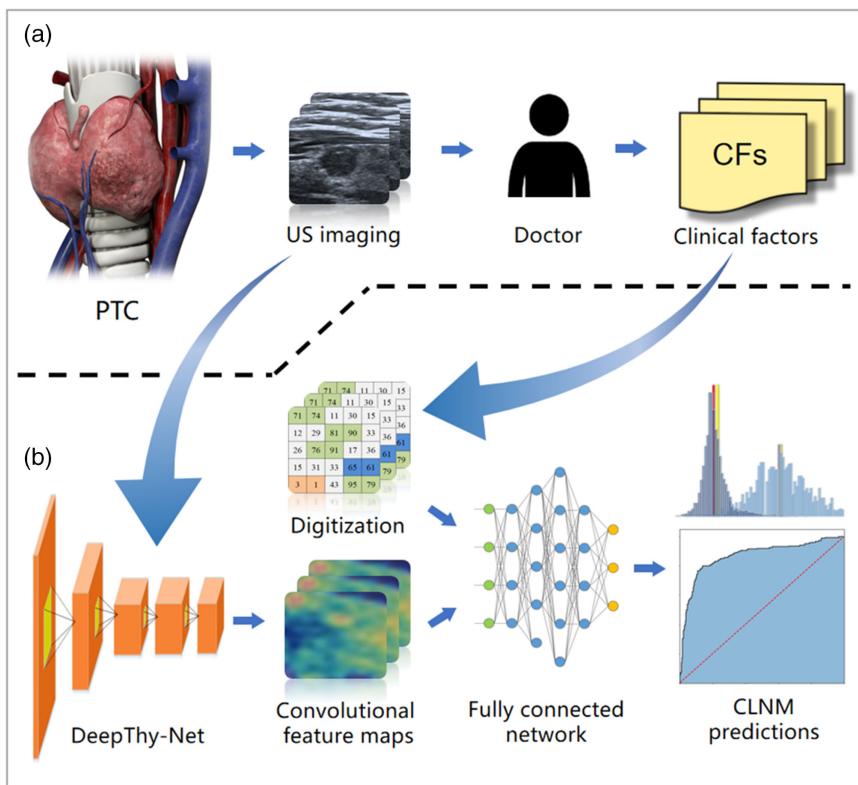


Figure 1. Overall schematics for the prediction of cervical lymph node metastasis (CLNM) in papillary thyroid cancer (PTC) using a deep learning (DL) model. a) The PTC patient underwent ultrasound examination, and after a doctor's evaluation, a series of high-risk clinical factors were recorded. b) The DeepThy-Net model was trained to extract the convolutional features of the ultrasound (US) images. The clinical factors recorded by doctors were digitized and input into a fully connected network with the convolutional features, and finally, the prediction results were obtained.

predictive value (NPV), accuracy (ACC), and F1 scores are shown in Table S2, Supporting Information. Specifically, for PTC cases with central-only CLNM, the sensitivity rate is 82.9% (141/170, 95% CI: 76.3–88.1%), the specificity is 83.0% (263/317, 95% CI: 78.3–86.8%), and the AUC value is 0.874 ($p < 0.01$). For cases involving both central and lateral CLNM, the overall sensitivity is 83.3% (105/126, 95% CI: 75.4–89.2%), the specificity is 88.6% (320/361, 95% CI: 84.8–91.6%), and the AUC value is 0.894 ($p < 0.01$). For non-metastatic cases, the sensitivity is 80.1% (153/191, 95% CI: 73.6–85.4%), the specificity is 83.1% (246/296, 95% CI: 78.2–87.1%), and the AUC value is 0.870 ($p < 0.01$). We also tried to introduce only one clinical factor at a time, and conducted experiments separately (train the model with only one clinical factor at a time) to test the contribution of each clinical factor to the AUC of our model. The experimental results show that the extrathyroidal extension, age, and tumor size are the top three clinical factors that contribute to the AUC, and their total contributions to the improvement of the AUC were 3.17, 3.00, and 2.96, respectively. Table S3, Supporting Information, shows the contributions of all clinical factors.

2.2. Performance in Independent Test Set 2

The independent test set 2 contained 2,390 ultrasound images from 416 patients. The test samples were divided into three

categories based on the pathological results: central-only CLNM (143 patients), both central and lateral CLNM (114 patients), and non-metastatic cases (159 patients). The ROC curves of the ResNet50, DenseNet121, InceptionV4, and CSAC-Net are shown in Figure 4, and the overall AUC was between 0.77 and 0.82. The single-modal DeepThy-Net (the green lines) obtained relatively better results, and the overall AUC was between 0.80 and 0.85. Our multimodal DeepThy-Net (the red curve with CFs) achieved the best performance, and the overall AUC was between 0.88 and 0.91.

Figure 5a shows US images of PTC with different types of CLNM and their corresponding convolutional feature maps. Columns 1 and 2 show the US images and convolutional feature maps of central-only CLNM cases; Columns 3 and 4 are both central and lateral CLNM cases; Columns 5 and 6 are non-metastatic cases. Similar to the results of an independent test set 1, the feature energies of samples without CLNM were lower than those of the other samples. Figure 5b,c shows the distribution of the identification results and the confusion matrix, respectively. The abscissa of Figure 5b represents the model output; the green histogram represents cases that should be predicted as 0; the blue histogram represents cases that should be predicted as 1. As can be seen, the experimental results of an independent test set 2 are consistent with an independent test set 1, which further confirms the performance and effectiveness of the proposed model.

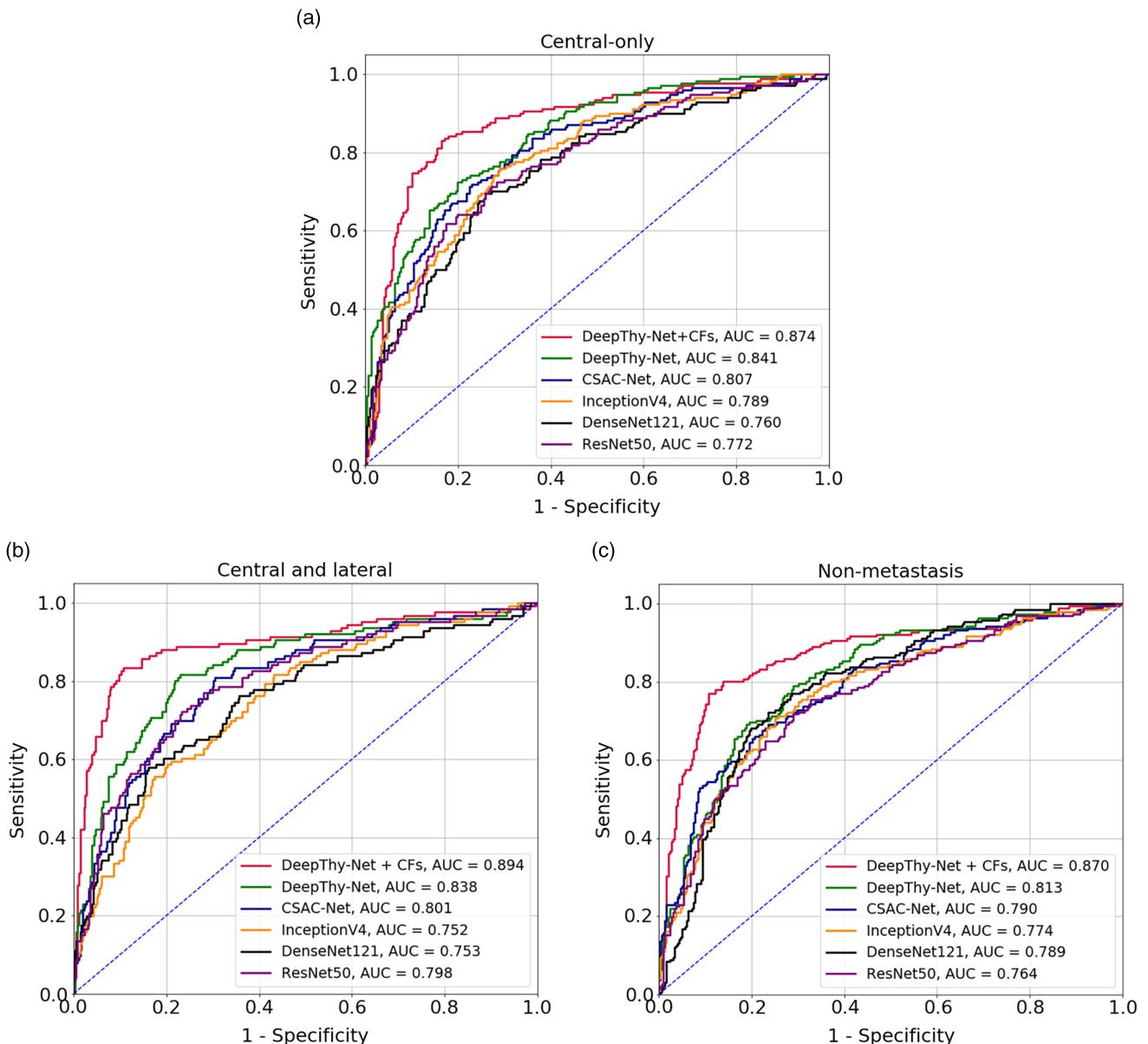


Figure 2. The receiver operating characteristic (ROC) curve of different methods in the independent test set 1. a) is the ROC curves of the central-only CLNM cases, CFs means clinical factors; b) is the ROC curves of both central and lateral CLNMs; c) is the non-metastatic cases.

Table 2 presents the prediction performance statistics for different methods. Specifically, for PTC cases with central-only CLNM, the sensitivity was 82.5% (118/143, 95% CI: 75.1–88.2%), the specificity was 82.1% (224/273, 95% CI: 76.9–86.3%), and the AUC value was 0.890 ($p < 0.01$). For cases involving both central and lateral CLNM, the overall sensitivity was 81.6% (93/114, 95% CI: 73.0–88.0%), the specificity was 87.7% (265/302, 95% CI: 83.4–91.1%), and the AUC value was 0.905 ($p < 0.01$). For non-metastatic cases, the sensitivity was 83.6% (133/159, 95% CI: 76.8–88.9%), the specificity was 82.5% (212/257, 95% CI: 77.2–86.8%), and the AUC value was 0.889 ($p < 0.01$). We also conducted an experiment to separately obtain the contribution of each clinical factor. The

results in the independent test set 2 are highly consistent with the independent test set 1, the extrathyroidal extension, age, and tumor size are the top three contribute to the AUC, and their contributions to the improvement of the AUC were 3.09, 2.64, and 2.48, respectively. Table S3, Supporting Information, shows the contributions of all clinical factors.

3. Discussion

The most common sites for CLNM in PTC are the central and lateral cervical regions.^[3,5,6] During a thyroidectomy, a decision regarding whether to perform a central and lateral neck LND

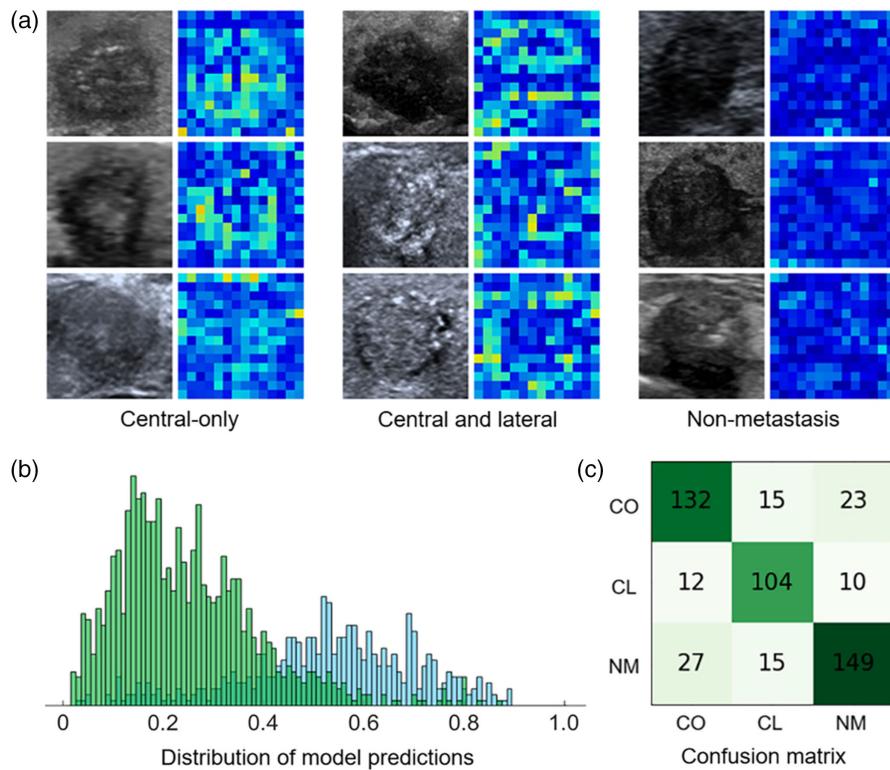


Figure 3. Experimental results of the independent test set 1. a) The US images and corresponding convolutional feature maps. b) The distribution of the output prediction of our model; c) The confusion matrix, where CO, CL, and NM represent the central-only CLNM, both central and lateral CLNMs and non-metastatic cases, respectively.

Table 1. The performance statistics of different methods in the independent test set 1.

Model	Central-only		Central and Lateral		Non-metastasis	
	TPR [%] (KLD)	TNR [%] (KLD)	TPR [%] (KLD)	TNR [%] (KLD)	TPR [%] (KLD)	TNR [%] (KLD)
ResNet50	68.2 (0.031)	74.1 (0.057)	69.8 (0.019)	77.8 (0.059)	66.0 (0.042)	73.0 (0.054)
DenseNet121	61.8 (0.047)	77.3 (0.044)	63.5 (0.029)	75.9 (0.069)	72.3 (0.027)	75.0 (0.046)
InceptionV4	75.9 (0.017)	71.0 (0.072)	65.1 (0.026)	70.1 (0.103)	74.9 (0.022)	69.6 (0.069)
CSAC-Net	74.1 (0.020)	71.9 (0.068)	73.0 (0.015)	74.0 (0.079)	73.8 (0.024)	68.9 (0.072)
DeepThy-Net without CFs	72.4 (0.023)	80.1 (0.034)	75.4 (0.012)	79.5 (0.051)	70.7 (0.031)	75.7 (0.044)
DeepThy-Net with CFs	82.9 (0.008)	83.0 (0.025)	83.3 (0.005)	88.6 (0.016)	80.1 (0.014)	83.1 (0.021)

requires an accurate metastasis risk assessment. Many studies have attempted to predict the CLNM in PTC, but these methods are limited. For example, the clinical factor-based methods suffer from relatively low accuracy, and in some cases, multi-center and multi-operator scenarios were not fully considered.^[21,22] Radiomic features and traditional ML methods (e.g., SVM and LR) for CLNM prediction are often influenced by unavoidable factors such as the degree of feature expression and limited model recognition capabilities.^[23,24]

To overcome these problems, we collected the US images and clinical factors of PTC from 3 hospitals for more than 11 years, and a total of 23 617 US images from 6,032 cases were collected from 23 operators and 19 pieces of US equipment. After that, a multimodal DL model called DeepThy-Net was developed to

predict different types of CLNMs in PTC using both US images and clinical factors. In addition, to avoid the potential central effect, we tested the model through independent test sets (from an independent center and not included in the training set) and obtained AUC values between 0.87 and 0.91. To the best of our knowledge, this is the first large-scale study to employ a DL model to assess the risks of multiple categories of CLNMs in PTC.

According to the 2021 Thyroid Cancer Clinical Practice guidelines issued by the National Comprehensive Cancer Network, PTC patients with CLNM need to undergo total thyroidectomy with LND, while PTC patients without evidence of CLNM can undergo a thyroid lobectomy alone when their tumors are $\leq 4\text{ cm}$.^[6] Moreover, due to the difficulty of preoperative

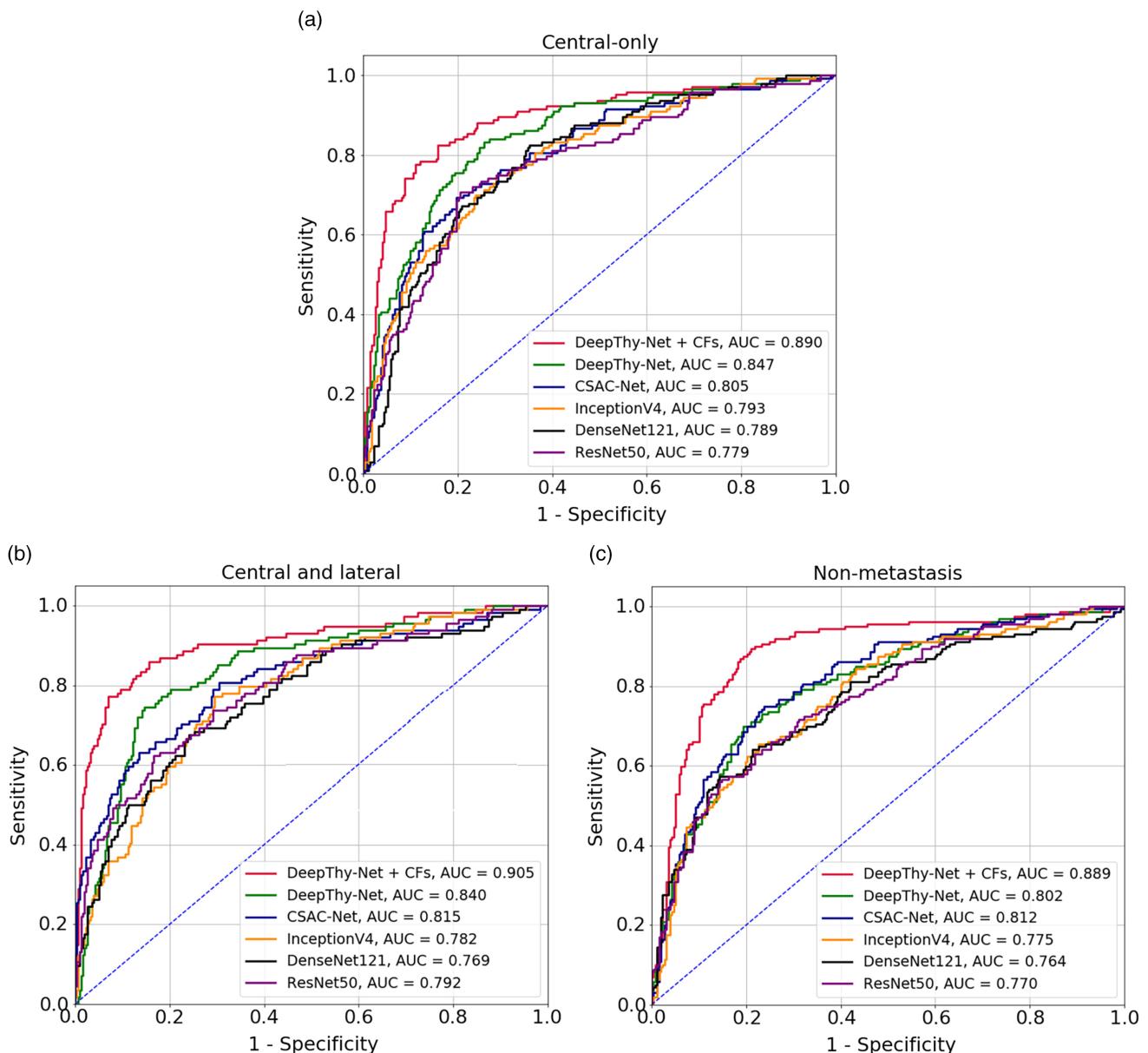


Figure 4. The ROC curve of different methods in the independent test set 2. a) is the ROC curves of the central-only CLNM cases; b) is the ROC curves of both central and lateral CLNMs; c) is the non-metastatic cases.

diagnosis of cervical lymph node metastasis, an active surveillance strategy instead of immediate surgery has received growing attention.^[33] Timely and accurate detection of the CLNM in PTC during the active surveillance process and application of effective treatment modalities are crucial for effective PTC management. Our DL-based diagnostic model can be easily incorporated into existing treatments and assist in the decision-making process. In particular, 1) For most tumors with size ≤ 1 cm and parts of tumors with size ≤ 4 cm, for which an active surveillance strategy is adopted in clinical practice, patients need to have routine ultrasound examinations every 3 to 6 months. If the DL model predicts lymph node metastasis during a regular ultrasound examination, the active surveillance strategy should

be suspended immediately. The examination level must be upgraded, including using CT/MRI with contrast for locally advanced disease, vocal cord mobility assessment, and performing FNA for suspicious lymph nodes. After confirming the specific metastatic lymph nodes, a total thyroidectomy and corresponding LND need to be performed; 2) For patients whose tumors are ≤ 4 cm in size and who are clinically treated by lobectomy, if no CLNM is found in the preoperative examination, but the DL model predicts that CLNM has occurred, the surgical level should be upgraded to total thyroidectomy with the corresponding LND. At the very least, the second round of US imaging (or other examination) should be performed to minimize the probability of missing the CLNMs; 3) For patients with tumor

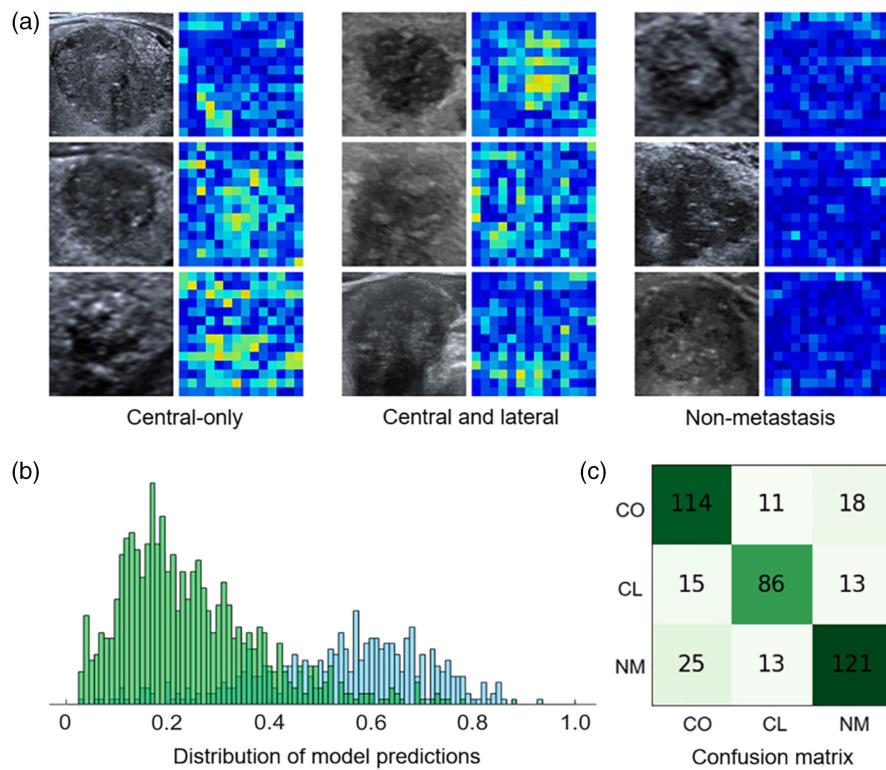


Figure 5. Experimental results of the independent test set 1. a) The US images and corresponding convolutional feature maps; b) The distribution of the output prediction of our model; c) The confusion matrix, where CO, CL, and NM represent the central-only, both central and lateral CLNMs and non-metastatic cases, respectively.

Table 2. The performance statistics of different methods in the independent test set 2.

Model	Central-only		Central and Lateral		Non-metastasis	
	TPR [%] (KLD)	TNR [%] (KLD)	TPR [%] (KLD)	TNR [%] (KLD)	TPR [%] (KLD)	TNR [%] (KLD)
ResNet50	70.6 (0.026)	78.0 (0.042)	63.2 (0.031)	79.5 (0.048)	73.6 (0.024)	64.2 (0.099)
DenseNet121	67.1 (0.033)	78.8 (0.040)	65.8 (0.027)	76.2 (0.064)	65.4 (0.042)	73.9 (0.052)
InceptionV4	74.8 (0.018)	69.2 (0.082)	74.6 (0.014)	70.5 (0.095)	71.7 (0.027)	65.8 (0.091)
CSAC-Net	72.7 (0.022)	72.9 (0.064)	71.9 (0.017)	73.8 (0.076)	76.7 (0.018)	70.0 (0.069)
DeepThy-Net without CFs	76.2 (0.016)	78.4 (0.041)	77.2 (0.011)	80.8 (0.042)	70.4 (0.03)	79.0 (0.034)
DeepThy-Net with CFs	82.5 (0.009)	82.1 (0.029)	81.6 (0.007)	87.7 (0.018)	83.6 (0.009)	82.5 (0.024)

size >4 cm who have undergone total clinical thyroidectomy, if the model predicts that CLNM has occurred, LND should be considered for the corresponding areas to further clean up potential cancer cells.

The study has some limitations. First, a very small number of PTCs have jump metastasis; that is, they directly metastasize across the central region to the lateral cervical lymph nodes. However, our study did not include such cases because the number of such cases is low. Second, although the amount of data was relatively substantial, the number of hospitals that contributed data was limited. In the future, we plan to encourage more hospitals to participate in the data collection tasks. This will not only increase the amount of data in the training set and improve the prediction accuracy, but also provide more opportunities to

further verify the robustness of the model. Third, we studied the CLNMs of PTC alone, but risk assessments for CLNM are also very important for other types of thyroid cancer. Extending the research to other types of thyroid cancer and further exploring the prediction of CLNMs in multiple pathological types will be of great value for the clinical management and treatment of thyroid cancer.

This retrospective study preliminarily verified the feasibility of using a multi-modal DL model to predict multiple CLNM patterns, however, it is still necessary to design prospective studies to further improve the generalization capability of the model. By prospectively designing the study protocol before data collection, more standard PTC ultrasound images and more abundant clinical information can be obtained, such as the addition of further

important gene mutations and proteomic information. In addition, the black-box properties of deep learning itself will also be a major challenge for future clinical applications. Although deep learning has achieved exciting diagnosis results in many diseases, such as lung cancer and fundus diseases, doctors still prefer transparent, understandable, and interpretable diagnostic models to better gain the confidence of supervisors and patients.^[34–36] In the future, we plan to introduce technologies such as explainable artificial intelligence (XAI) to improve the interpretability of our model. For example, we could use XAI and other technologies to analyze the changing rules of convolutional feature maps, and further obtain a more transparent and interpretable model.

4. Experimental Section

4.1. Patients

This study was approved by the ethics committees of all participating hospitals and the Chinese clinical trial registry center (ChiCTR2100053599), and all the patient-identifying information was removed. The Chinese Academy of Sciences, Zhejiang Cancer Hospital (IRB-2020-287); the Shanghai Tenth People's Hospital of Tongji University (IRB-2019-010-01); and the Affiliated Hangzhou First People's Hospital of Zhejiang University's School of Medicine (IRB-2019-200-01) were included in the study. **Figure 6** shows the data screening process,

and the inclusion criteria were as follows: 1) PTC patients who had received a thyroid US diagnosis and with available US images; 2) patients who had undergone thyroidectomy and LNDs; 3) the PTC and CLNMs were confirmed by a pathologist. The exclusion criteria were as follows: 1) incomplete or unqualified US images such as nodules that were too large for complete imaging; 2) patients who had received preoperative treatment; 3) for the training set, multifocal cases were excluded to ensure the certainty of the metastasis source; 4) patients with missing clinical information.

The PTC cases were randomly collected from January 16, 2009 to December 31, 2020. A total of 79 823 US images from 13 934 PTC patients were initially collected from the three centers, and after applying the exclusion criteria, 23 617 US images from 6,032 cases remained. The demographic data of the sample are summarized in **Table 3**. Instead of splitting the data set by the patient, we split it by hospital. This means that the training set and the test sets are independently collected from different hospitals, and each data set includes all categories. The data for the training set was obtained from Center 1 (Cancer Hospital of the University of Chinese Academy of Sciences, Zhejiang Cancer Hospital). Considering the challenge of determining the source of metastasis, we included only unifocal PTC cases to build the training set. The data for the two independent test sets were obtained from Centers 2 (Affiliated Hangzhou First People's Hospital, Zhejiang University School of Medicine) and 3 (the Shanghai Tenth People's Hospital). For the multifocal and multi-image cases in the test sets, we processed one lesion

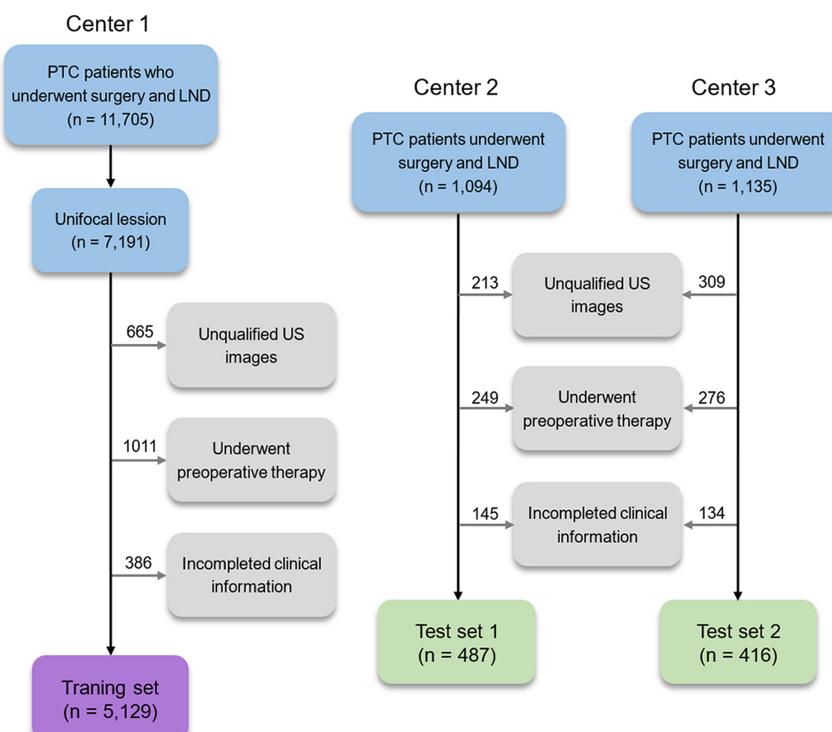


Figure 6. The patient enrollment process: the left and right sides show the enrollment processes for the training set and the independent test sets, respectively. Centre 1: Cancer Hospital of the University of Chinese Academy of Sciences (Zhejiang Cancer Hospital), Centre 2: Affiliated Hangzhou First People's Hospital, Zhejiang University School of Medicine, Centre 3: the Shanghai Tenth People's Hospital. For the training set, only a unifocal lesion was included, to ensure one-to-one correspondence between CLNM and primary focus.

Table 3. Descriptive statistics for the multi-center PTC cases.

Categories	Centre 1	Centre 2	Centre 3
	Training set	Test set 1	Test set 2
Total Number	5129	487	416
Age, mean(SD ^{a)})	43.1(13.3)	41.2(12.6)	44.5(11.4)
Sex			
Male	1443	137	110
Female	3686	350	306
CLNM status			
Only central	1463	170	143
Central & lateral	702	126	114
Non-metastasis	2964	191	159

^aSD: standard deviation.

and one image at a time and finally synthesized the recognition results according to the risk prediction of CLNMs.

4.2. Image Acquisition

The US images used were generated by 23 doctors using 19 pieces of equipment manufactured by the General Electric Company, Philips, Esaote, Siemens, and Toshiba (see Table S4, Supporting Information, for the detailed list). Patients were examined in the supine position with their necks fully exposed, and the thyroid glands were scanned. For each lesion, US images of the transverse and longitudinal sections of the target nodules were generated. All data were reviewed by three senior radiologists, each with more than 10 years of experience in the field of thyroid. Only quality-controlled data were included.

4.3. Pre-Processing and Performance Evaluations

First, the nodule areas were extracted from the US images and adjusted to 224×224 pixels. Clinical factors, such as extrathyroidal extension, capsular invasion, and lymphocytic thyroiditis, were collected from patients' examination reports. In addition

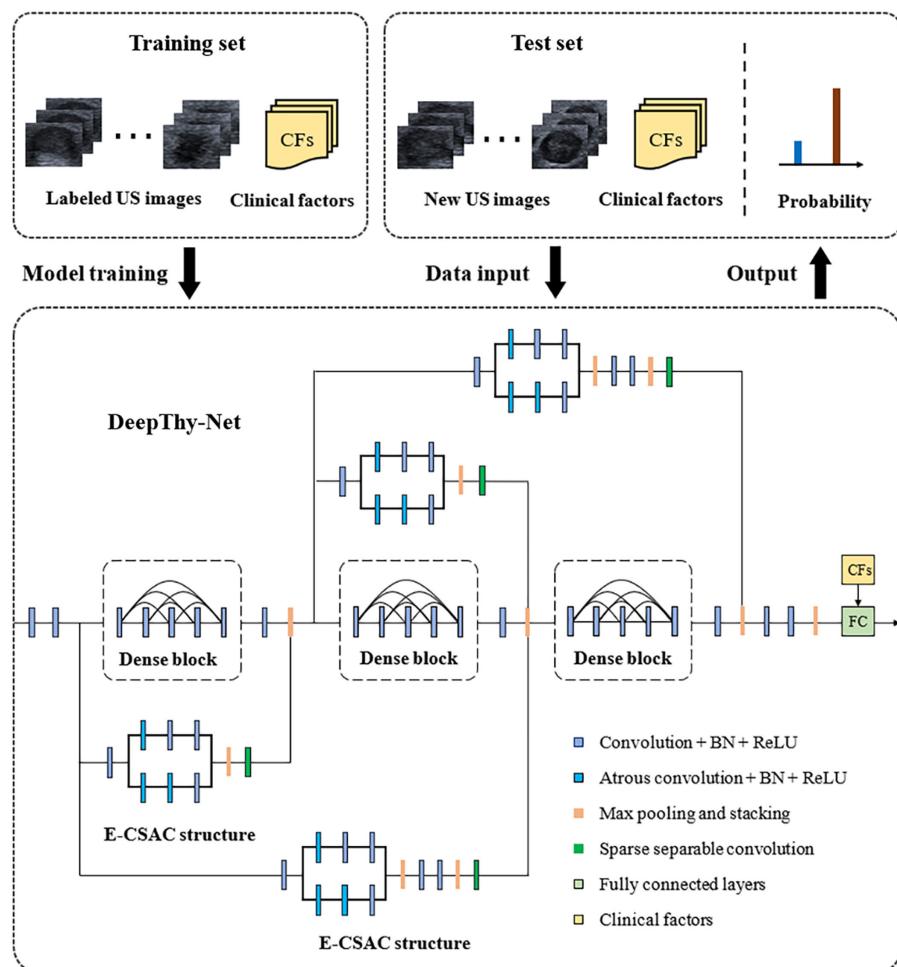


Figure 7. Training and testing procedure for the proposed DeepThy-Net. The top-left illustration shows the training set, and the top-right illustration shows the input test set and the output probability. The bottom illustration shows the framework of the DeepThy-Net; BN represents the batch normalization, ReLU represents the rectified linear units.

to image annotations, image enhancements, including random scaling and brightening by 0.9–1.1 times, horizontal or vertical reversals, and -10° to $+10^\circ$ random rotations, were also used to train the DL model.

The ROC, 95% confidence intervals (CIs), TPRs, TNRs, PPVs, NPVs, ACCs, and F1 score were calculated to assess performance. In addition, several typical samples were selected to generate convolutional feature maps. We built the model via TensorFlow 1.9 with Keras 2.1.4 and open-source Python 3.6.5 (the Python software foundation) on a Linux operating system (Ubuntu 16.04 operating system). The training process was executed on an NVIDIA GeForce TITAN Xp GPU with 11 G of memory and a 4.30 GHz Intel Core i7-7740X CPU with 64 G of memory. The statistical analyses were performed using the Python packages PyMC, PyLab, Scikit-learn, StatsModels, DeLong, and Seaborn.

4.4. The DL Model

A multimodal DL model called DeepThy-Net was proposed for predicting the risk of CLNM in PTC. The input for the model consisted of the nodules' US images and clinical factors. The model mainly consisted of three parts. First, several dense blocks were sequentially connected to build the baseline.^[31] Then four groups of cross-layer subnets were built, wherein the basic version of these subnets was from a cross-layer sparse atrous convolution (CSAC) network that we introduced earlier^[29]; it also performed well on the medical image analysis and the PASCAL VOC image classification benchmark dataset.^[37]

In this study, we improved the basic CSAC, and built four enhanced cross-layer sparse atrous convolution (E-CSAC) subnets with a new cross-layer hierarchy. After that, these E-CSAC subnets were integrated into the baseline to form the main framework of DeepThy-Net. Figure 7 shows the cross-layer structure, model training, and the testing process. Finally, in the full-connection layer, clinical factors were included as input nodes, and the output probabilities were calculated according to both clinical and convolutional image features. The loss function we used is the cross-entropy function with sparse coefficients

$$E(x, s) = - \sum_{x \in \Omega, c \in C} y_c(x) \log(p_c(x)) + \lambda \sum_{m_C} \|s\|_1 \quad (1)$$

where E denotes the loss function, x is the input sample, $c \in C$ represents the c^{th} channel, $y_c(x)$ is the label, s represents the sparse coefficient, $p_c(x)$ is the output probability, λ denotes the weight parameter and m_C represents the number of the E-CSAC module. The equation of SSC module which used the sparse constrained depthwise convolution is as follows

$$z_c^{l+1}(n_1, n_2) = s_c \left(\sum_{m_1 \in K_1} \sum_{m_2 \in K_2} k_c^{l+1}(m_1, m_2) f_c^l(n_1 - rm_1, n_2 - rm_2) + b_c^{l+1} \right) \quad (2)$$

where z denotes the convolutional result, l represents the l^{th} network layer, $c \in C$ represents the c^{th} channel, m and n are the

coordinates, $s_c \in S$ is the sparse weight coefficient for the c^{th} channel, k is the kernel of size $K_1 \times K_2$, f_c^l denotes the feature map of the l^{th} layer and c^{th} channel, r is the dilated rate and b is the bias. As is shown, the s_c term in Equation (2) is a coefficient that will constrain the weights of atrous convolution. Once the L_1 term $\|s\|_1$ in Equation (1) is minimized, this weight constraint will become a sparse constraint. Detailed parameter settings are presented in Table S5, Supporting Information. For each patient, after the images of all nodules were processed, the maximum probability of each region was recorded as the final risk probability of that region.

5. Conclusion

In this study, we collected US images and clinical factors of PTC patients from three hospitals for more than 11 years, and a multimodal deep learning model called DeepThy-Net was developed to predict different CLNM patterns in PTC. The experimental results in two independent test sets showed that the AUC values exceeded 0.87, indicating clinical applicability. The research can serve as an important supplement to fine-needle aspiration cytology and provide an important reference for treatment. For those PTC cases utilizing an active surveillance strategy, our method can also serve as an early warning tool.

Supporting Information

Supporting Information is available from the Wiley Online Library or from the author.

Acknowledgements

J.C.Y., Z.K.L., and W.W.Y. contributed equally to this work. This study was supported in part by the National Natural Science Foundation of China (82071946), the Natural Science Foundation of Zhejiang Province (LZY21F030001, LSD19H180001, LY20H180001), the funds from the University Cancer Foundation via the Sister Institution Network Fund at the University of Texas MD Anderson Cancer Center. The authors appreciate the valuable suggestions given by Dr. Xueqin Zhou from Esaote (Shenzhen) Medical Equipment Co., Ltd. and the Swiss Federal Institute of Technology in Zurich (ETHZ).

Conflict of Interest

The authors declare no conflict of interest.

Data Availability Statement

The data that support the findings of this study are available on request from the corresponding author. The data are not publicly available due to privacy or ethical restrictions.

Keywords

computational methods, deep learning, lymph-node metastasis, papillary thyroid cancer, ultrasound imaging

Received: April 23, 2022

Revised: June 20, 2022

Published online: August 10, 2022

- [1] H. Sung, J. Ferlay, R. L. Siegel, M. Laversanne, I. Soerjomataram, A. Jemal, F. Bray, *CA-Cancer J. Clin.* **2021**, *71*, 209.
- [2] R. L. Siegel, K. D. Miller, A. Jemal, *CA-Cancer J. Clin.* **2020**, *70*, 7.
- [3] M. Anastasios, D. Ramona, L. B. Naifa, J. R. Wang, R. Ferrarotto, C. Lu, M. D. Williams, G. B. Gunn, M. C. Hofmann, G. Cote, J. Sperling, N. D. Gross, E. M. Sturgis, R. P. Goepfert, S. Y. Lai, M. E. Cabanillas, M. Zafereo, *JAMA Oncol.* **2020**, *6*, 1397.
- [4] I. Lončar, S. P. J. van Dijk, M. J. H. Metman, J. F. Lin, S. Kruijff, R. P. Peeters, A. F. Engelsman, T. M. van Ginhoven, *Thyroid* **2021**, *31*, 1219.
- [5] A. Z. Kyle, M. H. Jerome, M. L. Angela, *JAMA* **2019**, *321*, 2010.
- [6] National Comprehensive Cancer Network, *NCCN Clinical Practice Guidelines in Oncology: Thyroid Carcinoma*, National Comprehensive Cancer Network **2020**, https://nccn.org/guidelines/category_1 (accessed: July 2020).
- [7] H. X. Nguyen, H. X. Nguyen, H. V. Nguyen, L. T. Nguyen, T. T. P. Nguyen, Q. V. Le, *Thyroid* **2021**, *31*, 410.
- [8] L. A. Shirley, N. B. Jones, J. E. Phay, *Front. Oncol.* **2017**, *7*, 00122.
- [9] J. H. Yu, Y. D. Deng, T. T. Liu, J. Zhou, X. Jia, T. Xiao, S. Zhou, J. Li, Y. Guo, Y. Wang, J. Zhou, C. Chang, *Nat. Commun.* **2020**, *11*, 4807.
- [10] L. Davies, C. H. Chang, B. Sirovich, R. M. Tuttle, M. Fukushima, Y. Ito, A. Miyauchi, *JAMA Otolaryngol. Head Neck Surg.* **2021**, *147*, 77.
- [11] Q. Z. Zhao, J. Ming, C. P. Liu, L. Shi, X. Xu, X. Nie, T. Huang, *Ann. Surg. Oncol.* **2013**, *20*, 746.
- [12] Z. Liu, J. Y. Lei, Y. Liu, Y. X. Fan, X. M. Wang, X. B. Lu, *Medicine* **2017**, *96*, e6240.
- [13] Y. S. Lim, J. C. Lee, Y. S. Lee, B. J. Lee, S. G. Wang, S. M. Son, I. J. Kim, *Surgery* **2011**, *150*, 116.
- [14] Q. Guan, Y. Wang, J. Du, Y. Qin, H. Lu, J. Xiang, F. Wang, *Ann. Transl. Med.* **2019**, *7*, 137.
- [15] C. X. Liu, C. Xiao, J. Chen, X. Y. Li, Z. J. Feng, Q. Y. Gao, Z. Liu, *BMC Cancer* **2019**, *19*, 622.
- [16] A. Nakajo, K. Minami, Y. Shinden, H. Toda, T. Hirashima, A. Nagata, Y. Nomoto, K. Maemura, S. Natsugoe, *Thyroid* **2020**, *50*, 778.
- [17] J. M. Stulak, C. S. Grant, D. R. Farley, G. B. Thompson, J. A. van Heerden, I. D. Hay, C. C. Reading, J. W. Charboneau, *Arch. Surg.* **2006**, *141*, 489.
- [18] J. Tran, M. Zafereo, *Thyroid* **2020**, *7*, ve.2020.0199.
- [19] Y. Tong, J. Li, Y. Huang, J. Zhou, T. Liu, Y. Guo, J. Yu, S. Zhou, Y. Wang, C. Chang, *Acad. Radiol.* **2021**, *28*, 1675.
- [20] Q. Wu, Y. Li, Y. Wang, B. Hu, *Clin. Transl. Oncol.* **2015**, *17*, 830.
- [21] J. L. Roh, J. M. Kim, C. I. Park, *Ann. Surg. Oncol.* **2011**, *18*, 2245.
- [22] X. Nie, Z. Tan, M. H. Ge, L. H. Jiang, J. F. Wang, C. M. Zheng, *Arch. Endocrinol. Metab.* **2016**, *60*, 492.
- [23] T. T. Liu, S. C. Zhou, J. H. Yu, Y. Guo, Y. Y. Wang, J. Zhou, C. Chang, *Technol. Cancer Res. Treat.* **2019**, *13*, 1617.
- [24] Y. Dai, J. Wang, S. Gao, *Adv. Intell. Syst.* **2022**, *3*, 2100263.
- [25] S. Jin, W. Y. T. Bao, Y. T. Yang, T. Bai, Y. B. Bai, *Sci. Rep.* **2018**, *8*, 17355.
- [26] L. Miccio, F. Cimmino, I. Kurelac, M. M. Villone, V. Bianco, P. Memmolo, F. Merola, M. Mugnano, M. Capasso, A. Iolascon, P. L. Maffettone, P. Ferraro, *View* **2020**, *1*, 20200034.
- [27] C. Pei, C. Liu, Y. Wang, D. Cheng, R. X. Li, W. K. Shu, C. Q. Zhang, W. L. Hu, A. H. Jin, Y. N. Yang, J. J. Wan, *Angew. Chem., Int. Ed.* **2020**, *59*, 10831.
- [28] Y. Su, X. X. Li, L. Huang, J. Cao, M. J. Zhang, V. Vedarethinam, W. Di, Z. Q. Hu, K. Qian, *Adv. Mater.* **2021**, *33*, 2007978.
- [29] J. C. Yao, T. Wang, G. H. Hou, D. Ou, W. Li, Q. D. Zhu, W. C. Chen, C. Yang, L. J. Wang, L. P. Wang, L. Y. Fan, K. Y. Shi, J. Zhang, D. Xu, Y. Q. Li, *Eur. Radiol.* **2021**, *31*, 7192.
- [30] K. M. He, X. Y. Zhang, S. Q. Ren, J. Sun, in *2015 IEEE Conference on Computer Vision and Pattern Recognition*, CVPR, Boston USA, Vol. 1 June, **2015**, p. 00385.
- [31] G. Huang, Z. Liu, L. V. D. Maaten, K. Q. Weinberger, in *2017 IEEE Conference on Computer Vision and Pattern Recognition*, CVPR, Honolulu USA, Vol. 3 July, **2017**, p. 06993.
- [32] C. Szegedy, S. Ioffe, V. Vanhoucke, A. Alemi, in *2016 IEEE Conference on Computer Vision and Pattern Recognition*, CVPR, Las Vegas USA, Vol. 2 June, **2016**, p. 07261.
- [33] S. Vaccarella, S. Franceschi, F. Bray, C. P. Wild, M. Plummer, L. D. Maso, *N. Engl. J. Med.* **2016**, *375*, 614.
- [34] L. Huang, L. Wang, X. M. Hu, S. Chen, Y. W. Tao, H. Y. Su, J. Yang, W. Xu, V. Vedarethinam, S. Wu, B. Liu, X. Z. Wan, J. T. Lou, Q. Wang, K. Qian, *Nat. Commun.* **2020**, *11*, 3556.
- [35] M. J. Zhang, L. Huang, J. Yang, W. Xu, H. Y. Su, J. Cao, Q. Wang, J. Pu, K. Qian, *Adv. Sci.* **2021**, *8*, 2101333.
- [36] W. Xu, J. X. Lin, M. Gao, Y. H. Chen, J. Cao, J. Pu, L. Huang, J. Zhao, K. Qian, *Adv. Sci.* **2020**, *7*, 2002021.
- [37] M. Everingham, S. M. A. Eslami, L. Van Gool, C. K. I. Williams, J. Winn, A. Zisserman, *Int. J. Comput. Vision* **2015**, *111*, 98.