



Article Evaluation of Cervical Lymph Node Metastasis in Papillary Thyroid Carcinoma Using Clinical-Ultrasound Radiomic Machine Learning-Based Model

Enock Adjei Agyekum ^{1,2}, Yong-Zhen Ren ^{1,2}, Xian Wang ¹, Sashana Sashakay Cranston ², Yu-Guo Wang ³, Jun Wang ⁴, Debora Akortia ⁵, Fei-Ju Xu ¹, Leticia Gomashie ⁶, Qing Zhang ¹, Dongmei Zhang ¹ and Xiaoqin Qian ^{1,*}

- ¹ Department of Ultrasound, Jiangsu University Affiliated People's Hospital, Zhenjiang 212002, China
- ² School of Medicine, Jiangsu University, Zhenjiang 212002, China
- ³ Department of Ultrasound, Nanjing Lishui District Hospital of Traditional Chinese Medicine, Nanjing 211200, China
- ⁴ Department of Biomedical Engineering, School of Communication and Information Engineering, Shanghai University, Shanghai 200444, China
- ⁵ Department of Clinical Microbiology, School of Medicine and Dentistry, Kwame Nkrumah University of Science and Technology, Kumasi 00233, Ghana
- ⁵ Department of Imaging, Klintaps University College, Accra 00233, Ghana
- Correspondence: yz_tyz1030@126.com

check for **updates**

Citation: Agyekum, E.A.; Ren, Y.-Z.; Wang, X.; Cranston, S.S.; Wang, Y.-G.; Wang, J.; Akortia, D.; Xu, F.-J.; Gomashie, L.; Zhang, Q.; et al. Evaluation of Cervical Lymph Node Metastasis in Papillary Thyroid Carcinoma Using Clinical-Ultrasound Radiomic Machine Learning-Based Model. *Cancers* 2022, 14, 5266. https://doi.org/10.3390/ cancers14215266

Academic Editors: Zhiyun Xue and Sameer Antani

Received: 15 September 2022 Accepted: 21 October 2022 Published: 26 October 2022

Publisher's Note: MDPI stays neutral with regard to jurisdictional claims in published maps and institutional affiliations.



Copyright: © 2022 by the authors. Licensee MDPI, Basel, Switzerland. This article is an open access article distributed under the terms and conditions of the Creative Commons Attribution (CC BY) license (https:// creativecommons.org/licenses/by/ 4.0/). **Simple Summary:** Accurate preoperative cervical lymph node metastasis (CLNM) prediction in papillary thyroid cancer (PTC) patients is essential for clinical treatment effectiveness, particularly for surgeons assessing the degree of surgical resection and the requirement for cervical lymph node dissection. As a result, a definite diagnosis of CLNM before surgery can assist the surgeon in selecting the best surgical technique and reducing the likelihood of reoperation. This research used several machine learning models based on clinical risk factors in conjunction with radiomics features to preoperatively evaluate CLNM in PTC patients, which can assist clinicians to choose a suitable treatment strategy for patients.

Abstract: We aim to develop a clinical-ultrasound radiomic (USR) model based on USR features and clinical factors for the evaluation of cervical lymph node metastasis (CLNM) in patients with papillary thyroid carcinoma (PTC). This retrospective study used routine clinical and US data from 205 PTC patients. According to the pathology results, the enrolled patients were divided into a non-CLNM group and a CLNM group. All patients were randomly divided into a training cohort (n = 143) and a validation cohort (n = 62). A total of 1046 USR features of lesion areas were extracted. The features were reduced using Pearson's Correlation Coefficient (PCC) and Recursive Feature Elimination (RFE) with stratified 15-fold cross-validation. Several machine learning classifiers were employed to build a Clinical model based on clinical variables, a USR model based solely on extracted USR features, and a Clinical-USR model based on the combination of clinical variables and USR features. The Clinical-USR model could discriminate between PTC patients with CLNM and PTC patients without CLNM in the training (AUC, 0.78) and validation cohorts (AUC, 0.71). When compared to the Clinical model, the USR model had higher AUCs in the validation (0.74 vs. 0.63) cohorts. The Clinical-USR model demonstrated higher AUC values in the validation cohort (0.71 vs. 0.63) compared to the Clinical model. The newly developed Clinical-USR model is feasible for predicting CLNM in patients with PTC.

Keywords: ultrasound radiomics; papillary thyroid carcinoma; cervical lymph node metastasis; thyroid neoplasms; machine learning

1. Introduction

Thyroid cancer is the most common endocrine neoplasm and the twelfth most common type of cancer [1]. Papillary thyroid carcinoma (PTC) is the most common subtype, accounting for approximately 80% of confirmed cases [2–6] and at least 85% of all welldifferentiated follicular thyroid cancers. Studies have revealed that lymph node metastasis is linked to disease recurrence and that it occurs in between 30% and 80% of PTC patients [7–9]. Cervical lymph node metastasis (CLNM) must be identified preoperatively and postoperatively in patients with thyroid cancer since its detection is crucial for limiting tumor recurrence and survival [10–13]. Ultrasound (US) is a safe and non-invasive imaging tool that can be used to count thyroid nodules as a first step [14]. A real-time dynamic exploration of the internal structure of the lesion and neighboring sections of the US has its own distinct advantages when compared to other examination methods. At the same time, high-frequency US is becoming a preferred method for detecting and diagnosing PTC patients' CLNM, as previous research has shown that preoperative US can aid in the diagnosis of PTC CLNM; however, some diagnostic limitations in US evaluation are unavoidable [15,16].

The accuracy of US diagnosis is heavily dependent on different operators' experience and subjective judgment, different interpretations of the same image, and different machine parameters. Moreover, it is worth noting that the human body contains approximately 800 lymph nodes, 300 of which are located in the head and neck. Because the lymphatic vessels and lymph nodes in the head and neck are so abundant, when PTC metastasizes along the lymphatic vessels to the cervical lymph nodes, there are often multiple metastases. It is difficult to detect all metastatic lymph nodes through imaging [15]. Furthermore, patients with suspicious lymph nodes must undergo fine-needle biopsy and preventive lymph node dissection to find lymph node metastasis. However, both processes are invasive and unnecessary for most patients without lymph node metastasis, and surgical complications such as hypoparathyroidism and recurrent laryngeal nerve injury will severely impact patients' quality of life [17]. As a result, it is critical to develop a more objective and accurate method for predicting CLNM in PTC patients prior to surgery in clinical practice. US radiomics (USR) is a new tool that can extract hundreds of quantitative features from medical images and combine them into a radiomic signature, an imagebased biomarker that can be used to diagnose diseases [18,19]. Machine learning based on radiomics is quickly gaining ground in the medical field [20]. It aims to detect patterns in imaging data and provide decision support by linking these patterns to treatment outcomes, allowing for greater precision in diagnosis and prognosis [21,22].

Clinical risk factors and the USR combined machine learning model, which incorporates clinical and radiomic features, can help provide complementary information for image features and collaborate to improve model performance by combining clinical and US image features. As a result, the combination of clinical risk factors and USR may be able to extract more valuable information from PTC US images, providing better prediction. The purpose of this research is to evaluate the diagnostic performance of a Clinical-USR machine learning model on conventional PTC US images in predicting CLNM in PTC patients.

2. Materials and Methods

The local ethics committee at Jiangsu University's Affiliated People's Hospital approved the retrospective study, and informed consent was waived. In total, 205 patients from our hospital were selected retrospectively between January 2015 and April 2020. The enrollment process is depicted in Figure 1.

All patients had routine 2-dimensional US. The training cohort (n = 143) and validation cohort (n = 62) were divided in a 7:3 ratio. The included cases were divided into CLNM and CLNM-free groups based on pathological results. The inclusion criteria were (1) nodules that had clear surgery- and puncture biopsy pathology-confirmed PTC; (2) routine US, with complete images, clear quality; and (3) did not receive chemoradiotherapy or other cancer treatment prior to surgery. Exclusion criteria were (1) pathology results that could not be identified as PTC nodules; (2) unclear images with incomplete nodules; (3) pregnant and lactating women; and (4) patients with a severe allergic history or severe cardiopulmonary disease.



Figure 1. Schematic diagram of the patient selection. PTC, papillary thyroid carcinoma.

2.1. Ultrasound Examination

All patients had a routine US examination prior to surgery, which was performed by well-trained technicians using a Philips Q5, Philips iU22 (both Healthcare, Eindhoven, The Netherlands) or a GE LOGIQ s8, LOGIQ E20, LOGIQ E9 (GE Medical Systems, American General, Boston, MA, USA) US system with a 5–12 MHz linear array transducer. The patient was positioned supine with the pillow removed in order to lower and slightly recline the head. This exposed as much of the neck as possible in order to perform a US examination of the thyroid and cervical area using longitudinal, horizontal continuous scanning and carefully scanning the lymph nodes in all areas of the neck.

This allowed for the measurement of thyroid tumor size or mass (maximum long axis of the nodule), tumor location (left lobe, right lobe, or isthmus), tumor position (upper, middle, or lower pole), aspect ratio (≤ 1 or >1), internal echo pattern (uniform or nonuniform), tumor border (clear, less clear), shape (regular or irregular), US ETE diagnosis (without ETE or with ETE), tumor peripheral blood flow (without or abundant), and tumor internal vascularization (without or abundant).

2.2. Region of Interests (ROIs) Segmentation

ROIs were manually drawn on US images by one radiologist with 15 years of experience in thyroid disease diagnosis using the software ITK-SNAP (version 3.8.0, http://www.itksnap.org. Accessed 10 August 2021) to indicate focal areas within the thyroid gland. The ROI was applied to the tumor's solid component, avoiding necrotic, hemorrhagic, and cystic areas. Thirty patients were chosen at random to assess the consistency of the ROI placements, and a second physician with 8 years of experience in thyroid US diagnosis independently placed ROIs on the relevant structures.

2.3. Radiomic Features Extraction

In total, 1046 USR features were extracted from these ROIs on US images using PyRadiomics (Version 2.2.0, https://github.com/Radiomics/pyradiomics. Accessed 10 August 2021). These features included first-order features, shape features, grey-level run length matrix (GLRLM) features, grey-level size zone matrix (GLSZM) features, grey-level dependence matrix (GLDM) features, neighborhood grey-tone dependency matrix (NGTDM) features, grey-level co-occurrence matrix (GLCM) features, and features derived from wavelet filter images containing first-order GLCM, GLRLM, GLSZM, GLDM, and NGTDM features.

2.4. Feature Selection and Construction of USR Model

To reduce bias and over-fitting in the study, the extracted features were normalized using a standard scalar. Because of the high dimensionality of the feature space, the number of features must be reduced to avoid the interference of a large number of redundant features in the data analysis, which has an impact on model construction and raises computational costs. The dataset was partitioned into training and validation cohorts.

We used Pearson's Correlation Coefficient (PCC) to reduce the row spatial dimension of the feature matrix so that each characteristic is relatively independent of the training data. Any two features with a PCC greater than 0.85 were considered redundant. Finally, in the training cohort, representative features were chosen using Recursive Feature Elimination (RFE) with 15-fold cross-validation. The Support Vector Machine with the linear kernel (SVM-L) [23], Support Vector Machine with radial basis function kernel (SVM-RBF) [23], LogisticRegressionCV (LRCV) [23], and Linear Discriminant Analysis (LDA) [23] classifiers were used to build the prediction models using the RFE's key features. In the validation procedure, the same feature sets were selected, and they were fed into the model.

The model's performance on the training and validation datasets was assessed using standard clinical statistics such as the area under the curve (AUC), sensitivity, specificity, negative predictive value (NPV), positive predictive value (PPV), and accuracy. The set of models that performed the best on the validation dataset was identified and evaluated by comparing diagnostic performance, and the best classifier was chosen. Figure 2 depicts the radiomic and machine learning workflow.



Figure 2. Schematic diagram of the radiomic workflow in building the machine learning models.

2.5. Construction of Clinical Model

In this study, basic clinical characteristics were analyzed, and then statistically significant characteristics with p < 0.05 were chosen to construct a Clinical Model, using SVM-L, SVM-RBF, LRCV, and LDA classifiers. The AUC and other relevant statistics were calculated to evaluate the diagnostic efficiency.

2.6. Development of the Clinical-USR Model

Statistically significant clinical factors were combined with extracted USR features to form a new feature set. A Clinical-USR model incorporating USR features and clinical characteristics was built using SVM-L, SVM-RBF, LRCV, and LDA classifiers, and the diagnostic efficacy of the combined prediction model was evaluated. The precision-recall curve shows the trade-off between precision and recall for different thresholds. High precision relates to a low false positive rate, and high recall relates to a low false negative rate.

High scores for both show that the classifier returns accurate results (high precision), as well as returning the majority of all positive results (high recall). Average precision (AP) summarizes such a plot as the weighted mean of precisions achieved at each threshold, with the increase in recall from the previous threshold used as the weight [23]. The precision–recall analysis of the model was assessed using the precision–recall curve. We also investigated the rate of false positives and false negatives, using the confusion matrix.

2.7. Statistical Analysis

Statistical analyses were carried out using Python (version 3.7 https://www.python. org/. Accessed 8 July 2021) and IBM SPSS Statistics for Windows version 26.0. (Armonk, NY, USA). To compare differences in categorical characteristics, Pearson's chi-square or Fisher's exact test were used. For continuous factors with normal distribution, the independent sample *t*-test was used, whereas, for continuous factors without normal distribution, the Mann–Whitney U test was used. A two-sided p < 0.05 indicated statistically significant differences.

PyRadiomics (version 2.2.0, https://github.com/Radiomics/pyradiomics. Accessed 10 August 2021) and scikit-learn version 1.2 [23] were used to extract USR features and build the prediction models. The AUC, sensitivity, specificity, accuracy, NPV, and PPV of each prediction model were calculated. The scikit-learn version 1.2 [23] was used to draw the precision–recall curve.

3. Results

3.1. Clinical Characteristics

A total of 205 patients with PTC were enrolled from ages 18–78 years, with an average age of 47.22 \pm 11.30 years, and a male-to-female ratio of 1:3.56. CLNM was excluded in 107 patients and confirmed in 98. Using stratified sampling, all patients were randomly assigned to a training group (n = 143) and a validation group (n = 62). Table 1 displays the clinical data for the training and validation groups. Table 2 displays the clinical data for the CLNM groups. Age, US CLNM diagnosis, mass, capsular invasion, shape, US ETE diagnosis, internal echo, aspect ratio, and tumor internal vascularization (all p < 0.05) did differ significantly between the two groups.

Table 1. Patient characteristics of the training and validation cohorts.

Characteristic	Training Cohort (<i>n</i> = 143)	Validation Cohort (<i>n</i> = 62)	p
Age , mean \pm SD, years	47.29 ± 11.31	47.06 ± 11.36	0.981
Sex, n			
Female	114	46	
Male	29	16	0.242
	Ultrasound Character	istic	
Tumor size (mass)	9.22 ± 6.85	9.61 ± 6.70	0.662
Tumor location			
Left lobe	60	18	
Right lobe	48	29	
Isthmus	35	15	0.142
Tumor position			
Upper pole	73	25	
Middle pole	5	2	
Inferior pole	65	35	0.346
Internal echo pattern			
Uniform	29	10	0.312
Nonuniform	114	52	
Tumor border			
Clear	48	11	0.015 *
unclear	95	51	

Characteristic	Training Cohort (<i>n</i> = 143)	Validation Cohort (<i>n</i> = 62)	р
Tumor internal			
vascularization			
Without	75	38	0.155
Abundant	68	24	
Tumor Peripheral blood			
flow			
Without	80	31	0.2(2
Abundant	63	31	0.263
Ultrasound ETE			
diagnosis			0.405
Without ETE	133	57	0.495
With ETE	10	5	
Aspect ratio			
≤ 1	91	38	0.434
>1	52	24	
Shape			
Regular	106	44	0.380
Irregular	37	18	
Ultrasound CLNM			
diagnosis			0.267
Without CLNM	80	37	0.367
With CLNM	63	25	
	Postoperative diagno	osis	
capsular invasion			
Negative	75	31	0.432
Positive	68	31	

Table 1. Cont.

ETE, extrathyroidal extension; CLNM, cervical lymph node metastasis; SD, standard deviation. *: statistically significant.

 Table 2. Patient characteristics of the PTC with CLNM and PTC without CLNM groups.

Characteristic	CLNM (-) (<i>n</i> = 107)	CLNM (+) (<i>n</i> = 98)	p
Age , mean \pm SD, years	48.97 ± 10.81	45.31 ± 11.56	0.035 *
Sex, n			
Female	80	80	
Male	27	18	0.154
I	Ultrasound Characteristic		
Tumor size (mass)	7.30 ± 5.01	11.57 ± 7.74	0.00 *
Tumor location			
Left lobe	44	34	
Right lobe	40	37	
Isthmus	23	27	0.515
Tumor position			
Upper pole	53	45	
Middle pole	4	3	
Inferior pole	50	50	0.818
Internal echo pattern			
Uniform	26	13	0.033 *
Nonuniform	81	85	
Tumor border			
Clear	30	29	0.463
unclear	77	69	
Tumor internal vascularization			0.001 *
Without	71	42	
Abundant	36	56	

Characteristic	CLNM (-)	CLNM (+)	р
	(n = 107)	(n = 98)	•
Tumor Peripheral blood flow			0.236
Without	61	50	
Abundant	46	48	
Ultrasound ETE diagnosis			
Without ETE	104	86	0.009 *
With ETE	3	12	
Aspect ratio			
≤ 1	55	74	0.000 *
>1	52	24	
Shape			
Regular	85	65	0.025 *
Irregular	22	33	
Ultrasound CLNM diagnosis			
Without CLNM	68	49	0.035 *
With CLNM	39	49	
	Postoperative diagnosis		
Capsular Invasion	1 0		
Negative	63	43	0.022 *
Positive	44	55	

Table 2. Cont.

_

ETE, extrathyroidal extension; CLNM, cervical lymph node metastasis; SD, standard deviation. *: statistically significant.

3.2. Clinical Model Construction

A clinical model was built using nine variables that were statistically significant. RFECV was also used to select the best variables, and the final four key variables that were chosen were age, mass, shape, and internal echo (Figures 3A–C and 4A–C).





(A)

Figure 3. Cont.



(C)

Figure 3. Selected features after RFE. (A) In the Clinical model, features were reduced to four features in the training cohort. (B) In the USR model, features were reduced to nine features in the training cohort. (C) In the Clinical-USR model, features were reduced to five features in the training cohort.



Figure 4. Cont.



0 50 100 150 200 Number of features selected

(C)

Cross validation score

Figure 4. Recursive feature elimination (RFE) with 15-fold cross-validation; number of features selected vs. cross-validation score. (A) Clinical model. (B) USR model. (C) Clinical-USR model.

The SVM-L classifier had the highest AUC (0.63) in the validation cohort compared with the other classifiers, so it was selected as the Clinical model (Table 3 and Figure 5).

Table 3.	Predictive performance	of the	machine	learning	models	for the	training	and	validation
cohorts.									

		Training	Cohort	Validation Cohort								
	ACC	AUC	SEN	SPEC	PPV	NPV	ACC	AUC	SEN	SPEC	PPV	NPV
Clinical model												
LDA	0.71	0.77	0.58	0.82	0.75	0.68	0.56	0.60	0.55	0.58	0.53	0.59
LRCV	0.71	0.77	0.61	0.81	0.75	0.69	0.55	0.59	0.55	0.55	0.52	0.58
SVM-L	0.72	0.77	0.61	0.82	0.76	0.69	0.61	0.63	0.59	0.64	0.59	0.64
SVM-RBF	0.73	0.79	0.78	0.69	0.70	0.77	0.56	0.60	0.69	0.45	0.52	0.63
					US	SR model						
LDA	0.76	0.83	0.75	0.77	0.75	0.77	0.66	0.74	0.52	0.79	0.68	0.65
LRCV	0.79	0.85	0.78	0.80	0.78	0.80	0.61	0.72	0.49	0.76	0.62	0.61
SVM-L	0.73	0.85	0.77	0.70	0.71	0.76	0.66	0.72	0.52	0.79	0.68	0.65
SVM-RBF	0.77	0.86	0.77	0.77	0.76	0.78	0.63	0.70	0.78	0.76	0.64	0.63
					Clinica	al-USR mo	del					
LDA	0.71	0.79	0.65	0.76	0.71	0.70	0.61	0.68	0.55	0.66	0.60	0.63
LRCV	0.70	0.78	0.61	0.78	0.72	0.68	0.68	0.71	0.62	0.73	0.67	0.69
SVM-L	0.71	0.80	0.65	0.76	0.71	0.70	0.65	0.69	0.59	0.70	0.63	0.66
SVM-RBF	0.73	0.80	0.64	0.82	0.77	0.71	0.63	0.69	0.55	0.70	0.62	0.63

AUC, area under the curve; ACC, accuracy; SEN, sensitivity; SPEC, specificity; NPV, negative predictive value; PPV, positive predictive value; USR, ultrasound radiomic; SVM-L, support vector machine with linear kernel; SVM-RBF, support vector machine with radial basis function kernel; LDA, linear discriminant analysis; LRCV, logistic Regression CV.



Figure 5. ROC curve of the classifiers used in building the Clinical model. ROC, receiver operating characteristic. (**A**) Support vector classifier with radial basis function kernel. (**B**) Support vector classifier with linear kernel. (**C**) Logistic regression classifier. (**D**) Linear discriminant analysis.

Based on US findings evaluated by an experienced sonographer, the US-reported status of CLNM by an experienced sonographer in directly detecting suspicious malignant cervical lymph node had an AUC of 0.568 (Figure 6), a sensitivity of 0.500, specificity of 0.636, NPV of 0.581, and PPV of 0.557.



Figure 6. ROC curve of the US reported status of CLNM by an experienced sonographer. ROC, receiver operating characteristic.

3.3. Construction of USR Model

A total of 1046 USR features were extracted from each 2-D US image. Any two features with a PCC > 0.85 were considered redundant. Nine features were selected in the training cohort using RFE and stratified 15-fold cross-validation (Figures 3B and 4B).

The features chosen were exponential_firstorder_Kurtosis, gradient_glszm_Large AreaEmphasis, lbp_2D_glrlm_RunLengthNonUniformity, lbp_2 D_glszm_SmallArea Emphasis, logarithm_glrlm_LongRunLowGrayLevelEmphasis, logarithm_glszm_Zone Entropy, square_glszm_SizeZoneNonUniformityNormalized, wavelet_LH_firstorder_ Skewness, and wavelet_HL_glcm_ClusterShade. In the validation cohort, the LDA classifier had the highest AUC (0.74) value compared with the other classifiers. The LDA was chosen for the USR model. Figure 7 shows the ROC curves of the USR model using various classifiers. Table 3 displays detailed information about the classifiers' prediction performance. LDA decision function is given as follows:

 $\label{eq:loss} \begin{array}{l} \mbox{Decision value} = 0.919 \times \mbox{exponential_firstorder_Kurtosis} - 1.327 \times \mbox{gradient_glszm} \\ \mbox{LargeAreaEmphasis} - 0.886 \times \mbox{lbp_2D_glrlm_RunLengthNonUniformity} + 0.736 \times \mbox{lbp_2D_glszm_SmallAreaEmphasis} + 0.596 \times \mbox{logarithm_glrlm_LongRunLowGrayLevel} \\ \mbox{Emphasis} + 1.273 \times \mbox{logarithm_glszm_ZoneEntropy} - 0.512 \times \mbox{square_glszm_SizeZone} \\ \mbox{NonUniformityNormalized} + 0.856 \times \mbox{wavelet_LH_firstorder_Skewness} - 0.520 \times \mbox{wavelet} \\ \mbox{HL_glcm_ClusterShade}. \end{array}$



Figure 7. ROC curve of the classifiers used in building the USR model. ROC, receiver operating characteristic. (**A**) Support vector classifier with radial basis function kernel. (**B**) Support vector classifier with linear kernel. (**C**) Logistic regression classifier. (**D**) Linear discriminant analysis.

3.4. Construction of Clinical-USR Model

We created a new feature set by combining the extracted USR features and nine clinical parameters.

Through PCC and RFE (Figure 4C), five key features, namely original_glszm_Zone Entropy, original_ngtdm_Busyness, exponential_glszm_LargeAreaEmphasis, wavelet-LH _glszm_LargeAreaLowGrayLevelEmphasis, and mass (Figure 3C, Table 4) were selected to construct the Clinical-USR model.

	Feature Name	Coefficient
original glszm ZoneEntropy	GLSZM	0.642
original ngtdm Busyness	NGTM	0.766
exponential glszm LargeAreaEmphasis	GLSZM	-0.652
waveletLH glszm LargeAreaLowGrayLevelEmphasis	GLSZM	-0.782
Tumor mass	N/A	1.132
N/A· Not applicable		

Table 4. Selected features and their coefficients during the recursive feature elimination procedure.

In the validation cohort, the LRCV classifier had the highest AUC (0.71) and ACC (0.68) values compared with the other three classifiers. The LRCV was chosen for the Clinical- USR model. Figure 8 shows the ROC curves of the Clinical- USR model using various classifiers.





Detailed information about the prediction performance of the classifiers is shown in Table 3.

LRCV decision function of the Clinical-USR model is given as follows:

 $\label{eq:loss} \begin{array}{l} \mbox{Decision value} = 0.642 \times \mbox{original_glszm_ZoneEntropy} + 0.766 \times \mbox{original_ngtdm_} \\ \mbox{Busyness} - 0.652 \times \mbox{exponential_glszm_LargeAreaEmphasis} - 0.782 \times \mbox{waveletLH_glszm_} \\ \mbox{LargeAreaLowGrayLevelEmphasis} + 1.132 \times \mbox{mass}. \end{array}$

On the training cohort containing 143 patients, the classifier correctly classified 42 (true positive) out of 69 without CLNM, and correctly classified 58 (true negative) out of 74 patients having CLNM (Figure 9A,B).





In the validation cohort containing 62 patients, the Clinical-USR model correctly classified 18 (true positive) out of 29 patients with no CLNM, and correctly classified 24 (true negative) out of 33 patients having CLNM (Figure 9B). The precision–recall curve of the Clinical-USR model is shown in Figure 10.



Figure 10. Precision-recall curve of the Clinical-USR model.

Further multivariate analysis of the features used in building the Clinical-USR model indicated that original glszm ZoneEntropy, original ngtdm Busyness, WavletLH glszm LargeAreaLow GrayLevelEmphasis, and mass are independent predictors of CLNM in PTC (Table 5).

Feature Name	В	SE	Wald	df	p	Exp (B)
Original glszm ZoneEntropy	0.696	0.237	8.627	1	0.003	2.006
Original ngtdm Busyness	1.248	0.546	5.215	1	0.022	3.482
exponential glszm LargeAreaEmphasis	-2.294	2.013	1.298	1	0.254	0.101
WaveletLH glszm						
LargeAreaLow	-896	0.459	3.812	1	0.051	0.408
GrayLevelEmphasis tumour size(mass)	1.264	0.300	17.726	1	0.000	3.541

Table 5. Results of multivariate regression analysis of selected features in diagnosis of CLNM in PTC patients.

4. Discussion

Early diagnosis of tumors with lymph node metastatic potential is crucial for preventing the fast progression of PTC. It also serves as a key marker for monitoring PTC progression and a predictor of a poor prognosis. Surgery is presently the first-line therapy for PTC patients who have received a clinical diagnosis. Individuals with a suspected risk of CLNM must have cervical preventative lymph node dissection; nevertheless, it is questionable what PTC patients can gain from prophylactic lymph node dissection. According to studies, early detection of lymph node metastases results in fewer recurrences and a better prognosis in PTC patients, who have a higher chance of surviving cancer [10,12]. People with PTC who have CLNM have higher rates of death and recurrence than patients without CLNM [24,25]. Furthermore, surgery might potentially impair postoperative parathyroid function and increase the chance of laryngeal recurrent nerve damage. It is also more strenuous on the surgeon and demands a greater degree of skill [26,27]. Therefore, a definitive diagnosis of CLNM before surgery can help the surgeon choose the optimal surgical strategy and reduce the chance of reoperation.

For clinical treatment to be effective, accurate preoperative CLNM prediction in PTC patients is crucial, especially for the surgeons to assess the extent of surgical resection and the need for cervical lymph node dissection [28,29]. In many patients, CLNM may not reveal any aberrant findings in preoperative US examination [30] due to US's unreliability in visualizing deep anatomic structures or structures that are acoustically shadowed by air and bone [31]. Additionally, because US examination is an empirical diagnostic that is greatly impacted by the operator's experience, it is subject to interobserver variability when determining CLNM [32]. We employed the Clinical-USR approach to identify CLNM before surgery in our study. The developed Clinical-USR model is a simple-to-use diagnostic and prognostic tool that could help patients without CLNM avoid unnecessary surgery. The AUC values revealed that the Clinical-USR model could discriminate between PTC patients with CLNM and PTC patients without CLNM in the training (AUC: 0.78) and validation (AUC: 0.71) cohorts. When compared to the Clinical model, the USR model had higher AUCs in both the training (0.83 vs. 0.77) and validation (0.74 vs. 0.63) cohorts. In both the training (AUC: 0.78 vs. 0.77) and validation (AUC: 0.71 vs. 0.63) cohorts, the Clinical-USR model showed higher AUC values than the Clinical model.

Original glszm ZoneEntropy, original ngtdm Busyness, exponential glszm LargeArea Emphasis, and wavelet LH glszm LargeAreaLowGrayLevelEmphasis are the selected USR features that were used in constructing the Clinical-USR model. In an image, GLSZM measures gray level zones. The number of related voxels with the same gray-level intensity is defined as a gray-level zone. If the distance between two voxels is 1 according to the infinity norm, they are considered linked (26 connected regions in 3D, 8 connected regions in 2D) [33]. The surrounding gray-tone difference matrix is used to calculate the difference between a gray value and its close neighbors' average gray values [34]. The distribution of zone widths and gray levels is measured for uncertainty/randomness using original glszm ZoneEntropy. More heterogeneity in the texture patterns is indicated by a higher value. Exponential glszm LargeAreaEmphasis is a measurement of how big area size zones are distributed, with a higher value indicating larger size zones and coarser textures.

Wavelet-LH glszm LargeAreaLowGrayLevelEmphasis glszm measures the fraction of the image's combined distribution of bigger size zones with lower gray-level values. Original ngtdm Busyness is a measure of the difference between a pixel and its neighbor. A high busyness rating suggests a 'busy' image, with fast changes in intensity between pixels and their surroundings.

This suggests that the newly constructed Clinical-USR model contains additional data that are significantly associated with PTC with CLNM but not a standard risk factor. PTC density and increased nonuniformity, for example, are features that are difficult to measure with the naked eye. However, in PTC, these characteristics are linked to tissue heterogeneity. For CLNM evaluation, the Clinical-USR model considers PTC heterogeneity, a measurable trait associated with the degree of malignancy in PTC. As a result, the quantitative Clinical-USR approach not only overcomes the subjectivity of standard US imaging diagnosis, but also makes use of a great deal of data that the naked eye cannot see.

According to postoperative pathology results, the US-reported status of CLNM by an experienced sonographer in directly detecting suspicious malignant cervical lymph node had an AUC of 0.568, a sensitivity of 0.500, specificity of 0.636, NPV of 0.581, and PPV of 0.557. There are approximately 200 lymph nodes distributed in the neck. Because the lymphatic vessels and lymph nodes in the neck are so numerous, when PTC metastasizes

to the cervical lymph nodes, multiple metastases are common, and it is difficult to detect all metastatic lymph nodes with US. In this study, the clinical-USR model outperformed an experienced sonographer's US-reported CLNM status.

Li et al. [35] developed a deep learning-based computer-aided model for the diagnosis of CLNM in patients with PTC, and then tested the model's accuracy with a validation set. In the validation cohort, the AUC of their model for the diagnosis of CLNM was higher than that of the current study (0.79 vs. 0.71). The reason for this could be the deep learning approach they used in their research. In our research, we used machine learning. Most of the applicable features in traditional machine learning approaches must be identified by a domain expert in order to minimize data complexity and make patterns more obvious for learning algorithms to work. Deep learning algorithms have the biggest benefit in that they aim to learn high-level features from data incrementally. This reduces the need for domain expertise and the extraction of hardcore features.

In their study, Tian et al. [36] developed a CLNM prediction model based on clinical risk variables. The AUC of their model for the diagnosis of CLNM in the validation cohort was slightly lower than that of the current study (0.70 vs. 71). Zou et al. [37] used machine learning models based on US data to determine the probability of CLNM. The AUC of their model for the diagnosis of CLNM was slightly higher in the validation cohort compared with the current study (0.73 vs. 0.71). The slightly higher AUCs in the preceding studies could be attributed to the larger number of patients included in their studies. Large datasets improve predictive model performance in machine learning. Many studies, including the abovementioned studies, have focused on predicting CLNM based on either US clinical risk factors or USR features, but not their combination. We developed a joint predictive model incorporating clinical risk factors and USR features, using four different machine learning algorithms in our study. In this study, we also built three different models: The Clinical model, the USR model, and the Clinical-USR model, and we employed and compared the performances of several machine learning classifiers in the construction of each model, and the best-performing classifier was chosen to represent the models. One statistic for evaluating the excellence of a model's output is precision-recall. While recall assesses the number of really relevant results returned, precision evaluates the relevance of the results. In the stairstep portion of Figure 6, recall and precision are inversely correlated; near the margins of these steps, a tiny adjustment in the threshold significantly affects precision while only slightly increasing recall. From the precision–recall analysis of the Clinical-USR model, the AP of the model was 0.63, which is satisfactory.

5. Conclusions

Because this was a retrospective study, there may have been a case selection bias that influenced the study results. Furthermore, we established and validated our Clinical-USR model for distinguishing CLNM in a single hospital and employed only greyscale US images in our Clinical-USR model. We do, however, intend to integrate multimodal USR characteristics in future studies. In summary, for the prediction of CLNM, a Clinical-USR model based on clinical factors and USR features was developed. The newly developed Clinical-USR model is non-invasive and feasible for predicting CLNM in PTC patients.

Author Contributions: Conceptualization, E.A.A. and X.Q.; methodology, E.A.A.; software, E.A.A. and D.A.; validation, S.S.C., L.G. and Y.-Z.R.; formal analysis, E.A.A.; investigation, E.A.A. and F.-J.X.; resources, E.A.A., J.W. and Y.-G.W.; data curation, E.A.A., X.W., Y.-Z.R., D.Z. and Q.Z.; writing—original draft preparation, E.A.A.; writing—review and editing, E.A.A., X.Q., D.A. and S.S.C.; visualization, E.A.A. and Y.-Z.R.; supervision, X.Q. and J.W.; project administration, E.A.A. and X.Q.; funding acquisition, X.Q. All authors have read and agreed to the published version of the manuscript.

Funding: This study was financially supported by the National Natural Science Foundation of China (Project No.: 81971629), Zhenjiang Commission of Science and Technology (Project No. SH2020046), and Postgraduate Research & Practice Innovation Program of Jiangsu Province (Project No. SJCX21 1728).

Institutional Review Board Statement: The study was conducted in accordance with the Declaration of Helsinki and approved by the Jiangsu University-Affiliated People's Hospital Ethics Committee. (SQK-20200146-Y, 30 December 2020).

Informed Consent Statement: Patient consent was waived by the ethics committee due to the retrospective nature of the study.

Data Availability Statement: The original contributions presented in the study are included in the article. Further inquiries can be directed to the corresponding authors.

Conflicts of Interest: The authors declare no conflict of interest.

References

- Bonjoc, K.-J.; Young, H.; Warner, S.; Gernon, T.; Maghami, E.; Chaudhry, A. Thyroid cancer diagnosis in the era of precision imaging. J. Thorac. Dis. 2020, 12, 5128–5139. [CrossRef]
- 2. Qiu, Z.; Li, H.; Wang, J.; Sun, C. miR-146a and miR-146b in the diagnosis and prognosis of papillary thyroid carcinoma. *Oncol. Rep.* **2017**, *38*, 2735–2740. [CrossRef]
- Lonjou, C.; Damiola, F.; Moissonnier, M.; Durand, G.; Malakhova, I.; Masyakin, V.; Calvez-Kelm, L.; Cardis, E.; Byrnes, G.; Kesminiene, A.; et al. Investigation of DNA re-pair-related SNPs underlying susceptibility to papillary thyroid carcinoma reveals MGMT as a novel candidate gene in Belarusian children exposed to radiation. *BMC Cancer* 2017, *17*, 328. [CrossRef]
- 4. Schneider, D.F.; Chen, H. New developments in the diagnosis and treatment of thyroid cancer. *CA Cancer J. Clin.* **2013**, *63*, 373–394. [CrossRef]
- Lopez-Campistrous, A.; Adewuyi, E.E.; Benesch, M.G.; Ko, Y.M.; Lai, R.; Thiesen, A.; Dewald, J.; Wang, P.; Chu, K.; Ghosh, S.; et al. PDGFRα Regulates Follicular Cell Differentiation Driving Treatment Resistance and Disease Recurrence in Papillary Thyroid Cancer. *eBioMedicine* 2016, *12*, 86–97. [CrossRef]
- 6. Chen, F.; Jin, Y.; Feng, L.; Zhang, J.; Tai, J.; Shi, J.; Yu, Y.; Lu, J.; Wang, S.; Li, X.; et al. RRS1 gene expression involved in the progression of papillary thyroid carcinoma. *Cancer Cell Int.* **2018**, *18*, 20. [CrossRef]
- Forest, V.I.; Clark, J.R.; Ebrahimi, A.; Cho, E.A.; Sneddon, L.; Gao, K.; O'brien, C.J. Central compartment dissection in thy-roid papillary carcinoma. *Ann. Surg.* 2011, 253, 123–130. [CrossRef] [PubMed]
- 8. Moo, T.A.; McGill, J.; Allendorf, J.; Lee, J.; Fahey, T.; Zarnegar, R. Impact of prophylactic central neck lymph node dissection on early recurrence in papillary thyroid carcinoma. *World J Surg.* **2010**, *34*, 1187–1191. [CrossRef] [PubMed]
- 9. Lundgren, C.I.; Hall, P.; Dickman, P.W.; Zedenius, J. Clinically significant prognostic factors for differentiated thyroid carcinoma: A population-based, nested case-control study. *Cancer* **2006**, *106*, 524–531. [CrossRef]
- 10. Sivanandan, R.; Soo, K.C. Pattern of cervical lymph node metastases from papillary carcinoma of the thyroid. *Br. J. Surg.* 2001, *88*, 1241–1244. [CrossRef]
- Robenshtok, E.; Fish, S.; Bach, A.; Domínguez, J.M.; Shaha, A.; Tuttle, R.M. Suspicious Cervical Lymph Nodes Detected after Thyroidectomy for Papillary Thyroid Cancer Usually Remain Stable Over Years in Properly Selected Patients. *J. Clin. Endocrinol. Metab.* 2012, 97, 2706–2713. [CrossRef] [PubMed]
- Randolph, G.W.; Duh, Q.-Y.; Heller, K.S.; LiVolsi, V.A.; Mandel, S.J.; Steward, D.L.; Tufano, R.P.; Tuttle, R.M. The Prognostic Significance of Nodal Metastases from Papillary Thyroid Carcinoma Can Be Stratified Based on the Size and Number of Metastatic Lymph Nodes, as Well as the Presence of Extranodal Extension. *Thyroid* 2012, 22, 1144–1152. [CrossRef] [PubMed]
- Johnson, N.A.; Tublin, M.E. Postoperative Surveillance of Differentiated Thyroid Carcinoma: Rationale, Techniques, and Controversies. Radiology 2008, 249, 429–444. [CrossRef] [PubMed]
- Zheng, B.-W.; Wang, J.-F.; Ju, J.-X.; Wu, T.; Tong, G.; Ren, J. Efficacy and safety of cooled and uncooled microwave ablation for the treatment of benign thyroid nodules: A systematic review and meta-analysis. *Endocrine* 2018, 62, 307–317. [CrossRef] [PubMed]
- 15. Zhao, H.; Li, H. Meta-analysis of ultrasound for cervical lymph nodes in papillary thyroid cancer: Diagnosis of central and lateral compartment nodal metastases. *Eur. J. Radiol.* **2019**, *112*, 14–21. [CrossRef]
- 16. Guang, Y.; He, W.; Zhang, W.; Zhang, H.; Zhang, Y.; Wan, F. Clinical Study of Ultrasonographic Risk Factors for Central Lymph Node Metastasis of Papillary Thyroid Carcinoma. *Front. Endocrinol.* **2021**, *12*, 1635. [CrossRef]
- 17. Xiaoyin, T.; Ping, L.; Dan, C.; Min, D.; Jiachang, C.; Tao, W.; Yaoping, S.; Zhi, W.; Bo, Z. Risk Assessment and Hydrodissection Technique for Radiofrequency Ablation of Thyroid Benign Nodules. *J. Cancer* **2018**, *9*, 3058–3066. [CrossRef]
- Limkin, E.J.; Sun, R.; Dercle, L.; Zacharaki, E.I.; Robert, C.; Reuzé, S.; Schernberg, A.; Paragios, N.; Deutsch, E.; Ferté, C. Promises and challenges for the implementation of computational medical imaging (radiomics) in oncology. *Ann. Oncol.* 2017, 28, 1191–1206. [CrossRef]
- Lambin, P.; Leijenaar, R.T.H.; Deist, T.M.; Peerlings, J.; de Jong, E.E.C.; van Timmeren, J.; Sanduleanu, S.; Larue, R.T.H.M.; Even, A.J.G.; Jochems, A.; et al. Radiomics: The bridge between medical imaging and personalized medicine. *Nat. Rev. Clin. Oncol.* 2017, 14, 749–762. [CrossRef]
- Gillies, R.J.; Kinahan, P.E.; Hricak, H. Radiomics: Images Are More than Pictures, They Are Data. Radiology 2016, 278, 563–577. [CrossRef] [PubMed]

- Aerts, H.J.W.L.; Velazquez, E.R.; Leijenaar, R.T.H.; Parmar, C.; Grossmann, P.; Carvalho, S.; Bussink, J.; Monshouwer, R.; Haibe-Kains, B.; Rietveld, D.; et al. Decoding tumour phenotype by noninvasive imaging using a quantitative radiomics approach. *Nat. Commun.* 2014, *5*, 4006. [CrossRef] [PubMed]
- Parmar, C.; Grossmann, P.; Bussink, J.; Lambin, P.; Aerts, H.J.W.L. Machine Learning methods for Quantitative Radiomic Biomarkers. Sci. Rep. 2015, 5, 13087. [CrossRef] [PubMed]
- Pedregosa, F.; Varoquaux, G.; Gramfort, A.; Michel, V.; Thirion, B.; Grisel, O.; Blondel, M.; Prettenhofer, P.; Weiss, R.; Dubourg, V.; et al. Scikit-learn: Machine Learning in Python. J. Mach. Learn. Res. 2011, 12, 2825–2830.
- 24. Siegel, R.L.; Miller, K.D.; Jemal, A. Cancer statistics, 2019. CA Cancer J. Clin. 2019, 69, 7–34. [CrossRef] [PubMed]
- Tran, B.; Roshan, D.; Abraham, E.; Wang, L.; Garibotto, N.; Wykes, J.; Campbell, P.; Ebrahimi, A. An Analysis of The American Joint Commit-tee on Cancer 8 th Edition T Staging System for Papillary Thyroid Carcinoma. *J. Clin. Endocrinol. Metab.* 2018, 103, 2199–2206. [CrossRef]
- Lee, D.Y.; Kwon, T.-K.; Sung, M.-W.; Kim, K.H.; Hah, J.H. Prediction of Extrathyroidal Extension Using Ultrasonography and Computed Tomography. Int. J. Endocrinol. 2014, 2014, 1–6. [CrossRef]
- Choi, J.S.; Kim, J.; Kwak, J.Y.; Kim, M.J.; Chang, H.S.; Kim, E.-K. Preoperative Staging of Papillary Thyroid Carcinoma: Comparison of Ultrasound Imaging and CT. *Am. J. Roentgenol.* 2009, 193, 871–878. [CrossRef]
- Li, T.; Li, H.; Xue, J.; Miao, J.; Kang, C. Shear wave elastography combined with gray-scale ultrasound for predicting central lymph node metastasis of papillary thyroid carcinoma. *Surg. Oncol.* 2021, 36, 1–6. [CrossRef]
- Zhang, H.; Hu, S.; Wang, X.; He, J.; Liu, W.; Yu, C.; Sun, Z.; Ge, Y.; Duan, S. Prediction of Cervical Lymph Node Metastasis Using MRI Radiomics Approach in Papillary Thyroid Carcinoma: A Feasibility Study. *Technol. Cancer Res. Treat.* 2020, 19, 1533033820969451. [CrossRef]
- Jiang, M.; Li, C.; Tang, S.; Lv, W.; Yi, A.; Wang, B.; Yu, S.; Cui, X.; Dietrich, C.F. Nomogram Based on Shear-Wave Elastography Radiomics Can Improve Preoperative Cervical Lymph Node Staging for Papillary Thyroid Carcinoma. *Thyroid* 2020, 30, 885–897. [CrossRef]
- Guo, L.; Ma, Y.-Q.; Yao, Y.; Wu, M.; Deng, Z.-H.; Zhu, F.-W.; Luo, Y.-K.; Tang, J. Role of ultrasonographic features and quantified BRAFV600E mutation in lymph node metastasis in Chinese patients with papillary thyroid carcinoma. *Sci. Rep.* 2019, *9*, 75. [CrossRef] [PubMed]
- Park, V.; Han, K.; Kim, H.J.; Lee, E.; Youk, J.H.; Kim, E.-K.; Moon, H.J.; Yoon, J.H.; Kwak, J.Y. Radiomics signature for prediction of lateral lymph node metastasis in conventional papillary thyroid carcinoma. *PLoS ONE* 2020, *15*, e0227315. [CrossRef] [PubMed]
- Thibault, G.; Fertil, B.; Navarro, C.; Pereira, S.; Cau, P.; Levy, N.; Sequeira, J.; Mari, J.-L. Texture Indexes and Gray Level Size Zone Matrix. Application to Cell Nuclei Classification. *Pattern Recognit. Inf. Process.* 2009, 2009, 140–145.
- 34. Amadasun, M.; King, R. Textural features corresponding to textural properties. *IEEE Trans. Syst. Man Cybern.* 1989, 19, 1264–1274. [CrossRef]
- Li, Y.Y.; Sun, W.X.; Liao, X.D.; Zhang, M.B.; Xie, F.; Chen, D.H.; Zhang, Y.; Luo, Y.K. A Thyroid Ultrasound Image-based Artificial Intel-ligence Model for Diagnosis of Central Compartment Lymph Node Metastasis in Papillary Thyroid Carcinoma. *Zhongguo* Yi Xue Ke Xue Yuan Xue Bao 2021, 43, 911–916. [CrossRef]
- 36. Tian, X.; Luo, Y.; Zhang, Y.; Song, Q.; Ren, L.; Zhang, Y.; Jiang, B. Application of Ultrasound Model for Predicting Central Compartment Lymph Node Metastasis in Papillary Thyroid Carcinoma. *Zhongguo Yi Xue Ke Xue Yuan Xue Bao* 2020, 42, 626–631.
- Zou, Y.; Shi, Y.; Liu, J.; Cui, G.; Yang, Z.; Liu, M.; Sun, F. A Comparative Analysis of Six Machine Learning Models Based on Ultrasound to Distinguish the Possibility of Central Cervical Lymph Node Metastasis in Patients With Papillary Thyroid Carcinoma. *Front. Oncol.* 2021, *11*, 1938. [CrossRef]