

## PREDICTION OF CERVICAL LYMPH NODE METASTASIS IN PAPILLARY THYROID CARCINOMA USING A MACHINE LEARNING APPROACH

## PREDIKCIJA METASTAZA U LIMFNIM ČVOROVIMA VRATA KOD PAPILARNOG KARCINOMA ŠTITASTE ŽLEZDE KORIŠĆENJEM PRINCIPA MAŠINSKOG UČENJA

Marina Popović Krneta<sup>1,2</sup>, Dragana Šobić Šaranović<sup>1,3</sup>, Ljiljana Mijatović Teodorović<sup>2,4</sup>

<sup>1</sup> Univerzitet u Beogradu, Medicinski fakultet, Beograd, Srbija

<sup>2</sup> Institut za onkologiju i radiologiju Srbije, Beograd, Srbija

<sup>3</sup> Univerzitetski klinički centar Srbije, Centar za nuklearnu medicinu sa pozitronskom emisionom tomografijom, Beograd, Srbija

<sup>4</sup> Univerzitet u Kragujevcu, Fakultet medicinskih nauka, Kragujevac, Srbija

**Correspondence:** marina.popovic1989@gmail.com

### Abstract

The incidence of papillary thyroid carcinoma (PTC) has been constantly increasing over the past three decades, establishing it as the most frequently diagnosed type of thyroid malignancy. While patients with PTC generally have a favorable outcome, the presence of lymph node metastases (LNM) may significantly impact their prognosis, leading to a higher likelihood of recurrence. The current pre-operative diagnosis of LNM primarily relies on cervical ultrasound examination, which is limited in sensitivity. As a result of low sensitivity, lymph node metastases remain undetected on the pre-operative staging and may later present as persistent or recurrent disease, necessitating further evaluation and potential reoperation.

To address the challenges of LNM diagnoses, various models have been developed to predict LNM in PTC patients. Among prediction models, special attention has been drawn to machine learning models that can predict disease outcomes with improved accuracy and enable individualized selection of optimal treatment for each patient. Therefore, this mini-review primarily focuses on explaining the fundamental principles of ML models through an example of LNM prediction in PTC patients. Additionally, an overview is provided on the most commonly used ML models in medicine, discussing their performance in studies employing such approaches for LNM prediction. Finally, the main challenges that limit the implementation of these models in clinical practice have been examined, and crucial areas for improvement have been identified.

Currently, ML models present a potentially useful tool for LNM prediction in PTC patients, but further research is necessary to fully leverage their capabilities and enable their implementation into decision support systems.

#### Keywords:

papillary thyroid carcinoma, machine learning, lymph node metastasis

## Sažetak

Incidencija papilarnog tireoidnog karcinoma (PTK) u poslednje tri decenije se konstantno uvećava, čineći ga najčešćim malignim oboljenjem štitaste žlezde. Iako pacijenti sa PTK generalno imaju povoljan ishod, prisustvo metastaza u limfnim čvorovima vrata može značajno uticati na njihovu prognozu, povećavajući verovatnoću ponovnog javljanja bolesti. Trenutna preoperativna dijagnostika limfonodalnih metastaza uglavnom podrazumeva ultrazvučni pregled vrata, ali osnovno ograničenje ove metode je niska senzitivnost. Kao rezultat niske senzitivnosti, limfonodalne metastaze ostaju neotkrivene tokom preoperativnog utvrđivanja stadijuma bolesti i mogu se kasnije manifestovati kao perzistentna ili rekurentna bolest, zahtevajući dalju evaluaciju i potencijalnu reoperaciju.

Kako bi se suočili sa izazovima dijagnostike limfonodalnih metastaza kod pacijenata sa PTK, razvijeni su različiti modeli za njihovu predikciju. Među modelima predikcije posebnu pažnju privlače modeli mašinskog učenja, koji mogu sa većom tačnošću predvideti ishod bolesti i omogućiti odabir optimalnog lečenja za svakog pacijenta individualno. Stoga je ovaj mini pregledni članak pretežno usmeren na objašnjenje osnovnih principa modela mašinskog učenja kroz primer predikcije metastaza u limfnim nodusima vrata kod pacijenata sa PTK. Pored toga, u radu je dat pregled najčešće korišćenih modela mašinskog učenja uz diskusiju njihove efikasnosti u studijama koje su predikciju limfonodalnih metastaza bazirale na ovom pristupu. Na samom kraju, razmatrani su i najčešći izazovi koji limitiraju implementaciju ovih modela u kliničkoj praksi i identifikovana su ključna područja za njihovo poboljšanje.

Trenutno, modeli mašinskog učenja predstavljaju potencijalno korisno sredstvo za predikciju limfonodalnih metastaza kod pacijenata sa PTK, ali su neophodna dalja istraživanja kako bi se potpuno iskoristile mogućnosti ovih modela i omogućila njihova implementacija u sisteme za podršku odlučivanju.

### Ključne reči:

papilarni tireoidni karcinom, mašinsko učenje, limfonodalne metastaze

## Introduction

With its incidence increasing rapidly over the last three decades, papillary thyroid carcinoma (PTC) has become the most commonly diagnosed form of thyroid malignancy (1). The presence of cervical lymph node metastases (LNM) is frequently observed, reaching up to 90% of patients diagnosed with PTC (2, 3). Moreover, at the time of diagnosis, central LNM is already present in 30 - 70% of patients, whereas the risk of lateral compartment node metastasis ranges from 20-50% (4, 5). Lymph node involvement might correlate to an increased local recurrence and diminished survival among PTC patients (6, 7). Consequently, accurate disease staging and optimal treatment decisions rely on determining the metastatic status of the cervical lymph nodes (8).

The preoperative assessment of cervical LNM in PTC patients commonly relies on ultrasound (US) as the recommended primary imaging technique. When a central or lateral LNM is diagnosed through the US or physical examination, therapeutic lymph node dissection (LND) is recommended (1, 9). However, despite its widespread use, meta-analyses have shown the sensitivity of the US is low, especially for central LNM (less than 33%) (10, 11). Further, the surgeon's ability to detect metastatic nodes intraoperatively is unreliable, with less than 30% accuracy in some cases (12, 13). As a result, nearly 40% of patients initially diagnosed as clinically node-negative (cN0) still have occult LNM (14). These findings underscore the limitations of current preoperative and intraoperative diagnostic techniques in accurately identifying the presence

of LNM. Undetected LNM may lead to persistent or recurrent disease, resulting in a second operation which poses greater challenges and carries a higher risk of complication (8).

Previous studies aiming to predict LNM risk generally applied statistical regression models due to their simplicity. However, such models presume a linear relationship drawn between patient characteristics and disease outcome, which may not accurately reflect the complexities present in the clinical information (15). To address these limitations, a steady introduction of machine learning (ML) technology has been undertaken in medical research. As a subfield of artificial intelligence, ML employs algorithms to analyze large datasets, identifying complex relationships between multiple predictors and outcomes. Such an approach goes beyond traditional statistical methods, providing medical professionals with new means to detect disease risk, diagnose and deliver treatment (16, 17). Building from here, this mini review aims to provide an overview of the main principles and the practical implementation of commonly used ML models in predicting LNM in PTC patients. To the best of our knowledge, this is the first review to discuss the application of ML in the prediction of LNM in PTC patients.

## Fundamentals of machine learning

Machine learning algorithms utilize previously collected data to establish patterns and relationships between them, thereby making predictions applicable to data not included in the original dataset (16). They are

divided into three main categories, depending on the type of research task to be solved: reinforcement, unsupervised and supervised learning. When applying ML in medicine, supervised learning is the predominant approach utilized for predictive modeling purposes. Supervised ML models are designed to identify complex patterns by analyzing known examples, allowing them to make informed predictions when presented with new instances (18).

The development of an ML model is initiated by first formulating a research question and defining the target outcome. In the context of LNM prediction studies, the outcome is typically a binary variable indicating whether LNM is present or not. Such a task may be addressed through supervised ML classification (15). Second step involves the data pre-processing phase. Through the pre-processing process, patient information (including demographic details, clinical factors, histopathological features, and, in certain cases, imaging data) is prepared for ML modeling. Diverse techniques of filling in missing values and cleaning up the data are applied to ensure all necessary information is complete and accurate. Therefore, pre-processing serves to optimize the data for subsequent ML analysis and ensure its quality for further modeling (19).

During the pre-processing phase, the data is usually divided into two distinct sets: the training and the test set. Typically, the ML model is trained on two-thirds of the data, and its performance is then evaluated on the remaining one-third of the sample (20). The purpose of training in ML models is to both analyze the data and uncover meaningful relationships that allow the model to effectively map the given inputs to their corresponding outputs (21). For instance, when predicting LNM in PTC patients, the ML model may learn that certain combinations of clinical factors and histopathological features are indicative of a higher likelihood of LNM. Such discovery allows the model to establish links between patient characteristics and the presence or absence of LNM, thereby facilitating accurate predictions. Throughout the training phase, multiple, candidate models are often tested on the data to assess their predictive performance (19). To compare these candidate models, researchers utilize multiple metrics, such as

sensitivity, specificity, accuracy, and the area under the receiver operating characteristic (ROC) curve (AUC) value. The evaluation of the candidate model parameters establishes the strengths and weaknesses of the models' predictive capabilities. The most promising model is then selected for validation, the next step of the ML process.

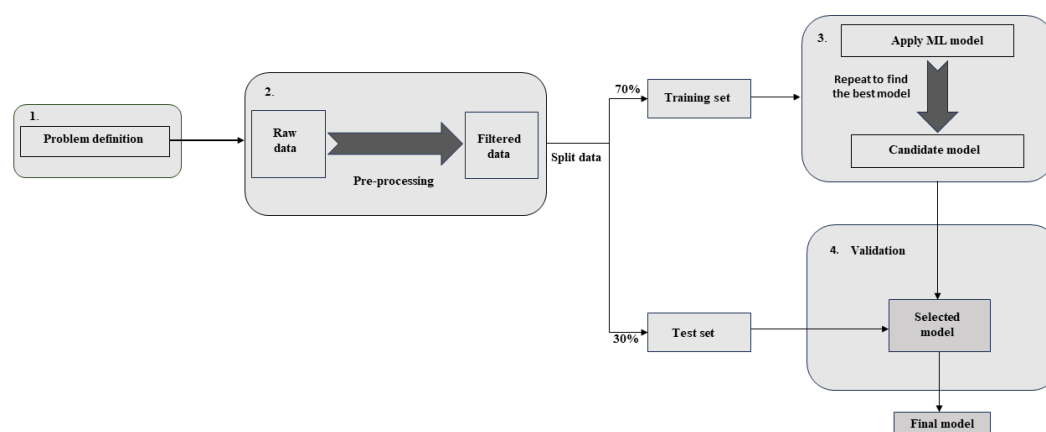
Validation is typically performed on a separate set, known as a test set, comprising new and unseen data that the model has not encountered while undergoing training (18). In our example of LNM prediction, the model selected through the training phase will be provided with input information on PTC patients without knowing their actual LNM diagnosis. The model then utilized learned patterns to predict whether the patient has LNM or not. The validation process compares the model's predictions on the test set to the true outcome, which is the histopathological confirmation of presence or absence of LNM in our study. This comparison allows the evaluation of the final model's performance using multiple classification metrics. Therefore, the final model obtained through the ML development process demonstrates its ability to generalize and make accurate predictions on new and previously unobserved patient data. The general process of building ML models is illustrated in **figure 1**.

## Machine learning algorithms commonly used in medical studies

### Logistic regression (LR)

Widely used and accepted, LR is able to predict an event's probability through fitting data to a logistic function (22). In basic terms, LR is constructed by using patient input characteristics whereafter, a logistic function is applied to differentiate between the two output classes (15). Owing to their interpretability, LR has long been prominent in modeling as it allows practitioners to easily identify the contribution that an independent variable has on the predicted outcome (17).

### K-Nearest Neighbor (KNN)



**Figure 1.** General machine learning building process: 1: Defining the study output; 2: Structuring and pre-processing the data required for model training; 3: Training multiple machine learning (ML) models to find the best-performing one; 4: Validating the final model's performance on new, unseen data (i.e., test data).

It is the most well-known instance-based learning algorithm. These models do not utilize training data to learn and form generalized rules, rather they draw direct comparisons between the training data and the novel data to be analyzed (18). By measuring distances to its neighbors, KNN identifies the closest ones and examines their known labels to make predictions (23). To illustrate, when using the KNN model to classify a new PTC patient, if the new patient has a greater number of neighbors in the training set labeled as LNM positive, the model would classify the patient as having LNM.

### Support Vector Machines (SVM)

As another popular ML model, SVM finds widespread application across diverse domains, including medicine. Its classifiers separate two classes with a linear boundary termed the hyperplane (24). Finding an optimal hyperplane in SVM is essential to minimize prediction errors on new data. This process begins by identifying the support vectors, which are the data points from both classes that are closest to the hyperplane. The optimal hyperplane may be determined by maximizing the distance between the support vectors and the hyperplane, known as the margin (18). Once the hyperplane is established, new data points may be assigned to one of the classes based on their proximity to the hyperplane.

### Decision tree (DT) models

A decision tree is a hierarchical model in which the dataset is progressively divided into subgroups through a sequence of repeated attribute-based splits. The algorithm then evaluates diverse attributes to select the best ones creating internal nodes at a decision point. This leads to subsequent splits and the formation of branches. The splitting process is repeated recursively as the algorithm further evaluates attributes to create additional internal nodes and refine the tree structure. The process continues until the terminal or the leaf node is reached. Then, the corresponding class is assigned to the sample (18,21).

### Classifier Ensembles

Ensemble learning is a method in which diverse classifiers are combined to build a more reliable predictive model (24). Bagging and boosting are two methods most frequently used to combine numerous classifiers. By using bootstrapping, multiple subgroups of the training data are created in the bagging approach. Each group is then used to train a different classifier, after which the predictions are combined. Boosting, on the other hand, focuses on training a number of weak classifiers repeatedly, where each following classifier is trained to correct the errors committed by those prior (25). The most prevalent ensemble classifiers are gradient boosting (GB), extreme gradient boosting (XGBoost), adaptive boosting (AdaBoost) and random forest (RF).

### Artificial neural networks (ANN)

These networks are computational systems able to process information and learn from prior experience, allowing them to acquire knowledge and generalize patterns to handle new situations (26). Consisting of nodes, known as neurons, an ANN weights specific inputs and generates an output value. The nodes are organized into multiple layers, including an input layer receiving the independent variables and an output layer where each node represents a possible outcome. The network is trained to generate an output by combining the input variables through multiple hidden layers. In cases where the ANNs have a numerous intermediate layer, the term deep learning is used. Deep learning has demonstrated impressive results in solving highly complex modeling problems (15,27).

---

## Practical Applications of machine learning models in lymph node prediction studies

---

There are a number of ML models recently constructed to predict LNM in PTC patients which aim to both improve classification accuracy and plan appropriate treatment management. **Table 1** details the studies utilizing ML models for LNM prediction.

In all included studies, the patients had undergone surgical treatment, in which the extent of the surgery was dependent on the specific outcome being predicted. For central LNM prediction, the surgical procedure typically involved a lobectomy with an accompanying isthmectomy and ipsilateral central LND. For bilateral PTC, a total thyroidectomy with a bilateral central LND was most commonly performed. Studies focusing on lateral LNM prediction, expanded the surgical approach to include the removal of the lateral lymph nodes. Of note is that the majority of guidelines do not generally recommend the prophylactic lateral LND (1,9). Consequently, the patients enrolled in such studies exhibited clinical or imaging findings which raised suspicion of lateral LNM. However, in one particular study involving cN0 patients, the decision to perform lateral LND resulted from the sentinel lymph node (SLN) biopsy of the lateral neck compartments. The SLNs removed during the biopsy were then subjected to a frozen section analysis (FSA). If the FSA confirmed metastatic SLNs, concurrent LND was performed (8).

The input of the models' training parameters primarily included patient preoperative characteristics such as demographic, clinical and US features as well as the histopathological data obtained during or subsequent to the thyroid surgery. Molecular biomarkers (i.e., BRAF mutation) were included in two studies (30,33). For all training models, tumor size and age were generally found to be most significant; notably, younger age and larger tumor size both had a positive association for LNM.

Comparing the performances of different ML models across the studies analyzed, no consistent pattern indicating a superior classifier or predicting an LNM outcome was observed. Nonetheless, ensemble-learning-based ML classifiers did generally exhibit better performance, with

the RF algorithm emerging as the best-performing model. The reported AUC values ranged from 0.67 to 0.85 in the studies reviewed. Additionally, some studies compared ML models with US and other methods, such as LR or nomograms. For instance, Wu Y et al. demonstrated that the ML classifier is able to yield better predictive performance in contrast to US (28). Similarly, ML models provided consistent superior performance in predicting LNM when compared to nomograms or logistic regression models (30,32,35).

The value and performance of AUC based on the confusion matrix (sensitivity, specificity and accuracy) were commonly used to evaluate ML models in order to predict LNM. However, the choice of metrics is determined by the classifier’s purpose. In LNM prediction studies, the primary objective should be to minimize false negative (FN) results to ensure that LNM patients are not missed in the predictive process. Nevertheless, metrics such as accuracy, though easy to interpret, do not provide data on the

prevalence of FNs or FPs. Likewise, neither sensitivity nor specificity alone summarizes completely the information conveyed by the confusion matrix. To address this limitation, two studies also incorporated Fβ metrics, such as the F1 score and F2 score, as part of their performance evaluation (8, 35). Parameter β is used by the F score to control the sensitivity balance and positive predictive value (PPV). The F1 score, used when β = 1, provides an equal balance between recall and precision. Increasing the β value, such as the F2 score, provides a higher weight to sensitivity. Consequently, when analyzing results it is essential to take into account the aggregate metric, such as the F β score, as they help to provide a better-confusion matrix overview (36). Since it is easy to calculate F-scores from the information provided by the confusion matrix, there is a notable trend of growing adoption of the F1 score in different oncology studies (37–39). This trend underscores its significance as a valuable metric for assessing model performance in real clinical scenarios where the detection of positive

**Table 1.** Machine learning models in prediction of lymph node metastasis in papillary thyroid carcinoma patients.

Study	Outcome	Sample size	LNM + (%)	LNM - (%)	Age/sex (male %)	Tumor size (cm)	ML models	Best ML model	AUC (%)	Sn (%)	Sp (%)	Accuracy (%)
Wu Y et al (28)	Central LNM	1103	51.3	48.7	41.0 (33-51)/ 26.9	1.0 (0.7-1.5)	RF, DT, GBM, XGB, AdaB, ANN	GBDT	73	64	72	-
Zhu J et al (29)	Central LNM	1271	55.5	44.5	41.38 ±11/ 26.7	0.99 ±0.57	LR, DT, RF, ANN, GBM, XGB	XGB	75	67	67	67
Feng JW et al (30)	Lateral LNM	1236	19.4	80.6	- / -	-	LR, RF, DT, SVM, BN, GBM, XGB, ANN	RF	85	89	78	-
Lai SW et al (31)	Lateral LNM	1815	62.5	37.5	42.0 (33-51) / 36.9	1.54 ±0.7	DT, RF, SVM, KNN, XGB, ANN	RF	80	89	49	74
Huang Y et al (32)	Central and lateral LNM	33285	89.5	10.5	- / 17.2	-	BN, RF XGB, AdaB, MLP	XGB	80	74	66	67
Yu Y et al (33)	Central LNM	1121	33.6	66.4	45.0 (37-51) / 21.7	0.70 (0.5-0.8)	LR, RF, DT, MARS, XGB, ANN	RF	79	56	87	77
Liu W et al (34)	Central LNM	1046	21.5	78.5	43.9 ±13 / 24.5	1.02 ±0.7	XGB, KNN, RF, GBM, AdaB, HEM	HEM	67	44	74	-
Popovic KM et al (8)	Central and lateral LNM	288	49.0	51.0	47.3 ±14 / 25.0	1 (0.1-4)	LR, KNN, SVM, DT	KNN	72	98	27	62

Values for age and tumor size are presented as median (minimum–maximum) or mean ± standard deviation. The abbreviations: - - negative; + - positive; LNM - lymph node metastasis; RF - random forest; DT - decision tree; GBM - gradient boosting machine; XGB - extreme gradient boosting; AdaB - adaptive boosting; ANN - artificial neural network; LR - logistic regression; SVM - support vector machine; BN - Bayesian network; MARS - multivariate adaptive regression splines; HEM - heterogeneous ensemble algorithm model; AUC - area under the receiver operating characteristic (ROC) curve; Sn - sensitivity; Sp - specificity.

class (such as detection of LNM) is of huge importance. A higher F-score indicates that the model performs better on the class of interest, highlighting its practical utility in clinical applications (40).

When analyzing medical studies, it is important to consider the common problem of imbalanced data, where one class or outcome differs significantly from others. In our review, three of the analyzed studies exhibited data imbalance (30, 32, 34). While these studies did not explicitly specify data imbalance techniques, they addressed this challenge by utilizing AUC, a valuable method for dealing with class imbalance. It addresses class distribution disparities by determining an optimum threshold for ML models and can play a role in improving diagnostic accuracy, particularly in imbalanced data sets.

In order to enhance the accessibility and usability of predictive models in a clinical setting, four studies developed user-friendly web calculators alongside the ML models. These calculators facilitate the models' practical application, allowing clinicians to make informed decisions regarding patient management and treatment planning (8, 29, 31, 33).

## Challenges, limitations and future perspectives of machine learning applications

While ML approaches have shown promise in predicting LNM in PTC patients, several challenges persist in their implementation. One key issue is that the majority of ML research is based on retrospective data, which might result in selection bias and complicate the analysis of important risk variables for LNM development (8). Within retrospective studies, both data availability and quality remain significant challenges with limited dataset size, missing values or errors potentially impacting the overall ML ability to predict outcomes (41). Future research should prioritize large, diverse and preprocessed datasets from prospective, multicenter studies to improve generalization and clinical usefulness of ML models.

Furthermore, all studies included in this review have established their ML model development on clinical, ultrasound and histopathological data. However, the limited data types have led to lower model performance of the examined ML models. To address this issue, the concept of radiomics has been introduced. Radiomics use digital images to extract multiple quantitative features and merged them with additional patient information to be integrated into ML models. These act as decision support tools which may improve diagnostic, prognostic and predictive accuracy (42). While the potential of radiomics is significant, the main limitation lies in its lack of standardization which must be addressed (43).

Machine learning models, often compared with traditional statistical methods, have shown that AI-based models are capable of producing significantly better predictive accuracy. However, unlike statistics, ML models are often regarded as non-interpretable, and for this reason, they are

frequently referred to as the "black boxes" (19). This lack of interpretability is a key factor contributing to physicians' hesitancy in adopting ML models for clinical decision-making. In response, various methods have recently been proposed to help users interpret the predictions of complex models, one of which is the SHapley Additive exPlanations (SHAP) analysis. It quantifies feature's contribution to the model's prediction for a specific patient, making the decision process of complex ML models more understandable (44, 45). These efforts aim to achieve better comprehension of ML models and, possibly, gain more confidence from physicians.

Finally, it is crucial to consider the ethical implications of the medical use of ML as healthcare technology advances. There are still numerous unresolved issues, including the extent to which patient privacy is affected by the large usage of data in ML research. Furthermore, it is important to reflect on who would be held responsible if ML decision-making results in patient injury (41). Clear guidelines are important to address these ethical issues and encourage the proper use of ML models in clinical practice.

## Conclusion

Pre-operative lymph node diagnostics in patients with PTC has its limitations. In the past year, ML techniques have become more widely used in order to address these challenges and improve the LNM prediction. Machine learning models have shown certain advantages, however, further work in explaining and optimizing data using standard multicenter databases is needed to ensure the successful use of these models in clinical practice.

## Literature

1. Filetti S, Durante C, Hartl D, Leboulleux S, Locati LD, Newbold K, et al. Thyroid cancer: ESMO Clinical Practice Guidelines for diagnosis, treatment and follow-up. *Annals of Oncology*. 2019; 30(12):1856–83.
2. Liu C, Xiao C, Chen J, Li X, Feng Z, Gao Q, et al. Risk factor analysis for predicting cervical lymph node metastasis in papillary thyroid carcinoma: a study of 966 patients. *BMC Cancer*. 2019; 19(1):622.
3. Genpeng L, Jianyong L, Jiaying Y, Ke J, Zhihui L, Rixiang G, et al. Independent predictors and lymph node metastasis characteristics of multifocal papillary thyroid cancer. *Medicine (Baltimore)*. 2018; 97(5):e9619.
4. Liu LS, Liang J, Li JH, Liu X, Jiang L, Long JX, et al. The incidence and risk factors for central lymph node metastasis in cN0 papillary thyroid microcarcinoma: a meta-analysis. *Eur Arch Otorhinolaryngol*. 2017; 274(3):1327–38.
5. Back K, Choe JH, Kim JS, Kim JH. Occult contralateral central neck metastasis in papillary thyroid carcinoma with preoperatively documented ipsilateral lateral neck metastasis. *European Journal of Surgical Oncology*. 2021; 47(6):1339–45.
6. Sapuppo G, Palermo F, Russo M, Tavarelli M, Masucci R, Squatrito S, et al. Latero-cervical lymph node metastases (N1b) represent an additional risk factor for papillary thyroid cancer outcome. *J Endocrinol Invest*. 2017; 40(12):1355–63.
7. Jiang L hao, Yin K xin, Wen Q liang, Chen C, Ge M hua, Tan Z. Predictive Risk-scoring Model For Central Lymph Node Metastasis and Predictors of Recurrence in Papillary Thyroid Carcinoma. *Sci Rep*. 2020; 10(1):710.

8. Popović Krneta M, Šobić Šaranović D, Mijatović Teodorović L, Krajčinović N, Avramović N, Bojović Ž, et al. Prediction of Cervical Lymph Node Metastasis in Clinically Node-Negative T1 and T2 Papillary Thyroid Carcinoma Using Supervised Machine Learning Approach. *JCM*. 2023; 12(11):3641.
9. Haugen BR, Alexander EK, Bible KC, Doherty GM, Mandel SJ, Nikiforov YE, et al. 2015 American Thyroid Association Management Guidelines for Adult Patients with Thyroid Nodules and Differentiated Thyroid Cancer: The American Thyroid Association Guidelines Task Force on Thyroid Nodules and Differentiated Thyroid Cancer. *Thyroid*. 2016; 26(1):1–133.
10. 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.
11. Alabousi M, Alabousi A, Adham S, Pozdnyakov A, Ramadan S, Chaudhari H, et al. Diagnostic Test Accuracy of Ultrasonography vs Computed Tomography for Papillary Thyroid Cancer Cervical Lymph Node Metastasis: A Systematic Review and Meta-analysis. *JAMA Otolaryngol Head Neck Surg*. 2022; 148(2):107.
12. Ji YB, Lee DW, Song CM, Kim KR, Park CW, Tae K. Accuracy of Intraoperative Determination of Central Node Metastasis by the Surgeon in Papillary Thyroid Carcinoma. *Otolaryngol Head Neck Surg*. 2014; 150(4):542–7.
13. Gonçalves Filho J, Zafereo ME, Ahmad FI, Nixon IJ, Shaha AR, Vander Poorten V, et al. Decision making for the central compartment in differentiated thyroid cancer. *Eur J Surg Oncol*. 2018; 44(11):1671–8.
14. Delgado-Oliver E, Vidal-Sicart S, Martínez D, Squarcia M, Mora M, Hanzu FA, et al. Applicability of sentinel lymph node biopsy in papillary thyroid cancer. *Q J Nucl Med Mol Imaging*. 2020; 64(4):400–5.
15. Richter AN, Khoshgoftaar TM. A review of statistical and machine learning methods for modeling cancer risk using structured clinical data. *Artif Intell Med*. 2018; 90:1–14.
16. Cuocolo R, Caruso M, Perillo T, Ugga L, Petretta M. Machine Learning in oncology: A clinical appraisal. *Cancer Lett*. 2020; 481:55–62.
17. Bertsimas D, Wiberg H. Machine Learning in Oncology: Methods, Applications, and Challenges. *JCO Clin Cancer Inform*. 2020; 4:885–94.
18. Lo Vercio L, Amador K, Bannister JJ, Crites S, Gutierrez A, MacDonald ME, et al. Supervised machine learning tools: a tutorial for clinicians. *J Neural Eng*. 2020; 17(6):062001.
19. Bur AM, Shew M, New J. Artificial Intelligence for the Otolaryngologist: A State of the Art Review. *Otolaryngol Head Neck Surg*. 2019; 160(4):603–11.
20. Bellazzi R, Zupan B. Predictive data mining in clinical medicine: Current issues and guidelines. *Int J Med Inform*. 2008; 77(2):81–97.
21. Alpaydin E. Introduction to machine learning, 2nd edition. MIT Press Cambridge; 2010.
22. Chao CM, Yu YW, Cheng BW, Kuo YL. Construction the Model on the Breast Cancer Survival Analysis Use Support Vector Machine, Logistic Regression and Decision Tree. *J Med Syst*. 2014; 38(10):106.
23. Badillo S, Banfai B, Birzele F, Davydov II, Hutchinson L, Kam - Thong T, et al. An Introduction to Machine Learning. *Clin Pharmacol Ther*. 2020; 107(4):871–85.
24. Sidey-Gibbons JAM, Sidey-Gibbons CJ. Machine learning in medicine: a practical introduction. *BMC Med Res Methodol*. 2019; 19(1):64.
25. Huang MW, Chen CW, Lin WC, Ke SW, Tsai CF. SVM and SVM Ensembles in Breast Cancer Prediction. *PLoS One*. 2017; 12(1):e0161501.
26. Marinkovic M, Popovic M, Stojanovic-Rundic S, Nikolic M, Cavic M, Gavrilovic D, et al. Comparison of Different Machine Learning Models in Prediction of Postirradiation Recurrence in Prostate Carcinoma Patients. *Biomed Res Int*. 2022; 2022:1–13.
27. Najafabadi MM, Villanustre F, Khoshgoftaar TM, Seliya N, Wald R, Muharemagic E. Deep learning applications and challenges in big data analytics. *J Big Data*. 2015; 2(1):1.
28. Wu Y, Rao K, Liu J, Han C, Gong L, Chong Y, et al. Machine Learning Algorithms for the Prediction of Central Lymph Node Metastasis in Patients With Papillary Thyroid Cancer. *Front Endocrinol*. 2020; 11:577537.
29. Zhu J, Zheng J, Li L, Huang R, Ren H, Wang D, et al. Application of Machine Learning Algorithms to Predict Central Lymph Node Metastasis in T1-T2, Non-invasive, and Clinically Node Negative Papillary Thyroid Carcinoma. *Front Med*. 2021; 8:635771.
30. Feng JW, Ye J, Qi GF, Hong LZ, Wang F, Liu SY, et al. A comparative analysis of eight machine learning models for the prediction of lateral lymph node metastasis in patients with papillary thyroid carcinoma. *Front Endocrinol*. 2022; 13:1004913.
31. Lai S wei, Fan Y long, Zhu Y hua, Zhang F, Guo Z, Wang B, et al. Machine learning-based dynamic prediction of lateral lymph node metastasis in patients with papillary thyroid cancer. *Front Endocrinol*. 2022; 13:1019037.
32. Huang Y, Mao Y, Xu L, Wen J, Chen G. Exploring risk factors for cervical lymph node metastasis in papillary thyroid microcarcinoma: construction of a novel population-based predictive model. *BMC Endocr Disord*. 2022; 22(1):269.
33. Yu Y, Yu Z, Li M, Wang Y, Yan C, Fan J, et al. Model development to predict central lymph node metastasis in cN0 papillary thyroid microcarcinoma by machine learning. *Ann Transl Med*. 2022; 10(16):892.
34. Liu W, Wang S, Xia X, Guo M. A Proposed Heterogeneous Ensemble Algorithm Model for Predicting Central Lymph Node Metastasis in Papillary Thyroid Cancer. *Int J Gen Med*. 2022; 15:4717–32.
35. Lai S wei, Fan Y long, Zhu Y hua, Zhang F, Guo Z, Wang B, et al. Machine learning-based dynamic prediction of lateral lymph node metastasis in patients with papillary thyroid cancer. *Front Endocrinol*. 2022; 13:1019037.
36. Lever J. Classification evaluation: It is important to understand both what a classification metric expresses and what it hides. *Nat Methods*. 2016; 13(8):603–5.
37. Alabi RO, Elmusrati M, Sawazaki Calone I, Kowalski LP, Haglund C, Coletta RD, et al. Comparison of supervised machine learning classification techniques in prediction of locoregional recurrences in early oral tongue cancer. *Int J Med Inform*. 2020; 136:104068.
38. Alkhadar H, Macluskey M, White S, Ellis I, Gardner A. Comparison of machine learning algorithms for the prediction of five-year survival in oral squamous cell carcinoma. *J Oral Pathol Med*. 2021; 50(4):378–84.
39. Alabi RO, Mäkitie AA, Pirinen M, Elmusrati M, Leivo I, Almangush A. Comparison of nomogram with machine learning techniques for prediction of overall survival in patients with tongue cancer. *Int J Med Inform*. 2021; 145:104313.
40. Bekkar M, Djemaa HK, Alitouche TA. Evaluation measures for models assessment over imbalanced data sets. *J Inf Eng Appl*. 2013; 3(10):27–38.
41. Rahmani AM, Yousefpoor E, Yousefpoor MS, Mehmood Z, Haider A, Hosseinzadeh M, et al. Machine Learning (ML) in Medicine: Review, Applications, and Challenges. *Mathematics*. 2021; 9(22):2970.
42. Gillies RJ, Kinahan PE, Hricak H. Radiomics: Images Are More than Pictures, They Are Data. *Radiology*. 2016; 278(2):563–77.
43. Lambin P, Leijenaar RTH, Deist TM, Peerlings J, De Jong EEC, Van Timmeren J, et al. Radiomics: the bridge between medical imaging and personalized medicine. *Nat Rev Clin Oncol*. 2017; 14(12):749–62.
44. El Jammal T, Guerber A, Prodel M, Fauter M, Sève P, Jamilloux Y. Diagnosing Hemophagocytic Lymphohistiocytosis with Machine Learning: A Proof of Concept. *J Clin Med*. 2022; 11(20):6219.
45. Lundberg SM, Lee S I. (2017). A Unified Approach to Interpreting Model Predictions. In: *Advances in Neural Information Processing Systems*. p.4765-4774.
46. Bur AM, Holcomb A, Goodwin S, Woodroof J, Karadaghy O, Shnyder Y, et al. Machine learning to predict occult nodal metastasis in early oral squamous cell carcinoma. *Oral Oncol*. 2019; 92:20–5.