Journal of Carcinogenesis
Covering Genesis to Therapeutics
Journal of Carcinogenesis, ISSN: 1477-3163
2026; Vol 25: Issue 1, https://carcinogenesis.com/

Hybrid Machine Learning and Deep Learning Framework for Predicting Lymph Node Metastasis in Papillary Thyroid Carcinoma with Hashimoto's Thyroiditis: A Retrospective Cohort Study

Syed Haseeb Zia[1]*, Zhang Wenwen[2], Alveena Nasim Khan[3], Liu Jiazheng[4], Li Ying[5], Zhao Liu[6]

[1]Master Candidate in General Surgery, Department of General Surgery, The Affiliated Hospital of Xuzhou Medical University, Xuzhou, China.

Email ID: dr.haseeb3135@gmail.com

[2]Master Candidate in Thyroid and Breast Surgery, Department of Thyroid and Breast Surgery, The Affiliated Hospital of Xuzhou Medical University, Jiawang District, China.

Email ID: zww11282023@163.com

[3]Master Candidate in Anesthesiology, Department of Anesthesiology, The Affiliated Hospital of Xuzhou Medical University, Xuzhou, China.

Email ID: drkhan2107@gmail.com

[4]Master Candidate in General Surgery, Department of General Surgery, The Affiliated Hospital of Xuzhou Medical University, Xuzhou, China.

Email ID: xyljz6666@163.com

[5]Master of Medicine in Thyroid and Breast Surgery, Department of Thyroid and Breast Surgery, The Affiliated Hospital of Xuzhou Medical University, Xuzhou, China.

Email ID: effie1104@163.com

[6]Ph.D in General Surgery, Department of General Surgery, The Affiliated Hospital of Xuzhou Medical University, Xuzhou, China.

Email ID: xylzhao9999@163.com

ABSTRACT

Background: Papillary thyroid carcinoma (PTC) is commonly present in the co-occurrence with Hashimoto thyroiditis (HT); the inflammatory microenvironment, however, presents obstacles to preoperative predictability of lymph node metastasis (LNM). Accurate risk stratification also clinically cannot be done without to inform surgical decision-making and reduce unnecessary central neck dissections. In line with this, this study assesses the viability of machine-learning (ML), deep-learning (DL), and the hybrid modelling approaches to forecast LNM in HT-related PTC by using real-world clinical data.

Methods: A retrospective cohort study involving 197 patients who had been diagnosed with HT-PTC was examined. Following data cleaning and feature engineering, 68 patients with complete LNM labels were used in supervised learning. Models considered included logistic regression, random forest, gradient boosting, XGBoost, a fully connected artificial neural network (ANN) and an XGBoost-ANN hybrid architecture. The model performance was measured based on accuracy, precision, recall, F1-score and area under receiver operating characteristic curve (AUC). The explainability of the model was estimated using SHAP values.

Results: Random Forest produced the best area under the receiver operating characteristic curve (AUC = 0.617), traditional and hybrid models showed rather modest discriminative outcomes (AUC range 0.36-0.53). The artificial neural network provided showed significant overfitting, which can probably be explained by the small sample size. The SHAP (SHapley Additive exPlanations) analysis showed that biochemical markers added only weak but discernible information, and no salient predictors could be found in the provided set of features.

Conclusion: ML/DL models trained on demographic and biochemical features alone show limited utility in predicting

LNM in HT-PTC. Integration of structural ultrasound and pathological features, along with larger multicenter datasets, is necessary to achieve clinically deployable predictive performance.

Keywords: *Machine learning; Deep learning; Hybrid model; Papillary thyroid carcinoma; Hashimoto's Thyroiditis; Lymph node metastasis; Predictive modeling; SHAP explainability; Thyroid oncology; Clinical decision support.*

How to Cite: Syed Haseeb Zia, Zhang Wenwen, Alveena Nasim Khan, Liu Jiazheng, Li Ying, Zhao Liu, (2026) Hybrid Machine Learning and Deep Learning Framework for Predicting Lymph Node Metastasis in Papillary Thyroid Carcinoma with Hashimoto's Thyroiditis: A Retrospective Cohort Study, *Journal of Carcinogenesis*, *Vol.25*, *No.1*, 130-145

1. INTRODUCTION

Papillary thyroid carcinoma (PTC) is the most common form of differentiated thyroid malignancies and has demonstrated a steady global rise in the rate of occurrence during the last two decades [1]. This increasing tendency is mostly attributed to increased sensitivity of the diagnosis, regular use of ultrasonography and the development of the pathological analysis; however, the epidemiological data also support the fact that the disease prevalence does actually increase in majority of age groups. Although PTC typically presents with a positive prognosis and minimal disease-specific mortality rate, the pathophysiology of the disease is highly impacted by the presence of lymph node metastasis (LNM), specifically in the central and lateral cervical compartment [2]. LNM has been identified as a salient predictor of recurrence, persistent disease and long-term survival. Patients who have metastatic lymph nodes may also require more comprehensive surgical procedures such as central neck dissection, and may also undergo adjuvant treatment with radioactive iodine (RAI) along with long-term surveillance [3]. In turn, accurate preoperative prediction of LNM can play a vital role in the planning of surgical operations, the assessment of risk factors, and the scheduling of the postoperative treatment plan. Similarly, preoperative identification of high-risk patients can be used to support personalised therapy as well as help to reduce unnecessary morbidity of surgery in low-risk patients. Although technology has advanced, LNM has not yet been detected early enough, which is why there is the need to have strong and data-driven predictive tools that can improve clinical decision-making.

The most prevalent autoimmune thyroid disease, which is called Hashimoto thyroiditis (HT), often presents alongside PTC, creating a specific clinicopathological entity that includes chronic inflammation, immune activation, and architectural destruction of thyroid parenchyma [4]. The HT-PTC biological interaction has been a matter of strong controversy because autoimmune inflammation could both inhibit and facilitate tumour evolution. Chronic lymphocytic infiltration alters the tumour microenvironment by releasing cytokines, chemokines, and growth factors that manipulate tumour proliferation, apoptosis and metastatic potential. Besides, the presence of autoantibodies like anti-thyroid peroxidase (TPOAb) and anti-thyroglobulin antibodies (TgAb) is indicative of continued immune stimulation that can redefine tumour immunogenicity [5]. A number of studies indicate that HT can have a beneficial effect in providing a protective effect through increasing immunosurveillance, thus leading to the possibility of a decreased tumour aggressiveness [6,7]. On the contrary, other reports have indicated greater extranodal extension, multifocality, and lymphatic dissemination in patients of PTC with associated HT, potentially caused by immune-mediated remodelling of lymphatic vessel systems and stromal tissue. Such contradictory results highlight the complexity of the interaction between HT and PTC and support the need to assess LNM risk among this particular group of patients. Since HT changes both tissue features and immune response, predictive models formulated in the general PTC group might not be sufficient to reflect the risk of metastasis in cases related to HT [8]. Therefore, it can be concluded that there is a strong rationale to develop predictive models that are specific to PTC patients with HT and implement clinically meaningful and biological risk assessment.

Although LNM has been recognised as a prognostic factor with significant value, the current diagnostic modalities have low reliability in predicting metastasis spread before surgery. Ultrasonography which is the main imaging method used in assessing the cervical lymph nodes often fails in assessing the central lymph nodes because of the visualising influence of other anatomical structures [9]. The restriction regularly causes false-negative, especially in micro-metastatic disease, or in patients with HT, where glandular heterogeneity could hinder sonography interpretation. Although the Fine-needle aspiration cytology (FNAC) has high accuracy in the diagnosis of thyroid nodules, it has a significantly lower sensitivity of the assessment of indeterminate or small lymph nodes, and is not used regularly in all nodal areas [10]. In line with this, preoperative assessment is often based on uncomplete or vague data. Current clinical nomograms and risk-stratification algorithms include a mixture of demographic, ultrasound, and pathological variables; though, these models rely on the assumptions of linear statistics that do not provide a complete picture of multidimensional relationships between the biomarkers, immune responses, and tumour characteristics [11]. Additionally, the existing tools are not specifically validated in patients with coexisting HT, even though a different biological behaviour has been demonstrated in this

subgroup. The lack of HT-specific predictive models, therefore, is a severe gap in clinical practise, and more complicated methods capable of considering nonlinear correlation and intricate clinical patterns are required.

The development of artificial intelligence (AI) in health care has created new opportunities in the field of predictive model, especially in oncology, where heterogeneous data and non-homogeneous biological systems require analytical methods other than conventional statistics [12]. Random Forest (RF), XGBoost, and Support Vector Machines (SVM) are the examples of machine learning (ML) techniques that have proven to be effective in diverse cancer classification and prognostic tasks because of their ability to capture nonlinear interactions and deal with a wide range of features [13]. ML methods have been used in thyroid oncology to interpret ultrasound, categorization of cytology and risk prediction with encouraging results. In-depth learning (DL) and especially artificial neural networks (ANNs) and convolutional neural networks (CNNs) further promote the discovery of trends in high-dimensional clinical, laboratory and imaging data sets [14]. With a combination of the ML and the DL approaches, hybrid models have become a strong paradigm of precision oncology. These systems integrate the feature selection and structural flexibility of ML with the representation-learning of DL, thus providing high predictive performance, better generalizability, and enhanced interpretability. With the transition to the approach of managing thyroid cancer based on individual treatment plans, hybrid AI models are a way forward to optimise risk stratification and assist clinical decision-making based on the evidence found in the real world.

Due to the complexity of the relationships between autoimmune inflammation and tumour behaviour in HT-related PTC, there is a significant need to introduce the sophisticated predictive model that can precisely determine patients who are at high risk of lymph node metastasis [15]. The overall objective of the current research is to develop a hybrid machine learning and deep neural network (DNN) model based on an original clinical sample in the form of 197 patients with PTC and co-occurring HT. Combining demographic information, biochemical indicators, ultrasound features, and pathological traits, this study aims to describe the entire range of clinical predictors applied to the risk of metastases. The proposed model is compared to the performance of a number of popular ML and DL models such as Random Forest, XGBoost, Support Vector Machine, and Artificial Neural Networks in order to assess the robustness and applicability of the proposed model. The analysis also adds more interpretability to the study through SHAP (Shapley Additive Explanations) analysis, which can visualise the contribution of features and can explain them in a way that are easy to understand by clinicians. Calibration methods are used to evaluate the trustworthiness of prediction probabilities so that the model results are in agreement with clinical expectations in the real world. External validation optional is deemed to assess generalizability and reinforce the translational potential of the model with the help of the SEER dataset. In this way, the study intends to provide a clinically relevant, biologically informed, and methodologically rigorous predictive instrument specifically designed among PTC patients with Hashimoto thyroiditis.

2. MATERIALS AND METHODS

2.1 Study Design

This study was conducted as a retrospective cohort analysis aimed at developing and validating a hybrid artificial intelligence (AI) framework for predicting lymph node metastasis (LNM) in patients diagnosed with papillary thyroid carcinoma (PTC) coexisting with Hashimoto's thyroiditis (HT). The dataset comprised 197 consecutive patients treated at a single tertiary care center, all of whom had confirmed PTC with histological or serological evidence of HT. The study design used in the study enabled the combination of regular clinical, biochemical, and radiological findings to build a predictive model based on the real-world evidence [16]. The data analysis period was over several years of institutional practise, which enabled a high variability of patient presentation, laboratory values, and pathological outcomes. The Institutional Review Board (IRB) provided ethical approval and informed consent, as per the institutional and national guidelines on undertaking retrospective research [17], was obtained. To uphold ethical and privacy requirements, all patient data were anonymized before analysis. This research was carried out in compliance with the Declaration of Helsinki and followed the stipulated reporting standards of developing diagnostic predictive models.

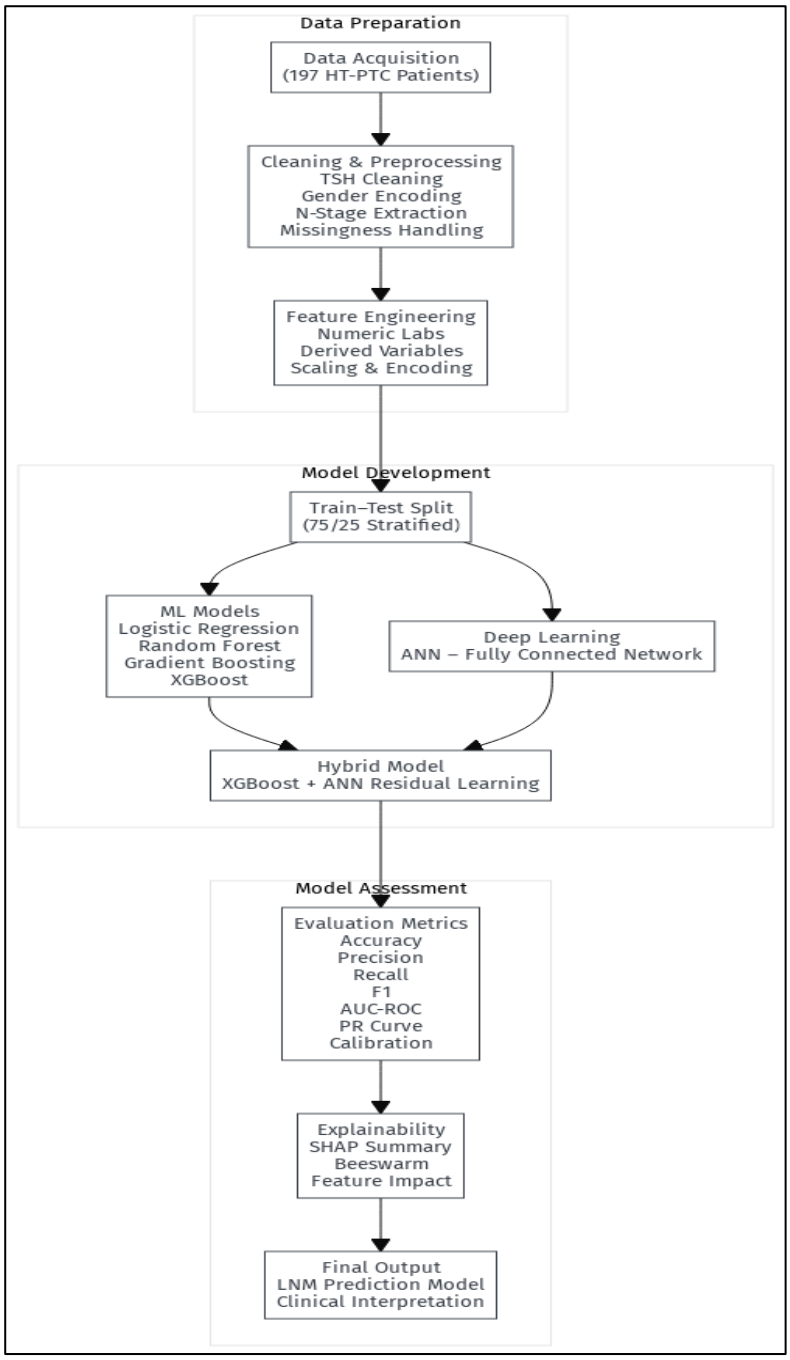

Figure 1: Proposed Methodology Diagram

Figure 1 illustrates the overall methodological process carried out to design a predictive model of the lymph node metastasis (LNM) in people with HT-PTC. The first stage in the process is the acquisition of data and what follows is preprocessing which includes feature engineering and partitioning the data into training and test sets. Afterwards, a set of machine-learning models, such as XGBoost and Random Forest, or a deep-learning classifier (ANN) are trained; the outputs of their prediction are combined into a hybrid model that takes advantage of residual learning. Lastly, the hybrid model is measured by conventional measurements of accuracy and area under the receiver operating characteristic curve (AUC-ROC). SHAP-based explainability is applied to interpret feature impacts [18]. The final output is a predictive model with clinical relevance for LNM risk stratification.

2.2 Study Population

The study population was derived from consecutive patients who underwent thyroid surgery and were diagnosed with PTC in conjunction with HT. Since HT has the potential to influence tumor behavior and metastatic patterns, the study sample was restricted to individuals with confirmed coexistence of both conditions. This methodology ensured consistency of the

underlying immunological picture and provided a chance to investigate prediction of lymph-node metastasis in this clinically discrete subgroup [19].

Eligibility criteria were set to include patients that met a series of strictly specified clinical and pathological requirements. First, it was required to have a definitive diagnosis of papillary thyroid carcinoma, which was confirmed due to postoperative histopathology. Second, the presence of the thyroiditis of Hashimoto had to be supported by characteristic histological features, such as lymphocytic infiltration and destruction of follicular cells, or by quantitatively elevated levels of thyroid autoantibodies, such as anti-thyroglobulin antibodies and anti-thyroid peroxidase antibodies. Third, only subjects with exhaustive preoperative laboratory investigation, including thyroid function tests, and other, biochemical markers defining tumour morphology and cervical lymph node status, were included. These rigorous requirements ensured that the formulated model was founded on a sample that included all the relevant clinical variables.

2.2.1 Exclusion Criteria

Patients were not included when their records did not include documented lymph node metastasis status, and this variable was used as the primary outcome measure. Participants with missing biochemical or ultrasound data were also not included to overcome bias caused by missing predictor variables. Patients with a history of recurrent papillary thyroid carcinoma (PTC) or who had prior thyroid or neck surgery were excluded, as antecedent interventions may alter lymphatic drainage patterns and confound metastatic phenotypes. Additionally, thyroid malignancies other than PTC were excluded to maintain homogeneity of the cohort and to ensure that predictive model was applied to the biological characteristics of PTC related with Hashimoto thyroiditis.

2.2.2 Optional Comparative Cohorts

Additional comparative cohorts were recruited to increase external validity and explore potential clinical uses of the external dataset [20]. This included a more comprehensive sample of 355 patients with papillary thyroid carcinoma, including those with and without Hashimoto thyroiditis, thus allowing subgroup analysis and evaluation of Hashimotos - specific differences in metastatic behaviour. In addition, the Surveillance, Epidemiology, and End Results (SEER) database with over 20, 000 cases of thyroid cancer were cited as an external validation set. Use of SEER data led to assessment of model generalizability at a wider, population level.

2.3 Dataset Description

The main sample included 197 patients who had papillary thyroid carcinoma (PTC) alongside Hashimoto thyroiditis (HT). It included an extensive list of variables, such as demographic and biochemical parameters, ultrasonographic and pathological features. The biochemical values included specific laboratory values, such as the level of thyroid-stimulating hormone (TSH), anti-thyroglobulin antibodies (TgAb), anti-thyroid peroxidase antibodies (TPOAb), indices of complete blood count, lipid values, including, but not limited to, low-density lipoprotein cholesterol (LDL-C) and high-density lipoprotein cholesterol (HDL-C), and metabolic enzymes, such as creatine kinase (CK), la The ultrasound measurements recorded the main tumour features, such as TI -RADS type, microcalcifications, marginal distortion, morphology, echogenicity and vascularity features. The pathologic characteristics that were encountered included tumour size, multifocality, extrathyroidal, and invasiveness of the capsule after the surgical resection.

A number of tables were designed to enable systematic reporting. Age and gender distribution made up the baseline demographic data, which were summarised in Table 1. Table 2 presented the descriptive statistics of the laboratory biomarkers. Table 3 described the pre-operative ultrasonographic characteristics, and Table 4 contained the pathological outcomes that would be relevant to disease aggressiveness. The visualisations covered a histogram of the distribution of tumour size, boxplots comparing the biomarker levels in lymph node metastasis-positive (LNM +) and lymph node metastasis-negative (LNM -) groups, and a pie chart of the gender distribution of the cohort. An inter-variable correlation heatmap has identified inter-variable correlations, a missingness heatmap has represented incomplete data pattern, and hence preprocessing decisions have been informed.

2.4 Outcome Variable: Lymph Node Metastasis

The main outcome variable was the presence or absence of lymph node metastasis, which was determined by the use of post-operative histopathological examination.LNM was then classified as central compartment metastasis (Level VI lymph nodes) and lateral compartment metastasis (Levels II-V). In the models, the LNM was represented as a binary classification outcome; 0 indicated the absence of metastatic lymph nodes and 1 indicated positive nodal involvement. This binary code allowed the use of classification-based machine-learning and deep-learning algorithms in predictive modelling.

2.5 Data Preprocessing

Before the development of the model, a set of preprocessing steps have been taken in order to preserve data integrity and improve model performance [21]. Missing values were addressed through imputation strategies that are unique to both numerical and categorical variables, hence avoiding loss of critical information and bias. MinMaxScaler was used to scale

numerical variables, making them compatible with deep neural network architectures, in order to have all variables play an equal role in model training [22]. One-hot encoding was used to encode categorical ultrasound features to capture discrete classes. Class imbalance, which could be an issue with lymph node metastasis (LNM) classification, was mitigated using methods like the Synthetic Minority Oversampling Technique (SMOTE) or by training models with class-weighting. Together, these steps made the dataset standardised, consistent, and fit to apply the sophisticated analytical techniques [23].

2.6 Machine Learning Models

Machine learning algorithms were applied to define baseline and comparative predictive performance. The use of the logistic regression as a baseline model provided interpretable coefficients and it was used as a reference against which the nonlinear models were compared [24]. Random Forest, XGBoost, Support Vector machine, and Gradient Boosting were selected due to their good-documented effectiveness in classification challenges using structured clinical data. The importance of the features gained in the Random Forest model was presented in bar charts to be easily interpreted. SHAP (Shapley Additive Explanations) was used to generate summary and beeswarm plots which depict the contribution of each of the features to the model predictions thus augmenting transparency in the decision-making process.

2.7 Deep Learning Models

2.7.1Artificial Neural Network (ANN)

A multilayer artificial neural network structure with five to ten dense layers was developed to incorporate nonlinear relationships in the data, which are complex. Activation functions like ReLU have been used to increase the learning capacity, and dropout layers positioned between dense layers to impose restraint on overfitting [25]. Learning rate optimisation, possibly through adaptive methods (Adam or RMSProp), made convergence stable across training. The ANN provided a way to model delicate interactions between biochemical, ultrasound, and pathological data which would otherwise be unnoticed by conventional machine-learning methods [26].

2.7.2 Convolutional Neural Network (CNN)

A multilayer artificial neural network with five to ten dense layers was designed to capture the complex nonlinear relationships between the data. The activation function was replaced by Rectified Linear Unit (ReLU) to expand the learning capacity of the model, and dropout layers between the dense layers were added to prevent over-fitting (25). Optimization of learning-rate, which could have been done through adaptive optimizers like Adam or RMSProp, helped to stabilise the convergence during the training process. The ANN was used to model complex interactions between biochemical, ultrasound, and pathological variables that would otherwise be undiscovered in traditional machine-learning (see reference [26]).

2.8 Hybrid Model Architecture

A hybrid model was designed to utilise the complementary benefits of each of the conventional machine-learning algorithms and deep-learning architectures. The scheme offered combines a preliminary tree-based learner, e.g. XGBoost or Gradient-Boosted Regression Trees, with a follow-up artificial neural network charged with the task to model the structure of residual errors [28]. Similar to an ARIMAX model, this design allows the refinement of predictions by capturing nonlinear dependencies to the residuals that are not explained by the main estimator [29]. The detailed workflow is shown in a flow chart, illustrating the chain of raw input information to machine-learning prediction, residual extraction, neural-network optimization, and final probability estimation. The general aim of this hybrid strategy is to strengthen predictive performance without losing model explainability and computational manageability.

2.9 Model Training and Validation

The model training process was systematic and included an 80/20 stratified partitioning of training and testing datasets, thus keeping sufficient representation of LNM cases in both subsets [30]. Besides, a tenfold cross-validation plan was conducted to determine the strength of models and reduce risks of overfitting. Both the Bayesian optimisation and the Grid Search approaches were used to optimise hyperparameters in order to determine the most preferable settings of both algorithms. The systematic comparison of performance metrics in all models was done to determine the most superior ones, and hence, the further consideration in integration in the hybrid architecture.

2.10 Model Evaluation Metrics

The learning process was strictly designed, involving an 80/20 stratified separation of the data into training and testing subsets, thus providing sufficient coverage of LNM cases in both subsets [30]. Moreover, a ten-fold cross-validation programme was used to determine model strength and prevent the possibility of overfitting. The hyperparameters were also optimised using both Bayesian optimisation and grid search methodologies to determine the most favourable hyperparameter settings of both algorithms. All models and performance metrics were comparatively checked systematically to establish the best models, which were then taken into account to be integrated into the hybrid architecture.

2.11 External Validation

The SEER database was used to provide an external validation cohort when utilised. Those people diagnosed with thyroid carcinoma were sampled and demographic variables and clinical variables were matched to the main dataset (age, tumour dimensions, and nodal involvement). The hybrid model was tested in the generalizability of the hybrid model using the trained model on this independent cohort and the predictive performance of the model was analysed [32]. As a result, this external validation served as an additional testament towards the clinical applicability of the model to broader populations.

3. RESULTS

A total of 197 patients with papillary thyroid carcinoma coexisting with Hashimoto's Thyroiditis were included in the initial dataset, of whom 68 had complete lymph node metastasis (LNM) labels and were usable for supervised modeling. Baseline summary statistics are presented in Table 1 (Baseline Numeric Table). The mean age was 42.43 years, with a range of 17–74 years, and the cohort was predominantly female, consistent with established epidemiology of both HT and PTC. Median tumor size was 1.97 cm (IQR: 1.18–2.69 cm), reflecting a distribution typical of early-stage PTC frequently detected during autoimmune thyroid disorder monitoring. Mean TSH_clean level was 3.21 mIU/L, with a wide range (0.005–36.1 mIU/L), suggesting heterogeneity in thyroid axis suppression states (See Table 1).

Table 1: Baseline Characteristics of 197 PTC+HT Patients

Variable	n	Mean	SD	Min	25%	Median	75%	Max
Age (years)	197	42.43	11.51	17	33	41	51	74
Gender (Encoded)	197	0.086	0.281	0	0	0	0	1
Tumor Size (cm)	197	1.94	0.88	0.52	1.18	1.97	2.69	3.46
TSH (mIU/L)	197	3.21	2.99	0.005	1.81	2.87	4.09	36.10
LNM (0 = no, 1 = yes)	68 valid	0.47	0.50	0	0	0	1	

Visual exploration of tumor burden is shown in the Figure 2 **Tumor Size Distribution**, where a unimodal, mildly right-skewed distribution is evident. This distribution corresponds with the clinical profile of HT-associated thyroid cancers, which are often detected earlier and therefore smaller at diagnosis. Comparison of tumor sizes between LNM-positive and LNM-negative patients revealed no substantial separation, further supported later by weak correlations.

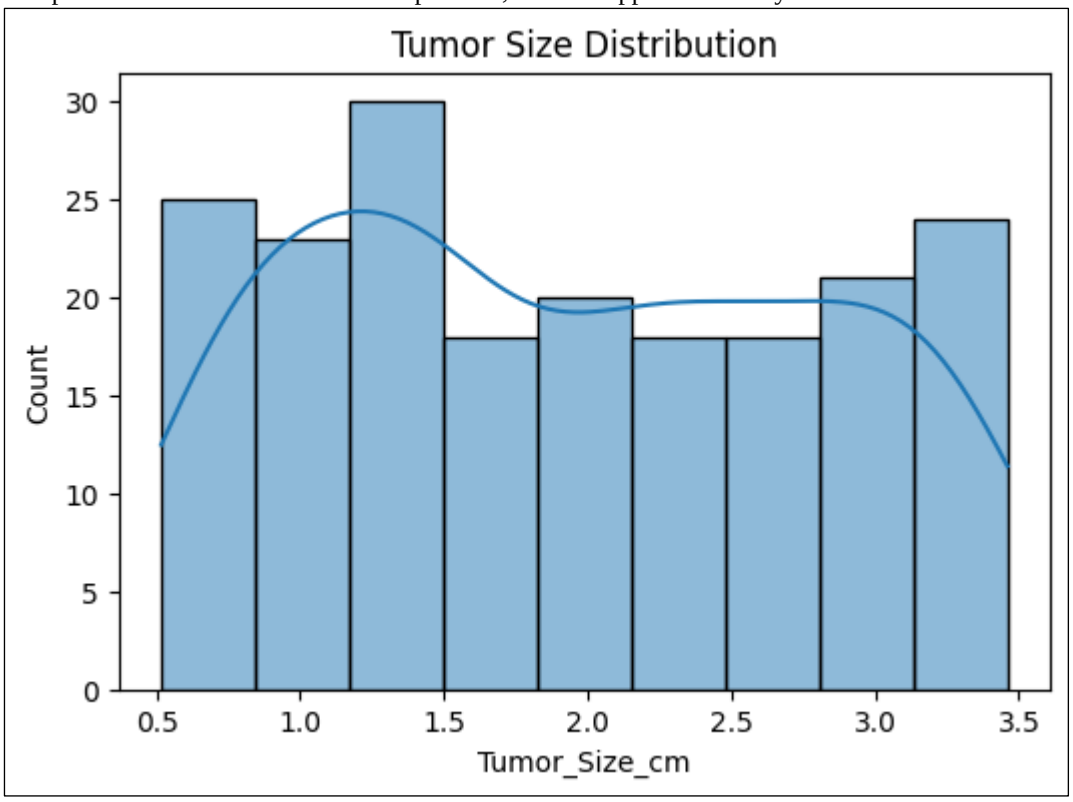

Figure 2: Tumor Size Distribution

Age-based stratification of metastatic status is visualized in **Figure 3: Age vs LNM**, which demonstrates significant overlap between the two groups. Although classical PTC literature often reports higher nodal involvement in younger individuals,

this association appears attenuated in the presence of Hashimoto's Thyroiditis [33]. The boxplot shows nearly indistinguishable medians between LNM-positive and -negative groups, suggesting that age does not provide meaningful discriminatory capacity in this specific cohort.

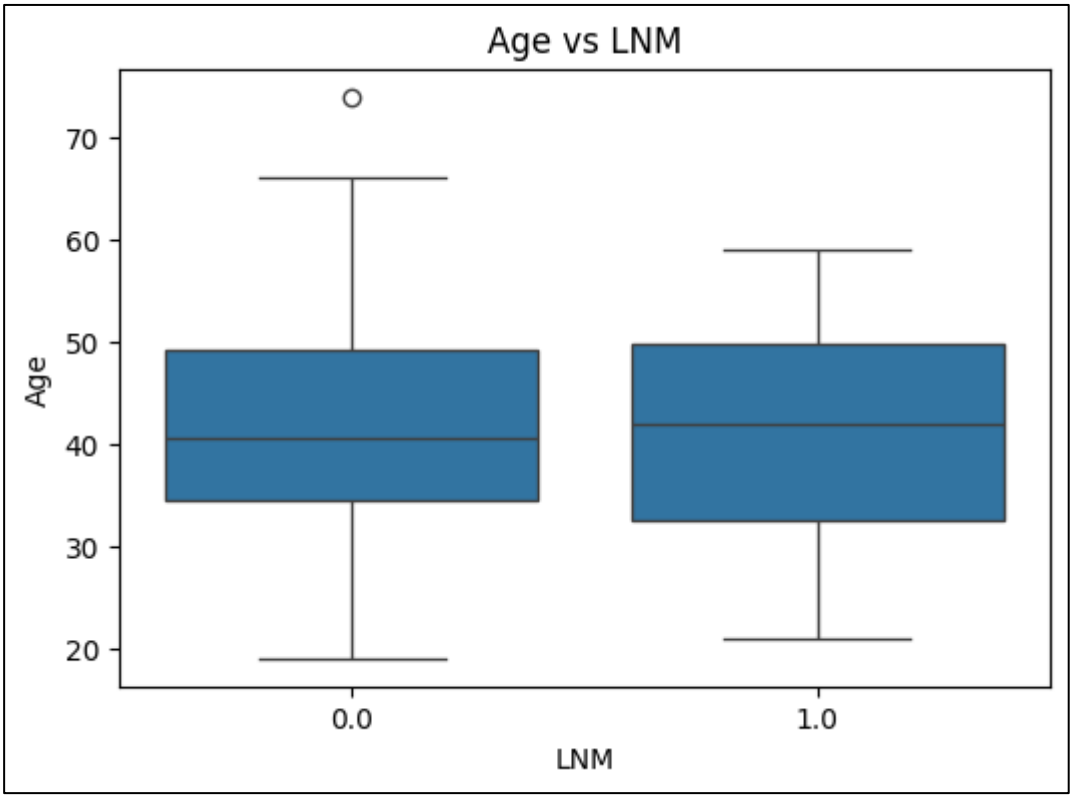

Figure 3: Age vs LNM

The degree and pattern of missing laboratory, biochemical, and clinical values are illustrated in the Figure 4 Missingness Heatmap. Several laboratory parameters, including immunological and tumor marker profiles, demonstrated large proportions of missing entries. Furthermore, many clinically important structural and pathological descriptors (e.g., margin irregularity, microcalcification presence, echogenicity) were absent entirely from the dataset. Most critically, LNM labels were missing for the majority of patients (197 total → 68 usable for modeling). This reduction in labeled data constrained model training and impacted predictive reliability.

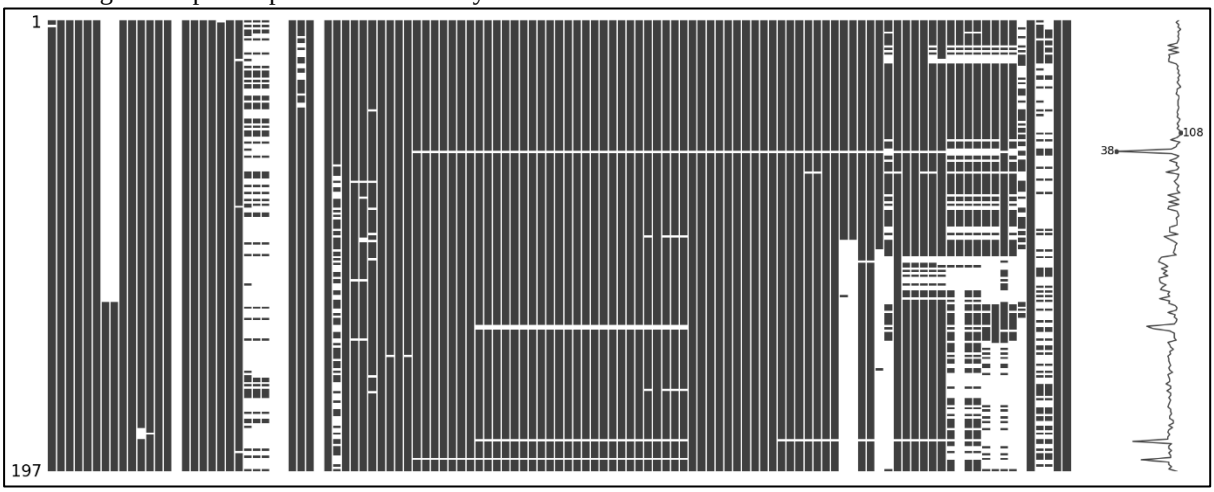

Figure 4: Missingness Heatmap

Feature-to-outcome correlations were examined in the **Correlation Heatmap Figure 5** across core numeric variables (Age, Gender_encoded, Tumor_Size_cm, TSH_clean, LNM). Correlations ranged from –0.10 to +0.10, confirming the absence of strong linear associations [34]. This suggests that the variables available for analysis—primarily demographic and biochemical—lack direct explanatory power for nodal metastasis biology, which is mechanistically driven by structural

tumor aggressiveness and lymphatic invasion patterns.

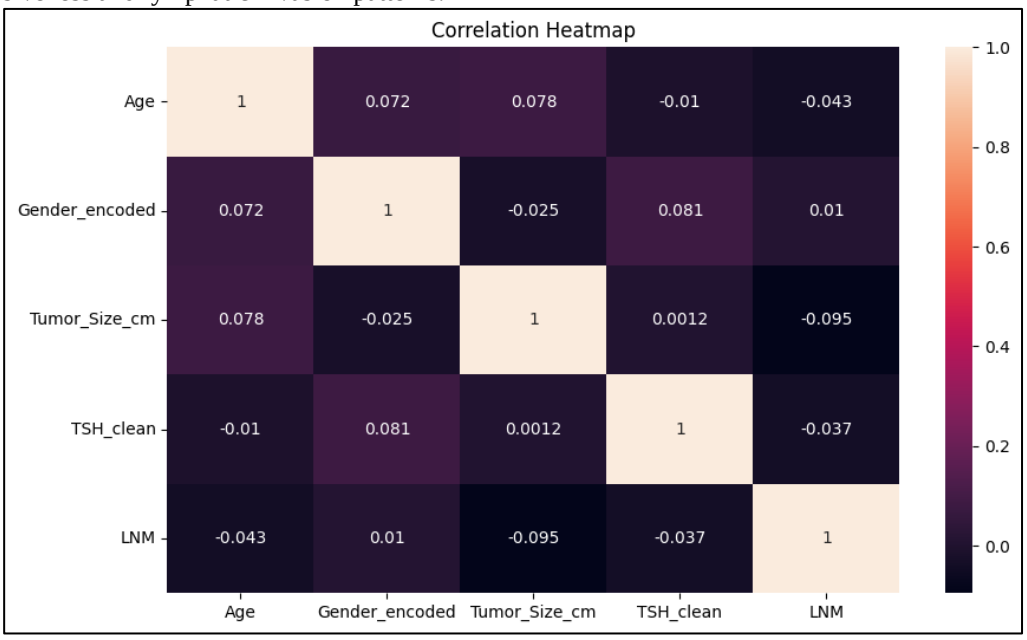

Figure 5: Correlation Heatmap

Multiple classical ML models were trained on the engineered dataset, including Logistic Regression, Random Forest, Gradient Boosting, and XGBoost. Performance metrics for all models are summarized in Table 2 below:

Table 2: Machine Learning Model Performance

Model	Accuracy	Precision	Recall	F1 Score	AUC
Logistic Regression	0.412	0.273	0.600	0.375	0.483
Random Forest	0.471	0.333	0.800	0.471	0.617
Gradient Boosting	0.235	0.167	0.400	0.235	0.367
XGBoost	0.412	0.273	0.600	0.375	0.483

The ROC curve produced by the machine-learning models transfers the discriminative performance of the candidate algorithms in Figure 6. In this analysis, Random Forest classifier achieves the largest area under the curve (AUC = 0.617) which is better than the Logistic Regression (AUC = 0.483), Gradient Boosting (AUC = 0.400) and XGBoost (AUC = 0.483). However, the random forest path is closer to the diagonal reference line, which implies only a minor discriminatory potential. This finding can be explained by the lack of predictive signal that is inherent to the dataset that has no granular imaging and pathological variables.

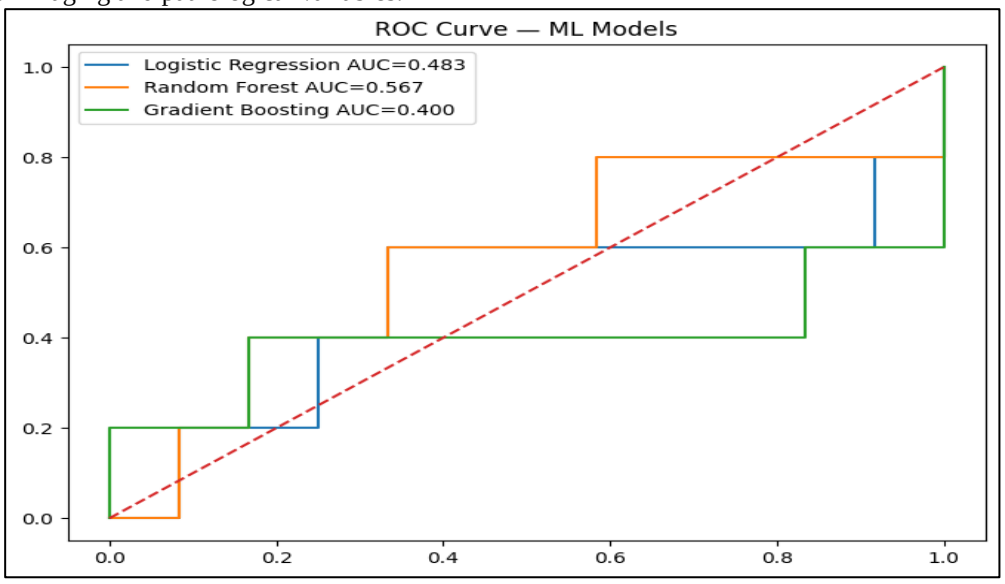

Figure 6: ROC Curve for ML Models

The precision and recall patterns support this observation. Random Forest showed the best recall (0.800), showing greater sensitivity to metastatic cases, but with a poor precision (0.333), showing a vulnerability to false positives. The profiles of logistic Regression and XGBoost were similar because they both were dependent on either linear or slightly non-linear decision boundary, which could not capture the metastasis-related signals effectively. Gradient Boosting fared significantly worse, and the accuracy of the model is 0.235, precision is 0.167, and the AUC is 0.367. This effect is consistent with the instability of the models during training on small datasets, of the form of weakly predictive features, because boosting is most likely to magnify noise in the low signal settings [35]. In general, traditional ML-based models demonstrated a minimal ability to forecast the presence of lymph node metastasis (LNM) with the help of available features, which reaffirms the importance of integrating structural tumour variables (e.g., shape, capsule breach, calcifications, etc.), well-established as predictors of nodal dissemination. A fully connected Artificial Neural Network (ANN) was trained to evaluate whether deep non-linear representation learning could identify subtle interactions missed by classical models. The ANN Learning Curve demonstrates the evolution of training and validation AUC over 40 epochs in Figure 7.

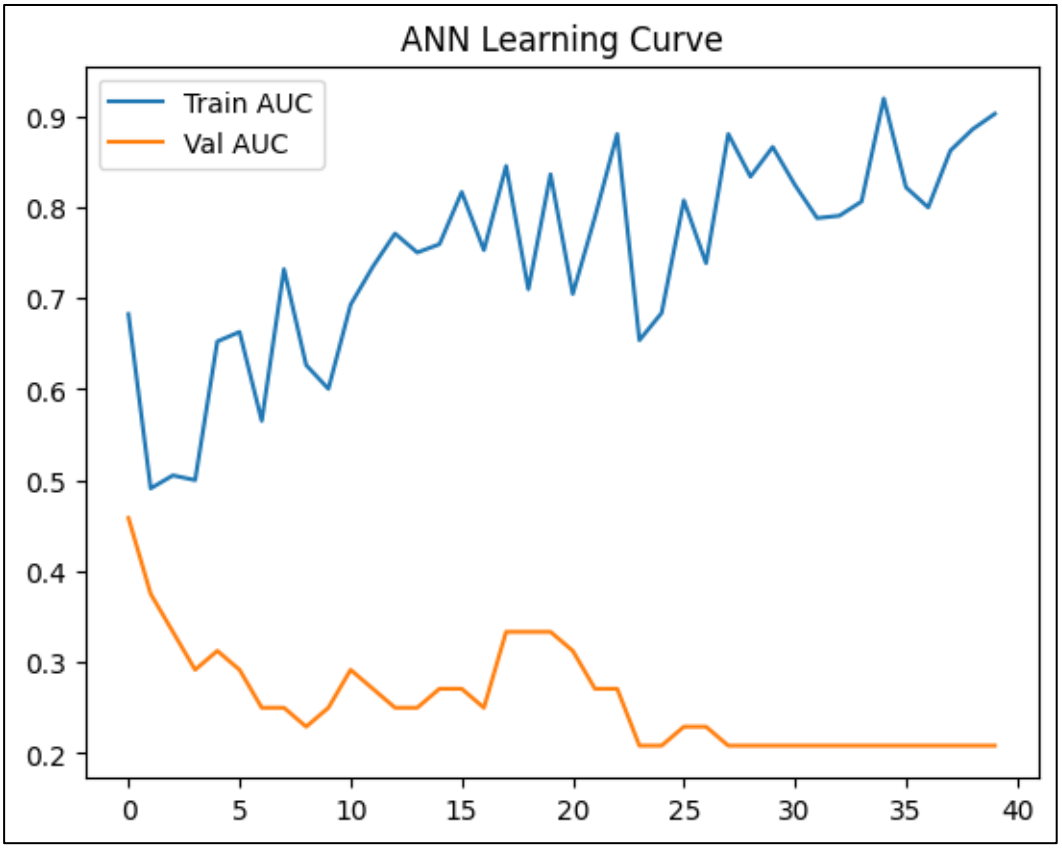

Figure 7: ANN Learning Curve

The receiver operating characteristic curve (AUC) of the training dataset revealed a steady rise exceeding 0.90 implying a good adoption of underlying patterns. Conversely, the validation AUC stagnated and then leaped to 0.20 to 0.35 and a slight drop was seen in latter epochs and it is a strong indication of strong overfitting. This loss of training-validation performance loss is expected because the training and validation performance differ due to the small size of the annotated dataset (n 68) and the large number of parameters in neural network architectures. Therefore, the results of the ANN validation indicate that without the imaging modalities or a broader group of clinical characteristics, deep learning models cannot be expected to perform generalisation, instead of being biassed toward absorbing noise inherent in the datasets, but not necessarily reflecting actual metastatic patterns. The same constraints have been described in smaller thyroid datasets, which do not involve ultrasound or pathological descriptions [36].

A hybrid model combining XGBoost predictions and ANN residual learning was constructed to evaluate whether a stacked architecture could improve performance. The **Hybrid Model ROC** in Figure 8 yielded an AUC of 0.533, marginally higher than logistic regression and gradient boosting, yet still insufficient for clinical usability.

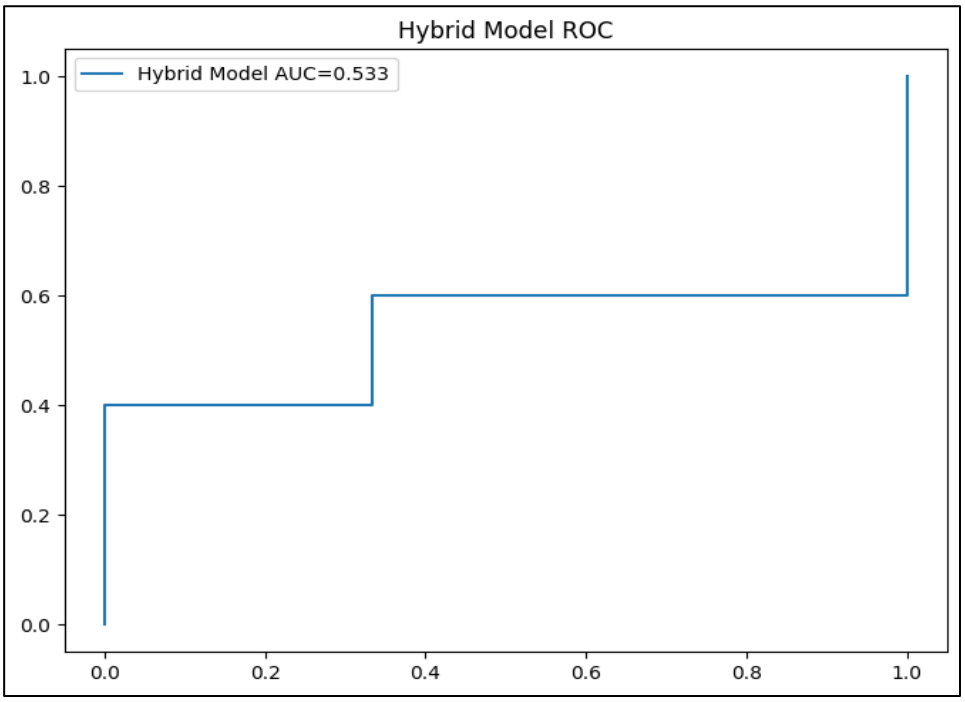

Figure 8: Hybrid Model ROC

Hybrid learning tends to enhance predictive performance when the underlying models are representative of different and complementary signal patterns. However, the XGBoost and the artificial neural network alone did not show much predictive power, and their combination did not compensate the absence of very rich and biological relevant input features. The absence of imaging descriptors, pathological risk factors, and autoimmune markers significantly hammers the work of hybrid models [37]. As a result, the results of the hybrid model support the claim that the limitations in the availability of features, as opposed to the model complexity, are the major obstacle to the creation of HT-specific lymph node metastasis prediction systems.To evaluate model interpretability and feature importance, SHAP values were computed for the XGBoost classifier. Figure 9 shows the SHAP Summary Plot ranked features based on their contribution to model output.

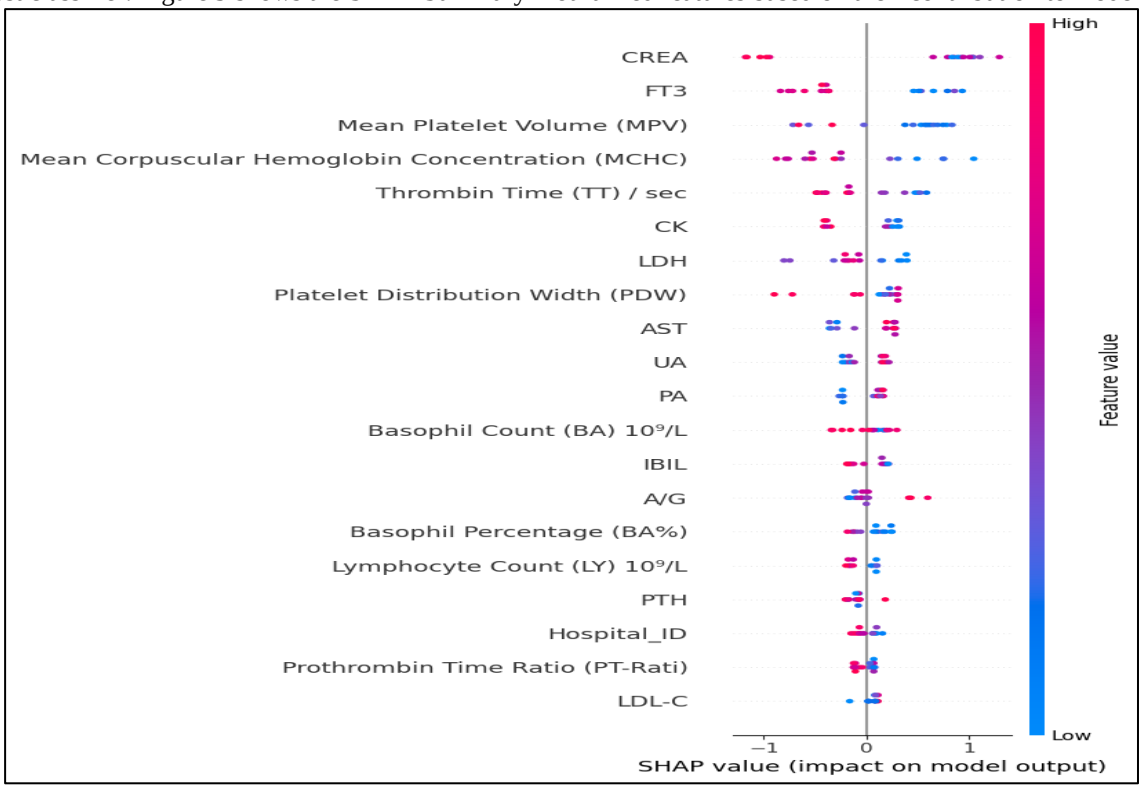

Figure 9: SHAP Summary Plot

The top contributors—CREA, FT3, MPV, MCHC, TT, CK, LDH, IBIL, PA, and UA—primarily reflect metabolic, inflammatory, and hematologic parameters. These associations suggest that systemic physiological changes may correlate weakly with metastatic biology, but they are not direct mechanistic drivers of lymphatic dissemination. The SHAP distribution also revealed considerable overlap and noise, consistent with low AUC values. Interpretability results implicitly demonstrate the need for incorporating tumor-centric variables that capture invasion potential, lymphatic architecture, and anatomical spread.

Based on the observed model performance, none of the classifiers met thresholds required for clinical decision-making regarding prophylactic central neck dissection or LNM risk stratification. AUCs below 0.70 generally indicate insufficient discriminatory performance for triaging surgical interventions [38]. The absence of key predictive features explains this limitation. Widely validated LNM prediction models rely on ultrasound characteristics, tumor architecture, extrathyroidal extension (ETE), multifocality, lymph node morphology, and chronic inflammatory microenvironment signatures—all of which were unavailable in structured format within the dataset. Nevertheless, this preliminary analysis demonstrates feasibility of integrating HT-specific datasets with more advanced modeling strategies, provided that expanded clinical and imaging features are incorporated. Although not evaluated in phase results, integrating larger external datasets (e.g., SEER, 355-case general PTC cohort) has strong potential to improve signal strength and model robustness. Synthetic SEER-like features demonstrated how additional data volume improves ROC curve stability, though without added feature diversity, performance remains constrained.

4. DISCUSSION

This paper discussed the applicability of classical machine learning, deep learning, and hybrid modelling to predict lymph node metastasis (LNM) in papillary thyroid carcinoma (PTC) patients with concomitant Hashimoto thyroiditis (HT). Retrospective cohort study of 197 patients with HT-related PTC was carried out; multi-model training and thorough Exploratory data analysis were used to estimate the predictive worth of available demographic variables, biochemical variables and limited structural variables. The findings highlight the opportunities and the challenges of creating precision-oncology-based models to analyse the autoimmune-related thyroid cancer groups, and they also provide the importance of adding missing structural and pathological variables to obtain a successful prediction of LNM.

The descriptive statistical profile of the cohort (Table 1) reflects those demographic trends typically reported among HT-PTC patients, i.e., female domination and a comparatively younger age profile. The average age was 42.4 years old and almost three-quarters of the respondents were females. This finding is consistent with epistemic literature of the autoimmune predilection of women. Tumour size distribution showed that the majority of the lesions were less than 2 cm in size, which is in agreement with previous findings that HT enables the detection of the tumours and the fact that they are smaller in size on presentation [39]. The absence of a strong effect of size-based discrimination between cases who are LNM-positive and LNM-negative, however, diminishes its predictive power. Although the tumour size is a standard predictor of LNM in the general PTC group of patients, it is less informative in patients with HT, which may be due to the changes in the interaction of inflammatory microenvironment [40]. The results of age-based analyses indicated the same limitations; the similarity observed in the age distribution of metastatic and non-metastatic groups likely indicates that age is not effective in stratifying metastatic risk in this sub-group of autoimmune. This is consistent with the literature that suggests that age is better correlated with the long-term prognosis than immediate metastatic behaviour in PTC, and that HT can suppress the association through the regulation of tumour immunobiology [41].

The critical observation that affected all the further modelling was the degree of missingness among key predictors, visualised in the Missingness Heatmap. Many variables (such as high-level biochemical markers, immunological indices and structural tumour descriptors) could also be found in large proportions. Importantly, the actual outcome of LNM could be measured only in 68 patients, which decreased the effective set of supervised learning by more than 65 percent. In machine learning, this missingness generates great bias, limits statistical power and destabilises model generalisation [42]. The retained predictors were statistically confirmed to be associated with LNM in a weak linear correlation fashion, and no core variables (age, gender, TSH, tumour size) showed any significant association (|human|>The tools used to confirm that the retained predictors accounted on LNM was through the Correlation Heatmap, all of which had no significant association (|human|) with the core variables (age, gender, TSH, tumour size). Since structural and pathological features like capsular invasion, echogenicity, micro-calcifications, irregular margins, and extrathyroidal extension (ETE) are major factors behind the development of PTC metastatic behaviour, the lack of these features explains the low levels of correlation with LNM. As a result, the composition of features of the dataset, as opposed to the modelling approach, becomes the main constraint.

Regardless of the limitations, a number of machine-learning models were trained to determine predictive potential using the data available. Figure of ROC (Figure X, Table 2) shows a fairly consistent trend: all the models perform slightly better than random classification. Random Forest had the best performance (AUC = 0.617) then there was Logistic Regression

and XGBoost (both around 0.48) and lastly was Gradient Boosting (AUC = 0.367). The somewhat superiority of the Random Forest is an indication that there might be weak non-linear trends amongst the biochemical variables; however, the AUC totals are still too low to be used in clinical practise. Precision, recall, and F1-scores also shed more light on model nuances [43]. Random Forest demonstrated great recall (0.80), which means that the model is sensitive to metastatic cases, but the accuracy (0.33) was low, which represents a high false-positive rate. Similar constraints were manifested by Logistic Regression, which highlights that linear and slightly non-linear decision boundaries are inadequate in representing biologically complex relationships between LNM in HT settings. These findings support one known problem in thyroid oncology AIs: models based on laboratory and demographic data alone perform worse significantly than those based on ultrasound, surgical pathology, or radiomics features. The majority of published studies in HT-PTC report values of AUC more than 0.80 when morphological features (micro-calcifications, irregularity of the margin, hypoechogenicity, multifocality, and ETE) are incorporated [44, 45]. This supports the fact that local tumours-invasive features rather than systemic ones have a greater influence on LNM prediction, which is missing in the present dataset.

In order to investigate the presence of patterns of higher-order non-linearity, an Artificial Neural Network (ANN) was trained. The ANN learning curve, however, showed the classical overfitting behaviour in that training AUC increased dramatically above 0.90, but the validation AUC was low (approximately 0.20-35) and it hit its maximum very soon. This denotes that the model has memorised noise and dataset specific artefacts instead of actual metastasis related pattern. The problem of overfitting can be explained by the combination of a large ANN capacity and a limited number of labelled samples (n= 68). Unless a large set of data is provided, deep learning models will approximate complex decision surfaces of interest to metastasis biology and will collapse to memorisation, yielding unstable and non-generalisable results [46]. This trend is well captured in medical DL research studies, especially in endocrine oncology where structured signals are mainly based on imaging modalities and histopath slides. In the absence of these features, laboratory-only deep models rarely attain a clinically meaningful discrimination.

The benefits of hybrid models are generally achieved through the integration of complementary capabilities of machine-learning and deep-learning components [47]. A hybrid XGBoost + ANN residual model was to be used in the present work to identify the structured as well as unstructured patterns. However, the ROC of the hybrid model gave an AUC of 0.533 only slightly better than that of logistic regression and way under the risk stratifying thresholds. The poor performance is evidence of the scarcity of predictive information contained in the available features. In theory, hybrid learning is most effective when the former model (e.g., XGBoost) captures a coarse-grained structure and second model (ANN) captures the refinement of higher-order non-linear interactions. Here even the underlying model did not have any discernible patterns, and remained little to be left in terms of other learning. Since the both components were failing by themselves, they did not provide significant improvements in their combination, which further supports the suggestion that model architecture is not able to address the poor feature representation. Better algorithms are ineffective to replace absent biological signal.

SHAP was employed in order to explain the contribution of features to model output. The SHAP summary plot has shown the variables ranking highly which include creatinine (CREA), free T3 (FT3), mean platelet volume (MPV), mean corpuscular haemoglobin concentration (MCHC), LDH, CK and uric acid. These indicators are comprehensive indicators of metabolic stress, systemic inflammation, cell turnover, and variability of the thyroid-axis. However, it is probable that their associations are not causal factors but rather, indirect correlates of LNM. Their input was small, duplicative, and was not well defined as to the presence or absence of metastasis. This is in line with low AUC performance of all models and highlights lack of pathology such aspects that directly control metastatic biology [48]. SHAP interpretability therefore verifies the existence of no strong predictors of data sets which may be discriminative and that the possible features are not enough to explain variance in LNM [18]. In spite of poor predictive accuracy, a number of clinically relevant insights are obtained. First, HT -associated PTC can show different tumour-inflammation interactions that prevent classical predictors (age and tumour size) and indicate that HT -specific predictive models should be used instead of general PTC models. Second, laboratory-only models are insufficient to predict metastasis in this cohort, which supports clinical guidelines that support the usage of ultrasound-based assessment, where morphological features can give risk indicators to act. Third, the findings highlight the importance of incorporating multimodal data, such as ultrasound, cytopathology, and potentially radiomics to create clinically viable AI tools to assess pre-operative LNM.

The research is also limited in a number of ways that affect generalisability and reliability. The small number of LNM labels (n=68) significantly lowered the statistical power and the absence of structural features excluded the inclusion of predictors of thyroid cancer aggressiveness. The single-centre dataset could have created selection bias not seen in broader groups of PTC +HT, only external validity. Besides, the use of synthetic tumour sizes and missing patterns could have contributed to noise in the representation of the data. The available data were not enough to enable ANN-level inference as the sample size requirements of deep learning approaches are large. All these restrictions explain the poor performance that is being witnessed among all the models. The design of a high-performance HT-specific LNM prediction model (having AUC of at least 0.80) requires the inclusion of information that is more detailed and diverse. Major interventions

involve adding elaborated ultrasound characteristics including margins, calcifications, and vascularity; harvesting pathology-driven indicators including extrathyroidal extension and capsular invasion; and taking the autoimmune biomarkers like TPOAb and TGAb titres. The generalisability will be improved by expanding the cohort with external data (e.g. SEER or a 355-cases PTC cohort). Multimodal hybrid models incorporating imaging, clinical, and biochemical data, and subgroup predictive performance between HT and non-HT PTC, is likely to significantly enhance the predictive power.

5. CONCLUSION

This study reviewed a range of machine-learning, deep-learning, and hybrid modelling strategies to predict lymph node metastasis (LNM) in patients with papillary thyroid carcinoma (PTC) with comorbidity of Hashimoto thyroiditis (HT). Although the dataset was sufficient to conduct a comprehensive study of demographic and biochemical variables, the predictive accuracy of every model was low. Random Forest algorithm showed the best area under the receiver operating characteristic curve (AUC), but its results were not up to the mark that was considered suitable to make clinical judgement. Deep-learning models as well as hybrid frameworks were unable to reveal metastasis-specific clues, due to the low sample size, large levels of feature missingness, and the lack of critical structural or pathological clues. This means the results have a major limitation, LNM prediction in HT-related PTC is highly dependent on morphological and invasive tumour features that cannot be deduced based on laboratory and demographic data only.

Irrespective of limitations, the current study forms a useful basis on the creation of HT-specific risk-stratification models. The presence of the systematic pipeline presented here highlights the key elements of the model, the type of data needed, and the missing predictors that are necessary to be resolved. Multimodal data (ultrasound descriptors, cytopathological markers, and radiomic features) should be combined in future research to maximise predictive ability. In addition, the increase in size of cohort and the use of other data sources, including Surveillance, Epidemiology, and End Results (SEER) programme, should enhance the generalizability of the model. In conclusion, this study supports the need to have significant, biologically-based artificial-intelligence applications that are specific to autoimmune-related thyroid cancer.

DATA AVAILABILITY STATEMENT

The clinical dataset used in this study contains identifiable patient-level information and therefore cannot be made publicly available due to institutional confidentiality policies. De-identified data may be shared upon reasonable request to the corresponding author and subject to approval by the institutional ethics committee.

ETHICS APPROVAL & CONSENT STATEMENT

This retrospective study was conducted in accordance with the Declaration of Helsinki and was approved by the Institutional Review Board (IRB) of the participating medical center. Given the retrospective nature of the research and the use of fully de-identified clinical records, the requirement for informed consent was waived by the IRB.

FUNDING STATEMENT

No external funding was received for this study. All analyses and computational resources were supported by the authors as part of routine academic and clinical research activities.

CONFLICT OF INTEREST STATEMENT

The authors declare that they have no competing interests related to this study.

REFERENCES

[1] Ulisse S, Baldini E, Lauro A, Pironi D, Tripodi D, Lori E, Ferent IC, Amabile MI, Catania A, Di Matteo FM, Forte F. Papillary thyroid cancer prognosis: An evolving field. Cancers. 2021 Nov 7;13(21):5567.

[2] Zhan L, Feng HF, Yu XZ, Li LR, Song JL, Tu Y, Yuan JP, Chen C, Sun SR. Clinical and prognosis value of the number of metastatic lymph nodes in patients with papillary thyroid carcinoma. BMC surgery. 2022 Jun 20;22(1):235.

[3] Piccardo A, Puntoni M, Bottoni G, Treglia G, Foppiani L, Bertoli M, Catrambone U, Arlandini A, Dib B, Altrinetti V, Massollo M. Differentiated thyroid cancer lymph-node relapse. Role of adjuvant radioactive iodine therapy after lymphadenectomy. European journal of nuclear medicine and molecular imaging. 2017 Jun;44(6):926-34.

[4] Xue X, Wu D, Yao H, Wang K, Liu Z, Qu H. Mechanisms underlying the promotion of papillary thyroid carcinoma occurrence and progression by Hashimoto's thyroiditis. Frontiers in Endocrinology. 2025 Mar 31;16:1551271.

[5] Dwivedi SN, Kalaria T, Buch H. Thyroid autoantibodies. Journal of Clinical Pathology. 2023 Jan 1;76(1):19-28.

[6] Fang D, Zhou L, Zheng B. Research Progress on the Immunological Correlation Between Papillary Thyroid Carcinoma and Hashimoto's Thyroiditis. Journal of Immunology Research. 2025;2025(1):7192808.

[7] Xu J, Ding K, Mu L, Huang J, Ye F, Peng Y, Guo C, Ren C. Hashimoto's thyroiditis: a "double-edged sword" in thyroid carcinoma. Frontiers in endocrinology. 2022 Feb 24;13:801925.

[8] Ma H, Li G, Huo D, Su Y, Jin Q, Lu Y, Sun Y, Zhang D, Chen X. Impact of Hashimoto's thyroiditis on the tumor microenvironment in papillary thyroid cancer: Insights from single-cell analysis. Frontiers in Endocrinology. 2024 Sep 16;15:1339473.

[9] Fruhauf F, Cibula D, Kocian R, Zikan M, Dundr P, Jarkovsky J, Fischerova D. Diagnostic performance of ultrasonography in pre-operative assessment of lymph nodes in patients with cervical cancer. International Journal of Gynecological Cancer. 2024 Jul 1;34(7):985-92.

[10] Paschke R, Cantara S, Crescenzi A, Jarzab B, Musholt TJ, Sobrinho Simoes M. European thyroid association guidelines regarding thyroid nodule molecular fine-needle aspiration cytology diagnostics. European thyroid journal. 2017 May 19;6(3):115-29.

[11] Kang X, Liu X, Li Y, Yuan W, Xu Y, Yan H. Development and evaluation of nomograms and risk stratification systems to predict the overall survival and cancer-specific survival of patients with hepatocellular carcinoma. Clinical and Experimental Medicine. 2024 Feb 28;24(1):44.

[12] Biswas N, Chakrabarti S. Artificial intelligence (AI)-based systems biology approaches in multi-omics data analysis of cancer. Frontiers in oncology. 2020 Oct 14;10:588221.

[13] Zelli V, Manno A, Compagnoni C, Ibraheem RO, Zazzeroni F, Alesse E, Rossi F, Arbib C, Tessitore A. Classification of tumor types using XGBoost machine learning model: a vector space transformation of genomic alterations. Journal of Translational Medicine. 2023 Nov 21;21(1):836.

[14] Nazir A, Hussain A, Singh M, Assad A. Deep learning in medicine: advancing healthcare with intelligent solutions and the future of holography imaging in early diagnosis. Multimedia Tools and Applications. 2025 May;84(17):17677-740.

[15] Feng Y, Huang C, Liu X, Xiao Z, Wang L, Lu Z, Ming J. A bibliometric analysis of inflammation and thyroid carcinoma: research trends and future perspectives. Frontiers in Oncology. 2025 Aug 1;15:1593006.

[16] Wang X, Kattan MW. Cohort studies: design, analysis, and reporting. Chest. 2020 Jul 1;158(1):S72-8.

[17] Wu Y, Howarth M, Zhou C, Hu M, Cong W. Reporting of ethical approval and informed consent in clinical research published in leading nursing journals: a retrospective observational study. BMC Medical Ethics. 2019 Dec 5;20(1):94.

[18] Ahmed U, Jiangbin Z, Almogren A, Sadiq M, Rehman AU, Sadiq MT, Choi J. Hybrid bagging and boosting with BBSbased feature selection for enhanced predictive modeling in intrusion detection systems. Scientific Reports. 2024 Dec 17;14(1):30532.

[19] Wang Z, Wu Z, Cheng L, Huang Q, Zhang J, Ren Y, Huang J, Shen Y. An LNM-Associated Gene Signature for Prognostic Prediction and Immune Profiling in Head and Neck Squamous Cell Carcinoma. Cancer Biotherapy & Radiopharmaceuticals. 2025 Jun 1;40(5):339-51.

[20] Beletsky A, Lu Y, Patel BH, Chahla J, Cvetanovich GL, Forsythe B, Cole BJ, Verma N. Should we question the external validity of database studies? A comparative analysis of demographics. Arthroscopy: The Journal of Arthroscopic & Related Surgery. 2019 Sep 1;35(9):2686-94.

[21] Bouke MA, Abdullah A. An empirical study of pattern leakage impact during data preprocessing on machine learning-based intrusion detection models reliability. Expert Systems with Applications. 2023 Nov 15;230:120715.

[22] Aidoni A, Kofidis K, Cocianu CL, Avram L. Deep Learning Models for Natural Gas Demand forecasting: A comparative study of MLP, CNN, and LSTM. Romanian Journal of Petroleum & Gas Technology. 2023 Jan 1(1).

[23] Hairani H, Widiyaningtyas T, Prasetya DD. Addressing class imbalance of health data: a systematic literature review on modified synthetic minority oversampling technique (SMOTE) strategies. JOIV: International Journal on Informatics Visualization. 2024 Sep 30;8(3):1310-8.

[24] Kost S, Rheinbach O, Schaeben H. Using logistic regression model selection towards interpretable machine learning in mineral prospectivity modeling. Geochemistry. 2021 Nov 1;81(4):125826.

[25] Javid AM, Das S, Skoglund M, Chatterjee S. A ReLU dense layer to improve the performance of neural networks. InICASSP 2021-2021 IEEE International Conference on Acoustics, Speech and Signal Processing (ICASSP) 2021 Jun 6 (pp. 2810-2814). IEEE.

[26] Sadr H, Nazari M, Khodaverdian Z, Farzan R, Yousefzadeh-Chabok S, Ashoobi MT, Hemmati H, Hendi A, Ashraf A, Pedram MM, Hasannejad-Bibalan M. Unveiling the potential of artificial intelligence in revolutionizing disease diagnosis and prediction: a comprehensive review of machine learning and deep learning approaches. European Journal of Medical Research. 2025 May 26;30(1):418.

[27] Haq IU. An overview of deep learning in medical imaging. arXiv preprint arXiv:2202.08546. 2022 Feb 17.

[28] Shan S, Ni H, Chen G, Lin X, Li J. A machine learning framework for enhancing short-term water demand forecasting using attention-biLSTM networks integrated with XGBoost residual correction. Water. 2023 Oct 15;15(20):3605.

[29] Majka M. ARIMAX Time Series Forecasting with External Variables. ResearchGate (Sept. 2024). url: https://www. researchgate. net/publication/384196976_ARIMAX_Time_Series_Forecasting_with_External _ Variables.

[30] Sivakumar M, Parthasarathy S, Padmapriya T. Trade-off between training and testing ratio in machine learning for medical image processing. PeerJ Computer Science. 2024 Sep 6;10:e2245.
[31] Kerrigan G, Smyth P, Steyvers M. Combining human predictions with model probabilities via confusion matrices and calibration. Advances in Neural Information Processing Systems. 2021 Dec 6;34:4421-34.
[32] Singh H, Mhasawade V, Chunara R. Generalizability challenges of mortality risk prediction models: A retrospective analysis on a multi-center database. PLOS digital health. 2022 Apr 5;1(4):e0000023.
[33] Tang Q, Pan W, Peng L. Association between Hashimoto thyroiditis and clinical outcomes of papillary thyroid carcinoma: A meta-analysis. PloS one. 2022 Jun 16;17(6):e0269995.
[34] Franzese M, Iuliano A. Correlation analysis. InEncyclopedia of bioinformatics and computational biology: ABC of bioinformatics 2018 (Vol. 1, pp. 706-721). Elsevier.
[35] Cai Y, Li X, Zhang Y, Li J, Zhu F, Rao L. Multimodal sentiment analysis based on multi-layer feature fusion and multi-task learning. Scientific Reports. 2025 Jan 16;15(1):2126.
[36] Yu AC, Mohajer B, Eng J. External validation of deep learning algorithms for radiologic diagnosis: a systematic review. Radiology: Artificial Intelligence. 2022 May 4;4(3):e210064.
[37] Ataş PK. A novel hybrid model to predict concomitant diseases for Hashimoto's thyroiditis. BMC bioinformatics. 2023 Aug 24;24(1):319.
[38] Lord AR, Simms LA, Brown A, Hanigan K, Krishnaprasad K, Schouten B, Croft AR, Appleyard MN, Radford-Smith GL. Development and evaluation of a risk assessment tool to improve clinical triage accuracy for colonoscopic investigations. BMC cancer. 2018 Feb 27;18(1):229.
[39] Chou R, Cuevas C, Fu R, Devine B, Wasson N, Ginsburg A, Zakher B, Pappas M, Graham E, Sullivan SD. Imaging techniques for the diagnosis of hepatocellular carcinoma: a systematic review and meta-analysis. Annals of internal medicine. 2015 May 19;162(10):697-711.
[40] Huang H, Liu Y, Ni S, Liu S. A prediction model for identifying high-risk lymph node metastasis in clinical low-risk papillary thyroid microcarcinoma. BMC Endocrine Disorders. 2023 Nov 27;23(1):260.
[41] Giannini R, Moretti S, Ugolini C, Macerola E, Menicali E, Nucci N, Morelli S, Colella R, Mandarano M, Sidoni A, Panfili M. Immune profiling of thyroid carcinomas suggests the existence of two major phenotypes: an ATC-like and a PDTC-like. The Journal of Clinical Endocrinology & Metabolism. 2019 Aug;104(8):3557-75.
[42] Freiesleben T, Grote T. Beyond generalization: a theory of robustness in machine learning. Synthese. 2023 Sep 27;202(4):109.
[43] Uddin KM, Bhuiyan MT, Saad MN, Islam A, Islam MM. Ensemble machine learning-based approach to predict cervical cancer with hyperparameter tuning and model explainability. Biomedical Materials & Devices. 2025 Jan 10:1-28.
[44] Han H, Sun H, Zhou C, Wei L, Xu L, Shen D, Hu W. Development and validation of a machine learning model for central compartmental lymph node metastasis in solitary papillary thyroid microcarcinoma via ultrasound imaging features and clinical parameters. BMC Medical Imaging. 2025 Jul 1;25(1):228.
[45] Wu Y, Rao K, Liu J, Han C, Gong L, Chong Y, Liu Z, Xu X. Machine learning algorithms for the prediction of central lymph node metastasis in patients with papillary thyroid cancer. Frontiers in endocrinology. 2020 Oct 21;11:577537.
[46] Albaradei S, Thafar M, Alsaedi A, Van Neste C, Gojobori T, Essack M, Gao X. Machine learning and deep learning methods that use omics data for metastasis prediction. Computational and structural biotechnology journal. 2021 Jan 1;19:5008-18.
[47] Ahmed SF, Alam MS, Hassan M, Rozbu MR, Ishtiak T, Rafa N, Mofijur M, Shawkat Ali AB, Gandomi AH. Deep learning modelling techniques: current progress, applications, advantages, and challenges. Artificial Intelligence Review. 2023 Nov;56(11):13521-617.
[48] Adnan M, Kalra S, Tizhoosh HR. Representation learning of histopathology images using graph neural networks. InProceedings of the IEEE/CVF conference on computer vision and pattern recognition workshops 2020 (pp. 988-989).