



BMJ Open Machine learning model for differentiating malignant from benign thyroid nodules based on the thyroid function data

Fuqiang Ma* ^{1,2}, Fengchang Yu*,³ Shenhui Lv*,⁴ Lihua Zhang,⁵ Zhilin Lu,² Quan Zhou,² He-Rong Mao,² Lele Zhang,¹ Nan Xiang ⁴

To cite: Ma* F, Yu* F, Lv* S, et al. Machine learning model for differentiating malignant from benign thyroid nodules based on the thyroid function data. *BMJ Open* 2025;**15**:e093466. doi:10.1136/bmjopen-2024-093466

► Prepublication history for this paper is available online. To view these files, please visit the journal online (<https://doi.org/10.1136/bmjopen-2024-093466>).

FM*, FY* and SL* contributed equally.

Received 07 September 2024
Accepted 22 April 2025



© Author(s) (or their employer(s)) 2025. Re-use permitted under CC BY-NC. No commercial re-use. See rights and permissions. Published by BMJ Group.

For numbered affiliations see end of article.

Correspondence to
Nan Xiang;
2727@hbucom.edu.cn,
Lele Zhang;
876758328@qq.com and
He-Rong Mao;
mhr1752@hbucom.edu.cn

ABSTRACT

Objectives To develop and validate a machine learning (ML) model to differentiate malignant from benign thyroid nodules (TNs) based on the routine data and provide diagnostic assistance for medical professionals.

Setting A qualified panel of 1649 patients with TNs from one hospital were stratified by gender, age, free triiodothyronine (FT3), free thyroxine (FT4) and thyroid peroxidase antibody (TPOAB).

Participants Thyroid function (TF) data of 1649 patients with TNs were collected in a single centre from January 2018 to June 2022, with a total of 273 males and 1376 females, respectively.

Measures Seven popular ML models (Random Forest, Decision Tree, Logistic Regression (LR), K-Neighbours, Gaussian Naive Bayes, Multilayer Perception and Gradient Boosting) were developed to predict malignant and benign TNs, whose performance indicators included area under the curve (AUC), accuracy, recall, precision and F1 score.

Results A total of 1649 patients were enrolled in this study, with the median age of 45.15±13.41 years, and the male to female ratio was 1:5.055. In the multivariate LR analysis, statistically significant differences existed between the TNs group and thyroid cancer group in gender, age, free triiodothyronine (FT3), free thyroxine (FT4) and TPOAB. Among the seven tested ML models, the best performance was achieved in the Gradient Boosting model in terms of precision, AUC, accuracy, recall and F1 score, with the AUC of 0.82, accuracy of 79.4% and precision of 0.814 after experimental verification. FT4, TPOAB and FT3 were validated as the top three features in the Gradient Boosting model.

Conclusions This study innovatively developed a predictive model for benign and malignant TNs based on the Gradient Boosting Decision Tree algorithm. For the first time, it validated the clinical predictive value of TF parameters (FT4, FT3) and TPOAB as key biomarkers.

BACKGROUND

Thyroid nodules (TNs), as one of the commonly found disorders in endocrinology, are mainly triggered by local abnormal growth of thyroid cells¹ and have become a clinical consensus for early and timely differentiation of malignant TNs. Varying according to

STRENGTHS AND LIMITATIONS OF THIS STUDY

- ⇒ Using common thyroid function data, this study employed seven popular machine learning (ML) models and provided a comparison of multiple prediction methods.
- ⇒ This study takes into consideration hyperparameters to optimise the performance of ML models.
- ⇒ This study employed a retrospective analysis solely based on the single-centre data, which may introduce potential selection bias.
- ⇒ The model lacks a graphical user interface.

countries and regions, it is estimated that the prevalence of TNs in the general population is higher than 67%,^{2,3} with a palpation detection rate of 4%–7%⁴ and an ultrasound (US) detection rate of 19%–68%.¹ Probably cancerated from the severe TNs, thyroid cancer (TC) is predicted to become the fourth most common type of cancer worldwide.⁵ Overall, China, the USA and India have witnessed the highest incidence of TC in the world,^{5,6} which has posed tremendous medical and financial burden on society, with the cost of well-identified TC care in the USA only having reached \$3.5 billion.⁷ To date, differentiating malignant from benign nodules has played a crucial role in the management of thyroid diseases, and early identification of malignant TNs is still regarded as a clinical consensus.⁸ Therefore, it is quite necessary for clinicians and researchers to devote sufficient energy to optimising the diagnostic and therapeutic process to screen malignant TNs.

Currently, US, fine needle aspiration biopsy (FNAB) examination, frozen section (FS), molecular detection, surgery⁹ or these in combination are often utilised to screen benign and malignant TNs. However, the limitations of the above-mentioned technologies can be clearly seen. For instance, US

diagnosis of thyroid diseases can be influenced by the clinical experience of physicians,¹⁰ while the clinical value and necessity of FNAB and FS are still of controversy as they are still sophisticated and time-consuming.^{11 12} In addition, molecular means are still not easily accessible so far in many countries like in China,¹³ which necessitates the development of a diagnostic tool that can collect data from existing diagnostic tests to improve the predictability of benign and malignant risk of TNs.¹⁴

As artificial intelligence (AI) has been widely utilised in analysing the healthcare data,¹⁵ machine learning (ML) models are adopted today to predict various diseases by learning the embedded relationships between features. Researchers worldwide have reported a variety of AI applications in the field of thyroid,¹⁶ but most of them focus on the prediction of TNs using US, radiology and genomics data.^{17–20} The ML algorithms, however, have not been adequately used to predict thyroid function (TF)-based malignant TNs through numerical data with the most common, readily available and cost-effective characteristics in today's clinical practice.

Some malignant TNs may not affect TF as TF indicators are subject to the low sensitivity to malignant TNs. It may also be associated with the accuracy of prediction results that can be easily affected by a variety of factors such as disease and medication. However, an increasing number of studies have shown that TF is significantly correlated with TC.²¹ For some patients, the chance of success for preferred treatment protocols would be greatly impacted.²² Other studies suggest that patients would feel more satisfied when treatments are given with consideration of their preferences.²³

This study aims to investigate the utility of numerical TF data obtained during initial consultations of patients with TNs to train an optimised ML model to predict the occurrence of malignant TNs. The proposed predictive tool is designed to assist clinical decision-making by providing patients with valuable prognostic information, potentially enhancing the successful treatment rates through timely medical intervention and facilitating informed shared decision-making regarding the optimal treatment strategies in the current medical management system.²⁴ Participants' consents are unnecessary, since the data we utilised in this study were anonymous and all patients' information was deidentified at the point of collection.

METHODS

Participants

A total of 1649 cases with TNs were collected from January 2018 to June 2022 in the Endocrinology Department of Hubei Provincial Hospital of Traditional Chinese Medicine. Among them, 1096 patients (150 males and 946 females) were diagnosed with benign TNs, with the median age of 45.31±13.34 years; while the other 553 (123 males and 430 females) were pathologically confirmed with TC, with the median age of 46.35±14.08 years.

Inclusion criteria

The included cases are those conforming to TNs and TC diagnostic criteria designated in the 2015 American Thyroid Association Management Guidelines for Adult Patients with Thyroid Nodules and Differentiated Thyroid Cancer.⁹

Exclusion criteria

The excluded cases are the pregnant and lactating patients, and those with acute infections, or with immune system diseases, or with neurological diseases, and without follow-up information. This study strictly observed the Helsinki Declaration, and all patients' information was kept confidential.

Clinical data collection

Clinical data of patients were collected by well-trained professional medical workers from the Information Department of Hubei Provincial Hospital of Traditional Chinese Medicine, including demographic information (gender and age), TF profiles (five biochemical parameters), thyroid US findings, and surgical pathology reports.

Patient and public involvement statement

None.

Data processing

Data validity assessment

The original data were cross-checked by two workers, and researchers then randomly selected 5% of the processed data for comparison.

Data processing and cleansing

In the classification framework, patients with pathologically confirmed TC were designated as positive cases, whereas those with TNs but without malignant tumours were classified as negative ones. Regarding the evaluation of model performance, when multiple area under the curve (AUC) values were generated from a single predictive model, specifically, multiple test results derived from the same model within a specific study, priority was given to selecting the predictive occurrence data that demonstrated optimal model performance.

Machine learning (ML) training

The seven types of ML algorithms employed for data modelling are Random Forest, Decision Tree, Logistic Regression (LR), K-Neighbours, Gaussian Naive Bayes (GNB), Multilayer Perception (MLP) and Gradient Boosting. Random Forest (RF) is an integrated learning method that performs classification or regression tasks by constructing multiple Decision Trees and aggregating their outputs to improve accuracy and reduce overfitting. Decision Tree is an interpretable classification and regression method that partitions data into subsets based on feature values through a series of binary decisions, typically represented as 'yes' or 'no' questions, to assign data points to specific categories or predicted numerical values. LR is a statistical method used for binary or

multiclass classification problems, estimating probabilities by applying a logistic function to maximise the likelihood of observed data. k-nearest neighbours (k-NN) is a non-parametric classification and regression method that predicts outcomes by measuring the similarity between instances based on distance metrics, such as Euclidean distance, and assigning labels based on the majority vote of its k-NN. GNB is a probabilistic classifier based on Bayes' theorem with the assumption of feature independence, which simplifies computations while maintaining reasonable performance for many applications. MLP is a type of feedforward artificial neural network consisting of an input layer, one or more hidden layers and an output layer, where information propagates through fully connected layers to learn complex patterns in the data. Gradient Boosting Trees is an ensemble learning algorithm that iteratively adds weak prediction models, typically Decision Trees, to minimise a loss function and enhance predictive performance. To evaluate the predictability of each model, the receiver operating characteristic (ROC) curve was plotted to reflect the model's performance,²⁵ while accuracy, AUC, precision, recall and F1 score were used to determine the predictability in comparison with the performance of ML algorithms. Then, to eliminate the impact of data division on the model performance, five-fold Cross-Validation was used for the ML model in the experiment.²⁶ Finally, for different ML models, the importance of each feature in the models was further explored.²⁷

To optimise the performance of ML models, hyperparameters should be taken into consideration. In our study, the employed ML algorithms have their own specific hyperparameters, which are detailed as follows:

For the RF algorithm, 100 trees with the Gini criterion were used. The minimum number of samples required to split an internal node was set to 2, and the minimum number of samples required to be at a leaf node was set to

1. For the Decision Tree, the Gini criterion was also used to select the best split strategy at each node. The minimum number of samples required to split an internal node was 2. The LR model was configured with an L2 norm penalty and a tolerance for stopping criteria set to 1e-4. The K Neighbours algorithm was set with five neighbours, with equal weight of all points in each neighbourhood, and the Euclidean distance (L2) was used as the Minkowski Distance. For the GNB, the portion of the largest variance of all features added to variances for calculation stability was set to 1e-9. The MLP model used the ReLU activation function for the hidden layer, the Adam optimizer, an L2 regularisation strength of 0.0001 and a constant learning rate of 0.001. The Gradient Boosting algorithm utilised the binomial and multinomial deviance loss functions, with a learning rate of 0.1 and 100 boosting stages. All these algorithms were implemented using the Scikit-Learn library in Python.

Statistical analysis

The number of cases (%) was adopted for describing the count data, the χ^2 test for comparison between groups, the t-test for measuring data accorded with a normal distribution and the 'mean \pm SD' (X \pm S) for expressing results. The median (M) and IQR were exploited for those not conforming to a normal distribution, and the Mann-Whitney U test for comparison between groups. The LR analysis was applied for variables with statistical differences in univariate analysis, so as to further screen variables. The influence of each prognostic factor was judged by calculating the OR of each variable. The two-tailed test was utilised to examine all statistical tests, and $p < 0.05$ was considered statistically significant.

Table 1 Baseline distribution of 1649 patients

Variables	Number of cases (1649)	Benign nodules (1096)	Malignant nodules (553)	t/z/x ²	95% CI	P value
Age	44.72 \pm 12.99	45.58 \pm 13.85	43.54 \pm 12.21	-2.628	-3.106 to -0.451	0.009
≤ 50	1067 (64.7%)	688 (64.5%)	379 (35.5%)	5.849	1.052 to 1.624	0.016
50	582 (35.3%)	408 (70.1%)	174 (29.9%)			
Male	273 (16.6%)	150 (13.7%)	123 (22.2%)	19.478	0.426 to 0.722	0.000
Female	1376 (83.4%)	946 (86.3%)	430 (77.8%)			
FT3	4.56 (0.80)	4.58 (0.67)	4.49 (1.64)	-5.305	-0.29 to 0.12	0.000
FT4	15.29 (3.81)	15.47 (3.35)	14.8 (15.58)	-5.782	-1.42 to 0.70	0.000
TSH	1.90 (1.98)	1.98 (2.11)	1.81 (1.78)	-2.479	-0.326 to -0.034	0.013
TGAB	48.8 (166.8)	54.0 (201.00)	46.1 (79.48)	-2.238	-15.40 to -1.50	0.025
TPOAB	46.05 (1243.3)	62.0 (1268.7)	37.0 (85.1)	-8.197	-21.60 to -11.10	0.000

FT3, free triiodothyronine; FT4, free thyroxine; TGAB, thyroglobulin antibody; TPOAB, thyroid peroxidase antibody; TSH, third generation thyroid stimulating hormone.

Table 2 Univariate and multivariate LR analysis of the influence on benign and malignant TNs

Variables	Control group	Univariate LR		Multivariate LR	
		95% CI	P value	95% CI	P value
Age	Thyroid nodule	0.982 to 0.997	0.009	0.975 to 0.992	0.000
Gender	Female	1.390 to 2.355	0.000	1.478 to 2.665	0.000
FT3	Thyroid nodule	0.490 to 0.650	0.000	0.662 to 0.965	0.020
FT4	Thyroid nodule	0.863 to 0.903	0.000	0.871 to 0.926	0.000
TSH	Thyroid nodule	0.975 to 1.012	0.488		
TGAB	Thyroid nodule	0.998 to 0.999	0.000	0.999 to 1.000	0.139
TPOAB	Thyroid nodule	0.999 to 0.999	0.000	0.999 to 0.999	0.000

FT3, free triiodothyronine; FT4, free thyroxine; LRs, logistic regression; TGAB, thyroglobulin antibody; TNs, thyroid nodules; TPOAB, thyroid peroxidase antibody; TSH, thyroid stimulating hormone.

RESULTS

Demographic characteristics

A total of 1649 patients participated in the study, of which 553 suffered from TC confirmed with postsurgery pathological findings, and 1096 were randomly selected with benign TNs. Ranging from 9 to 89 years, the median age of patients was 45.15 ± 13.41 years, with the male to female ratio of 1:5.055. Details are shown in [table 1](#).

Univariate and multivariate LR analysis of benign and malignant TNs

In the univariate LR analysis, statistically significant differences existed between the two groups in age, gender, FT3, FT4, TGAB and thyroid peroxidase antibody (TPOAB) ($p < 0.05$). In the multivariate LR analysis, there were statistically significant differences between the two groups in age, gender, FT3, FT4 and TPOAB ($p < 0.05$). Details are shown in [table 2](#).

Development and validation of the predictive model

Seven distinct ML algorithms were constructed to develop predictive models, with precision scores ranging from 0.571 to 0.814. Comparative analysis revealed that the Gradient Boosting model outperformed other algorithms across multiple evaluation metrics, including precision, AUC, accuracy, recall and F1 score, as detailed in [table 3](#). Based on its superior performance, the

Gradient Boosting algorithm was consequently selected as the optimal predictive model. The corresponding ROC curves, illustrating the performance characteristics of all developed models, are presented in [figure 1](#).

Relative importance of features in the ML algorithms

[Figure 2](#) illustrates the variable importance profiles across different ML algorithms for distinguishing between benign and malignant TNs. The analysis reveals consistent patterns in feature importance rankings. While minor variations in variable significance were observed among the algorithms, thyroid-related biomarkers including FT4, FT3 and TPOAB consistently demonstrated high predictive importance across all the models. Notably, in the optimal Gradient Boosting model, these three biomarkers maintained their prominence, with FT4, TPOAB and FT3 emerging as the top-ranked features, respectively.

DISCUSSIONS

In this study, seven ML algorithms were established for differentiating benign and malignant TNs using TF profiles from patients with TNs. Through the comprehensive analysis, the Gradient Boosting algorithm was identified as the most robust predictive model, demonstrating

Table 3 Comparative validation of the predictive performance of seven ML algorithms

Model type	Precision	AUC	Accuracy	Recall	F1
K-Neighbours	0.572	0.71	0.700	0.436	0.493
Gradient Boosting	0.814	0.82	0.794	0.499	0.616
Multilayer Perceptron	0.518	0.66	0.671	0.364	0.423
Gaussian Naive Bayes	0.369	0.62	0.434	0.962	0.533
Decision Tree	0.571	0.67	0.711	0.570	0.568
Random Forest	0.771	0.81	0.784	0.506	0.608
Logistic Regression	0.723	0.70	0.727	0.302	0.425

AUC, area under the curve; ML, machine learning.

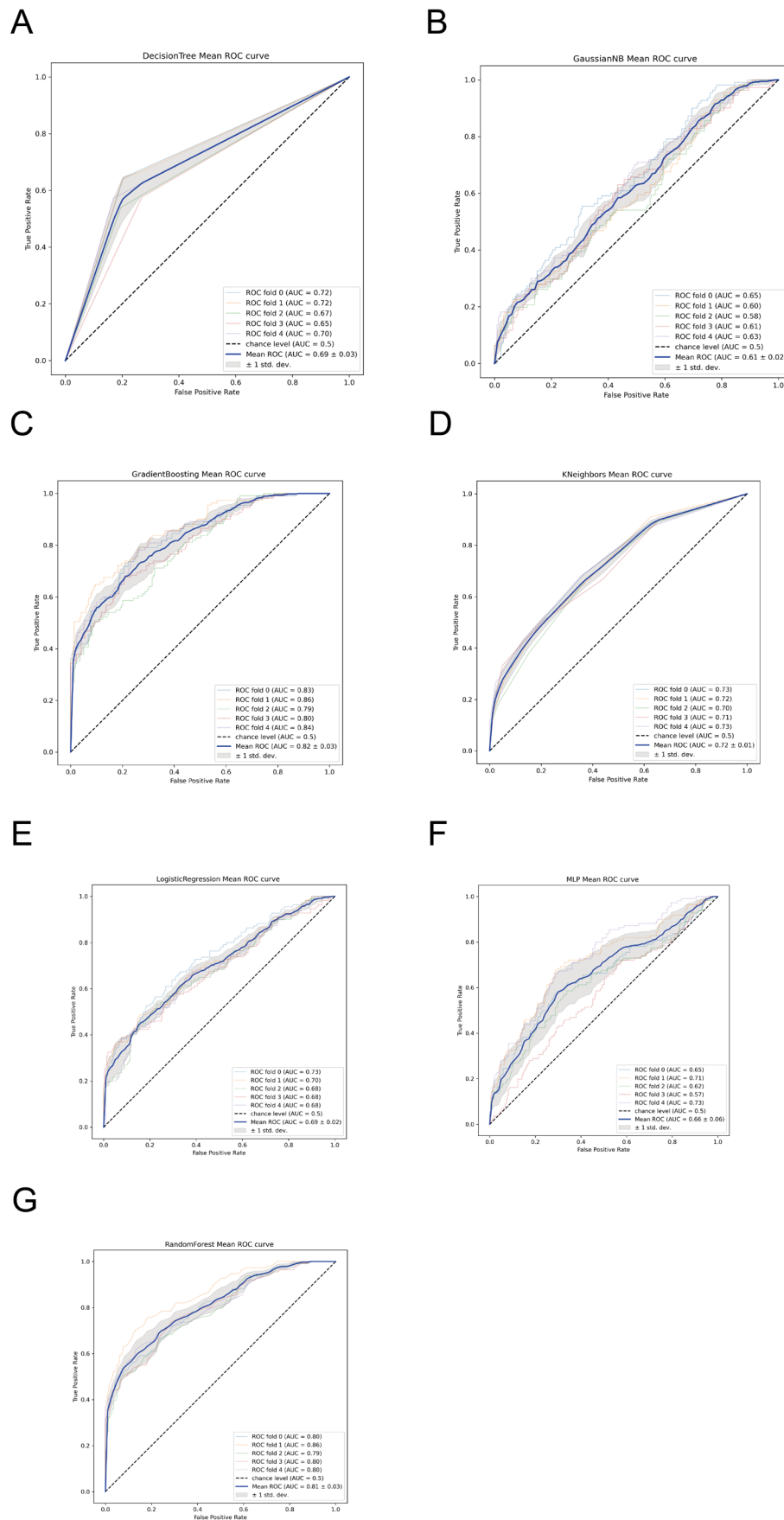


Figure 1 ROC curve of the ML models. (A) Decision Tree; (B) Gaussian Naive Bayes; (C) Gradient Boosting; (D) K-Neighbours; (E) Logistic Regression; (F) MLP; (G) Random Forest. MLP, Multilayer Perception; ROC, receiver operating characteristic.

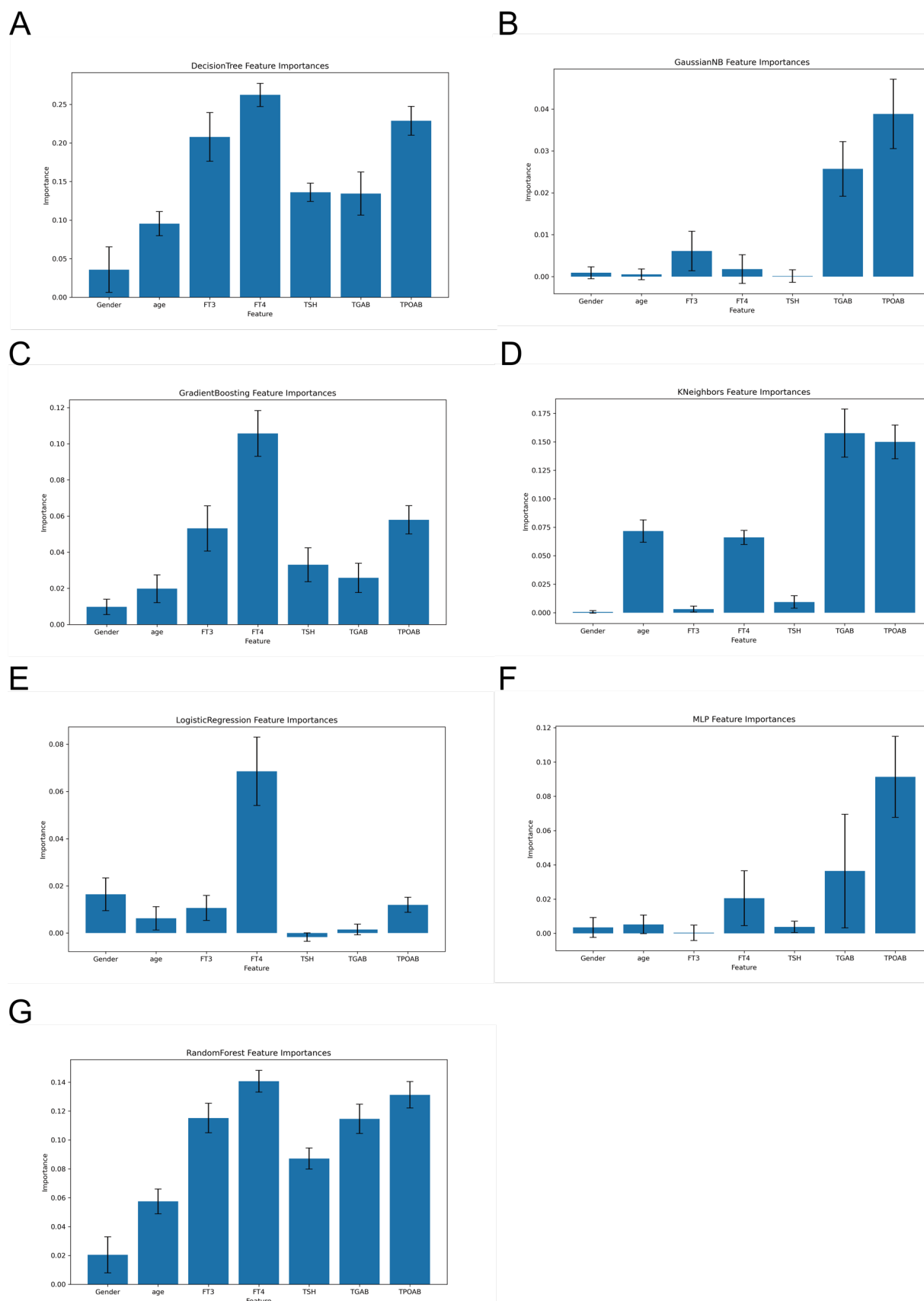


Figure 2 Importance ranking of features in the ML models. (A) Decision Tree; (B) Gaussian Naive Bayes; (C) Gradient Boosting; (D) K-Neighbours; (E) Logistic Regression; (F) MLP; (G) Random Forest. FT3, free triiodothyronine; FT4, free thyroxine; ML, machine learning; MLP, Multilayer Perception; TGAB, thyroglobulin antibody; TPOAB, thyroid peroxidase antibody; TSH, thyroid stimulating hormone.

superior accuracy and reliability in TC prediction. Furthermore, an in-depth feature importance analysis was conducted to elucidate the relative contribution of various clinical parameters within the optimal model.

Previous studies on ML-based prediction of malignant TNs have predominantly focused on imaging and histological characteristics, including morphological features, nodule dimensions and textural patterns as primary predictive indicators.¹⁸ This preference stems from the perceived limited diagnostic utility of TF indexes, which are often considered less reliable due to their susceptibility to various confounding factors. However, emerging evidence has increasingly demonstrated a significant association between TF profiles and TC risk.²⁸ Previous epidemiological studies have shown that thyroid dysfunction is associated with elevated TC risk,²⁹ while specific thyroid hormone levels show distinct correlation patterns: FT3 exhibits an inverse relationship with TC risk,³⁰ and enhanced FT4 levels demonstrate positive correlation with TC incidence. These results actually underscore the clinical relevance of our study, providing valuable insights for enhancing patient counselling and clinical decision-making. By providing more accurate risk stratification, our findings may inform critical decisions regarding therapeutic approaches, potentially guiding the choice between pharmacological intervention and surgical management.³¹

The ML models in this study demonstrated significant predictive capabilities for malignant TNs by leveraging TF data, offering a more direct and objective assessment of patients' conditions. This approach laid a scientifically robust and precise foundation for the clinical diagnosis of thyroid disorders. In countries with a high incidence of TC, such as China and India, predicting the availability and affordability indicators can reduce expenditures on further medical examinations, thereby mitigating the financial burden on society. Given that evidence-based consultations have been shown to yield better outcomes followed by shared decision-making and the application of patients' preferred treatment options,^{23 31} this personalised and patient-oriented predictive model has the potential to significantly enhance treatment efficacy utilising simple parameters.

To minimise the impact of confounding factors on the TF data, this study collected and analysed the TF data from patients' initial visits. Multivariate LR analysis revealed significant differences in age, gender, FT3, FT4 and TPOAB between the two groups, identifying these variables as independent risk factors for both benign and malignant TNs. Previous studies have found that the incidence of TC is age-dependent, with a notably higher prevalence among women compared with men.³² In our study, the TC group exhibited a younger age distribution than the TN group, potentially attributable to earlier detection and intervention in TC cases. While some studies have linked lower levels of FT3 and FT4 to an increased risk of TC,³³ others have found no significant association²⁸ or even reported an inverse correlation between elevated

FT3 levels and TC risk.³⁴ Furthermore, the relationship between FT4 levels and TC appears to be non-linear, with a marked increase in TC risk observed when FT4 levels exceed approximately 2.2 ng/dL.³⁵ Our findings align with certain prior studies, underscoring the intricate relationship between TNs, TF and TC.³⁶ Although TPOAB is primarily utilised for diagnosing Hashimoto's thyroiditis and Graves' disease,³⁷ two fine needle aspiration cytology (FNAC) studies from the same institution have established TPOAB as an independent risk factor for malignant thyroid tumours, highlighting a critical association between TPOAB and TC.³⁸ This conclusion is consistent with the results of our study.

This study demonstrated distinct advantages over previous studies in predicting the risk of malignant TNs. First, while significant differences in the TF data have been observed between patients with benign TNs and TC, few studies have focused on TF as a predictive marker. In contrast, our study leveraged the TF data routinely collected and readily accessible in clinical practice. Furthermore, although ML methods have shown diagnostic potential in distinguishing benign from malignant TNs using genomics, radiological and US data, there are still scarce studies employing ML algorithms with TF as a key feature. To the best of our knowledge, this study is among the first to develop an ML-based predictive model utilising TF data for real-time risk assessment of TNs, thus achieving high levels of accuracy and reliability.

However, this study still has its usual limitations. First, all data in the study were collected from a single centre, and the findings lack validation using external data sets. Second, the retrospective design of this work may introduce selection and information bias, potentially affecting the robustness of the results. Third, the model lacks a graphical user interface. All these factors should be carefully considered in future studies so that the generalisability and precision of the findings can be elevated.

CONCLUSIONS

This study innovatively developed a predictive model for benign and malignant TNs based on the gradient boosting decision tree algorithm. For the first time, it validated the clinical predictive value of TF parameters (FT4, FT3) and TPOAB as key biomarkers. By leveraging ML interpretability techniques, the dose-response relationship between these indicators and the malignant risk of nodules was elucidated, providing a quantitative basis for early TC screening. This study will facilitate the transition of TN diagnosis and treatment from static assessment to a dynamic intelligent decision-making framework, offering a novel paradigm for the application of precision medicine in endocrine tumour management.

Author affiliations

¹Department of Integrated Traditional and Western Medicine, The First Affiliated Hospital, and College of Clinical Medicine of Henan University of Science and Technology, Luoyang, China

²Hubei University of Chinese Medicine, Wuhan, Hubei, China

³Wuhan University, School of Information Management, Wuhan, China

⁴Hubei Provincial Hospital of Traditional Chinese Medicine, Wuhan, Hubei, China

⁵Huanggang Hospital of Chinese Medicine, Hubei University of Chinese Medicine, Huanggang, Hubei, China

Acknowledgements We express our heartfelt thanks to Director Qing Liu and Director Wei Yan from Hubei Hospital of Traditional Chinese Medicine for their constructive suggestions in the writing of this paper.

Contributors NX, LLZ and H-RM conceptualised and designed the study and NX has full access to all the data used in the study; FQM and H-RM wrote the article, FCY did machine learning and SHL did statistical analysis. LHZ, ZLL and QZ collected and analysed the clinical data. All authors had revised and agreed with publication before the manuscript submission. FQM, FCY and SHL contributed equally to this work. NX is responsible for the overall content as guarantor.

Funding The authors have not declared a specific grant for this research from any funding agency in the public, commercial or not-for-profit sectors.

Competing interests None declared.

Patient and public involvement Patients and/or the public were not involved in the design, or conduct, or reporting or dissemination plans of this research.

Patient consent for publication Not applicable.

Ethics approval This study had been reviewed and approved by the Ethics Committee of Hubei Provincial Hospital of Traditional Chinese Medicine (approval no: HBZY2023-C27-02). All methods were adopted in line with relevant guidelines and regulations.

Provenance and peer review Not commissioned; externally peer reviewed.

Data availability statement Data are available upon reasonable request. The data sets used and analysed in this study are available from the corresponding author NX on reasonable request.

Open access This is an open access article distributed in accordance with the Creative Commons Attribution Non Commercial (CC BY-NC 4.0) license, which permits others to distribute, remix, adapt, build upon this work non-commercially, and license their derivative works on different terms, provided the original work is properly cited, appropriate credit is given, any changes made indicated, and the use is non-commercial. See: <http://creativecommons.org/licenses/by-nc/4.0/>.

ORCID iDs

Fuqiang Ma* <http://orcid.org/0000-0002-7792-7432>

Nan Xiang <http://orcid.org/0000-0003-0347-1580>

REFERENCES

- Durante C, Grani G, Lamartina L, et al. The Diagnosis and Management of Thyroid Nodules: A Review. *JAMA* 2018;319:914–24.
- Huang Y, Li Z, Yang K, et al. The association of uric acid with the development of thyroid nodules: a retrospective cohort study. *BMC Endocr Disord* 2022;22:197.
- Uppal N, Collins R, James B. Thyroid nodules: Global, economic, and personal burdens. *Front Endocrinol (Lausanne)* 2023;14:1113977.
- Wilhelm S. Evaluation of thyroid incidentaloma. *Surg Clin North Am* 2014;94:485–97.
- Kim J, Gosnell JE, Roman SA. Geographic influences in the global rise of thyroid cancer. *Nat Rev Endocrinol* 2020;16:17–29.
- Schuster-Bruce J, Jani C, Goodall R, et al. A Comparison of the Burden of Thyroid Cancer Among the European Union 15+ Countries, 1990–2019 Estimates From the Global Burden of Disease Study. *Jama Otolaryngol* 2022;148:350–9.
- Lubitz CC, Kong CY, McMahon PM, et al. Annual financial impact of well-differentiated thyroid cancer care in the United States. *Cancer* 2014;120:1345–52.
- Sorrenti S, Dolcetti V, Radzina M, et al. Artificial Intelligence for Thyroid Nodule Characterization: Where Are We Standing? *Cancers (Basel)* 2022;14:3357.
- Haugen BR, Alexander EK, Bible KC, 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;26:1–133.
- Lu WW, Zhang D, Ni XJ. A Review of the Role of Ultrasound Radiomics and Its Application and Limitations in the Investigation of Thyroid Disease. *Med Sci Monit* 2022;28:e937738.
- Nie W, Zhu L, Yan P, et al. Thyroid nodule ultrasound accuracy in predicting thyroid malignancy based on TIRADS system. *Adv Clin Exp Med* 2022;31:597–606.
- Huang J, Luo JL, Chen JS, et al. Intraoperative frozen section can be reduced in thyroid nodules classified as Bethesda categories V and VI. *Sci Rep* 2017;7:5244.
- Mao ZC, Ding YF, Wen LP, et al. Combined fine-needle aspiration and selective intraoperative frozen section to optimize prediction of malignant thyroid nodules: A retrospective cohort study of more than 3000 patients. *Front Endocrinol* 2023;14:14.
- Fresilli D, Grani G, De Pascali ML, et al. Computer-aided diagnostic system for thyroid nodule sonographic evaluation outperforms the specificity of less experienced examiners. *J Ultrasound* 2020;23:169–74.
- Wu Y, Rao K, Liu J, et al. Machine Learning Algorithms for the Prediction of Central Lymph Node Metastasis in Patients With Papillary Thyroid Cancer. *Front Endocrinol (Lausanne)* 2020;11:577537.
- Wildman-Tobriner B, Taghi-Zadeh E, Mazurowski MA. Artificial Intelligence (AI) Tools for Thyroid Nodules on Ultrasound, From the AJR Special Series on AI Applications. *AJR Am J Roentgenol* 2022;219:1–8.
- Qin H, Que Q, Lin P, et al. Magnetic resonance imaging (MRI) radiomics of papillary thyroid cancer (PTC): a comparison of predictive performance of multiple classifiers modeling to identify cervical lymph node metastases before surgery. *Radiol Med* 2021;126:1312–27.
- He J, Zhang H, Wang X, et al. A pilot study of radiomics signature based on biparametric MRI for preoperative prediction of extrathyroidal extension in papillary thyroid carcinoma. *J Xray Sci Technol* 2021;29:171–83.
- Yang F, Zhang J, Li BK, et al. Identification of Potential lncRNAs and miRNAs as Diagnostic Biomarkers for Papillary Thyroid Carcinoma Based on Machine Learning. *Int J Endocrinol* 2021;2021:3984463.
- Zhang H, Zhang ZZ, Liu XD, et al. DNA Methylation Haplotype Block Markers Efficiently Discriminate Follicular Thyroid Carcinoma from Follicular Adenoma. *J Clin Endocrinol Metab* 2021;106:1011–21.
- Huang H, Rusiecki J, Zhao N, et al. Thyroid-Stimulating Hormone, Thyroid Hormones, and Risk of Papillary Thyroid Cancer: A Nested Case-Control Study. *Cancer Epidemiol Biomarkers Prev* 2017;26:1209–18.
- Hentzen JEK, Verschuur MA, Lemmers M, et al. Factors influencing women's preferences for subsequent management in the event of incomplete evacuation of the uterus after misoprostol treatment for miscarriage. *Hum Reprod* 2017;32:1674–83.
- Wallace RR, Goodman S, Freedman LR, et al. Counseling women with early pregnancy failure: utilizing evidence, preserving preference. *Patient Educ Couns* 2010;81:454–61.
- Charon R. The patient-physician relationship. Narrative medicine: a model for empathy, reflection, profession, and trust. *JAMA* 2001;286:1897–902.
- Rizzo V, Salmasi MY, Sabetai M, et al. Infective endocarditis: Do we have an effective risk score model? A systematic review. *Front Cardiovasc Med* 2023;10:1093363.
- Zhu J, Zheng J, Li L, 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 (Lausanne)* 2021;8:635771.
- Xu Q, Xie W, Liao B, et al. Interpretability of Clinical Decision Support Systems Based on Artificial Intelligence from Technological and Medical Perspective: A Systematic Review. *J Healthc Eng* 2023;2023:9919269.
- Rinaldi S, Plummer M, Biessy C, et al. Thyroid-stimulating hormone, thyroglobulin, and thyroid hormones and risk of differentiated thyroid carcinoma: the EPIC study. *J Natl Cancer Inst* 2014;106:dju097.
- Liang L, Zheng XC, Hu MJ, et al. Association of benign thyroid diseases with thyroid cancer risk: a meta-analysis of prospective observational studies. *J Endocrinol Invest* 2019;42:673–85.
- Hu M, Zhang C, Liang L, et al. Fasting serum glucose, thyroid-stimulating hormone, and thyroid hormones and risk of papillary thyroid cancer: A case-control study. *Head Neck* 2019;41:2277–84.
- Hamel CC, Vart P, Vandenbussche FPHA, et al. Predicting the likelihood of successful medical treatment of early pregnancy loss: development and internal validation of a clinical prediction model. *Hum Reprod* 2022;37:936–46.
- Kwong N, Medici M, Angell TE, et al. The Influence of Patient Age on Thyroid Nodule Formation, Multinodularity, and Thyroid Cancer Risk. *J Clin Endocrinol Metab* 2015;100:4434–40.
- Gul K, Ozdemir D, Dirikoc A, et al. Are endogenously lower serum thyroid hormones new predictors for thyroid malignancy in addition to higher serum thyrotropin? *Endocrine* 2010;37:253–60.

- 34 Hu MJ, Zhang C, Liang L, *et al.* Fasting serum glucose, thyroid-stimulating hormone, and thyroid hormones and risk of papillary thyroid cancer: A case-control study. *Head Neck* 2019;41:2277–84.
- 35 Wang Z, Lin YX, Jiang YX, *et al.* The associations between thyroid-related hormones and the risk of thyroid cancer: An overall and dose-response meta-analysis. *Front Endocrinol* 2022;13:13.
- 36 Zafón C, Obiols G, Mesa J. Preoperative TSH level and risk of thyroid cancer in patients with nodular thyroid disease: nodule size contribution. *Endocrinol Nutr* 2015;62:24–8.
- 37 Soh SB, Aw TC. Laboratory Testing in Thyroid Conditions - Pitfalls and Clinical Utility. *Ann Lab Med* 2019;39:3–14.
- 38 Boi F, Minerba L, Lai ML, *et al.* Both thyroid autoimmunity and increased serum TSH are independent risk factors for malignancy in patients with thyroid nodules. *J Endocrinol Invest* 2013;36:313–20.