

RESEARCH

Open Access



Prediction of Seronegative Hashimoto's thyroiditis using machine learning models based on ultrasound radiomics: a multicenter study

Wenjun Wu¹, Shengsheng Yao², Daming Liu³, Yuan Luo⁴, Yihan Sun⁵, Ting Ruan¹, Mengyou Liu⁶, Li Shi⁷, Chang Liu⁸, Mingming Xiao⁴, Qi Zhang³, Zhengshuai Liu⁹, Xingai Ju¹⁰, Jiahao Wang⁸, Xiang Fei², Li Lu¹¹, Yang Gao², Ying Zhang², Liying Gong¹², Xuanyu Chen¹³, Wanli Zheng¹³, Xiali Niu¹⁰, Xiao Yang⁴, Huimei Cao⁷, Shijie Chang^{8*}, Jianchun Cui^{2*} and Zuoxin Ma^{12*}

Abstract

Background Seronegative Hashimoto's thyroiditis is often underdiagnosed due to the lack of antibody markers. Combining ultrasound radiomics with machine learning offers potential for early detection in patients with normal thyroid function.

Methods Data from 164 patients with single thyroid lesions and normal thyroid function, treated surgically between 2016 and 2024, were retrospectively collected from four hospitals. Radiomics features were extracted from ultrasound images of non-tumorous hypoechoic areas. Pathological lymphocytic infiltration and hypoechoic ratios were evaluated by senior pathologists and ultrasound physicians.

A machine learning model, CCH-NET, was developed using a random forest classifier after feature selection with Least Absolute Shrinkage and Selection Operator (LASSO) regression. The model was trained and tested with an 80:20 split and compared to senior ultrasound physicians.

Results The CCH-NET model achieved a sensitivity of 0.762, specificity of 0.714, and an area under the curve (AUC) of 0.8248, outperforming senior ultrasound physicians (AUC = 0.681). It maintained consistent accuracy across test sets, with F1 scores of 0.778 and 0.720 in Test_1 and Test_2, respectively, and exhibited superior predictive rates.

Conclusion The CCH-NET model enhances accuracy in detecting early Seronegative Hashimoto's thyroiditis over senior ultrasound physicians.

*Correspondence:

Shijie Chang
sjchang@cmu.edu.cn
Jianchun Cui
cjc7162003@aliyun.com
Zuoxin Ma
mzx533@163.com

Full list of author information is available at the end of the article



© The Author(s) 2025. **Open Access** This article is licensed under a Creative Commons Attribution-NonCommercial-NoDerivatives 4.0 International License, which permits any non-commercial use, sharing, distribution and reproduction in any medium or format, as long as you give appropriate credit to the original author(s) and the source, provide a link to the Creative Commons licence, and indicate if you modified the licensed material. You do not have permission under this licence to share adapted material derived from this article or parts of it. The images or other third party material in this article are included in the article's Creative Commons licence, unless indicated otherwise in a credit line to the material. If material is not included in the article's Creative Commons licence and your intended use is not permitted by statutory regulation or exceeds the permitted use, you will need to obtain permission directly from the copyright holder. To view a copy of this licence, visit <http://creativecommons.org/licenses/by-nc-nd/4.0/>.

Ethics No. [2023] H013

Trial registration Chinese Clinical Trial Registry; CTR2400092179; 12 November 2024.

Keywords Radiomics, Hashimoto's thyroiditis (HT), Seronegative Hashimoto thyroiditis, Machine learning, Ultrasound imaging, Hypothyroidism

Introduction

Seronegative Hashimoto's thyroiditis (Seronegative HT) is a distinct subtype of Hashimoto's thyroiditis (HT), characterized by the absence of thyroid antibodies, a more insidious disease course, and lower invasiveness [1]. However, it can still lead to hypothyroidism, frequently resulting in misdiagnosis or underdiagnosis, with a prevalence reported to reach as high as 20.8% [2]. In 2020, L. Croce et al. established the diagnostic criteria for Seronegative HT [2], which include negative TPO-Ab (thyroid peroxidase antibody) and TG-Ab (thyroglobulin antibody), the presence of subclinical or clinical hypothyroidism, and diffuse hypoechogenicity on thyroid ultrasound. Notably, Seronegative HT is often diagnosed after hypothyroidism has already developed, and it accounts for 34.6% of primary hypothyroidism cases. Thus, improving the early diagnostic rate of Seronegative HT during the euthyroid phase and reducing missed diagnoses are critical clinical challenges.

According to L. Croce's criteria, the diagnosis of Seronegative HT during the early euthyroid phase is challenging due to the lack of typical antibody presentation and distinct ultrasound features. Studies by M. Rotondi et al. have shown that hypoechoic areas observed on thyroid ultrasound are associated with lymphocytic infiltration, which may serve as a key diagnostic marker for early Seronegative HT [3–5]. However, research in this area remains limited. In recent years, machine learning has been widely adopted in the medical field [6, 7]. Machine learning has been widely applied in the medical field, particularly in analyzing large volumes of imaging data to develop diagnostic models. These models enable the comprehensive evaluation of complex ultrasound features, facilitating the identification of subtle early-stage disease changes. By assisting clinicians, machine learning models significantly enhance diagnostic efficiency and accuracy, demonstrating notable advantages in early detection and differential diagnosis. This study included patients with a pathological diagnosis of HT following thyroid tumor surgery. The research focused on the "hypoechoic" features of thyroid ultrasound and analyzed ultrasound images within a 0.5 cm perimeter surrounding the tumor lesion. An AI-assisted ultrasound diagnostic model was developed for HT, using lymphocytic infiltration ratios from pathological slides as a reference to refine the model's detection algorithms and thresholds. The aim was to enhance the diagnostic

accuracy of early Seronegative HT during the euthyroid phase, reduce missed diagnoses, and assist in clinical decision-making.

Methods

Ethical approval for this retrospective case-control study was granted by the Ethics Committee of Liaoning Provincial People's Hospital (Ethics No. [2023] H013). The study adhered to the principles of the Declaration of Helsinki, and eligible patients were identified through medical records with written informed consent obtained from each patient after full explanation of the purpose and nature of all procedures used. This study was registered in the Chinese Clinical Trial Registry (CTR2400092179; 12 November 2024.).

Study subjects and subgroups

The study included a total of 164 patients undergoing thyroid surgery for tumor treatment across four hospitals in China (Liaoning Provincial People's Hospital: 101 cases, Linghai Dalinghe Hospital, Liaoning Province: 9 cases, Lixin County People's Hospital, Anhui Province: 6 cases, and Fengcheng Phoenix Hospital, Liaoning Province: 18 cases) from November 2016 to January 2024, with a total of 298 ultrasound images collected. The training set consisted of 110 patients with 220 images, while the external test set included 24 patients with 48 images, and the normal control group comprised 30 patients with 30 images. All cases were categorized into antibody-positive (training set, internal test, external test: $n=50$, $n=9$, $n=12$) and antibody-negative groups (training set, internal test, external test: $n=42$, $n=9$, $n=12$), along with a non-Hashimoto's thyroiditis normal control group ($n=30$), as detailed in Fig. 1. Inclusion criteria: (1) Patients with normal preoperative thyroid function tests. (2) Patients who underwent thyroidectomy for thyroid tumor disease.

(3) Comprehensive transverse and longitudinal ultrasound images centered on the tumor, with reports provided by senior ultrasound physicians with over 10 years of experience.

(4) Postoperative pathological results confirming the presence of HT. Exclusion criteria: (1) Pregnant or breastfeeding women; (2) History of neck irradiation or thyroid surgery prior to this study; (3) Current and/or previous treatment with thyroid hormones; (4) Use of corticosteroids, amiodarone, lithium, oral contraceptives, or other medications that interfere with thyroid function [8]; (5)

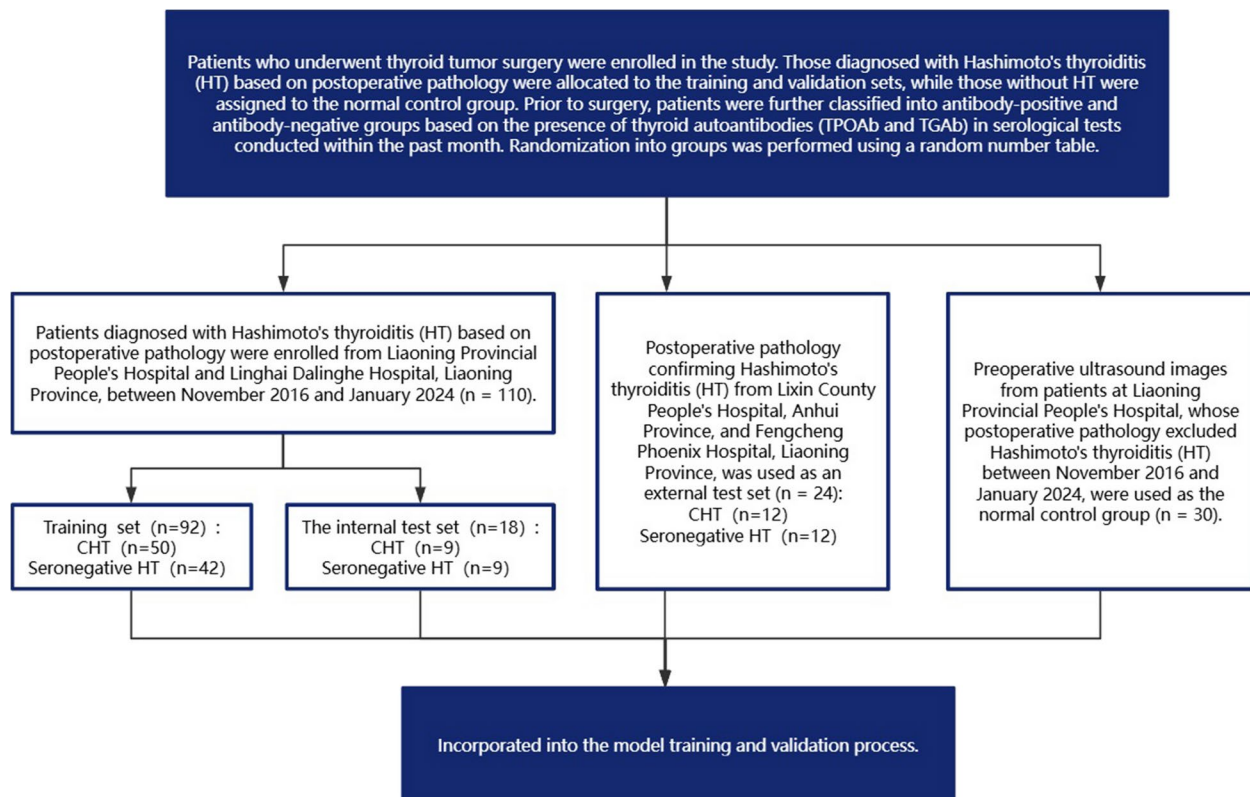


Fig. 1 Flow chart regarding the participants in this study

Coexisting chronic diseases that severely affect bodily function; [6] Previous diagnosis of other thyroid diseases, particularly Graves' disease and subacute thyroiditis; (7) Presence of severe obesity (BMI ≥ 30 kg/m²); (8) Incomplete thyroid function tests and ultrasound image data within one month prior to surgery.

Data collection

Preoperative data: including patient demographics (age, sex), thyroid function, and ultrasound images, along with postoperative pathology slides, were collected for analysis. Thyroid ultrasound images were captured using PHILIPS EPIQ7, SIEMENS-S3000, or GE Voluson color Doppler ultrasound machines, each equipped with 5–12 MHz linear array probes. Patients were positioned supine with the head tilted backward to expose the anterior neck for optimal imaging. Longitudinal and transverse images of the thyroid lobes and isthmus were acquired by senior ultrasound physicians with over 10 years of clinical experience, following the American College of Radiology (ACR) accreditation standards [9]. The largest cross-sectional view of the thyroid tumor, in both longitudinal and transverse planes, was selected, and ultrasound images within 0.5 cm of the tumor margin were utilized for machine learning analysis. The pathological slide containing the largest tumor diameter was selected, and lymphocyte infiltration within a 0.5 cm radius around the

nodule was assessed and recorded by a senior pathologist [10], serving as the reference standard for training and testing of the radiomic-histologic model.

Machine learning models construction

Data processing and model development

Image preprocessing and Region of Interest (ROI) segmentation (Mask)

In this study, ROI segmentation was performed using the Labelme v5.3.1 software to obtain JSON files, and these files were processed in a Python 3.6 environment to generate mask images. Main Process: The ROI, defined as the 0.5 cm ring-shaped area surrounding the nodule in the ultrasound image, was manually delineated using the LabelMe software. The central nodular lesion was delineated as the "exclusion region," while the anterior neck muscles were marked as the "hypoechoic reference region." After delineating these three regions, each was separately labeled and saved as a JSON file, creating the final delineation data. The JSON file was then input into the program, and functions from the LabelMe library were used to convert the image into a mask image, saved as a PNG file with the suffix "mask." The original data image was also converted into a grayscale image and saved as a PNG file with the suffix "original."

A grayscale quantization method was employed to discretize the grayscale levels of the image into a limited

number of levels. This approach reduces computational complexity and helps minimize the impact of noise on feature extraction. Quantization refers to the process of converting the continuous brightness variations of image pixels into discrete values. In other words, the amplitude values of the spatial coordinates in the original grayscale image are discretized. A greater number of quantization levels results in richer image layers, higher grayscale resolution, and improved image quality. Conversely, fewer quantization levels result in less detailed image layers, lower grayscale resolution, and a layered contour phenomenon in the image, reducing its quality.

High-throughput radiomic feature extraction and selection

The processed mask images were imported into the PyRadiomics library, and a feature extractor was defined for radiomics feature extraction using PyRadiomics version 3.0.1. Subsequently, the mask and original grayscale images were loaded into the program, and the extractor.execute() method was utilized to extract features from both. In the PyRadiomics configuration file (yaml), the setting section was customized with a bin width of 25, the interpolator set to "sitkBSpline," sigma set to 1.0, and the image type specified as "Original" for direct extraction. The extracted results were stored for subsequent analysis.

The extracted features were then converted into a list and saved to an Excel file, yielding 94 high-throughput radiomics features [11]. These features included First Order Statistics (19 features), Gray Level Co-occurrence Matrix (24 features), Gray Level Run Length Matrix (16 features), Gray Level Size Zone Matrix (16 features), Gray Level Dependence Matrix (14 features), and Neighboring Gray Tone Difference Matrix (5 features). Additionally, the ratio of hypoechoic areas was included as a radiomics feature, resulting in a total of 95 features for subsequent modeling and selection.

Feature selection, statistical analysis, and model development

Ratio feature computation

The grayscale mean of the muscle tissue region was used to standardize the hypoechoic region's pixel intensity, ensuring uniform feature extraction within the ROI for subsequent analysis. The model development was conducted in a Python 3.6 environment using PyCharm. Input data included the original grayscale images ("original") and the mask images ("mask image"), and the ratio of the hypoechoic region and the optimal threshold were computed through comparative analysis.

Main process The original grayscale image files ("original") and the mask files ("mask image") were read and converted into pixel matrices. The number of non-zero pixels in the muscle matrix ('musl') and the mean value of these

pixels were calculated, respectively. Similarly, the number of non-zero pixels and the average value for the ROI were computed. The pixel values of the ROI were then subtracted by the grayscale mean of the muscle matrix to obtain the calibrated ROI. The Pearson correlation coefficient (PCC) was introduced as an evaluation metric.

$$PCC(x, y) = \frac{\sum (x_i - \bar{x})(y_i - \bar{y})}{\sqrt{\sum (x_i - \bar{x})^2 \cdot \sum (y_i - \bar{y})^2}}$$

The similarity between the hypoechoic region and the ROI was calculated by iteratively checking the PCC between the overlapping area of the ROI and the hypoechoic region. The optimal threshold was determined by maximizing the correlation between the computer-identified and manually delineated hypoechoic regions. This threshold was continuously refined by comparing it with the gold standard from pathological assessments provided by the pathology department, thus improving the accuracy of threshold selection and recognition results.

At this threshold, the proportion of the hypoechoic region within the ROI, referred to as the ratio, was calculated. This ratio represented the proportion of the hypoechoic area within the ROI as identified by the computer and was used as a feature. As shown in Fig. 3 A1 and A2, the red region indicates the irrelevant area (tumor area), the green region represents the study area (hypoechoic region), and the yellow area is the reference region (muscle region).

Feature Selection Feature selection was performed on the ultrasound radiomics features from the training set. Pearson correlation analysis was initially used to exclude features with a correlation coefficient greater than 0.9. Subsequently, the t-distributed stochastic neighbor embedding (t-SNE) nonlinear dimensionality reduction algorithm was applied to the extracted multidimensional features for dimensionality reduction and visualization. T-SNE was used to visualize high-dimensional data in a two- or three-dimensional space, as it models similar samples by nearby points and dissimilar samples by distant points.

At a higher level, t-SNE constructs a probability distribution for high-dimensional samples, where similar samples are likely to be selected, while dissimilar points have a low probability of selection. t-SNE then defines a similar distribution for points in the low-dimensional embedding. Finally, it minimizes the Kullback-Leibler divergence between the two distributions concerning the embedding point positions [12].

After dimensionality reduction, Least Absolute Shrinkage and Selection Operator (LASSO) regression was applied to select the most informative features. LASSO regression, a linear regression model, incorporates L1

regularization to constrain the coefficients during model training, automatically selecting features by shrinking the coefficients of less important features to zero, thereby simplifying the model. The objective of LASSO regression is to minimize the following loss function:

$$\text{Loss} = \frac{1}{2m} \sum_{i=1}^m (y_i - \hat{y}_i)^2 + \alpha \sum_{j=1}^n |\beta_j|$$

The loss function includes a least-squares term and a regularization term (specifically, the L1 norm). The goal is to find a set of regression coefficients (β) that minimize the loss function.

In Python, the LASSO process was implemented using the LassoCV function from `sklearn.linear_model`, along with the StandardScaler function from `sklearn.preprocessing` for feature normalization. The StandardScaler ensured that all features were scaled to comparable ranges, enabling fair coefficient comparisons. For the hyperparameter α (corresponding to λ , the penalty term), a grid search was conducted. The range of α values was defined from 10^{-4} to 10^0 , on a logarithmic scale, divided into 50 intervals using the `np.logspace` function. Through iterative adjustment of the penalty term α , the regression coefficients were progressively reduced. Only the non-zero coefficients corresponding to significant features were retained. Cross-validation was then used to determine the optimal α value (α_{\min}), ensuring the selection of the best feature set.

Machine learning classifier

After feature selection using LASSO regression, a Random Forest (RandomForestClassifier) machine learning model was constructed based on the selected features. In machine learning, Random Forest is an ensemble classifier comprising multiple decision trees, where the final output is determined by the majority vote of individual trees. Among existing algorithms, RandomForestClassifier demonstrates excellent accuracy, operates efficiently on large datasets, handles high-dimensional feature inputs without requiring dimensionality reduction, and effectively evaluates the importance of each feature in classification tasks.

Data preprocessing

The necessary libraries and modules were first imported, including pandas for data processing and components from sklearn, such as RandomForestClassifier, DictVectorizer for feature extraction, and GridSearchCV for hyperparameter tuning and cross-validation. The dataset was then loaded, and the selected features—original_glcM_ClusterShade, original_glcM_Idmn, original_ngtdm_Busyness, original_glcM_MCC, and ratio—were extracted.

These features were subsequently transformed into a dictionary-based representation using DictVectorizer.

Model construction

A RandomForestClassifier from `sklearn.ensemble` was defined with key hyperparameters, including `n_estimators` (number of trees in the forest), `max_depth` (maximum depth of each tree), `min_samples_split` (minimum number of samples required to split an internal node, set to 2), `min_samples_leaf` (minimum number of samples required to be at a leaf node, set to 1), `criterion` (set to gini for impurity-based feature splitting), and `bootstrap` (set to True, indicating that bootstrap sampling was used when building trees). Among these, `n_estimators` was the most critical hyperparameter, initialized at 50 with an increment of 50 up to a maximum of 1000. A grid search with cross-validation (GridSearchCV) was employed to optimize the hyperparameters, particularly `n_estimators` (number of decision trees) and `max_depth` (maximum depth of decision trees). The Random Forest classifier was trained using the training dataset (`X_train`, `y_train`), and the optimal hyperparameter combination (`best_params`) and best-performing estimator (`best_estimator`) were identified.

Figure 2 illustrates the construction process of the machine learning model CCH-NET. The flowchart includes image acquisition, image preprocessing, feature extraction, feature selection, calculation of ratios, model building and analysis, and input of images for AI-based diagnosis. The CCH-NET model was developed using the selected features.

Figure 3 presents the ultrasound images and their corresponding pathological images from the test set, which were used for predictive modeling and performance evaluation.

Figure 4 illustrates the trajectory of a coefficient for an individual variable. The y-axis indicates the coefficient values, while the x-axis represents $\log(\lambda)$. As $\log(\lambda)$ increases, the regression coefficients (y-axis values) progressively converge to zero. Variables with non-zero coefficients are the important features selected for the model.

Figure 5 shows the relationship between the logarithm of the penalty term, $\log(\lambda)$, and the mean squared error (MSE). The x-axis represents $\log(\lambda)$, while the y-axis shows the MSE. A smaller y-axis value indicates better model fitting. The left dashed line corresponds to λ_{\min} , which is the λ value with the minimum bias and represents the optimal model fitting. The right dashed line corresponds to λ_{1se} , which is one standard error to the right of λ_{\min} .

Figure 6 illustrates how the coefficients of the selected features evolve as $\log(\lambda)$ increases. As $\log(\lambda)$ increases, the coefficients of the selected features gradually approach zero. Features with slower coefficient reduction

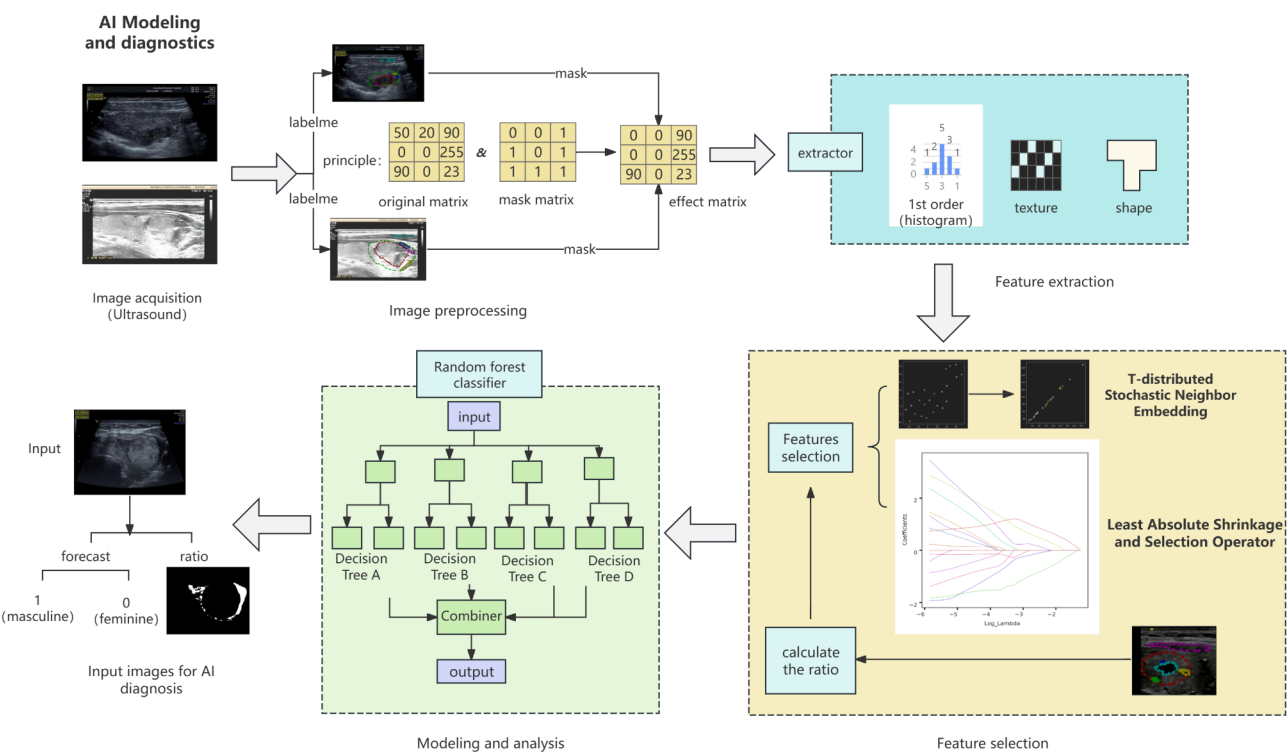


Fig. 2 Method and process of establishing CCH-NET

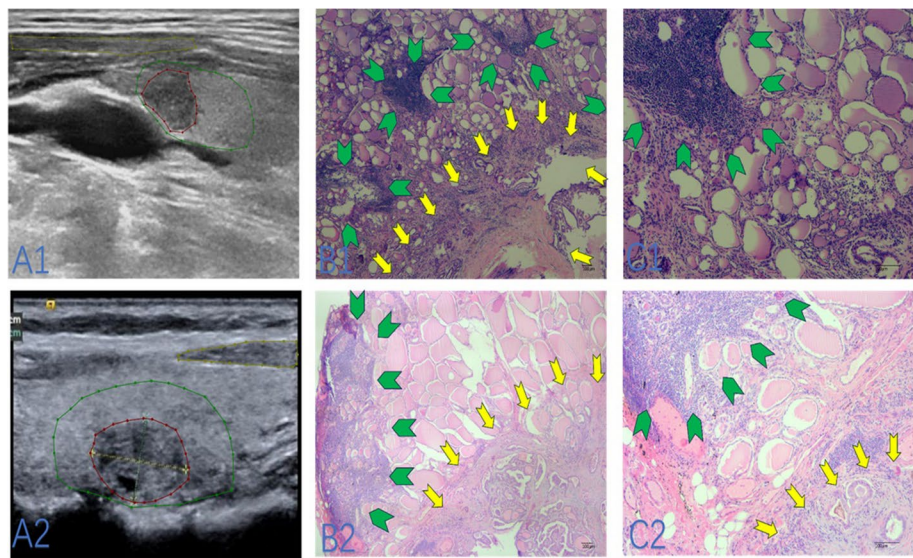


Fig. 3 Ultrasound and pathological images of antibody-positive and antibody-negative cases. A1: Antibody-positive ultrasound: red = excluded area (tumor), green = ROI, yellow = reference area. A2: Antibody-negative ultrasound: same color codes as A1. B1: Antibody-positive pathology (40x): 10–20 % lymphocyte infiltration; yellow arrow = thyroid tumor, green arrow = lymphocyte cluster. B2: Antibody-negative pathology (40x): 5% lymphocyte infiltration; same arrow indicators as B1. C1: Antibody-positive pathology (100x): 10–20 % lymphocyte infiltration. C2: Antibody-negative pathology (100x): 5% lymphocyte infiltration; green arrow = lymphocyte cluster

and later convergence to zero contribute more significantly to the predictive model, highlighting their greater importance in the model.

Statistical methods

Data analysis was performed using SPSS 26.0 software. Baseline characteristics were confirmed to follow a normal distribution. Continuous variables were expressed as Mean ± Standard Deviation (Mean ± SD) and compared

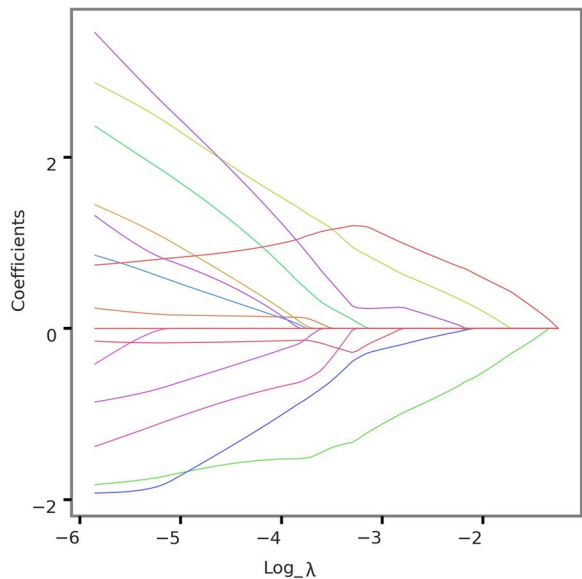


Fig. 4 Coefficient path diagram for regression

between groups using independent sample t-tests. Categorical data were presented as frequencies and percentages (n, %) and analyzed using chi-square (χ^2) tests. A P -value < 0.05 was considered statistically significant. The diagnostic performance of the model was evaluated using receiver operating characteristic (ROC) curves, with pathological results serving as the gold standard. The AUC was used to quantitatively assess the diagnostic accuracy of the model, and its performance was compared to diagnostic outcomes assessed by senior ultrasound physicians with extensive experience.

Results

Clinical profiles of patients included in this study

Table 1 summarizes the baseline characteristics of HT patients (training set: $n= 92$, test set: $n= 42$). Significant differences were observed in age, TSH, TPOAb, and TGAb levels between classic Hashimoto’s thyroiditis(CHT) and Seronegative HT patients across all sets. Specifically, Seronegative HT patients were significantly older than CHT patients (training set: $P= 0.05$; internal test set: $P= 0.01$; external test set: $P= 0.01$). In the training set, TSH levels in Seronegative HT patients were

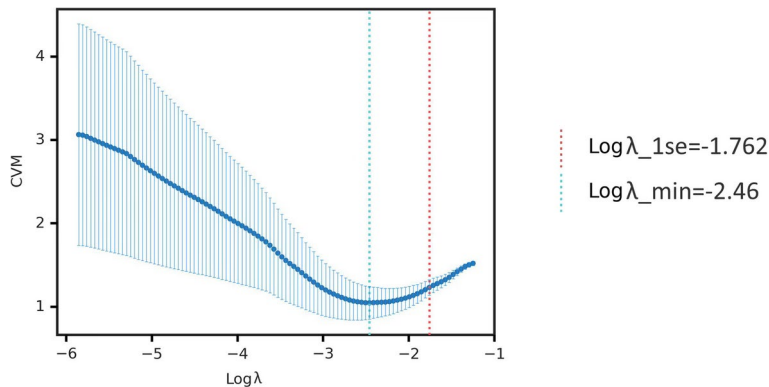


Fig. 5 Cross-validation curve for LASSO regression

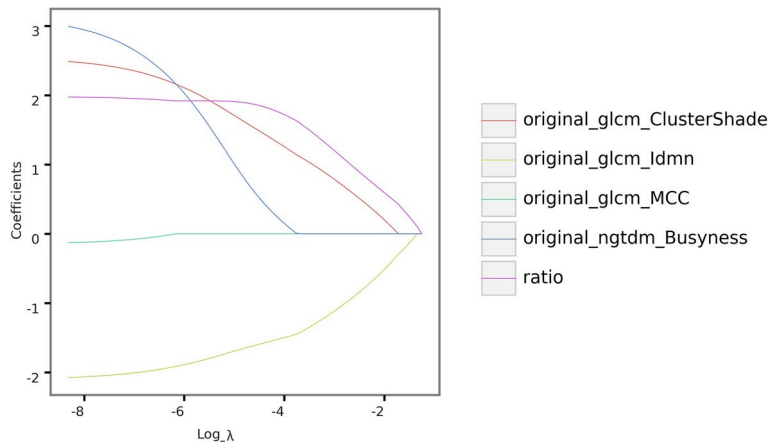


Fig. 6 Coefficient trajectories of the selected features

Table 1 Baseline characteristics of HT patients

| Items | Training set | | P-value | Internal test set | | P-value | External test set | | P-value |
|-------------|---------------------|---------------------------------|---------|--------------------|--------------------------------|---------|---------------------|---------------------------------|---------|
| | CHT(<i>n</i> = 50) | Seronegative HT(<i>n</i> = 42) | | CHT(<i>n</i> = 9) | Seronegative HT(<i>n</i> = 9) | | CHT(<i>n</i> = 12) | Seronegative HT(<i>n</i> = 12) | |
| Female/Male | 47/3 | 41/1 | 0.40 | 8/1 | 9/0 | 0.50 | 11/1 | 12/0 | 0.50 |
| Age(years) | 48.4 ± 13.1 | 53.7 ± 9.6 | 0.05 | 51.3 ± 12.7 | 54.1 ± 8.3 | 0.01 | 44.2 ± 16.7 | 55.5 ± 8.3 | 0.01 |
| FT3(pg/ml) | 3.35 ± 0.74 | 3.24 ± 0.43 | 0.09 | 3.15 ± 0.24 | 3.41 ± 0.57 | 0.13 | 3.74 ± 1.15 | 3.01 ± 0.41 | 0.08 |
| FT4(ng/dl) | 1.23 ± 0.22 | 1.29 ± 0.23 | 0.39 | 1.14 ± 0.09 | 1.27 ± 0.21 | 0.04 | 1.32 ± 0.29 | 1.26 ± 0.26 | 0.76 |
| TSH(μU/ml) | 3.10 ± 3.63 | 2.22 ± 1.23 | 0.02 | 3.47 ± 2.77 | 1.57 ± 1.00 | 0.13 | 3.12 ± 4.02 | 2.76 ± 0.89 | 0.09 |
| TPOAb(U/ml) | 557.81 ± 530.12 | 37.18 ± 11.86 | 0.00 | 1038.96 ± 385.41 | 35.49 ± 8.65 | 0.00 | 136.87 ± 138.76 | 37.69 ± 13.72 | 0.02 |
| TGAb(U/ml) | 242.18 ± 194.71 | 20.75 ± 14.87 | 0.00 | 252.51 ± 196.84 | 27.07 ± 20.08 | 0.00 | 347.56 ± 231.57 | 20.64 ± 14.09 | 0.00 |

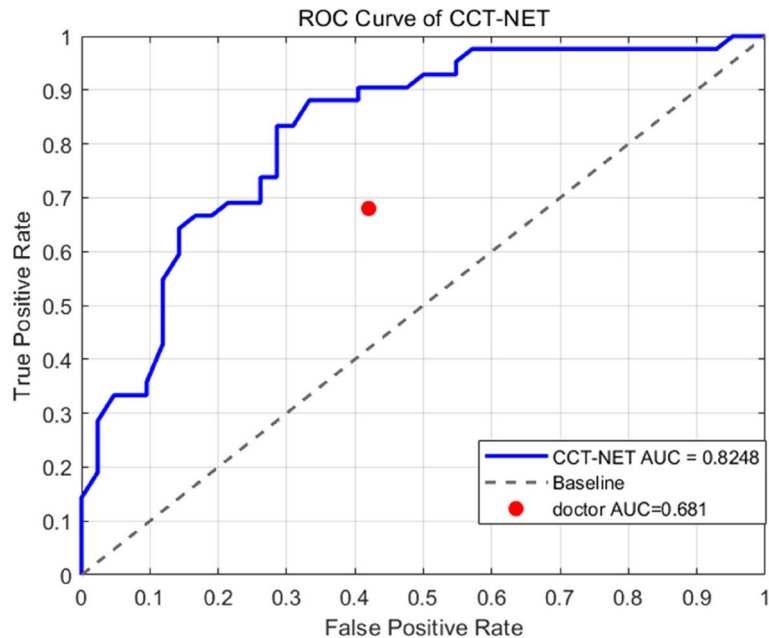


Fig. 7 The ROC curve and AUC for CCH-NET and senior ultrasound physicians

significantly lower compared to CHT patients ($P=0.02$). Additionally, FT4 levels in the Seronegative HT group were significantly higher than those in the CHT group in the internal test set ($P=0.04$).

Comparison of diagnostic efficacy between the CCH-NET Model and senior ultrasound physicians

The hypoechoic regions identified by CCH-NET were compared to the lymphocytic infiltration percentages from pathological analysis. Diagnostic accuracy was defined as a match or a discrepancy within 3%. Among 9 HT patients in the test sets with discrepancies, 8 showed AI-detected hypoechoic areas within 10% of pathological findings, with an average discrepancy of 6.10%, while 1 patient had a discrepancy of 11%.

Figure 7 illustrates the ROC curves comparing the diagnostic performance of the CCH-NET model and senior ultrasound physicians. The AUC of CCH-NET was 0.8482, surpassing that of the senior ultrasound

physicians (0.681), demonstrating the superior diagnostic efficacy of the CCH-NET model.

CCH-NET demonstrated significantly higher diagnostic accuracy than s senior ultrasound physicians, with internal (88.89% vs. 22.22%, $p<0.01$). For the external test set, CCH-NET also showed higher accuracy (75.00% and 66.67% vs. 50.00%), though the difference was not statistically significant ($p>0.05$). For the normal group, both methods achieved identical diagnostic accuracy of 93.33% ($P=1.00$).

The confusion matrix for CCH-NET's diagnostic performance is presented in Fig. 8. The model correctly identified 32 True Positives (TP) and 30 True Negatives (TN), with 12 False Positives (FP) and 10 False Negatives (FN).

Table 2 presents the diagnostic performance metrics of the CCH-NET model, including Accuracy, Precision, Recall, Sensitivity, Specificity, and F1 Score. The model demonstrated superior predictive performance in Test_1, achieving an F1 Score of 0.778. In Test_2, which

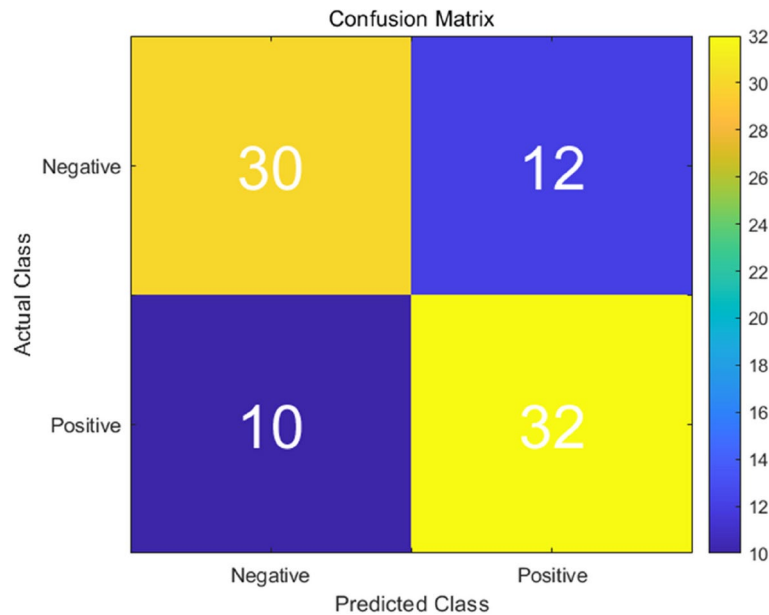


Fig. 8 Confusion matrix for CCH-NET in diagnosing Hashimoto's thyroiditis

Table 2 Diagnostic performance metrics of the CCH-NET model in test_1 and test_2

| Type | Total sample size | Acc | Pre | Rc | F1-Score | Sen | Sc | Positive predictive rate | Negative predictive rate |
|------------|-------------------|-------|-------|-------|----------|-------|-------|--------------------------|--------------------------|
| Test_1 | 36 | 0.778 | 0.778 | 0.778 | 0.778 | 0.778 | 0.778 | 0.778 | 0.778 |
| Test_2 | 48 | 0.727 | 0.692 | 0.750 | 0.720 | 0.750 | 0.667 | 0.692 | 0.727 |
| Test_total | 84 | 0.738 | 0.727 | 0.762 | 0.744 | 0.762 | 0.714 | 0.727 | 0.750 |

comprised data from different hospitals, the model maintained consistent diagnostic accuracy with an F1 Score of 0.720, indicating robust generalization across diverse datasets. Additionally, the model's Positive Predictive Rate and Negative Predictive Rate were consistently higher than those of senior ultrasound physicians in both test sets, further underscoring its diagnostic reliability.

Discussion

A key pathological hallmark of Seronegative HT is early focal or diffuse lymphocytic infiltration [13]. While fine needle aspiration biopsy (FNAB) is an effective diagnostic tool for Seronegative HT, its invasive nature often leads to poor patient acceptance [14]Ultrasound, a widely used non-invasive diagnostic modality, plays a pivotal role in the detection of thyroid disorders. Studies indicate that most Seronegative HT patients exhibit "hypoechoic" features in their ultrasound images [5], which are regarded as early ultrasonographic markers of Seronegative HT [15, 16]. Leveraging this feature, the diagnostic sensitivity of ultrasound physicians for HT can be enhanced. Hence, ultrasound offers significant clinical utility for the early screening of both HT and Seronegative HT.

The application of artificial intelligence (AI)-assisted ultrasound diagnostics has advanced considerably in the assessment of thyroid nodules [17–19]; however, it remains in the nascent stages of development within the domain of HT. Zhao et al. [20] developed a convolutional neural network-based computer-aided diagnosis model for Hashimoto's thyroiditis (CAD-HT), achieving a diagnostic efficacy of 89%. Zhang et al. [21]devised a dual-branch deep learning architecture capable of concurrently processing serological markers and ultrasound images, culminating in the development of a diagnostic model known as HTNet, which attained a diagnostic accuracy of 83.2% for HT and an AUC value of 94.9%. Both studies represent the latest advancements in current HT research; however, neither included Seronegative HT within their scope of investigation.

Our patient data suggest that the prevalence age of seronegative HT is higher than that of CHT, indicating that the course of seronegative HT may be more insidious, with milder symptoms or slower progression. This underscores the importance of early diagnosis, particularly in screening older age groups. Furthermore, seronegative HT patients exhibited lower TSH levels and higher FT4 levels (Table 1), aligning with the findings reported by Mario Rotondi [13]. These findings suggest

that seronegative HT causes less thyroid damage, which may be a characteristic feature of the condition. However, distinguishing between seronegative HT and CHT based on this feature alone is not feasible. Therefore, ultrasound analysis, augmented by imaging histology and machine learning techniques, may serve as a more accurate tool for early diagnosis, particularly in patients with atypical or antibody-negative clinical presentations. Importantly, the model demonstrates robustness and outperforms the diagnostic performance of senior ultrasound physicians, regardless of varying antibody statuses. It is important to note that the model demonstrates robust performance, outperforming senior ultrasound physicians, regardless of antibody status. This further validates the conclusions drawn by Zhao et al. on the CAD-HT model and Zhang et al. on HTNet. These findings indicate that radiomics-based machine learning models can reliably and effectively diagnose HT and related conditions without reliance on antibody expression, adapting well to diverse patient groups and exhibiting strong robustness. A retrospective cross-sectional study revealed that Seronegative HT accounts for up to 34.6% of primary hypothyroidism cases [22]. Given this proportion, many Seronegative HT cases may remain undetected. Therefore, the core of this study is to improve the detection rate of early Seronegative HT in patients with normal thyroid function or subclinical hypothyroidism by applying machine learning models. In our internal test set, senior ultrasound physicians achieved a low diagnostic accuracy (39.09%), likely due to the milder autoimmune inflammatory infiltration and smaller thyroid volume in Seronegative HT compared to CHT [13], as well as the less evident coarse internal structure on ultrasound [22], making it challenging to identify typical features. Thus, improving the early detection of Seronegative HT is critical. Our study demonstrates that the CCH-NET model can enhance the diagnostic accuracy of ultrasound physicians for Seronegative HT, facilitating timely intervention for early-stage patients and potentially preventing the onset of hypothyroidism.

In this study, the multidimensional features of ultrasound images are extracted by using pyradiomics correlation, and Scikit-Learn correlation methods, such as T-SNE dimensionality reduction method and PCC evaluation metrics, are introduced to filter the features that may be relevant and effective, so as to simplify the model by significantly reducing the computational time complexity, and at the same time, to ensure the simplicity of the model and the validity of the model. In the process of constructing the model, negative case images and positive case images are grouped multiple times, and leave-one-out cross-validation is used. This cross-validation method is unique in that it uses a single observation in the original sample as the test data, and the rest of the

Table 3 The diagnostic accuracy of CCH-NET and senior ultrasound physicians for different antibodies in the test set

| Items | | CCH-NET | Senior ultrasound physicians | X ² | P-value |
|-----------------------------|-------------------------------|---------|------------------------------------|----------------|---------|
| Internal test accuracy | CHT(9cases) | 88.89% | 22.22% | 8.10 | < 0.01 |
| | Seronegative HT (9 cases) | 88.89% | 22.22% | 8.10 | < 0.01 |
| External test accuracy | CHT (12 cases) | 75.00% | 50.00% | 1.60 | 0.21 |
| | Seronegative HT (12 cases) | 66.67% | 50.00% | 0.69 | 0.41 |
| Normal group accuracy | Normal group (30 cases) | 93.33% | 93.33% | 0.00 | 1.00 |

observations as the training data, which avoids the narrowness and overfitting of fewer number of cases and fewer number of positive cases [23], and effectively ensures that the model can be tested in case of limited number of cases, which effectively guarantees the model's validity in case of limited number of cases. accuracy of CCH-NET diagnosis in limited cases.

First, it is based on retrospectively collected data from thyroid surgery cases, which were pathologically confirmed as seronegative. The relatively small number of eligible samples may limit the model's generalization ability. Furthermore, while the majority of patients in this study were female, reflecting the known female predominance in Hashimoto's Thyroiditis, the sex distribution between the Seronegative HT and CHT groups was not significantly different ($P > 0.05$, Table 3). This suggests that sex-related variability did not influence the diagnostic performance of the model. However, the overrepresentation of female patients in our dataset may still limit the generalizability of the findings to male patients. Future studies with more balanced sex distributions are warranted to validate the model across diverse populations. From an algorithmic perspective, during feature selection, Lasso regression can be unstable when the number of feature variables greatly exceeds the number of samples, as it may not perform group selection, meaning that related variables may not be selected or excluded together. During the dimensionality reduction process using the t-SNE algorithm, the results can be highly sensitive to initialization, leading to different outcomes depending on the initialization. Additionally, t-SNE has high computational complexity, requiring significant processing time for large-scale datasets. Although certain variables, such as TSH and age, did not significantly impact model performance, future research should increase the sample size to further optimize the model. Second, nine HT patients showed bias in both the internal and external test sets. This may be due to the fact that the ultrasound images used contained only two-dimensional images of the longitudinal and transverse sections

of the largest diameter of the thyroid nodule, whereas the periphery of the nodule is a three-dimensional space. In the case of lymphocytic infiltration distribution, there exists the possibility of missed detection [24]. The application of AI and dynamic 3D imaging histology should be explored in subsequent studies [25]. Third, this study only focused on the ultrasonographic features of early Seronegative HT "hypoechoic." In other stages of HT, such as "pseudonodule," "mild enlargement of the gland" in the progressive stage, "lattice-like fibrous segregation" in the active stage, and "echogenic enhancement" in the recovery stage [26], were not used as indicators for model learning. Therefore, in future research, we will focus on expanding the sample size of the study, introducing the AI learning of dynamic ultrasound [27], and combining with ultrasound imaging histology multi-feature analysis to pursue the further improvement of the diagnostic accuracy of CCH-NET for Seronegative HT.

In conclusion, the CCH-NET model, integrating radiomics and machine learning methods, offers a promising tool for the early diagnosis of SN-HT, with potential value in clinical applications.

Abbreviations

| | |
|--------|---|
| HT | Hashimoto's thyroiditis |
| AUC | Area Under the Curve |
| CHT | Classic Hashimoto's Thyroiditis |
| TPO-Ab | Thyroid Peroxidase Antibody |
| TG-Ab | Thyroglobulin Antibody |
| TSH | Thyroid-Stimulating Hormone |
| FT3 | Free Triiodothyronine |
| FT4 | Free Thyroxine |
| LASSO | Least Absolute Shrinkage and Selection Operator |
| ROI | Region of Interest |
| PCC | Pearson Correlation Coefficient |
| FNAB | Fine Needle Aspiration Biopsy |
| ROC | Receiver Operating Characteristic |

Acknowledgements

We thank all those who participated in the study.

Authors' contributions

Jianchun Cui, Zuoxin Ma, and Shijie Chang designed the study. Yuan Luo and Mingming Xiao performed the pathological slide analysis. Daming Liu and Qi Zhang reviewed the ultrasound images. Chang Liu, Jiahao Wang, and Shijie Chang developed the machine learning model. Mengyou Liu, Li Shi, Zhengshuai Liu, Huimei Cao, Xiang Fei, Yang Gao, Ying Zhang, Xuanyu Chen, Wanli Zheng, Xiali Niu, and Xiao Yang collected the images and clinical information. Wenjun Wu and Shengsheng Yao conducted statistical analysis and manuscript editing. Xingai Ju, Yihan Sun, Li Lu, and Liying Gong discussed and reviewed the manuscript. All authors read and approved the final manuscript. All authors had full access to the data and took full responsibility for the decision to submit for publication. Wenjun Wu, Shengsheng Yao, Daming Liu, Yuan Luo, Yihan Sun, Ting Ruan, Mengyou Liu, Li Shi, and Chang Liu contributed equally to this work.

Funding

This work was supported by the Applied Basic Research Program of the Liaoning Provincial Science and Technology Department (Project No. 2022020247-JH2/1013) and the Science and Technology Bureau of the 10th Division of the Xinjiang Production and Construction Corps in Beitun City (Project No. 2022 - 179 - 03).

Data availability

No datasets were generated or analysed during the current study.

Declarations

Ethics approval and consent to participate

Ethical approval was granted by the Ethics Committee of Liaoning Provincial People's Hospital (Ethics No. [2023] H013). Written informed consent was obtained from all patients after an explanation of the study procedures. This study was registered in the Chinese Clinical Trial Registry (CTR2400092179; 12 November 2024).

Consent for publication

This study does not contain any identifiable patient data or images requiring consent for publication. All authors have reviewed and approved the final manuscript and agree to its submission to the journal.

Competing interests

The authors declare no competing interests.

Author details

¹Liaoning University of Chinese Medicine, Shenyang, China

²Department of Thyroid and Breast Surgery, People's Hospital of China Medical University (Liaoning Provincial People's Hospital), Shenyang, China

³Department of Ultrasound Medicine, People's Hospital of China Medical University (Liaoning Provincial People's Hospital), Shenyang, China

⁴Department of Pathology, People's Hospital of China Medical University (Liaoning Provincial People's Hospital), Shenyang, China

⁵Department of Cardiology, People's Hospital of China Medical University (Liaoning Provincial People's Hospital), Shenyang, China

⁶Department of Thyroid Surgery, Lixin County People's Hospital, Haozhou, China

⁷Jinzhou Medical College Linghai Dalinghe Hospital, Jinzhou, China

⁸School of Intelligent Medicine, China Medical University, Shenyang, China

⁹Fengcheng Phoenix Hospital, Dandong, China

¹⁰Department of General Medicine, People's Hospital of China Medical University (Liaoning Provincial People's Hospital), Shenyang, China

¹¹Department of Endocrinology, People's Hospital of China Medical University (Liaoning Provincial People's Hospital), Shenyang, China

¹²Department of Medical Laboratory, People's Hospital of China Medical University (Liaoning Provincial People's Hospital), Shenyang, China

¹³Jinzhou Medical University, Jinzhou, China

Received: 10 November 2024 / Accepted: 28 March 2025

Published online: 07 April 2025

References

1. Baker JR Jr, Saunders NB, Wartofsky L, Tseng YC, Burman KD. Seronegative Hashimoto thyroiditis with thyroid autoantibody production localized to the thyroid. *Ann Intern Med.* 1988;108(1):26–30. <https://doi.org/10.7326/0003-4819-108-1-26>.
2. Croce L, De Martinis L, Pinto S, Coperchini F, Dito G, Bendotti G, et al. Compared with classic Hashimoto's thyroiditis, chronic autoimmune serum-negative thyroiditis requires a lower substitution dose of L-thyroxine to correct hypothyroidism. *J Endocrinol Invest.* 2020;43(11):1631–6. <https://doi.org/10.1007/s40618-020-01249-x>.
3. Vitti P, Rago T. Thyroid ultrasound as a predictor of thyroid disease. *J Endocrinol Invest.* 2003;26(7):686–9. <https://doi.org/10.1007/BF03347031>.
4. Willms A, Bieler D, Wieler H, Willms D, Kaiser KP, Schwab R. Correlation between sonography and antibody activity in patients with Hashimoto thyroiditis. *J Ultrasound Med.* 2013;32(11):1979–86. <https://doi.org/10.7863/ultra.32.11.1979>.
5. Rotondi M, Coperchini F, Magri F, Chiovato L. Serum-negative autoimmune thyroiditis: what's in a name? *J Endocrinol Invest.* 2014;37(6):589–91. <https://doi.org/10.1007/s40618-014-0083-8>.
6. Shehab M, Abualigah L, Shambour Q, Abu-Hashem MA, Shambour MKY, Alsalihi AI, Gandomi AH. Machine learning in medical applications: A review

- of state-of-the-art methods. *Comput Biol Med.* 2022;145:105458. <https://doi.org/10.1016/j.combiomed.2022.105458>.
7. Zhao C, Sun Z, Yu Y, Lou Y, Liu L, Li G, et al. A machine learning-based diagnosis modeling of IgG4 Hashimoto's thyroiditis. *Endocrine.* 2024;86(2):672–81. <https://doi.org/10.1007/s12020-024-03889-y>.
 8. Burch HB. Drug Effects on the Thyroid. *N Engl J Med.* 2019;381(8):749–61. <http://doi.org/10.1056/NEJMr1901214>.
 9. Fogh SE, Pope CH, Rosenthal SA, Conway PD, Hulick PR, Johnson JL, et al. American College of Radiology (ACR) Radiation Oncology Practice Accreditation: A pattern of change. *Pract Radiat Oncol.* 2016;6(5):e171–7. <https://doi.org/10.1016/j.prro.2016.01.010>.
 10. Ragusa F, Fallahi P, Elia G, Gonnella D, Paparo SR, Giusti C, et al. Hashimoto's thyroiditis: Epidemiology, pathogenesis, clinic and therapy. *Best Pract Res Clin Endocrinol Metab.* 2019;33(6):101367. <https://doi.org/10.1016/j.beem.2019.101367>.
 11. Jundong L, Kewei C, Suhang W, Fred M, Robert PT, Jiliang T, et al. Feature selection: a data perspective. *ACM Comput Surv.* 2017;50(6):1–45. <https://doi.org/10.1145/3136625>.
 12. Hinton G, Van Der Maaten L. Visualizing data using t-sne journal of machine learning research. *J Machine Learn Res.* 2008;9:2579–605.
 13. Rotondi M, de Martinis L, Coperchini F, Pignatti P, Pirali B, Ghilotti S, et al. Serum negative autoimmune thyroiditis displays a milder clinical picture compared with classic Hashimoto's thyroiditis. *Eur J Endocrinol.* 2014;171(1):31–6. <https://doi.org/10.1530/EJE-14-0147>.
 14. Jankovic B, Le KT, Hershman JM. Hashimoto's Thyroiditis and Papillary Thyroid Carcinoma: Is There a Correlation? *J Clin Endocrinol Metabol.* 2013;98(2):474–82. <https://doi.org/10.1210/jc.2012-2978>.
 15. Rago T, Chiovato L, Grasso L, Pinchera A, Vitti P. Thyroid ultrasonography as a tool for detecting thyroid autoimmune diseases and predicting thyroid dysfunction in apparently healthy subjects. *J Endocrinol Invest.* 2001;24(10):763–9. <https://doi.org/10.1007/bf03343925>.
 16. Vejbjerg P, Knudsen N, Perrild H, Laurberg P, Pedersen IB, Rasmussen LB, et al. The association between hypoechogenicity or irregular echo pattern at thyroid ultrasonography and thyroid function in the general population. *Eur J Endocrinol.* 2006;155(4):547–52. <https://doi.org/10.1530/eje.1.02255>.
 17. Xu D, Sui L, Zhang C, Xiong J, Wang VY, Zhou Y, et al. The clinical value of artificial intelligence in assisting junior radiologists in thyroid ultrasound: a multicenter prospective study from real clinical practice. *BMC Med.* 2024;22(1):293. <https://doi.org/10.1186/s12916-024-03510-z>.
 18. Wang J, Zheng N, Wan H, Yao Q, Jia S, Zhang X, et al. Deep learning models for thyroid nodules diagnosis of fine-needle aspiration biopsy: a retrospective, prospective, multicentre study in China. *Lancet Dig Health.* 2024;6(7):e458–69. [https://doi.org/10.1016/S2589-7500\(24\)00085-2](https://doi.org/10.1016/S2589-7500(24)00085-2).
 19. Zhou T, Xu L, Shi J, Zhang Y, Lin X, Wang Y, et al. US of thyroid nodules: can AI-assisted diagnostic system compete with fine needle aspiration? *Eur Radiol.* 2024;34(2):1324–33. <https://doi.org/10.1007/s00330-023-10132-1>.
 20. Zhao W, Kang Q, Qian F, Li K, Zhu J, Ma B. Convolutional Neural Network-Based Computer-Assisted Diagnosis of Hashimoto's Thyroiditis on Ultrasound. *J Clin Endocrinol Metab.* 2022;107(4):953–63. <https://doi.org/10.1210/clinem/dgab870>.
 21. Zhang Q, Zhang S, Pan Y, Sun L, Li J, Qiao Y, et al. Deep learning to diagnose Hashimoto's thyroiditis from sonographic images. *Nat Commun.* 2022;13(1):3759. <https://doi.org/10.1038/s41467-022-31449-3>.
 22. Iwamoto Y, Kimura T, Itoh T, Mori S, Sasaki T, Sugisaki T, et al. Structural and functional differences in auto-antibody positive compared to auto-antibody negative hypothyroid patients with chronic thyroiditis. *Sci Rep.* 2023;13(1):15542. <https://doi.org/10.1038/s41598-023-42765-z>.
 23. Hirose TA, Arimura H, Ninomiya K, Yoshitake T, Fukunaga JI, Shioyama Y. Radiomic prediction of radiation pneumonitis on pretreatment planning computed tomography images prior to lung cancer stereotactic body radiation therapy. *Sci Rep.* 2020;10(1):20424. <https://doi.org/10.1038/s41598-020-77552-7>.
 24. Kollorz EK, Hahn DA, Linke R, Goecke TW, Hornegger J, Kuwert T. Quantification of thyroid volume using 3-D ultrasound imaging. *IEEE Trans Med Imaging.* 2008;27(4):457–66. <https://doi.org/10.1109/TMI.2007.907328>.
 25. Cai S, Chen Y, Zhao S, He D, Li Y, Xiong N, et al. Dynamic 3D radiomics analysis using artificial intelligence to assess the stage of COVID-19 on CT images. *Eur Radiol.* 2022;32(7):4760–70. <https://doi.org/10.1007/s00330-021-08533-1>.
 26. Wu G, Zou D, Cai H, Liu Y. Ultrasonography in the diagnosis of Hashimoto's thyroiditis. *Front Biosci (Landmark Ed).* 2016;21(5):1006–12. <https://doi.org/10.2741/4437>.
 27. Tiyyarattanachai T, Apiparakoon T, Marukatat S, Sukcharoen S, Yimsawad S, Chaichuen O, et al. The feasibility to use artificial intelligence to aid detecting focal liver lesions in real-time ultrasound: a preliminary study based on videos. *Sci Rep.* 2022;12(1):7749. <https://doi.org/10.1038/s41598-022-11506-z>.

Publisher's Note

Springer Nature remains neutral with regard to jurisdictional claims in published maps and institutional affiliations.