Check for updates

OPEN ACCESS

EDITED BY Trevor Edmund Angell, University of Southern California, United States

REVIEWED BY Rosaria Maddalena Ruggeri, University of Messina, Italy Cari Kitahara, National Cancer Institute (NIH), United States

*CORRESPONDENCE Yuanyuan Wu yywu202207@gmail.com Yi Sun sunyi19700420@163.com

[†]These authors have contributed equally to this work

SPECIALTY SECTION

This article was submitted to Thyroid Endocrinology, a section of the journal Frontiers in Endocrinology

RECEIVED 01 March 2022 ACCEPTED 07 July 2022 PUBLISHED 08 August 2022

CITATION

Li P, Liu F, Zhao M, Xu S, Li P, Cao J, Tian D, Tan Y, Zheng L, Cao X, Pan Y, Tang H, Wu Y and Sun Y (2022) Prediction models constructed for Hashimoto's thyroiditis risk based on clinical and laboratory factors. *Front. Endocrinol.* 13:886953. doi: 10.3389/fendo.2022.886953

COPYRIGHT

© 2022 Li, Liu, Zhao, Xu, Li, Cao, Tian, Tan, Zheng, Cao, Pan, Tang, Wu and Sun. This is an open-access article distributed under the terms of the Creative Commons Attribution License (CC BY). The use, distribution or reproduction in other forums is permitted, provided the original author(s) and the copyright owner(s)

are credited and that the original publication in this journal is cited, in accordance with accepted academic practice. No use, distribution or reproduction is permitted which does not comply with these terms.

Prediction models constructed for Hashimoto's thyroiditis risk based on clinical and laboratory factors

Peng Li^{1†}, Fang Liu^{2†}, Minsu Zhao^{3†}, Shaokai Xu¹, Ping Li¹, Jingang Cao¹, Dongming Tian¹, Yaopeng Tan¹, Lina Zheng², Xia Cao², Yingxia Pan^{4,5}, Hui Tang^{4,5}, Yuanyuan Wu^{4,5*} and Yi Sun^{1*}

¹Department of Breast Surgery, Xuchang Central Hospital, Xuchang, China, ²Health Management Center, Kaifeng Central Hospital, Kaifeng, China, ³Department of Endocrinology, Jincheng People's Hospital, Jincheng City, China, ⁴Department of Medicine, Shanghai Biotecan Pharmaceuticals Co., Ltd., Shanghai, China, ^sShanghai Zhangjiang Institute of Medical Innovation, Shanghai, China

Background: Hashimoto's thyroiditis (HT) frequently occurs among autoimmune diseases and may simultaneously appear with thyroid cancer. However, it is difficult to diagnose HT at an early stage just by clinical symptoms. Thus, it is urgent to integrate multiple clinical and laboratory factors for the early diagnosis and risk prediction of HT.

Methods: We recruited 1,303 participants, including 866 non-HT controls and 437 diagnosed HT patients. 44 HT patients also had thyroid cancer. Firstly, we compared the difference in thyroid goiter degrees between controls and patients. Secondly, we collected 15 factors and analyzed their significant differences between controls and HT patients, including age, body mass index, gender, history of diabetes, degrees of thyroid goiter, UIC, 25-(OH)D, FT3, FT4, TSH, TAG, TC, FPG, low-density lipoprotein cholesterol, and high-density lipoprotein cholesterol. Thirdly, logistic regression analysis demonstrated the risk factors for HT. For machine learning modeling of HT and thyroid cancer, we conducted the establishment and evaluation of six models in training and test sets.

Results: The degrees of thyroid goiter were significantly different among controls, HT patients without cancer (HT-C), and HT patients with thyroid cancer (HT+C). Most factors had significant differences between controls and patients. Logistic regression analysis confirmed diabetes, UIC, FT3, and TSH as important risk factors for HT. The AUC scores of XGBoost, LR, SVM, and MLP models indicated appropriate predictive power for HT. The features were arranged by their importance, among which, 25-(OH)D, FT4, and TSH were the top three high-ranking factors.

Conclusions: We firstly analyzed comprehensive factors of HT patients. The proposed machine learning modeling, combined with multiple factors, are

efficient for thyroid diagnosis. These discoveries will extensively promote precise diagnosis, personalized therapies, and reduce unnecessary cost for thyroid diseases.

KEYWORDS

Hashimoto's thyroiditis, thyroid cancer, precise diagnosis, machine learning modeling, risk factors

Introduction

Hashimoto's thyroiditis (HT) is a chronic disease of autoimmune thyroid disorder (AITD) with high incidence, which is a common cause of hypothyroidism. It is mainly manifested by thyroid lymphocytic infiltration and fibrosis, as well as positive thyroid-specific antibodies in peripheral blood (1). Targeting the imaging diagnosis of the thyroid, the Thyroid Imaging Reporting and Data System (TIRADS) was established in many countries (2-4), including China (5). The best imaging method for diagnosing thyroid diseases is ultrasonography which provides precise dimensions of thyroid goiters (6). HT is the most frequent cause of goiters which are initially firm, bumpy, symmetric, and painless (7). However, HT is usually not recognized and diagnosed in the early stage and therefore, HT is more likely to deteriorate into papillary thyroid carcinoma (PTC) than nodular goiter (8, 9). Thus, it is necessary to establish prediction models for HT diagnosis at an early stage.

According to numerous reports, the development of thyroid diseases, especially thyroid cancer, is correlated with many risk factors, among which the main factors may optimize diagnostic techniques and treatment methods. For instance, iodine is an essential raw material for the synthesis of thyroid hormones (10). Studies have shown that the harm of iodine deficiency to individuals is significantly more serious than the harm caused by excessive iodine consumption (11). Insufficient iodine intake can

cause elevated thyroid stimulating hormone (TSH) levels, leading to clinical hypothyroidism by the regulation through intact hypothalamus-pituitary-thyroid axis (12, 13). The longterm iodine deficiency can lead to HT and even thyroid cancer. However, thyroid cancer is seemingly more prevalent in areas with excessive iodine intake which will induce the reduction of thyroid hormone synthesis. Susceptible population with excessive iodine intake may either induce hypothyroidism, such as HT, or cause hyperthyroidism, such as Grave's disease (GD) (14, 15), which suggests a complex relationship between iodine levels with AITD (16). The iodine uptakes of HT patients are usually low and can be tested by urinary iodine concentration (UIC) (17). Moreover, vitamin D is usually lacking in patients suffering from HT (18). It is reported that the risk of developing HT and thyroid cancer will increase with low 25-hydroxy vitamin D (25-(OH)D) levels, which can be improved by the supplementation of vitamin D (19).

In addition to malnutrition, the high levels of antibodies against thyroid antigens have been positively correlated with the clinical diagnosis of HT, especially for the thyroglobulin antibody (TGAb) against thyroid peroxidase (TPO) and the thyroid peroxidase antibody (TPOAb) against thyroglobulin (TG) (20–22). Also, the proper levels of inherent hormone metabolism, such as free thyroxine (FT4) and free triiodothyronine (FT3), along with fatty acid intake play important roles in the course of HT and thyroid cancer (19, 23). It is reported that other endocrine diseases, such as diabetes and hyperlipemia, are often concurrent in HT patients whose fasting plasma glucose (FPG) may be disordered (24, 25), while triglyceride (TAG), total cholesterol (TC) and low-density lipoprotein cholesterol (LDL-C) may be significantly lower than in healthy people (26).

In recent years, the applications of novel computer-aided diagnosis based on bioinformatic engineering has increasingly played appealing roles in the clinical diagnoses of multiple disorders. Many research studies on computer-aided diagnoses of disorders, including thyroid diseases, have been carried out. It is reported that computer-aided diagnosis can offer efficient dataset processing, accurate diagnosis, clinical decision-making support, and even potential treatment options for thyroid diagnosis (27–30). Computer-aided diagnosis of HT has been

Abbreviations: HT, Hashimoto's thyroiditis; PTC, papillary thyroid carcinoma; TIRADS, Thyroid Imaging Reporting and Data System; ICP-MS, inductively coupled plasma-mass spectrometry; ROC, receiver operating characteristic curve; AUC, area under curve; HT-C, HT patients without cancer; HT+C, HT patients with thyroid cancer; BMI, body mass index; UIC, urinary iodine concentrations; 25-(OH)D, 25 hydroxyvitamin D; FT3, free triiodothyronine; FT4, free thyroxine; TSH, thyroid stimulating hormone; TPOAb, thyroid peroxidase antibody; TAG, triglyceride; TC, total cholesterol; LDL-C, low density lipoprotein cholesterol; HDL-C, high density lipoprotein cholesterol; FPG, fasting plasma glucose; TGAb, thyroglobulin antibody; OR, odds ratio; AUC, area under curve; KNN, k-nearest neighbor classifier; LR, logistic regression; SVM, a support vector machine; DT, the decision tree model; MLP, the multilayer perceptron network; XGBoost, eXtreme Gradient Boosting.

carried out since 2009 (31). However, current computer-aided studies basically focus on thyroid imaging and only a few research studies involve molecular levels. According to a 2019 study, thyroid antibodies and hormones were involved in establishing a machine learning model for the quantitative and convenient diagnosis of HT distinguished from the Graves' disease (32), but these factors were not comprehensive enough. Recent studies on HT found that mathematical modeling highlighted T helper lymphocytes TH1, TH17, and the bacterial balance of the gut microbiota activities as risk factors for HT development (33).

In this study, we collected fifteen clinical characteristics and laboratory data to study their significant correlations with HT and thyroid cancer. Furthermore, we established machine learning models of HT development and thyroid cancer based on twelve factors. The study may offer reference in the diagnosis and prevention of HT, which also effectively reduces the economic burden on patients and save limited medical resources.

Methods

Recruitment and examination of participants

From July 2018 to January 2021, total 1,344 participants were recruited from Xuchang Central Hospital, Kaifeng Central Hospital, and Jincheng People's Hospital. This total included 885 non-HT controls and 459 diagnosed HT patients. All participants signed informed consents. The exclusion criteria in this study were incomplete clinical information, controls with thyroid cancer, and abnormal values. After screening, the cases of controls and HT patients were 866 and 437, respectively. The total factors were collected as follows: age, body mass index (BMI), gender, history of diabetes, degrees of thyroid goiter, thyroid cancer, UIC, 25-(OH)D, FT3, FT4, TSH, TAG, TC, FPG, LDL-C, and high-density lipoprotein cholesterol (HDL-C).

Inclusion criteria of HT patients and controls

The clinical criteria of HT are diffuse thyroid goiter, tough tissue, hypoechoic, disordered echoes on the thyroid, nodular changes with regular borders diagnosed by B-mode ultrasound imaging, along with increased TPOAb and/or TGAb in sera (34– 36). The controls had none or merely part of those changes in the thyroid gland. By means of clinical inspection and palpation, thyroid goiter was divided into three degrees (I, II, III) according to their sizes (Degree I: touchable but invisible; Degree II: touchable and visible, limited to posterior margin of sternocleidomastoid muscle; Degree III: touchable and visible, greater than posterior margin of sternocleidomastoid muscle). After data screening, 1,303 non-HT controls and HT patients were enrolled in the cohort from the initial 1,344 participants.

Ultrasound examination of thyroid glands

Ultrasound examinations were performed and pictured with the Vinno E30 scanner (Vinno Technology, Jiangsu, China). The frequency bandwidth of broadband linear array transducer (X6-16L) is $10 \sim 14$ MHz. In comparison with the controls, ultrasound features of HT patients were diffusely hypoechoic nodules on thyroid glands, visible sheet-like hypoechoic areas, and fibrous hyperechoic areas (Additional file 1). Ultrasound features of PTC differentiated from benign goiter, including irregular and rough borders, aspect ratios of nodules, microcalcification, increased blood flow, disordered and hypoechoic appearance on fibrous hyperplasia (37).

Microscopic examination of cytopathological sections

Microscopic examinations were only available for HT patients who potentially had thyroid cancer. The specimens of patients' thyroid glands were taken from biopsy or surgery and made into intraoperative cytology smears. The histopathological specimens were fixed in formalin solution and observed by Hematoxylin-Eosin staining. By means of microscope (Leica, DM 2000) examination, the postoperative diagnostic results were classified into normal, HT-type inflamed, and cancerous cells by their cytological appearance (Additional file 2). The pathological photographs were taken under magnification of 200. The pathological features of specimens taken from HT patients are thyroid follicles with Hürthle-cell metaplasia, lymphoid follicles with germinal centers, minimal colloid in follicles, and polymorphic lymphoid cells (38). The typical cytological recognition of PTC is an extensive papillary pattern, lymphoplasmacytic infiltrate, follicular variant, tall cell variant, and psammoma bodies. The typical nuclear characteristics of PTC are ground-glass nuclei, nuclear grooves, and round intranuclear pseudoinclusions (39, 40).

Collection and detection of urine and blood samples

Urine and blood samples were collected and detected from all participants at the time of initial diagnosis. The UIC in urine samples were detected by inductively coupled plasma-mass spectrometry (ICP-MS) (Agilent Technologies, Inc., Tokyo, Japan) (41). The vitamin D detection includes vitamin D3 (cholecalciferol) and vitamin D2 (ergocalciferol) levels, which were measured by 25-(OH)D in serum. The 25-(OH)D levels were measured using Ultrahigh Performance Liquid Chromatography-Tandem Mass Spectrometry (UPLC-MS/MS; UPLC-MS/MS ACQUITY UPLC I-class Xevo TQD, Waters Corporation) on the basis of the manufacturer's protocol. Serum was separated from fasting blood samples by spinning at 3,000 rpm for 15 min in 4°C. The laboratory data of 25-(OH) D, FT3, FT4, TSH, FPG, and lipid profiles were detected using an automatic biochemical analyzer (Beckman Coulter, Inc., CA, USA). TPOAb, TGAb and TG were detected using an automatic biochemical analyzer (Roche Ltd., Switzerland). All the laboratory data were measured in serum, except for UIC. All the factors mentioned above are summarized in Table 1.

Machine learning modeling of HT

A total of 275 controls and 361 HT patients were recruited to construct the HT machine learning model with Python Software (version 3.8). Multiple factors were selected including gender, age, BMI, diabetes, 25-(OH)D, UIC, FT3, FT4, TSH, FPG, TC, and TAG.

Six machine learning models, including K-nearest neighbor classifier (KNN), logistic regression (LR), support vector machine (SVM), decision tree model (DT), multilayer perceptron network (MLP), and eXtreme Gradient Boosting (XGBoost) models were utilized. StratifiedKFold (k= 5) was used to divide samples into the training set and the test set. After the training, the six models were evaluated in the test set. The accuracy, precision, recall (sensitivity), F1 scores, and area under curve (AUC) scores of each model were calculated, respectively. The selected features were ranked in sequence of their importance for the XGBoost model. The receiver operating

TABLE 1 Clinical characteristics of controls and HT patients.

characteristic curve (ROC) of each model displayed the judgment ability of the model.

Statistical analysis

Statistical analyses were performed using SPSS 27.0 and Graphpad Prism 8.0 software. Significant difference analysis of unordered variables and rank variables in distribution between independent groups were separately assessed with the Chisquare test and the Ridit test. Significant difference analysis of continuous variables between the two groups were separately compared by the Mann–Whitney U test for large sample sizes and the Kolmogorov-Smirnov test for small sample sizes. Significant difference analysis of continuous variables among triple groups were compared by the Kruskal–Wallis H test. The quantitative data were displayed as mean ± standard deviation.

Binary logistic regression analysis was performed to investigate the risk factors for HT development by SPSS software with the Hosmer-Lemeshow test. The odds ratio (OR) corresponded to the correlation between factors and HT development. OR > 1 represented the risk factor triggering HT, while OR < 1 represented the protective factor restricting HT. The analysis results were adjusted for each factor and considered statistically significant when *P*-values of tests were less than 0.05.

Results

Cohorts of participants

Among total 1,344 participants, 1,303 non-HT controls and first-visit HT patients were enrolled after screening. According

Character	Control (n=866)	HT-C (n=393)	HT+C (n=44)	P-value
Goiter degree				<0.0001
None	859 (99.19%)	36 (9.16%) ↓	1 (2.27%) ↓	
Ι	3 (0.35%)	183 (46.56%) ↑	13 (29.55%) ↑	
II	2 (0.23%)	165 (41.98%) ↑	29 (65.91%) ↑	
III	2 (0.23%)	9 (2.29%) ↑	1 (2.27%) ↑	
Age (years)	54.06 ± 13.50	47.66 ± 13.09 ↓	51.45 ± 11.86 ↓	< 0.0001
BMI (kg/m ²)	24.49 ± 3.31	23.15 ± 3.55 ↓	24.64 ± 3.31 ↑	< 0.0001
Gender				< 0.0001
Male	508 (58. 66%)	49 (12.47%) ↓	8 (18.18%) ↓	
Female	358 (41.34%)	344 (87.53%) ↑	36 (81.82%) ↑	
Diabetes				0.0433
Yes	28 (3.23%)	24 (6.11%) ↑	3 (6.82%) ↑	
No	838 (96.77%)	369 (93.89%)↓	41 (93.18%) ↓	

The table shows the statistics of clinical characteristics and laboratory results of controls, HT patients and patients with thyroid cancer. There were significant differences in most factors between controls and HT patients. BMI, body mass index. P-values were calculated by Kruskal–Wallis H test or Chi-square test among the triple groups. Symbols " \downarrow and \uparrow " represent "reduced and increased" change in mean values compared to corresponding controls.

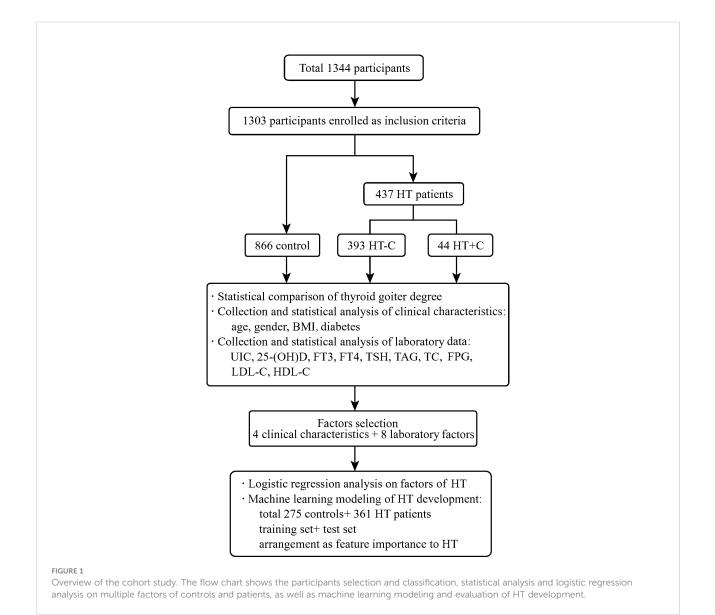
to ultrasound imaging and cytopathological examinations, they were divided into 866 controls and 437 patients. The 437 diagnosed HT patients were divided into 393 HT patients without cancer (HT-C) and 44 HT patients with thyroid cancer (HT+C). Then, multiple clinical characteristics and laboratory test data were collected as shown in Figure 1.

Clinical characteristics of participants

Compared to controls, the significant differences in goiter degrees between HT-C and HT+C patients are shown in Table 1. It indicates that 357 HT-C patients (90.83%) and 43 HT+C patients (97.73%) developed goiter while only 7 controls (0.81%) had goiter. Specifically, Degree I of goiter mostly appeared in 183

HT-C patients (46.56%), Degree II of goiter mostly appeared in 29 HT+C patients (65.91%), Degree III of goiter mostly appeared in 9 HT-C patients (2.29%). In contrast, Degrees I to III of goiter among controls were 3, 2, and 2, which all accounted for the least (0.35%, 0.23%, and 0.23%, respectively). Therefore, the proportions of goiter degrees were significantly different between the HT-C and HT+C patients compared to the controls.

Among HT patients, 344 of 393 (87.53%) HT-C and 36 of 44 (81.82%) HT+C patients were women while 358 of 866 (41.34%) controls were women. 24 of 393 (6.11%) HT-C and 3 of 44 (6.82%) HT+C patients were incorporated with diabetes while 28 of 866 (3.23%) controls had diabetes. Statistics on significant tests indicated that women (P < 0.0001) and diabetic patients (P = 0.0433) were more likely to develop HT disease. Besides, compared to controls, the clinical characteristics, including age, BMI, gender,



and diabetes all played significant roles in HT-C patients. However, only BMI had a statistical difference between HT-C and HT+C patients. The average BMI of HT-C was significantly lower than those in controls and HT+C patients. Significant difference in continuous variables were compared by the Mann–Whitney U test between the two groups and the Kruskal–Wallis H test among the three groups. Significant difference in categorical variables were compared by the Chi-square test.

Laboratory test data of participants

Compared to controls, the 437 HT patients significantly had advantageous distributions in averages of the following factors: UIC (162.13 ± 110.58) µg/L vs. (243.99 ± 203.27) µg/L, FT3 (4.71 ± 1.01) nmol/L vs. (5.04 ± 2.01) nmol/L, TSH (2.86 ± 2.31) uIU/mL vs. (7.07 ± 14.69) uIU/mL, LDL-C (2.67 ± 0.80) mmol/L vs. (2.81 ± 0.82) mmol/L, separately. On the contrary, the significant inferior factors in HT patients were as follows: age (54.06 ± 13.50) years vs. (48.04 ± 13.01) years, BMI (24.49 \pm 3.31) kg/m² vs. (23.30 \pm 3.55) kg/m², 25-(OH)D (20.72 \pm 7.35) ng/mL vs. (15.03 ± 10.37) ng/mL, FT4 (16.41 ± 2.35) nmol/L vs. (16.36 ± 7.71) nmol/L, TAG (1.66 ± 1.09) mmol/L vs. (1.36 \pm 1.13) mmol/L, TC (4.70 \pm 0.99) mmol/L vs. (4.44 \pm 0.89) mmol/L and FPG (5.66 \pm 1.34) mmol/L vs. (5.27 \pm 1.13) mmol/L. However, the HDL-C level of (1.35 ± 0.33) mmol/L in controls had no statistical significance with (1.44 ± 0.97) mmol/L in HT patients. Among laboratory data of HT-C patients, UIC, TSH, and LDL-C are significantly higher while 25-(OH)D, TAG, TC, and FPG are significantly lower than those of the controls. Nevertheless, only TSH was significant different compared to HT+C patients (Figure 2). In conclusion, apart from HDL-C, most clinical characteristics indeed have diagnostic values for HT while they have no indicative meaning for thyroid cancer.

Logistic regression analysis on risk factors for HT development

The selected factors collected from 275 controls and 361 HT patients were independently analyzed for the HT risk factors with binary logistic regression. The OR corresponded to the probability that factors may promote HT (OR > 1) or restrict HT development (OR < 1). The adjusted result shows that diabetes (OR = 6.617, P < 0.001), UIC (OR = 1.001, P = 0.005), FT3 (OR = 1.721, P < 0.001), and TSH (OR = 1.128, P < 0.001) significantly have risk effects on HT development. In contrast, gender (OR = 0.264, P < 0.001), 25-(OH)D (OR = 0.945, P < 0.001) and FPG (OR = 0.664, P < 0.001) may protect HT development. Thus, diabetes, UIC, FT3, and TSH can be identified as important risk factors for HT development (Table 2).

Machine learning modeling of HT and thyroid cancer development

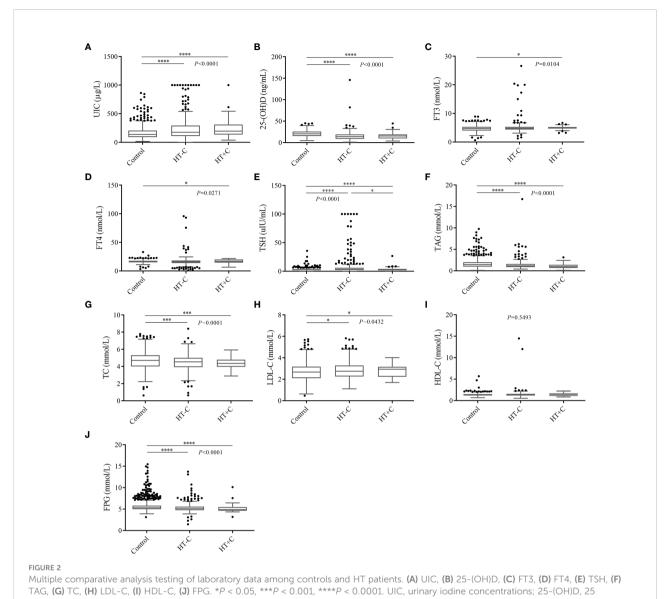
A total 275 controls and 361 HT patients with complete clinical characteristics and laboratory results were selected for modeling. Six models were established by different machine learning algorithms. The ROC curves of six models are displayed in Figure 3B, among which the XGBoost model has the highest AUC score (0.781673). Also, the AUC scores of LR (0.767496), SVM (0.766069), and MLP (0.775333) models were superior. However, the AUC scores of KNN and DT models were lower than 0.75, which were not practicable for HT prediction. The performance of six machine learning models verifies that the XGBoost model is the best in accuracy (0.729774) and precision (0.770717) (Table 3). The following features were arranged according to their importance for the XGboost model, including 25-(OH)D, FT4, TSH, FT3, TC, UIC, FPG, gender, TAG, BMI, age, and diabetes (Figure 3A). Among all features, 25-(OH)D played the most important role in HT risk model construction.

For the purpose of further studying whether these risk factors are also available for thyroid cancer developing from HT, machine learning models were established between 318 HT-C patients and 43 HT+C patients using the similar methods above. However, the AUC score of the most optimal model in the test set was lower than 0.65, which indicated that these factors are not suitable for machine learning of thyroid malignancy in our cohort (Additional file 3).

Discussion

Presently, HT is the most common chronic autoimmune disease worldwide (42), which has the risk of occurring with thyroid cancer or deteriorating into thyroid cancer, such as PTC (43-45). Apart from thyroid cancer, HT may also lead to other disorders (46, 47). However, some difficulties still exist in the diagnosis of HT. For example, ultrasonography is an essential and prevalent noninvasive tool for diagnosis of HT (48), but the correct diagnosis of HT is based on the extensive clinical experience of physicians. Besides, it is difficult for ultrasonography to distinguish HT from other diseases when HT coexists with other thyroid disorders (49). Conversely, the diagnosis of HT based on an invasive cytological examination of ultrasound-guided fine-needle biopsy or surgical samples should also be optimized (49), since the examination process needs at least three days and the recognition of marked cells needs more time as well as specialized knowledge of cytology. Thus, it is urgent to develop new precise diagnostic methods for earlystage HT.

In our study, the cases of thyroid goiter accounted for 0.81%, 90.83%, and 97.73% of the 866 non-HT controls, 357 HT-C

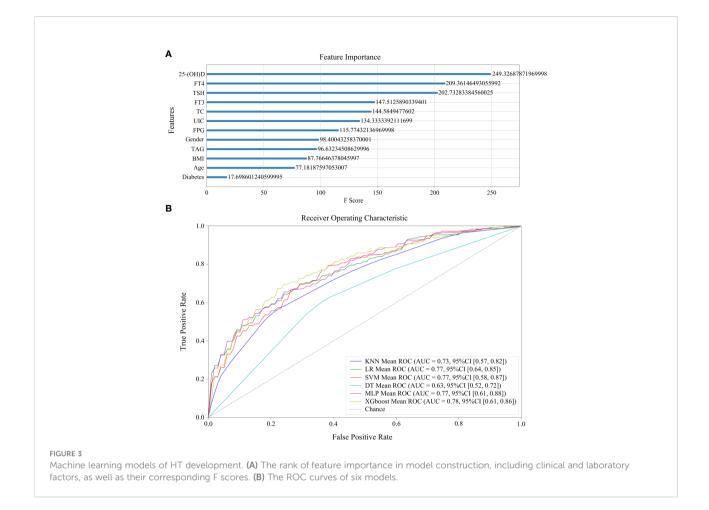


hydroxyvitamin D; FT3, free triiodothyronine; FT4, free thyroxine; TSH, thyroid stimulating hormone; TAG, triglyceride; TC, total cholesterol; LDL-C, low density lipoprotein cholesterol; HDL-C, high density lipoprotein cholesterol; FPG, fasting plasma glucose.

TABLE 2 Logistic regression analysis on risk factors for HT.

Characters	Wald	P-value	OR	95% CI for OR	
				Lower	Upper
Gender	33.053	< 0.001	0.264	0.167	0.415
Diabetes	12.962	< 0.001	6.617	2.365	18.512
UIC	7.804	0.005	1.001	1.000	1.003
25-(OH)D	19.761	< 0.001	0.945	0.922	0.969
FT3	15.929	< 0.001	1.721	1.318	2.246
TSH	11.203	< 0.001	1.128	1.051	1.211
FPG	12.804	<0.001	0.664	0.531	0.831

The adjusted quantification of relative risk relations between factors and HT. OR, odds ratio; UIC, urinary iodine concentrations; 25-(OH)D, 25 hydroxyvitamin D; FT3, free triiodothyronine; TSH, thyroid stimulating hormone; FPG, fasting plasma glucose.



patients, and 43 HT+C patients, respectively. Besides, 44 (9.59%) out of 49 HT patients also had thyroid cancer. Thus, the thyroid goiter significantly has a positive relationship with HT development. The incidence of HT is clinically insidious and sometimes asymptomatic and it has been reported that a number of HT patients (38.80%) were significantly asymptomatic (50). If the thyroid goiter is not adequately investigated or patients are asymptomatic, it is likely to remain undiagnosed though being in an early phase of HT. Therefore, correctly distinguishing benign and malignant thyroid goiters is important for the treatment and prognosis assessment of HT.

There were multiple factors that significantly changed between HT patients and controls in our study. Firstly, the levels of TSH elevated likewise in HT patients compared to controls. The FT4 level significantly reduced while the FT3 level significantly elevated in HT patients compared to controls. In order to diagnose HT more conveniently and quickly, it is preferred to simultaneously inspect the thyroid functions, hormones, and antibodies in the blood, including TSH, FT3, and FT4. It was reported that in the HT patients, there may be a clinical manifestation of elevated TSH, along with poor FT3 and FT4 levels (12, 13, 19, 23, 32). Secondly, there has been much

Models	Accuracy	Precision	Recall	F1 Score	AUC Score
KNN	0.681029	0.746247	0.681507	0.710569	0.728307
LR	0.701378	0.730701	0.780822	0.749290	0.767496
SVM	0.710839	0.720756	0.825304	0.766319	0.766069
DT	0.643135	0.680359	0.706202	0.691899	0.633101
MLP	0.721863	0.746918	0.797603	0.767313	0.775333
XGBoost	0.729774	0.770717	0.772755	0.767228	0.781673

TABLE 3 Performance summary of six machine learning models of HT.

AUC, area under curve; KNN, k-nearest neighbor classifier; LR, logistic regression; SVM, a support vector machine; DT, the decision tree model; MLP, the multilayer perceptron network; XGBoost, eXtreme Gradient Boosting.

research on the physiological changes in the blood and urine of HT patients (18, 51, 52). For the HT patients we studied, the average 25-(OH)D level was significantly reduced while UIC was significantly elevated compared to controls. Micronutrients, such as iodine and vitamin D, are often found to be deficient in HT patients (17, 53). In our study, the levels of FPG, TAG, and TC were all significantly decreased while LDL-C was significantly increased in HT patients compared to controls. Blood glucose and blood lipid, including FPG, TAG, TC, and LDL-C are indeed found to be decreased in HT patients (24-26). Thirdly, except for laboratory results, many clinical characteristics are also associated with HT development. In our study, we found that gender, age, and BMI were significantly correlated with HT. Diabetes also had a significant positive correlation with HT. According to previous studies, middle-aged women are more likely to suffer from HT than men (32). In general, the average BMI of HT patients tends to be lower than controls (53). Also, patients with an endocrine disorder, such as diabetes, also have a potential HT risk (25). However, many other autoimmune disorders are associated to HT, such as rheumatic disorders, autoimmune hemolytic anemia, and immune thrombocytopenic purpura, which may become potential risk factors for HT development (54-57).

As for thyroid cancer, in our study, only BMI and TSH had significantly positive impacts on the development of thyroid cancer compared to controls. Most thyroid cancer patients had the same altered tendency as HT patients, except for BMI and FT4. Previous studies show there are obvious elevated levels of TG, TGAb, TPOAb, and disordered TSH in patients with thyroid cancer (58–60), which were not analyzed in this study. Besides, poor prognostic clinical characteristics of thyroid cancer also include older age and large tumor size, which are usually detected in middle-aged women (39). However, these factors had no difference in our study.

Moreover, among the factors above, we found four factors that can be considered risk factors for HT development by adjusted logistic regression analysis: diabetes, UIC, FT3, and TSH. Finally, we utilized multiple factors to conduct machine learning modeling of HT. The verified risk factors inducing HT were analyzed in this study. Six predictive models were evaluated in the test set and the XGBoost model had the highest accuracy, precision, and AUC score. Also, LR, SVM, and MLP models also showed good performance. To date, there have been few studies on machine learning modeling of HT with multiple factors (32, 42). However, the machine learning modeling was not available to distinguish HT+C from HT with these factors. The arrangement of feature importance showed that 25-(OH)D was the best risk factor to indicate HT in comparison with the others. FT4 and TSH were the next two important factors of HT. We thereby suggest that deficient nutrition uptakes, including vitamin D and iodine, along with multiple factors related to thyroid

hormones and lipid profiles, have great significance in the early diagnosis and risk prediction of HT. The usual diagnostic procedure of HT is long and it is difficult to confirm HT at early stage by ultrasound or cytological examinations. This study was the first time comprehensive factors of HT were taken into consideration. These factors made a superior contribution to the reliable performance of our models. From the perspective of the AUC score (0.781673), accuracy (0.729774), and precision (0.770717) of the XGBoost model, it is relatively valuable for the evaluation of HT risk and clinical decision. The AUC scores of LR, SVM, and MLP models are all 0.77, which indicate robust predictive performance of these multiple factors. The previous diagnosis models of HT were based on TPOAb and TGAb, which prompted AUC scores to 0.8 (32). However, since TPOAb and TGAb are the diagnostic criteria for HT, they were not suitable risk factors in our prediction modeling of HT. Moreover, the research may help identify novel preventative and therapeutic factors for HT patients and even in patients presenting with asymptomatic euthyroidism or subclinical hypothyroidism.

Nevertheless, there are still some further challenges to address. Because of the limited cohort study, a more clinical database in a larger cohort of HT patients should be collected to verify and improve the performance of the proposed models of machine diagnosis. In addition, for optimal machine learning modeling that distinguishes thyroid cancer from HT disease, other clinical characteristics and laboratory test data should be taken into account. Considering that HT is an autoimmune disease, lymphocytes and intestinal microbiota should be meaningful risk factors for clinical reference (33).

Conclusions

In summary, we collected clinical information of HT patients and non-HT controls. Firstly, we found that the counts of thyroid goiter were significantly different among controls, patients with HT and thyroid cancer. Secondly, comprehensive factors including age, BMI, gender, diabetes, UIC, 25-(OH)D, FT3, FT4, TSH, TAG, TC, FPG, and LDL had significant correlations with HT, among which diabetes, UIC, FT3, and TSH were confirmed as important risk factors for HT development. Besides, XGBoost, LR, SVM and MLP models displayed appropriate predictive power in the machine learning modeling of HT. Among all the features, 25-(OH)D, FT4, and TSH played important roles in HT risk model construction. Our study firstly demonstrates that comprehensive examinations along with machine learning modeling can enhance the precision and efficiency of thyroid diagnosis, which may also enlighten better prevention and new therapeutic schemes for thyroid disorders.

Data availability statement

The raw data supporting the conclusions of this article will be made available by the authors, without undue reservation.

Ethics statement

The studies involving human participants were reviewed and approved by the Ethics Committee of Xuchang Central Hospital, Kaifeng Central Hospital and Jincheng People's Hospital. Written informed consent to participate in this study was provided by the participants' legal guardian/next of kin.

Author contributions

PeL, FL, and MZ conceived the study, collected samples and wrote the manuscript. SX, PiL, JC, DT, YT, LZ, and XC performed laboratory experiments and collected data. YP analyzed data and established the machine learning models. HT provided guidance on data analysis and writing revision. YW was in charge of clinical database analysis and wrote the manuscript. YS edited and revised the manuscript. All the authors read and approved the final manuscript.

Conflict of interest

Authors YP and HT were employed by company Shanghai Biotecan Pharmaceuticals Co., Ltd.

The remaining authors declare that the research was conducted in the absence of any commercial or financial relationships that could be construed as a potential conflict of interest.

Publisher's note

All claims expressed in this article are solely those of the authors and do not necessarily represent those of their affiliated organizations, or those of the publisher, the editors and the reviewers. Any product that may be evaluated in this article, or claim that may be made by its manufacturer, is not guaranteed or endorsed by the publisher.

Supplementary material

The Supplementary Material for this article can be found online at: https://www.frontiersin.org/articles/10.3389/ fendo.2022.886953/full#supplementary-material

ADDITIONAL FILE 1

Thyroid ultrasound images of HT and PTC patients compared to healthy controls. (A) The normal thyroid ultrasound image of controls. (B) The thyroid ultrasound image of HT patients. (C) The thyroid ultrasound image of PTC patients.

ADDITIONAL FILE 2

Specimens from healthy controls, patients with HT and PTC (HE staining, x200). (A) Control shows normal thyroid follicular epithelial cells (EC), thyroid parafollicular cells (PC) and follicles filled with colloid (FC). (B) Typical changes of HT, including thyroid follicles with Hürthle-cell metaplasia (H), plasma cells (P) and lymphoid follicles with germinal centers (G). (C) Typical changes of PTC, such as papillary areas (PA) and follicular variant (FV).

ADDITIONAL FILE 3

Machine learning modeling of thyroid cancer deteriorated from HT. The ROC curves of six models were shown with different colors. KNN, knearest neighbor classifier; LR, logistic regression; SVM, a support vector machine; DT, the decision tree model; MLP, the multilayer perceptron network; XGBoost, eXtreme Gradient Boosting.

References

1. Rotondi M, Carbone A, Coperchini F, Fonte R, Chiovato L. Diangnosis of endocrine disease: IgG4-related thyroid autoimmune disease. *Eur J Endocrinol* (2019) 180:R175-83. doi: 10. 1530/EJE-18-1024

2. Sánchez JF. TI-RADS classification of thyroid nodules based on a score modified according to ultrasound criteria for malignancy. (2014) 78(3):138-48.

3. Horvath E, Majlis S, Rossi R, Franco C, Niedmann JP, Castro A, et al. An ultrasonogram reporting system for thyroid nodules stratifying cancer risk for clinical management. *J Clin Endocrinol Metab* (2009) 94(5):1748–51. doi: 10.1210/ jc.2008-1724

4. Zayadeen AR, Abu-Yousef M, Berbaum K. JOURNAL CLUB: Retrospective evaluation of ultrasound features of thyroid nodules to assess malignancy risk: A step toward TIRADS. *AJR Am J Roentgenol* (2016) 207(3):460–9. doi: 10.2214/AJR.15.15121

5. Zhou J, Yin L, Wei X, Zhang S, Song Y, Luo B, et al. 2020 Chinese guidelines for ultrasound malignancy risk stratification of thyroid nodules: the c-TIRADS. *Endocrine* (2020) 70(2):256–79. doi: 10.1007/s12020-020-02441-y

6. Bomeli SR, LeBeau SO, Ferris RL. Evaluation of a thyroid nodule. Otolaryngol Clin North Am (2010) 43(2):229–38. doi: 10.1016/j.otc.2010.01.002

7. Pearce EN, Farwell AP, Braverman LE. Thyroiditis. N Engl J Med (2003) 348 (26):2646-55. doi: 10.1056/NEJMra021194

8. Fiore E, Rago T, Latrofa F, Provenzale MA, Piaggi P, Delitala A, et al. Hashimoto's thyroiditis is associated with papillary thyroid carcinoma: role of TSH and of treatment with 1-thyroxine. *Endocr Relat Cancer* (2011) 18(4):429–37. doi: 10.1530/ERC-11-0028

9. Antonelli A, Ferrari SM, Corrado A, Di Domenicantonio A, Fallahi P. Autoimmune thyroid disorders. *Autoimmun Rev* (2015) 14(2):174–80. doi: 10.1016/j.autrev.2014.10.016

10. Rousset B, Dupuy C, Miot F, Dumont JJE. Thyroid hormone synthesis and secretion. In: *Endotext*. (2015) South Dartmouth (MA): MDText.com, Inc. (2015).

11. Caron P. Neurocognitive outcomes of children secondary to mild iodine deficiency in pregnant women. *Ann Endocrinol (Paris)* (2015) 76(3):248–52. doi: 10.1016/j.ando.2015.01.001

12. Kawicka A, Regulska-Ilow B, Regulska-Ilow B. Metabolic disorders and nutritional status in autoimmune thyroid diseases. *Postepy Hig Med Dosw (Online)* (2015) 69:80–90. doi: 10.5604/17322693.1136383

13. Feldt-Rasmussen U, Klose M, Benvenga S. Interactions between hypothalamic pituitary thyroid axis and other pituitary dysfunctions. *Endocrine* (2018) 62(3):519–27. doi: 10.1007/s12020-018-1738-6

14. Leung AM, Braverman LE. Consequences of excess iodine. Nat Rev Endocrinol (2014) 10(3):136-42. doi: 10.1038/nrendo.2013.251

15. Zimmermann MB, Galetti V. Iodine intake as a risk factor for thyroid cancer: a comprehensive review of animal and human studies. *Thyroid Res* (2015) 8:8. doi: 10.1186/s13044-015-0020-8

16. Luo J, Hendryx M, Dinh P, He K. Association of iodine and iron with thyroid function. *Biol Trace Elem Res* (2017) 179(1):38–44. doi: 10.1007/s12011-017-0954-x

17. Knezevic J, Starchl C, Tmava Berisha A, Amrein K. Thyroid-Gut-Axis: How does the microbiota influence thyroid function? *Nutrients* (2020) 12(6). doi: 10.3390/nu12061769

18. Liontiris MI, Mazokopakis EE. A concise review of hashimoto thyroiditis (HT) and the importance of iodine, selenium, vitamin d and gluten on the autoimmunity and dietary management of HT patients. points that need more investigation. *Hell J Nucl Med* (2017) 20(1):51–6. doi: 10. 1967/s002449910507

19. Ihnatowicz P, Drywien M, Wator P, Wojsiat J. The importance of nutritional factors and dietary management of hashimoto's thyroiditis. *Ann Agric Environ Med* (2020) 27(2):184–93. doi: 10.26444/aaem/112331

20. Moodley K, Botha J, Raidoo DM, Naidoo S. Immuno-localisation of antithyroid antibodies in adult human cerebral cortex. *J Neurol Sci* (2011) 302(1-2):114–7. doi: 10.1016/j.jns.2010.11.027

21. Visser WE, Peeters RP. Interpretation of thyroid function tests during pregnancy. *Best Pract Res Clin Endocrinol Metab* (2020) 34(4):101431. doi: 10.1016/j.beem.2020.101431

22. Hadizadeh Riseh S, Abbasalizad Farhang M, Mobasseri M, Asghari Jafarabadi M. The relationship between thyroid hormones, antithyroid antibodies, anti-tissue transglutaminase and anti-gliadin antibodies in patients with hashimoto's thyroiditis. *Acta Endocrinol (Buchar)* (2017) 13(2):174–9. doi: 10.4183/aeb.2017.174

23. Korevaar TIM, Minguez-Alarcon L, Messerlian C, de Poortere RA, Williams PL, Broeren MA, et al. Association of thyroid function and autoimmunity with ovarian reserve in women seeking infertility care. *Thyroid* (2018) 28(10):1349–58. doi: 10.1089/thy.2017.0582

24. Fenneman AC, Rampanelli E, Yin YS, Ames J, Blaser MJ, Fliers E, et al. Gut microbiota and metabolites in the pathogenesis of endocrine disease. *Biochem Soc Trans* (2020) 48(3):915–31. doi: 10.1042/BST20190686

25. Watanabe T, Temma Y, Okada J, Yamada E, Saito T, Osaki A, et al. In patients with type 2 diabetes the presence of hashimoto's thyroiditis reduces the beneficial effect of dipeptidyl peptidase-4 inhibitor on plasma glucose control. *Endocr J* (2021) 68(5):599–603. doi: 10.1507/endocrj.EJ20-0620

26. Lei Y, Yang J, Li H, Zhong H, Wan Q. Changes in glucose-lipid metabolism, insulin resistance, and inflammatory factors in patients with autoimmune thyroid disease. *J Clin Lab Anal* (2019) 33(7):e22929. doi: 10.1002/jcla.22929

27. Ma L, Ma C, Liu Y, Wang X. Thyroid diagnosis from SPECT images using convolutional neural network with optimization. *Comput Intell Neurosci* (2019) 2019:1–11. doi: 10.1155/2019/6212759

28. Chen Y, Li D, Zhang X, Jin J, Shen Y. Computer aided diagnosis of thyroid nodules based on the devised small-datasets multi-view ensemble learning. *Med Image Anal* (2021) 67:101819. doi: 10.1016/j.media.2020.101819

29. Kang S, Lee E, Chung CW, Jang HN, Moon JH, Shin Y, et al. A beneficial role of computer-aided diagnosis system for less experienced physicians in the diagnosis of thyroid nodule on ultrasound. *Sci Rep* (2021) 11(1):20448. doi: 10. 1038/s41598-021-99983-6

30. Han M, Ha EJ, Park JH. Computer-aided diagnostic system for thyroid nodules on ultrasonography: Diagnostic performance based on the thyroid imaging reporting and data system classification and dichotomous outcomes. *AJNR Am J Neuroradiol* (2021) 42(3):559–65. doi: 10.3174/ajnr.A6922

31. Chiou SC, Hsieh MH, Chen HY, Lin JD, Chen CC, Hsu WH, et al. The reproducibility of the virtual organ computer-aided analysis program for evaluating 3-dimensional power Doppler ultrasonography of diffuse thyroid disorders. J Endocrinol Invest (2009) 32(2):139–46. doi: 10.1007/BF03345703

32. Cui Z, Wang Z, Liu X, Cai Y, Xu X, Yang T. Establishment of clinical diagnosis model of graves' disease and hashimoto's thyroiditis. *J Transl Med* (2019) 17(1):11. doi: 10.1186/s12967-018-1765-3

33. Salazar-Viedma M, Vergano-Salazar JG, Pastenes L, D'Afonseca V. Simulation model for hashimoto autoimmune thyroiditis disease. *Endocrinology* (2021) 162(12):bqab190. doi: 10.1210/endocr/bqab190

34. Effraimidis G, Wiersinga WM. Mechanisms in endocrinology: autoimmune thyroid disease: old and new players. *Eur J Endocrinol* (2014) 170(6):R241–52. doi: 10.1530/EJE-14-0047

35. Weetman A. Autoimmune thyroid disease. *Endocrine* (2020) 68(2):258–60. doi: 10.1007/s12020-020-02188-6

36. Wiersinga WM. Hashimoto's thyroiditis. In: P Vitti and L Hegedüs, editors. *Thyroid diseases, endocrinology.* Cham: Springer (2018).

37. Lin J-D. Thyroid cancer in thyroid nodules diagnosed using ultrasonography and fine needle aspiration cytology. *J Med Ultrasound* (2010) 18 (3):91–104. doi: 10.1016/S0929-6441(10)60014-8

38. Harvey AM, Truong LD, Mody DR. Diagnostic pitfalls of hashimoto's/ lymphocytic thyroiditis on fine-needle aspirations and strategies to avoid overdiagnosis. *Acta Cytol* (2012) 56(4):352-60. doi: 10.1159/000338738

39. LiVolsi VA. Papillary thyroid carcinoma: an update. *Mod Pathol* (2011) 24 (Suppl 2):S1–9. doi: 10.1038/modpathol.2010.129

40. Hunt JL, Barnes EL. Non-tumor-associated psammoma bodies in the thyroid. Am J Clin Pathol (2003) 119(1):90–4. doi: 10.1309/RWPPYCBYT2JVA023

41. Lee JH, Ji OJ, Song MJ, Park HD, Kim HK, Kim SW, et al. Determination of urinary iodine concentration by inductively coupled plasma-mass spectrometry in thyroid cancer patients on low-iodine diet. *Korean J Lab Med* (2010) 30(4):351–6. doi: 10.3343/kjlm.2010.30.4.351

42. Gan T, Randle RW. The role of surgery in autoimmune conditions of the thyroid. Surg Clin North Am (2019) 99(4):633–48. doi: 10.1016/j.suc.2019.04.005

43. Noureldine SI, Tufano RP. Association of hashimoto's thyroiditis and thyroid cancer. *Curr Opin Oncol* (2015) 27(1):21-5. doi: 10.1097/CCO.0000000000000150

44. Del Rio P, Montana Montana C, Cozzani F, Rossini M, Loderer T, Dall'Aglio E, et al. Is there a correlation between thyroiditis and thyroid cancer? *Endocrine* (2019) 66(3):538–41. doi: 10.1007/s12020-019-01935-8

45. Feldt-Rasmussen U. Hashimoto's thyroiditis as a risk factor for thyroid cancer. *Curr Opin Endocrinol Diabetes Obes* (2020) 27(5):364–71. doi: 10.1097/MED.00000000000570

46. Xerfan EMS, Facina AS, Andersen ML, Tufik S, Tomimori J. Hashimoto thyroiditis as a cause or consequence of obstructive sleep apnea. J Clin Sleep Med (2019) 15(11):1703. doi: 10.5664/jcsm.8054

47. Kumar V, Abbas AK, Fausto N, Aster JC. Robbins And cotran pathologic basis of disease, professional edition e-book. Oxford: Elsevier health sciences (2014).

48. Kangelaris GT, Kim TB, Orloff LA. Role of ultrasound in thyroid disorders. Otolaryngol Clin North Am (2010) 43(6):1209-27. doi: 10.1016/j.otc.2010.08.006

49. Wu G, Zou D, Cai H, Liu Y. Ultrasonography in the diagnosis of hashimoto's thyroiditis. Front Biosci (Landmark Ed) (2016) 21:1006-12. doi: 10.2741/4437

50. Saraf SR, Gadgil NM, Yadav S, Kalgutkar AD. Importance of combined approach of investigations for detection of asymptomatic hashimoto thyroiditis in early stage. *J Lab Physicians* (2018) 10(3):294–8. doi: 10.4103/JLP.JLP_72_17

51. Caturegli P, De Remigis A, Rose NR. Hashimoto thyroiditis: clinical and diagnostic criteria. *Autoimmun Rev* (2014) 13(4-5):391-7. doi: 10.1016/j.autrev.2014.01.007

52. Rayman MP. Multiple nutritional factors and thyroid disease, with particular reference to autoimmune thyroid disease. *Proc Nutr Soc* (2019) 78 (1):34-44. doi: 10.1017/S0029665118001192

53. Chao G, Zhu Y, Fang L. Correlation between hashimoto's thyroiditis-related thyroid hormone levels and 25-hydroxyvitamin d. *Front Endocrinol (Lausanne)* (2020) 11:4. doi: 10.3389/fendo.2020.00004

54. Giuffrida G, Bagnato G, Campenni A, Giovinazzo S, Keller KP, Alibrandi A, et al. Non-specific rheumatic manifestations in patients with hashimoto's thyroiditis: a pilot cross-sectional study. *J Endocrinol Invest* (2020) 43(1):87–94. doi: 10.1007/s40618-019-01083-w

55. Ruggeri RM, Giuffrida G, Campenni A. Autoimmune endocrine diseases. *Minerva Endocrinol* (2018) 43(3):305-22. doi: 10.23736/S0391-1977.17.02757-2

56. Ruggeri RM, Trimarchi F, Giuffrida G, Certo R, Cama E, Campenni A, et al. Autoimmune comorbidities in hashimoto's thyroiditis: different patterns of association in adulthood and childhood/adolescence. *Eur J Endocrinol* (2017) 176(2):133–41. doi: 10.1530/EJE-16-0737

57. Weetman AP. An update on the pathogenesis of hashimoto's thyroiditis. J Endocrinol Invest (2021) 44(5):883–90. doi: 10.1007/s40618-020-01477-1

58. Guo X, Chen X, Zhang C, Zhang J, Zhang C. Hyperinsulinemia and thyroid peroxidase antibody in Chinese patients with papillary thyroid cancer. *Endocr J* (2019) 66(8):731–7. doi: 10.1507/endocrj.EJ18-0358

59. Ritter A, Mizrachi A, Bachar G, Vainer I, Shimon I, Hirsch D, et al. Detecting recurrence following lobectomy for thyroid cancer: Role of thyroglobulin and thyroglobulin antibodies. *J Clin Endocrinol Metab* (2020) 105 (6):dgaa152. doi: 10.1210/clinem/dgaa152

60. Schiffmann L, Kostev K, Kalder M. Association between various thyroid gland diseases, TSH values and thyroid cancer: a case-control study. *J Cancer Res Clin Oncol* (2020) 146(11):2989–94. doi: 10.1007/s00432-020-03283-x