Cross-sectional Study

ANNALS OF MEDICINE & SURGERY

OPEN

Identification of important symptoms and diagnostic hypothyroidism patients using machine learning algorithms

Salahuddin Rakhshani Rad, MSc[a], Zahra H. Mohammadi, MD[b], Mahdieh J. Zadeh, MD[c], Mohammad A. Mosleh-Shirazi, PhD[f,e], Tania Dehesh, PhD[d,*]

Background: Hypothyroidism is one of the most common endocrine diseases. It is, however, usually challenging for physicians to diagnose due to nonspecific symptoms. The usual procedure for diagnosis of Hypothyroidism is a blood test. In recent years, machine learning algorithms have proved to be powerful tools in medicine due to their diagnostic accuracy. In this study, the authors aim to predict and identify the most important symptoms of Hypothyroidism using machine learning algorithms.

Method: In this cross-sectional, single-center study, 1296 individuals who visited an endocrinologist for the first time with symptoms of Hypothyroidism were studied, 676 of whom were identified as patients through thyroid-stimulating hormone testing. The outcome was binary (with Hypothyroidism /without Hypothyroidism). In a comparative analysis, random forest, decision tree, and logistic regression methods were used to diagnose primary Hypothyroidism.

Results: Symptoms such as tiredness, unusual cold feeling, yellow skin (jaundice), cold hands and feet, numbness of hands, loss of appetite, and weight Hypothyroidism gain were recognized as the most important symptoms in identifying Hypothyroidism. Among the studied algorithms, random forest had the best performance in identifying these symptoms (accuracy = 0.83, kappa = 0.46, sensitivity = 0.88, specificity = 0.88).

Conclusions: The findings suggest that machine learning methods can identify Hypothyroidism patients who show relatively simple symptoms with acceptable accuracy without the need for a blood test. Greater familiarity and utilization of such methods by physicians may, therefore, reduce the expense and stress burden of clinical testing.

Keywords: decision tree, hypothyroidism, logistic regression, machine learning algorithms, random Forest

Introduction

Hypothyroidism is one of the most common diseases in the world, in which insufficient thyroid hormone is produced[1]. Due to the wide variation in clinical symptoms, the definition of Hypothyroidism is mainly biochemical[2]. Ninety nine percent of primary cases of Hypothyroidism are related to deficiency of thyroxine (T4) and triiodothyronine (T3) hormones[3–5]. Deficiency in T4 and T3 hormones, which are produced by the thyroid gland, leads to increasing thyroid-stimulating hormone (TSH) production through a negative feedback mechanism[4].

HIGHLIGHTS

- Diagnosis of primary hypothyroidism using simple symptoms.
- Classification of the most important symptoms of hypothyroidism by decision tree and random forest algorithms.
- The random forest algorithm as a supervised machine learning algorithm performed better in classifying patients.

Hypothyroidism has nonspecific symptoms such as weight Hypothyroidism gain, fatigue, insufficient concentration, depression, menstrual irregularities, and constipation, which change with age, sex, and other factors[1,6]. Autoimmune thyroiditis (Hashimoto's disease) is the most common symptom of this disorder[7].

The prevalence of Hypothyroidism is 2% in the world, even in the existence of enough iodine in daily food[8]. In a cohort study that was conducted in Iran in 2017, a significant increase in the prevalence of thyroid dysfunction was reported, from 1.4 to 10.5, which was attributed to several factors such as geographical areas, aging, ethnicity, and the amount of iodine intake[9].

Increasing serum cholesterol levels and the risk of coronary artery disease and cardiovascular mortality are the most common

[a]Department of Biostatistics and Epidemiology, School of Public Health, [b]Endocrinology and Metabolism Research Center, Institute of Basic and Clinical Physiology Sciences, [c]Clinical Research Development Unit, Shahid Bahonar Hospital, [d]Modeling in Health Research Center, Institute for Futures Studies in Health, Kerman University of Medical Sciences, Kerman, [e]Department of Radio-Oncology and [f]Ionizing and Non-Ionizing Radiation Protection Research Center (INIRPRC), School of Paramedical Sciences, Shiraz University of Medical Sciences, Shiraz, Iran

Sponsorships or competing interests that may be relevant to content are disclosed at the end of this article.

*Corresponding author. Address: Kerman, The beginning of Haft Bagh Alavi Axis, University of Medical Sciences Campus, Kerman 7616913555, Islamic Republic of Iran. Tel.: +34 313 25700. E-mail: Tania_dehesh@yahoo.com (T. Dehesh).

Annals of Medicine & Surgery (2024) 86:3233–3241

Received 2 November 2023; Accepted 30 March 2024

Published online 17 April 2024

http://dx.doi.org/10.1097/MS9.0000000000002068

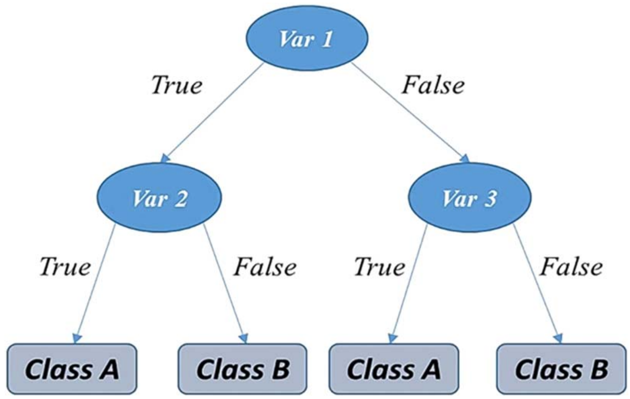

Figure 1. The decision tree structure.

complications of hypothyroidism[10]. The economic burden of Hypothyroidism is fairly high, especially in patients with other underlying diseases such as diabetes and hemodialysis[11–13]. The common clinical method for diagnosing equally primary Hypothyroidism is to check the serum concentration of TSH; people with TSH and T4 levels above the reference age range are diagnosed as hypothyroid[2,14]. The upper limit of the TSH reference range usually increases with age in adults[2].

In recent years, artificial intelligence and machine learning techniques have attracted increasing attention from medical researchers[15,16]. Among the most attractive features of machine learning in medicine are disease prediction and diagnosis of simple symptoms[17]. The prediction models such as support vector machine, decision tree (DT), random forest (RF), and artificial neural network, are among the most popular machine learning methods.

As accurate diagnosis of Hypothyroidism is currently based on the TSH level obtained by a blood test, it creates some expense burden and anxiety for patients. The first aim of the present study is to diagnose Hypothyroidism in new cases that have no history of Hypothyroidism symptoms with three statistical machine learning methods [logistic regression (LR), DT, and RF]. The diagnosis is performed using simple and widely-accepted visual symptoms of Hypothyroidism that endocrinologists identify[18]. Second, the most important visual features of Hypothyroidism that can help physicians in diagnosis, are also ranked using DT and RF methods.

Methods

In total, 1296 individuals (1088 women and 208 men) aged 18 years or over participated in this cross-sectional, single-center study from September to December 2022 at our main public clinic for thyroid treatment in Kerman, southern Iran. Inclusion criteria: physical referral to the endocrinologist because of Hypothyroidism symptoms for the first time (new case), aged 18 years or more. Exclusion criteria: having a history of Hypothyroidism treatment and thyroid gland surgery, having Hypothyroidism during previous pregnancies. We carried out no sample size calculation; The study included all people who passed the inclusion and exclusion criteria.

The data was divided into two parts, called train and test sets. Three models were first built with the train set and then their classification efficiency was investigated on the test set. The checklist included 11 questions and six demographic variables. The checklist questions each had four choices for selection: very little, little, somewhat, and much (Table 2). These questions were based on visual Hypothyroidism symptoms that an endocrinologist could diagnose without laboratory tests.

The interviewing team was trained by the endocrinologist about the concept and purpose of the checklist questions before dispatching them to the clinic. The participants completed a written informed consent form before enrollment in the study. The actual status of people was diagnosed by laboratory TSH test as Hypothyroidism or not having Hypothyroidism. The outcome was binary (people with Hypothyroidism /without Hypothyroidism). Two classification methods of machine learning (DT and RF) and a traditional classification method (LR) were used in order to diagnose new Hypothyroidism cases based on the 11 checklist questions. This study was conducted in accordance with the Declaration of Helsinki. The registration unique identifying number (UIN) was NCT06112886[19]. This study was approved by the local ethics committee of Kerman University of Medical Sciences (reference number: IR.KMU.REC.1401.292). The method section was in accordance with the strengthening the reporting of cohort, cross-sectional, and case–control studies in surgery (STROCSS) criteria[20].

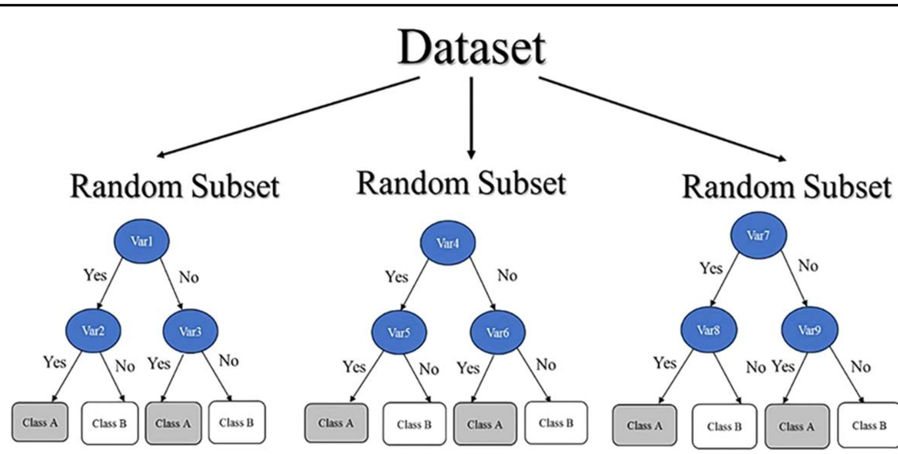

Figure 2. The random forest structure.

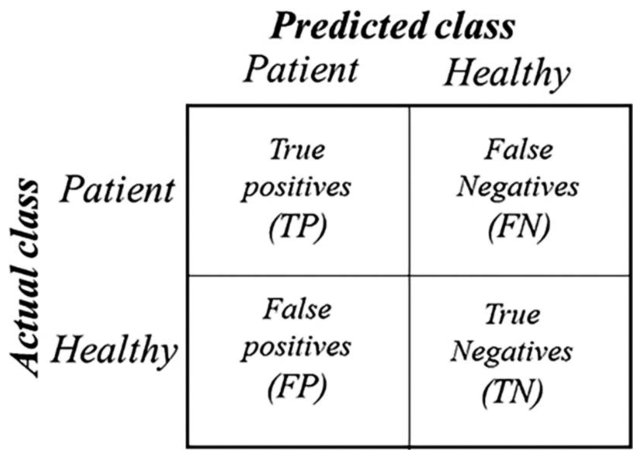

Figure 3. The basic framework of the confusion matrix.

Logistic regression method

The traditional and most well-established method for classification is LR[21]. In this method, the response variable is binary, which usually indicates the occurrence or nonoccurrence of an event, here in two classes; patient class and nonpatient (healthy) class. The LR method calculated the probability of being in the patient class as follows:

$$P(x) = \frac{\exp(\alpha + \beta x)}{1 + \exp(\alpha + \beta x)}$$

x is a vector of patients' features (checklist questions) and α, β, are LR parameters. The LR method has a cut point to discriminate between patients and nonpatients. For example, a person that has a probability higher than the cut point is in the patient class[22].

Table 1

Demographic variables in both patient and nonpatient groups

Variables	Total (1296)	Without hypothyroidism ($n=620$)	With hypothyroidism ($n=676$)	*P*
Age (mean ± SD)	36.38 ± 13.73	36.17 ± 14.38	36.6 ± 13.11	0.78
Sex, *n* (%)				
Female	1088 (84)	508 (81.9)	580 (85.8)	0.344
Male	208 (16.0)	112 (18.1)	96 (14.2)	
Education, *n* (%)				
Illiterate	196 (15.1)	84 (13.5)	112 (16.6)	0.621
Prehigh-school diploma	168 (13.0)	240 (38.7)	284 (42.0)	
High-school diploma	524 (40.4)	216 (34.8)	192 (28.4)	
College education	408 (31.5)	80 (12.9)	88 (13)	
Marital status, *n* (%)				
Single	264 (20.4)	152 (24.5)	112 (16.6)	0.076
Married	1032 (79.6)	468 (75.5)	564 (83.4)	
Family history of hypothyroidism, *n* (%)				
Yes	636 (49.1)	280 (45.2)	356 (52.7)	0.177
No	660 (50.9)	340 (54.8)	320 (47.3)	
Employment, *n* (%)				
Unemployed	760 (58.6)	340 (54.8)	420 (62.1)	0.407
Public sector job	188 (14.5)	100 (16.1)	88 (13)	
Private sector job	348 (26.9)	180 (29.0)	168 (24.9)	

Table 2

Checklist questions

Checklist's items	With hypothyroidism, *n* (%)	Without hypothyroidism, *n* (%)	*P*
Q1: Have you had unusual fatigue or weakness during the last few months?			
Very little	36 (15.8)	192 (84.2)	< 0.001
Little	72 (34.0)	140 (66.0)	
Somewhat	244 (60.4)	160 (39.6)	
Much	324 (71.7)	128 (28.3)	
Q2: Have you had dry skin during the last few months?			
Very little	52 (22.0)	184 (78.0)	< 0.001
Little	76 (33.9)	148 (66.1)	
Somewhat	264 (62.9)	156 (37.1)	
Much	280 (68.0)	132 (32.0)	
Q3: Have you recently felt unusually cold?			
Very little	108 (31.4)	236 (68.6)	< 0.001
Little	196 (50.0)	196 (50.0)	
Somewhat	196 (60.5)	128 (39.5)	
Much	176 (74.6)	60 (25.4)	
Q4: Have you had a loss of appetite and weight gain in the past months?			
Very little	92 (30.7)	208 (69.3)	< 0.001
Little	124 (38.8)	196 (61.3)	
Somewhat	220 (60.4)	144 (39.6)	
Much	240 (76.9)	72 (23.1)	
Q5: Have you had cold hands and feet in recent months?			
Very little	72 (23.4)	236 (76.6)	< 0.001
Little	92 (37.7)	152 (62.3)	
Somewhat	296 (69.2)	132 (30.8)	
Much	216 (68.4)	100 (31.6)	
Q6: Have you had excessive menstrual bleeding in recent months?			
Very little	300 (50.7)	292 (49.3)	0.014
Little	92 (42.6)	124 (57.4)	
Somewhat	88 (66.7)	44 (33.3)	
Much	108 (73.0)	40 (27.0)	
Q7: Have you had numbness in your hands in recent months?			
Very little	60 (20.0)	240 (80.0)	< 0.001
Little	56 (31.8)	120 (68.2)	
Somewhat	224 (62.9)	132 (37.1)	
Much	336 (72.4)	128 (27.6)	
Q8: Have you experienced Yellowness in the color of your skin in recent months?			
Very little	124 (31.6)	268 (68.4)	< 0.001
Little	120 (37.5)	200 (62.5)	
Somewhat	240 (69.0)	108 (31.0)	
Much	188 (81)	44 (19)	
Q9: Have you had ingrown hairs and brittle nails in recent months?			
Very little	116 (35.8)	208 (64.2)	< 0.001
Little	116 (35.8)	208 (64.2)	
Somewhat	212 (65.4)	112 (34.6)	
Much	232 (71.6)	92 (28.4)	
Q10: Have you had swelling around the eyes or swelling in the legs in recent months?			
Very little	172 (36.1)	304 (63.9)	< 0.001
Little	144 (50.0)	144 (50.0)	
Somewhat	176 (61.1)	112 (38.9)	
Much	184 (75.4)	60 (24.6)	
Q11: Have you been constipated in recent months?			
Very little	156 (40.6)	228 (59.4)	< 0.001
Little	104 (38.2)	168 (61.8)	
Somewhat	204 (60.0)	136 (40.0)	
Much	212 (70.7)	88 (29.3)	

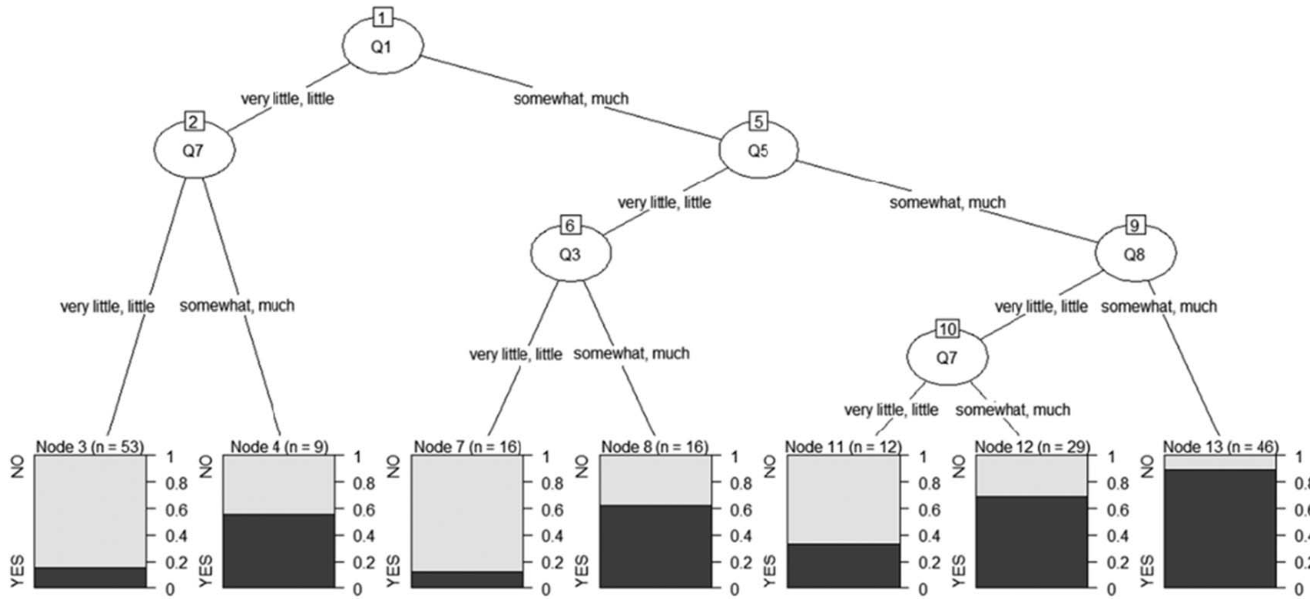

Figure 4. The classification tree.

Decision tree

DT is a practical machine learning algorithm, which is used in various fields such as industry, marketing, and medicine. A tree diagram includes root, branches, nodes, and leaves. This diagram is drawn from top to bottom so that the root node is located at the highest part of the tree. Then, this node is divided into several branches. Variables that are considered for this task have the most homogeneity within the nodes and the greatest difference between them. Nodes that do not have any branches are called terminal nodes, which are considered as leaves for a tree diagram.

The tree model is divided into two categories based on the type of dependent variable: (i) regression tree for a quantitative dependent variable, in which case, mean and SD are reported for each node, and (ii) classification tree for a qualitative dependent variable, for which frequency and percentage are reported for each node (Fig. 1).

Random forest

RF algorithm is a combination of several DT algorithms[23]. RF builds multiple DTs and merges them together to produce more accurate and stable predictions.

RF adds additional randomness to the model as the trees grow. Instead of searching for the most important features when splitting a node, this algorithm looks for the best features among a

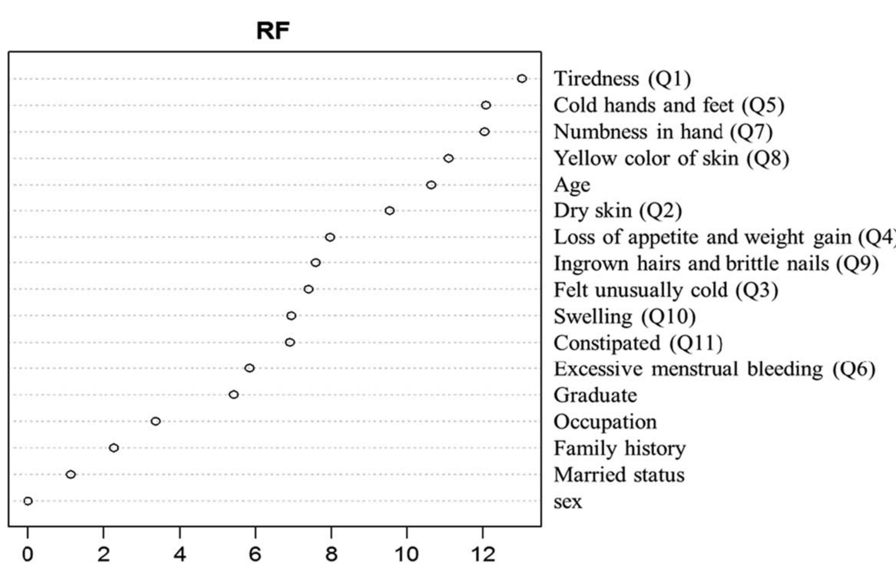

Figure 5. The ranks of important variables in random forest.

random set of features. Therefore, in RF, only a subset of features is considered by the algorithm to split a node (Fig. 2).

Goodness of fit with classifier indices

In order to compare the classifying ability of models, goodness of fit classifier indices was used. The confusion matrix is known as a classifier index. In Figure 3, the main frame of the confusion matrix is displayed. In this framework, true positive (TP) are actual patients that the classifier has correctly identified, true negative (TN) are nonpatients that the classifier has correctly identified, false positive (FP) are nonpatients that the classifier has wrongly identified as patients, and false negative (FN) are patients that the classifier has wrongly identified as healthy.

The following measures, which are based on the confusion matrix, are commonly used to explore the discrimination ability of classifiers:

$$Accuracy = \frac{TP + TN}{TP + TN + FP + FN}$$

$$kappa = \frac{observed \quad accuracy - expected \quad accuracy}{1 - expected \quad accuracy} = \frac{p_0 - p_e}{1 - p_e}$$

$$Sensitivity = \frac{TP}{TP + FN}$$

$$specificity = \frac{TN}{TN + FP}$$

A higher value of these indices indicates better discrimination.

Results

Table 1 describes the demographic characteristics. Mean and SD were reported for the quantitative variable (age) and frequency and percentage were reported for the qualitative variables. Also, *t*-test and χ^2 tests were used to investigate the relationship between demographic variables and the response variable (Hypothyroidism). In comparison of two groups, there were no statistically significant differences in all demographic variables.

Table 2 lists the checklist questions. As can be seen, all questions have a statistically significant relationship with hypothyroidism. Question 6 was investigated only in women.

Figure 4 shows a classification tree with six nodes and seven terminal nodes. This figure shows that questions number 1, 7, 5, 3, and 8 are, respectively, the most important features for diagnosing. According to this chart, if someone had severe ('somewhat' or 'much') tiredness (Q1), and also feeling coldness in hands and feet (Q5), and had yellow skin (jaundice) (Q8), she or he would have Hypothyroidism with more than 90%. The probability of Hypothyroidism in someone who had severe tiredness or fatigue (Q1) and feeling coldness in hand and feet (Q5), but did not have yellow skin (Q8) and a had high degree of numbness in the hands (Q7) was more than 0.65 (or 65%). The other branches of the tree could be followed as above.

Figure 5 shows the importance of each feature in diagnosis of Hypothyroidism with the RF method. The higher the point in this plot, the more important that feature was in diagnosis of Hypothyroidism. The existence of tiredness and feeling coldness in the hands and feet were the most important features in diagnosis of HYPOTHYROIDISM based on the RF method with the Gini index.

Table 3 also shows the importance (rank) of each feature in both RF and DT methods. It can be seen that question 1 was the most important diagnostic question in both methods. Questions 8, education, occupation, family history, marital status, and sex also had the same ranks.

Table 4 shows the results of the LR model. None of the demographic variables had a significant effect on HYPOTHYROIDISM. According to the OR values in Table 4, the chance of having HYPOTHYROIDISM was: (a) 4.91 times [OR = 4.91, 95% CI = (1.31–18.36), *P*-value = 0.018] in people who had much felling tiredness (Q1) compared to very little, (b) 5.98 times [OR = 5.98, 95% CI = (1.84–19.40), *P*-value = 0.003] in individuals who had much loss of appetite and weight Hypothyroidism gain (Q4) compared to very little, (c) 5.96 times [OR = 5.96, 95% CI = (1.89, 18.73), *P*-value = 0.002] in those who had somewhat cold hands and feet (Q5) compared to very little, (d) 3.78 times [OR = 3.78, 95% CI = (1.12–12.69), *P*-value = 0.031] in individuals who had somewhat numbness in hands (Q7) compared to very little and (d) 3.41 times [OR = 3.41, 95% CI = (1.03–11.33), *P*-value = 0.045] in those who had much yellow color of skin (Q8) compared to very little.

Table 5 shows the results of comparing four criteria (Sensitivity, Specificity, Kappa, and Correct classification rate) in two-part data (train set and test set) between three classification methods. Eighty percent of the data are training data and 20% are test data. As can be seen, the RF algorithm had the highest CCR, KC, SE, and SP in the train and test data sets.

Table 3

The importance variables of HYPOTHYROIDISM based on the DT and RF algorithms

Checklist questions	Importance (rank) based on DT	Importance (rank) based on RF
1. Unusual fatigue or weakness (Q1)	**1**	**1**
2. Dry skin (Q2)	3	6
3. Felt unusually cold (Q3)	8	9
4. Loss of appetite and weight gain (Q4)	5	7
5. Cold hands and feet (Q5)	7	2
6. Excessive menstrual bleeding (Q6)	11	12
7. Numbness in hand (Q7)	6	3
8. Yellow color of skin (Q8)	4	4
9. Ingrown hairs and brittle nails (Q9)	10	8
10. Swelling (Q10)	9	10
11. Constipated (Q11)	2	11
12. Age	15	5
13. Education	13	13
14. Occupation	14	14
15. Family history	15	15
16. Marital status	16	16
17. Sex	17	17

Bold values indicate the checklist's question number that is the most important diagnostic characteristic of hypothyroidism according to both DT and RF methods.

Table 4

The effects of risk factors on HYPOTHYROIDISM

	Univariable		Multiple	
Variables	**OR (95% CI)**	***P***	**OR (95% CI)**	***P***
Age	1.00 (0.98–1.01)	0.785	—	—
Sex (woman)	1.33 (0.73–2.41)	0.345		
Education				
Illiterate	Ref		—	—
Middle school	1.21 (0.52–2.77)	0.649		
Diploma	1.07 (0.53–2.15)	0.837		
College education	0.80 (0.36–1.66)	0.562		
Family history	1.35 (0.87–2.09)	0.178	—	—
Employment				
Unemployed	Ref		—	—
Government job	1.32 (0.79–2.02)	0.280		
Nongovernment	0.94 (0.46–1.91)	0.871		
Marital status (married)	1.63 (0.94–2.82)	0.078	—	—
Q1: unusual fatigue or weakness				
Very little	Ref		Ref	
Little	2.74 (1.10–6.82)	0.030	0.98 (0.24–3.91)	0.979
Somewhat	8.13 (3.59–18.39)	< 0.001	3.27 (0.92–11.57)	0.066
Much	13.5 (5.93–30.68)	< 0.001	**4.91** (1.31–18.36)	0.018
Q2: dry skin				
Very little	Ref		Ref	
Little	0.14 (0.06–0.29)	< 0.001	0.27 (0.06–1.16)	0.07
Somewhat	0.24 (0.12–0.48)	< 0.001	1.43 (0.43–4.77)	0.55
Much	0.79 (0.45–1.41)	0.439	1.24 (0.37–4.11)	0.72
Q3: felt unusually cold				
Very little	Ref		Ref	
Little	0.15 (0.07–0.32)	<0.001	1.88 (0.67–5.27)	0.22
Somewhat	0.34 (0.16–0.69)	0.003	1.40 (0.48–4.09)	0.53
Much	0.52 (0.25–1.09)	0.083	1.62 (0.49–5.33)	0.42
Q4: loss of appetite and weight gain				
Very little	Ref		Ref	
Little	1.43 (0.73–2.78)	0.292	1.16 (0.40–3.32)	0.776
Somewhat	3.45 (1.81–6.59)	< 0.001	2.73 (0.91–8.15)	0.071
Much	7.53 (3.66–15.48)	< 0.001	**5.98** (1.84–19.40)	0.003
Q5: cold hands and feet				
Very little	Ref		Ref	
Little	1.98 (0.94–4.15)	0.069	2.23 (0.66–7.57)	0.195
Somewhat	7.35 (3.76–14.34)	<0.001	**5.96** (1.89–18.73)	0.002
Much	7.08 (3.48–14.39)	<0.001	2.47 (0.78–7.87)	0.124
Q6: excessive menstrual bleeding				
Very little	Ref		Ref	
Little	0.38 (0.17–0.84)	0.017	0.72 (0.37–1.39)	0.182
Somewhat	0.27 (0.10–0.69)	0.006	1.94 (0.88–4.29)	0.074
Much	0.74 (0.26–2.06)	0.566	2.62 (1.18–5.81)	0.725
Q7: numbness in your hands				
Very little	Ref		Ref	
Little	1.86 (0.79–4.36)	0.150	2.40 (0.67–8.54)	0.176
Somewhat	6.78 (3.33–13.81)	< 0.001	**3.78** (1.12–12.69)	0.031
Much	10.50 (5.22–21.08)	< 0.001	2.73 (0.84–8.79)	0.092
Q8: Yellowness in the color of your skin				
Very little	Ref		Ref	
Little	1.25 (0.67–2.33)	0.470	0.43 (0.15–1.23)	0.117
Somewhat	4.65 (2.50–8.64)	< 0.001	2.05 (0.72–5.80)	0.174
Much	8.94 (4.10–19.51)	< 0.001	**3.41** (1.03–11.33)	0.045
Q9: ingrown hairs and brittle nails				
Very little	Ref		Ref	
Little	1.00 (0.52–1.90)	0.998	0.57 (0.19–1.66)	0.306
Somewhat	3.39 (1.78–6.46)	< 0.001	1.53 (0.51–4.58)	0.446
Much	4.52 (2.33–8.77)	< 0.001	1.13 (0.38–3.34)	0.820
Q10: swelling around the eyes or legs				
Very little	Ref		Ref	
Little	1.76 (0.97–3.20)	0.060	1.37 (0.49–3.79)	0.539
Somewhat	2.77 (1.51–5.07)	0.001	1.14 (0.42–3.09)	0.786

Table 4

(*Continued*)

Variables	Univariable OR (95% CI)	Univariable P	Multiple OR (95% CI)	Multiple P
Much	5.42 (2.71–10.83)	< 0.001	0.96 (0.28–3.21)	0.949
Q11: constipated				
Very little	Ref		Ref	
Little	0.90 (0.47–1.71)	0.758	0.31 (0.10–0.96)	0.044
Somewhat	2.19 (1.20–3.97)	0.010	0.48 (0.16–1.37)	0.172
Much	3.52 (1.85–6.69)	< 0.001	0.98 (0.31–3.03)	0.977

Bold values are statistical significant.

Discussion

Among the 11 symptoms of Hypothyroidism in this study, tiredness was the most important clinical feature that discriminated between Hypothyroidism and healthy individuals with the three classification methods. Numbness of hands, cold hands and feet, jaundice (yellow skin), loss of appetite, and weight Hypothyroidism gain, and feeling abnormally cold followed in importance. Also, increased level of menstruation in women increased the chance of Hypothyroidism. In this study, the DT, RF, and LR algorithms were used to diagnose Hypothyroidism, among which the RF algorithm had the best performance in diagnosing the disease.

In the study of Carlé *et al.*[24], it was observed that more than 80% of primary cases of Hypothyroidism suffer from tiredness, which was one of the most important symptoms for diagnosing Hypothyroidism in our study in the three algorithms, too. Also, many other studies showed that tiredness is one of the most common symptoms of Hypothyroidism[25–27].

Another symptom of Hypothyroidism that many patients suffered from was feeling unusually cold. In a study conducted by Khurram *et al.*[28], about 50% of patients with primary Hypothyroidism felt unusually cold, while in our study about 74.6% of people complained of feeling coldness in their hands and feet.

In a study conducted by Canaris *et al.*, symptoms such as tiredness, dry skin, and hoarse voice were significant in univariate analysis, but in multivariate analysis, among 13 symptoms of Hypothyroidism, only two symptoms of puffy eyes and constipation were significant. In our study, the symptoms of tiredness and dry skin were statistically significant in univariate analysis, but neither puffy eyes nor constipation were significant[29].

Several studies have been conducted on the most common symptoms of Hypothyroidism, such as weight Hypothyroidism gain, and muscle cramps, which show that these symptoms are very important in the diagnosis of Hypothyroidism, and these results are in line with the present study[1,30,31]. To the best of our knowledge, there has been no study that shows the numbness of hands and jaundice as important symptoms of Hypothyroidism.

In a diagnostic study of thyroid disease conducted by Alyas *et al.*[32] in 2022, it was observed that the RF algorithm had the best performance (sensitivity = 94.8) in diagnosing patients and nonpatients, which is in line with our study.

In a systematic review, it was reported that out of 17 published articles that used the RF algorithm, in nine (53%) articles, the RF algorithm had the highest prediction accuracy compared to other machine learning algorithms such as support vector machine, Naïve Bayes, artificial neural network, and K-nearest neighbor[17]. In the present study, the RF algorithm had the best performance in diagnosis of Hypothyroidism, and this adds further evidence that the RF algorithm can help clinicians as a complementary tool in diagnosis of diseases, in addition to or perhaps instead of clinical tests.

This study was conducted at a single center with a relatively uniform population ethnicity. As a further study, testing the robustness of these tested algorithms on different populations comprising more varied ethnicities (and their dependence on the studied population) is of interest.

Limitations

Firstly, machine learning algorithms shows greater efficiency in large sample sizes, therefore, RF may show better goodness of fit indices if our sample size was larger. Secondly, there may be more symptoms that are important in the diagnosis process, but most of them need measurement instruments. We wanted to use only visual and simple symptoms.

Conclusions

The importance of diagnostic clinical features of Hypothyroidism was ranked, which could help physicians investigate them in patients based on their order of importance. The findings suggest that machine learning methods can identify Hypothyroidism patients who show relatively simple symptoms with acceptable accuracy, without the need for a blood test. Greater familiarity

Table 5

Comparing RF, DT, and LR with goodness of fit indices

Train set	LR	DT	RF
Accuracy	0.52	0.84	0.98
Kappa	0.041	0.67	0.95
Sensitivity	0.54	0.86	0.96
Specificity	0.50	0.81	0.97
Test set			
Accuracy	0.64	0.70	0.83
Kappa	0.29	0.40	0.46
Sensitivity	0.72	0.64	0.88
Specificity	0.59	0.75	0.88

and utilization of such methods by physicians may, therefore, reduce the expense and stress burden of clinical testing. Such statistical classification methods could be helpful in classifying patients and promise to help physicians increasingly in future.

Ethical approval

All procedures performed in this study were in accordance with the 1964 Helsinki Declaration and its later amendments. Participants' written informed consent was obtained before starting the interview and collecting information from the respondents.

The proposal of the present study was approved by the ethics committee of Kerman University of Medical Sciences, Kerman, Iran (reference 401000296). The ethical approval code is IR. KMU.REC.1401.292.

Consent

All procedures performed in this study were in accordance with the 1964 Declaration of Helsinki and its subsequent amendments. Informed written consent of the participants was obtained from the respondent before starting the interview and data collection. The informed consent was in the form of a form and each of the literate participants read and signed it, and for those who were illiterate, the form was read orally. Also, how to access the results of this research was explained to the participants.

Sources of funding

None.

Author contribution

S.R.R.: data analysis and drafting of manuscript; Z.H.M.: diagnosis of new patients by clinical examination and checklist design; M.J.Z.: diagnosis of new patients by clinical examination; M.A.M.S.: reviewed, edited, and approved the final manuscript; T.D.: supervised, reviewed, and edited the manuscript.

Conflicts of interest disclosure

The authors declare that they have no conflicts of interest.

Research registration unique identifying number (UIN)

Our study was registered in ClinicalTrials.gov. ClinicalTrials.gov Identifier: NCT06112886.

Guarantor

Tania Dehesh: correspond author; Salahuddin Rakhshani Rad: first author.

Data availability statement

The datasets used and/or analyzed during the current study are available from the corresponding author on reasonable request.

Provenance and peer review

Not commissioned, externally peer-reviewed.

Acknowledgements

The authors wish to thanks all those who helped to carry this study, especially Bessat clinic in Kerman, for their cooperation.

References

[1] Gaitonde DY, Rowley KD, Sweeney LB. Hypothyroidism: an update. South Afr Family Pract 2012;54:384–90.
[2] Chaker L, Bianco AC, Jonklaas J, *et al*. Hypothyroidism and hypertension: fact or myth?–Authors' reply. Lancet 2018;391:30.
[3] Guglielmi R, Grimaldi F, Negro R, *et al*. Shift from levothyroxine tablets to liquid formulation at breakfast improves quality of life of hypothyroid patients. Endocr Metab Immune Disord Drug Targets 2018;18:235–40.
[4] Malaty W. Primary hypothyroidism. 2017. In:2017.
[5] Melmed S, Polonsky KS, Larsen PR, *et al*. Williams textbook of endocrinology E-Book. Elsevier Health Sciences; 2015.
[6] Taylor PN, Albrecht D, Scholz A, *et al*. Global epidemiology of hyperthyroidism and hypothyroidism. Nat Rev Endocrinol 2018;14: 301–16.
[7] Chiovato L, Magri F, Carlé A. Hypothyroidism in context: where we've been and where we're going. Adv Ther 2019;36:47–58.
[8] Talebi S, Ghaedi E, Sadeghi E, *et al*. Trace element status and hypothyroidism: a systematic review and meta-analysis. Biol Trace Elem Res 2020;197:1–14.
[9] Aminorroaya A, Meamar R, Amini M, *et al*. Incidence of thyroid dysfunction in an Iranian adult population: the predictor role of thyroid autoantibodies: results from a prospective population-based cohort study. Eur J Med Res 2017;22:1–10.
[10] Cappola AR, Fried LP, Arnold AM, *et al*. Thyroid status, cardiovascular risk, and mortality in older adults. JAMA 2006;295: 1033–41.
[11] Malik BA, Butt MA. Is delayed diagnosis of hypothyroidism still a problem in Faisalabad, Pakistan. Life 2008;11:12.
[12] Smallridge RC, Ladenson P. Hypothyroidism in pregnancy: consequences to neonatal health. J Clin Endocrinol Metabol 2001;86: 2349–53.
[13] Raval AD, Sambamoorthi U. Incremental healthcare expenditures associated with thyroid disorders among individuals with diabetes. J Thyroid Res 2012;2012:1–10.
[14] Okosieme O, Gilbert J, Abraham P, *et al*. Management of primary hypothyroidism: statement by the British Thyroid Association Executive Committee. Clin Endocrinol (Oxf) 2016;84:799–808.
[15] Sidey-Gibbons JA, Sidey-Gibbons CJ. Machine learning in medicine: a practical introduction. BMC Med Res Methodol 2019;19:1–18.
[16] Abbasi S, Tavakoli M, Boveiri HR, *et al*. Medical image registration using unsupervised deep neural network: a scoping literature review. Biomed Signal Process Control 2022;73:103444.
[17] Uddin S, Khan A, Hossain ME, *et al*. Comparing different supervised machine learning algorithms for disease prediction. BMC Med Inform Decis Mak 2019;19:1–16.
[18] Kasper D, Fauci A, Hauser S, *et al*. Harrison's principles of internal medicine, 19e, 1. Mcgraw-hill; 2015.
[19] https://clinicaltrials.gov/study/NCT06112886 2023.
[20] Mathew G, Agha R, Albrecht J, *et al*. STROCSS 2021: strengthening the reporting of cohort, cross-sectional and case-control studies in surgery. Int J Surg Open 2021;37:100430.
[21] Hosmer DW Jr, Lemeshow S, Sturdivant RX. Applied logistic regression, Vol 398. John Wiley & Sons; 2013.
[22] Kleinbaum DG, Klein M, Pryor ER. Logistic regression: a self-learning text, 94. Springer; 2002.
[23] Breiman L. Random forests. Machine learn 2001;45:5–32.
[24] Carlé A, Pedersen IB, Knudsen N, *et al*. Hypothyroid symptoms and the likelihood of overt thyroid failure: a population-based case–control study. Eur J Endocrinol 2014;171:593–602.
[25] Barker D, Bishop J. Computer-based screening system for patients at risk of hypothyroidism. The Lancet 1969;294:835–8.

[26] Doucet J, Trivalle C, Chassagne P, *et al*. Does age play a role in clinical presentation of hypothyroidism? J Am Geriatr Soc 1994;42: 984–6.
[27] Watt T, Hegedüs L, Rasmussen ÅK, *et al*. Which domains of thyroid-related quality of life are most relevant? Patients and clinicians provide complementary perspectives. Thyroid 2007;17:647–54.
[28] Khurram IM, Choudhry KS, Muhammad K, *et al*. Clinical presentation of hypothyroidism: a case control analysis. J Ayub Med Coll Abbottabad 2003;15:45–9.
[29] Canaris GJ, Steiner JF, Ridgway EC. Do traditional symptoms of hypothyroidism correlate with biochemical disease? J Gen Intern Med 1997; 12:544–50.
[30] Almandoz JP, Gharib H. Hypothyroidism: etiology, diagnosis, and management. Med Clin 2012;96:203–21.
[31] Wilson SA, Stem LA, Bruehlman RD. Hypothyroidism: diagnosis and treatment. Am Fam Physician 2021;103:605–13.
[32] Alyas T, Hamid M, Alissa K, *et al.* Empirical method for thyroid disease classification using a machine learning approach. Biomed Res Int 2022;2022:1–10.