

Using The Random Forest Algorithm To Detect The Activity of Thyroid-Associated Ophthalmopathy

Minghui Wang

Beijing Tongren Eye Center, Beijing Tongren Hospital, Beijing Ophthalmology and Visual Science Key Lab, Capital Medical University, Beijing

Hanqiao Zhang

Beijing Tsinghua Changgung Hospital, School of Clinical Medicine, Tsinghua University

Li Dong

Beijing Tongren Eye Center, Beijing Tongren Hospital, Beijing Ophthalmology and Visual Science Key Lab, Capital Medical University, Beijing

Yang Li

Beijing Tongren Eye Center, Beijing Tongren Hospital, Beijing Ophthalmology and Visual Science Key Lab, Capital Medical University, Beijing

Zhijia Hou

Beijing Tongren Eye Center, Beijing Tongren Hospital, Beijing Ophthalmology and Visual Science Key Lab, Capital Medical University, Beijing

Dongmei Li (✉ ldmlily@x263.net)

Beijing Tongren Eye Center, Beijing Tongren Hospital, Beijing Ophthalmology and Visual Science Key Lab, Capital Medical University, Beijing

Research Article

Keywords: Thyroid-associated ophthalmopathy(TAO), Clinical activity score, Random forest algorithm

Posted Date: September 14th, 2021

DOI: <https://doi.org/10.21203/rs.3.rs-787674/v1>

License:   This work is licensed under a Creative Commons Attribution 4.0 International License.

[Read Full License](#)

Abstract

Objective: The aim of this study is to establish a random forest model to detect active and quiescent phases of patients with Thyroid-associated ophthalmopathy (TAO) and to evaluate its diagnostic performance.

Methods: A total of 146 patients (292 eyes) who were diagnosed with TAO and were treated in the Ophthalmology Outpatient Clinic of Beijing TongRen hospital were retrospectively included in the study. We took the clinical activity score of TAO as the target; took gender, age, smoking status, I-131 treatment history, thyroid nodules, thyromegaly, thyroid hormone and TSH-receptor antibodies (TRAb) as predictive characteristic variables to establish a random forest model. The proportion of the training group to the testing group was 7:3. We analyzed the model's accuracy, precision, sensitivity, specificity, positive predictive value (PPV), negative predictive value (NPV), F1 score and out-of-bag (OOB) error, with the accuracy, the brier loss and the area under the receiver operating characteristic curve compared with logistic regression model.

Results: Our model has an accuracy of 0.93, a sensitivity of 0.88, a specificity of 0.96, a positive predictive value of 0.94, a negative predictive value of 0.93, an F1 score of 0.91 and an OOB error of 0.12. The accuracy of the random forest model and the logistic regression model were 0.93 and 0.79, respectively, the brier loss were 0.06 and 0.20, and the area under the receiver operating characteristic curve were 0.95 and 0.86.

Conclusion: By integrating these high-risk factors, the random forest algorithm can be used as a complementary diagnostic method to determine the activity of TAO, showing prominent diagnostic performance.

Introduction

Thyroid-associated ophthalmopathy (TAO) is an autoimmune disorder affecting the orbital and periorbital tissues, which can cause dysthyroid optic neuropathy, exposure keratitis, even corneal ulcer in severe cases. Therefore, timely interventions and treatments are particularly crucial to the prognosis of patients with TAO. The natural course of TAO is presented as a U-shaped curve, which can be divided into two stages: the active phase and the quiescent phase [1,2]. Treatments during these two phases are very different [3]. As a result, detecting the clinical activity phase of patients with TAO is critical to the choice of treatment [4,5].

Nowadays, the Clinical Activity Score (CAS) is the most common method to evaluate the phase of the disease [6]. However, such evaluation standards require professional ophthalmologists to conduct a comprehensive examination of the patient's eye manifestation, which may cause the delay in monitoring the patient's progression. Also, due to the subjectivity of CAS, the score evaluated by different doctors are not exactly the same. Therefore, the CAS has limitations, and it is necessary to find a more objective method to predict the activity of patients.

The random forest algorithm is superior to traditional statistical methods. It is based on a large number of decision tree classifiers, which helps to reduce deviations, tolerant outliers, and avoid overfitting [7]. Although this algorithm has been widely applied in many fields, no one has used it to study the activity of TAO yet.

It has been well-documented that the patient's gender, age, smoking status, I-131 treatment history, thyroid function status, and thyroid hormone receptor antibody (TRAb) are high-risk factors for TAO which are closely related to the development of the disease[8,9]. In this study, we integrate the above indicators and plan to use the random forest algorithm to detect the activity of TAO patients. The purpose of the study is to investigate whether we can predict the activity of patients with TAO through the random forest algorithm, a machine learning method without professional eye examinations.

Material And Methods

Patients:

The study was approved by the Ethics Committee of Beijing Tongren Hospital, Capital Medical University and carried out strictly in accordance with the ethical guidelines of the declaration of Helsinki (Approval No.TRECKY2018-056). From April 2019 to January 2021, 146 patients (292 eyes in total) who visited Ophthalmology Outpatient Clinic of Beijing TongRen hospital were examined. The inclusion criteria were as follows: (1) age more than 18 years, (2) meeting the internationally recognized diagnostic criteria for TAO, and (3) integral clinical demographic data. The enrolled patients were registered with the European Group on Graves' orbitopathy (EUGOGO) first diagnosis form to collect the patient's gender, age, smoking status, I-131 treatment history, thyroid nodules, thyromegaly, serum thyroid hormone, TRAb, ocular symptoms and signs. The exclusion criteria were as follows: (1) underage patients; (2) eye inflammation caused by other eye diseases; (3) patients with other autoimmune diseases.

Clinical thyroid associated ophthalmopathy evaluation

Based on the EUGOGO guidelines, patients were divided into two groups according to their clinical activity of TAO [10]. Professional ophthalmologists from Beijing Tongren Hospital conducted eye examinations on the patients. Their clinical activity was evaluated based on the clinical activity score (CAS), which consists of 7 orbital signs and symptom variables (including spontaneous retrobulbar pain, pain on attempted upward or downward gaze, swelling of the eyelids, redness of the eyelids, redness of conjunctiva, chemosis, swelling of caruncle or plica), each variable is worth 1 point [10]. The sum of the scores of the 7 variables can reflect the degree of disease activity, and $CAS \geq 3$ points is used as a sign of activity phase [11]. (Fig.1)

Laboratory measurements

Levels of serum TT3, TT4, FT3, FT4 and TSH were measured by chemiluminescence using a UniCel DXI800 analyzer (Beckman Coulter, USA). TSHR Ab levels, were measured by electrochemical

luminescence immunoassays (ECLIA) using a Cobas801 analyzer (Roche Diagnostics GmbH, Germany). Reference values were as follows: TT3, 0.92-2.79 nmol/L; TT4, 57.9-140.3 nmol/L; FT3, 3.5-6.5 pmol/L; FT4, 11.5-22.7 g/dL; TSH, 0.55-4.78 μ IU/mL; TRAb, <1.75 IU/L.

Statistical analysis

The t-test was used for continuous variables, and Fisher's exact test for categorical variables. The IBM SPSS Statistics software (Version 20.0. IBM Corporation, NY, USA) was used for analyses.

Random forest modeling

We used the sklearn package in python3.9 to build a random forest model, including the patient's gender, age, I-131 treating history, smoking status, thyromegaly, thyroid nodules, FT3, FT4, TT3, TT4, TSH, TRAb, and clinical activity, altogether 13 characteristic variables. All variables were uniformly quantified and coded (such as table 1). We took the CAS of TAO as the target, and other 12 factors as predictive characteristic variables. We randomly selected 30% of the sample from the total data as the test set, and the remaining 70% as the training set.

In order to address the class-imbalanced dataset, we applied the Synthetic Minority Oversampling Technique (SMOTE) to oversample the minority class. SMOTE considered a sample from the minority class and its k-nearest neighbors, taking vectors between the sample and neighbors in the feature space. New records were then generated by picking points randomly on these vectors.

In this model, the data was randomly sampled from the training set repeatedly and replaced by bootstrapping/bagging. A subset was generated for each decision tree, and then the decision tree model was trained using this subset [7,12]. This method can reduce the variance of random forest model, that is, reduce the impact of individual abnormal samples on the overall prediction performance of the model. Specifically, in order to minimize the classification error of the model, the random forest model randomly selects a subset of feature variables (the number is mtry). We took the recommended value of mtry=3, which is the square root of the total number of features [13].

We chose out-of-bag (OOB) error that is generally used to estimate the generalization error of the model and use the number of decision trees (ntree) in the random forest model as a variable to repeatedly train the model and calculate the OOB error until it becomes stable [14]. We chose the model with the lowest OOB error and took the corresponding ntree as the optimal value. The input variables in the optimized model were ranked by relative importance in predicting clinical activity based on the mean decrease accuracy (MDA) and mean decrease gini (MDG) [13].

On the basis of this model, we calculated the model's accuracy, sensitivity, specificity, positive predictive value, negative predictive value, and F1 score of the test set. F1 score is an index to measure the accuracy of binary classification model, which conveys the balance between precision and sensitivity. Finally, we

integrated the variables into the logistics regression model, comparing the accuracy, the area under the receiver operating characteristic curve and the brier loss of two models [15].

Results

Baseline characteristic

146 patients with 292 eyes were analyzed in this study, 116 (39.7%) of whom were diagnosed with active phase of TAO (Table 1).

Table1 Demographical and clinical features.

Variables	Data type	Quiescent phase [□]	Active phase [□]	P value
Female	Binary Classification	98(55.7%)	66 (56.9%)	0.838
Smoke	Binary Classification	32 (18.2%);	15 (12.9%) [□]	0.258
I131	Binary Classification	21 (11.9%);	10 (8.6%) [□]	0.440
Nodule	Binary Classification	43 (24.4%);	33 (28.4%) [□]	0.340
Thyromegaly	Binary Classification	64 (36.4%);	70 (60.3%) [□]	<0.001
Age	Continuous Variable	46.04±13.20	50.10±14.10	0.014
FT3	Continuous Variable	5.08±1.26	6.19±2.23	<0.001
FT4	Continuous Variable	13.52±4.27	16.58±6.95	<0.001
TT3	Continuous Variable	1.91±0.84	2.06±0.81	0.131
TT4	Continuous Variable	100.34±32.82	110.79±39.39	0.019
TSH	Continuous Variable	4.01±9.85	1.75±3.14	0.005
TRAb	Continuous Variable	3.73±4.52	17.91±12.98	<0.001
Clinical Activity	Binary Classification	176	116	

Random forest prediction of thyroid associated ophthalmopathy

Combined with the patient's gender, age, I-131 treating history, smoking status, thyromegaly, thyroid nodules, FT3, FT4, TT3, TT4, TSH, TRAb, we successfully constructed a random forest model and tested it to predict the clinical activity of TAO. When mtry= 3, the out-of-bag error decreases as ntree increases (Fig.1) and remains stable until about ntree=350. Our final random forest model is built with 350 trees (ntree=350), and its out-of-bag error is as low as 0.12 (Fig.2).

Based on the MDA and MDG of the predictive characteristic variables in the model, we ranked the relative importance of the variables respectively (Fig.3). It can be seen that the TRAb level is the most important

feature for the model, and its importance is significantly higher than other features. Five thyroid hormones all rank high and show similar feature importance, no matter the model was based on the MDA and MDG. The accuracy of this random forest model to predict the clinical activity of TAO is 0.93, the sensitivity is 0.88, the specificity is 0.96, the PPV is 0.94, the NPV is 0.93, F1 Score of 0.91, and the OOB error is 0.12 (Table 2). The accuracy of the random forest model is higher than that of the logistic regression model, which are 0.93 and 0.79, respectively; the brier losses are 0.06 and 0.20, respectively; the AUC are 0.95 and 0.86, respectively (Fig.4).

Table2 Prediction of Clinical activity based on a random forest model.

Prediction	Based on CAS		Total	Accuracy	Sensitivity	Specificity	PPV	NPV	F1 Score
	Active	Quiescent							
Active	29	4	33	0.93	0.88	0.96	0.94	0.93	0.91
Quiescent	2	41	43						
Total	31	45							

CAS: clinical activity score; PPV: positive predictive value; NPV: negative predictive value; F1 Score= $2 \times \text{Specificity} \times \text{Sensitivity} / (\text{Specificity} + \text{Sensitivity})$

Discussion

Although CAS is gradually being adopted by ophthalmologists around the world, it is often the endocrinologist or the primary care physician who first evaluates patients with TAO. Thus, accurate CAS evaluation can be time-consuming and laborious. In this study, we constructed a random forest model containing 146 patients with a total of 292 eyes. Based on the patient's gender, age, smoking, I-131 treating history, thyroid hormones, TRAb, thyromegaly and thyroid nodules, we predicted the clinical activity of patients with TAO.

The most important clinical significance of the CAS score is to guide the treatment of diseases. We analyzed the CAS score as a continuous variable instead of binarizing, because the binary classification is sufficient for our purpose. For patients in the active phase, immunomodulatory therapy can suppress inflammation and eye diseases caused by autoimmune reactions. Once the inflammation is static, immunomodulatory treatments is no longer effective, and we should consider reconstructive surgery to improve the appearance of the eyes and eliminate symptoms.

Unlike classic statistical models, random forest models can provide unbiased importance rankings of input variables. In our model, TRAb is the most important predictor of clinical activity, which is consistent with the opinion of previous study that serum TRAb is a laboratory test marker for judging the clinical activity of TAO [4]. The recommendation of the American Association of Clinical Endocrinologists and EUGOGO agrees that TRAb is the central pathogenic event of TAO, and can be used to distinguish patients with active and quiescent phase [16- 20].

In our model we also found that the thyroid hormone ranks high both based on the mean decrease in accuracy and mean decrease in the Gini coefficient, representing the importance of maintaining normal thyroid function among all factors except for TRAb. It has been reported that the thyroid hormone, like autoimmune inflammation, is a significant contributor to oxidative stress in patients with TAO, which manifests as upregulated lipid peroxidation and antioxidant system activation [21,22]. It is found that euthyroid state restoration leads to a relative reduction in activity and levels of most antioxidant parameters [23,24]. Thus, the autoimmune inflammation of the orbital tissue and thyroid hormone status seem to be mutually independent modifiers of TAO activity which may explain the fact that CAS score is related to hormone levels.

Our random forest model predicts the clinical activity of TAO with an accuracy of 0.93, a sensitivity of 0.88, a specificity of 0.96, a PPV of 0.94, a NPV of 0.93, an F1 Score of 0.91, and an OOB error as low as 0.12, which is better than the traditional multi-logistic regression and shows good prediction ability. These findings indicate that the random forest model can provide us with a satisfactory alternative of CAS to evaluate the clinical activity of TAO. In particular, its high specificity may make our method useful for detecting.

In this study, we chose the random forest instead of other algorithms, since the random forest algorithm does not require strict assumptions of the data, and it is less sensitive to outliers than classic methods (such as decision trees). Random forest models can handle a large number of input variables that are common in clinical situations, and they are not prone to overfitting bias [25, 26]. In the training process, not only random samplings, but also different features can be randomly selected by bootstrapping/bagging methods to generate training data sets, which helps to reduce the correlation between decision trees in the final random forest model.

Besides, we used out-of-bag (OOB) error to estimate the generalization error of the model [15]. OOB error provides an unbiased estimation of error and predictive power, which means that random forest algorithms do not require cross-validation, and this has proven to be unbiased in many tests [7]. It is also observed that our sample size is not very large and the number of quiescent phase and active phase of the target in the data is quite different (class-imbalanced). If you train directly, it may cause the trained model to be biased. A common solution to this problem is SMOTE. After SMOTE, the number of records of two types of data will be the same, so that the accuracy of the trained model will be more convincing.

These results indicate that for patients with TAO who cannot meet professional ophthalmologists in time, we can use the random forest model to distinguish patients who may need treatments from patients in quiescent phase, arranging ophthalmology consultation for them. In this way, we could successfully alleviate the overburden of the health system in some countries due to the long waiting time for expert consultation.

There are some limitations in our research. First of all, the sample size was quite small, and patients included in this study are all Chinese, thus the application of the results is restricted by ethnicity. Besides,

this study was a cross-sectional study without long-term follow-up of patients, while we hope that the model can be used to predict the response to treatment in subsequent studies. We will continuously collect more patients' information to further improve the accuracy of the model. In future research, we hope we can add these factors to our model and expand our samples to observe whether it can be improved.

Conclusion

This research lays the foundation for the application of the random forest model in TAO. Random forest algorithm can generate a model that combines age, gender, smoking status, I-131 treatment history, thyroid function and TRAb to predict the clinical activity of TAO, having a good diagnostic performance. We suggest that the random forest model can be used as an auxiliary method to judge the clinical activity of patients with TAO, so as to distinguish patients who need urgent eye consultations from patients who only need regular follow-ups.

Declarations

Acknowledgement:

This work was supported by Biomechanical Research of Human Orbital development Based on the Finite Element [82071005]; and the special fund of the pediatric medical coordinated development center of Beijing hospitals authority [No.XTCX201824].

Author information

Authors: Minghui Wang¹, Hanqiao Zhang², Li Dong¹, Yang Li¹, Zhijia Hou¹, Dongmei Li^{1*}

¹ Beijing Tongren Eye Center, Beijing Tongren Hospital, Beijing Ophthalmology and Visual Science Key Lab, Capital Medical University, Beijing, China

² Beijing Tsinghua Changgung Hospital, School of Clinical Medicine, Tsinghua University

Correspondence to: Dongmei Li. Beijing Tongren Eye Center, Beijing Tongren Hospital, Capital Medical University and Beijing Ophthalmology Visual Science Key Lab, 1# Dong Jiao Min Xiang, Beijing 100730, China.

E-mail: Idmlily@x263.net

Address: 1# Dong Jiao Min Xiang, Beijing, China.

Contributors:

Design and conduction of the study: Minghui Wang, Hanqiao Zhang, Zhijia Hou, Li Dong, Yang Li, Dongmei Li

Acquisition, analysis and interpretation of data: Minghui Wang

Drafted the article: Minghui Wang.

Revised it critically for important intellectual content: Dongmei Li.

Approved the version to be published: Dongmei Li.

All authors have read and approved the final version of the manuscript.

Consent for publication:

All co-authors have seen and approved the final version of the paper and have agreed to its submission for publication. Written informed consent was obtained from the patient for publication of this report and any accompanying images.

Ethics approval and consent to participate

The study was approved by the Ethics Committee of Beijing Tongren Hospital, Capital Medical University and carried out strictly in accordance with the ethical guidelines of the declaration of Helsinki (Approval No.TRECKY2018-056).

Competing interests:

The Author(s) declare(s) that there is no conflict of interest.

Abbreviations

TRAb: thyroid hormone receptor antibody

CAS: clinical activity score

PPV: positive predictive value

NPV: negative predictive value

OOB: out-of-bag

References

1. Bartley GB. Rundle and His Curve. *Arch Ophthalmol* 2011; 129: 356.
2. Kalmann R, Mourits MP. Late recurrence of unilateral graves orbitopathy on the contralateral side. *Am J Ophthalmol* 2002; 133: 727–729.

3. Hiromatsu Y, Eguchi H, Tani J, et al. Graves' ophthalmopathy: epidemiology and natural history. *Intern Med* 2014; 53(5): 353-360.
4. Tvito-Green R, Reich E, Robenshtok E, et al. Third-Generation Elisa Thyrotropin-Receptor Antibody Levels as an Adjuvant Tool to Guide Management of Patients with Graves Orbitopathy. *Endocr Pract* 2014; 20: 145–149.
5. Dolman PJ. Evaluating Graves' Orbitopathy. *Best Pract Res Cl En* 2012; 26: 229–248.
6. Bahn RS, Burch HB, Cooper DS, et al. Hyperthyroidism and Other Causes of Thyrotoxicosis: Management Guidelines of the American Thyroid Association and American Association of Clinical Endocrinologists. *Thyroid* 2011; 21: 593–646.
7. Breiman L. Random forests. *Mach Learn* 2001; 45: 5–32.
8. Bartalena L. Prevention of Graves' ophthalmopathy. *Best Pract Res Cl En* 2012; 26: 371–379.
9. Bhatti MT, Dutton JJ. Thyroid Eye Disease: Therapy in the Active Phase. *J Neuro-Ophthalmol* 2014; 34: 186–97.
10. Bartalena L, Baldeschi L, Boboridis K, et al. The 2016 European Thyroid Association/European Group on Graves' Orbitopathy Guidelines for the Management of Graves' Orbitopathy. *Eur Thyroid J* 2016; 5: 9–26.
11. Wiersinga WM, Perros P, Kahaly GJ, et al. Clinical assessment of patients with Graves' orbitopathy: the European Group on Graves' Orbitopathy recommendations to generalists, specialists and clinical researchers. *Eur J Endocrinol* 2006; 155:387–389.
12. Biau G. Analysis of a random forests model. *J Mach Learn Res* 2012; 13: 1063–1095.
13. Malley JD, Malley KG, Pajevic S. Statistical Learning for Biomedical Data. *Cambridge Univ Press* 2011; 137–154.
14. Bylander T. Estimating Generalization Error on Two-Class Datasets Using Out-of-Bag Estimates. *Mach Learn* 2002; 48: 287-297.
15. Couronné R, Probst P, Boulesteix A-L. Random forest versus logistic regression: a large-scale benchmark experiment. *BMC Bioinformatics* 2018; 19: 270.
16. Bartalena L, Tanda ML. Clinical practice. Graves' ophthalmopathy. *N Engl J Med* 2009; 360: 994–1001.
17. Bahn RS. Graves' ophthalmopathy. *N Engl J Med* 2010; 362: 726–738.

18. Lytton SD, Ponto KA, Kanitz M, et al. A novel thyroid stimulating immunoglobulin bioassay is a functional indicator of activity and severity of graves' orbitopathy. *J Clin Endocrinol Metab* 2010; 95: 2123–2131.
19. Gerding MN, van der Meer JWC, Broenink M, et al. Association of thyrotrophin receptor antibodies with the clinical features of Graves' ophthalmopathy: TSH receptor antibodies and Graves' ophthalmopathy. *Clin Endocrinol* 2000; 52: 267–271.
20. Nicolì F, Lanzolla G, Mantuano M, et al. Correlation between serum anti-TSH receptor autoantibodies (TRAbs) and the clinical feature of Graves' orbitopathy. *J Endocrinol Invest* 2021; 44: 581–585.
21. Turck N, Eperon S, De Los Angeles Gracia M, et al. Thyroid-Associated Orbitopathy and Biomarkers: Where We Are and What We Can Hope for the Future. *Dis Markers* 2018; 2018: 1–9.
22. Venditti P, Meo SD. Thyroid hormone-induced oxidative stress. *Cell Mol Life Sci* 2006; 63: 414–434.
23. Choi W, Li Y, Ji YS, et al. Oxidative stress markers in tears of patients with Graves' orbitopathy and their correlation with clinical activity score. *BMC Ophthalmol* 2018; 18: 303.
24. Londzin-Olesik M, Kos-Kudla B, Nowak A, et al. The effect of thyroid hormone status on selected antioxidant parameters in patients with Graves' disease and active thyroid-associated orbitopathy. *Endokrynol Pol* 2020;71(5):418-424.
25. Askland KD, Garnaat S, Sibrava NJ, et al. Prediction of remission in obsessive compulsive disorder using a novel machine learning strategy. *Int J Methods Psychiatr Res* 2015; 24(2): 156.
26. Stekhoven DJ, Bühlmann P. MissForest – non-parametric missing value imputation for mixed-type data. *Bioinformatics* 2012; 28: 112–118.

Figures

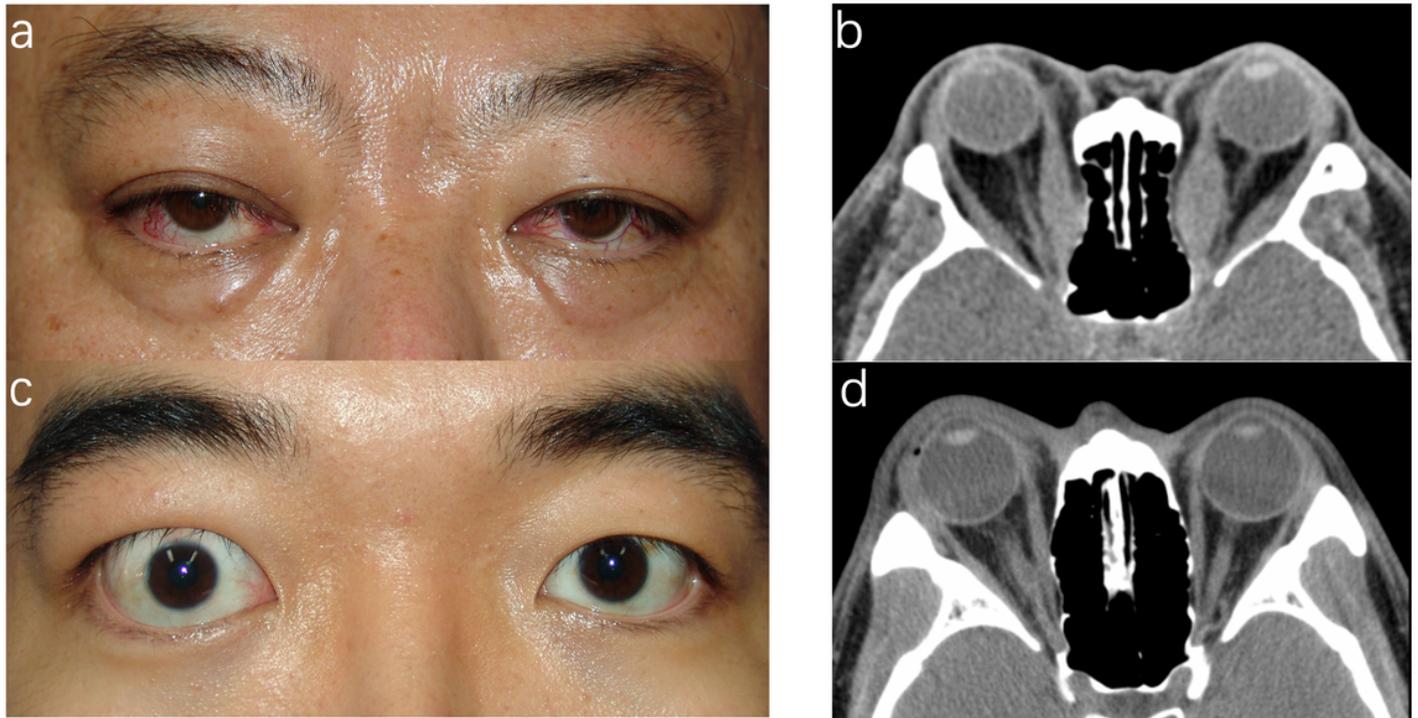


Figure 1

Patients during active phase and the quiescent phase. (a) A patient during active phase with swelling eyelids, redness conjunctiva, chemosis, caruncle inflammation, spontaneous retrobulbar pain of both two eyes and (b) its orbital CT showing hypertrophy extraocular muscles with the volume of 8.729cm³ and 8.488cm³, respectively. (c) A patient during quiescent phase with proptosis and mild swelling eyelids of the right eye and (d) its orbital CT with the extraocular volume of 2.056cm³ and 2.143cm³, respectively.

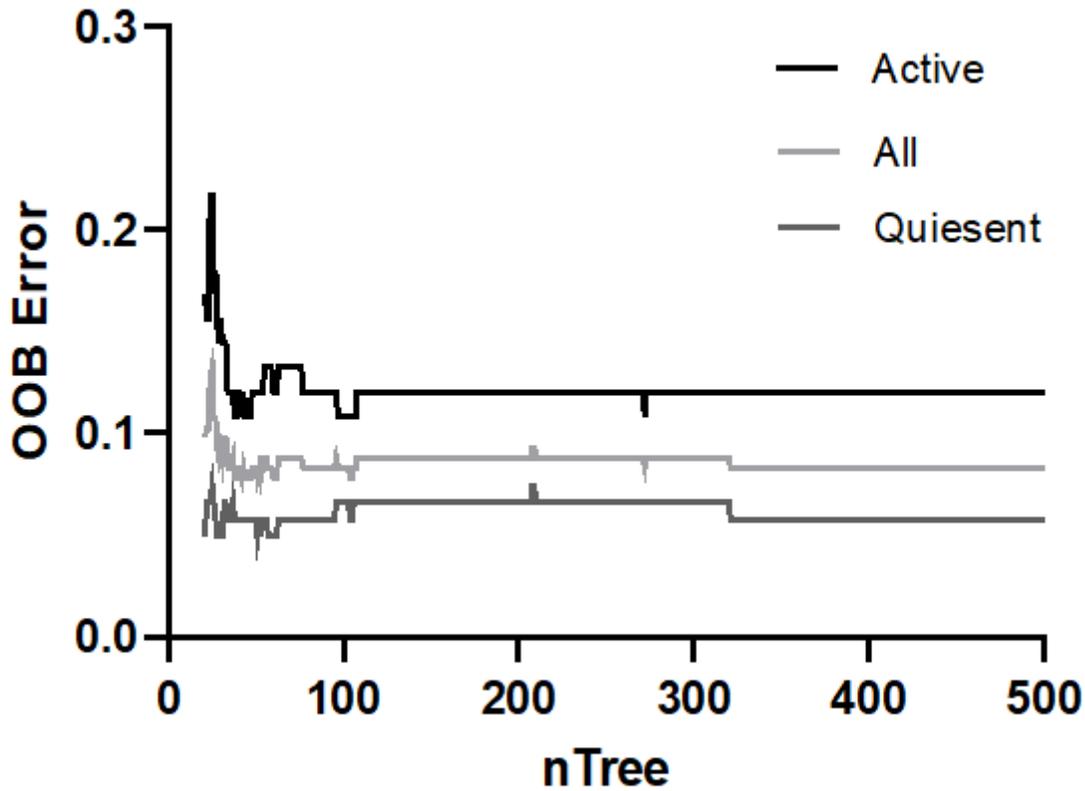


Figure 2

The Out-of-bag (OOB) error. The middle line depicts the OOB error of all data. The top line depicts the OOB error in patients with active phase, and the bottom line shows the OOB error in patients with quiescent phase. For all data, $mtry = 3$.

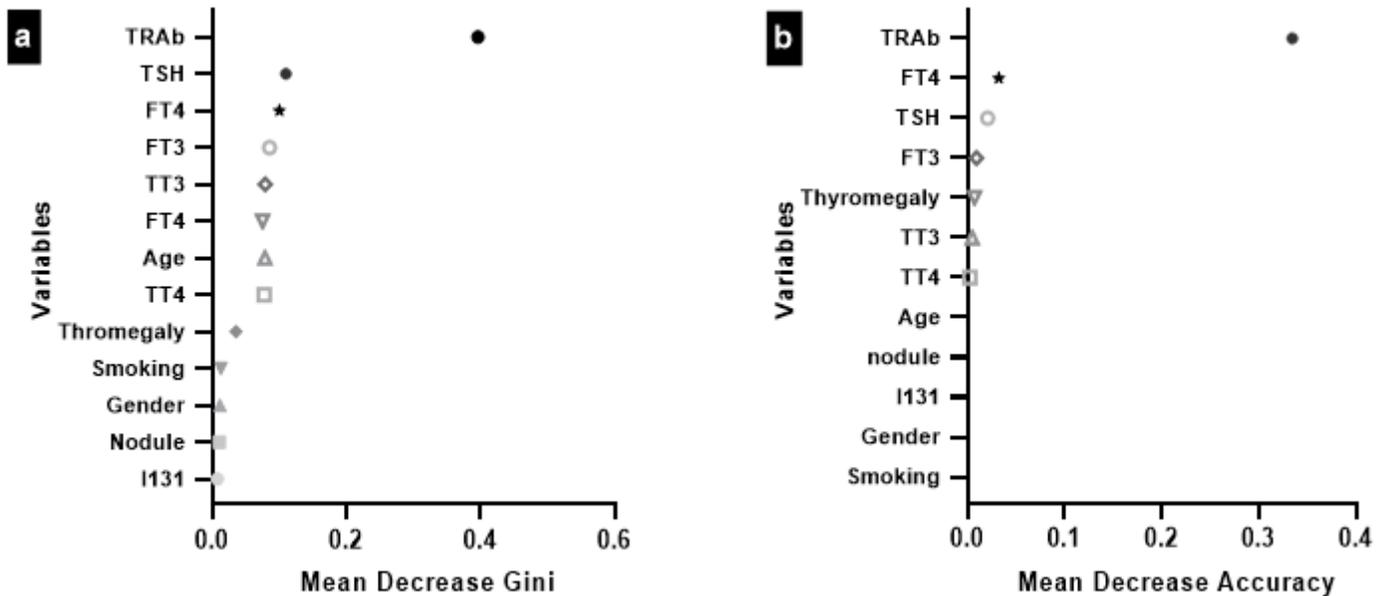


Figure 3

Ranking of input variables in the model to predict the clinical activity of TAO. (a) Mean decrease accuracy. (b) Mean decrease gini.

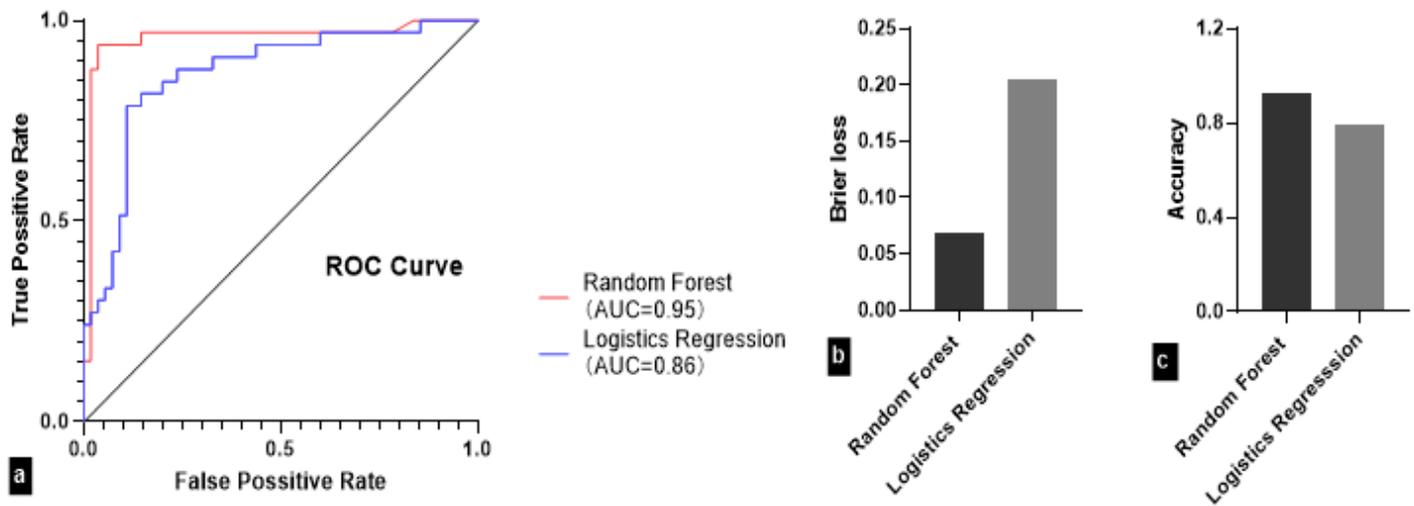


Figure 4

Comparison between random forest model and multivariate logistic regression model. (a) Area under the receiver operating characteristic curve; (b) Brier loss; c. Accuracy