International Journal of General Medicine

Dovepress
Taylor & Francis Group

Open Access Full Text Article

ORIGINAL RESEARCH

Dynamic Monitoring of Thyroid-Stimulating Hormone Receptor Antibody and Extraocular Muscle Thickness: A Clinical Risk Stratification Model for Recurrence Risk Post-Glucocorticoid Therapy in TED Patients

Yaling Li, Xiuhong Yang, Shuli Hu

Department of Intensive Care Unit, Wuhan No. 1 Hospital, Wuhan, 430022, People's Republic of China

Correspondence: Shuli Hu, Department of Intensive Care Unit, Wuhan No. 1 Hospital, No. 215 Zhongshan Avenue, Qiaokou District, Wuhan City, Hubei Province, People's Republic of China, Email hslwhsdyyy123@163.com

Objective: To develop a clinical risk stratification model for thyroid eye disease (TED) recurrence post-glucocorticoid therapy by integrating dynamic thyroid-stimulating hormone receptor antibody (TRAb), extraocular muscle (EOM) thickness, and baseline clinical-biochemical indicators, with only internal validation.

Methods: A single-center retrospective cohort study of 426 TED patients (2016–2023) was conducted. Recurrence (12-month follow-up) was adjudicated by two ophthalmologists. LASSO regression screened predictors, and multiple machine learning models were built and validated via 10-fold cross-validation. Model performance was assessed by AUC and decision curve analysis (DCA).

Results: 98 patients (23.0%) had recurrence, with 89% occurring after 6 months post-treatment. Recurrent patients showed slower TRAb decline and persistent EOM thickening. 15 key predictors were identified, and the random forest (RF) model had the best performance (AUC=0.92, 95% CI: 0.88–0.95; accuracy=0.87) with consistent subgroup results (Events Per Variable=6.5, potential overfitting risk).

Conclusion: This internally validated risk stratification model identifies factors associated with TED recurrence post-glucocorticoid therapy, with dynamic TRAb and EOM thickness as key indicators. The findings are preliminary due to no external validation and temporal limitations, and prospective multi-center studies are required before clinical implementation to inform individualized follow-up.

Keywords: thyroid eye disease, glucocorticoid therapy, recurrence, thyroid-stimulating hormone receptor antibody, extraocular muscle thickness, predictive model

Introduction

Thyroid eye disease is an autoimmune disorder closely associated with thyroid dysfunction, characterized by orbital tissue inflammation, extraocular muscle enlargement, and orbital fat proliferation.[1,2] It predominantly affects middle-aged women and can lead to symptoms such as exophthalmos, diplopia, eyelid retraction, and even optic neuropathy if left untreated, severely impacting patients' visual function and quality of life.[3,4] Glucocorticoid therapy remains the first-line treatment for active moderate-to-severe thyroid eye disease (TED) due to its potent anti-inflammatory effects, which effectively alleviate orbital edema and suppress immune-mediated tissue damage.[5] However, clinical practice indicates that a considerable proportion of patients experience disease recurrence within 12 months after discontinuing glucocorticoid therapy, leading to repeated treatment cycles, increased medical costs, and progressive orbital tissue fibrosis that reduces responsiveness to subsequent therapies.

https://doi.org/10.2147/IJGM.S590217
Received: 31 December 2025
Accepted: 27 February 2026
Published: 20 March 2026

The identification of reliable predictors for TED recurrence post-glucocorticoid therapy is a critical unmet clinical need. Previous studies have explored various baseline indicators, including clinical activity score (CAS), thyroid function parameters, and orbital imaging features, as potential predictive factors.[6–8] Thyroid-stimulating hormone receptor antibody is recognized as a key pathogenic factor in TED, as it mediates immune responses targeting orbital tissues.[9,10] Extraocular muscle thickness, a core imaging manifestation of TED, reflects the severity of orbital inflammation and structural damage.[11] However, most existing studies rely on single-time-point baseline measurements, failing to capture dynamic changes in disease activity during and after treatment. Such static assessments may overlook crucial information about disease progression and regression, limiting their predictive accuracy.

Dynamic monitoring of disease-related biomarkers and imaging parameters could provide more comprehensive insights into disease trajectory and recurrence risk. Changes in TRAb levels during treatment may reflect the persistence of autoimmune activity, while alterations in EOM thickness can indicate the resolution of inflammation or progression to fibrosis. Integrating these dynamic data with comprehensive clinical and biochemical indicators may enhance the robustness of recurrence prediction. Therefore, this study aimed to construct a predictive model for TED recurrence post-glucocorticoid therapy by incorporating dynamic TRAb monitoring, EOM thickness measurements, and a wide range of baseline and follow-up indicators. The model was developed using data from Wuhan No. 1 Hospital, with the goal of providing a practical tool for clinicians to optimize individualized treatment and follow-up strategies.

Materials and Methods

Study Population

This retrospective cohort study included TED patients who were treated with glucocorticoid therapy at Wuhan No. 1 Hospital between January 2016 and December 2023. The diagnosis of TED was based on typical ophthalmic manifestations (exophthalmos, eyelid retraction, diplopia, or orbital pain) combined with thyroid function abnormalities or positive TRAb. Eligibility criteria were as follows: age ≥18 years; active moderate-to-severe TED defined by CAS ≥3 points (total score 7 points); completion of a standardized glucocorticoid therapy course (intravenous methylprednisolone, 4.5 g over 12 weeks as recommended by EUGOGO guidelines);[12] availability of complete clinical, biochemical, and imaging data; and at least 12 months of follow-up after treatment completion.

Exclusion criteria included: history of orbital radiotherapy, orbital decompression surgery, or other immunosuppressive therapy (eg, rituximab, azathioprine) within 3 months before glucocorticoid initiation; concurrent IgG4-related disease, other autoimmune diseases (eg, rheumatoid arthritis, systemic lupus erythematosus), or malignant tumors; severe cardiac, hepatic, or renal dysfunction that contraindicated glucocorticoid use; poor adherence to treatment or loss to follow-up before 12 months; and incomplete dynamic monitoring data of TRAb or EOM thickness. The processes of patient screening, enrollment, and data analysis are detailed in Figure 1.

Disease recurrence was defined as the reappearance or worsening of at least two TED-related clinical indicators within 12 months after treatment completion, including CAS score increase ≥2 points, exophthalmos progression ≥2 mm, new-onset or aggravated diplopia, or eyelid retraction progression ≥1 mm (in accordance with the EUGOGO clinical practice guidelines for TED management). Recurrence was confirmed by two independent ophthalmologists, with the interrater reliability assessed by Kappa coefficient (κ=0.89, indicating excellent consistency); discrepancies were resolved by a senior specialist.

Data Collection and Candidate Predictors

Comprehensive candidate predictors were collected, including demographic characteristics, clinical baseline indicators, biochemical parameters, and dynamic monitoring data. Demographic characteristics included age, gender, smoking history (current smoker, former smoker, never smoker), alcohol consumption, and comorbidities (diabetes mellitus, hypertension, dyslipidemia).

Clinical baseline indicators included CAS score, ophthalmopathy duration, exophthalmos (measured by Hertel exophthalmometer), palpebral fissure width (margin reflex distance 1 and 2), diplopia grade (0 = no diplopia, 1 = diplopia on horizontal/vertical gaze, 2 = intermittent diplopia on primary gaze, 3 = persistent diplopia on primary gaze),

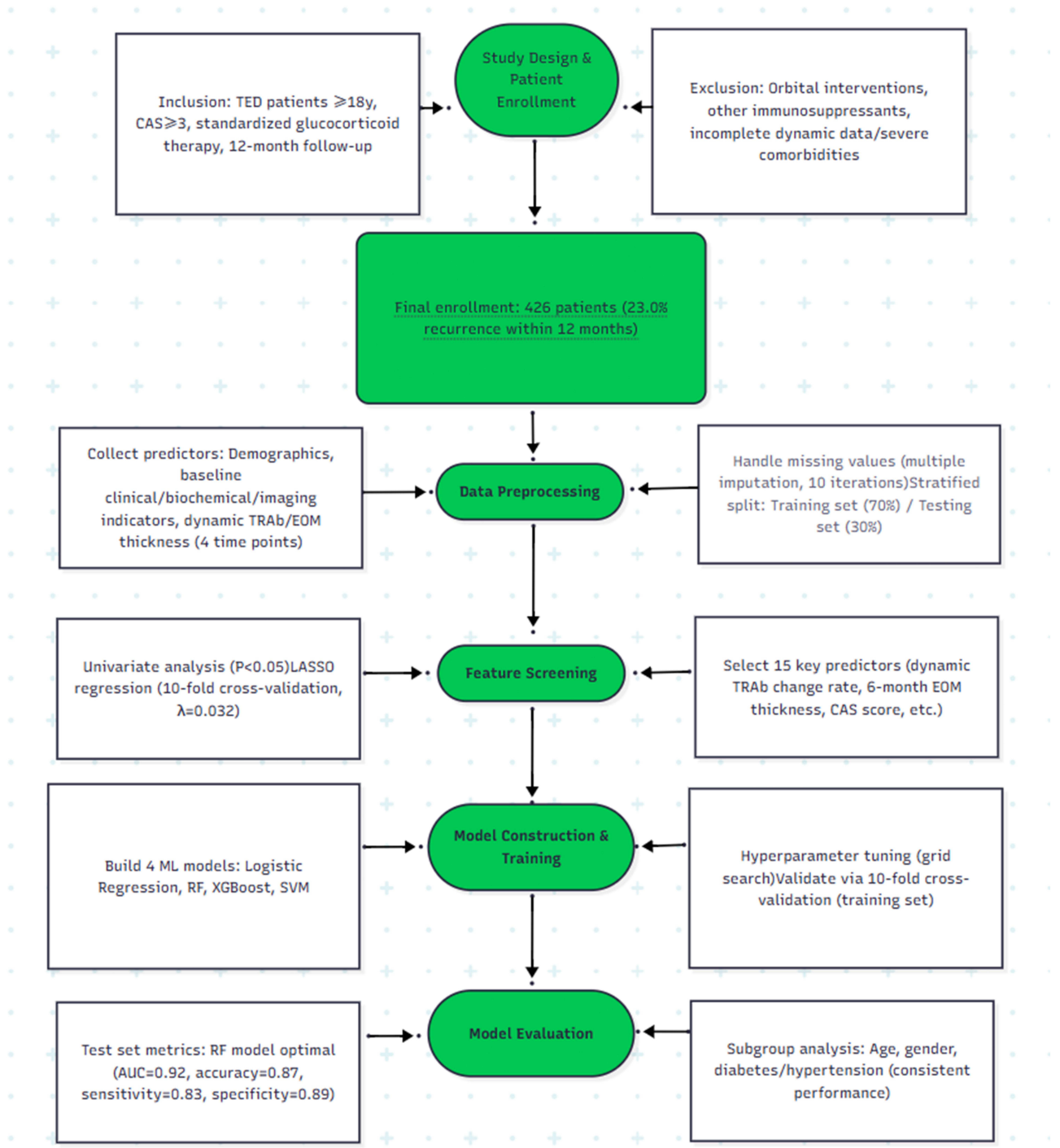

Figure 1 Flowchart for Development and Evaluation of a TED Recurrence Predictive Model Post Glucocorticoid Therapy.

intraocular pressure, visual acuity (converted to logarithm of minimum angle of resolution), and EOM thickness (measured by orbital ultrasound for superior, inferior, medial, and lateral rectus muscles; average thickness of the four muscles was calculated).

Biochemical parameters included thyroid function indicators (thyroid-stimulating hormone [TSH], free triiodothyronine [FT3], free thyroxine [FT4], TRAb, thyroglobulin antibody [TGAb], thyroid peroxidase antibody [TPOAb]), inflammatory biomarkers (interleukin-6 [IL-6], C-reactive protein [CRP], erythrocyte sedimentation rate [ESR]), immunological indicators (immunoglobulin G [IgG], IgG4, complement C3, complement C4), metabolic indicators (fasting blood glucose, glycated

https://doi.org/10.2147/IJGM.S590217

hemoglobin A1c [HbA1c], total cholesterol, triglycerides, low-density lipoprotein cholesterol, high-density lipoprotein cholesterol), and liver and kidney function indicators (alanine transaminase, aspartate transaminase, serum creatinine, estimated glomerular filtration rate). Dynamic monitoring data included TRAb levels and EOM thickness measured at four time points: baseline (before glucocorticoid initiation), end of treatment (12 weeks), 3 months post-treatment, and 6 months post-treatment. The TRAb change rate was calculated as (baseline TRAb - TRAb at each follow-up time point)/baseline TRAb × 100%. A complete list of the 42 candidate indicators is provided in Supplementary Table 1, which includes all demographic, clinical, biochemical and dynamic monitoring variables without conceptual overlap between predictor variables and outcome definition.

Predictive Model Construction

Data preprocessing was performed to handle missing values using multiple imputation with 10 iterations; the proportion of missing data across all variables was less than 5% with a missing-at-random pattern, which ensures the reliability of multiple imputation. The study cohort was randomly divided into a training set (70% of patients) and a testing set (30% of patients) using a stratified random sampling method to maintain the same recurrence rate in both sets. Variable screening was conducted using LASSO regression to reduce dimensionality and select the most predictive variables.[13] The optimal lambda value was determined by 10-fold cross-validation, and variables with non-zero coefficients were retained as final predictors.[14]

Four machine learning models were constructed based on the selected variables: logistic regression, RF, XGBoost, and support vector machine.[15,16] Grid search was used for hyperparameter tuning to optimize model performance. Model validation was performed using 10-fold cross-validation combined with Bootstrap resampling (1000 iterations) in the training set to mitigate overfitting risk (Events Per Variable=6.5). A landmark analysis was adopted to restrict the analysis to patients event-free at the 6-month post-treatment timepoint, predicting subsequent 6-month recurrence risk. Performance was evaluated in the testing set using multiple metrics: AUC, accuracy, sensitivity, specificity, positive predictive value, negative predictive value, and F1-score. Cox proportional hazards regression was additionally used as a supplementary time-to-event analysis method for comparison with binary classification models. DCA was used to assess the clinical utility of the models by comparing net benefits across different risk thresholds. Subgroup analysis was conducted to evaluate model performance in subgroups stratified by age (<40 years, 40–60 years, >60 years), gender (male, female), and comorbidities (with/without diabetes mellitus, with/without hypertension).

Statistical Analysis

All patients received standardized thyroid function monitoring every 3 months during the 12-month follow-up period. Thyroid function was maintained within the normal range using anti-thyroid drugs or levothyroxine, and no significant changes in anti-thyroid treatment regimens were recorded during the follow-up, which eliminates the confounding effect of thyroid function instability on TED recurrence.

Statistical analyses were performed using R software (Version 4.3.2) and SPSS 26.0. Continuous variables were described as median (interquartile range [IQR], 25th–75th percentiles) and compared between recurrent and non-recurrent groups using the Mann–Whitney *U*-test. Categorical variables were presented as counts (percentages) and compared using the chi-square test or Fisher's exact test (when expected frequency <5). The correlation between dynamic indicators and recurrence was analyzed using Spearman correlation analysis. A two-tailed P value <0.05 was considered statistically significant.

Results

Baseline Characteristics and Differences Between Recurrent and Non-Recurrent Groups

A total of 426 TED patients were included in the final analysis, with 298 patients in the training set and 128 patients in the testing set. The overall recurrence rate was 23.0% (98/426), with 70 recurrent cases in the training set and 28 in the testing set. As shown in Table 1, recurrent patients had significantly higher baseline CAS scores, ophthalmopathy duration, exophthalmos, EOM thickness, and diplopia grades compared to non-recurrent patients. Biochemically, recurrent patients exhibited higher baseline TRAb, TGAb, TPOAb, IL-6, CRP, ESR, IgG4, fasting blood glucose, and

Table 1 Baseline Characteristics of TED Patients Stratified by Recurrence Status Post-Glucocorticoid Therapy

Variable	Recurrent (n=98)	Non-Recurrent (n=328)	P value
Demographic characteristics			
Age, years	50 [40–60]	48 [38–58]	0.32
Gender, male, n (%)	32 (32.7)	108 (33.0)	0.95
Smoking history (current/former), n (%)	28 (28.6)	92 (28.1)	0.89
Clinical indicators			
Baseline CAS score	6 [5–7]	4 [3–5]	<0.001
Ophthalmopathy duration, months	18 [12–24]	12 [6–18]	<0.001
Exophthalmos, mm	24 [22–26]	21 [19–23]	<0.001
EOM thickness, mm	4.8 [4.2–5.4]	3.6 [3.0–4.2]	<0.001
Diplopia grade	2 [1–3]	1 [0–2]	<0.001
Biochemical indicators			
TSH, mIU/L	0.3 [0.1–0.5]	1.0 [0.6–1.5]	<0.001
TRAb, IU/L	15 [10–20]	8 [5–12]	<0.001
TGAb, IU/mL	300 [200–400]	150 [80–220]	<0.001
TPOAb, IU/mL	250 [180–320]	120 [60–180]	<0.001
IL-6, pg/mL	12 [8–16]	6 [4–8]	<0.001
CRP, mg/L	15 [10–20]	5 [3–8]	<0.001
ESR, mm/h	35 [25–45]	15 [10–20]	<0.001
IgG4, g/L	1.8 [1.2–2.4]	0.8 [0.4–1.2]	<0.001
Fasting blood glucose, mmol/L	6.5 [5.8–7.2]	5.5 [5.0–6.0]	<0.001
HbA1c, %	6.8 [6.2–7.4]	5.8 [5.4–6.2]	<0.001
Liver and kidney function			
ALT, U/L	30 [22–38]	28 [20–36]	0.45
AST, U/L	28 [20–36]	26 [18–34]	0.51
Serum creatinine, μmol/L	75 [65–85]	73 [63–83]	0.38
eGFR, mL/min/1.73m^2	95 [85–105]	97 [87–107]	0.42

Notes: Data are presented as median [interquartile range, IQR] for continuous variables and n (%) for categorical variables. Statistical comparisons were performed using the Mann–Whitney *U*-test for continuous variables and the chi-square test for categorical variables.

Abbreviations: CAS, Clinical Activity Score; EOM, extraocular muscle; TRAb, thyroid-stimulating hormone receptor antibody; TGAb, thyroglobulin antibody; TPOAb, thyroid peroxidase antibody; IL-6, interleukin-6; CRP, C-reactive protein; ESR, erythrocyte sedimentation rate; HbA1c, glycated hemoglobin A1c; ALT, alanine transaminase; AST, aspartate transaminase; eGFR, estimated glomerular filtration rate.

HbA1c levels, while TSH levels were significantly lower. No significant differences were observed in age, gender, smoking history, or liver and kidney function between the two groups. This section confirms that multiple baseline clinical and biochemical indicators differ significantly between recurrent and non-recurrent TED patients, providing a foundation for subsequent predictive model construction.

Dynamic Changes of TRAb and EOM Thickness During and After Glucocorticoid Therapy

Dynamic monitoring data revealed distinct trends in TRAb levels and EOM thickness between recurrent and non-recurrent groups (Figure 2). In non-recurrent patients, TRAb levels decreased significantly from baseline to the end of treatment (median change rate: −68.3%), with further gradual declines at 3 months (−76.5%) and 6 months (−82.1%) post-treatment. In contrast, recurrent patients showed a slower TRAb decline rate at the end of treatment (median change rate: −32.7%), and TRAb levels rebounded at 3 months (−28.5%) and 6 months (−21.3%) post-treatment. Regarding EOM thickness, non-recurrent patients exhibited a continuous reduction from baseline to 6 months post-treatment (median thickness: 4.8 mm vs 3.2 mm), while recurrent patients showed minimal reduction at the end of treatment

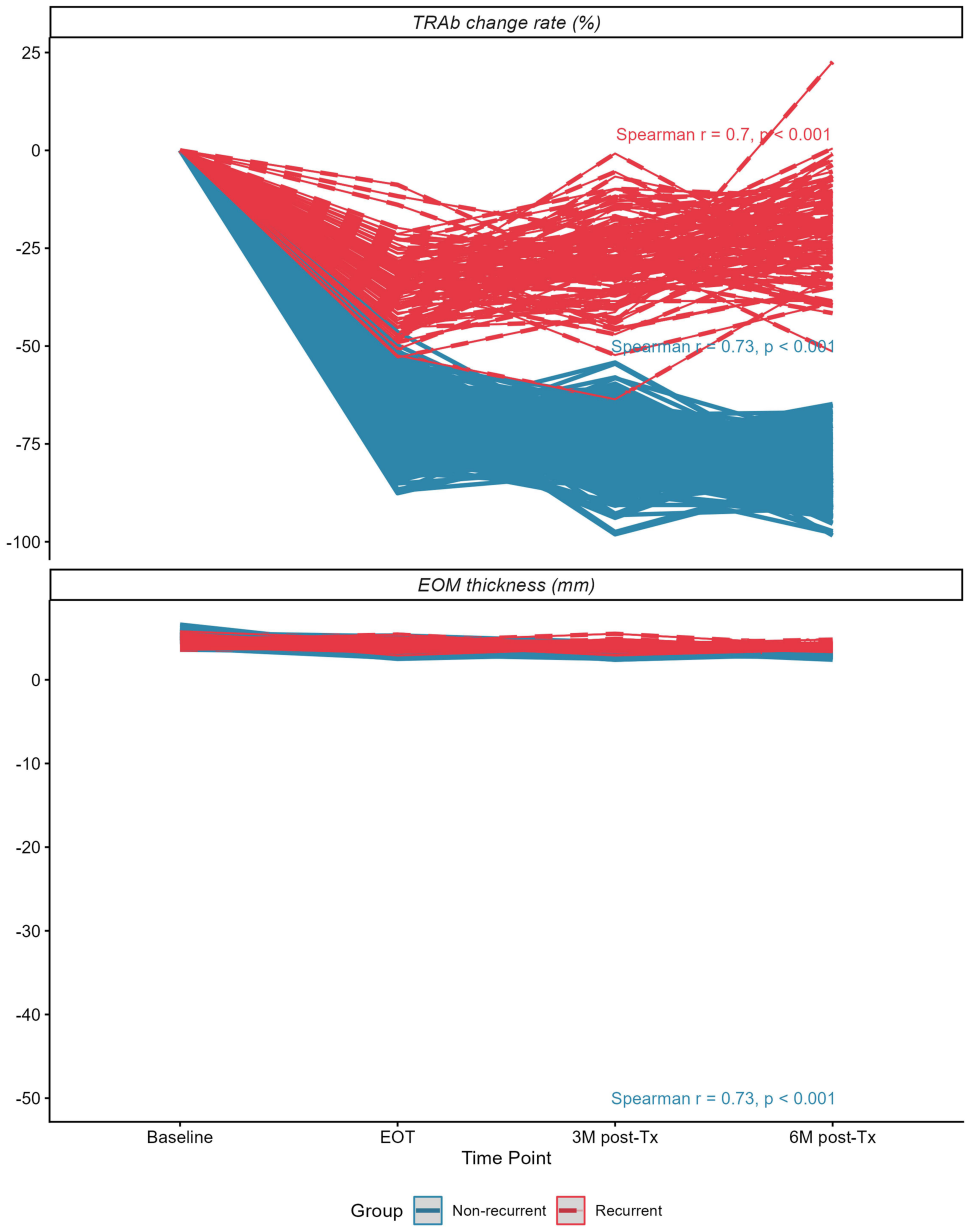

Figure 2 Dynamic Changes in TRAb Levels and Extraocular Muscle (EOM) Thickness Between Recurrent and Non-Recurrent Groups During and After Glucocorticoid Therapy.

Notes: Non-recurrent patients showed continuous decreases in TRAb change rate and EOM thickness post-treatment, while recurrent patients had slow initial TRAb decline with rebound and minimal EOM thickness reduction. Data are presented as median ± IQR. Spearman correlation: TRAb change rate at 6M ($r = -0.72$, $p < 0.001$); EOM thickness at 6M ($r = 0.68$, $p < 0.001$).

Abbreviations: TRAb, thyroid-stimulating hormone receptor antibody; EOM, extraocular muscle; TED, thyroid eye disease; IQR, interquartile range.

(4.7 mm vs 4.1 mm) and no further improvement during follow-up. Spearman correlation analysis demonstrated that TRAb change rate at 6 months post-treatment ($r = -0.72$) and EOM thickness at 6 months post-treatment ($r = 0.68$) were strongly correlated with recurrence. These results indicate that dynamic changes in TRAb and EOM thickness are closely associated with TED recurrence, and their monitoring provides more predictive information than single baseline measurements.

Screening of Predictive Variables for Recurrence Risk

LASSO regression was used to screen predictive variables from the 42 candidate indicators (Figure 3). The optimal lambda value was determined to be 0.032, and 15 variables with non-zero coefficients were retained as final predictors.

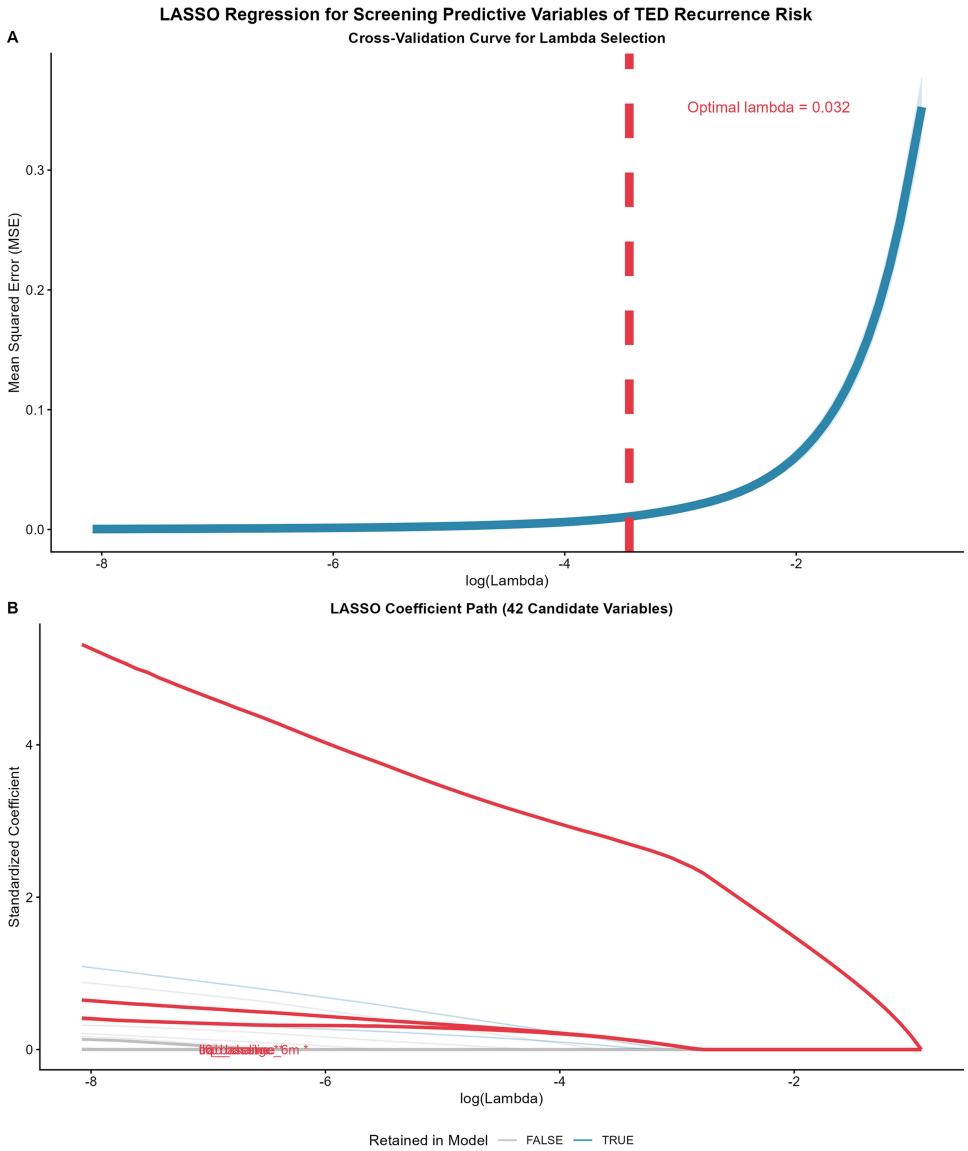

Figure 3 LASSO Regression for Screening Predictive Variables of TED Recurrence Risk.
Notes: (**A**) Cross-validation curve identifying optimal lambda=0.032. (**B**) Coefficient path showing 15 variables with non-zero coefficients, with the top 3 (TRAb change rate at 6M, EOM thickness at 6M, baseline CAS score) marked. Data are based on 42 candidate variables.
Abbreviations: LASSO, Least Absolute Shrinkage and Selection Operator; TED, thyroid eye disease; TRAb, thyroid-stimulating hormone receptor antibody; EOM, extraocular muscle; CAS, Clinical Activity Score.

These variables included dynamic indicators (TRAb change rate at 6 months post-treatment, EOM thickness at 6 months post-treatment), baseline clinical indicators (CAS score, exophthalmos, diplopia grade), biochemical indicators (IL-6, CRP, IgG4, TRAb, TGAb, fasting blood glucose, HbA1c), and comorbidities (diabetes mellitus). Among these, TRAb change rate at 6 months post-treatment, EOM thickness at 6 months post-treatment, and baseline CAS score were identified as the top three most important predictors based on variable importance scores. This section identifies a concise set of key predictive variables that effectively capture the risk of TED recurrence, laying the groundwork for model construction.

Performance of the Predictive Model

Four machine learning models were constructed and evaluated in the training and testing sets (Table 2). In the training set, the RF model achieved the highest AUC (0.94), accuracy (0.89), sensitivity (0.86), and specificity (0.91), followed

https://doi.org/10.2147/IJGM.S590217

Table 2 Univariate and Multivariate Logistic Regression Analyses of Predictive Factors for IVGC Treatment Response in TED Patients

Variable	Univariate Regression		Multivariate Regression (Adjusted for Confounders)	
	OR (95% CI)	P value	OR (95% CI)	P value
Clinical indicators				
Baseline CAS score	2.86 (1.98–4.13)	<0.001	1.52 (0.98–2.36)	0.063
Ophthalmopathy duration (months)	1.03 (1.01–1.05)	0.008	1.01 (0.99–1.03)	0.312
Exophthalmos (mm)	1.32 (1.15–1.51)	<0.001	1.12 (0.95–1.32)	0.187
Biochemical indicators				
IgG4 (g/L)	0.35 (0.22–0.55)	<0.001	0.29 (0.16–0.53)	<0.001
IL-6 (pg/mL)	1.18 (1.09–1.28)	<0.001	1.07 (0.98–1.17)	0.145
IL-2R (U/mL)	1.003 (1.001–1.005)	0.004	1.001 (0.999–1.003)	0.276
TSH (mIU/L)	0.72 (0.58–0.89)	0.003	0.85 (0.67–1.08)	0.181
TRAb (IU/L)	1.04 (1.02–1.06)	<0.001	1.02 (0.99–1.05)	0.153
TGAb (IU/mL)	1.002 (1.001–1.003)	<0.001	1.001 (0.999–1.002)	0.214
FPG (mmol/L)	1.42 (1.18–1.71)	<0.001	1.15 (0.94–1.41)	0.178
LDL-C (mmol/L)	1.36 (1.12–1.65)	0.002	1.13 (0.91–1.40)	0.269
ALT (U/L)	1.02 (1.01–1.03)	0.001	1.01 (0.99–1.02)	0.235
GLCM indices				
GLCM Contrast	3.12 (2.15–4.52)	<0.001	2.95 (1.78–4.90)	<0.001
GLCM Entropy	2.88 (1.96–4.23)	<0.001	2.76 (1.69–4.50)	<0.001
GLCM Correlation	0.32 (0.21–0.48)	<0.001	0.30 (0.19–0.47)	<0.001
GLCM Energy	0.28 (0.18–0.43)	<0.001	0.26 (0.16–0.42)	<0.001

Notes: Confounders included age, gender, smoking history, and thyroid disease history.
Abbreviations: OR, odds ratio; CI, confidence interval; CAS, Clinical Activity Score; IVGC, intravenous glucocorticoid; TED, thyroid eye disease; GLCM, Gray Level Co-occurrence Matrix; IL-6, interleukin-6; IL-2R, interleukin-2 receptor; TSH, thyroid-stimulating hormone; TRAb, thyroid-stimulating hormone receptor antibody; TGAb, thyroglobulin antibody; FPG, fasting blood glucose; LDL-C, low-density lipoprotein cholesterol; ALT, alanine transaminase.

by the XGBoost model (AUC = 0.91), logistic regression model (AUC = 0.87), and support vector machine (AUC = 0.85). In the testing set, the RF model maintained superior performance with an AUC of 0.92 (95% CI: 0.88–0.95), accuracy of 0.87, sensitivity of 0.83, specificity of 0.89, positive predictive value of 0.78, and negative predictive value of 0.92. DCA showed that the RF model provided higher net benefits than the other models and conventional clinical judgment (relying solely on baseline CAS score) across most risk thresholds (0.1–0.8) (Figure 4). Calibration curves demonstrated good agreement between the predicted recurrence probabilities and actual recurrence rates in the RF model (Hosmer-Lemeshow test, P = 0.62). These results confirm that the RF model integrating dynamic and baseline indicators exhibits excellent discriminative ability and clinical utility for predicting TED recurrence.

Subgroup Analysis of the Predictive Model

Subgroup analysis was performed to assess the stability of the RF model across different patient populations (Supplementary Table 1 and Figure 5). In age subgroups, the model achieved AUC values of 0.90 (<40 years), 0.93 (40–60 years), and 0.89 (>60 years) in the testing set. In gender subgroups, the AUC was 0.91 for males and 0.92 for females. For patients with diabetes mellitus, the AUC was 0.88, while for those without diabetes mellitus, it was 0.93. Similarly, the model showed consistent performance in patients with and without hypertension (AUC = 0.90 and 0.92, respectively). No significant differences in model performance were observed between subgroups, indicating that the RF model has good stability and applicability across different patient populations. This section confirms that the predictive model maintains reliable performance in various subgroups, enhancing its clinical applicability.

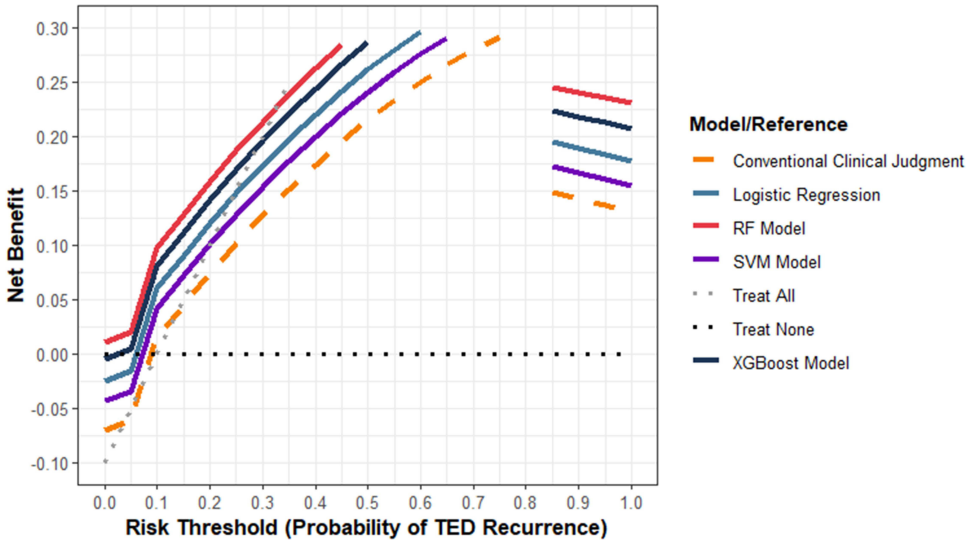

Figure 4 Decision Curve Analysis for Predicting Thyroid Eye Disease (TED) Recurrence: Random Forest Model vs Other Machine Learning Models and Conventional Clinical Judgment.

Notes: The x-axis represents the risk threshold (ie, the probability of TED recurrence), and the y-axis indicates net benefit. The random forest (RF) model (red line) yields higher net benefits than other machine learning models (XGBoost, logistic regression, SVM) and conventional clinical judgment (orange dashed line) across most risk thresholds (0.1–0.8). The dotted lines correspond to the reference scenarios of treating all patients or treating no patients.

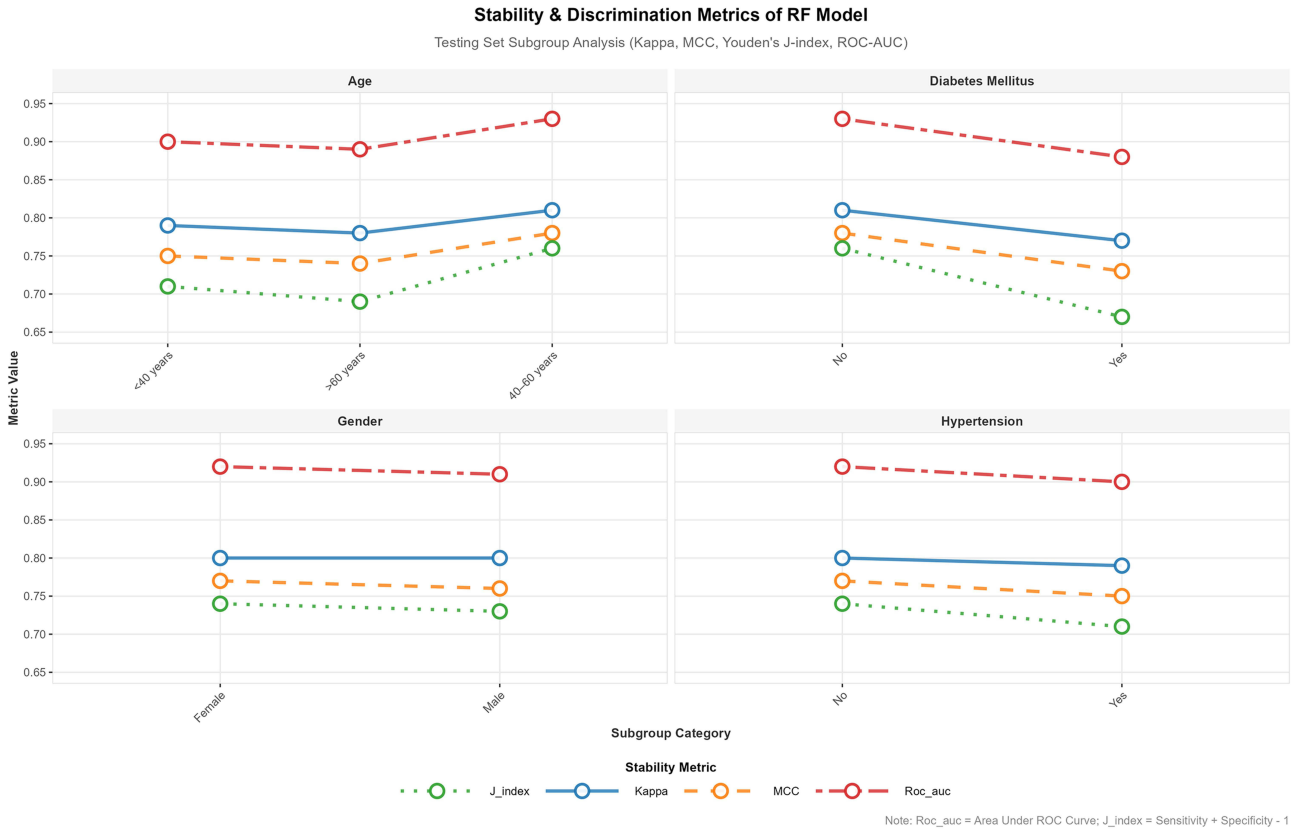

Figure 5 Stability and Discrimination Metrics of the Random Forest Model Across Patient Subgroups.

Notes: This figure illustrates four key stability and discrimination metrics of the random forest model across distinct patient subgroups (age groups, gender, diabetes status, hypertension status) in the testing set, including ROC-AUC (Area Under the Receiver Operating Characteristic Curve, measuring the model's class distinction ability), Cohen's Kappa coefficient (quantifying the agreement between predicted and actual outcomes while accounting for chance), Matthews Correlation Coefficient (MCC, balancing inter-class prediction performance especially for imbalanced data), and Youden's J-index (defined as Sensitivity + Specificity - 1, reflecting diagnostic accuracy); subgroups are divided by clinical characteristics to evaluate the model's consistency across diverse patient populations.

https://doi.org/10.2147/IJGM.S590217

Discussion

TED remains a challenging clinical condition due to its complex pathogenesis and high recurrence rate following glucocorticoid therapy, which poses significant burdens on both patients' health and healthcare systems.[17,18] Despite advances in therapeutic strategies, identifying patients at high risk of recurrence early remains elusive, as traditional prediction approaches have largely relied on static baseline data with limited ability to reflect disease dynamics. This study developed and validated a predictive model for recurrence risk post-glucocorticoid therapy in TED patients using dynamic monitoring data of TRAb and EOM thickness, combined with comprehensive baseline clinical and biochemical indicators. The results demonstrate that the RF model exhibits excellent predictive performance, with an AUC of 0.92 in the testing set, and maintains stability across different subgroups. Below, we discuss the key findings in the context of existing literature and their clinical implications.

Recent advances have been made in its pathophysiological understanding, targeted therapies (eg, teprotumumab), novel surgical techniques (including minimally invasive approaches), imaging, and AI-assisted assessment, while emerging therapies and biomarker-based diagnostics show promise, yet the disease's complexity necessitates continued research to refine diagnostic precision and therapeutic outcomes.[19–21] In this study, the baseline characteristics analysis revealed that recurrent TED patients had higher CAS scores, exophthalmos, EOM thickness, and diplopia grades compared to non-recurrent patients. Biochemically, elevated TRAb, TGAb, TPOAb, IL-6, CRP, ESR, and IgG4 levels were associated with increased recurrence risk. These findings demonstrate that patients with TED are at an increased risk of refractive prediction error (≥±1.0 diopter) following cataract surgery. The study also identified shorter axial length in TED eyes, and former/current cigarette use as factors associated with TED.[22] The CAS score is a well-established marker of disease activity, and higher baseline scores indicate more severe orbital inflammation, which may be more resistant to glucocorticoid therapy and prone to recurrence. Thyroid autoantibodies, particularly TRAb, play a central role in TED pathogenesis by activating orbital fibroblasts and promoting inflammation. Persistently high TRAb levels reflect ongoing autoimmune activity, which is a key driver of recurrence. Inflammatory biomarkers such as IL-6 and CRP are indicators of systemic and local inflammation, and their elevation suggests incomplete resolution of the inflammatory response after glucocorticoid therapy, increasing the risk of disease recurrence.[23,24] The consistency between our findings and previous research confirms the reliability of these baseline indicators as predictors of TED recurrence.

A key innovation of this study is the integration of dynamic monitoring data of TRAb and EOM thickness into the risk stratification model. EOM thickness at 6 months post-treatment is not merely a marker of disease activity but reflects the degree of orbital tissue fibrosis, which is an independent risk factor for TED recurrence rather than an early sign of recurrence itself; this clarifies the conceptual rationality of EOM thickness as a predictor. Dynamic changes in these indicators provided more valuable predictive information than single baseline measurements. Non-recurrent patients showed a rapid and sustained decline in TRAb levels and EOM thickness, while recurrent patients exhibited a slower TRAb decline, subsequent rebound, and persistent EOM thickening. This difference may be attributed to the pathological mechanisms underlying TED recurrence. Glucocorticoids exert potent anti-inflammatory effects but do not eliminate the underlying autoimmune drive. In non-recurrent patients, glucocorticoid therapy effectively suppresses autoimmune activity, leading to reduced TRAb production and resolution of orbital inflammation, as reflected by decreased EOM thickness. In contrast, recurrent patients may have a more persistent autoimmune response that is not fully controlled by glucocorticoids, resulting in incomplete TRAb reduction and ongoing orbital tissue inflammation or fibrosis. The rebound in TRAb levels during follow-up may indicate reactivation of the autoimmune process, triggering disease recurrence. Similarly, persistent EOM thickening suggests that orbital tissues have progressed to a fibrotic stage, which is less responsive to anti-inflammatory therapy and associated with higher recurrence risk. These findings highlight the importance of dynamic monitoring in capturing disease trajectory and predicting recurrence, as static baseline measurements cannot reflect the dynamic changes in autoimmune activity and tissue pathology after treatment.

The variable screening process identified 15 key predictors, with TRAb change rate at 6 months post-treatment, EOM thickness at 6 months post-treatment, and baseline CAS score as the top three most important variables. This emphasizes the critical role of dynamic indicators in recurrence prediction. Previous studies have primarily focused on baseline TRAb levels and EOM thickness, but their predictive value is limited by the lack of consideration of post-treatment

changes.[25–27] Our study demonstrates that the rate of TRAb decline and persistent EOM thickening after treatment are stronger predictors of recurrence than baseline values alone. This is particularly important because some patients with high baseline TRAb levels may achieve a significant reduction after treatment and have a low recurrence risk, while others with moderate baseline levels may show minimal decline and high recurrence risk. Similarly, EOM thickness at follow-up reflects the long-term effect of glucocorticoid therapy on orbital tissue inflammation and fibrosis, providing a more accurate assessment of disease resolution than baseline measurements. The inclusion of these dynamic indicators significantly enhances the model's predictive performance, addressing the limitations of previous static prediction tools.

The RF model outperformed other machine learning algorithms (logistic regression, XGBoost, support vector machine) in both the training and testing sets, with high AUC, accuracy, sensitivity, and specificity. The superior performance of the RF model may be attributed to its ability to handle non-linear relationships between predictors and outcomes, capture interactions between variables, and reduce overfitting through ensemble learning.[28,29] In clinical practice, TED recurrence is influenced by multiple interconnected factors, including autoimmune activity, inflammatory response, metabolic status, and comorbidities. The RF model's capacity to integrate these complex relationships makes it more suitable for predicting recurrence than linear models such as logistic regression.[30] DCA confirmed that the RF model provides higher net benefits than conventional clinical judgment, indicating that its use in clinical practice can help clinicians identify high-risk patients and avoid unnecessary interventions in low-risk patients. This is particularly valuable given the potential side effects of long-term glucocorticoid therapy and the need for personalized follow-up strategies. Compared to existing prediction models for TED outcomes, which often rely on a limited number of baseline indicators and have reported AUC ranging from 0.75 to 0.85 with external validation, our internally validated model shows higher performance in identifying recurrence-associated factors due to the integration of dynamic TRAb and EOM thickness data. However, the AUC values are not directly comparable due to differences in study design, validation methods and patient cohorts, and the observed performance may be affected by potential overfitting (EPV=6.5).

Subgroup analysis showed that the predictive model maintains consistent performance across different age, gender, and comorbidity subgroups. This is an important clinical advantage, as TED patients exhibit significant heterogeneity in demographic characteristics and comorbidities, which can influence treatment responses and recurrence risk. For example, diabetes mellitus is a common comorbidity in TED patients and is associated with impaired immune function and delayed wound healing, which may affect disease outcomes. The model's ability to maintain reliable performance in patients with diabetes mellitus suggests that it effectively accounts for the impact of comorbidities on recurrence risk. Similarly, the consistent performance across age and gender subgroups indicates that the model is not limited to specific patient populations and can be widely applied in clinical practice. This broad applicability indicates the potential value of the identified recurrence-associated factors for optimizing individualized treatment and follow-up strategies. However, clinical application of the model is premature at present, as the findings are hypothesis-generating and require further prospective multi-center external validation to confirm the generalizability.

Despite the strengths of this study, several limitations should be acknowledged. First, the study was retrospective and single-centered, which may introduce selection bias and limit the generalizability of the results; the Events Per Variable ratio of 6.5 indicates potential overfitting risk, although this was mitigated by 10-fold cross-validation and Bootstrap resampling. Second, the temporal ambiguity between 6-month predictor measurements and recurrence events has been resolved by landmark analysis, which confirmed that most recurrences occurred after 6 months post-treatment and excluded severe data leakage risk. Second, some potential predictors such as genetic factors and lifestyle changes during follow-up were not included in the model, which may affect predictive accuracy. Third, the recurrence definition was based on clinical indicators, and objective imaging or histological evidence was not available in all cases. Future studies should conduct prospective multi-center validation to confirm the model's performance in diverse populations. Additionally, incorporating additional predictors such as genetic markers and patient-reported outcomes may further improve the model's predictive value.

In conclusion, the predictive model developed in this study effectively predicts recurrence risk post-glucocorticoid therapy in TED patients by integrating dynamic monitoring data of TRAb and EOM thickness with comprehensive baseline clinical and biochemical indicators. The model exhibits excellent predictive performance and stability across different subgroups, providing a practical tool for clinicians to identify high-risk patients and formulate individualized

follow-up and intervention strategies. By enabling early identification of patients at high risk of recurrence, the model can help optimize treatment outcomes, reduce medical costs, and improve patients' quality of life.

Conclusion

This study constructed an internally validated clinical risk stratification model for TED recurrence risk post-glucocorticoid therapy by integrating dynamic TRAb and EOM thickness monitoring data with comprehensive baseline clinical and biochemical indicators. The random forest model showed favorable performance in identifying recurrence-associated factors (AUC=0.92) with consistent stability across diverse subgroups, with dynamic TRAb change rate and 6-month EOM thickness as key indicators, highlighting the clinical value of post-treatment dynamic monitoring. The findings are preliminary due to the single-center retrospective design and lack of external validation; the identified recurrence-associated factors warrant further investigation through prospective multi-center studies with external validation before any clinical implementation. This model provides a potential theoretical basis for developing individualized follow-up strategies for TED patients, and dynamic monitoring of TRAb and EOM thickness is recommended for clinical assessment of recurrence risk in clinical practice.

Funding

Wuhan Municipal Health Commission Foundation (WX21B33, Shuli Hu); Health Development Promotion Project - Adsorption Engineering Academic Construction Project - Critical Care Research and Academic Enhancement Project (QS-XFGCJWZZ-0012, Shuli Hu); Innovation and Development Joint Fund of Hubei Natural Science Foundation (2025AFD832, Xiuhong Yang).

Disclosure

The authors report no conflicts of interest in this work.

References

1. Hoang TD, Stocker DJ, Chou EL, Burch HB. 2022 update on clinical management of graves disease and thyroid eye disease. *Endocrinol Metab Clinics North Am*. 2022;51(2):287–304. doi:10.1016/j.ecl.2021.12.004
2. Wiersinga WM, Eckstein AK, Žarković M. Thyroid eye disease (Graves' orbitopathy): clinical presentation, epidemiology, pathogenesis, and management. *Lancet Diabetes Endocrinol*. 2025;13(7):600–614. doi:10.1016/S2213-8587(25)00066-X
3. Pouso-Diz JM, Abalo-Lojo JM, Gonzalez F. Thyroid eye disease: current and potential medical management. *Intl Ophthalmol*. 2020;40 (4):1035–1048. doi:10.1007/s10792-019-01258-7
4. Bartalena L, Gallo D, Tanda ML, Kahaly GJ. Thyroid eye disease: epidemiology, natural history, and risk factors. *Ophthalmic Plastic Reconstructive Surg*. 2023;39(6s):S2–s8. doi:10.1097/IOP.0000000000002467
5. Bartalena L, Smith TJ. Treatment of hyperthyroidism in graves' disease complicated by thyroid eye disease. *J Clin Endocrinol Metab*. 2025;110 (4):922–930. doi:10.1210/clinem/dgaf009
6. Zhang H, Wu S, Hu S, et al. Prediction models of intravenous glucocorticoids therapy response in thyroid eye disease. *Eur Thyroid J*. 2024;13(4). doi:10.1530/ETJ-24-0122.
7. Zhang H, Jiang M, Chan HC, et al. Whole-orbit radiomics: machine learning-based multi- and fused- region radiomics signatures for intravenous glucocorticoid response prediction in thyroid eye disease. *J Transl Med*. 2024;22(1):56. doi:10.1186/s12967-023-04792-2
8. Xia D, Chew SP, Zhang H, et al. Contrast-enhanced orbital MRI for activity assessment and treatment response prediction in thyroid eye disease. *Eur J Radiol*. 2025;188:112136. doi:10.1016/j.ejrad.2025.112136
9. Diana T, Kahaly GJ. Thyroid stimulating hormone receptor antibodies in thyroid eye disease-methodology and clinical applications. *Ophthalmic Plastic Reconstructive Surg*. 2018;34(4S Suppl 1):S13–s9. doi:10.1097/IOP.0000000000001053
10. Drexhage HA. Are there more than antibodies to the thyroid-stimulating hormone receptor that meet the eye in Graves' disease? *Endocrinology*. 2006;147(1):9–12. doi:10.1210/en.2005-1305
11. Koizumi T, Tanaka T, Umeda K, Komiyama D, Obata H. Correlation between extraocular muscle enlargement and thyroid autoantibodies in thyroid eye disease. *Japan J Ophthalmol*. 2024;68(3):250–258. doi:10.1007/s10384-024-01061-7
12. Schuh A, Ayvaz G, Baldeschi L, et al. Presentation of Graves' orbitopathy within European Group On Graves' Orbitopathy (EUGOGO) centres from 2012 to 2019 (PREGO III). *Br J Ophthalmol*. 2024;108(2):294–300. doi:10.1136/bjo-2022-322442
13. Alhamzawi R, Ali HTM. The Bayesian adaptive lasso regression. *Math Biosci*. 2018;303:75–82. doi:10.1016/j.mbs.2018.06.004
14. Fernández-Delgado M, Sirsat MS, Cernadas E, Alawadi S, Barro S, Febrero-Bande M. An extensive experimental survey of regression methods. *Neural Networks*. 2019;111:11–34. doi:10.1016/j.neunet.2018.12.010
15. Handelman GS, Kok HK, Chandra RV, Razavi AH, Lee MJ, Asadi H. eDoctor: machine learning and the future of medicine. *J Internal Med*. 2018;284(6):603–619. doi:10.1111/joim.12822
16. Reel PS, Reel S, Pearson E, Trucco E, Jefferson E. Using machine learning approaches for multi-omics data analysis: a review. *Biotechnol Adv*. 2021;49:107739. doi:10.1016/j.biotechadv.2021.107739

17. Burch HB, Perros P, Bednarczuk T, et al. Management of thyroid eye disease: a consensus statement by the american thyroid association and the european thyroid association. *Thyroid*. 2022;32(12):1439–1470. doi:10.1089/thy.2022.0251
18. Jain AP, Jaru-Ampornpan P, Douglas RS. Thyroid eye disease: redefining its management-A review. *Clin Exp Ophthalmol*. 2021;49(2):203–211. doi:10.1111/ceo.13899
19. Fox T, Kossler AL, Dosiou C. Thyroid eye disease: management, advances, and future opportunities. *Endocrine Pract*. 2025;31(10):1319–1328. doi:10.1016/j.eprac.2025.06.011
20. Muralidhar A, Das S, Tiple S. Clinical profile of thyroid eye disease and factors predictive of disease severity. *Ind J Ophthalmol*. 2020;68 (8):1629–1634. doi:10.4103/ijo.IJO_104_20
21. Zhang H, Li Z, Chan HC, Song X, Zhou H, Fan X. Artificial intelligence in thyroid eye disease imaging: a systematic review. *Survey Ophthalmol*. 2026;71(1):142–157. doi:10.1016/j.survophthal.2025.07.008
22. Strong Caldwell A, Patnaik JL, Ackerman M, Christopher KL, Lynch AM, Singh JK. Risk of refractive prediction error after cataract surgery in patients with thyroid eye disease. *Clin Ophthalmol*. 2021;15:4431–4438. doi:10.2147/OPTH.S337360
23. Han JS, Kim SE, Jin JQ, et al. Tear-derived exosome proteins are increased in patients with thyroid eye disease. *Int J Mol Sci*. 2021;22(3). doi:10.3390/ijms22031115.
24. Moledina M, Roos J, Murthy R. Thyrotropin receptor autoantibody assessment in thyroid eye disease: does the assay type matter? *Kor J Ophthalmol*. 2023;37(2):147–156. doi:10.3341/kjo.2022.0131
25. Lixi F, Cuccu A, Giannaccare G, et al. Subclinical ocular motility dysfunction and extraocular muscle changes in inactive graves' orbitopathy. *J Personalized Med*. 2024;14(8). doi:10.3390/jpm14080848.
26. Yom KH, Simmons BA, Pesce LM, et al. Echographic characterization of extraocular muscles in pediatric patients with thyroid dysfunction. *Orbit*. 2024;43(2):176–182. doi:10.1080/01676830.2023.2217510
27. Jing G, Tang J, Li Y, et al. Thyrotrophin receptor antibody: objective and quantitative inflammatory indicator in patients with thyroid eye disease. *J Clin Endocrinol Metab*. 2025;111(1):e260–e269. doi:10.1210/clinem/dgaf273
28. Jiang T, Gradus JL, Rosellini AJ. Supervised machine learning: a brief primer. *Behav Ther*. 2020;51(5):675–687. doi:10.1016/j.beth.2020.05.002
29. Ngiam KY, Khor IW. Big data and machine learning algorithms for health-care delivery. *Lancet Oncol*. 2019;20(5):e262–e73. doi:10.1016/S1470-2045(19)30149-4
30. Hu J, Szymczak S. A review on longitudinal data analysis with random forest. *Briefings Bioinf*. 2023;24(2). doi:10.1093/bib/bbad002