Machine Learning for Predicting Recurrent Course in Uveitis Using Baseline Clinical Characteristics

William Rojas-Carabali,[1–3] Carlos Cifuentes-González,[1,3] Anna Utami,[4] Manisha Agarwal,[5] John H. Kempen,[6–9] Edmund Tsui,[10] Robert P. Finger,[11] Alok Sen,[12] Anita Chan,[13] Ariel Schlaen,[14] Vishali Gupta,[15] Alejandra de-la-Torre,[16] Bernett Lee,[2,17–20] and Rupesh Agrawal[1–3,21,22]; on behalf of the PROTON Study Group

[1]Programme for Ocular Inflammation & Infection Translational Research, Department of Ophthalmology, National Healthcare Group Eye Institute, Tan Tock Seng Hospital, Singapore, Singapore
[2]Lee Kong Chian School of Medicine, Nanyang Technological University, Singapore, Singapore
[3]National Healthcare Group Eye Institute, Tan Tock Seng Hospital, Singapore, Singapore
[4]Jakarta Eye Centre, Jakarta, Indonesia
[5]Dr Shroff's Charity Eye Hospital, New Delhi, India
[6]Department of Ophthalmology and Schepens Eye Research Institute, Massachusetts Eye and Ear/ Harvard Medical School, Boston, Massachusetts, United States
[7]Sight for Souls, Bellevue, Washington, United States
[8]Department of Ophthalmology, Addis Ababa University, Addis Ababa, Ethiopia
[9]MyungSung Christian Medical Center (MCM) Eye Unit, MCM Comprehensive Specialized Hospital, MyungSung Medical School, Addis Ababa, Ethiopia
[10]Stein Eye Institute, Department of Ophthalmology, David Geffen School of Medicine, Los Angeles, California, United States
[11]Department of Ophthalmology, Universitätsmedizin Mannheim, University of Heidelberg, Heidelberg, Germany
[12]SNC Eye Hospital, Jagadhri, India
[13]Singapore National Eye Centre, Singapore
[14]Hospital de Clinicas Jose de San Martín Universidad de Buenos Aires/Hospital Universitario Austral, Buenos Aires, Argentina
[15]Advanced Eye Centre, Post graduate Institute of Medical Education and Research (PGIMER), Chandigarh, India
[16]Neuroscience (NEUROS) Research Group, Neurovitae Research Center, Institute of Translational Medicine (IMT), Escuela de Medicina y Ciencias de la Salud, Universidad del Rosario, Bogotá, Colombia
[17]Singapore Immunology Network (SIgN), Agency for Science, Technology and Research (A*STAR), Singapore, Singapore
[18]Centre for Biomedical Informatics, Lee Kong Chian School of Medicine, Nanyang Technological University, Singapore, Singapore
[19]Infectious Disease Labs (ID Labs), Agency for Science, Technology and Research (A*STAR), Singapore, Singapore
[20]Khoo Teck Puat Hospital, National Healthcare Group, Singapore, Singapore
[21]Singapore Eye Research Institute, Singapore, Singapore
[22]Duke NUS Medical School, Singapore, Singapore

Correspondence: Rupesh Agrawal, National Healthcare Group Eye Institute, Tan Tock Seng Hospital, Singapore 308433, Singapore; rupeshttsh@gmail.com.

Received: June 9, 2025
Accepted: August 4, 2025
Published: August 27, 2025

Citation: Rojas-Carabali W, Cifuentes-González C, Utami A, et al. Machine learning for predicting recurrent course in uveitis using baseline clinical characteristics. *Invest Ophthalmol Vis Sci.* 2025;66(11):67. https://doi.org/10.1167/iovs.66.11.67

Purpose. We developed and evaluated machine learning models for predicting the risk of recurrent uveitis using baseline clinical characteristics, to inform clinical decision-making and risk stratification.

Methods. A retrospective analysis was conducted using the Ocular Autoimmune Systemic Inflammatory Infectious Study registry, including 966 patients (1432 eyes) with uveitis. Three machine learning classifiers—random Forest, eXtreme Gradient Boosting, and radial basis function support vector classifier—were trained on preprocessed baseline demographic and clinical data. Predictors were selected through bivariate analysis with false discovery rate correction. Models were optimized using grid search with five-fold stratified cross-validation. Performance was evaluated on a hold-out test set using accuracy, sensitivity, specificity, area under the receiver operating characteristic curve, and Shapley additive explanations values for feature importance.

Results. The random Forest model achieved the highest test accuracy (0.77), with high specificity (0.93) but modest sensitivity (0.44) for identifying recurrences. eXtreme Gradient Boosting and radial basis function support vector classifier showed comparable accuracies (0.73 and 0.74, respectively) but slightly lower sensitivities. Shapley additive explanation analysis identified vitreous haze, retrolental cells, and noninfectious etiology as key predictors. Learning curves indicated that model performance stabilized with the available sample size, suggesting adequate training data.

Conclusions. Machine learning models, particularly random Forest, effectively identified patients at low risk of uveitis recurrence, offering high specificity. However, sensitivity

remained limited, highlighting challenges in predicting infrequent events in a heterogeneous disease population.

Keywords: machine learning, uveitis recurrence, risk stratification, ocular inflammation, SHAP analysis

Uveitides encompass a heterogeneous group of inflammatory ocular disorders that pose a significant risk of vision loss if inadequately managed.[1,2] It is estimated to account for up to 10% of blindness in developed countries, underscoring the necessity for early intervention and accurate prognosis.[3] The disease progression can follow an acute, chronic, or recurrent course, as classified by the Standardization Uveitis Nomenclature working group.[4] In cases of recurrence, the relapsing nature of uveitis significantly complicates clinical decision-making, because disease reactivation is driven by multifactorial etiologies, immune dysregulation, and patient-specific factors (i.e., aging and emotional factors).[5]

Predicting the course of uveitis remains a significant challenge. Although certain clinical characteristics, such as etiology, offer some insight, they are not definitive predictors, because the same disease can manifest differently across patients.[5] For instance, HLAB27$^+$ uveitis, traditionally classified as recurrent, can follow a chronic course in up to 23.1% of cases.[6] Similarly, toxoplasmosis, a common infectious cause of uveitis, is associated with recurrence rates as high as 42%.[7] Other etiologies, such as cytomegalovirus anterior uveitis and Behçet disease uveitis, also demonstrate variable recurrence rates, with cytomegalovirus anterior uveitis recurring in approximately 37% of cases[8] and Behçet disease uveitis in about 12% of cases.[9]

Recent advances in machine learning (ML) have highlighted its potential to model complex biomedical processes, offering a data-driven approach to risk stratification.[10,11] These techniques have shown promise in predicting treatment responses and disease progression in various ophthalmic conditions.[12,13] We hypothesize that baseline clinical characteristics may serve as reliable predictors of disease recurrence, independent of the underlying pathology. The ability to stratify patients based on recurrence risk at the time of initial presentation could enable more precise therapeutic interventions and improve long-term outcomes.

This study aims to develop an etiology-agnostic predictive framework for uveitis recurrence using supervised ML techniques and clinical data from the Ocular Autoimmune Systemic Inflammatory Infectious Study (OASIS) registry. By evaluating multiple ML algorithms, we aimed to identify the most effective predictive model and determine the baseline clinical features most strongly associated with a recurrent course of the disease.

Methods

Data Sources and Collection

This study used clinical data from the OASIS registry, a comprehensive database containing longitudinal records of patients with uveitis. Data collection spanned from May 2021 to March 2024, encompassing demographic, clinical, and laboratory parameters at baseline and follow-up visits. More details about the registry have been published in Ng et al.[14] The dataset was divided into development and evaluation cohorts to ensure robust model validation.

Participants

As of the time of analysis (March 14, 2025), the registry included 3173 patients diagnosed with inflammatory ocular diseases by a uveitis specialist. For this study, we included patients for whom the variable recurrences was recorded as a binary variable (yes or no) at the end of follow-up. Patients with incomplete clinical histories, insufficient follow-up, or other inflammatory ocular diseases were excluded. This resulted in a final cohort of 966 patients (1,432 eyes) from 11 clinical centers (Supplementary Table S1). Figure 1 summarizes the study methodology.

Outcome Definition

In this study, recurrence was defined as the reappearance of inflammatory ocular disease after a period of inactivity, as determined by a uveitis specialist, most of whom follow the Standardization Uveitis Nomenclature Working Group Criteria.[15] The outcome was recorded as a binary variable (yes or no) at the end of follow-up, indicating whether a recurrence had occurred at any point during the observation period. The follow-up duration varied by patient, with a median of 12 months (interquartile range, 3–24 months).

Data Preparation

Data underwent preprocessing, including handling missing values through multiple imputation techniques, normalization of continuous variables, and encoding categorical features. Quality control checks ensured consistency across sociodemographic groups. Missing data were handled using multiple imputation and mean/mode substitution techniques, depending on the variable type and missingness mechanism. Supplementary Material S1 contains detailed information about Data Extraction and Cleaning Pipeline.

Predictors Selection

All 450 baseline variables were considered, and a bivariate analysis was conducted against the outcome variable. The Kruskal–Wallis test was used for continuous variables, while Fisher's exact test was applied to categorical variables. A false discovery rate adjustment was performed on the P values to account for multiple comparisons. Only variables with an adjusted P value of less than 0.05 were selected for model training. Supplementary Material S1 contains detailed information about Bivariate Analysis for Feature Selection.

Investigative Ophthalmology & Visual Science

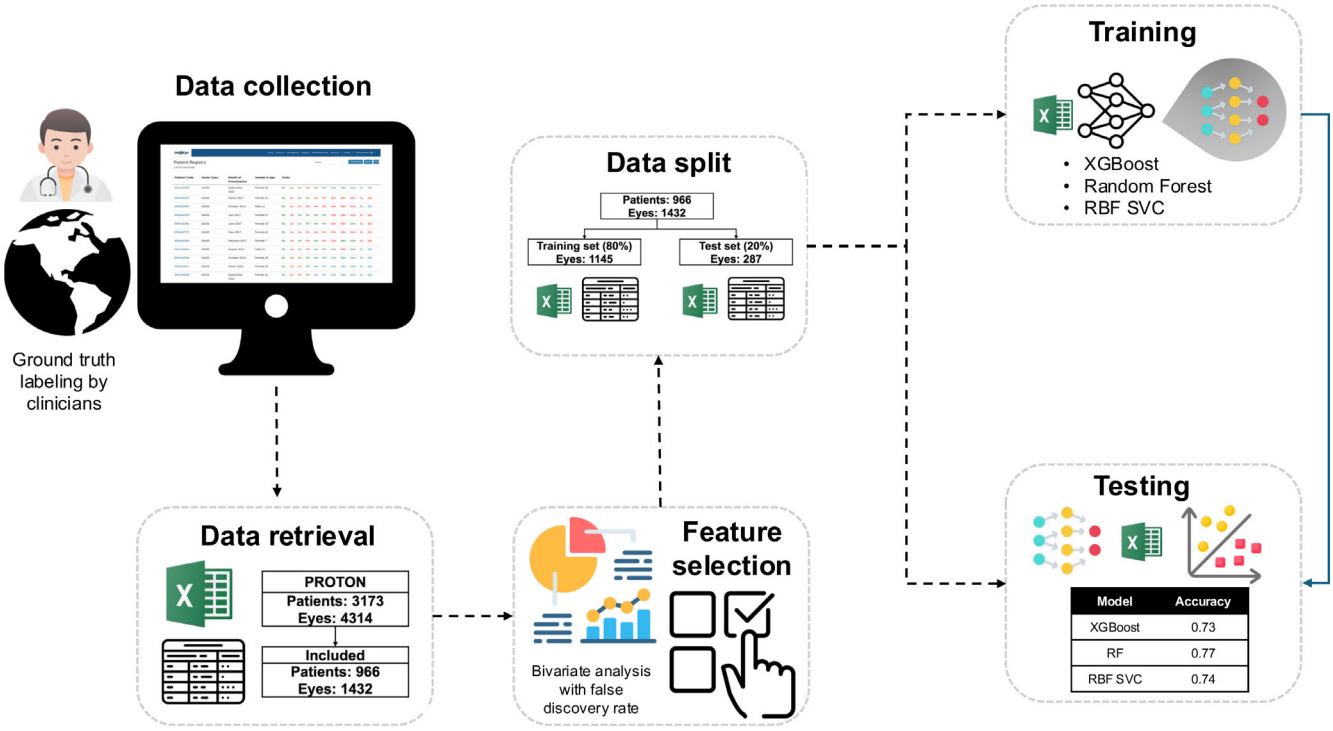

Figure 1. Workflow of data collection, feature selection, and ML model evaluation.

Handling of Bilateral Cases and Unit of Analysis

In this study, the unit of analysis was the eye, because recurrence and inflammation patterns can vary significantly between eyes, even within the same patient. This is particularly true for diseases such as HLA-B27–associated uveitis, which often exhibit asymmetric bilateral involvement. To evaluate the potential dependency between eyes, we conducted a preliminary correlation analysis of key inflammatory parameters across both eyes at multiple timepoints. The correlation coefficients ranged from 0.5 to 0.9, indicating variable but non-negligible intereye dependence. Given these findings, and to maintain clinical relevance—because recurrence prediction often guides eye-specific management—we opted for a per-eye analytic approach.

ML Models Training

We obtained a dataset (containing predictors and outcome variables) in Microsoft Excel format and imported it into Python using the pandas library. A set of independent variables (predictors) was defined, including clinical and demographic factors (e.g., diagnostic etiology, type of etiology [infectious, noninfectious, idiopathic, etc.], anatomical diagnosis, age, etc.). Then predictors (x) were isolated from the outcome column (y). Because the outcome was categorical, we applied a LabelEncoder to transform it into numeric labels for ML.

Next, we identified which predictor columns were numeric (e.g., integer or float based) vs. categorical (object or string based). This step guided our use of different preprocessing pipelines. For numeric features, we used median imputation (via SimpleImputer(strategy='median')) and standardization (via StandardScaler). For categorical features, we used most-frequent imputation (via SimpleImputer(strategy='most_frequent')) and one-hot encoding (via OneHotEncoder, with handle_unknown='ignore' to gracefully handle unseen categories).

We split the dataset into a training and a hold-out test set using an 80/20 ratio, stratifying by outcome variable to preserve class distribution. Both eyes from individual patients were included and treated as independent samples during model development. Data splitting was performed at the eye level, without constraining both eyes from a single patient to be in the same dataset (training or test). As such, it is possible that both eyes from a given patient were assigned to different sets. The design decision was made in advance based on two key factors: the clinical relevance of analyzing each eye independently, because eyes can differ in diagnosis and treatment. Additionally, splitting data at the eye level maximized the use of a limited and diverse dataset for training and evaluation.

Three ML models were configured, each placed inside a pipeline that included the data preprocessing steps (ColumnTransformer for numeric and categorical pipelines) followed by the classifier. The models examined were:

1. eXtreme Gradient Boosting (XGBoost) Classifier (XGBClassifier),
2. Random Forest (RF) Classifier (RandomForestClassifier), and
3. Support vector classifier (SVC) with an radial basis function (RBF) kernel.

To tune model hyperparameters, we performed a grid search using five-fold stratified cross-validation (StratifiedKFold) with accuracy as the scoring metric. For each model, we defined a search grid over parameters such as the

number of estimators and maximum depth (for XGBoost and RF), or cost and gamma parameters (for SVC). The pipeline (preprocessing + classifier) was then trained and evaluated across folds; the best parameter combination was identified based on mean cross-validation accuracy.

After selecting the best parameters for each model, we refitted them on the entire training set and evaluated them on the hold-out test set. Performance metrics included accuracy, confusion matrix, precision, recall, F1-score (via classification_report), and area under the receiver operating characteristic (ROC) curve (roc_auc_score). We also generated ROC curves (roc_curve) and precision–recall curves (precision_recall_curve), visualizing each model's ability to distinguish between outcome classes across various probability thresholds. We computed the learning curves (learning_curve) to assess how model performance evolves with increasing training sample sizes and whether adding more data might improve outcomes.

Interpretability

For model interpretability, we used SHapley Additive exPlanations (SHAP) analysis on our XGBoost classifier. Test data were preprocessed using our established pipeline—imputing and standardizing numeric variables and one-hot encoding categorical variables—to yield a transformed feature set with corresponding feature labels. SHAP values were computed using SHAP's TreeExplainer, and a beeswarm summary plot was generated using a red–blue color scale to illustrate the direction and magnitude of feature contributions for the top 10 predictors. Figure dimensions and margins were optimized to enhance clarity.

Ethical Considerations

OASIS 1, OASIS 2, and PROTON initially received approval from the IRB of the National Healthcare Group (Ethics ID: OASIS 1 - 2020/00301, OASIS 2 - 2021/00655, PROTON - 2020/00944). For OASIS 1, the requirement for informed consent was waived owing to the study's retrospective nature. In contrast, participants in OASIS 2 and PROTON provided signed informed consents. Additionally, each collaborative center secured approval from its respective local institutional review board or ethics committee, ensuring that all studies were conducted following the principles of the Declaration of Helsinki.

Results

A total of 1432 eyes, from 966 patients, were included in this study. The training set contained 1145 and the hold-out test had 287 observations. Most cases were idiopathic (479 [33.4%]) and patients were 17 to 40 years old (668 [46.6%]). The median follow-up duration was 12 months (interquartile range, 3–24 months). The etiologies and clinical characteristics list are reported in Tables 1 and 2, respectively.

Model Selection and Tuning

All models underwent hyperparameter tuning via five-fold stratified cross-validation. XGBoost achieved a best mean cross-validation accuracy of 0.72 (95% confidence interval [CI], 0.68–0.78) with max_depth=6 and n_estimators=200. RF produced a best mean cross-validation accuracy of 0.74 (95% CI, 0.72–0.82) using max_depth=None and n_estimators=200. RBF SVM attained a best mean cross-validation accuracy of 0.73 (95% CI, 0.684–0.786) with C = 1 and gamma = 1. Based on these hyperparameters, each pipeline was retrained on the entire training partition and subsequently tested on the hold-out set (20% of the data).

Hold-Out Test Performance

RF demonstrated the highest test accuracy (0.77), with a confusion matrix of 180 true negatives, 14 false positives, 52 false negatives, and 41 true positives. This yielded an overall sensitivity for the positive class of 0.44 and a precision of 0.75. XGBoost achieved an accuracy of 0.73, classifying 173 true negatives, 21 false positives, 56 false negatives, and 37 true positives. RBF SVM yielded an accuracy of 0.74, with 186 true negatives, 8 false positives, 68 false negatives, and 25 true positives. All models had higher recall for the majority (negative) class compared with the minority (positive) class. RF showed the most balanced performance between precision and recall among the three, albeit with modest sensitivity for the positive class (Table 3).

ROC and Precision–Recall Analyses

ROC curves plotted on the hold-out data (Fig. 2, ROC curves) revealed that RF had the highest area under the curve, followed closely by XGBoost and then the RBF SVM. Precision recall curves (Fig. 2, precision recall curves) demonstrated similar patterns, with RF exhibiting the highest average precision. However, in all models, precision and recall were lower for the positive (recurrence) class than for the negative class (Fig. 2).

Learning Curves

Learning curves indicated that all three algorithms stabilized in performance as sample size increased, suggesting that although incremental gains might be achievable with larger datasets, most benefit was already realized with the available training size. None of the models showed significant overfitting in the training range evaluated.

Combined, RF provided the best overall accuracy in identifying recurrences, although all models demonstrated some degree of class imbalance in sensitivity for positive cases (Fig. 3).

SHAP Analysis of Feature Importance

The SHAP summary plot (Fig. 4) illustrates the impact of individual features on the model's predictions of recurrence in inflammatory ocular diseases. Features such as vitreous haze, retrolental cells, and noninfectious etiology exhibited notable contributions to the model's output. Age group (17–40 and 41–60 years), intermediate uveitis, and the presence of synechiae also played a role in prediction. The SHAP values indicate both positive and negative influences on the outcome, with the spread of points reflecting variability in feature importance across the dataset.

Subgroup Analysis: Idiopathic Uveitis

To further explore feature importance in a clinically relevant subgroup, we conducted a separate analysis restricted to patients with idiopathic uveitis ($n = 479$ eyes [33.4% of the

Table 1. Diagnoses Included in the Cohort

	Training		Test		
	Recurrent Course		Recurrent Course		
Diagnosis	**No**	**Yes**	**No**	**Yes**	**Total**
ANCA associated vasculitis	–	–	1	0	1
AZOOR	–	–	0	2	2
Alkaptonuria	0	1			1
Alport's syndrome	1	0	1	0	2
Ankylosing spondylitis	9	8	2	3	22
Antiphospholipid syndrome	2	0	–	–	2
Autoimmune retinopathy	4	3	–	–	7
Behcet's disease	8	7	3	1	19
Birdshot retinochoroidopathy	2	7	2	1	12
Blau syndrome	2	0	–	–	2
Brucellosis	2	0	–	–	2
Cytomegalovirus	5	1	–	–	6
Chron disease	0	1	0	1	2
Dermatomyositis	4	0	–	–	4
Drug induced	3	3	–	–	6
Endogenous endophthalmitis	2	1	1	1	5
Epstein–Barr virus	1	0	–	–	1
HIV	3	0	–	–	3
HLA-B27 associated	28	23	4	5	60
Herpes Virus	16	5	3	1	25
Idiopathic	270	117	68	24	479
JIA	7	23	3	8	41
Multiple sclerosis	3	2	–	–	5
Posner Schlossman syndrome	2	1	0	1	4
Post surgical	0	2	–	–	2
Psoriasis	3	6	–	–	9
Psoriatic arthritis	0	2	1	0	3
Reactive arthritis	1	3	–	–	4
Relapsing polychondritis	2	0	1	0	3
Related to MPO vasculitis	1	0	–	–	1
Rheumatoid arthritis	4	0	0	2	6
Rickettsia	1	0	–	–	1
SLE	9	2	1	0	12
Sarcoidosis	21	14	9	3	47
Secondary to COVID-19 vaccine	2	0	–	–	2
Serpiginous choroidopathy	1	0	3	0	4
Sjogren syndrome	1	0	–	–	1
Spondyloarthropathy	0	1	0	1	2
Sympathetic ophthalmia	1	0	–	–	1
Syphilis	9	0	1	0	10
TINU	2	0	1	0	3
Toxoplasmosis	54	14	15	4	87
Traumatic	2	0	–	–	2
Tuberculosis	122	43	23	13	201
Undetermined	141	57	45	19	262
Undifferentiated immune phenotype	3	0	–	–	3
Viral unspecific	2	4	1	0	7
Vogt–Koyanagi–Harada disease	20	18	5	3	46
Subtotal	776	369	194	93	1432
Total	1145		287		1432

cohort]). There was a decrease in the accuracy of all models (XGBoost = 0.67; RF = 0.72; RBF SMV = 0.66) and RBF SVM overfitted. Moreover, the top predictors of recurrence in this subgroup differed slightly from the full cohort. Notably, the anatomical diagnosis, the presence of vitreous haze, and age group remained highly influential features, whereas the importance of retinal vasculitis was reduced. The top 10 features for the idiopathic subgroup are presented in Supplementary Material S2.

Discussion

Although artificial intelligence (AI) has emerged as a transformative tool in ophthalmology, particularly in diagnosing and managing complex diseases, its application in uveitis has been relatively slow.[16,17] This is mainly due to the scarcity of high-quality datasets and the relatively low prevalence of the disease (<730 cases per 100,000 people),[18] which limits the availability of large-scale, well-annotated

Table 2. Distribution of Predictors in Training and Test Cohorts

	Training		Test			
	Recurrent Course		Recurrent Course			
	No	Yes	No	Yes	Total	Adjusted *P* Value
Age group						
0–16	72	58	19	12	161	0.01
17–40	369	157	95	47	668	
41–60	231	111	49	22	413	
>60	104	43	31	12	190	
Category-specific diagnosis						
Idiopathic	270	117	68	24	479	0.004
Infectious	217	68	44	20	349	
Noninfectious	136	121	37	30	324	
Others	10	5			15	
Undetermined	143	58	45	19	265	
Anatomic diagnosis						
Anterior + intermediate	15	9	2	1	27	0.003
Anterior scleritis	0	1			1	
Anterior uveitis–granulomatous	49	25	12	9	95	
Anterior uveitis–nongranulomatous	182	134	48	36	400	
Episcleritis	1	0			1	
Intermediate uveitis	96	27	27	3	153	
Keratouveitis	2	0	2	1	5	
Others	29	5	3	2	39	
Panuveitis	153	83	34	19	289	
Posterior scleritis	0	2	0	3	5	
Posterior uveitis	84	34	27	5	150	
Retinal vasculitis	143	42	36	11	232	
Perivascular infiltrates						
No	686	355	172	87	1300	0.003
Yes	69	10	14	5	98	
Anterior chamber flare						
1+ (faint)	66	57	16	11	150	0.004
2+ (moderate, iris and lens details clear)	35	32	8	7	82	
3+ (marked, iris and lens details hazy)	9	4	3	0	16	
4+ (intense, fibrin or plastic aqueous)	1	0			1	
None	618	253	159	72	1102	
Vitreous haze						
0 (nil)	487	197	133	54	871	0.004
1 (post pole clearly visible)	97	67	31	17	212	
2 (post pole details slightly hazy)	107	46	22	9	184	
3 (post pole details very hazy)	34	20	4	2	60	
4 (post pole details barely visible)	13	8	1	2	24	
5 (fundus details not visible)	15	18	1	6	40	
Synechiae						
No	624	269	160	69	1122	0.007
Yes	147	99	33	24	303	
Retrolental cells						
No	633	329	163	78	1203	0.016
Yes	128	33	27	13	201	
Chorioretinal scar						
No	686	348	162	86	1282	0.019
Yes	72	17	24	6	119	
Vitreous bleed						
No	720	358	179	91	1348	0.028
Yes	38	8	13	2	61	
Retinal vasculitis						
No	594	310	143	78	1125	0.029
Yes	151	51	41	15	258	
Cornea edema						
No	746	364	188	93	1391	0.032
Yes	26	4	6	0	36	
Keratic precipitates						
No	598	263	148	67	1076	0.05
Yes	171	105	45	26	347	

Investigative Ophthalmology & Visual Science

Table 3. Performance of Three ML Models on the Hold-Out Test Set (n = 287)

Model	Best CV Accuracy*	Test Accuracy†	TN‡	FP‡	FN‡	TP‡	Sensitivity (TPR)	Specificity (TNR)	Precision (PPV)	Recall	F1-Score
XGB	0.72	0.73	173	21	56	37	0.40	0.89	0.64	0.39	0.49
RF	0.74	0.77	180	14	52	41	0.44	0.93	0.75	0.44	0.55
RBF SVC	0.73	0.74	186	8	68	25	0.27	0.96	0.76	0.26	0.39

CV, cross-validation; FN, false negatives; FP, false positives; PPV, positive predictive value (TP / [TP + FP]); TN, true negatives; TNR, true negative rate (TN / [TN + FP]); TP, true positives; TPR, true positive rate (TP / [TP + FN]).

* Best CV accuracy was obtained via five-fold stratified cross-validation with an accuracy scoring metric.

† Test accuracy was computed on the 20% hold-out set.

‡ Confusion matrix cells (TN, FP, FN, TP) are from the hold-out test. From these values, we computed sensitivity, specificity, and precision for the positive class (recurrence).

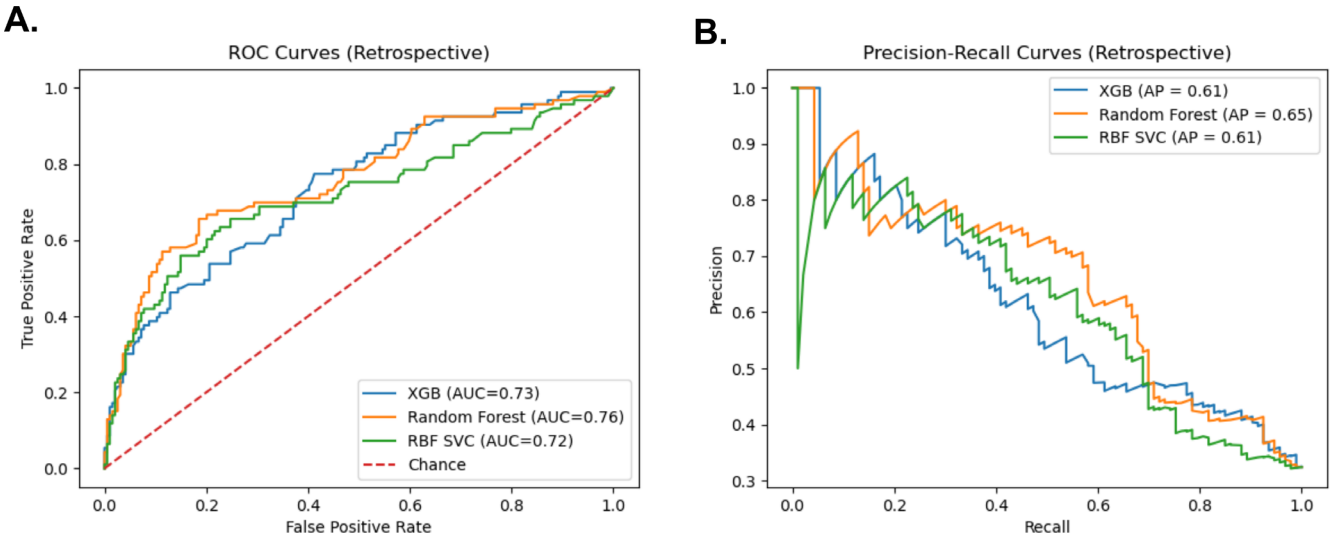

Figure 2. ROC and precision–recall curves for the three ML models in predicting recurrent course. Each panel compares the performance of XGBoost (XGB), RF, and an RBF-kernel SVC on the hold-out test set for identifying patients at risk of recurrent uveitis. (**A**) The ROC curves plot the true-positive rate (sensitivity) against the false-positive rate; the diagonal red dashed line represents chance performance. Numbers in parentheses reflect the area under the ROC curve (AUC) for each model. (**B**) The precision– recall curves demonstrate the models' trade-offs between precision (positive predictive value) and recall (sensitivity). The values in parentheses indicate the average precision (AP). Higher curves and larger AUC/AP values indicate better discrimination. The RF model exhibited the highest AUC (0.76) and AP (0.65), followed by XGB and the RBF SVC.

data for model training. Most AI applications in ophthalmology have focused on imaging-based diagnostics, such as automated detection of diabetic retinopathy and AMD.[19] However, in the field of uveitis, the use of AI has been more limited and mainly focused on ocular imaging,[20] with only a few studies applying ML to structured tabular data. Hammam et al.[21] developed ML models to predict vision-threatening Behçet's disease, demonstrating that XGBoost outperformed traditional logistic regression in identifying patients at high risk of vision-threatening Behçet's disease. Their study highlighted key predictive factors, including disease activity, thrombocytosis, smoking history, and daily steroid use, providing valuable insights into how AI can enhance risk stratification in uveitis management.

This study represents a novel contribution to the application of ML in uveitis by leveraging a large, multicenter, and ethnically diverse registry (OASIS)[14] to predict disease recurrence—an outcome highly relevant to long-term clinical management. Unlike previous ML efforts that primarily focused on diagnostic classification,[22] our work addresses a prognostic question: predicting recurrence at the eye level, which directly informs treatment intensity and monitoring strategies. The heterogeneity of the dataset enhances the generalizability of our findings across clinical settings. The SHAP analysis results are consistent with previous findings in specific uveitis subtypes, such as tubercular uveitis, where vitreous haze has been identified as an independent risk factor for recurrence in multivariate analyses.[23] Moreover, the use of SHAP-based interpretability allows transparent, feature-level insights into model decision-making,[24] improving clinical trust and potential adoption.

In this study, we used several ML algorithms—RF, XGBoost, and an RBF-kernel support vector machine—to predict the recurrent course of uveitis using 14 baseline clinical and demographic variables. The selected predictors captured a broad range of anterior, intermediate, posterior, and panuveitic findings (e.g., vitreous haze, perivascular infiltrates, and retinal vasculitis)[25] alongside key patient-level diagnostic categorizations and age groups. Such diversity in predictors aligns with the multifactorial nature of uveitis pathophysiology and its varied anatomical presentations.

Investigative Ophthalmology & Visual Science

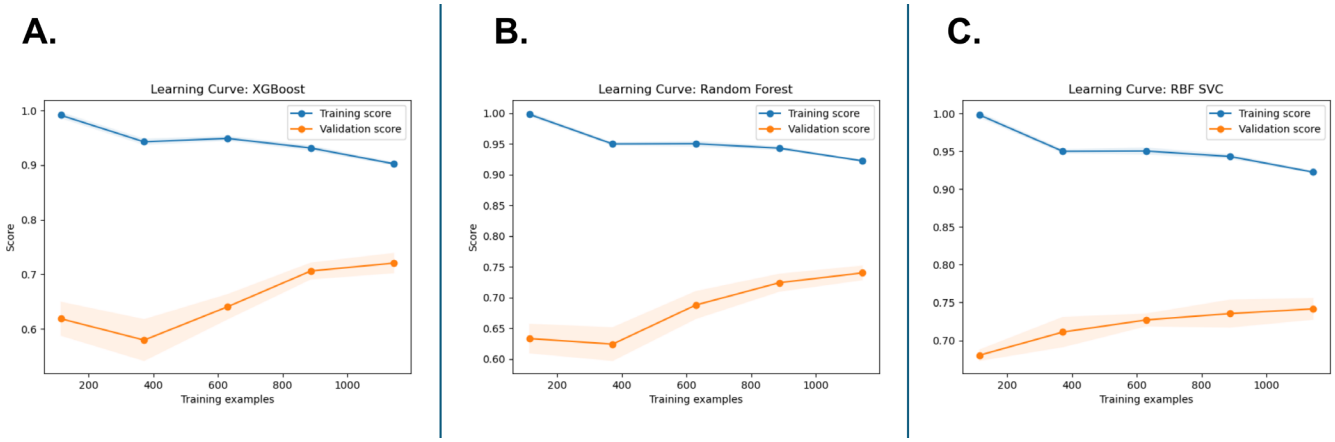

Figure 3. Learning curves for XGBoost, RF, and RBF SVC in predicting recurrent course. Each panel depicts training (*blue*) and validation (*orange*) performance as the number of training examples increases, based on five-fold cross-validation. The shaded regions represent ±1 SD. (**A**) XGBoost's validation accuracy gradually increases and narrows the gap with the training curve. (**B**) RF follows a similar pattern, showing improved validation scores as the sample size grows. (**C**) RBF SVC exhibits the same overall trend of rising validation accuracy, but with a somewhat larger gap to the training curve at smaller sample sizes. These learning curves collectively suggest that larger training sets improve model generalization, although each model's convergence patterns differ in how quickly they reach performance stability.

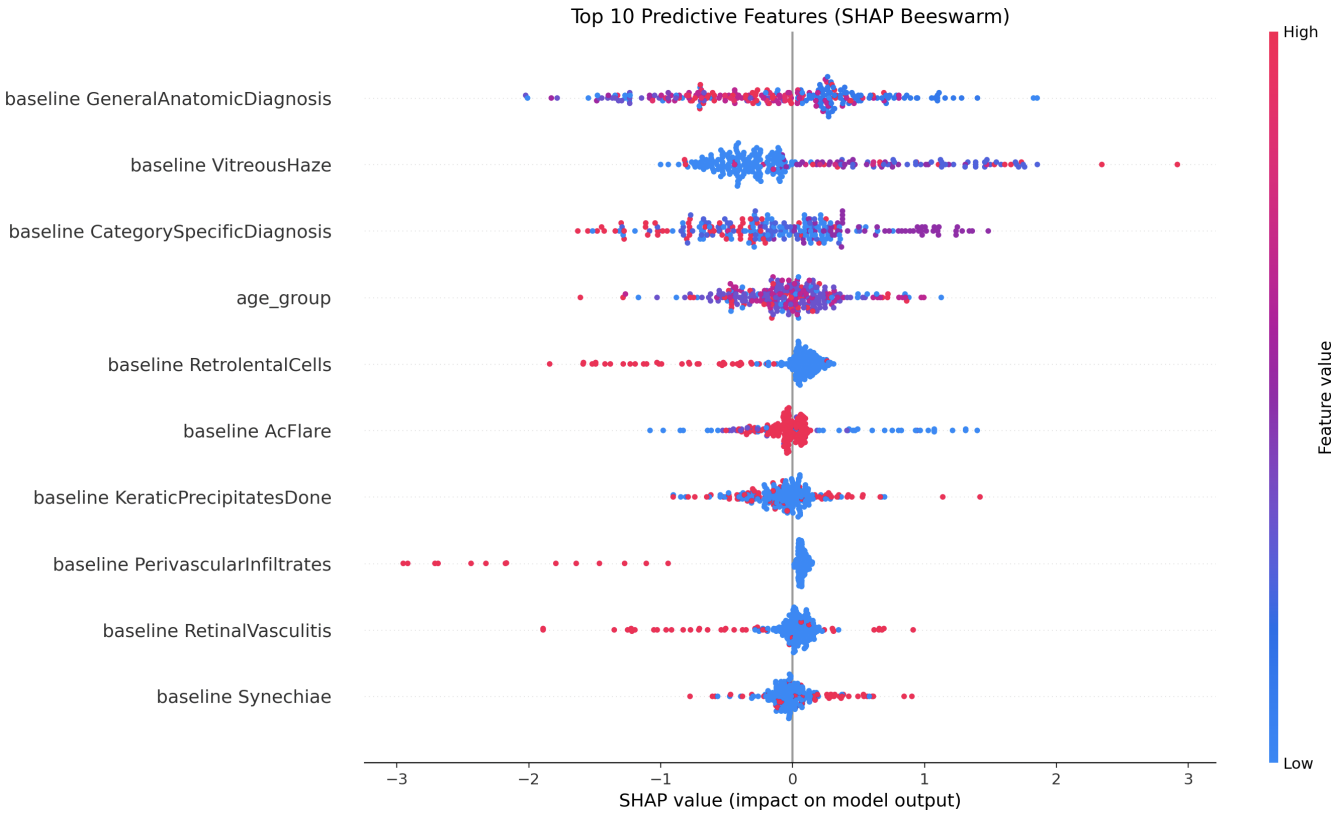

Figure 4. Top 10 relevant features derived from SHAP analysis. This figure presents how specific clinical features identified through SHAP contribute to the model's classification.

Through rigorous cross-validation and hyperparameter tuning,[26] RF models achieved the highest accuracy on the retrospective hold-out set (77%), followed closely by XGBoost (73%) and the RBF SVM (74%). Notably, each model demonstrated relatively high specificity, indicating a strong ability to correctly identify patients unlikely to experience a relapse. The negative predictive value (NPV) reflects the probability that a patient predicted as negative by the model truly does not experience recurrence. Among the evaluated models, the RBF SVC demonstrated the highest NPV at 73%, indicating that, when the model classified a patient as negative, there was a 73% chance that the patient would not have the outcome. The RF model followed closely with an NPV of 78%; XGBoost had the lowest NPV at 76%. These find-

ings suggest that all models performed relatively well in ruling out recurrence in patients predicted as negative, with RF providing the most reliable negative predictions. From a clinical standpoint, a high specificity can limit unnecessary escalation of therapy or follow-up intensity among individuals at low risk of recurrence, thereby decreasing potential adverse effects and conserving resources.[27,28]

The model could function most effectively as a reassurance and triage tool, helping to identify individuals with a low likelihood of disease recurrence. In the context of uveitides—a group of conditions with uncertain progression—this capability is clinically meaningful, as it provides mid- to high-confidence reassurance to low-risk patients. The model supports more efficient use of health care resources by decreasing unnecessary anxiety and minimizing the burden of excessive follow-up investigations. Its strength in accurately ruling out recurrence makes it well-suited for guiding clinical decisions around the intensity of monitoring and enabling a more personalized approach to patient care. Further refinement, including exploring cost-sensitive learning or more extensive feature engineering, may be warranted to enhance the detection of patients at risk for recurrences.

However, the sensitivity for detecting actual cases that will recur was more modest. This finding reflects the inherent challenges of predicting a relatively infrequent event, often compounded by clinical heterogeneity across etiologies and a smaller proportion of relapse cases within the dataset. Although high specificity helps clinicians to avoid overtreating patients at lower risk, the lower sensitivity highlights an opportunity to enhance detection of high-risk cases. Techniques such as cost-sensitive learning, oversampling methods (e.g., SMOTE), or incorporating additional biomarkers and imaging data may further refine minority class recognition.[26,29] We recognize that integrating ML models with clinician expertise could further enhance predictive accuracy and usability. Future work may explore hybrid frameworks combining algorithmic predictions with structured clinical input to optimize real-world implementation.

Certain limitations must be acknowledged. First, although ML models demonstrate promising predictive performance, external validation on independent datasets is required to confirm their robustness. Second, imbalances in the dataset, particularly regarding recurrence cases, may have influenced model performance. Third, the definitions of indeterminate and idiopathic as groups of etiologies may vary between centers owing to the lack of standardization. Finally, the use of both eyes from the same patient as independent samples may introduce bias owing to intereye correlation—potentially inflating model performance by allowing patient-specific features to appear in both training and test sets. Although this approach maximized data use and captured eye-specific variation, it may decrease the generalizability of the findings.

Despite these limitations, the present approach demonstrates important strengths. First, we used a consistent data-preprocessing pipeline (including imputation, scaling, and one-hot encoding) across all models, ensuring reproducibility and decreasing bias from manual feature transformations. Second, thorough cross-validation minimized overfitting risk and provided robust estimates of model generalizability. Finally, the predictors themselves reflect clinically accessible information, suggesting that these models could be readily integrated into routine practice if validated prospectively. Moreover, this study uses a data-driven, etiology-agnostic approach that enhances generalizability across diverse patient populations.

Future research should focus on exploring alternative approaches to improve recall rates, further and integrating additional data sources, such as imaging biomarkers and genetic predisposition factors, to enhance model accuracy. Additionally, prospective validation in clinical settings will be necessary to assess real-world applicability. Developing interpretable ML frameworks may also help to facilitate clinician adoption, ensuring that predictive insights are effectively integrated into decision-making workflows.

Conclusions

Our findings suggest that ML models, particularly RF, can achieve high specificity in identifying uveitis patients who are less likely to experience recurrences. This work may help clinicians to tailor their follow-up schedules and therapy intensities more efficiently. Nonetheless, improvements in sensitivity are needed to detect those at greatest risk of flare reliably. Future efforts should focus on advanced techniques to address class imbalance and incorporate richer clinical data, with prospective validation in diverse uveitis populations. Such work will help to solidify the role of ML in personalizing the management of this complex, sight-threatening condition.

Acknowledgments

RA is supported by grants awarded by the National Medical Research Council (NMRC), Ministry of Health, Republic of Singapore grant number NRMC/CSAINV22jul-0004, NMRC/CSAINV19nov-0007, and NMRC/CIRG21nov-0023. The funders had no role in study design, data collection and analysis, publication decisions, or manuscript preparation.

Disclosure: **W. Rojas-Carabali**, None; **C. Cifuentes-González**, None; **A. Utami**, None; **M. Agarwal**, None; **J.H. Kempen**, None; **E. Tsui**, None; **R.P. Finger**, None; **A. Sen**, None; **A. Chan**, None; **A. Schlaen**, None; **V. Gupta**, None; **A. de-la-Torre**, None; **B. Lee**, None; **R. Agrawal**, None

References

1. Apivatthakakul A, Tantraworasin A, Rothova A, et al. Prevalence and risk factors for ocular complications in new-onset uveitis: a study from a tertiary referral center in Northern Thailand. *Clin Ophthalmol*. 2025;19:543–551.
2. Prieto Del Cura MDM, Gonzalez-Guijarro JJ. Impact of ocular complications on visual outcomes in adult patients with uveitis. *Cureus*. 2022;14:e21370.
3. Suttorp-Schulten MS, Rothova A. The possible impact of uveitis in blindness: a literature survey. *Br J Ophthalmol*. 1996;80:844–848.
4. Jabs D, Nussenblatt R, Rosenbaum J. Standardization of uveitis nomenclature for reporting clinical data. Results of the First International Workshop. *Am J Ophthalmol*. 2005;140:509–516.
5. O'Connor GR. Factors related to the initiation and recurrence of uveitis. *Am J Ophthalmol*. 1983;96:577–599.
6. Pineda-Sierra JS, Cifuentes-González C, Rojas-Carabali W, et al. Clinical characterization of patients with HLA-B27-associated uveitis and evaluation of the impact of systemic treatment on the recurrence rate: a cross-sectional study. *J Ophthal Inflamm Infect*. 2023;13:38.

7. Cifuentes-González C, Rojas-Carabali W, Pérez ÁO, et al. Risk factors for recurrences and visual impairment in patients with ocular toxoplasmosis: a systematic review and meta-analysis. *PLoS One*. 2023;18:e0283845.
8. Standardization of Uveitis Nomenclature (SUN) Working Group. Classification criteria for cytomegalovirus anterior uveitis. *Am J Ophthalmol*. 2021;228:89–95.
9. Standardization of Uveitis Nomenclature (SUN) Working Group. Classification criteria for Behçet disease uveitis. *Am J Ophthalmol*. 2021;228:80–88.
10. Wang L, Zhao Z, Luo Y, et al. Classifying 2-year recurrence in patients with dlbcl using clinical variables with imbalanced data and machine learning methods. *Comput Methods Programs Biomed*. 2020;196:105567.
11. Venerito V, Emmi G, Cantarini L, et al. Validity of machine learning in predicting giant cell arteritis flare after glucocorticoids tapering. *Front Immunol*. 2022;13:860877.
12. Wan Q, Wang Q, Wei R, et al. Machine learning-based progress prediction in accelerated cross-linking for keratoconus. *Graefes Arch Clin Exp Ophthalmol*. 2025, doi:10.1007/s00417-025-06792-y.
13. Wang Y, Du R, Xie S, et al. Machine learning models for predicting long-term visual acuity in highly myopic eyes. *JAMA Ophthalmol*. 2023;141:1117.
14. Ng SMS, Low R, Pak C, et al. The role of a multicentre data repository in ocular inflammation: the Ocular Autoimmune Systemic Inflammatory Infectious Study (OASIS). *Eye*. 2023;37:3084–3096.
15. Standardization of Uveitis Nomenclature (SUN) Working Group. Standardization of uveitis nomenclature for reporting clinical data. Results of the First International Workshop. *Am J Ophthalmol*. 2005;140:509–516.
16. Rojas-Carabali W, Cifuentes-González C, Gutierrez-Sinisterra L, et al. Managing a patient with uveitis in the era of artificial intelligence: current approaches, emerging trends, and future perspectives. *Asia-Pac J Ophthalmol*. 2024;13:100082.
17. Murugan SRB, Sanjay S, Somanath A, et al. Artificial intelligence in uveitis: innovations in diagnosis and therapeutic strategies. *Clin Ophthalmol*. 2024;18:3753–3766.
18. García-Aparicio Á, García De Yébenes MJ, Otón T, Muñoz-Fernández S. Prevalence and incidence of uveitis: a systematic review and meta-analysis. *Ophthalmic Epidemiol*. 2021;28:461–468.
19. Yang J, Fong S, Wang H, et al. Artificial intelligence in ophthalmopathy and ultra-wide field image: a survey. *Expert Syst Appl*. 2021;182:115068.
20. Lu A, Li K, Guo S, et al. Development and validation of novel retina biomarkers and artificial intelligence models for Behçet disease uveitis prediction. *Biomed Signal Process Control*. 2024;94:106271.
21. Hammam N, Bakhiet A, El-Latif EA, et al. Development of machine learning models for detection of vision threatening Behçet's disease (BD) using Egyptian College of Rheumatology (ECR)–BD cohort. *BMC Med Inform Decis Mak*. 2023;23:37.
22. Jacquot R, Ren L, Wang T, et al. Neural networks for predicting etiological diagnosis of uveitis. *Eye*. 2025;39:992–1002.
23. Agrawal R, Gunasekeran DV, Grant R, et al. Clinical Features and Outcomes of Patients With Tubercular Uveitis Treated With Antitubercular Therapy in the Collaborative Ocular Tuberculosis Study (COTS)–1. *JAMA Ophthalmol*. 2017;135:1318.
24. Nohara Y, Matsumoto K, Soejima H, Nakashima N. Explanation of machine learning models using SHapley Additive exPlanation and application for real data in hospital. *Comput Methods Programs Biomed*. 2022;214:106584.
25. Zierhut M, Pavesio C, Ohno S, et al. Eds. *Intraocular Inflammation*. Berlin, Heidelberg: Springer Berlin Heidelberg; 2016. Available at: https://link.springer.com/10.1007/978-3-540-75387-2. Accessed March 19, 2025.
26. Mitchell TM. *Machine Learning*. New York: McGraw-Hill; 1997.
27. Schlattmann P. Statistics in diagnostic medicine. *Clin Chem Lab Med*. 2022;60:801–807.
28. Zhang L, Rojas-Carabali W, Choo SS, et al. Validation of the Online Collaborative Ocular Tuberculosis Study calculator for tubercular uveitis. *JAMA Ophthalmol*. 2024;142:1140.
29. Kelleher JD, MacNamee B, D'Arcy A. *Fundamentals of Machine Learning for Predictive Data Analytics: Algorithms, Worked Examples, and Case Studies*. Cambridge, MA; London, UK: The MIT Press; 2015.

Investigative Ophthalmology & Visual Science