Journal of Biostatistics and Epidemiology

J Biostat Epidemiol. 2023;9(2):189-200

Original Article

A Wegner's Granulomatosis Risk Prediction Model Based on Machine Learning Algorithms

Jaleh Shoshtarian Malak[1], Samira Alsaeidi[2*], Fatemeh Haji Ali Asgari[1], Fahimeh Khedmatkon[1]

[1]Department of Electronic Health, Virtual School, Tehran University of Medical Sciences, Tehran, Iran.
[2]Rheumatology Research Center, Tehran University of Medical Sciences, Tehran, Iran.

ARTICLE INFO

Received 27.12.2022
Revised 28.01.2023
Accepted 14.02.2023
Published 15.06.2023

Key words:
Wegener's granulomatosis relapse;
Relapse prediction;
Machine learning;
Clinical decision-making;
Xgboost algorithm;
Birmingham vasculitis activity score;
Predictive modeling;
Healthcare analytics;
Autoimmune diseases;
Precision medicine

ABSTRACT

Introduction: Prediction of Wegener's granulomatosis diagnosis and relapse is a complex process. In this study, we applied machine learning algorithms to predict Wegener's granulomatosis relapse.
Methods: In this research, 189 patients admitted to Amiralam Hospital were studied and followed for approximately 2 years. Patient features included demographics, organ involvement, symptoms, and other clinical data. Different popular machine learning algorithms were applied for predicting Wegener's granulomatosis relapse, including Support Vector Machines, Random Forest, Gradient Boosting, and XGBoost algorithms. The prediction model performance was measured for the different candidate prediction algorithms using accuracy, precision, recall, and F1-measure. The selected prediction model performance was calculated based on different relapse rates and major relapse occurrence according to Birmingham Vasculitis Activity Score (BVAS) fields.
Results: Applying different machine learning algorithms, the XGBoost algorithm performed the best. The results indicated that the prediction model's performance increased when calculating higher relapse rate possibilities. The XGBoost model had 82% accuracy while predicting more than one relapse rate and 92% accuracy in predicting more than twice the relapse rate. We also calculated the SHAP value for the prediction model. The results indicated that Cr, BVAS, lymphocyte percentage, vitamin D, nose involvement, alkaline phosphatase, diagnosis age, white blood cell count, erythrocyte sedimentation rate, and initial nose presentation are the 10 most important features according to SHAP value.
Conclusion: In this study, we have developed Wegener's granulomatosis relapse prediction model using machine learning algorithms. We achieved reasonable precision and recall for early prediction and decision-making regarding Wegener's granulomatosis relapse

Introduction

Granulomatosis with Polyangiitis (Wegener's) disease, also known as GPA, is an autoimmune disease of unknown etiology. It causes symptoms in the sinuses, lungs, and kidneys, as well as other organs. GPA is diagnosed based on a combination of clinical manifestations

*.Corresponding Author: s-alesaeidi@sina.tums.ac.ir, salesaeidi@gmail.com

including blood tests, X-rays, CT scans, and physical exams. Tissue biopsies of the affected organ are most often necessary to confirm the diagnosis. Antineutrophil cytoplasmic antibodies directed against proteinase 3 (PR3-ANCA) highly specify GPA.[1,2] GPA can be categorized as mild, moderate, or severe depending on the extent of organ involvement based on the Birmingham Vasculitis Activity Score (BVAS). According to the American College of Rheumatology's classification criteria, GPA as a separate disease from other forms of systemic vasculitis has limitations because it may not clearly distinguish GPA from microscopic polyangiitis (MPA) and vasculitis mimics.[3] Some researchers have indicated that ANCA is a valuable measure to predict the relapse of GPA,[4] while others stated that ANCAs do not predict disease activity in Wegener's granulomatosis.[5]

Nowadays, machine learning techniques have been widely used for disease prognosis, diagnosis, and management. Predicting a chronic disease and diagnosing it at early stages plays a pivotal role in medical scenarios.[6] Previous studies on GPA classification and risk prediction were limited by conventional statistical methods; hence, models for severity prediction and risk factor analysis of GPA had not yet been developed. A prediction model and risk stratification model for lupus nephritis renal flare developed with Gradient Boosting and Simplified Risk Score Prediction Model showed good predictive performance of 0.81 C-index.[7] A simple, clinician-friendly machine learning-based model introduced to assist with the diagnosis of systemic lupus erythematosus showed 94.8% accuracy.[8] While several different prediction algorithms have been proposed for different chronic disease management, less work has been done in GPA data and risk factor analysis/prediction. High-dimensional data with small sample sizes are a common problem in clinical research where case numbers are limited. Therefore, machine learning algorithms should be applied to avoid optimistic performance estimates. Due to integrated technology, ensemble feature selection algorithms have better stability/robustness than other methods in handling high-dimensional data with multiple optimal feature subsets. In this research, we used XGBoost ensemble algorithms.[9]

Considering the high relapse rate of Wegener's disease, several studies have sought to analyze associated risk factors.[10,11] Immunosuppressive drugs are prescribed based on periodic ANCA evaluation. Patient at cessation stage of disease who received no higher dose of immunosuppressive drugs reported more recurrence rate than those who received ongoing immunosuppression according to ANCA value.[12] Clinicians are concerned about the toxic effect of immunosuppressive drugs used for Wegener's patients. Although ANCA is identified one of the main biomarkers to predict GPA relapse,[13] Studies had shown that beside ANCA, heart involvement and age could highly increase the risk of the relapse.[14] Routinely measuring ANCA titer within 2- or 3-time intervals, had shown only 50 to 60 percent association to disease severity. Accordingly measuring early ANCA level before symptomatic change showed even less relationship.[15,16] ANCA evaluation along with determining potential organ involvement could not be a promising method for recurrence evaluation. Instead using predictive methods might provide better decision support to maintain the patient.

Methods

Figure 1 shows a Schematic overview of relapse prediction model construction.

Study cohort

In this research, we conducted a longitudinal study consisting of 189 patients followed for approximately 2 years. The patients were admitted to Amiralam Hospital between 2013 and 2018 and their data were registered in the hospital's data registry system. During each patient admission, the Birmingham Vasculitis Activity Score (BVAS) was calculated for all patients based on symptoms at the time of diagnosis and recurrence of the disease. Patient demographics, organ involvement, symptoms, and other clinical data were stored electronically. The ANCA titer was measured at different disease phases for each patient at different times. Considering the data set characteristics, 57 percent of the patients were male and 42 percent were female. We randomly selected a study cohort comprising 70 percent of patients with Wegener's disease identified and 30 percent of controls to construct, train, and compare the machine learning models. 115 patients experienced at least one relapse, while 63 cases had more than one relapse and 28 cases relapsed more than twice.

Feature Selection

Initially, we had 249 features including patient demographics, organ involvement, symptoms, and other clinical data. Redundant features were removed, including identification features related to the patient. Organ involvement sub-features were merged into single features indicating each organ's involvement status. For example, considering the sinuses, we had different involvements including sinusitis, mastoid, erosion, and sclerosis. All of these symptoms were merged into one sinus-related feature based on rheumatology specialists'

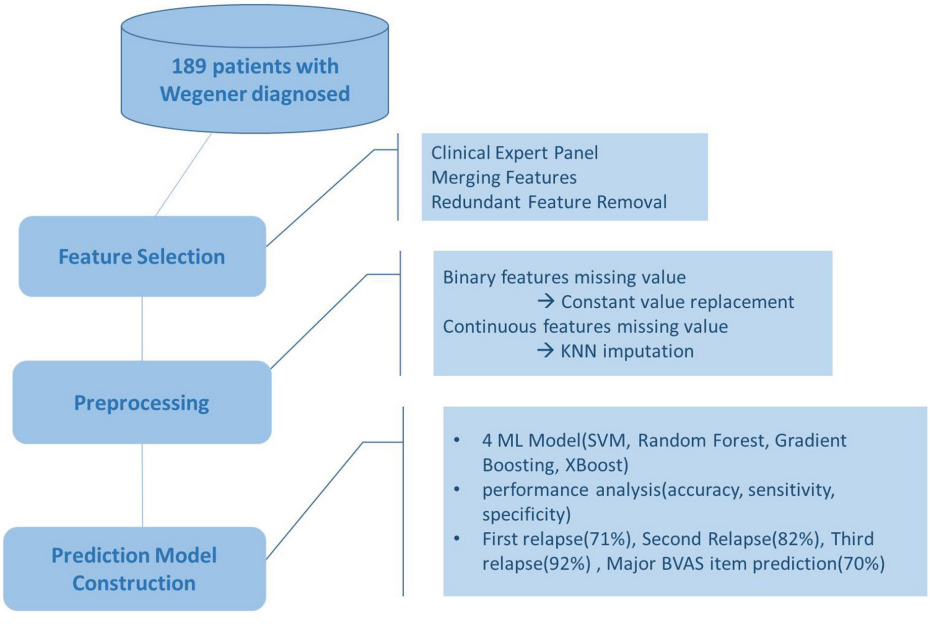

Figure 1. Schematic overview of relapse prediction model construction

opinions. According to the specialists' opinions, 57 features were removed, leaving a final selection of 87 features. Considering the machine learning techniques we applied for prediction, these models intrinsically perform feature selection automatically as part of the learning process and only include predictors that help maximize accuracy.

Clinical Outcomes

In this study, patients' severity of disease was calculated based on the Birmingham Vasculitis Activity Score (BVAS) according to symptoms present at the time of diagnosis. The number of patient relapse times at subsequent visits was also calculated. The prediction models' performance was tested to predict different relapse rates. We also analyzed the occurrence of major BVAS/WG items and applied prediction models to forecast major relapse events.

Data preprocessing

Redundant features were removed, including patient identification features. Ninety-five percent of the data had less than 5% missing values, while 5% had approximately 25-30% missing values. Some data items were not filled based on the fact that during patient assessment sessions, patients did not have any signs for that field. In these cases, the missing value was replaced by zero. Continuous and categorical features with missing values were imputed using the K-NN imputation algorithm.

Prediction model

Different popular prediction models were applied for WG relapse prediction, including SVM, Random Forest, Gradient Boosting, and XGBoost algorithms. The model with the highest accuracy was selected for prediction. We used 70% of the data for model training and 30% as test data to determine model performance. We evaluated the following metrics: Accuracy, Precision, Recall, and F1-score (Table 1). Ensemble machine learning methods including Gradient Boosting methods are widely used for classification and predictive modeling. They are constructed from decision tree models. Decision trees are added one at a time to the

Table 1. Accuracy, Precision, Recall and fi-score of prediction methods for different relapse rate

#Relapse	Model	Accuracy	Precision	Recall	Fi-score
Replapse>1	SVM	0.71	0.5	0.71	0.59
	RandomForest	0.74	0.72	0.74	0.72
	GradientBoosting	0.76	0.77	0.76	0.77
	XGBoost	0.82	0.81	0.82	0.81
Replapse>2	SVM	0.92	0.85	0.92	0.88
	RandomForest	0.92	0.85	0.92	0.88
	GradientBoosting	0.79	0.87	0.79	0.83
	XGBoost	0.92	0.85	0.92	0.88
Replapse>3	SVM	1	1	1	1
	RandomForest	1	1	1	1
	GradientBoosting	0.84	1	0.84	0.91
		1	1	1	1

ensemble and fit to correct prediction errors made by prior models. Built models are then fit using arbitrary differentiable loss functions and gradient descent optimization algorithms. Extreme Gradient Boosting (XGBoost) is an efficient implementation of gradient boosting algorithms that applies approximate computation algorithms. Feature importance of the input parameters was calculated based on the fitted XGBoost prediction model. All prediction model implementations were done using Python 3.6 in the Google Colab environment (packages: sklearn, XGBoost, graphviz, matplotlib).

Our classifier is a binary classifier predicting relapse occurrence, hence we did not classify patients based on their relapse rate. Instead, in each scenario, we changed the relapse rate threshold and considered patients with lower relapse rates as normal. For instance, the scenario with relapse > 2 had patients with a relapse rate greater than 2 as the outcome True (1) and others as the outcome False (0). So we trained and tested the model with different data sets where the outcome was different based on their relapse rate. Table 2 demonstrates the number of samples in each relapse frequency (Relapse >0, Relapse>1, Relapse>2, Relapse>3).

Prediction model tuning

In this research, we tested different values for XGBoost's common influential parameters to achieve higher model performance. These hyperparameters included tree-specific and learning task-specific values and weights that determine the learning process of the algorithm. Important tree-specific features tested were maximum tree depth, percentage rows and columns of rows used for each tree. Learning task-specific tree parameters controlled included the learning rate and early stopping rounds to improve the algorithm's accuracy.

Prediction model performance analysis

The prediction model performance was measured for the different candidate prediction algorithms using accuracy, precision, recall, and F1-score measures. The selected prediction model performance was calculated based on different relapse rates and occurring major relapse according to BVAS score fields.

Results

Considering the relapse rate of the case study instances 61% had at least one-time relapse, 33% percent had more than twice relapse, 15% had more than three-time relapse and 60% had major relapse according to BVAS feature set. Applying different machine learning algorithms Including SVM,Random Forest, Gradient Boosting and XGBoost algorithms, the results indicated the XGBoost algorithm had the highest performance (Figure2). Table1 shows accuracy, precision, recall and fi-score of prediction methods for different relapse rates.

The results indicated that the performance of

Table 2. Number of samples in each relapse frequency (Relapse >0, Relapse>1, Relapse>2, Relapse>3)

Class	Relapse 0	Relapse 1	Relapse 2	Relapse 3
0	71	123	158	176
1	115	63	28	10

XGBoost prediction model increased when calculating higher relapse rate possibilities. XGBoost model had 82% accuracy while predicting more than once relapse rate and 92% accuracy predicting more than twice relapse rate.

The feature importance was calculated based on the XGBoost prediction model; Figure 3 demonstrates the top 15 features according to importance. We also calculated SHAP (SHapley Additive exPlanations) values for a more detailed analysis of features (Figure 4). For each top important feature, we plotted split value histograms to gain better insight into their values and cutoff points (Figure 5). Considering the important features' prediction model split values, Hemoglobin (Hb) was 12.5 gm/dl, Birmingham Vasculitis Activity Score (BVAS) was 12 and 15, Vitamin D was 15 or 23 ng/ml, Polyglandular autoimmune syndrome (PGA) was 6.5, Creatinine (Cr) was 105 and 120 µmol/L, aspartate aminotransferase (AST) was 18.2 U/L, platelet count (PLT) was 280,000, Alkaline Phosphatase (ALP) was 174 IU/L, and lymphocyte percentage was 12.5. The results also indicated that Cr, BVAS, lymphocyte percentage, vitamin D, nose involvement, ALP, diagnosis age, white blood cell count, erythrocyte sedimentation rate, and initial nose presentation are the 10 most important features according to SHAP value (Figure 4).

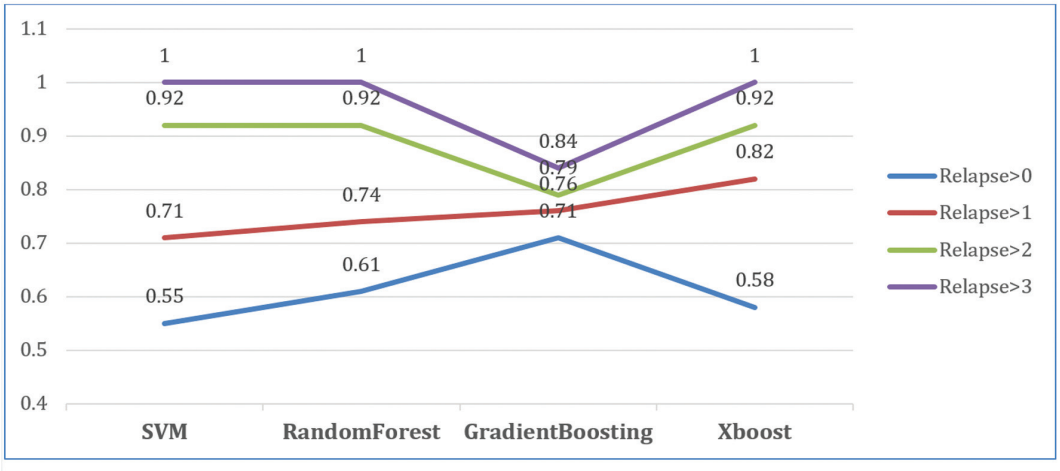

Figure 2. Prediction models accuracy for different relapse rates

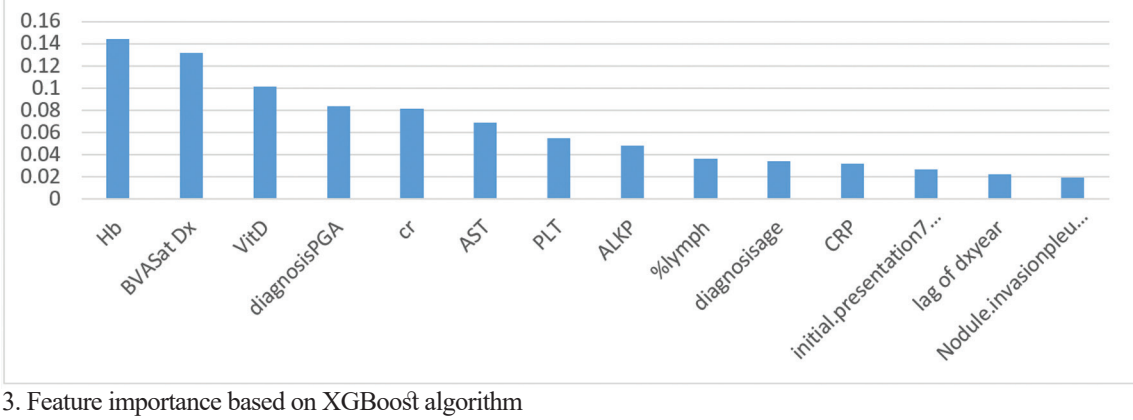

Figure 3. Feature importance based on XGBoost algorithm

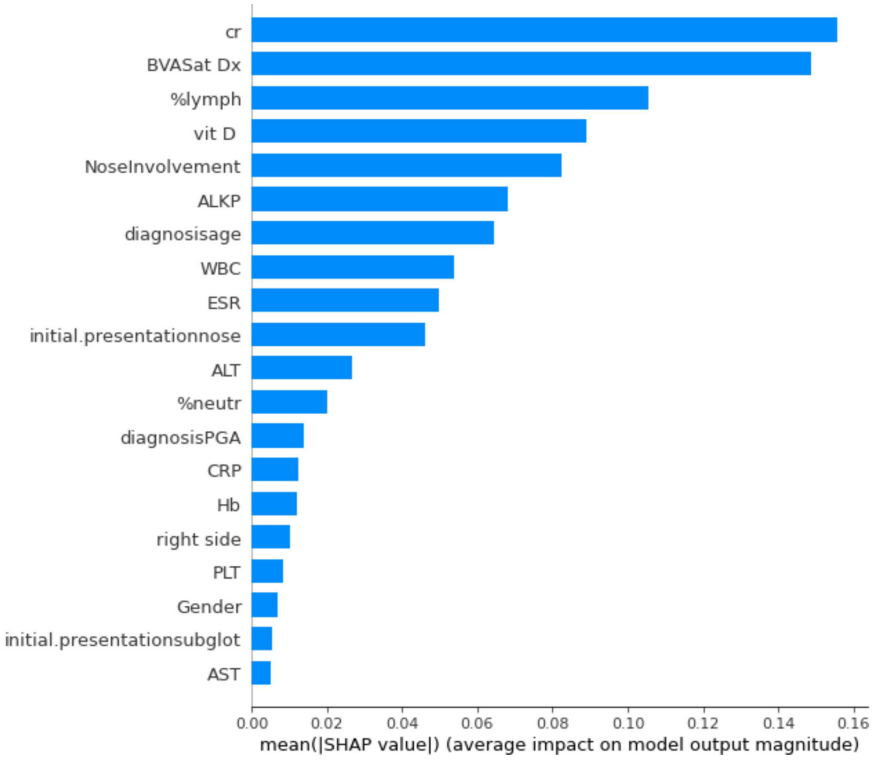

Figure 4. Calculating the SHAB value for XGBoost prediction model

Discussion

Identifying patients with a higher probability of relapse has been investigated in several studies.[16-19] Multiple correspondence analysis (MCA) was used for cluster analysis based on a k-means algorithm. One cluster model included nine clinical baseline variables as input variables, and a second cluster model additionally included ANCA specificities. Five partial clusters were distinguished for each relapse rate: cardiovascular AAV (highest relapse risk), non-renal AAV, renal AAV with PR3-ANCA, renal AAV without PR3-ANCA, and gastrointestinal AAV (lowest risk).[18] A stepwise algorithm was proposed for the classification of antineutrophil cytoplasmic antibody-associated vasculitis (AAV) and polyarteritis nodosa (PAN).[8] There has been much discussion about relapse prediction based on ANCA measurements, with studies showing ANCA is a more accurate measure for GPA patients with renal involvement but less efficient for non-renal disease.[20] There is controversy regarding whether the course of GPA is different between PR3-ANCA-associated vasculitis (AAV) versus MPO-AAV. Studies indicated that more organs are affected in PR3-AAV, whereas renal-limited vasculitis occurs more often in patients with MPO-AAV.[21] This is the first study to apply machine learning methods to predict relapse in GPA patients. We applied several different machine learning algorithms and found the XGBoost method showed the best performance for predicting relapse rates. Since two years of follow-up data were available for patients, the prediction models were tested on different relapse rates. The results indicated the selected prediction model had better performance for higher relapse

rates. We also analyzed important features and SHAP (SHapley Additive exPlanations) values for the major relapse prediction model. The important features for predicting first-time relapse were the same as for major relapse prediction. However, the SHAP value analysis differed - it indicated that Hb, BVAS, Vitamin D, Cr, PGA, PLT, ALP, ESR, AST and lymphocyte percentage had the highest SHAP values for major relapse prediction. In many medical applications, the number of samples is small while the number of descriptive dimensions is high. This "high dimension, low sample size" problem commonly arises in medicine where small patient populations are described by many characteristics. There are several ways to address this problem - feature selection is a popular method for reducing dimensions. In this study, we removed 57 features during feature selection. Additionally, some regularized machine learning algorithms like ensemble methods can automatically calculate feature importance from trained predictive models.[9,22,23]

A score for estimating the probability of relapse in granulomatosis with polyangiitis (GPA) and microscopic polyangiitis (MPA) was developed called the Five Factor Score (FFS), and univariate and multivariate analyses were conducted to identify variables associated with relapse.[24] The area under the receiver operating characteristic (ROC) curve of their score showed approximately 0.71 performance in patients monitored within 36 months. However, the study was influenced by major usage of rituximab before the study, which significantly changed the treatment of GPA and MPA patients. In another study, the Least Absolute Shrinkage and Selection Operator method and Cox proportional hazards regression were used to develop a predictive model for long-term survival of ANCA-associated vasculitis.[25] The optimism-corrected concordance index and integrated Brier score of their prediction model were 0.728 and 0.109, respectively. Their predictive model demonstrated higher net benefits compared to the revised FFS (rFFS) and the Birmingham Vasculitis Activity Score (BVAS) system. O. Burkhardt et al.[26] searched for risk factors predicting mortality in patients with Wegener's granulomatosis (WG) treated in the intensive care unit (ICU). They studied 17 patients admitted to the ICU and monitored their clinical and laboratory variables over 4 years. Negative prognostic factors were identified using contingency table analyses, univariate logistic regression, and discriminant analysis. The discriminant analysis and application of the modified Ridge rule with selection produced a satisfactory result with an accuracy of 82.4% and a 90.0% probability for a correct prognosis of survival. In another study,[27] logistic regression models, Kaplan–Meier method, and Cox proportional hazards regression models were applied to analyze relapse rate, MPO-ANCA status, and remission-maintenance therapies. A XGBoost-based model was built to predict the risk of myelodysplastic syndrome using vital signs, lab results, and demographics from the prior two years of patient data.[28] The model achieved an area under the receiver operating characteristic curve (ROC AUC) value of 0.87 for prediction of MDS one year prior to diagnosis, with a sensitivity of 0.79 and specificity of 0.80. The XGBoost model was compared against logistic regression and artificial neural network models, which achieved ROC AUC values of 0.838 and 0.832, respectively.

In this study, we developed a GPA relapse

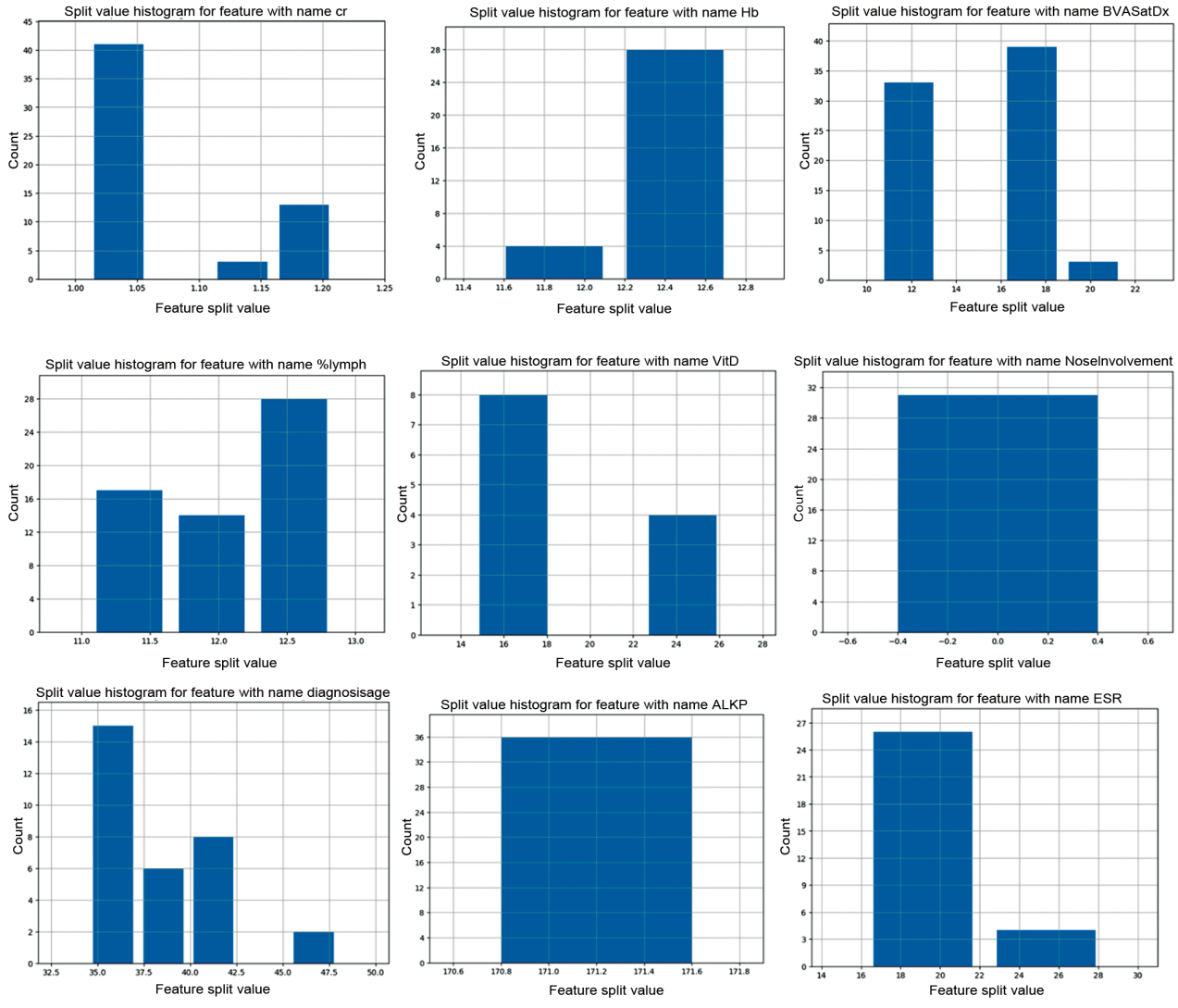

Figure 5. XGBoost GPA prediction model Important features split values histogram

prediction model using machine learning algorithms. We achieved reasonable precision and recall for the relapse prediction model. Nowadays, machine learning techniques are widely used to simulate human decision-making and handle complex tasks. Additional studies should prospectively validate the proposed prediction model to analyze its validity in different patient populations and assess health outcomes. We are also working to enrich our dataset with non-GPA patients to build a prediction model for Wegener's diagnosis.

Conflict of interest statement

The authors declare no conflicts of interest regarding this article

References

1. Lamprecht P, Gross WL. Wegener's granulomatosis. Herz. 2004;29(1):47-56.

doi:10.1007/s00059-004-2525-0.

2. https://www.rheumatology.org/I-Am-A/Patient-Caregiver/Diseases-Conditions/Granulomatosis-with-Polyangitis-Wegners

3. Lutalo PM, D'Cruz DP. Diagnosis and classification of granulomatosis with polyangiitis (aka Wegener's granulomatosis). Journal of autoimmunity. 2014 Feb 1;48:94-8.

4. Boomsma MM, Stegeman CA, Van Der Leij MJ, Oost W, Hermans J, Kallenberg CG, Limburg PC, Tervaert JC. Prediction of relapses in Wegener's granulomatosis by measurement of antineutrophil cytoplasmic antibody levels: a prospective study. Arthritis & Rheumatism: Official Journal of the American College of Rheumatology. 2000 Sep;43(9):2025-33.

5. Benjamin M, McGonagle D. Histopathologic changes at "synovio–entheseal complexes" suggesting a novel mechanism for synovitis in osteoarthritis and spondylarthritis. Arthritis & Rheumatism: Official Journal of the American College of Rheumatology. 2007 Nov;56(11):3601-9.

6. Jain D, Singh V. Feature selection and classification systems for chronic disease prediction: A review. Egyptian Informatics Journal. 2018 Nov 1;19(3):179-89.

7. Chen Y, Huang S, Chen T, Liang D, Yang J, Zeng C, Li X, Xie G, Liu Z. Machine learning for prediction and risk stratification of lupus nephritis renal flare. American Journal of Nephrology. 2021;52(2):152-60.

8. Adamichou C, Genitsaridi I, Nikolopoulos D, Nikoloudaki M, Repa A, Bortoluzzi A, Fanouriakis A, Sidiropoulos P, Boumpas DT, Bertsias GK. Lupus or not? SLE Risk Probability Index (SLERPI): a simple, clinician-friendly machine learning-based model to assist the diagnosis of systemic lupus erythematosus. Annals of the rheumatic diseases. 2021 Jun 1;80(6):758-66.

9. Wang J, Xu J, Zhao C, Peng Y, Wang H. An ensemble feature selection method for high-dimensional data based on sort aggregation. Systems Science & Control Engineering. 2019 Nov 29;7(2):32-9.

10. Pierrot-Deseilligny Despujol C, Pouchot J, Pagnoux C, Coste J, Guillevin L. Predictors at diagnosis of a first Wegener's granulomatosis relapse after obtaining complete remission. Rheumatology. 2010 Nov 1;49(11):2181-90.

11. Sanders JS, Stassen PM, Van Rossum AP, Kallenberg CG, Stegeman CA. Risk factors for relapse in anti-neutrophil cytoplasmic antibody (ANCA)-associated vasculitis: tools for treatment decisions. Clin Exp Rheumatol. 2004 Jan 1;22(6 Suppl 36):S94-101.

12. Han, W. K., Choi, H. K., Roth, R. M., McCluskey, R. T., & Niles, J. L. (2003). Serial ANCA titers: useful tool for prevention of relapses in ANCA-associated vasculitis. Kidney international, 63(3), 1079-1085.

13. G. Tomasson, P. C. Grayson, A. D. Mahr, M. LaValley, and P. A. Merkel, "Value of ANCA measurements during remission to predict a relapse of ANCA-associated vasculitis—a meta-analysis," Rheumatology, vol. 51, no. 1, pp. 100–109, 2012.

14. Hogan PC, O'Connell RM, Scollard S, Browne E, Hackett EE, Feighery C. Biomarkers predict relapse in granulomatosis with polyangiitis. Journal of biomarkers. 2014;2014.

15. Daouk, G. H., Palsson, R., & Arnaout, M. A. (1995). Inhibition of proteinase 3 by ANCA and its correation with disease activity in Wegener's granulomatosis. Kidney international, 47(6), 1528-1536.

16. Dabrowski, A., &Droszcz, W. (2000). Wegener's granulomatosis-clinical analysis of 18 patients group. Medical Science Monitor, 6(1), CS151-CS157.

17. Kemna MJ, van Paassen P, Damoiseaux JG, Cohen Tervaert JW. Maintaining remission in patients with granulomatosis with polyangiitis or microscopic polyangiitis: the role of ANCA. Expert Opinion on Orphan Drugs. 2017 Mar 4;5(3):207-18.

18. Mahr A, Katsahian S, Varet H, Guillevin L, Hagen EC, Höglund P, Merkel PA, Pagnoux C, Rasmussen N, Westman K, Jayne DR. Revisiting the classification of clinical phenotypes of anti-neutrophil cytoplasmic antibody-associated vasculitis: a cluster analysis. Annals of the rheumatic diseases. 2013 Jun 1;72(6):1003-10.

19. Watts R, Lane S, Hanslik T, Hauser T, Hellmich B, Koldingsnes W, Mahr A, Segelmark M, Cohen-Tervaert JW, Scott D. Development and validation of a consensus methodology for the classification of the ANCA-associated vasculitides and polyarteritis nodosa for epidemiological studies. Annals of the rheumatic diseases. 2007 Feb 1;66(2):222-7.

20. Kemna MJ, Damoiseaux J, Austen J, Winkens B, Peters J, van Paassen P, Tervaert JW. ANCA as a predictor of relapse: useful in patients with renal involvement but not in patients with nonrenal disease. Journal of the American Society of Nephrology. 2015 Mar 1;26(3):537-42.

21. Hilhorst M, van Paassen P, Tervaert JW. Proteinase 3-ANCA vasculitis versus myeloperoxidase-ANCA vasculitis. Journal of the American Society of Nephrology. 2015 Oct 1;26(10):2314-27.

22. Vabalas A, Gowen E, Poliakoff E, Casson AJ. Machine learning algorithm validation with a limited sample size. PloS one. 2019 Nov 7;14(11):e0224365.

23. Yata K, Aoshima M. Effective PCA for high-dimension, low-sample-size data with noise reduction via geometric representations. Journal of multivariate analysis. 2012 Feb 1;105(1):193-215.

24. Samson M, Devilliers H, Thietart S, Charles P, Pagnoux C, Cohen P, Karras A, Mouthon L, Terrier B, Puéchal X, Guillevin L. Score to assess the probability of relapse in granulomatosis with polyangiitis and microscopic polyangiitis. RMD open. 2023 Mar 1;9(1):e002953.

25. Chen Z, Tian X, Qu J, Chen J, Yang Y, Li J. Development and internal validation of a

model to predict long-term survival of ANCA associated vasculitis. Rheumatology and Immunology Research. 2023 Apr 18;4(1):30-9.

26. Burkhardt O, Köhnlein T, Wrenger E, Lux A, Neumann KH, Welte T. Predicting outcome and survival in patients with Wegener's granulomatosis treated on the intensive care unit. Scandinavian journal of rheumatology. 2007 Jan 1;36(2):119-24.

27. Casal Moura M, Specks U, Tehranian S, Sethi S, Zubidat D, Nardelli L, Dos Santos FG, Sousa C, León-Róman J, Bobart SA, Greene E. Maintenance of Remission and Risk of Relapse in Myeloperoxidase-Positive ANCA-Associated Vasculitis with Kidney Involvement. The Clinical Journal of the American Society of Nephrology. 2023 Jan 1;18(1):47-59.

28. Radhachandran A, Garikipati A, Iqbal Z, Siefkas A, Barnes G, Hoffman J, Mao Q, Das R. A machine learning approach to predicting risk of myelodysplastic syndrome. Leuk Res. 2021 Oct;109:106639. doi: 10.1016/j.leukres.2021.106639. Epub 2021 Jun 8. PMID: 34171604.