의학박사 학위논문

설명 가능한 인공지능을 사용하여 류마티스관절염 환자에서 생물학적제제에 대한 개인별 관해 예측

Individualized prediction of remission for biologics using explainable artificial intelligence

울 산 대 학 교 대 학 원

의 학 과

구 본 산

Individualized prediction of remission for biologics using explainable artificial intelligence

지 도 교 수　　유　　빈

이 논문을 의학박사 학위 논문으로 제출함

2020 년　8 월

울 산 대 학 교 대 학 원

의　　학　　과

구　본　산

구본산의 의학박사학위 논문을 인준함

심사위원 이 창 근 (인)

심사위원 유 빈 (인)

심사위원 김 용 길 (인)

심사위원 오 지 선 (인)

심사위원 고 은 미 (인)

울 산 대 학 교 대 학 원

2020 년 8 월

국문요약

연구배경: 많은 생물학적제제가 류마티스관절염 환자의 치료에 사용되고 있으나 효과가 있는 생물학적제제를 선택하는데 도움이 되는 요인은 알려지지 않았다. 본 연구의 목적은 류마티스관절염 환자에서 생물학적제제의 반응을 예측하는 머신러닝 모델을 구축하고 설명가능한 인공지능을 이용하여 류마티스관절염의 관해에 필요한 변수들의 중요한 특징을 찾는 것이다.

연구방법: 2012 년 12 월부터 2019 년 6 월까지 대한류마티스학회 생물학적제제 등록사업에서 생물학적제제인 에타너셉트, 아달리무맙, 골리무맙, 인플릭시맙, 토실리주맙으로 치료받은 1,204 명의 환자에서 1,397 개의 추적관찰데이터를 연구에 이용하였다. 생물학적제제를 처음 사용한 시점을 baseline 으로 하고, 이 시점의 임상변수 중에서 데이터 충실도와 임상적 중요도를 고려하여 64 개의 변수를 예측 변수로 선정하였다. 반응변수로는 생물학적제제를 시작한 시점의 baseline 과 동일 약물을 유지하는 다음 추적관찰시점의 DAS28-ESR 을 비교하여 ACR/EULAR response criteria 와 질병활성도 카테고리에 기반하여 다음과 같이 개별 약물에 대한 치료반응을 분류하였다. 1) good response, 2) good response without increasing prednisolone dose, 3) good or moderate response, 4) low disease activity, 5) remission 을 outcome 으로 정의하였고 각 outcome 들로 머신러닝 모델을 구축하였다. 머신러닝 방법은 Lasso, ridge, support vector machine, random forest, XGBoost 의 5 가지 방법을 사용하였다. 다섯개의 치료반응을 예측하는 머신 러닝 모델들 중에서 정확도와 area under receiver operating characteristics (AUROC)가 가장 좋은 한가지를 outcome 으로 선택하기로 하였다. 선택된 outcome 을 예측하는 머신러닝 모델에서 예측에 필요한 변수들의 중요한 특징(important feature)을 알아보기 위하여 Shapley additive explanation (SHAP) 가치 (예측 결과를 바꿀 수 있는 영향력, 절대값이 높을수록 예측결과를 바꾸는데 기여도가 높음)를 사용하였다.

결과: 다섯가지 머신러닝 방법을 이용하여 다섯개의 치료반응을 예측하였다. 이 중에서 생물학적제제 치료 후 remission 과 remission 에 실패한 수의 비율이 비슷하고 정확도와 AUROC 가 높았던 remission 을 outcome 으로 선택하였다. 각각의 생물학적제제에서 remission 예측의 정확도는 52.7%-73.7%, AUROC 는 0.547-0.747 였다. 머신러닝 모델로 remission 을 예측하는데 필요한 important feature

를 알아보기 위하여 Shapley plot 을 작성하고 SHAP 가치를 계산하였다. 모든 생물학적제제, TNF 억제제, 비 TNF 억제제에서 첫번째 important feature 는 프레드니솔론 용량(SHAP 가치 -0.268)이었으며 용량이 적을수록 remission 과 관련이 있고 많을수록 remission 실패와 관련이 있었다.

각각의 생물학적제제에서도 Shapley plot 을 확인하였으며 생물학적제제마다 변수의 영향력과 방향성에서 차이가 있었다. adalimumab 에서는 나이(SHAP 가치 -0.290)가 가장 높은 important feature 로서 어릴수록 remission 과 관계가 있었고 나이가 많으면 remission 실패와 관련이 있었다. 다음으로 혈색소(SHAP 가치 0.160)가 높으면, 프레드니솔론 용량(SHAP 가치 0.143)이 적으면 remission 과 관련이 있었다. etanercpet 에서는 프레드니솔론 용량(SHAP 가치 -0.967) 적으면, C-반응단백(SHAP 가치 0.790)이 높으면, 적혈구침강속도(SHAP 가치 -0.789)가 낮으면 remission 과 관련 있었다. infliximab 에서는 적혈구침강속도(SHAP 가치 -0.507)가 낮으면, 프레드니솔론 용량(SHAP 가치 -0.443)이 낮으면 혈색소(SHAP 가치 0.391)이 높으면 remission 과 관련이 있었다. golimumab 에서는 적혈구침강속도(SHAP 가치 -1.347)이 낮으면, 프레드니솔론 용량(SHAP 가치 -1.203)이 낮으면, 항 CCP 항체(SHAP 가치 -0.864)가 낮으면 remission 과 관련이 있었다. abatacept 에서는 유병기간(SHAP 가치 -0.845)이 짧으면, 프레드니솔론 용량(SHAP 가치 -0.817)가 적으면, 혈색소(SHAP 가치 0.773)이 높으면 remission 과 관련이 있었다. tocilizumab 에서는 프레드니솔론 용량(SHAP 가치 -0.175)가 낮으면, C-반응단백(SHAP 가치 0.134)이 높으면, 적혈구침강속도(SHAP 가치 -0.097)이 낮으면 remission 과 관련이 있었다.

각각의 생물학적제제에서 변수별로 SHAP 가치의 평균을 구했을 때 가장 높은 important feature 는 프레드니솔론 용량(평균 SHAP 가치 -0.535)였고 모든 생물학적제제에서 각 생물학적제제별 프레드니솔론 용량의 SHAP 가치는 remission 과 음의 상관관계였다. 그 다음은 적혈구침강속도(평균 SHAP 가치 -0.448)이고 adalimumab 에서만 remission 과 양의 상관관계였고 나머지 생물학적제제에서는 모두 remission 과 음의 상관관계였다. 세번째는 혈색소(평균 SHAP 가치 0.337)였으며 golimumab 에서만 remission 과 음의 상관관계였고 나머지 생물학적제제에서는 모두 remission 과 양의 상관관계였다. 항 CCP 항체(평균 SHAP 가치 -0.285)의 경우 tocilizumab 에서 remission 에 영향력이 없었으나 나머지 생물학

적제제에서는 모두 remission과 음의 상관관계가 있었다. 머신러닝을 이용하여 remission을 예측할 때 각 생물학적제제마다 변수들의 영향력이 다르게 작용한다는 것을 확인하였다.

결론: 머신러닝을 이용하여 각각의 생물학적제제에서 remission을 예측하였다. 또한 설명가능한 인공지능을 이용하여 생물학적제제마다 remission으로 예측된 환자들의 임상적 특징의 차이를 확인하였다. 이러한 머신러닝 방법을 활용하면 생물학적제제의 선택에 도움을 줄 뿐만 아니라 류마티스관절염 환자의 치료결과를 향상시키는데 도움을 줄 수 있다.

Table of contents

List of Tables

List of Figures

List of Appendix

Introduction

Rheumatoid arthritis (RA) is a chronic inflammatory disease affecting the synovial tissue in multiple joints. Biological therapies, known as biologics, are considered in patients who have high disease activity despite being treated with conventional disease modifying anti-rheumatic drugs (cDMARDs), including methotrexate, leflunomide, and sulfasalazine.[1-3] Recently, various biologics, such as tumor necrosis factor (TNF) inhibitors, abatacept, tocilizumab, and rituximab, have become available for the treatment of RA.

Considering various factors, such as disease activity, side effects, and cost-effectivity, biologics are generally prescribed by referring to recommendations in clinical practice.[1-3] However, treatment with biologics may fail as a result of differences in clinical characteristics among individuals. Indeed, in clinical trials, 30%–40% of patients treated with a biologics do not respond, and the reaction rate decreases with subsequent biologics.[4, 5] Approximately 6% of patients who started TNF inhibitors as a first biologic were subsequently classified as refractory.[6]

Treatment failure due to empirical selection of biologics not only increases the patient's pain but also increases the cost of healthcare.[7] Therefore, it is necessary to develop a biomarker that can identify which biologics will have a good effect on a case-by-case basis. Although several clinical, genetic, and proteomic studies have tried to identify biomarkers that could predict the response to biologics in patients with RA, it is impossible to represent the features of all patients with a single biomarker. In addition, most of the biomarkers are impractical for clinicians and patients to use in routine clinics. Therefore, it may be useful to predict the response of biologics using various information from clinical data that are easily accessible in routine clinics.

Statistical methods such as regression show simplicity rather than complexity aiming to explain the relationship between variables, while machine learning focuses on accurate predictions. Although machine learning methods are difficult to interpret, they can incorporate many more variables, are generalizable across a much broader array of data types, and can produce results in more complex situations.[8] We are in the process of collecting various pieces of clinical information from patients with RA in the study cohort.[9] However, there are limitations in explaining complex relationships between various clinical variables, disease

activity, and drug effects by statistical methods alone. In addition, it is difficult to predict the effect of a drug using a statistical method due to varying genetic and environmental risk factors.[10] Therefore, machine learning methods are superior in terms of analyzing the effects of variables, such as clinical information of RA patients, on the effect of the drug.

In the machine learning method, outcomes are predicted using complex variables. However, it is difficult for humans to understand how machine learning predicts outcomes from relationships among numerous variables. In order to overcome these shortcomings, several methods have been proposed to improve the uncertainty and increase the reliability of machine learning methods. Explainable artificial intelligence (AI), which presents the reason for prediction in a way that humans can understand, suggest the relationship between various variables necessary to predict an outcome.[11]

In terms of the complex clinical data of patients with RA, the study aim was to determine the response to each biologic using machine learning in combination with many outpatient variables. This method may be sufficient for predicting the effects of various biologics for multifactorial RA. In addition, using explainable AI, determining the most important factors that contribute to a good response to each biologic would allow the application of personalized treatment, as well as an understanding of the characteristics of biologics by class.

The purpose of this study was to establish a machine learning model that can predict remission in RA patients treated with biologics using data of RA patients from the Korean College of Rheumatology Biologics and Targeted Therapy Registry (KOBIO). This study also aimed to identify the factors that have the greatest influence on the response to biologics in the machine learning model using explainable AI.

Methods

Study population

This retrospective cohort study used data from the KOBIO registry, which is a nationwide multicenter cohort, to evaluate the effectiveness and side effects of biologics therapy in patients with RA.[9] Patients were recruited from 38 hospitals in 2012. Demographics, medications, comorbidities, extra-articular manifestations, disease activity, radiographic findings, and laboratory findings were recorded by the investigators. The data of patients who were followed up annually were recorded on the KOBIO website (http://www.kobio.or.kr/kobio/), and the patients provided informed consent before registration. The study was approved by the Institutional Review Board of Inje University Seoul Paik Hospital (PAIK 2018-11-005). The clinical trial committee under the Korean Society of Rheumatology approved the study and provided the data.

Data collection

From December 2012 to June 2019, a total of 2,122 patients who started biologic disease modifying anti-rheumatic drugs (bDMARDs) and targeted synthetic disease modifying anti-rheumatic drugs (tsDMARDs) were registered (Figure 1 and 2). At the first visit, the baseline data were registered, and follow-up data were registered each subsequent year. tsDMARDs such as tofacitinib were excluded because the aim was to predict patients with a response to bDMARDs. Among the bDMARDs, rituximab was excluded from the analysis due to the small number of samples. Of the 2,122 patients, 464 who did not follow-up after initial enrollment in the cohort, and 353 patients who previously used bDMARDs were excluded. Excluding 35 patients who were treated with rituximab and tofacitinib, and 66 who had missing values, a total of 1,204 patients treated with bDMARDs, such as adalimumab, etanercept, infliximab, golimumab, tocilizumab, and abatacept, were investigated. The disease activity scores in 28 joints using the erythrocyte sedimentation rate (DAS28-ESR) was measured at baseline and follow-up (Figure 3), and 1,397 bDMARDs with follow-up DAS28-ESR data were found in 1,204 patients.

Outcome construction

Prior to constructing a model that predicts treatment response using machine learning, an appropriate outcome with high accuracy and reliability in response definition was explored.

Five outcomes with binary variables were defined to determine the response to each bDMARD using baseline and current DAS28-ESR (Appendix Table 1):[12, 13] 1) 'Good response' (current DAS28-ESR at follow-up ≤ 3.2 and improvement from baseline > 1.2); 2) 'good response without increasing prednisolone dose'; 3) 'good and moderate response' (improvement from baseline > 1.2 or improvement from baseline > 0.6 to ≤ 1.2 with current DAS28-ESR ≤ 5.1); 4) 'low disease activity' (current DAS28-ESR ≤ 3.2); and 5) 'remission' (current DAS28-ESR ≤ 2.6).

Machine learning methodology

Various machine learning models were used to predict outcomes; these included lasso and ridge[1] based on a linear relationship,[14] support vector machine[2] using kernel methods,[15] tree-based random forest[3],[16] and the Xgboost model[4] [17]. These five machine learning models were

---

[1]Lasso (least absolute shrinkage and selection operator) and Ridge regression are methods to reduce model complexity and prevent over-fitting. Lasso regression helps to reduce over-fitting with feature selection. Ridge regression shrinks the coefficients, which helps to reduce the model complexity and multi-collinearity.

[2]Support vector machine is learning algorithms that analyze data used for classification and regression analysis. The objective of the support vector machine algorithm is to determine a hyperplane in an N-dimensional space that distinctly classifies the data points.

[3]Random forests are an ensemble learning method that operate by constructing a multitude of decision trees at training and outputting the class that is classification or regression of individual trees.

[4]XgBoost is a software library that provides a gradient boosting framework. Gradient boosting is a machine learning technique for regression and classification problems that produces a prediction model in the form of an ensemble of weak prediction models, such as decision trees. Gradient boosting builds the model in a stage-wise fashion, similar to other boosting methods, and generalizes them by allowing optimization of an arbitrary differentiable loss function.

used to predict all five outcome criteria in each bDMARDs, and the most appropriate outcome among the five outcome criteria was selected.

K-fold cross-validation and dimension reduction

When constructing a predictive model using a training set with too many variables, the production of model corresponds too closely or exactly to a training set (overfitting) and may therefore fail to predict outcomes with test data reliably. To avoid this overfitting problem, the training set and the test set were divided at a 7:3 ratio, the models were trained with the training set, and the prediction results were verified with the rest of the test set. For the training data set, a five-fold cross-validation was performed in order to tune the hyperparameters that were determined as outside models (Figure 4). In this procedure, a grid search was conducted in order to evaluate all possible combinations of hyperparameters. The grid search found optimal hyperparameters with the objective function of determining the area under the receiver operating characteristics (AUROC) in each model. Bootstrapping (random sampling with replacement) was also performed in order to obtain a median value for the AUROC curve, as well as to determine the accuracy to reduce the variance of the measurements caused by small samples when dividing the training and test sets.

Dimension reduction was conducted to avoid the "curse of dimensionality" caused by a large number of variables compared to the size of the data. Among 64 variables, 15 variables that were known to be of medical importance were preselected (sex, age, baseline DAS28-ESR, methotrexate dose, steroid dose, erythrocyte sedimentation rate [ESR], C-reactive protein [CRP], rheumatoid factor [RF], anti-cyclic citrullinated peptide antibody [ACPA], anti-nuclear antibody, and five comorbidities). Subsequently, 20 variables that were highly correlated with the drug response of each bDMARD were added to train the models.

Explainable AI for important features in response to bDMARDs

From this machine learning model, the most important variables, and their impact on predicting the selected response criteria using explainable AI were determined. By focusing on model performance, including model accuracy, machine learning models have become complex and lost their interpretability, as shown in Xgboost and deep learning. Although there is some feature importance in random forest and Xgboost, these models provide an inconsistent measure depending on the tree structure; in addition, they only show the overall

importance, and not the direction of the effect of independent variables.[11] The Shapley additive explanations (SHAP) method was utilized with the intention to improve these problems.[18] The SHAP method approximates a complex model to a linear model and interprets the feature importance in that linear model, in order to demonstrate how much a given feature changes the prediction. This methodology satisfies three conditions; that the approximated linear model has similar accuracy to the original model in the local domain, that meaningless variables have no impact in explanatory power, and that feature importance is consistent in the model structure. Using the methodology, consistent feature importance, regardless of model structure and direction of effect of predictive variables was demonstrated, suggesting clinician intuitive insights to increase or decrease the response outcome and find potential variables affecting the selection of bDMARDs.

Predicting subsequent bDMARDs using machine learning models

Remission of the subsequent bDMARDs in patients who failed remission with bDMARDs in the original data were predicted. The ensemble method of five machine learning models based on the majority vote were used to determine whether the prediction models were available in clinical practice.

Statistical analysis

All data were summarized as mean (standard deviation) or percentage. To evaluate the machine learning performance, the accuracy and AUROC curve were used. The no information rate[5] was used as a baseline to determine the overall distribution of the classification and compare it with the machine learning models. Statistical analyses were performed using R statistical language version 3.6.1 (The R Project for Statistical Computing, Vienna, Austria), and model training was performed using the caret package and SHAPforxgboost package in R.

[5]The no information rate is the largest proportion of the observed classes. A hypothesis test is computed to evaluate whether the overall accuracy rate is greater than the rate of the largest class.

Results

Clinical characteristics of the patients

Table 2 shows the clinical characteristics of the study cohort. The mean age at baseline was 54.02 (12.76) years of age, the majority of patients were female (82.62%), and the disease duration was 7.07 (7.16) years. RF and ACPA positivity were 83.20% and 73.43%, respectively, while the mean DAS28-ESR at 1,204 baseline and 1,397 follow-ups were 5.62 (1.02) and 4.34 (1.28]), respectively.

Outcome setting according to response to biologics

Table 2 shows the response to bDMARDs with baseline and follow-up disease activity for 1,397 follow-ups. The follow-ups were groups according to the five response outcome criteria (Table 3). Among the follow-ups; 1) the number of follow-ups corresponding to the ‘good response’ was 791 (56.6%), 2) the number of follow-ups corresponding to the ‘good response without increasing prednisolone dose’ was 739 (52.9%), 3) the number of follow-ups corresponding to the ‘good or moderate response’ was 1,203 (86.1%), 4) the number of follow-ups corresponding to the ‘low disease activity’ was 821 (58.8%), and 5) the number of follow-ups corresponding to ‘remission’ was 564 (40.4%).

Selection of best response outcome

Five machine learning methods were used to predict five response criteria for each bDMARD: 1) ‘good response’ (Appendix Table 2), 2) ‘good response without increasing prednisolone dose’ (Appendix Table 3), 3) ‘moderate or good response’ (Appendix Table 4), 4) ‘low disease activity’ (Appendix Table 5), and 5) ‘remission’ (Table 4). The no information rate, accuracy, and AUROC curve were 50.0% - 74.6%, 54.2% - 74.6%, and 0.529 – 0.734 in ‘good response’, 52.6% - 69.7%, 56.2% - 73.0%, and 0.531 – 0.788 in ‘good response without increasing prednisolone dose’, 82.3% - 94.3%, 72.2% - 94.3%, and 0.570 – 0.832 in ‘good or moderate response’, 50.0% - 76.2%, 53.8% - 76.2%, and 0.5000 – 0.756 in ‘low disease activity’, and 51.7% - 68.6%, 57.2% - 73.7%, and 0.547 – 0.767 in ‘remission’.

Among the response definitions, the ‘remission’ showed a lower no information rate than 'moderate or good response', and higher accuracy and AUROC curve than 'good response', 'good response without increasing prednisolone dose', and 'low disease activity'. Therefore, because the ‘remission’ was less affected by the measurement error of outcome than the other

response criteria, remission was defined as the final response outcome.

Prediction of remission in machine learning models

In all machine learning methods for predicting remission, the accuracy and AUROC curve were 57.2%–73.7% and 0.547–0.767, respectively (Table 3). The accuracy and AUROC curve of remission prediction were 60.2%–64.0% and 0.652–0.674 in all bDMARDs, 71.3%–72.6% and 0.693–0.714 in TNF inhibitor, and 57.2%–62.2% and 0.609–0.663 in non-TNF inhibitor, respectively. Among bDMARDs, the accuracy and AUROC curve were 66.3%–69.8% and 0.626–0.707 in adalimumab, 66.2%–70.8% and 0.695–0.743 in etanercept, 66.7%–69.4% and 0.701–0.767 in golimumab, 66.7%–72.9% in infliximab, 66.7%–73.7% and 0.663–0.759 in abatacept, and 59.8%–61.0% and 0.548–0.603 in tocilizumab, respectively. For each bDMARD, the accuracy and AUROC curve were similar across different machine learning models.

Important features for remission in explainable AI

The SHAP method for remission was performed in order to determine the influence of variables that contributed to remission in the prediction model. The interpretation of feature importance with the Shapley plot is shown in Figure 5. Figure 5 lists the SHAP values in the order of the absolute values, which are the order of the important variables that contribute to decide the predicted outcome of remission. For each variable, the feature value change from light to dark indicates that the value of the variable changes from low to high. In the example of age, the lighter feature value color represents younger age, while the darker color represents older age. In adalimumab, the increasing SHAP value with lighter color indicates that the relationship with remission increases with decreasing age, while the decreasing SHAP value with darker color indicates that the relationship with remission failure increases with increasing age. Therefore, the linear change of the feature value is expressed in color, and the SHAP value (impact on model output) indicates whether it is close to the remission or the remission failure according to the feature value.

In all bDMARDs, prednisolone dose was the variable that had the highest impact on remission depending on the feature value (SHAP value = -0.268). A low prednisolone dose was associated with remission, while a high dose was associated with remission failure; a low ESR level (SHAP value = -0.103) was associated with remission, while a high ESR level was

associated with remission failure; and a high hemoglobin level (SHAP value = 0.068) was associated with remission, while a low hemoglobin level was associated with remission failure. Furthermore, low levels of DAS28-ESR (SHAP value = -0.064), triglyceride (SHAP value = -0.049), disease duration (SHAP value = -0.028), ACPA (SHAP value = -0.025), cholesterol (SHAP value = -0.014), age (SHAP value = -0.012), and low density lipoprotein (SHAP value = -0.011) were also associated with remission. However, a high CRP level (SHAP value = 0.028) was associated with remission.

Important features for remission in TNF inhibitors and non-TNF inhibitors

In TNF inhibitors, prednisolone dosage was the variable with the greatest impact (SHAP value = -0.343), and patients with a high prednisolone dose had a lower rate of remission. Low levels of ESR (SHAP value = -0.176) and DAS28-ESR (SHAP value = -0.136) were associated with remission, while a high hemoglobin level (SHAP value = 0.135) was associated with remission. Furthermore, low levels of triglyceride (SHAP value = -0.121), disease duration (SHAP value = -0.112), age (SHAP value = -0.103), and ACPA (SHAP value = -0.099) were associated with remission. In CRP levels (SHAP value = -0.085), the feature value showed no linear relationship with remission or remission failure. Low CRP was associated with both remission and remission failure, but high CRP showed relatively weak association with remission failure. In non-TNF inhibitors, the prednisolone dose was the variable with the greatest impact (SHAP value = -0.284), and low prednisolone dose was associated with remission. High levels of methotrexate dose (SHAP value = 0.184) and CRP (SHAP value = 0.156) were associated with remission, as were low ESR level (SHAP value = -0.127) and high hemoglobin level (SHAP value = 0.104). Furthermore, low levels of platelet (SHAP value = -0.096), low density lipoprotein (SHAP value = -0.085), age (SHAP value = -0.078), and RF (SHAP value = -0.075) were associated with remission.

Important features for remission in each bDMARD

In adalimumab, the most important feature was age (SHAP value = -0.290), and older patients had a lower rate of remission, while younger patients had a higher rate of remission. A high hemoglobin level (SHAP value = 0.160) and low prednisolone doses (SHAP value = -0.143) were associated with remission, as were low levels of disease duration (SHAP value = -0.097), aspartate aminotransferase (SHAP value = -0.071), ACPA (SHAP value = -0.065), DAS28-

ESR (SHAP value = -0.063), triglyceride (SHAP value = -0.052), and cholesterol (SHAP value = -0.045).

In etanercept, the most important feature was prednisolone dosage (SHAP value = -0.967); a low prednisolone dose was associated with remission, and a high dose was associated with remission failure. Moreover, a high CRP level (SHAP value = 0.790), low levels of ESR (SHAP value = -0.789), RF (SHAP value = -0.717), DAS28-ESR (SHAP value = -0.582), platelet (SHAP value = -0.565), age (SHAP value = -0.518), and cholesterol (SHAP value = -0.448) were associated with remission.

In infliximab, the most important feature was ESR (SHAP value = -0.507); a low ESR level was associated with remission, and a high ESR level was associated with remission failure. Moreover, a low prednisolone dose (SHAP value = -0.443), high hemoglobin level (SHAP value = 0.391), short disease duration (SHAP value = -0.287), low ACPA level (SHAP value = -0.287), and low cholesterol level (SHAP value = -0.244) were associated with remission. Treatment of LTBI (SHAP value = 0.227), a high glucose level (SHAP value = 0.225), and a high alanine aminotransferase level (SHAP value = 0.218) were associated with remission.

In golimumab, the most important feature was ESR (SHAP value = -1.347); a low ESR was associated with remission, and a high ESR was associated with remission failure. A low prednisolone dose (SHAP value = -1.203), low ACPA (SHAP value = -0.864), low DAS28-ESR (SHAP value = -0.715), and low hemoglobin (SHAP value = -0.3642) level were associated with remission. Moreover, a high RF level (SHAP value = 0.630), young age (SHAP value = -0.582), and short disease duration (SHAP value = -0.484) were associated with remission.

In abatacept, the most important feature was disease duration (SHAP value = -0.845); a short disease duration was associated with remission or remission failure, while a long disease duration was associated with remission failure. A low prednisolone dose (SHAP value = -0.817), high hemoglobin level (SHAP value = 0.773), high methotrexate dose (SHAP value = 0.671), low ANA titer (SHAP value = 0.596), low platelet level (SHAP value = 0.499), and low ACPA level (SHAP value = -0.487) were associated with remission.

In tocilizumab, the most important feature was prednisolone dosage (SHAP value = -0.175); a low prednisolone dose was associated with remission, and a high dose was associated with

remission failure. Furthermore, a high CRP level (SHAP value = 0.134), low ESR level (SHAP value = -0.097), low DAS28-ESR (SHAP = -0.076), low RF level (SHAP value = -0.066), young age (SHAP value = -0.065), low platelet level (SHAP value = -0.064), high hemoglobin level (SHAP value = 0.056), and low cholesterol level (SHAP value = -0.056) were associated with remission.

Impact of each bDMARD in the ranking of important features

Table 5 shows the mean SHAP values of variables in all types of bDMARDs. The degree and direction of the contribution of the variable to remission was different with each bDMARD. In all bDMARDs, the prednisolone dose was negatively associated with remission and had the highest mean absolute SHAP value (mean SHAP value = –0.535) among variables. Furthermore, the ESR was negatively associated with remission and had the second highest mean absolute SHAP value (mean SHAP value = –0.448). The ESR was negatively associated with remission in etanercept, infliximab, golimumab, abatacept, and tocilizumab, but was positively associated with remission in adalimumab. Hemoglobin was positively associated with remission and had the third highest mean absolute SHAP value (mean SHAP value = 0.337). Furthermore, Hemoglobin was positively associated with adalimumab, etanercept, infliximab, abatacept, and tocilizumab, but was negatively associated with remission in golimumab. ACPA was negatively associated with remission and had the fourth highest mean absolute SHAP value (mean SHAP value = -0.285); however, it was not negatively or positively associated with remission in tocilizumab (SHAP value = 0). RF was positively associated with remission in adalimumab, infliximab, and golimumab, and negatively associated with remission in etanercept and tocilizumab. However, it showed a non-linear relationship in the Shapley plot in abatacept (Figure 5). In order of mean SHAP value, age, CRP, disease duration, DAS28-ESR, platelet, methotrexate dose, ANA, cholesterol, triglyceride, and alanine aminotransferase were either positively or negatively associated with remission.

Prediction remission of subsequent bDMARDs

The combination of various important features was shown to contribute to determination of the remission from the explainable AI SHAP value. A combination of variables in machine learning for each bDMARD was used to predict the remission rate when patients who failed

remission were treated with the other bDMARDs. Clinical data of follow-ups with remission failure in the real world were applied to the ensemble of the five machine learning models for each bDMARD, and the remission rate for the hypothetical bDMARD treatment was obtained (Table 6).

When the clinical data of follow-ups that failed remission in adalimumab treatment were applied to the machine learning model, the remission rate was 20.2%, 13.4%, 11.1%, 24.7%, and 76.3% in etanercept, golimumab, infliximab, abatacept, and tocilizumab, respectively. When failed remission occurred in etanercept, the remission rate was 9.0%, 10.3%, 9.0%, 26.2%, and 73.8% in adalimumab, golimumab, infliximab, abatacept, and tocilizumab, respectively. When failed remission occurred in golimumab, the remission rate was 8.6%, 19.8%, 12.3%, 25.9%, and 83.4% in adalimumab, etanercept, infliximab, abatacept, and tocilizumab, respectively. When failed remission occurred in infliximab, the remission rate was 5.1%, 14.5%, 7.7%, 26.5%, and 79.5% in adalimumab, etanercept, golimumab, abatacept, and tocilizumab, respectively. When failed remission occurred in abatacept, the remission rate was 6.8%, 15.9%, 9.1%, 10.2%, and 68.2% in adalimumab, etanercept, golimumab, infliximab, and tocilizumab, respectively. When failed remission occurred in tocilizumab, the remission rate was 9.4%, 18.8%, 13.1%, 8.8%, and 17.5% in adalimumab, etanercept, golimumab, infliximab, and abatacept, respectively. The remission rate was higher in the prediction model of non-TNF inhibitors compared to that of TNF inhibitors. Moreover, among the prediction model of all bDMARDs, tocilizumab had the highest remission rate of all bDMARDs, while among the prediction model of TNF inhibitors, etanercept had the highest remission rate of all TNF inhibitors.

Table 1. Clinical characteristics of 1,204 patients treated with biologics.

Variable	Value
Age, mean (SD), year	54.03 (12.79)
Female (%)	82.62
Disease duration, mean (SD), year	7.07 (7.16)
Non-smoking (%)	84.35
History of cardiovascular diseases (%)	3.89
History of lung diseases (%)	6.13
History of hemato-oncologic diseases (%)	1.32
HBsAg positivity (%)	3.48
HBsAb_positivity (%)	46.69
HBcAb positivity (%)	7.62
HCV Ab positivity (%)	0.75
Rheumatoid factor positivity (%)	83.20
Rheumatoid factor, mean (SD), mg/dL	141.60 (216.78)
Anti-CCP antibody positivity (%)	73.43
Anti-CCP antibody, mean (SD), mg/dL	190.28 (242.05)
Erythrocyte sedimentation rate, mean (SD), mm^3/hr	48.63 (26.79)
C-reactive protein, mean (SD), mg/dL	2.36 (3.02)
Anti-nuclear antibody (%)	35.35
Methotrexate dose, mean (SD), mg	10.47 (5.43)
Prednisolone* dose, mean (SD), mg	3.01 (2.73)
DAS28-ESR at baselines, mean (SD)	5.64 (1.01)
Period from start of biologics therapy to first follow-up, mean (SD), year	0.97 (0.31)
DAS28-ESR at follow-up, mean (SD) (n=1,397)	4.34 (1.28)

*Glucocorticoid dose such as prednisolone, methylprednisolone, deflazacort, and dexamethasone was converted to prednisolone doses.
DAS28-ESR: disease activity score in 28 joints using erythrocyte sedimentation rate; HBsAg: Hepatitis B surface antigen; HBsAb: Hepatitis B surface antibody; HBcAb: Hepatitis B core antibody; HCV: hepatitis C virus

Table 2. Distribution of response to bDMARDs

Current DAS28	DAS28: Improvement vs baseline			
	>1.2	>0.6 to ≤1.2	≤0.6	Total
≤3.2	791	16	14	821
>3.2 to ≤5.1	312	65	53	430
>5.1	19	14	113	146
Total	1,122	95	180	1,397

DAS28-ESR: disease activity score in 28 joints using erythrocyte sedimentation rate.

Table 3. The construction and descriptive statistics of output variables

Response to biologics	**Follow-ups (n=1,397)**	**Percent**
Good Response	791	56.6%
Good Response without increasing prednisolone dose	739	52.9%
Good or moderate Response	1,203	86.1%
Low disease activity or remission	821	58.8%
Remission	564	40.4%

bDMARDs, biologics disease modifying anti-rheumatic drugs.

Table 4. Prediction of remission for biologic DMARD

	Remission / total follow-ups	Measure	Baseline*	Lasso	Ridge	SVM	Random Forest	Xgboost
All bDMARDs	564/1,397	Accuracy	59.6%	63.5%	64.0%	62.9%	60.2%	63.2%
		AUROC	0.500	0.660	0.674	0.664	0.654	0.652
TNF inhibitors	252/793	Accuracy	68.6%	71.9%	72.6%	71.7%	71.3%	71.3%
		AUROC	0.500	0.713	0.714	0.705	0.694	0.693
Non-TNF inhibitors	312/604	Accuracy	51.7%	61.9%	62.2%	61.7%	58.3%	57.2%
		AUROC	0.500	0.661	0.663	0.659	0.648	0.609
Adalimumab	91/289	Accuracy	68.6%	69.8%	69.8%	66.3%	68.6%	69.8%
		AUROC	0.500	0.680	0.707	0.665	0.668	0.626
Etanercept	75/220	Accuracy	66.2%	69.2%	70.8%	69.2%	69.2%	67.7%
		AUROC	0.500	0.735	0.743	0.721	0.695	0.681
Golimumab	41/122	Accuracy	66.7%	69.4%	69.4%	69.4%	66.7%	69.4%
		AUROC	0.500	0.767	0.759	0.729	0.716	0.701
Infliximab	45/162	Accuracy	72.9%	71.9%	72.9%	66.7%	72.9%	66.7%
		AUROC	0.500	0.685	0.707	0.652	0.648	0.547
Abatacept	62/194	Accuracy	68.4%	73.7%	73.7%	71.1%	68.4%	66.7%
		AUROC	0.500	0.729	0.759	0.728	0.691	0.663
Tocilizumab	250/410	Accuracy	61.0%	61.0%	61.0%	61.0%	61.0%	59.8%
		AUROC	0.500	0.602	0.603	0.595	0.581	0.548

* no information rate for bDMARDs

bDMARDs, biologics disease modifying anti-rheumatic drugs; SVM, support vector machine; TNF, Tumor necrosis factor inhibitors; AUROC, area under the receiver operating characteristic.

Table 5. Order of important features for prediction of remission in each biologic.

Clinical Variable	Mean SHAP value	Adalimumab	Etanercept	Infliximab	Golimumab	Abatacept	Toclilzumab
Prednisolone dose	-0.535	-0.143	-0.967	-0.443	-1.203	-0.817	-0.175
Erythrocyte sedimentation rate	-0.448	+0.041	-0.789	-0.507	-1.347	-0.355	-0.097
Hemoglobin	+0.337	+0.160	+0.335	+0.391	-0.642	+0.773	+0.056
Anti-CCP antibody	-0.285	-0.065	-0.324	-0.253	-0.864	-0.487	0
Rheumatoid factor	0.266*	+0.037	-0.717	+0.105	+0.630	0.306*	-0.068
Age	-0.263	-0.290	0.518*	-0.167	-0.582	-0.221	-0.065
C-reactive protein	+0.253	0	+0.790	+0.065	+0.326	+0.454	+0.134
Disease duration	-0.245	-0.097	0	-0.287	-0.484	-0.845	0
DAS28-ESR	-0.229	-0.063	-0.582	-0.072	-0.715	-0.093	-0.076
Platelet	-0.228	0	-0.565	0	0.469*	-0.499	-0.064
Methotrexate dose	+0.162	0	-0.372	+0.044	0	+0.671	+0.048
Anti-nuclear antibody	-0.116	0	0	-0.066	-0.152	-0.596	0
Cholesterol	-0.114	-0.045	-0.448	-0.244	0	0	-0.061
Triglyceride	-0.065	-0.052	-0.212	0	-0.189	0	0
Alanine aminotransferase	0.055*	0	0	+0.218	-0.167	0	0

* a non-linear relationship such as quadratic effect or mixed effect between drugs and variables
DAS28-ESR: disease activity score in 28 joints using erythrocyte sedimentation rate

Table 6. Predicting remission of each DMARDs for remission failures using machine learning model

bDMARDs that have failed remission	Remission failure / total	Prediction models of each bDMARDs for remission failures					
		Adalimumab	Etanercept	Golimumab	Infliximab	Abatacept	Tocilizumab
Adalimumab	198/289	-	20.2%	13.4%	11.1%	24.7%	76.3%
Etanercept	145/220	9.0%	-	10.3%	9.0%	26.2%	73.8%
Golimumab	81/122	8.6%	19.8%	-	12.3%	25.9%	83.4%
Infliximab	107/162	5.1%	14.5%	7.7%	-	26.5%	79.5%
Abatacept	132/194	6.8%	15.9%	9.1%	10.2%	-	68.2%
Tocilizumab	160/410	9.4%	18.8%	13.1%	8.8%	17.5%	-

* Ensemble models of all of five machine learning models based on majority vote for each bDMARDs

Fig. 1. Study overview

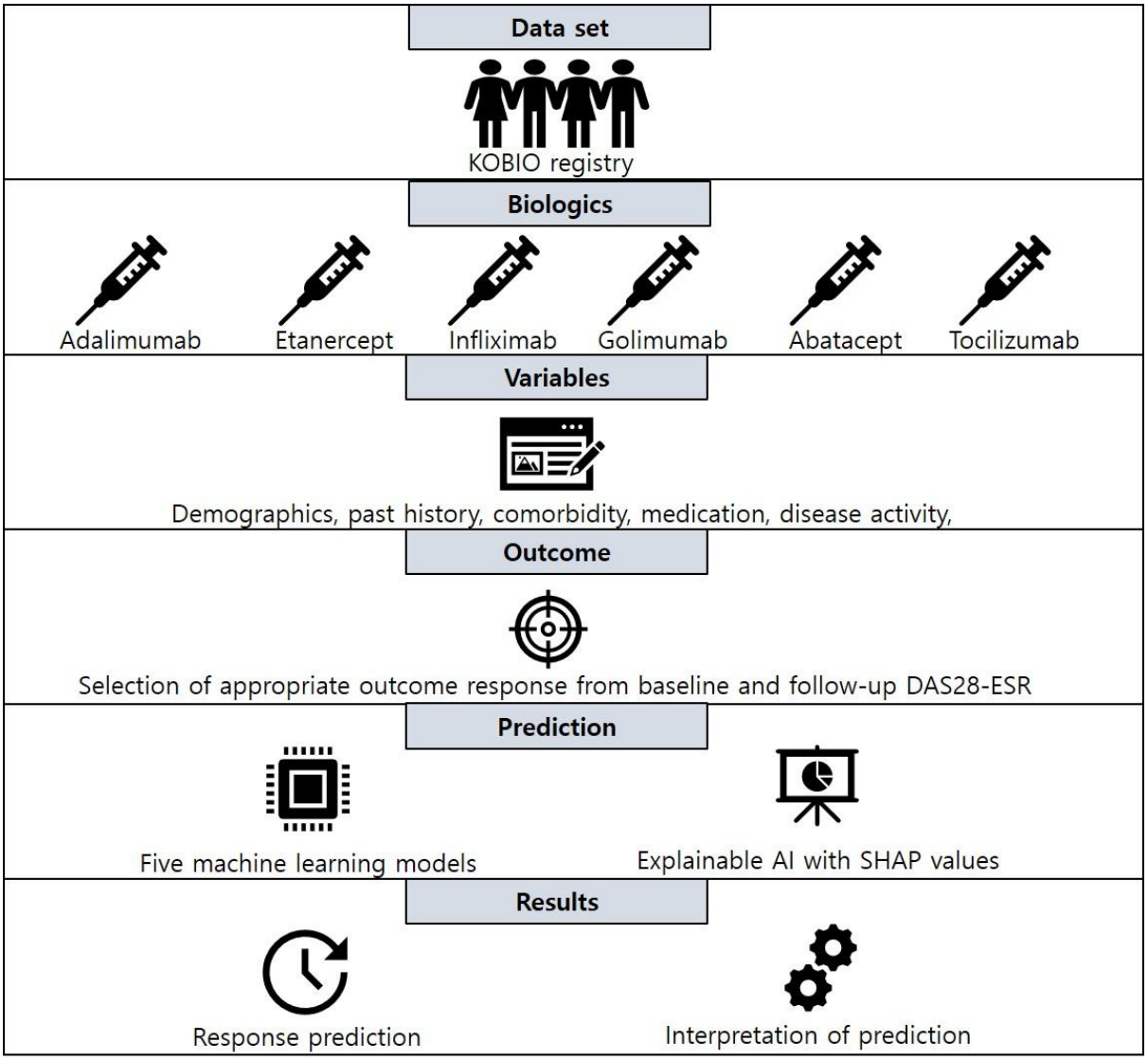
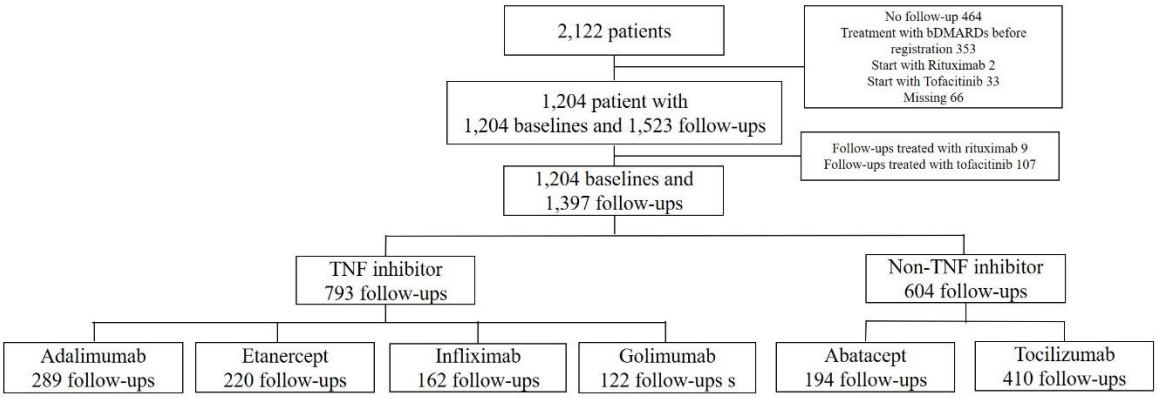
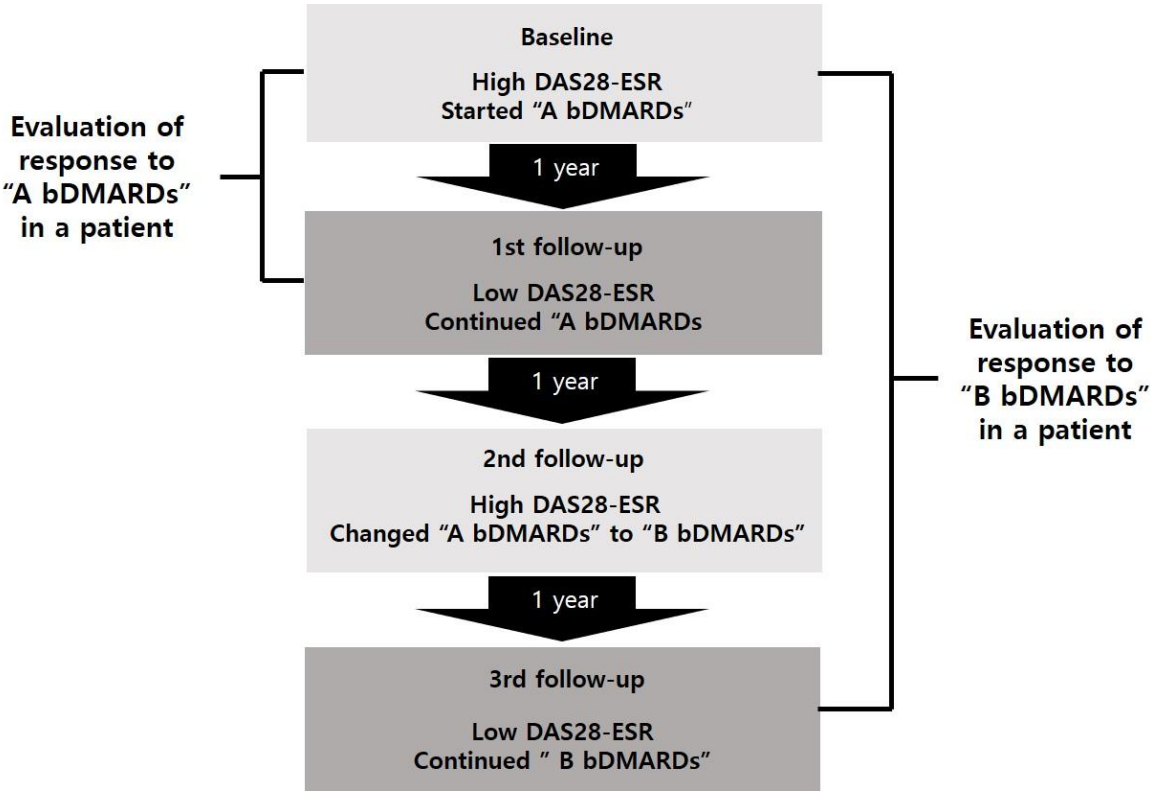


Fig. 2. Schematic diagram that defines the treatment response of a bDMARDs.



bDMARDs, biologics disease modifying anti-rheumatic drugs; TNF, tumor necrosis factor

Fig. 3. Flow-chart of patient selection. Current DAS28-ESR and improvement were determined using baseline DAS28-ESR and follow-up DAS28-ESR after approximately 1 year of treatment.



bDMARDs, biologics disease modifying anti-rheumatic drugs.

Fig. 4. 5-fold cross-validation to tune hyperparameters. Each bDMARDs data were divided as the training set and the test set with 7:3 ratio.

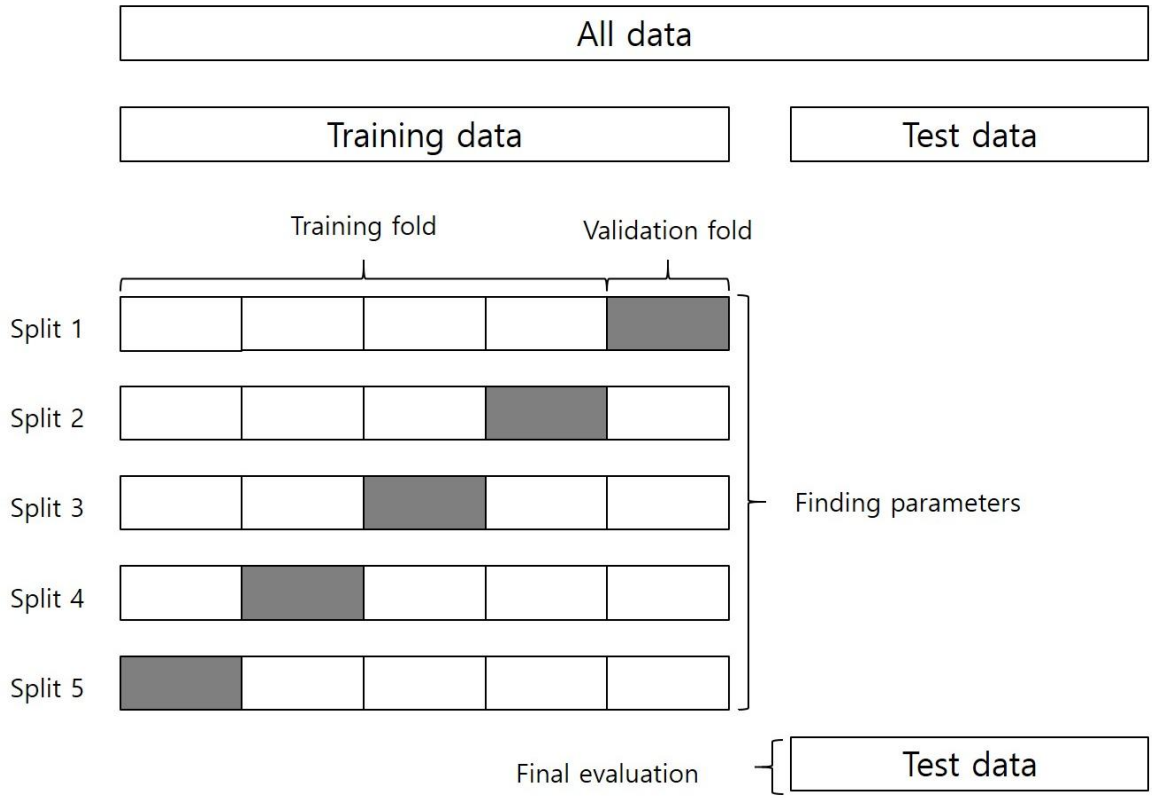
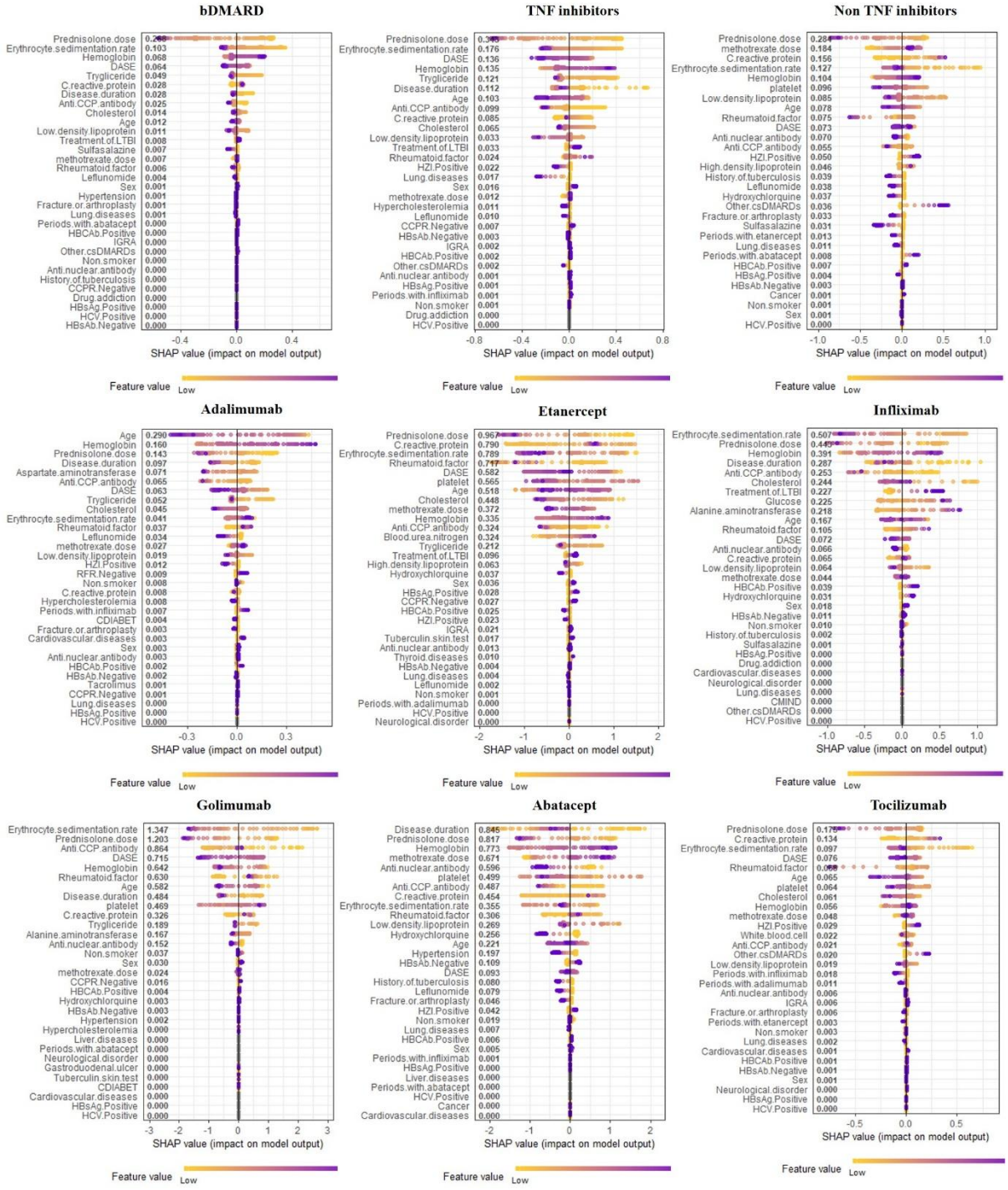


Fig. 5. Shapley plots for bDMARDs.



bDMARDs, biologics disease modifying anti-rheumatic drugs; TNF, tumor necrosis factor

Discussion

This study identified the best model to predict the response of biologics treatment for patients with RA, using clinical features in the outpatient setting. Machine learning models were built to enable the selection of the bDMARD that could best achieve remission. In all machine learning models, remission was estimated with an accuracy of approximately 57.2%–73.7%. In addition, important clinical features that explain the reason for prediction of remission or remission failure in the machine learning model were presented.

Recently, studies have attempted to predict the disease activity or treatment response of RA using machine learning.[19-21] Although these studies showed high accuracy in predictions, they did not explain the reasons for the predictions. In addition, genetic information which is difficult to obtain from patients in routine clinical practice was used. Although studies using big data should consider the implementation of findings in clinical practice, the feasibility and applicability remain low.[22] To compensate for these shortcomings, this study used information from a registry that was available in clinical practice and predicted remission at next follow-up after the start of treatment with each biologic. In addition, predictions, as well as the importance of variables, were presented, which was necessary for predictions to be understood by clinicians using explainable AI.

Norgeot et al. predicted the disease activity (controlled vs. uncontrolled) of RA patients at the next clinical visit using a longitudinal deep learning model and electronic health record data from a single cohort from a university hospital.[21] This deep learning model demonstrated excellent forecasting performance (AUROC, 0.91; 95% CI, 0.86–0.96); however, in a public safety-net hospital test cohort, this trained model showed an AUROC curve of 0.74 (95% CI, 0.65–0.83). Although the accuracy and AUROC curve were found to be excellent in training, the performance may degrade in datasets from other sources. Guan et al. used machine learning to predict TNF inhibitor responses in patients with RA using clinical and genetic markers.[19] They created Gaussian process regression model to predict changes in DAS28 and to classify them into either the responder or the non-responder group. They showed a correlation coefficient of 0.405 in DAS28 change and 78% accuracy in response during cross-validation tests. However genetic single-nucleotide polymorphism biomarkers showed a small contribution in the prediction using an independent data set. Therefore, this result showed that

the clinical features may be most predictive of treatment response.

In terms of interpretability in machine learning models, the prediction using most machine learning methods does not help in actual clinical practice because the reasons underpinning the predicted response is unknown. To overcome these limitations, explainable AI was applied to find variables that influence prediction and to provide more practical information than previous studies. In the Sharpley plot, the majority of the variables had a linear relationship with remission, and it was demonstrated that machine learning based on regression as well as deep learning was sufficient for prediction of remission.

Although a machine learning model such as deep learning was a "black-box" model that could not explain the reason for prediction, explainable AI refers to AI that provides reasons for prediction in a way that humans can understand.[15] Using this method, a model was built that best-predicted the effectiveness of bDMARDs in real world clinical practice, while identifying the characteristics of important variables necessary for the prediction. Although there were some differences among bDMARDs, approximately 15 common variables were found to be important in predicting remission. The most important feature was the prednisolone dose, which indicated that patients with high prednisolone dosage are farthest from remission after a year. Considering that biologics have glucocorticoid-sparing effects,[23, 24] this result showed that it may be difficult to control high disease activity with biologics in patients with high glucocorticoid requirements. In addition, a high ESR was an important feature for remission.[25, 26] However, in the case of adalimumab, a high ESR affected remission failure but had a low absolute SHAP value. Therefore, the same variable in each biologic can have different impact on remission.

Several laboratory findings were determined to be important features for remission in treatment with bDMARDs. RF and ACPA were important indicators in the prediction of remission and contributed to the determination of the direction of treatment in patients with RA.[27, 28] However, several studies have shown conflicting results on the relationship between RF and ACPA in response to TNF inhibitors.[29, 30] In this study, low levels of ACPA were associated with remission, except with tocilizumab. Furthermore, a high level of RF was associated with remission in adalimumab, infliximab, and golimumab; however, a low level of RF was associated with remission in etanercept and tocilizumab. In abatacept, the RF level

and remission were not linearly related to each other. Moreover, among the laboratory variables, one of the most important findings for remission in this study was hemoglobin. Considering anemia is associated with disease activity[31] and erosion progression,[32] hemoglobin may be an important factor indicating the disease state in patients with RA. Interestingly, comorbidities had little effect on remission or remission failure compared to laboratory findings.

However, in terms of the machine learning model, it is difficult to generalize each of these variables as an absolute predictor of remission. Since the prediction model was generated by combining various clinical variables, it is likely that these variables contributed to remission to varying degrees, rather than the absolute contribution of one variable. It would be feasible and applicable to predict remission with a combination of features using machine learning, as opposed to finding a generalized single biomarker in various manifestations of patients with RA. Considering that RA is a complex disease of interaction between environmental and genetic factors, this approach would be appropriate for predicting remission in patients treated with biologics.[33]

The established machine learning model was also tested on several types of hypothetical bDMARD treatments never used for patients who failed bDMARDs. As in the real world, some of the patients who failed with a certain bDMARD were predicted to have remission with treatment using other bDMARDs. In addition, patients who failed with TNF inhibitors had higher remission rates in the prediction model of non-TNF inhibitors, such as abatacept and tocilizumab, than the other TNF inhibitors. These results suggested that certain classes of biologics may be more effective in some patients; thus, prescribing biologics based on machine learning predictions may reduce patient pain and medical costs by reducing remission failure. However, in this prediction model, the remission rate in tocilizumab was higher than that in other bDMARDs, and it may be due to the overestimated treatment response based on DAS28-ESR of tocilizumab.[34, 35] Therefore, further development of this model is needed to improve accuracy and reliability.

This study had some limitations. First, the dosage intervals or doses of biologics were not considered. Second, this study did not distinguish between the primary response failure related to failure of clinical improvement and the secondary response failure related to loss of response

after clinical improvement.[30] Third, this study did not provide evidence as to how the important features of each biologic relates to their mechanism of action. Fourth, in these machine learning models, it was not possible to suggest a more effective bDMARD among several biologics that predicted remission. Fifth, the performance of this machine learning could not be confirmed with the test set of other RA cohort data. Therefore, it is necessary to check whether this machine learning model is applicable to other cohorts.

Conclusion

Explainable AI was able to identify some important clinical features in treatment with biologics. Various features of patients can affect remission, and there are differences in important features that affect remission between biologics. This study suggested that an advanced machine learning approach may support clinical decisions to improve treatment outcomes in biological therapy in patients with RA.

Reference

1. Smolen JS, Landewe RBM, Bijlsma JWJ, Burmester GR, Dougados M, Kerschbaumer A, et al. EULAR recommendations for the management of rheumatoid arthritis with synthetic and biological disease-modifying antirheumatic drugs: 2019 update. Ann Rheum Dis. 2020. doi: 10.1136/annrheumdis-2019-216655.
2. Lau CS, Chia F, Dans L, Harrison A, Hsieh TY, Jain R, et al. 2018 update of the APLAR recommendations for treatment of rheumatoid arthritis. Int J Rheum Dis. 2019;22(3):357-75.
3. Singh JA, Saag KG, Bridges SL, Jr., Akl EA, Bannuru RR, Sullivan MC, et al. 2015 American College of Rheumatology Guideline for the Treatment of Rheumatoid Arthritis. Arthritis Rheumatol. 2016;68(1):1-26.
4. Keystone EC, Kavanaugh AF, Sharp JT, Tannenbaum H, Hua Y, Teoh LS, et al. Radiographic, clinical, and functional outcomes of treatment with adalimumab (a human anti-tumor necrosis factor monoclonal antibody) in patients with active rheumatoid arthritis receiving concomitant methotrexate therapy: a randomized, placebo-controlled, 52-week trial. Arthritis Rheum. 2004;50(5):1400-11.
5. Weinblatt ME, Kremer JM, Bankhurst AD, Bulpitt KJ, Fleischmann RM, Fox RI, et al. A trial of etanercept, a recombinant tumor necrosis factor receptor:Fc fusion protein, in patients with rheumatoid arthritis receiving methotrexate. N Engl J Med. 1999;340(4):253-9.
6. Kearsley-Fleet L, Davies R, De Cock D, Watson KD, Lunt M, Buch MH, et al. Biologic refractory disease in rheumatoid arthritis: results from the British Society for Rheumatology Biologics Register for Rheumatoid Arthritis. Ann Rheum Dis. 2018;77(10):1405-12.
7. Kievit W, Adang EM, Fransen J, Kuper HH, van de Laar MA, Jansen TL, et al. The effectiveness and medication costs of three anti-tumour necrosis factor alpha agents in the treatment of rheumatoid arthritis from prospective clinical practice data. Ann Rheum Dis. 2008;67(9):1229-34.
8. Panch T, Szolovits P, Atun R. Artificial intelligence, machine learning and health systems. J Glob Health. 2018;8(2):020303.

9. Shin K, Choi SJ, Kwak S-K, Park Y-B, Sung Y-K, Choi C-B, et al. AB1066 Korean Biologics Registry of Patients with Systemic Rheumatic Disease (KOBIO): A Nationwide Registry to Assess Adverse Events Associated with Biologic Treatment in Korea. Annals of the Rheumatic Diseases. 2014;73(Suppl 2):1153-4.
10. Deane KD, Demoruelle MK, Kelmenson LB, Kuhn KA, Norris JM, Holers VM. Genetic and environmental risk factors for rheumatoid arthritis. Best Pract Res Clin Rheumatol. 2017;31(1):3-18.
11. Lundberg SM, Erion GG, Lee S-I. Consistent Individualized Feature Attribution for Tree Ensembles. arXiv e-prints [Internet]. 2018 February 01, 2018.[arXiv:1802.03888 p.]. Available from: https://ui.adsabs.harvard.edu/abs/2018arXiv180203888L.
12. Prevoo ML, van 't Hof MA, Kuper HH, van Leeuwen MA, van de Putte LB, van Riel PL. Modified disease activity scores that include twenty-eight-joint counts. Development and validation in a prospective longitudinal study of patients with rheumatoid arthritis. Arthritis Rheum. 1995;38(1):44-8.
13. van Gestel AM, Prevoo ML, van 't Hof MA, van Rijswijk MH, van de Putte LB, van Riel PL. Development and validation of the European League Against Rheumatism response criteria for rheumatoid arthritis. Comparison with the preliminary American College of Rheumatology and the World Health Organization/International League Against Rheumatism Criteria. Arthritis Rheum. 1996;39(1):34-40.
14. Tibshirani R. Regression Shrinkage and Selection via the Lasso. Journal of the Royal Statistical Society Series B (Methodological). 1996;58(1):267-88.
15. Suykens JAK, Vandewalle J. Least Squares Support Vector Machine Classifiers. Neural Process Lett. 1999;9(3):293-300.
16. Liaw A, Wiener M. Classification and Regression by RandomForest. Forest. 2001;23.
17. Chen T, Guestrin C. XGBoost: A Scalable Tree Boosting System. Proceedings of the 22nd ACM SIGKDD International Conference on Knowledge Discovery and Data Mining; San Francisco, California, USA: Association for Computing Machinery; 2016. p. 785–94.
18. Lundberg SM, Lee S-I. A Unified Approach to Interpreting Model Predictions. 2017.4765--74.

19. Guan Y, Zhang H, Quang D, Wang Z, Parker SCJ, Pappas DA, et al. Machine Learning to Predict Anti-Tumor Necrosis Factor Drug Responses of Rheumatoid Arthritis Patients by Integrating Clinical and Genetic Markers. Arthritis Rheumatol. 2019;71(12):1987-96.
20. Kim KJ, Kim M, Adamopoulos IE, Tagkopoulos I. Compendium of synovial signatures identifies pathologic characteristics for predicting treatment response in rheumatoid arthritis patients. Clin Immunol. 2019;202:1-10.
21. Norgeot B, Glicksberg BS, Trupin L, Lituiev D, Gianfrancesco M, Oskotsky B, et al. Assessment of a Deep Learning Model Based on Electronic Health Record Data to Forecast Clinical Outcomes in Patients With Rheumatoid Arthritis. JAMA Netw Open. 2019;2(3):e190606.
22. Gossec L, Kedra J, Servy H, Pandit A, Stones S, Berenbaum F, et al. EULAR points to consider for the use of big data in rheumatic and musculoskeletal diseases. Ann Rheum Dis. 2020;79(1):69-76.
23. Duquenne C, Wendling D, Sibilia J, Job-Deslandre C, Guillevin L, Benichou J, et al. Glucocorticoid-sparing effect of first-year anti-TNFalpha treatment in rheumatoid arthritis (CORPUS Cohort). Clin Exp Rheumatol. 2017;35(4):638-46.
24. Seror R, Dougados M, Gossec L. Glucocorticoid sparing effect of tumour necrosis factor alpha inhibitors in rheumatoid arthritis in real life practice. Clin Exp Rheumatol. 2009;27(5):807-13.
25. Tanaka Y, Hirata S, Kubo S, Fukuyo S, Hanami K, Sawamukai N, et al. Discontinuation of adalimumab after achieving remission in patients with established rheumatoid arthritis: 1-year outcome of the HONOR study. Ann Rheum Dis. 2015;74(2):389-95.
26. Vastesaeger N, Kutzbach AG, Amital H, Pavelka K, Lazaro MA, Moots RJ, et al. Prediction of remission and low disease activity in disease-modifying anti-rheumatic drug-refractory patients with rheumatoid arthritis treated with golimumab. Rheumatology (Oxford). 2016;55(8):1466-76.
27. Canhao H, Rodrigues AM, Mourao AF, Martins F, Santos MJ, Canas-Silva J, et al. Comparative effectiveness and predictors of response to tumour necrosis factor

inhibitor therapies in rheumatoid arthritis. Rheumatology (Oxford). 2012;51(11):2020-6.

28. Potter C, Hyrich KL, Tracey A, Lunt M, Plant D, Symmons DP, et al. Association of rheumatoid factor and anti-cyclic citrullinated peptide positivity, but not carriage of shared epitope or PTPN22 susceptibility variants, with anti-tumour necrosis factor response in rheumatoid arthritis. Ann Rheum Dis. 2009;68(1):69-74.

29. Cuppen BV, Welsing PM, Sprengers JJ, Bijlsma JW, Marijnissen AC, van Laar JM, et al. Personalized biological treatment for rheumatoid arthritis: a systematic review with a focus on clinical applicability. Rheumatology (Oxford). 2016;55(5):826-39.

30. Tak PP. A personalized medicine approach to biologic treatment of rheumatoid arthritis: a preliminary treatment algorithm. Rheumatology (Oxford). 2012;51(4):600-9.

31. Wilson A, Yu HT, Goodnough LT, Nissenson AR. Prevalence and outcomes of anemia in rheumatoid arthritis: a systematic review of the literature. Am J Med. 2004;116 Suppl 7A:50S-7S.

32. Moller B, Scherer A, Forger F, Villiger PM, Finckh A, Swiss Clinical Quality Management Program for Rheumatic D. Anaemia may add information to standardised disease activity assessment to predict radiographic damage in rheumatoid arthritis: a prospective cohort study. Ann Rheum Dis. 2014;73(4):691-6.

33. McInnes IB, Schett G. The pathogenesis of rheumatoid arthritis. N Engl J Med. 2011;365(23):2205-19.

34. Kawashiri SY, Kawakami A, Iwamoto N, Fujikawa K, Aramaki T, Tamai M, et al. Disease activity score 28 may overestimate the remission induction of rheumatoid arthritis patients treated with tocilizumab: comparison with the remission by the clinical disease activity index. Mod Rheumatol. 2011;21(4):365-9.

35. Smolen JS, Aletaha D. Interleukin-6 receptor inhibition with tocilizumab and attainment of disease remission in rheumatoid arthritis: the role of acute-phase reactants. Arthritis Rheum. 2011;63(1):43-52.

Abstract

Individualized prediction of remission for biologics using explainable artificial intelligence

Division of Rheumatology, Department of Internal Medicine, University of Ulsan College of Medicine, Asan Medical Center, Soul, Korea

Background: This study aimed to establish a machine learning model that can predict the response to biologic disease modifying anti-rheumatic drugs (bDMARDs) in patients with rheumatoid arthritis (RA). In addition, we also sought to identify important features in remission using explainable artificial intelligence (AI).

Methods: Follow-up data of 1,204 patients who started treatment with bDMARDs, such as etanercept, adalimumab, golimumab, infliximab, abatacept, and tocilizumab, from December 2012 to June 2019 in the Korean College of Rheumatology Biologics and Targeted Therapy Registry, were enrolled. We selected the best of the five response criteria as the outcome for the next follow-up (about a year later) with bDMARDs treatment. Machine learning methods, such as lasso, ridge, support vector machine, random forest, and XGBoost, were used for prediction of remission. For interpretability, the Shapley additive explanation (SHAP) value (impact on model output with positive or negative relationship) was used.

Results: In predicting remission, the accuracy and area under the receiver operating characteristic were 57.2%–74.5% and 0.547–0.747, respectively. The Shapley plot showed that the impact of the variables on predicting remission differed for each bDMARD. The most important features were age in adalimumab, prednisolone dose in etanercept and tocilizumab, erythrocyte sedimentation rate in infliximab and golimumab, and disease duration in abatacept, with mean SHAP values of –0.290, –0.967, –0.175, –0.507, –1.347, and –0.845, respectively.

Conclusions: Using machine learning models and explainable AI, we could predict remission and explain the clinical features of patients with remission for each bDMARD. This approach may help to improve treatment outcomes in patients with RA.

Appendix

Table 1. Definition of response criteria

Current DAS28	DAS28: Improvement vs baseline		
	>1.2	>0.6 to ≤1.2	≤0.6
≤3.2	Good response	Moderate response	No response
>3.2 to ≤5.1			
>5.1			

DAS28-ESR: disease activity score in 28 joints using erythrocyte sedimentation rate.

Table 2. Prediction of good response for biologic DMARD

	N	Measure	Baseline*	Lasso	Ridge	SVM	Random Forest	Xgboost
Biologic DMARD	1,397	Accuracy	56.7%	64.1%	64.1%	64.4%	58.1%	64.4%
		AUROC	0.500	0.678	0.68	0.674	0.67	0.679
TNF inhibitors	793	Accuracy	51.9%	62.9%	63.3%	63.7%	62.4%	62.9%
		AUROC	0.500	0.678	0.678	0.677	0.682	0.680
Non-TNF inhibitors	604	Accuracy	68.3%	70.3%	70.6%	70.6%	68.3%	68.3%
		AUROC	0.500	0.685	0.684	0.678	0.643	0.598
Abatacept	289	Accuracy	54.4%	66.7%	66.7%	64.9%	61.4%	57.9%
		AUROC	0.500	0.716	0.731	0.716	0.734	0.608
Adalimumab	220	Accuracy	52.3%	65.1%	65.1%	64.0%	61.6%	61.6%
		AUROC	0.500	0.686	0.694	0.682	0.68	0.647
Etanercept	122	Accuracy	50.0%	66.7%	66.7%	65.2%	60.6%	58.3%
		AUROC	0.500	0.728	0.728	0.709	0.693	0.612
Golimumab	162	Accuracy	50.0%	63.9%	66.7%	66.7%	66.7%	66.7%
		AUROC	0.500	0.682	0.704	0.698	0.726	0.736
Infliximab	194	Accuracy	56.2%	64.6%	66.7%	62.5%	58.3%	54.2%
		AUROC	0.500	0.689	0.708	0.66	0.659	0.529
Tocilizumab	410	Accuracy	74.6%	72.1%	73.0%	73.0%	74.6%	74.6%
		AUROC	0.500	0.613	0.617	0.605	0.608	0.551

* no information rate for bDMARDs

bDMARDs, biologics disease modifying anti-rheumatic drugs; SVM, support vector machine; TNF, Tumor necrosis factor inhibitors; AUROC, area under the receiver operating characteristic.

Table 3. Prediction of good response without increasing prednisolone dose for biologic DMARD

Biologics	N	Measure	Baseline*	Lasso	Ridge	SVM	Random Forest	Xgboost
All biologic	1,397	Accuracy	52.9%	63.6%	63.6%	63.4%	62.2%	62.7%
		AUROC	0.500	0.692	0.692	0.688	0.673	0.679
TNF inhibitors	793	Accuracy	54.4%	66.0%	66.5%	66.2%	63.5%	64.3%
		AUROC	0.500	0.705	0.707	0.708	0.673	0.696
Non-TNF inhibitors	604	Accuracy	62.8%	68.9%	68.3%	68.3%	67.8%	67.2%
		AUROC	0.500	0.690	0.693	0.684	0.680	0.665
Abatacept	289	Accuracy	52.6%	61.4%	63.2%	61.4%	62.3%	58.8%
		AUROC	0.500	0.665	0.678	0.657	0.660	0.615
Adalimumab	220	Accuracy	54.7%	65.1%	66.3%	65.1%	65.1%	64.0%
		AUROC	0.500	0.699	0.713	0.704	0.675	0.683
Etanercept	122	Accuracy	53.8%	66.2%	67.7%	66.2%	60.0%	61.5%
		AUROC	0.500	0.728	0.736	0.720	0.664	0.658
Golimumab	162	Accuracy	50.0%	66.7%	66.7%	66.7%	69.4%	66.7%
		AUROC	0.500	0.710	0.738	0.708	0.788	0.721
Infliximab	194	Accuracy	58.3%	64.6%	64.6%	62.5%	56.2%	57.3%
		AUROC	0.500	0.666	0.696	0.659	0.589	0.531
Tocilizumab	410	Accuracy	69.7%	72.1%	72.1%	73.0%	69.7%	71.3%
		AUROC	0.500	0.698	0.703	0.695	0.640	0.627

* no information rate for bDMARDs

bDMARDs, biologics disease modifying anti-rheumatic drugs; SVM, support vector machine; TNF, Tumor necrosis factor inhibitors; AUROC, area under the receiver operating characteristic.

Table 4. Prediction of moderate or good response for biologic DMARD

	N	Measure	Baseline*	Lasso	Ridge	SVM	Random Forest	Xgboost
All biologic	1,397	Accuracy	86.1%	86.4%	86.4%	86.1%	84.9%	86.1%
		AUROC	0.500	0.732	0.727	0.617	0.666	0.713
TNF inhibitors	793	Accuracy	82.3%	82.3%	82.3%	82.3%	81.4%	82.3%
		AUROC	0.500	0.700	0.704	0.687	0.66	0.669
Non-TNF inhibitors	604	Accuracy	91.7%	91.7%	91.7%	91.7%	91.7%	91.7%
		AUROC	0.500	0.789	0.754	0.749	0.709	0.731
Abatacept	289	Accuracy	87.7%	87.7%	87.7%	84.2%	87.7%	87.7%
		AUROC	0.500	0.761	0.810	0.766	0.676	0.650
Adalimumab	220	Accuracy	82.6%	81.4%	81.4%	77.9%	80.2%	82.6%
		AUROC	0.500	0.651	0.680	0.646	0.594	0.582
Etanercept	122	Accuracy	86.2%	84.6%	86.2%	83.1%	84.6%	86.2%
		AUROC	0.500	0.700	0.740	0.730	0.672	0.570
Golimumab	162	Accuracy	80.6%	80.6%	77.8%	72.2%	77.8%	80.6%
		AUROC	0.500	0.754	0.739	0.660	0.696	0.682
Infliximab	194	Accuracy	79.2%	79.2%	81.2%	77.1%	79.2%	79.2%
		AUROC	0.500	0.688	0.700	0.686	0.634	0.616
Tocilizumab	410	Accuracy	94.3%	94.3%	94.3%	92.6%	94.3%	94.3%
		AUROC	0.500	0.832	0.830	0.735	0.773	0.679

* no information rate for bDMARDs

bDMARDs, biologics disease modifying anti-rheumatic drugs; SVM, support vector machine; TNF, Tumor necrosis factor inhibitors; AUROC, area under the receiver operating characteristic.

Table 5. Prediction of low disease activity for biologic DMARD

	N	Measure	Baseline*	Lasso	Ridge	SVM	Random Forest	Xgboost
All Biologic	1,397	Accuracy	58.9%	65.6%	65.6%	65.6%	59.3%	64.4%
		AUROC	0.500	0.688	0.690	0.686	0.663	0.676
TNF inhibitors	793	Accuracy	50.6%	65.6%	66.5%	65.8%	64.6%	62.9%
		AUROC	0.500	0.708	0.712	0.708	0.703	0.682
Non-TNF inhibitors	604	Accuracy	70.0%	70.6%	71.1%	70.6%	70.0%	70.0%
		AUROC	0.500	0.674	0.681	0.669	0.650	0.611
Abatacept	289	Accuracy	56.1%	64.9%	66.7%	63.2%	59.6%	56.1%
		AUROC	0.500	0.684	0.706	0.673	0.701	0.500
Adalimumab	220	Accuracy	50.0%	67.4%	68.6%	66.3%	65.1%	59.3%
		AUROC	0.500	0.720	0.736	0.716	0.697	0.625
Etanercept	122	Accuracy	53.8%	67.7%	67.7%	66.9%	61.5%	53.8%
		AUROC	0.500	0.756	0.750	0.727	0.726	0.537
Golimumab	162	Accuracy	54.3%	65.7%	68.6%	62.9%	65.7%	57.1%
		AUROC	0.500	0.717	0.737	0.701	0.789	0.677
Infliximab	194	Accuracy	55.3%	66.0%	67.0%	63.8%	63.8%	55.3%
		AUROC	0.500	0.706	0.739	0.691	0.690	0.500
Tocilizumab	410	Accuracy	76.2%	73.4%	75.4%	74.6%	76.2%	76.2%
		AUROC	0.500	0.630	0.640	0.613	0.620	0.500

* no information rate for bDMARDs

bDMARDs, biologics disease modifying anti-rheumatic drugs; SVM, support vector machine; TNF, Tumor necrosis factor inhibitors; AUROC, area under the receiver operating characteristic.