Heliyon 10 (2024) e22925

Contents lists available at ScienceDirect

Heliyon

journal homepage: www.cell.com/heliyon

Predicting anti-TNF treatment response in rheumatoid arthritis: An artificial intelligence-driven model using cytokine profile and routine clinical practice parameters

Juan Luis Valdivieso Shephard [a,*], Enrique Josue Alvarez Robles [b], Carmen Cámara Hijón [c], Borja Hernandez Breijo [d], Marta Novella-Navarro [e], Patricia Bogas Schay [e], Ricardo Cuesta de la Cámara [c], Alejandro Balsa Criado [e], Eduardo López Granados [c], Chamaida Plasencia Rodríguez [e]

[a] *Immunology Unit, Hospital Universitario La Paz-Idipaz, 28046, Madrid, Spain*
[b] *Spaik Technologies, S.L. Madrid, Spain*
[c] *Immunology Unit, Hospital Universitario La Paz-Idipaz, Madrid, Spain*
[d] *Immunology-Rheumatology Research Group, Hospital Universitario La Paz-Idipaz, Madrid, Spain*
[e] *Rheumatology Service, Hospital Universitario La Paz-Idipaz, Madrid, Spain*

ARTICLE INFO

Keywords:
Rheumatoid arthritis
Cytokines
Artificial intelligence
Anti-TNF response prediction
Interleukin 2
Interleukin 6

ABSTRACT

Introduction: Rheumatoid arthritis (RA) is a heterogeneous disease in which therapeutic strategies used have evolved dramatically. Despite significant progress in treatment strategies such as the development of anti-TNF drugs, it is still not possible to differentiate those patients who will respond from who will not. This can lead to effective-treatment delays and unnecessary costs. The aim of this study was to utilize a profile of the patient's characteristics, clinical parameters, immune status (cytokine profile) and artificial intelligence to assess the feasibility of developing a tool that could allow us to predict which patients will respond to treatment with anti-TNF drugs.
Methods: This study included 38 patients with RA from the RA-Paz cohort. Clinical activity was measured at baseline and after 6 months of treatment. The cytokines measured before the start of anti-TNF treatment were IL-1, IL-12, IL-10, IL-2, IL-4, IFNg, TNFa, and IL-6. Statistical analyses were performed using the Wilcoxon-Rank-Sum Test and the Benjamini-Hochberg method. The predictive model viability was explored using the 5-fold cross-validation scheme in order to train the logistic regression models.
Results: Statistically significant differences were found in parameters such as IL-6, IL-2, CRP and DAS-ESR. The predictive model performed to an acceptable level in correctly classifying patients (ROC-AUC 0.804167 to 0.891667), suggesting that it would be possible to develop a clinical classification tool.
Conclusions: Using a combination of parameters such as IL-6, IL-2, CRP and DAS-ESR, it was possible to develop a predictive model that can acceptably discriminate between remitters and

* Corresponding author.
E-mail addresses: juanluis.valdivieso@salud.madrid.org (J.L. Valdivieso Shephard), ejarkm@gmail.com (E.J. Alvarez Robles), ccamarah@salud.madrid.org (C. Cámara Hijón), borjahdezbreijo@gmail.com (B. Hernandez Breijo), marta.novella@salud.madrid.org (M. Novella-Navarro), patricia.bogas.sc@gmail.corec (P. Bogas Schay), ricardo.cuesta@salud.madrid.org (R. Cuesta de la Cámara), alejandro.balsa@salud.madrid.org (A. Balsa Criado), elgranados@salud.madrid.org (E. López Granados), chamiplaro@gmail.com (C. Plasencia Rodríguez).

https://doi.org/10.1016/j.heliyon.2023.e22925
Received 27 June 2023; Received in revised form 18 November 2023; Accepted 22 November 2023
Available online 28 November 2023

non-remitters. However, this model needs to be replicated in a larger cohort to confirm these findings.

1. Introduction

Rheumatoid arthritis (RA) is considered a chronic and heterogeneous disease characterized by immunological abnormalities, systemic inflammation, and persistent synovitis [1]. In recent years, it has become clear that the disease is much more common than previously thought, with a global prevalence of 0.46%–1% [2]. Delays in therapy initiation and/or poor controls can ultimately result in joint damage, disabilities, decreased quality of life, and cardiovascular and/or other comorbidities [1,3]. The exact pathogenesis of RA is not fully understood, though one multifactorial theory that the interactions between the environment and genes result in a tolerance loss for self-antigens. Inflammatory cells, including lymphocytes, and secreted inflammatory products such as cytokines, notably TNF-a, IL-6, IL-2, IL-17 and GM-CSF, among others, contribute to the development of RA [4–6]. While in the early stages, inflammation affects mainly the joints, the persistence of this process via a positive loop results in RA becoming a systemic disorder [1].

The therapeutic strategies used in patients with RA have evolved significantly over the years, from the use of non-steroid anti-inflammatory drugs and corticoid steroids to biologic or target specific disease modifying anti-rheumatic drugs (b/tsDMARDs) [5]. According to EULAR recommendations, there is a clear trend towards stricter control of disease activity in order to achieve remission or low disease activity as quickly as possible ("treat to target" strategy) [5,6]. However, despite the large therapeutic arsenal available for treating patients with RA, approximately 20–40 % of patients do not respond to the first administered b/tsDMARD and less than 50 % achieve remission [7,8]. In addition, these drugs represent a significant cost to the health system and are not without adverse effects. Thus, it would be very useful if clinicians had the ability to identify the best target for each patient.

All of these data highlight the need to identify biomarkers to predict clinical response at the start of treatment, which would greatly aid therapeutic decisions. Several biomarkers have been proposed as predictors of good clinical response in patients with RA [9,10]. However, no robust biomarkers are currently available to predict the responses to different types of DMARDs [11,12]. Some international and national groups are also focusing on finding predictors and/or biomarkers to therapy response in order to develop personalized medicine for patients with RA.

On the other hand, in recent years the integration of artificial intelligence (AI) into the medical field has marked a transformative shift in patient care, diagnostic and therapeutic interventions. AI's ability to recognize patterns and derive insights from data is a key tool for improving precision medicine. Particularly in chronic diseases such as RA, where patient responses to treatments such as TNF inhibitors (anti-TNF) are heterogeneous, AI can illuminate nuanced patterns in data that would otherwise be missed by traditional analysis.

Given the recent advances in technology and the current heterogeneous evidence, the main aim of this study is to explore the utility of using serum cytokines values and clinical practice parameters to predict therapy response in patients with RA treated with anti-TNF drugs.

2. Patients and methods

This is an observational retrospective study that included 38 patients with RA from the RA-Paz cohort [13]. Patients were diagnosed by the treating rheumatologist and fulfilled ACR/EULAR 2010 classification criteria [14] (Fig. 1). All included patients met the following inclusion criteria: ≥18 years of age; had one positive result for positive rheumatoid factor (FR) or anti-citrullinated peptide

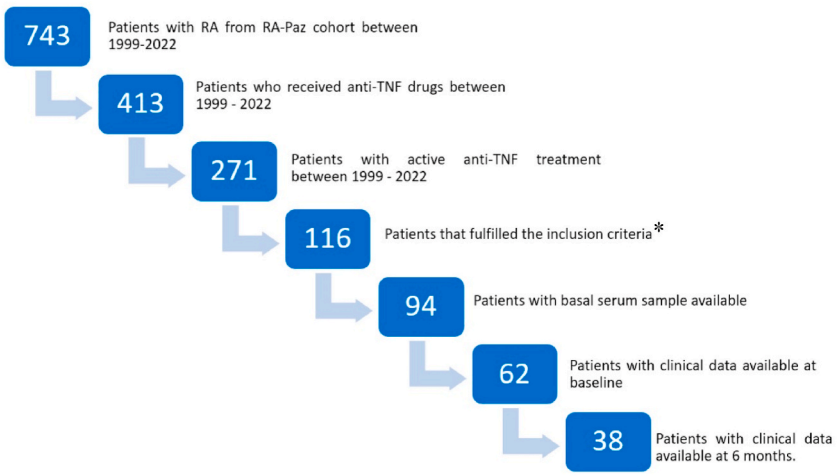

Fig. 1. Patients' selection flowchart. * Positive for FR or ACPA, use of DMARD, first anti-TNF treatment and DAS28-ESR ≥3.2.

antibodies (ACPA); prior treatment with methotrexate or another disease-modifying antirheumatic drugs; no previous bDMARD (anti-TNF) treatment and presented active disease activity as defined by the Disease Activity Score in the 28 joints-Erythrocyte Sedimentation Rate (DAS28-ESR) $\geq$3.2 at the start of treatment. Exclusion criteria were: any absence of clinical data or serum samples at baseline (prior to treatment initiation) and at 6 months. Each patient signed a written consent before enrollment.

Demographic, clinical, and serological variables were collected from the electronic medical records and the database of the Complex Therapies Unit at the La Paz University Hospital's Rheumatology Service. Clinical activity was measured by DAS28-ESR at baseline (before starting the anti-TNF) and after 6 months. At the 6-month visit, patients were stratified according to therapy response as follows: I) remitters (R): patients with a DAS28-ESR $\leq$2.6 and; II) non-remitters (non-R): patients with DAS28-ESR >2.6.

2.1. Methods

2.1.1. Rheumatoid factor and anti–citrullinated protein antibody determination

Serum samples were collected at time of diagnosis and prior to anti-TNF initiation in order to perform other determinations as part of routine laboratory practice. Rheumatological factor (RF) concentrations were measured by nephelometry using the BNII System and N Latex RF Kit (Siemens Healthcare Diagnostics Inc., Newark, NJ, USA) according to the instruction manual. Serum anti–citrullinated protein antibody was determined by ELISA as anti-CCP2 using the ImmunoscanCCPlus kit (Svar, Malmö, Sweden).

2.1.2. Cytokine assay

Cytokine determination was performed by immunoassays in microfluidic simple plex cartridges using the Ella system (Bio-Techne, Abingdon, United Kingdom). The cytokines measured were as follows: IL-1, IL-12, IL-10, IL-2, IL-4, IFNg, TNFa, and IL-6; regrettably, IL-17 and GM-CSF could not be analyzed due to their unavailability from the provider. The selection of the cytokine panel, which encompassed both pro-inflammatory and immunomodulatory cytokines, was determined by their influence on the pathogenesis of the disease, their potential influence on the response to anti-TNF treatment and the provider's availability during the research. The samples used were comprised of patient serum prior to initiation of anti-TNF treatment. The cytokine assay results are shown in Table 2.

2.1.3. Statistical analysis

Analyses were performed using the Wilcoxon Rank Sum Test (P VALUE). To demonstrate that the significance between the different outcomes was not due to chance, the correction of multiple comparisons using the Benjamini-Hochberg (BH) method with a predefined value of FDR = 0.1 was applied (P VALUE ADJ) (free software used MEV2.0).

2.1.4. Viability assessment for predictive models

The potential use of a machine learning predictive model to estimate the probability of a patient to be a remitter or a non-remitter was evaluated using an iterated cross-validation schema.

In order to construct the machine learning model, in which weights are adjusted iteratively until the model converges on a solution that minimizes error, the data was partitioned into training (80 %) and validation (20 %) sets. Given the limited number of samples, employing a repeated 5-fold cross-validation scheme became the more robust method for training the logistic regression models. In this

Table 1

Baseline characteristics of patients included in the study. This table shows the mean ± SD, median (IQR) or absolute number (percentage) for all patients included. RF, rheumatoid factor; ACPA, anti-citrullinated protein antibodies; DAS28-ESR, Disease Activity Score-28, Erythrocyte Sedimentation Rate; csDMARD, conventional synthetic disease-modifying anti-rheumatic drug; MTX, methotrexate; OD, other conventional synthetic disease-modifying anti-rheumatic drugs than methotrexate; CRP: C-reactive protein; ESR: Erythrocyte Sedimentation Rate. The differences between groups were analyzed considering a p-value <0.05 as a statistically significant result.

Baseline patients' characteristics	Total patients (n = 38)	Non-remitters (n = 20; 53 %)	Remitters (n = 18; 47 %)	p-value
Age (years)	54 ± 13	57 ± 14	52 ± 11	0.2
Female	30 (78 %)	15 (75 %)	15 (83 %)	0.5
Disease duration (years)	7 (3–10)	8 (3–13)	6 (3–10)	0.5
RF positive	33 (87 %)	17 (85 %)	26 (89 %)	0.7
RF max titers	167(86.7–326)	191(113.3–536)	131(56.5–196)	0.03
ACPA positive	31 (81.5 %)	15 (75 %)	15 (83.3 %)	0.2
ACPA max titer	525 (141–1624)	485.5(84.3–843.7)	708(142.5–1732)	0.2
Smoking habit				0.492
Non-smokers	17 (45 %)	10 (50 %)	7 (39 %)	
Smoker	21 (55 %)	10 (50 %)	11 (61 %)	
Body mass index (kg/m2)	25.9 ± 4.9	27.0 ± 4.4	24.6 ± 5.2	0.06
DAS28-ESR	4.9 ± 1.1	5.5 ± 0.9	4.2 ± 0.8	<0.0001
CRP	6.9(1.81–28.2)	15(5.9–35.6)	2.4(0.98–7.23)	0.003
ESR	28(10.7–41)	40.5(33.2–53)	12.5(8.5–24.25)	0.000
Concomitant csDMARDs	37 (97 %)	20 (100 %)	17 (94 %)	0.2
MTX [±OD]	30 (79 %)	18 (90 %)	12 (67 %)	0.07
Only OD	7 (18 %)	2 (10 %)	5 (28 %)	0.1
Prednisone use	15 (39 %)	9 (45 %)	6 (33 %)	0.4

Table 2
Baseline cytokine results in patients classified as remitters and non-remitters after 6 months of anti-TNF therapy. Section A: Remitters. RP: Remitter patients (number). **Section B:** Non-remitters. NRP: Non-Remitter patients (number). A red scale was used to indicate the relationship. between values in the same column..

Section A	IL-1	IL-12	IL-10	IL-2	IL-4	IFNg	TNFa	IL-6
RP1	0,842	3,45	3,18	1,09	0,721	0,875	6,64	3,42
RP2	1,49	1,04	2,89	0,764	0,756	14,4	5,7	15,7
RP3	0,354	0,0919	4,41	0,61	0,382	0,611	10,9	1,63
RP4	0,784	0,771	1,85	0,637	0,34	0,334	5,2	1,22
RP5	1	1,37	2,92	0,877	0,321	0,9	4,39	2,58
RP6	1,25	3,68	3,53	1,14	0,028	2,11	9,83	104
RP7	1,65	1,71	3,09	0,943	0,402	1,18	6,92	6,81
RP8	0,903	2,27	2,55	0,867	0,351	0,978	8,43	3,16
RP9	0,82	1,07	2,8	0,711	0,303	2,04	7,54	2,91
RP10	0,781	0,232	1,81	0,419	0,154	1,77	9,19	5,45
RP11	1,02	1,52	10,1	0,482	0	7,26	2,91	10,7
RP12	2,07	0,348	1,71	0,39	0,215	1,12	13,3	3,24
RP13	3,84	0,732	2,33	1,25	0,364	1,59	7,2	39,6
RP14	5,78	1,54	1,73	0,685	0,867	0,247	0,449	2,43
RP15	0,866	0,191	7,91	0,732	0,318	0,424	5,04	1,36
RP16	0,246	0,596	2,52	0,803	0,365	1,38	5,19	3,44
RP17	0,944	1,14	2,4	0,815	0,201	1,34	26,6	151
RP18	1,45	2,03	2,56	0,646	0,292	0,756	9,53	3,67
Median	0,972	1,105	2,68	0,748	0,3305	1,15	7,06	3,43

Section B	IL-1	IL-12	IL-10	IL-2	IL-4	IFNg	TNFa	IL-6
NRP1	0,743	1,42	3,24	0,777	0,906	1,1	4,91	4,96
NRP2	0,755	1,15	2,97	1,28	5,97	4,4	6,6	21,1
NRP3	1,07	0,613	2,85	0,397	1,47	1,43	7,33	77,9
NRP4	36	0,945	1,85	0,503	0,063	0,888	10,5	118
NRP5	0,23	0,342	1,28	0,048	0,129	0,348	4,69	19,3
NRP6	1,48	6,09	6,28	0,7	0,713	2,34	15,4	31,9
NRP7	1,79	0,498	3,71	0,238	0,258	0,277	7,02	14,2
NRP8	0,863	0,552	2,52	0,424	0,278	0,661	9,03	6,32
NRP9	2,49	1,2	1,83	0,431	0,133	4,36	7,31	15,7
NRP10	0,909	0,482	2,05	0,49	0,328	0,279	4,97	8,91
NRP11	0,593	0,277	3,07	0,339	0,289	1,11	8,54	29,1
NRP12	0,924	1,24	3,06	1,15	4,23	11,5	11	15,3
NRP13	0,408	0,533	1,92	0,538	0,219	0,543	6,8	10,1
NRP14	0,833	0,544	2,8	1,08	1,73	6,1	6,57	15,2
NRP15	0,652	0,282	2,35	0,308	0,584	1,71	9,21	6,95
NRP16	0,197	0,737	1,72	0,101	0,047	1,35	5,59	1,95
NRP17	0,593	0,3	6,41	0,499	2,42	2,02	21	5,96
NRP18	0,263	0,03	2,23	0,181	0,067	3,49	7,27	13,6
NRP19	3,88	3,22	6,34	0,261	3,12	2,51	6,45	40,4
NRP20	0	1,75	7,26	2,48	0,201	1,21	2,52	7,55
Median	0,794	0,582	2,82	0,46	0,30	1,39	7,14	14,7

technique, the dataset is randomly divided into 5 equal-sized subsets. Among these, one subset serves as the validation set while the remaining subsets are used for training. This ensures that every data point has a chance to be in the validation set, thus maximizing the use of all available data.

Since the key feature of repeated k-fold cross-validation is its repetition, the 5-fold cross-validation process was reiterated 1000 times, with different data splits in each cycle. This ensured that the model's performance was not contingent upon a specific random data split. Essentially, it provided a more comprehensive assessment of the model's stability and generalizability.

Utilizing the repeated 5-fold cross-validation (Fig. 2), we were able to compute a confidence interval for each of the model evaluation metrics, including ROC-AUC, Accuracy, Precision, Recall, and F1. These metrics constituted the preferred approach for evaluating the performance of the classification models. They not only provided a measure of the model's central tendency, but also provided insights into its variability, presenting a more comprehensive perspective on the model's robustness and reliability. To be more specific, the Receiver Operating Characteristic Area Under the Curve (ROC-AUC) stood out as a foundational metric that provided a thorough evaluation of a model's capacity to differentiate between positive and negative instances across various classification thresholds. This metric was incorporated to assess the model's overall discriminative prowess, which held particular significance when appraising the model's performance across an extensive array of threshold selections.

Accuracy, a widely acknowledged metric, denotes the proportion of accurately classified instances out of the total instances. Its inclusion is imperative in furnishing a general overview of the model's correctness in predicting both positive and negative classes, especially when class distribution maintains equilibrium. With respect to precision, this metric gauges the proportion of accurately predicted positive instances among all instances forecasted as positive. Precision assumes relevance when the cost of false positives is considerable, serving as an indicator of the model's capacity to limit erroneous positive predictions.

Recall quantifies the proportion of accurately predicted positive instances among all actual positive instances. Its significance comes to the forefront when the cost of false negatives is substantial, as it measures the model's efficacy in capturing a substantial portion of positive instances. The F1 Score, a harmonic mean of precision and recall, furnishes a well-rounded assessment that takes into account both false positives and false negatives. This metric is crucial in studies that seek to strike a balance between precision and recall, thereby gaining insights into the model's comprehensive performance while navigating the trade-offs between different error types.

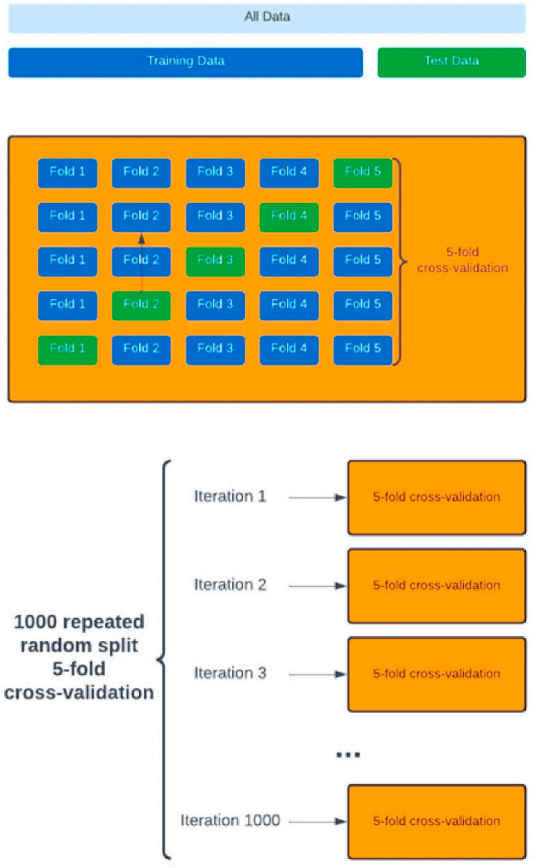

Fig. 2. 5-fold cross-validation scheme.

3. Results

3.1. Patient characteristics

All of the baseline demographic, clinical and serological characteristics of patients are shown in Table 1. Remitters (47 %) presented lower RF titers, BMI and basal clinical activity parameters (DAS28-ESR, CRP and ESR) compared with non-remitters. Most patients were treated concomitantly with MTX as a conventional DMARD.

3.2. Cytokines levels according to clinical response

As can be seen in Table 2, cytokine values were very heterogeneous. Interestingly, the median value for basal IL-6 was higher for non-remitters, while that of basal IL-2 was higher for remitters. Thus, it makes sense to study their interactions.

All study variables were analyzed with the BH method to discriminate any associations that led to remission not due to chance. After performing this statistical test, IL-6, IL-2, CRP and DAS-ESR remained statistically significant in terms of achieving remission (Table 3 and Fig. 3). Surprisingly, the IL-6/IL-2 ratio (IL-6/IL-2R) proved more significant than the results recorded for each cytokine alone.

3.3. Predictive models result

To gauge the performance of the predictive model, only variables that were found to be statistically significative and of practical interest were used (IL-6, IL-2, CRP, DAS28-ESR). The resulting metrics were as follows: ROC-AUC 0.804167 to 0.891667; Accuracy 0.789286 to 0.896429; Precision 0.833000 to 1.000; Recall 0.7 to 0.818333 and F1 0.74981 to 0.86666 (Fig. 4 shows the ROC-AUC for the predictive model).

Despite the wide range obtained, possibly due to the limited number of subjects, the range of possible values in the ROC-AUC (0.804167–0.891667) shows that the construction of a machine learning model could achieve good performance and boast high predictive power. Even using the most pessimistic value of the interval (0.804167), the model would still have acceptable discriminatory power (Table 4).

At the following link (https://spaik-technologies-streamlit-apps-srcapp-cggs3m.streamlit.app/PRERA%20calculator) it is possible to access the response prediction simulation generated from the data obtained in this study, which utilized using IL-6, IL-2, DAS28-ESR and CRP values.

4. Discussion

This study explored the capacity of a specific pre-treatment cytokine profile and the use of artificial intelligence to predict the response to anti-TNF drugs in patients with RA. The main purpose of this work was to assess whether cytokine assessment in tandem with and AI could help develop a tool that can be implemented in real clinical practice. In addition, our study also found that some characteristics between remitters and non-remitters differ prior to starting an anti-TNF treatment. Although anti-TNF therapy has been the cornerstone of biological treatment for patients with RA, currently there are many other therapeutic options with different mechanism of action. Thus, the ability to differentiate those patients who had a greater likelihood to achieve remission would be very

Table 3
Statistically significant variables.

	P-VALUE	P VALUE ADJ
DAS-ESR	2,85E-04	0,006248809
CRP	0,002602	0,02421764
IL-6/IL-2 R	0,002602	0,02421764
IL-6	0,008878	0,04811428
IL-2	0,022587	0,075878344

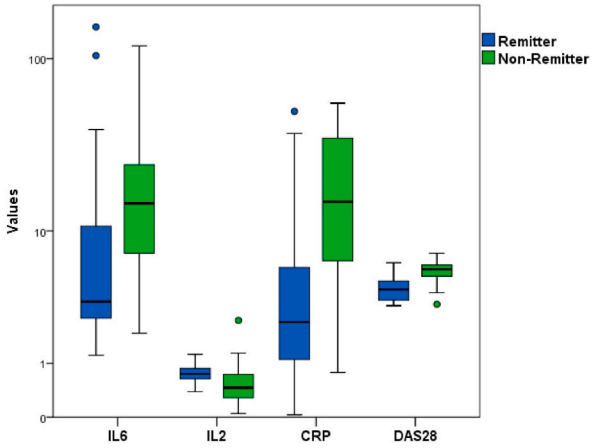

Fig. 3. Boxplot indicating differences between remitters and non-remitters for significant variables.

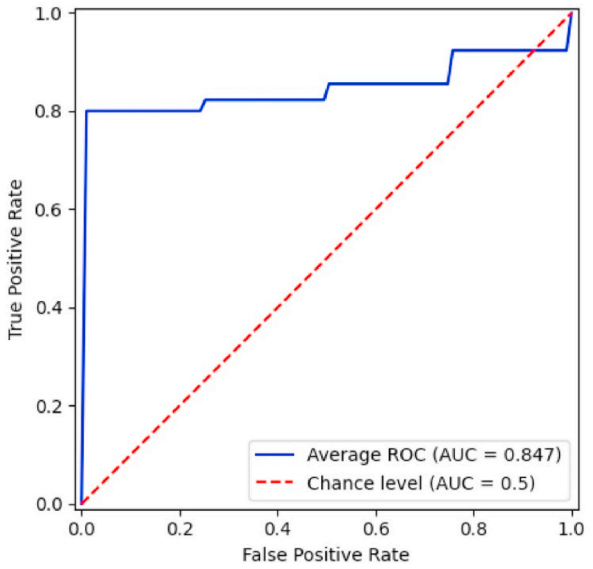

Fig. 4. ROC-AUC for remission predictive model in patients with RA prior to start of anti-TNF therapy. This model has been carried out using the following variables: DAS28-ESR, CRP, IL-6, and IL-2R.

helpful to physicians, enabling them to personalize and enhance their patients' treatments. Therefore, it is crucial to focus research efforts on the development of response predictors.

In our cohort, the remitter group showed statistically significative differences compared to the non-remitter group in such values as baseline DAS28-ESR, CRP, ESR, IL-2, and IL-6. Common exploratory methods such as DAS28-ESR, CRP or ESR are widely used to assess therapy response. In our study these three variables were found at higher levels in non-remitters patients prior to start of anti-TNF. In concordance with our results, Novella et al. observed that patients with higher DAS28-ESR at baseline were more likely to experience biologic disease-modifying antirheumatic drugs (bDMARDs) failure. As Chandrashekara et al. has described, while the DAS28 Index has been used successfully in clinical practice to assess the current state of RA, it is important to note that this index is composed of several variables some of which are subjective and dependent upon patient perception or limited due to the number of articulations. Regarding the CRP and ESR parameters, unless their rise can be linked to increased AR activity, both are highly unspecific and could be affected by other circumstances. These include obesity and physical activity in the case of CRP and gender or age in the case ESR. In addition, other non-specific situations such as infections or traumas can affect both parameters. Therefore, although these parameters are very useful and widely applied in clinical practice, their role as predictors of response to treatment remains unclear [13,15].

As regards others reports, one recent publication described the biding between rapid radiological progression and treatment response with baseline CRP. However, as its predictive role was not consistent, further studies are needed it to explore this association [16].

Concerning cytokine assessment, our non-remitter cohort showed lower levels of IL-2 compared to the remitters. It's important to

Table 4
Viability metrics.

Metric	Lower bound	Upper bound
ROC-AUC	0.804167	0.891667
Accuracy	0.789286	0.896429
Precision	0.833000	1.000
Recall	0.7	0.818333
F1	0.74981	0.86666

note that in recent years IL-2 has been described as possessing a pleotropic function. This suggests that IL-2, depending on its serum level, is important for activation or immune regulation due to its capacity to stimulate the proliferation of regulator lymphocytes (Treg) at a certain sera threshold. This discovery has been of such importance that low doses of IL-2 as RA therapy are currently being studied. In another study of great relevance, anti-bodies against IL-2 were found to be present in a large percentage of RA patients. This supports the contention that, in some cases, abnormally low levels of IL-2 can be a determining factor in RA pathogeny [17,18]. Following this thought line, we hypothesized that the difference in baseline IL-2 levels between remitters and non-remitters could stem from a different pathogenic mechanism, resulting in different outcomes, and that each group could thus benefit from different treatments. Moving forward, in our cohort higher levels of IL-6 were found in non-remitter patients. The pathogenic role of IL-6 in RA is widely known, as is its correlation with disease activity and that its reduction after star of treatment could correlate with patient prognosis [19]. However, its role as a response predictor was mostly studied vis-à-vis IL-6 pathway acting drugs. Recent studies have shown that patients with higher levels of IL-6 are those most likely to benefit from sarilumab (anti-IL-6 receptor) treatment rather than anti-TNF (adalimumab) [20]. Intriguingly, some studies have published values opposite from our own, e.g., higher values of IL-6 in remitter patients [21]. However, some important differences in the criteria for classifying a patient as a remitter were incorporated in our design to ensure a more homogeneous population. These included the necessity of seropositivity or previous use of csDMARDs, among others. These considerable design discrepancies make inter-study comparisons quite difficult.

Given the fact that the cytokine results for IL-2 and IL-6 were statistically significative but inversed between remitters and non-remitters, we hypothesized that the relationship between them, specifically their ratio, could be an interesting area of study. Surprisingly, the results from a cytokine cumulative analysis were statistically more significative than that for each cytokine alone. This approach could have important clinical implication, as is mentioned above. Taking into account the pleotropic characteristics of IL-2 and its regulatory functions, our method investigated the various relationships between both inflammation and regulation factors.

The use of artificial intelligence and machine learning could be considered groundbreaking in clinical practice, increasing its relevance and impact in the field as is readily apparent by the steady stream of recent publications [22–24]. The arrival of this technology in rheumatology, and more specifically in rheumatoid arthritis, is already having considerable impact [25]. Therefore, our study explores the use of this technology in conjunction with clinical parameters and cytokine profiles, a dynamic that thus far has not been properly studied. However, while there are many approaches for developing a predictive model, we chose the option best suited to our data characteristics and one that it is in widespread use across the field [26–29].

Although this technology opens a whole new spectrum of possibilities, it has some limitations. In our case involves the small sample size that resulted strict inclusion criteria in order to obtain the most homogenized population. Nevertheless, a repeated 5-fold cross-validation scheme to train the logistic regression models was performed, taking into account the limited sample size, thus ensuring an acceptable ROC curve was obtained.

In summary, we show how the use of machine learning can generate a predictive model that incorporates clinical and analytical parameters associated with the response to anti-TNF treatment. In addition, our study provides readers with a free simulation tool, affording a glimpse into the potential benefits of this approach. While our findings are encouraging, this model must be validated across a broader patient cohort to solidify the robustness of these outcomes.

Key messages

- The prediction of response to treatment with biologic drugs is essential to optimize time and resources.
- Interleukins 2 and 6 appear to have great potential for identifying those patients who will respond.
- This study provides a web application for predicting response to anti-TNF treatment and demonstrates the feasibility of using these resources.

Funding

No specific funding was received from any bodies in the public, commercial or not-for-profit sectors to carry out the work described in this article.

Ethical approval information, institution(s) and number(s)

This study conforms to the Declaration of Helsinki, the Ethics Committee of the University Hospital of La Paz approved the research, and written informed consent was obtained from the subjects.

Data availability statement

No data from this study is deposited in a public repository. Relevant data is contained within the article or cited therein.

CRediT authorship contribution statement

Juan Luis Valdivieso Shephard: Conceptualization, Data curation, Formal analysis, Investigation, Methodology, Writing – original draft, Writing – review & editing, Validation. **Enrique Josue Alvarez Robles:** Methodology, Software, Validation, Visualization. **Carmen Cámara Hijón:** Conceptualization, Investigation. **Borja Hernandez Breijo:** Data curation, Investigation, Writing – original draft. **Marta Novella-Navarro:** Conceptualization, Investigation, Software, Writing – original draft, Writing – review & editing. **Patricia Bogas Schay:** Data curation, Investigation. **Ricardo Cuesta de la Cámara:** Methodology. **Alejandro Balsa Criado:** Funding acquisition, Investigation, Methodology, Project administration, Resources. **Eduardo López Granados:** Funding acquisition, Investigation, Project administration, Resources. **Chamaida Plasencia Rodríguez:** Conceptualization, Funding acquisition, Investigation, Methodology, Resources, Visualization, Writing – review & editing.

Declaration of competing interest

The authors declare that they have no known competing financial interests or personal relationships that could have appeared to influence the work reported in this paper.

Acknowledgements

We would like to express our gratitude to all the patients for their strength and to the Rheumatology and Immunology services of the Hospital Universitario La Paz for making this work possible. Special thanks to the HULP Biostatistics Service, with a particular mention to Mariana Diaz-Almirón for her invaluable support with database management and data analysis. Additionally, our appreciation goes to the Spanish Foundation of Rheumatology for their expert medical writing and editorial assistance during the manuscript's preparation.

References

[1] I.B. McInnes, G. Schett, The pathogenesis of rheumatoid arthritis, N. Engl. J. Med. 365 (23) (2011) 2205–2219.
[2] D.T. Felson, Epidemiology of the rheumatic diseases, Curr. Opin. Rheumatol. 2 (2) (1990) 301–308.
[3] K. Almutairi, J. Nossent, D. Preen, H. Keen, C. Inderjeeth, The global prevalence of rheumatoid arthritis: a meta-analysis based on a systematic review, Rheumatol. Int. 41 (5) (2021) 863–877.
[4] H.Y. Yap, S.Z. Tee, M.M. Wong, S.K. Chow, S.C. Peh, S.Y. Teow, Pathogenic role of immune cells in rheumatoid arthritis: implications in clinical treatment and biomarker development, Cells 7 (10) (2018) 161.
[5] N. Kondo, T. Kuroda, D. Kobayashi, Cytokine networks in the pathogenesis of rheumatoid arthritis, Int. J. Mol. Sci. 22 (20) (2021), 10922.
[6] Y.J. Lin, M. Anzaghe, S. Schülke, Update on the pathomechanism, diagnosis, and treatment options for rheumatoid arthritis, Cells 9 (4) (2020) 880.
[7] J.S. Smolen, R.B. Landewé, J.W. Bijlsma, G.R. Burmester, M. Dougados, A. Kerschbaumer, et al., EULAR recommendations for the management of rheumatoid arthritis with synthetic and biological disease-modifying antirheumatic drugs: 2019 update, Ann. Rheum. Dis. 79 (6) (2020) 685–699.
[8] R. van Vollenhoven, Treat-to-target in rheumatoid arthritis—are we there yet? Nat. Rev. Rheumatol. 15 (3) (2019) 180–186.
[9] K.J. Aaltonen, S. Ylikylä, J. Tuulikki Joensuu, P. Isomäki, L. Pirilä, M. Kauppi, et al., Efficacy and effectiveness of tumour necrosis factor inhibitors in the treatment of rheumatoid arthritis in randomized controlled trials and routine clinical practice, Rheumatology 56 (5) (2017) 725–735.
[10] E.G. Favalli, M.G. Raimondo, A. Becciolini, C. Crotti, M. Biggioggero, R. Caporali, The management of first-line biologic therapy failures in rheumatoid arthritis: current practice and future perspectives, Autoimmun. Rev. 16 (12) (2017) 1185–1195.
[11] T. Lequerre, F. Jouen, M. Brazier, S. Clayssens, N. Klemmer, J.F. Menard, et al., Autoantibodies, metalloproteinases and bone markers in rheumatoid arthritis patients are unable to predict their responses to infliximab, Rheumatology 46 (3) (2007) 446–453.

[12] I. González-Alvaro, A.M. Ortiz, E.G. Tomero, A. Balsa, J. Orte, A. Laffon, et al., Baseline serum RANKL levels may serve to predict remission in rheumatoid arthritis patients treated with TNF antagonists, Ann. Rheum. Dis. 66 (12) (2007) 1675–1678.
[13] X.M. Teitsma, J.W. Jacobs, P.H. De Jong, J.M. Hazes, A.E. Weel, P.M. Welsing, et al., Adding baseline protein biomarkers to clinical predictors does not enhance prediction of treatment response to a methotrexate strategy in early rheumatoid arthritis, Ann. Rheum. Dis. 78 (1) (2019) 142–144.
[14] W. Katchamart, S. Johnson, H.J. Lin, V. Phumethum, C. Salliot, C. Bombardier, Predictors for remission in rheumatoid arthritis patients: a systematic review, Arthritis Care Res. 62 (8) (2010) 1128–1143.
[15] M. Novella-Navarro, C. Plasencia, C. Tornero, V. Navarro-Compán, J.L. Cabrera-Alarcón, et al., Clinical predictors of multiple failure to biological therapy in patients with rheumatoid arthritis, Arthritis Res. Ther. 22 (2020) 1–8.
[16] D. Aletaha, T. Neogi, A.J. Silman, J. Funovits, D.T. Felson, C.O. Bingham, et al., 2010 rheumatoid arthritis classification criteria: an American College of Rheumatology/European League against Rheumatism collaborative initiative, Arthritis Rheum. 62 (9) (2010) 2569–2581.
[17] A. Finckh, A. Ciurea, L. Brulhart, B. Möller, U.A. Walker, D. Courvoisier, et al., Which subgroup of patients with rheumatoid arthritis benefits from switching to rituximab versus alternative anti-tumour necrosis factor (TNF) agents after previous failure of an anti-TNF agent? Ann. Rheum. Dis. 69 (2) (2010) 387–393.
[18] J.E. Pope, E.H. Choy, C-reactive protein and implications in rheumatoid arthritis and associated comorbidities, in: Seminars in Arthritis and Rheumatism, vol. 51, WB Saunders, 2021, pp. 219–229. No. 1.
[19] R. Wu, N. Li, X. Zhao, T. Ding, H. Xue, C. Gao, et al., Low-dose Interleukin-2: Biology and therapeutic prospects in rheumatoid arthritis, Autoimmun. Rev. 19 (10) (2020), 102645.
[20] M. Bo, M. Niegowska, G.L. Erre, M. Piras, M.G. Longu, P. Manchia, et al., Rheumatoid arthritis patient antibodies highly recognize IL-2 in the immune response pathway involving IRF5 and EBV antigens, Sci. Rep. 8 (1) (2018) 1789.
[21] Y. Yoshida, T. Tanaka, Interleukin 6 and rheumatoid arthritis, BioMed Res. Int. (2014) 2014.
[22] A. Boyapati, S. Schwartzman, J. Msihid, E. Choy, M.C. Genovese, G.R. Burmester, et al., Association of high serum Interleukin-6 levels with severe progression of rheumatoid arthritis and increased treatment response differentiating Sarilumab from adalimumab or methotrexate in a post hoc analysis, Arthritis Rheumatol. 72 (9) (2020) 1456–1466.
[23] R. Shi, M. Chen, B. Litifu, Serum interleukin-6 and survivin levels predict clinical response to etanercept treatment in patients with established rheumatoid arthritis, Mod. Rheumatol. 28 (1) (2018) 126–132.
[24] A.A. de Hond, A.M. Leeuwenberg, L. Hooft, I.M. Kant, S.W. Nijman, H.J. van Os, et al., Guidelines and quality criteria for artificial intelligence-based prediction models in healthcare: a scoping review, NPJ digital medicine 5 (1) (2022) 2.
[25] A. Javaid, O. Shahab, W. Adorno, P. Fernandes, E. May, S. Syed, Machine learning predictive outcomes modeling in inflammatory bowel diseases, Inflamm. Bowel Dis. 28 (6) (2022) 819–829.
[26] J. Lyu, W. Cui, J. Finkelstein, Use of artificial intelligence for predicting COVID-19 outcomes: a scoping review, Stud. Health Technol. Inf. 289 (2022) 317–320.
[27] S. Momtazmanesh, A. Nowroozi, N. Rezaei, Artificial intelligence in rheumatoid arthritis: current status and future perspectives: a state-of-the-Art review, Rheumatology and Therapy 9 (5) (2022) 1249–1304.
[28] H.R. Gosselt, M.M. Verhoeven, M. Bulatović-Ćalasan, P.M. Welsing, M.C. de Rotte, J.M. Hazes, et al., Complex machine-learning algorithms and multivariable logistic regression on par in the prediction of insufficient clinical response to methotrexate in rheumatoid arthritis, J. Personalized Med. 11 (1) (2021) 44.
[29] B.S. Koo, S. Eun, K. Shin, H. Yoon, C. Hong, D.H. Kim, et al., Machine learning model for identifying important clinical features for predicting remission in patients with rheumatoid arthritis treated with biologics, Arthritis Res. Ther. 23 (1) (2021), 1-0.