

Article

Forecasting Achievement of Inactive Disease in Juvenile Idiopathic Arthritis with Artificial Intelligence †

Ana I. Rebollo-Giménez ^{1,2,3,*} , Francesca Ridella ⁴ , Silvia Maria Orsi ⁴ , Elena Aldera ⁴ , Marco Burrone ^{1,4} ,
Valentina Natoli ^{1,4} , Silvia Rosina ¹ , Alessandro Consolaro ^{1,4}, Esperanza Naredo ⁵, Angelo Ravelli ^{4,6} 
and Davide Cangelosi ⁷ 

- ¹ UOC Reumatologia e Malattie Autoinfiammatorie, IRCCS Istituto Giannina Gaslini, 16147 Genoa, Italy
 - ² Department of Rheumatology, Gregorio Marañón University Hospital, Gregorio Marañón Health Research Institute (IiSGM), 28007 Madrid, Spain
 - ³ Faculty of Medicine, Autonomous University of Madrid (UAM), 28029 Madrid, Spain
 - ⁴ Dipartimento di Neuroscienze, Riabilitazione, Oftalmologia, Genetica e Scienze Materno-Infantili (DINOEMI), Università degli Studi di Genova, 16132 Genoa, Italy
 - ⁵ Department of Rheumatology, Joint and Bone Research Unit, Fundación Jiménez Díaz University Hospital, Health Research Institute Fundación Jiménez Díaz (IIS-FJT, UAM), 28040 Madrid, Spain
 - ⁶ Direzione Scientifica, IRCCS Istituto Giannina Gaslini, 16147 Genoa, Italy
 - ⁷ Unità di Bioinformatica Clinica, Direzione Scientifica, IRCCS Istituto Giannina Gaslini, 16147 Genoa, Italy
- * Correspondence: ana.rebollogimenez@gmail.com; Tel.: +39-010-56362796
- † Preliminary results of this work have been previously submitted in abstract format at the 2024 EULAR Congress.

Abstract: Objective: to seek for predictors of inactive disease (ID) in juvenile idiopathic arthritis (JIA) with artificial intelligence. **Methods:** The clinical charts of patients seen within 6 months after disease onset between 2007 and 2019 and with follow-up visits at 6, 12, 18, and 24 months were reviewed retrospectively. Sixty-eight potential predictors were recorded at each visit. The primary endpoint was ID at 24 months by 2004 Wallace criteria. Data obtained from diverse combinations of visits were examined to identify the best forecasting model. After pre-processing, the cohort was divided into training (50%) and testing (50%) cohorts. Multivariate time series forecasting, coupled with the Random Forest method, was used to train the machine learning (ML) forecasting model. Predictive performance was assessed through the Matthews correlation coefficient (MCC). **Results:** A total of 414 patients were included. The best performance in predicting ID at 24 months in the training cohort was provided by the 0–12 months interval (MCC = 0.68). In the testing cohort, the same ML model confirmed a high forecasting performance (MCC = 0.65). Assessment of feature importance and impact analysis showed that the most relevant predictor of ID was the physician’s global assessment (PhGA), followed by the count of active joints (AJC). **Conclusions:** PhGA and AJC values over the first 12 months were the strongest predictors of ID at 24 months. This finding highlights the importance of regular quantitative assessment of disease activity by the caring physician in monitoring the course of the patient toward achievement of complete disease quiescence.

Keywords: juvenile idiopathic arthritis; pediatric rheumatology; outcome predictors; inactive disease; artificial intelligence; machine learning



Academic Editor: Emanuela Del Giudice

Received: 25 April 2025

Revised: 20 May 2025

Accepted: 3 June 2025

Published: 7 June 2025

Citation: Rebollo-Giménez, A.I.; Ridella, F.; Orsi, S.M.; Aldera, E.; Burrone, M.; Natoli, V.; Rosina, S.; Consolaro, A.; Naredo, E.; Ravelli, A.; et al. Forecasting Achievement of Inactive Disease in Juvenile Idiopathic Arthritis with Artificial Intelligence. *Children* **2025**, *12*, 741. <https://doi.org/10.3390/children12060741>

Copyright: © 2025 by the authors. Licensee MDPI, Basel, Switzerland. This article is an open access article distributed under the terms and conditions of the Creative Commons Attribution (CC BY) license (<https://creativecommons.org/licenses/by/4.0/>).

1. Introduction

Juvenile idiopathic arthritis (JIA) is a chronic rheumatic disease characterized by prolonged synovial inflammation that may lead to progressive joint damage and disability. Permanent changes may also develop in extraarticular organs/systems, such as the eye (as

a complication of chronic anterior uveitis), or may result from side effects of medications, especially systemic glucocorticoids [1]. This morbidity may have a marked impact on the quality of life of patients and their families [2,3].

Over the past two decades, there has been a major advance in the management of JIA, which has made remission an achievable goal for the vast majority of patients [4]. Complete disease quiescence is regarded as the ideal therapeutic objective because its attainment may prevent long-term articular and extra-articular damage and functional impairment [5]. The recommendations for the treat-to-target strategy in JIA have set inactive disease (ID) as the primary target for treatment [6].

Predicting the achievement of ID in JIA is vital to guiding and optimizing treatment decisions. It would be desirable to differentiate early in the course of the illness those patients who are likely to have potentially destructive arthritis from those with self-limiting or non-erosive disease. A better understanding of disease trajectories may facilitate early intervention strategies tailored to individual patient characteristics. However, JIA is a heterogeneous condition which varies widely from one patient to another in terms of disease manifestation, course, phenotype, and response to therapies. Although several prognostic factors have been identified across the various categories of JIA [7], reliable and consistent predictors of therapeutic response and outcome are not yet available for use in routine clinical practice. Notably, most studies have focused on the search for baseline predictors, whereas the development of longitudinal prediction models has been seldom attempted.

In recent years, artificial intelligence (AI) has emerged as a powerful tool that can facilitate screening, diagnosis, monitoring, risk assessment, prognosis determination, achievement of optimal treatment outcome, and de novo drug discovery for patients with rheumatic disorders [8–11]. Investigation of the potential applications of AI, including machine learning (ML) and deep learning techniques, is an exponentially growing field in medicine and healthcare. It has been suggested that incorporation of these methods can be critical to providing high-quality care to patients with chronic rheumatic diseases who lack an optimal treatment [12]. Notably, AI technologies are potentially well suited for constructing longitudinal models of ID prediction in JIA due to their ability to analyze complex, multivariate, and temporal data. They can help identify patterns and trends across multiple time points, which is essential for tracking disease progression. Their capacity to capture non-linear relationships and to select the most relevant features enhances both the accuracy and interpretability of predictions. These properties make AI a valuable tool for personalized and dynamic modeling of disease outcomes in JIA. Against this background, the primary aim of this study was to develop and validate a multivariable forecasting model of ID in JIA through the use of longitudinal data.

2. Materials and Methods

2.1. Study Design and Patient Selection

The clinical charts of all consecutive patients with JIA as defined by the International League of Associations for Rheumatology (ILAR) criteria [13], who were first seen at the Gaslini Institute of Genoa, Italy, in the first 6 months following disease onset between 2007 and 2019 and who had a follow-up visit with available information at 6, 12, 18, and 24 months after initial evaluation were reviewed retrospectively.

2.2. Clinical Assessment

Baseline information included sex, age, disease duration from occurrence of the first symptoms consistent with JIA, and an ILAR category. For simplicity, we grouped patients with different ILAR categories into four “functional” disease phenotypes: systemic arthritis,

polyarthritis (including extended oligoarthritis and rheumatoid factor, RF-positive and RF-negative polyarthritis), oligoarthritis (including persistent oligoarthritis), and other arthritis (including enthesitis-related arthritis, psoriatic arthritis and undifferentiated arthritis). The following data were extracted for each study visit: physician global assessment (PhGA) of overall disease activity using a 21-numbered circle numerical rating scale (NRS), ranging from 0 (=no activity) to 10 (=maximum activity) [14]; active joint count (AJC), assessed in 73 joints [15]; type of affected joints; presence of active systemic manifestations (fever, rash, hepatosplenomegaly, lymphadenopathy, serositis); and presence of active uveitis by the examining ophthalmologist. A joint was defined as active if it displayed swelling or, in case swelling was absent or not detectable (as in the case of cervical spine or hip), pain/tenderness plus restricted motion [16]. Laboratory tests included the erythrocyte sedimentation rate (ESR) and C-reactive protein (CRP). Patients were considered ANA-positive if they had a minimum of two positive ANA test results, obtained at least three months apart during follow-up, using indirect immunofluorescence on Hep-2 cells at a titer of $\geq 1:160$ [17]. The medications administered between study visits were recorded.

Data collection was carried out by five pediatric rheumatology fellows (AIRG, SO, FR, EA, and VN), with oversight provided by an experienced investigator (AR).

2.3. Study Endpoint

The study endpoint was the achievement of ID at 24 months from the baseline evaluation. The state of ID was defined according to the 2004 Wallace criteria, i.e., as joint(s) with active arthritis, no systemic manifestations attributable to JIA, no active uveitis, normal acute-phase reactants, and PhGA indicating no disease activity (defined as a score of 0 on the 0–10 VAS) [18]. However, in a subset of patients, the full application of the Wallace criteria was not possible due to missing PhGA data. In visits where this parameter was unavailable, but all other Wallace criteria were fulfilled, inactive disease status was inferred, following the approach adopted in previous studies [19,20], by reviewing the patient chart until a consensus by two investigators was reached (AIRG and VN). To support this assessment, the attending physician who had initially evaluated the patient during the visit was independently asked to review their clinical notes and verify the state of inactive disease. Any discrepancies between the treating physician and the investigators were resolved through consensus involving the two investigators and a senior author (AR) [21].

2.4. Statistical Analysis

Descriptive statistics were first employed to provide an overview of the baseline characteristics of the study population. Continuous variables were summarized using medians and interquartile ranges (1st–3rd quartiles), while categorical variables were described by their absolute frequencies and corresponding percentages.

2.5. Machine Learning Analysis

The Last Observation Carried Forward (LOCF) method and subsequently the Baseline Observation Carried Forward (BOCF) method, implemented in the Pandas python package [22], were employed for imputing missing values. Patients with remaining missing values after imputation were excluded from the analysis. The mlforecast package, version 0.15.0, was utilized to train the multivariate time series model and generate forecasts. Exogenous features with available values at 24 months were used for forecasting. The recursive method was employed with the horizon parameter set to 1.

Hyperparameter tuning was conducted using the Optuna package [23] with a TPE-Sampler seed set to 10. A total of 100 Optuna trials were executed to optimize the model's parameters. The parameter search space for Random Forests included `n_estimators` ranging from 2 to 500, `max_depth` from 2 to 32, `min_samples_split` from 2 to 10, `min_samples_leaf`

from 3 to 10, and the `random_state` set to 0. Random Forests [24], penalized logistic regression, K-nearest neighbors (KNN), and a support vector machine (SVM) [25] were used in combination with `MLforecast` for binary classification. To ensure an unbiased assessment of model performance, the dataset was randomly split into training (50%) and testing sets (50%). The training set was used to build forecast models, while the testing set was used to estimate prediction performance. The Matthews correlation coefficient (MCC) was chosen as the metric to assess model performance. The MCC ranged from -1 to $+1$, where $+1$ indicates perfect prediction and -1 indicates imperfect prediction. The scale for evaluating prediction performance was defined as follows: $0 \leq \text{MCC} \leq 0.19$ (very low), $0.2 \leq \text{MCC} \leq 0.39$ (low), $0.4 \leq \text{MCC} \leq 0.59$ (moderate), $0.6 \leq \text{MCC} \leq 0.79$ (high), and $0.8 \leq \text{MCC} \leq 1.0$ (very high) [26]. Feature importance in the training set for Random Forests was estimated using the `feature_importances_` attribute of the fitted `RandomForestClassifier` function, calculated as the mean and standard deviation of the impurity decrease across each tree in the trained model. To expedite model training, parallelization was implemented using the `Dask` package [27]. Feature impact in the testing set was estimated using `shap` package by setting up `shap` via the `Explainer` function [28]. Beeswarm plots were utilized to visualize the results for global explainability of the model. Penalized logistic regression, KNN, and SVM model were implemented in `scikit-learn` package version 0.15 [25].

2.6. Patients and Public Involvement

This study was conducted in accordance with the ethical principles outlined in the Declaration of Helsinki. All parents or guardians, or patients themselves if age appropriate, were routinely asked to provide written consent to the use of patients' clinical data for research purposes during the first observation at the study center. The study protocol was approved by the Ethics Committee of Regione Liguria (Genoa, Italy) under protocol number 642/2022—DB id 12828, dated 16 June 2023 [21].

3. Results

3.1. Patient Population and Dataset Creation

A total of 414 patients, whose main clinical features at study entry are shown in Table 1, were included in this study. Table 2 reports the cumulative frequency of the medications administered during the 24 months of follow-up. Our patient population, which is largely represented by children with oligoarticular onset disease, and by our policy of administering intra-articular glucocorticoid therapy in all affected joints as first-line treatment in most children with either oligoarthritis or with polyarthritis and the predominant involvement of large joints. The majority of children who were not receiving any DMARDs were those who experienced sustained remission after such a therapeutic approach. Figure 1 illustrates the schematic representation of the analysis workflow. Patients without a follow-up visit at 24 months, which precluded assessment of the study endpoint, were excluded from the analysis. To evaluate the ability of clinical features to forecast the ID status at 24 months and to identify the earliest follow-up visit capable of accurately predicting the ID status, the dataset was divided into four distinct subsets: T0-T24, T0-T6-T24, T0-T6-T12-T24, and T0-T6-T12-T18-T24. These subsets comprised 339, 339, 302, and 201 patients, respectively, who were retained for subsequent analyses. A total of 68 clinical features, listed in the Supplementary Table S1, were assessed at each time point for their predictive ability. Supplementary Table S2 reports the number of patients for each feature, time point, and feature value. After data imputation and subsequent exclusion of patients with remaining missing values, 317, 327, 294 and 197 patients, respectively, were retained in each of the above datasets for further analysis.

Table 1. Baseline features of the 414 JIA patients *.

	N (%) or Median (IQR)	N with Available Information
Demographic features		
Females	308 (74.4)	414
Median (IQR) age at disease onset (years)	3.1 (1.8–7.0)	414
Median (IQR) age at study entry (years)	3.2 (2.0–7.1)	414
Median (IQR) disease duration at study entry (months)	1.9 (1–3.4)	414
Functional phenotype		
Systemic arthritis	29 (7)	29
Polyarthritis ^a	211 (51)	211
Oligoarthritis	158 (38.2)	158
Other arthritis ^b	16 (3.9)	16
Antinuclear antibody-positive	268 (65.1)	414
Uveitis	15 (3.6)	414
Clinical outcome measures		
Median (IQR) physician’s global assessment	4 (3–6)	414
Active joint count	2 (1–4)	414
Acute phase reactants		
Median (IQR) erythrocyte sedimentation rate	33 (17–51)	356
Median (IQR) C-reactive protein	0.8 (0.5–2.3)	359
Joints involved		
Temporomandibular	16 (3.9)	414
Cervical spine	12 (2.9)	414
Shoulder	17 (4.1)	414
Elbow	59 (14.3)	414
Wrist	77 (18.6)	414
Small hand joints	116 (28)	414
Sacroiliac	2 (0.5)	414
Hip	23 (5.6)	414
Knee	319 (77.1)	414
Ankle	187 (45.2)	414
Small foot joints	81 (19.6)	414

* Data are the number (percentage) unless otherwise indicated. IQR, interquartile range. ^a 11 rheumatoid factor-positive; ^b 6 enthesitis-related arthritis, 5 psoriatic arthritis, and 5 undifferentiated arthritis.

Table 2. Cumulative frequency of the medications administered during the 24-month follow-up period *.

Treatments	N (%)
NSAIDs	346 (83.6)
Intra-articular glucocorticoids	291 (70.3)
Systemic glucocorticoids	113 (27.3)
Methotrexate	273 (65.9)
Sulfasalazine	5 (1.2)
Cyclosporine	3 (0.7)
Etanercept	58 (14)
Adalimumab	18 (4.3)
Infliximab	1 (0.2)
Tocilizumab	3 (0.7)
Abatacept	0 (0)
Anakinra	14 (3.4)
Canakinumab	5 (1.2)
Tofacitinib	0 (0)
Baricitinib	1 (0.2)

* Data are the number (percentage). NSAIDs, non-steroidal anti-inflammatory drugs.

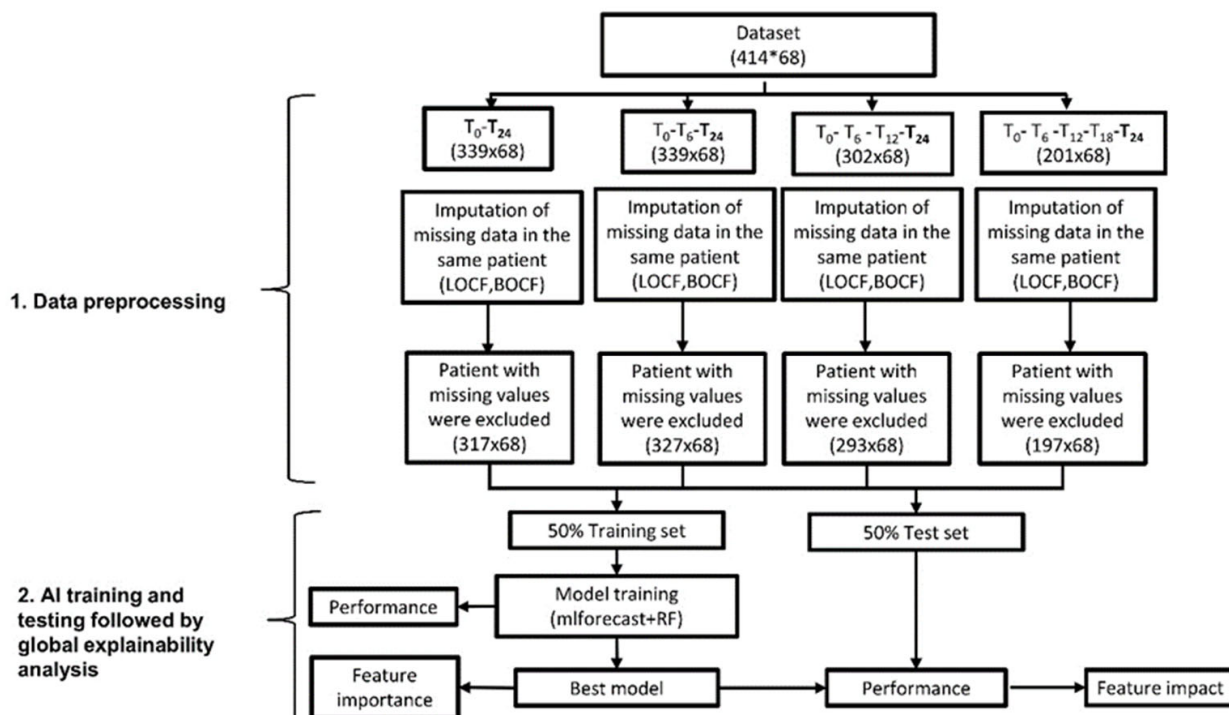


Figure 1. Workflow of the analysis illustrating data pre-processing, AI training, and testing, followed by a global explainability analysis. LOCF = Last Observation Carried Forward. BOCF = Baseline Observation Carried Forward; RF = Random Forests; AI = artificial intelligence.

3.2. Forecast of the State of ID at 24 Months

Each dataset was randomly split into training and testing sets to train the time series forecasting model and generate independent forecasts for the testing set. The MLforecast method, coupled with the RandomForests algorithm, exhibited a heterogeneous forecasting performance in both the training and the testing sets across different datasets, as summarized in Table 3.

Table 3. Forecasting performance of Random Forests coupled with the MLforecast model based on all clinical features and selected time points.

Dataset	MCC Training Set	MCC Testing Set
T0-T6-T12-T18-T24	0.70	0.42
T0-T6-T12-T24	0.68	0.65
T0-T6-T24	0.57	0.50
T0-T24	0.0	0.0

MCC = Matthews Correlation Coefficient. $0 \leq \text{MCC} \leq 0.19$ (Very low), $0.2 \leq \text{MCC} \leq 0.39$ (low), $0.4 \leq \text{MCC} \leq 0.59$ (Moderate), $0.6 \leq \text{MCC} \leq 0.79$ (high), and $0.8 \leq \text{MCC} \leq 1.0$ (very high).

The following most accurate Random Forests models were identified for each dataset:
 T0-T6-T12-T18-T24: ‘n_estimators’: 160, ‘max_depth’: 32, ‘min_samples_split’: 4, ‘min_samples_leaf’: 10.
 T0-T6-T12-T24: ‘n_estimators’: 437, ‘max_depth’: 30, ‘min_samples_split’: 10 ‘min_samples_leaf’: 4.
 T0-T6-T24: ‘n_estimators’: 262, ‘max_depth’: 20, ‘min_samples_split’: 5, ‘min_samples_leaf’: 5.
 T0-T24: ‘n_estimators’: 157, ‘max_depth’: 11, ‘min_samples_split’: 6, ‘min_samples_leaf’: 6.

In the training set, the performance was high for both the T0-T6-T12-T18-T24 and T0-T6-T12-T24 datasets, which achieved MCC scores of 0.70 and 0.68, respectively. The T0-T6-T24 dataset showed a moderate MCC of 0.57. However, the forecasting performance was 0.0 for the T0-T24 dataset, indicating that the model trained solely on baseline features was unable to accurately predict the ID status at T24. Based on the training set performance, the forecasting model developed for the T0-T6-T12-T24 dataset demonstrated the best balance between performance and early prediction capability. In the testing set, the T0-T6-T12-T24 dataset achieved the highest MCC of 0.65, indicating its superior effectiveness in predicting the ID status at 24 months. Lower or comparable MCC were obtained with penalized logistic regression, KNN and SVM models on the training and test sets (Supplementary Table S3).

Feature importance is crucial for post hoc explainability of complex models like Random Forests and for identifying the most relevant clinical features [29]. The relevance of all clinical features was assessed in each training and testing set, and they were ranked in decreasing order of importance, as depicted in Figures 2 and 3. In forecasting the ID status at 24 months, the most relevant features were found to be PhGA and AJC, as illustrated in Figure 2. Acute phase reactants, such as erythrocyte sedimentation rate and C-reactive protein, and the involvement of knee and ankle joints, also exhibited some importance, whereas the remaining clinical features proved less relevant.

PhGA and AJC consistently showed significantly greater importance compared to other features across the T0-T6-T24 (Figure 2a), T0-T6-T12-T24 (Figure 2b), and T0-T6-T12-T18-T24 datasets (Figure 2c). This finding underscores their robust capability to forecast the achievement of the ID status at 24 months.

The role of features in forecasting the ID status at 24 months was further investigated using SHAP (SHapley Additive exPlanations) values in the testing sets, which facilitates a global explainability analysis (Figure 3). Across the T0-T6-T24 (Figure 3a), T0-T6-T12-T24 (Figure 3b), and T0-T6-T12-T18-T24 (Figure 3c) datasets, PhGA and AJC were confirmed as the features with the highest impact. The analysis indicated that lower values of PhGA and AJC were closely related to the attainment of ID at 24 months. The T0-T6 dataset is not included in Figure 3 because for this time interval no feature was found to be important in the training set or had an impact greater than zero in the testing dataset.

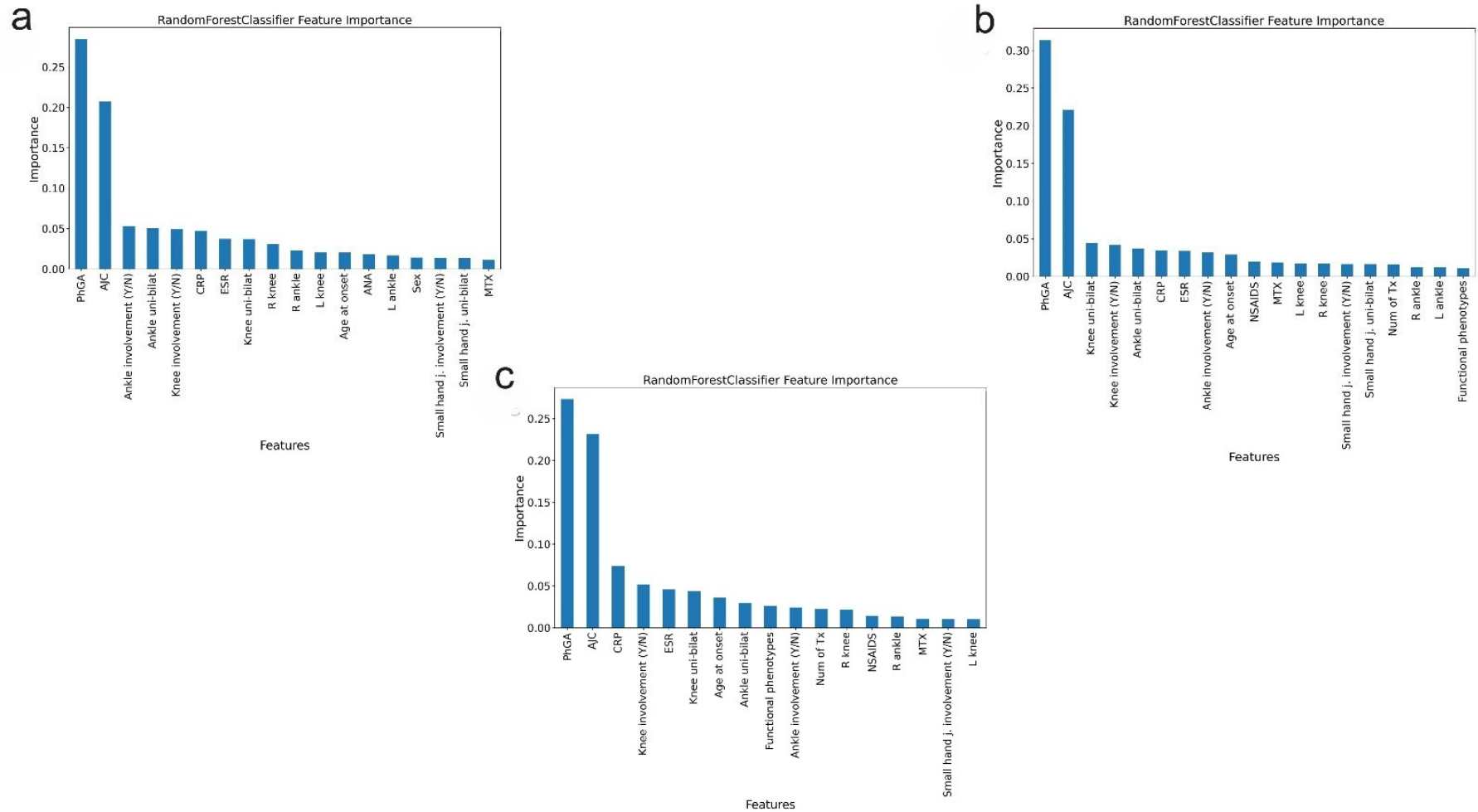


Figure 2. Feature importance in the training set (global explainability): Random Forests. (a) Random Forests using the T0-T6-T24 dataset. (b) Random Forests using the T0-T6-T12-T24 dataset. (c) Random Forest using the T0-T6-T12-T18-T24 dataset. The inclusion of features was restricted to those with an importance greater than 0.01. The most relevant predictors of ID status at 24 months were PhGA and AJC. ID = inactive disease; PhGA = physician global assessment; AJC = active joint count; CRP = C-reactive protein; (Y/N) = Yes/No. Uni-bilat = unilateral–bilateral; ESR = Erythrocyte Sedimentation Rate; R = Right; NSAIDS = non-steroidal anti-inflammatory drugs; L = left; MTX = methotrexate; IAC = intraarticular corticosteroids, j = joint, GS = glucocorticoids, ETA = etanercept; ANA = antinuclear antibody; RF = rheumatoid factor; TMJ = temporomandibular joint.

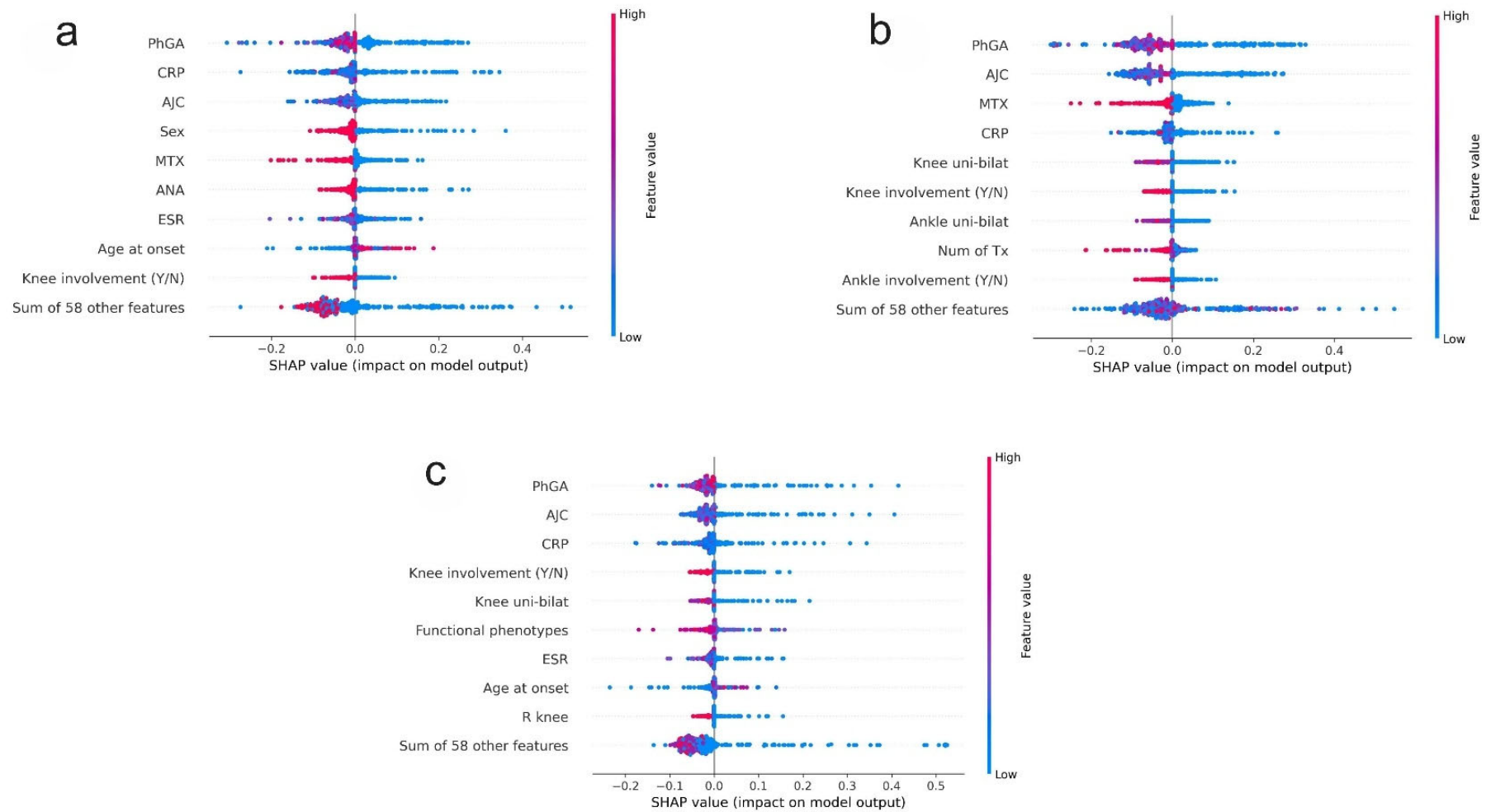


Figure 3. Feature impact on the test set (global explainability): SHAP values. (a) shows the SHAP values using the T0-T6-T24 dataset. (b) the SHAP values using the T0-T6-T12-T24 dataset. (c) the SHAP values using the T0-T6-T12-T18-T24 dataset. AJC and CRP were the features with the highest predictive impact for the T0-T6-T24, T0-T6-T12-T24 and T0-T-T12-T18-T24 datasets. SHAP values = SHapley Additive exPlanations; PhGA = physician global assessment; AJC = active joint count; CRP = C-reactive protein; (Y/N) = Yes/No. Uni-bilat = unilateral–bilateral; ESR = erythrocyte sedimentation rate; R = right; L = left; MTX = methotrexate; Num of Tx = number of treatments, ANA = antinuclear antibody; RF = rheumatoid factor.

4. Discussion

The use of ML techniques for longitudinal data analysis can reveal hidden patterns that may be difficult to detect in cross-sectional studies. This is because historical patient data can be utilized to build the model. Forecasting involves fitting a model to historical, time-stamped data in order to predict future values of a single variable by leveraging the sequential nature of the data. Traditional statistical approaches, such as Moving Average, Exponential Smoothing, or AutoRegressive Integrated Moving Average [30], have been successfully applied in the healthcare domain [31]. However, these methods were not suitable for our longitudinal study because our model needed to include multiple features, along with an output variable for a large set of patients. In contrast, multivariate time series forecasting methods, such as MLforecast, allowed us to include both binary reference variables and multiple input features for a set of patients longitudinally. ML forecast provides publicly available open source libraries that simplify AI programming and could be coupled with Random Forests or other ML models such as penalized logistic regression, KNN, or SVM. However, Random Forests was the model that obtained the best performance across datasets and was selected for further analyses. Despite Random Forests being a black-box model with limited interpretability, it is a well-known ML technique suitable for prediction and feature importance analysis [27]. For these reasons, we employed MLforecast in our analysis. We used real or estimated values as exogenous features for predicting ID at 2 years after the first observation.

Among the different combinations of time points examined, the forecasting model developed for the T0-T6-T12-T24 dataset demonstrated the best effectiveness in predicting the ID status at 24 months. This finding indicates that the 12-month time frame is the optimal interval for evaluating the capability of clinical outcome measures to forecast the achievement of ID.

Our analysis revealed that the PhGA was consistently the best predictor of achievement of ID at 24 months. This finding indicates that the subjective and objective quantitative estimations of the overall level of disease activity by the caring physician at the time of the visit plays a major role in forecasting the future achievement of ID, and that its decrease over time ensures that the patient is in the best trajectory to reach the state of ID. It also underscores the fundamental importance of regularly performing and recording the PhGA during clinical follow-up of patients with JIA.

The PhGA is a key outcome measure for evaluating disease activity in JIA. It is a complex construct that integrates the information obtained from a clinical history with the findings of a clinical assessment. The PhGA has been found to possess strong responsiveness to clinically important change [32] and to be a valid and reliable indicator of overall JIA activity across all stages of the illness [33]. PhGA scores at disease onset predicted disease trajectory after five years [34]. Owing to its good measurement properties, the PhGA has been incorporated into the main composite endpoints for JIA [16,35,36]. However, previous analyses have highlighted a frequent heterogeneity in rating the PhGA across physicians [37,38]. A recent multinational effort has developed recommendations for scoring the PhGA in children with JIA, aiming to enhance the reliability and comparability of disease activity measurement for clinical care and international clinical trials [39].

The second predictive factor in order of importance was the AJC, which is another physician-centered outcome measure. Backström and co-workers found that the most important factors affecting the PhGA score in patients with non-systemic JIA were the swollen and tender joint count [40]. Guidelines for performing a standardized joint assessment in JIA have been provided [13,14]. However, concern was raised by Alongi et al., who found that many pediatric rheumatologists did not mark a score of 0 for patients who lacked active joints. The presence of pain in those joints not meeting the definition

of active arthritis used in JIA was the main determinant of this phenomenon [41]. This observation is consistent with the above considerations about the usefulness of practical guidance for scoring the PhGA in JIA. The number of patients with sJIA included in our cohort was relatively low; we acknowledge that the use of active joint count as a predictive variable may still be relevant for this subgroup. Nevertheless, sJIA can occasionally present with predominant polyarticular joint involvement, especially in patients with a chronic disease course. In such cases, the active joint count may serve as a relevant and clinically meaningful parameter for monitoring disease activity and guiding therapeutic decisions, similarly to other JIA categories. However, due to the distinct pathophysiology and clinical heterogeneity of sJIA, further studies specifically designed to validate predictive models in this subgroup are warranted.

Our study is not without limitations. Patient data were collected through the retrospective review of clinical charts. A retrospective analysis is subject to missing and possibly erroneous data. We did not investigate whether children who were taken off DMARDs within the first year were more likely to experience active disease or relapse. We could not investigate the role of novel biomarkers or imaging methods, especially ultrasound, either in predicting ID or in assessing the state of ID at 24 months. It has been argued that these methodologies may establish disease remission more reliably than clinical assessment [42,43]. The methodology used in our study did not enable us to assess the relationship between variables and the study outcome by calculating numerical thresholds but only allowed us to assess the relative importance of each variable in forecasting ID. Given the low number of patients with sJIA, we could not conclude that our prediction model is as accurate for patients with this JIA category as for those with non-systemic forms. Due to the lack of data in a sizeable number of patients and visits, we could not incorporate the parent/patient assessment of well-being and pain intensity among the study predictors. Therefore, we should recognize that our results do not reflect the potential predictive role of the parent/patient perception of disease status and course. We should finally acknowledge that other items not captured in this study, such as intolerance of medication, might have influenced achievement of ID. The main strength of our study lies in demonstrating, through ML analysis, the capability of our AI model to predict the state of ID at 24 months using empirical exogenous features defined with data obtained from 0 to 12 months. This finding, together with the observed lack of any predictive value of the baseline data, underscores the potential superiority of assessing longitudinal data from multiple follow-up visits, over the simple evaluation of data collected at disease onset, in forecasting long-term achievement of ID.

In conclusion, our study is the first to investigate the role of ML techniques in the prediction of ID in patients with JIA. Through the analysis of longitudinal models built with data collected at subsequent longitudinal visits, we found that the PhGA and the AJC over the first 12 months were the strongest predictors of achieving ID at 24 months. This finding underscores the fundamental importance of regularly measuring the level of disease activity in daily clinical practice using standardized and well-validated outcome measures. These preliminary results highlight the potential of ML techniques in predicting ID in JIA patients; however, further multicenter studies with larger cohorts are necessary to confirm and generalize these results. Future applications of AI methods should integrate clinical information with genetic, multiomic, and imaging data in order to develop robust predictive clinical algorithms that help to optimize the management of JIA through a more precise and personalized therapeutic approach.

Supplementary Materials: The following supporting information can be downloaded at <https://www.mdpi.com/article/10.3390/children12060741/s1>, Table S1. Predictor variables assessed in the analyses; Table S2. Patient features by time points for each dataset (T0-T24, T0-T6-T24, T0-T6-T12-

T24, and T0-T6-T12-T18-T24); Table S3: Forecasting MCC of additional AI algorithms coupled with the MLforecast model based on all clinical features and selected time points.

Author Contributions: A.I.R.-G.: Conceptualization, Data curation, Formal analysis, Investigation, Methodology, Project Administration, Resources, Writing—original draft. F.R., S.M.O., E.A., M.B., V.N., S.R., A.C., and E.N.: Data Curation, Investigation, Writing—Review and Editing. A.R.: Conceptualization, Data curation, Methodology, Project Administration, Supervision, Writing—Review and Editing. D.C.: Conceptualization, Data curation, Methodology, Project Administration, Software, Resources, Supervision, Writing—Review and Editing. A.I.R.-G. confirms that all authors have expressly approved the version to be published and take responsibility for the affirmations regarding article submission (e.g., not under consideration by another journal), the integrity of the data presented, and the statements regarding compliance with institutional review board/Helsinki Declaration requirements. All authors have read and agreed to the published version of the manuscript.

Funding: This work was supported by the Italian Ministry of Health with “2024 Ricerca Corrente” funds—ID n. RRC-2024-2787194 and “2020 5x1000” funds—ID n. 5M-2018-23680429. Funders had no active role in the present study.

Institutional Review Board Statement: This study was conducted in accordance with the Declaration of Helsinki and approved by the Ethics Committee of Regione Liguria (Genoa, Italy) under protocol number 642/2022—DB id 12828, dated 16 June 2023.

Informed Consent Statement: Informed consent was obtained from all subjects involved in the study.

Data Availability Statement: The anonymized clinical data used in this manuscript have been deposited on the FigShare platform and are available at DOI: 10.6084/m9.figshare.28429931.

Acknowledgments: Ana Isabel Rebollo-Gimenez holds a fellowship supported by the Fundación Española de Reumatología. The authors thank the Spanish Foundation of Rheumatology for providing medical writing/editorial assistance during the preparation of the manuscript-ID FERBT2025. The results of this work have been previously submitted in abstract format at the 2024 EULAR Congress [44].

Conflicts of Interest: AR reports a relationship with AbbVie, Novartis, Pfizer, Reckitt Benckiser, Alexion, Galapagos, and Sobi that includes speaking and lecture fees. AC reports a relationship with Pfizer that includes speaking and lecture fees. AIRG, SMO, FR, EA, EN, MB, VN, SR, and DC declare no conflicts of interest.

Abbreviations

The following abbreviations are used in this manuscript:

AI	Artificial Intelligence
AJC	Active Joint Count
ANA	Antinuclear Antibody
BOCF	Baseline Observation Carried Forward
CRP	C-Reactive Protein
ESR	Erythrocyte Sedimentation Rate
ID	Inactive Disease
ILAR	International League of Associations for Rheumatology
JIA	Juvenile Idiopathic Arthritis
LOCF	Last Observation Carried Forward
MCC	Matthews Correlation Coefficient
ML	Machine Learning
PhGA	Physician Global Assessment
RF	Rheumatoid Factor
SHAP	SHapley Additive exPlanations

References

1. Martini, A.; Lovell, D.J.; Albani, S.; Brunner, H.I.; Hyrich, K.L.; Thompson, S.D.; Ruperto, N. Juvenile idiopathic arthritis. *Nat. Rev. Dis. Primers* **2022**, *8*, 5. [[CrossRef](#)] [[PubMed](#)]
2. Miller, M.L.; LeBovidge, J.; Feldman, B. Health-related quality of life in children with arthritis. *Rheum. Dis. Clin. N. Am.* **2002**, *28*, 493–501. [[CrossRef](#)]
3. Brunner, H.I.; Giannini, E.H. Health-related quality of life in children with rheumatic diseases. *Curr. Opin. Rheumatol.* **2003**, *15*, 602–612. [[CrossRef](#)]
4. Lovell, D.J.; Ruperto, N.; Giannini, E.H.; Martini, A. Advances from clinical trials in juvenile idiopathic arthritis. *Nat. Rev. Rheumatol.* **2013**, *9*, 557–563. [[CrossRef](#)] [[PubMed](#)]
5. Shoop-Worrall, S.J.W.; Verstappen, S.M.M.; Baidam, E.; Chieng, A.; Davidson, J.; Foster, H.; Ioannou, Y.; McErlane, F.; Wedderburn, L.R.; Thomson, W.; et al. How common is clinically inactive disease in a prospective cohort of patients with juvenile idiopathic arthritis? The importance of definition. *Ann. Rheum. Dis.* **2017**, *76*, 1381–1388. [[CrossRef](#)]
6. Ravelli, A.; Consolaro, A.; Horneff, G.; Laxer, R.M.; Lovell, D.J.; Wulffraat, N.M.; Akikusa, J.D.; Al-Mayouf, S.M.; Antón, J.; Avcin, T.; et al. Treating juvenile idiopathic arthritis to target: Recommendations of an international task force. *Ann. Rheum. Dis.* **2018**, *77*, 819–828. [[CrossRef](#)] [[PubMed](#)]
7. van Dijkhuizen, E.H.P.; Wulffraat, N.M. Early predictors of prognosis in juvenile idiopathic arthritis: A systematic literature review. *Ann. Rheum. Dis.* **2015**, *74*, 1996–2005. [[CrossRef](#)]
8. Momtazmanesh, S.; Nowroozi, A.; Rezaei, N. Artificial Intelligence in Rheumatoid Arthritis: Current Status and Future Perspectives: A State-of-the-Art Review. *Rheumatol. Ther.* **2022**, *9*, 1249–1304. [[CrossRef](#)]
9. Venerito, V.; Bilgin, E.; Iannone, F.; Kiraz, S. AI am a rheumatologist: A practical primer to large language models for rheumatologists. *Rheumatology* **2023**, *62*, 3256–3260. [[CrossRef](#)]
10. Benavent, D.; Madrid-García, A. Large language models and rheumatology: Are we there yet? *Rheumatol. Adv. Pr.* **2024**, *9*, rkae119. [[CrossRef](#)]
11. La Bella, S.; Gupta, L.; Venerito, V. AI am the future: Artificial intelligence in pediatric rheumatology. *Curr. Opin. Rheumatol.* **2025**, 10–1097. [[CrossRef](#)] [[PubMed](#)]
12. Dubey, S.; Chan, A.; O Adebajo, A.; Walker, D.; Bukhari, M.; Adebajo, A. Artificial intelligence and machine learning in rheumatology. *Rheumatology* **2024**, *63*, 2040–2041. [[CrossRef](#)] [[PubMed](#)]
13. Petty, R.E.; Southwood, T.R.; Manners, P.; Baum, J.; Glass, D.N.; Goldenberg, J.; He, X.; Maldonado-Cocco, J.; Orozco-Alcala, J.; Prieur, A.-M.; et al. International League of Associations for Rheumatology classification of juvenile idiopathic arthritis: Second revision, Edmonton, 2001. *J. Rheumatol.* **2004**, *31*, 390–392. [[PubMed](#)]
14. Filocamo, G.; Davì, S.; Pistorio, A.; Bertamino, M.; Ruperto, N.; Lattanzi, B.; Consolaro, A.; Magni-Manzoni, S.; Galasso, R.; Varnier, G.C.; et al. Evaluation of 21-numbered circle and 10-centimeter horizontal line visual analog scales for physician and parent subjective ratings in juvenile idiopathic arthritis. *J. Rheumatol.* **2010**, *37*, 1534–1541. [[CrossRef](#)]
15. Bazso, A.; Consolaro, A.; Ruperto, N.; Pistorio, A.; Viola, S.; Magni-Manzoni, S.; Malattia, C.; Buoncompagni, A.; Loy, A.; Martini, A.; et al. Development and testing of reduced joint counts in juvenile idiopathic arthritis. *J. Rheumatol.* **2009**, *36*, 183–190. [[CrossRef](#)]
16. Ravelli, A.; Viola, S.; Ruperto, N.; Corsi, B.; Ballardini, G.; Martini, A. Correlation between conventional disease activity measures in juvenile chronic arthritis. *Ann. Rheum. Dis.* **1997**, *56*, 197–200. [[CrossRef](#)]
17. Ravelli, A.; Varnier, G.C.; Oliveira, S.; Castell, E.; Arguedas, O.; Magnani, A.; Pistorio, A.; Ruperto, N.; Magni-Manzoni, S.; Galasso, R.; et al. Antinuclear antibody-positive patients should be grouped as a separate category in the classification of juvenile idiopathic arthritis. *Arthritis Rheum.* **2011**, *63*, 267–275. [[CrossRef](#)] [[PubMed](#)]
18. Wallace, C.A.; Ruperto, N.; Giannini, E. Childhood Arthritis and Rheumatology Research Alliance, Pediatric Rheumatology International Trials Organization, Pediatric Rheumatology Collaborative Study Group. Preliminary criteria for clinical remission for select categories of juvenile idiopathic arthritis. *J. Rheumatol.* **2004**, *31*, 2290–2294.
19. Bava, C.; Mongelli, F.; Pistorio, A.; Bertamino, M.; Bracciolini, G.; Dalprà, S.; Davì, S.; Lanni, S.; Muratore, V.; Pederzoli, S.; et al. A prediction rule for lack of achievement of inactive disease with methotrexate as the sole disease-modifying antirheumatic therapy in juvenile idiopathic arthritis. *Pediatr. Rheumatol. Online J.* **2019**, *17*, 50. [[CrossRef](#)]
20. Bava, C.; Mongelli, F.; Pistorio, A.; Bertamino, M.; Bracciolini, G.; Dalprà, S.; Davì, S.; Lanni, S.; Muratore, V.; Pederzoli, S.; et al. Analysis of arthritis flares after achievement of inactive disease with methotrexate monotherapy in juvenile idiopathic arthritis. *Clin. Exp. Rheumatol.* **2021**, *39*, 426–433. [[CrossRef](#)]
21. Rebollo-Giménez, A.I.; Pistorio, A.; Orsi, S.M.; Ridella, F.; Aldera, E.; Carlini, L.; Natoli, V.; Burrone, M.; Rosina, S.; Naddei, R.; et al. Frequency of remission achievement in the pre-treat-to-target decade in juvenile idiopathic arthritis. *Pediatr. Rheumatol. Online J.* **2025**, *23*, 8. [[CrossRef](#)] [[PubMed](#)]

22. McKinney, W. Data Structures for Statistical Computing in Python. In Proceedings of the 9th Python in Science Conference, Austin, TX, USA, 28 June–3 July 2010; Volume 445, pp. 56–61. [\[CrossRef\]](#)
23. Akiba, T.; Sano, S.; Yanase, T.; Ohta, T.; Koyama, M. Optuna: A Next-generation Hyperparameter Optimization Framework. In Proceedings of the 25th ACM SIGKDD International Conference on Knowledge Discovery & Data Mining, Anchorage, AK, USA, 4–8 August 2019; pp. 2623–2631. [\[CrossRef\]](#)
24. Breiman, L. Random Forests. *Mach. Learn.* **2001**, *45*, 5–32. [\[CrossRef\]](#)
25. Pedregosa, F.; Varoquaux, G.; Gramfort, A.; Michel, V.; Thirion, B.; Grisel, O.; Blondel, M.; Prettenhofer, P.; Weiss, R.; Dubourg, V.; et al. Scikit-learn: Machine learning in Python. *J. Mach. Learn. Res.* **2011**, *12*, 2825–2830.
26. Matthews, B.W. Comparison of the predicted and observed secondary structure of T4 phage lysozyme. *Biochim. Biophys. Acta* **1975**, *405*, 442–451. [\[CrossRef\]](#)
27. Rocklin, M. Dask: Parallel computation with blocked algorithms and task scheduling. In Proceedings of the 14th Python in Science Conference, Austin, TX, USA, 6–12 July 2015; pp. 126–132.
28. Lundberg, S.M.; Lee, S.I. A Unified Approach to Interpreting Model Predictions. In *Advances in Neural Information Processing Systems 30*; Curran Associates Inc.: Red Hook, NY, USA, 2017; pp. 4765–4774. [\[CrossRef\]](#)
29. Chicco, D.; Haupt, R.; Garaventa, A.; Uva, P.; Luksch, R.; Cangelosi, D. Computational intelligence analysis of high-risk neuroblastoma patient health records reveals time to maximum response as one of the most relevant factors for outcome prediction. *Eur. J. Cancer* **2023**, *193*, 113291. [\[CrossRef\]](#)
30. Fatima, S.S.W.; Rahimi, A. A Review of Time-Series Forecasting Algorithms for Industrial Manufacturing Systems. *Machines* **2024**, *12*, 380. [\[CrossRef\]](#)
31. Soyiri, I.N.; Reidpath, D.D. Evolving forecasting classifications and applications in health forecasting. *Int. J. Gen. Med.* **2012**, *5*, 381–389. [\[CrossRef\]](#)
32. Moretti, C.; Viola, S.; Pistorio, A.; Magni-Manzoni, S.; Ruperto, N.; Martini, A.; Ravelli, A. Relative responsiveness of condition specific and generic health status measures in juvenile idiopathic arthritis. *Ann. Rheum. Dis.* **2005**, *64*, 257–261. [\[CrossRef\]](#)
33. Palmisani, E.; Solari, N.; Magni-Manzoni, S.; Pistorio, A.; Labò, E.; Panigada, S.; Martini, A.; Ravelli, A. Correlation between juvenile idiopathic arthritis activity and damage measures in early, advanced, and longstanding disease. *Arthritis Rheum.* **2006**, *55*, 843–849. [\[CrossRef\]](#)
34. Guzman, J.; Henrey, A.; Loughin, T.; Berard, R.A.; Shiff, N.J.; Jurencak, R.; Benseler, S.M.; Tucker, L.B. Predicting Which Children with Juvenile Idiopathic Arthritis Will Have a Severe Disease Course: Results from the ReACCh-Out Cohort. *J. Rheumatol.* **2017**, *44*, 230–240. [\[CrossRef\]](#)
35. Wallace, C.A.; Giannini, E.H.; Huang, B.; Itert, L.; Ruperto, N.; Childhood Arthritis Rheumatology Research Alliance (CARRA); Pediatric Rheumatology Collaborative Study Group (PRCSG); Paediatric Rheumatology International Trials Organisation (PRINTO). American College of Rheumatology provisional criteria for defining clinical inactive disease in select categories of juvenile idiopathic arthritis. *Arthritis Care Res.* **2011**, *63*, 929–936. [\[CrossRef\]](#) [\[PubMed\]](#)
36. Trincianti, C.; Consolaro, A. Outcome Measures for Juvenile Idiopathic Arthritis Disease Activity. *Arthritis Care Res.* **2020**, *72* (Suppl. S10), 150–162. [\[CrossRef\]](#) [\[PubMed\]](#)
37. Falcone, A.; Cassone, R.; Rossi, E.; Pistorio, A.; Martini, A.; Ravelli, A. Inter-observer agreement of the physician’s global assessment of disease activity in children with juvenile idiopathic arthritis. *Clin. Exp. Rheumatol.* **2005**, *23*, 113–116.
38. Taylor, J.; Giannini, E.H.; Lovell, D.J.; Huang, B.; Morgan, E.M. Lack of Concordance in Interrater Scoring of the Provider’s Global Assessment of Children With Juvenile Idiopathic Arthritis With Low Disease Activity. *Arthritis Care Res.* **2018**, *70*, 162–166. [\[CrossRef\]](#)
39. Rypdal, V.; Brunner, H.I.; Feldman, B.M.; Ruperto, N.; Aggarwal, A.; Angeles-Han, S.T.; Backström, M.; Balay-Dustrude, E.; Bracaglia, C.; De Benedetti, F.; et al. Physician’s global assessment of disease activity in juvenile idiopathic arthritis: Consensus-based recommendations from an international task force. *Ann. Rheum. Dis.* **2025**. [\[CrossRef\]](#) [\[PubMed\]](#)
40. Backström, M.; Tarkiainen, M.; Gottlieb, B.S.; Trincianti, C.; Qiu, T.; Morgan, E.; Lovell, D.J.; Bovis, F.; Löyttyniemi, E.; Ruperto, N.; et al. Paediatric rheumatologists do not score the physician’s global assessment of juvenile idiopathic arthritis disease activity in the same way. *Rheumatology* **2023**, *62*, 3421–3426. [\[CrossRef\]](#)
41. Alongi, A.; Giancane, G.; Naddei, R.; Natoli, V.; Ridella, F.; Burrone, M.; Rosina, S.; Chedeville, G.; Alexeeva, E.; Horneff, G.; et al. Drivers of non-zero physician global scores during periods of inactive disease in juvenile idiopathic arthritis. *RMD Open* **2022**, *8*, e002042. [\[CrossRef\]](#)
42. De Lucia, O.; Giani, T.; Caporali, R.; Cimaz, R. Ultrasound versus physical examination in predicting disease flare in children with juvenile idiopathic arthritis: A systematic literature review and qualitative synthesis. *Med. Ultrason.* **2022**, *24*, 473–478. [\[CrossRef\]](#)

43. Rosina, S.; Natoli, V.; Santaniello, S.; Trinciante, C.; Consolaro, A.; Ravelli, A. Novel biomarkers for prediction of outcome and therapeutic response in juvenile idiopathic arthritis. *Expert. Rev. Clin. Immunol.* **2021**, *17*, 853–870. [[CrossRef](#)]
44. Giménez, A.I.R.; Cangelosi, D.; Ridella, F.; Orsi, S.; Aldera, E.; Natoli, V.; Rosina, S.; Naredo, E.; Ravelli, A. POS0762 Seeking for predictors of inactive disease in juvenile idiopathic arthritis with artificial intelligence. *Ann. Rheum. Dis.* **2024**, *83* (Suppl. S1), 1172–1173. [[CrossRef](#)]

Disclaimer/Publisher’s Note: The statements, opinions and data contained in all publications are solely those of the individual author(s) and contributor(s) and not of MDPI and/or the editor(s). MDPI and/or the editor(s) disclaim responsibility for any injury to people or property resulting from any ideas, methods, instructions or products referred to in the content.