



# Machine learning models with time-series clinical features to predict radiographic progression in patients with ankylosing spondylitis

Bon San Koo, M.D., Ph.D.<sup>1\*</sup>, Miso Jang, M.D., Ph.D.<sup>2,3\*</sup>, Ji Seon Oh, M.D., Ph.D.<sup>4</sup>, Keewon Shin, Ph.D.<sup>2</sup>, Seunghun Lee, M.D., Ph.D.<sup>5</sup>, Kyung Bin Joo, M.D., Ph.D.<sup>5</sup>, Namkug Kim, Ph.D.<sup>6,7</sup>, Tae-Hwan Kim, M.D., Ph.D.<sup>8</sup>

<sup>1</sup>Department of Internal Medicine, Inje University Ilsan Paik Hospital, Inje University College of Medicine, <sup>2</sup>Department of Biomedical Engineering, Asan Medical Institute of Convergence Science and Technology, Asan Medical Center, University of Ulsan College of Medicine, <sup>3</sup>Department of Medicine, Asan Medical Center, University of Ulsan College of Medicine, <sup>4</sup>Department of Information Medicine, Big Data Research Center, Asan Medical Center, <sup>5</sup>Department of Radiology, Hanyang University Hospital for Rheumatic Diseases, Departments of <sup>6</sup>Radiology and <sup>7</sup>Convergence Medicine, Asan Medical Center, University of Ulsan College of Medicine, <sup>8</sup>Department of Rheumatology, Hanyang University Hospital for Rheumatic Diseases, Seoul, Korea

**Objective:** Ankylosing spondylitis (AS) is chronic inflammatory arthritis causing structural damage and radiographic progression to the spine due to repeated and continuous inflammation over a long period. This study establishes the application of machine learning models to predict radiographic progression in AS patients using time-series data from electronic medical records (EMRs).

**Methods:** EMR data, including baseline characteristics, laboratory findings, drug administration, and modified Stoke AS Spine Score (mSASSS), were collected from 1,123 AS patients between January 2001 and December 2018 at a single center at the time of first ( $T_1$ ), second ( $T_2$ ), and third ( $T_3$ ) visits. The radiographic progression of the ( $n+1$ )th visit ( $P_{n+1}=(mSASSS_{n+1}-mSASSS_n)/(T_{n+1}-T_n) \geq 1$  unit per year) was predicted using follow-up visit datasets from  $T_1$  to  $T_n$ . We used three machine learning methods (logistic regression with the least absolute shrinkage and selection operation, random forest, and extreme gradient boosting algorithms) with three-fold cross-validation.

**Results:** The random forest model using the  $T_1$  EMR dataset best predicted the radiographic progression  $P_2$  among the machine learning models tested with a mean accuracy and area under the curves of 73.73% and 0.79, respectively. Among the  $T_1$  variables, the most important variables for predicting radiographic progression were in the order of total mSASSS, age, and alkaline phosphatase.

**Conclusion:** Prognosis predictive models using time-series data showed reasonable performance with clinical features of the first visit dataset when predicting radiographic progression.

**Keywords:** Ankylosing spondylitis, Machine learning, Disease progression

Received September 4, 2023; Revised October 15, 2023; Accepted October 30, 2023, Published online December 20, 2023

**Corresponding author:** Namkug Kim, <https://orcid.org/0000-0002-3438-2217>  
Department of Convergence Medicine, Asan Medical Center, University of Ulsan College of Medicine, 88 Olympic-ro 43-gil, Songpa-gu, Seoul 05505, Korea. **E-mail:** namkugkim@gmail.com  
Tae-Hwan Kim, <https://orcid.org/0000-0002-3542-2276>  
Department of Rheumatology, Hanyang University Hospital for Rheumatic Diseases, 222-1 Wangsimni-ro, Seongdong-gu, Seoul 04763, Korea. **E-mail:** thkim@hanyang.ac.kr

\*These authors contributed equally to this work.

Copyright © The Korean College of Rheumatology.



This is an Open Access article distributed under the terms of the Creative Commons Attribution Non-Commercial License (<http://creativecommons.org/licenses/by-nc/4.0>) which permits unrestricted non-commercial use, distribution, and reproduction in any medium, provided the original work is properly cited.

## INTRODUCTION

Patients with ankylosing spondylitis (AS), a chronic inflammatory arthritis, have chronic inflammatory back pain and gradually develop ankylosis of the spine [1], limiting their movement. Because structural changes due to inflammation may impact normal functioning and quality of life. Identifying key predictors that contribute to the acceleration of vertebral ankylosis in AS patients is of paramount importance.

Previous studies mostly used statistical methods to investigate patient features related to spinal structural changes shown by radiography. They identified that radiographic progression significantly correlated with men, tobacco, inflammation, and HLA-27 [2-4]. However, predicting radiographic progression in an individual patient is challenging because of various indirectly related factors over time. Because numerous data of various types have been accumulated over time in electronic medical records (EMRs) of AS patients under clinical care, statistical methods may have limitations in analyzing and predicting AS radiographic progression. However, machine learning methods can help predict radiographic progression using these accumulated data and facilitate understanding the complex relationships between variables in big data.

Using machine learning methods in relation to big data has increased in the medical field [5]. This approach not only predicts disease outcomes through data-analysis but also highlights the significance of key features required to forecast disease onset or activity. Therefore, EMRs stored over time may be the best source for use in machine learning models [6]. However, there are challenges to using big data analytics in the medical field. It is necessary to consider whether big data analytics offer evidence of help in clinical practice and whether it can overcome the quality, inconsistency, observational data limitations, and validation issues of big data in terms of the approach [7-9].

A major strength of machine learning models is that they can handle complex and heterogeneous data such as time-series EMRs. This study explored applying machine learning models to predict radiographic progression in AS patients based on time-series data from earlier visits and identify predictive datasets and key features contributing to radiographic progression in these models.

## MATERIALS AND METHODS

### Patients

This paper describes a retrospective study conducted at the Hanyang University Seoul Hospital. The dataset comprised reviewed EMR data from January 2001 to December 2018 of 1,280 patients. All patients were diagnosed with AS according to the following modified New York criteria [10]; 1) clinical criteria; lower back pain, limited range of motion of the lumbar spine, and limitation of chest expansion for at least three months, 2) radiological criteria; Sacroiliac arthritis is bilateral grade 2~4 or unilateral grade 3~4. If any criteria from both the clinical and radiographical criteria are fulfilled, it is classified as AS. Out of the 1,280 patients, 157 were excluded due to a lack of clinical and/or radiologic data. The study was approved by the institutional review board at the Hanyang University Seoul Hospital (HYUH 2020-03-012-003). Informed consent was waived because this study retrospectively reviewed the EMRs. This study included only anonymized patient data and was performed in accordance with the Declaration of Helsinki.

### Clinical data

Patients in this cohort had radiographs taken every 2 years to evaluate modified Stoke AS Spine Score (mSASSS) using spinal radiographic changes. Clinical characteristics, including age, sex, disease duration from the first to the last follow-up, HLA-B27 positivity, eye involvement with uveitis, and peripheral joint involvement with arthritis other than axial joints, were investigated. Baseline laboratory results comprised hemoglobin, hematocrit, blood urea nitrogen, creatinine, aspartate transaminase, alanine transaminase (ALT), alkaline phosphatase (ALP), albumin, cholesterol, protein, creatine phosphokinase, gamma glutamyl peptidase, lactate dehydrogenase, erythrocyte sedimentation rate (ESR), and C-reactive protein (CRP) levels. The prescribed drugs were classified as nonsteroidal anti-inflammatory drugs (NSAIDs), methotrexate, steroids, sulfasalazine, and biological disease-modifying antirheumatic drugs (bDMARDs). The mean values of laboratory tests, the total number of prescribed medications from the first visit to the current time point, and clinical characteristics were used as machine learning features.

### Radiographic progression assessment

The mSASSS is a tool used to assess changes in spinal stiff-

ness in AS patients [11,12]. In the lateral view of the cervical and lumbar spine, sclerosis, erosion, syndesmophyte, and complete ankylosis at 24 corners can be scored from 0 to 3, totaling 72 points. Although the criteria for radiographic progression in AS patients vary among studies, it is generally defined as an increase of 2 or more in the total mSASSS score after two years [12].

Two radiologists (SL and KBJ) independently assessed the images and scored them according to the mSASSS (0~72) [11]. Intraobserver reliability with consistency for a reader (intraclass coefficient [ICC]=0.978, 95% confidence interval [CI]: 0.976 to 0.979) and interobserver reliability with the agreement between two readers (ICC=0.946, 95% CI: 0.941 to 0.950) were also excellent [13,14].

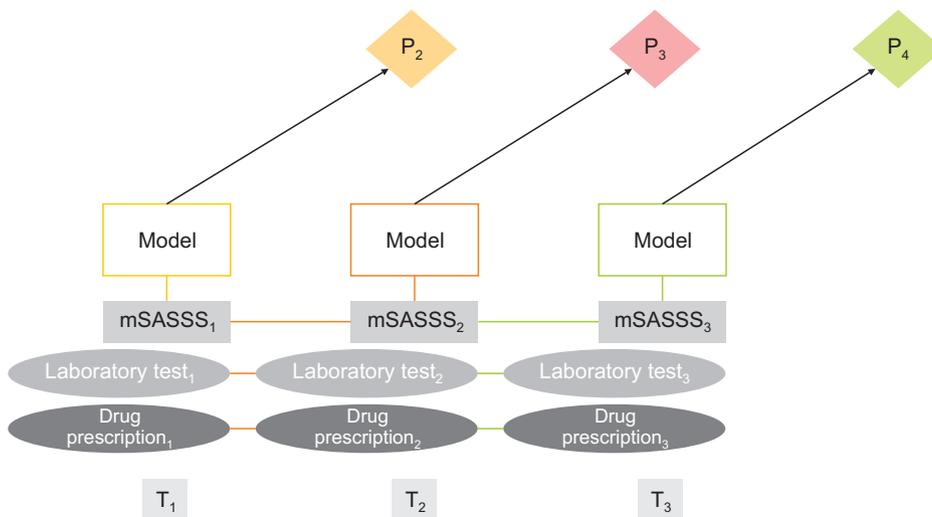
### Model design

Although there is a correlation between the onset of inflammation and spinal radiographic changes after two years, the evidence is inconclusive [12]. Therefore, this study presented models to predict radiographic progression with clinical variables of visits at various time points. The first ( $T_1$ ), second ( $T_2$ ), and third visits ( $T_3$ ) were defined as the time points at which the first, second, and third radiographs were taken, respectively. In addition, the radiographic progression at each time point was calculated as follows:  $P_{n+1}=(mSASSS_{n+1}-mSASSS_n)/(T_{n+1}-T_n)$ . In other words, mSASSS change is calculated as the difference between the current time point and the previous time point in mSASSS, divided by the time, and presented as the rate of

change over one year. A radiographic progressor was defined as an individual whose mean mSASSS worsened by more than one unit over one year [15]. AS patients were categorized into progressor and non-progressor groups. The model uses a binary classifier with progressor and non-progressor groups labeled 1 and 0, respectively.

We composed three clinical datasets for predicting radiographic progression: baseline dataset at the first visit ( $T_1$ ) with radiographic progression at the second visit ( $P_2$ ), two-point dataset at first and second visits ( $T_1+T_2$ ) with radiographic progression at the third visit ( $P_3$ ), and three-point dataset at first, second, and third visits ( $T_1+T_2+T_3$ ) with radiographic progression at fourth visit ( $P_4$ ). The three clinical dataset matrixes were used to train the three prediction models for progressor and non-progressor groups (Figure 1). Three machine learning classifiers were applied: logistic regression with least absolute shrinkage and selection operation (LASSO) using Python in the Scikit-learn package (<https://github.com/scikit-learn/scikit-learn>) [16], random forest (RF) using the Scikit-learn package [17], and extreme gradient boosting (XGBoost) using the Xgboost package (<https://github.com/dmlc/xgboost>) [18]. The algorithms were selected based on their superior performance and application readiness. All continuous clinical features were centered and scaled to a mean of zero and a standard deviation of one (z-score transformation was performed before feature selection). The results of the three models were compared to determine the best combination for determining progressor or non-progressor in the three clinical datasets. All possible

Radiographic progression ( $P_{n+1}$ ):  $(mSASSS_{n+1}-mSASSS_n)/(T_{n+1}-T_n)$



**Figure 1.** Time points for prediction of radiographic progression. The datasets including the clinical information of the first, second, and third visits were  $T_1$ ,  $T_2$ , and  $T_3$ , respectively. The radiographic progressions of the second, third, and fourth visits were  $P_2$ ,  $P_3$ , and  $P_4$ , respectively. mSASSS: modified stoke ankylosing spondylitis spine score.

combinations of the model's hyperparameters were investigated through grid search using the GridSearchCV library in the Scikit-learn package [16].

A LASSO regression model uses a linear combination of the selected features weighted by their respective coefficients for prediction. RF, a representative ensemble method, is widely used because it is powerful and lighter than other ensemble methods. RF constructs several tree-type base models and forms an ensemble through the bootstrap aggregating or bagging technique. XGBoost is a gradient-boosted decision tree algorithm for large datasets. Detailed hyperparameters of the three models in the three datasets are described in the supporting information (Supplementary File).

## Performance evaluation

We evaluated the prediction models in three rounds of three-fold cross-validation [19]. The operations, including z-normalization and machine learning classification, were executed separately on the training data during each cross-validation. Because of the unequal distribution of the progressor and non-progressor groups in the dataset, we used stratified cross-validation to divide the dataset. In each round, an entire dataset was randomly and equally divided into three parts with stratified probability. Two were used as the training dataset, and the third as the test dataset. The process was repeated three times in the three datasets in the three models. The one-point dataset for predicting radiographic progression at the second visit ( $T_1$  for  $P_2$ ) had 29 features, two-point dataset for predicting radiograph-

**Table 1.** Baseline characteristics in patients with non-progression and progression

Variable	Total patients (n=1,123)	Non-progressor (n=830)	Progressor (n=293)	p-value
Male	993 (88.42)	718 (86.51)	275 (93.86)	0.001
Age (yr)	32.01±9.41	30.98±9.46	34.93±8.65	<0.001
Eye involvement	363 (32.32)	245 (29.53)	118 (40.27)	<0.001
Peripheral involvement	401 (35.71)	319 (38.43)	82 (27.99)	0.002
HLA-B27	1,079 (96.08)	793 (95.54)	286 (97.61)	0.163
ALP (IU/L)	79.51±32.98	77.82±32.16	84.28±34.82	0.005
ALT (IU/L)	21.55±16.64	21.06±16.91	22.95±15.81	0.084
AST (IU/L)	19.96±9.24	19.94±9.36	20.00±8.89	0.921
Albumin (g/dL)	4.33±1.07	4.38±1.01	4.19±1.21	0.019
BUN (mg/dL)	12.94±4.69	13.14±4.54	12.37±5.06	0.022
CPK (IU/L)	96.40±231.63	99.21±243.98	88.43±192.56	0.444
CRP (mg/dL)	1.74±2.09	1.61±2.00	2.10 ±2.31	0.001
Cholesterol (mg/dL)	162.48±50.70	162.31±48.81	162.94±55.79	0.864
Creatinine (mg/dL)	0.83±0.31	0.83±0.22	0.83±0.48	0.999
ESR (mm/hr)	28.57±27.30	26.89±26.93	33.32±27.82	<0.001
GGT (IU/L)	14.97±30.47	13.52±26.00	19.09±40.31	0.028
Hb (g/dL)	13.40±3.17	13.46±3.02	13.24±3.57	0.346
Hct (%)	40.64±9.36	40.78±8.88	40.23±10.60	0.427
LDH (IU/L)	114.39 ±77.68	115.65±77.13	110.81±79.23	0.366
NSAIDs	880 (78.36)	650 (78.31)	230 (78.50)	0.987
bDMARDs	246 (21.91)	185 (22.29)	61 (20.82)	0.659
Methotrexate	151 (13.45)	122 (14.70)	29 (9.90)	0.049
Steroids	260 (23.15)	197 (23.73)	63 (21.50)	0.485
Sulfasalazine	283 (25.20)	228 (27.47)	55 (18.77)	0.004
mSASSS	14.57±16.28	12.36±16.07	20.84±15.25	<0.001

Values are presented as number (%) or mean±standard deviation. HLA: human leukocyte antigen, ALP: Alkaline phosphatase, AST: aspartate aminotransferase, ALT: alanine aminotransferase, BUN: blood urea nitrogen, CPK: creatine phosphokinase, CRP: C-reactive protein, ESR: erythrocyte sedimentation rate, GGT: gamma glutamyl peptidase, Hb: hemoglobin, Hct: hematocrit, LDH: lactate dehydrogenase, NSAIDs: nonsteroidal anti-inflammatory drugs, bDMARDs: biologic disease-modifying anti-rheumatic drugs, mSASSS: modified stoke ankylosing spondylitis spine score.

ic progression at the third visit ( $T_1+T_2$  for  $P_3$ ) had 53 features, and three-point dataset for predicting radiographic progression at the fourth visit ( $T_1+T_2+T_3$  for  $P_4$ ) had 77 features. Each average of the three models of the three-fold cross-validation in the one-point dataset is the estimated performance of the models. The same was the case for the other two datasets. We used the receiver-operator characteristics (ROC) to assess the predictive power of each predictor.

**Feature selection**

We performed feature importance analysis using RF and XGBoost to verify the robustness of the results. Features with greater contributions to the LASSO regression model were selected for analysis. Variable importance was evaluated using the model-based variable importance scores. The important variables (particularly those informative to radiographic progression) were captured when fitting the models to the training dataset [20,21].

**Statistical analysis**

For continuously distributed data, the results are shown as mean±standard deviation; between-group comparisons were

performed using Student’s t-test. Categorical or dichotomous variables were expressed as frequencies and percentages and were compared using the chi-squared test. Area under receiver operating characteristic curve (AUCs) were used to determine the diagnostic performance, with optimal thresholds of the clinical parameters determined by maximizing the sum of the sensitivity and 1–specificity, i.e., the Youden index values. Machine learning model training and statistical analysis were performed using Python (version 3.5.2; Python Software Foundation, Wilmington, DE, USA).

**RESULTS**

**Differences between non-progressor and progressor groups**

Out of the 1,280 patients, 157 lacked clinical and/or radiologic prescription and laboratory data; therefore, 1,123 patients were included in the study. The average time intervals between  $T_1$  and  $T_2$  and  $T_2$  and  $T_3$  were  $2.27\pm 1.38$  years and  $2.12\pm 1.58$  years, respectively. The baseline characteristics of the non-progressor and progressor groups at the first visit ( $T_1$ ) are shown in Table 1. The datasets of 1,123 patients at the first visit, 1,115 patients at

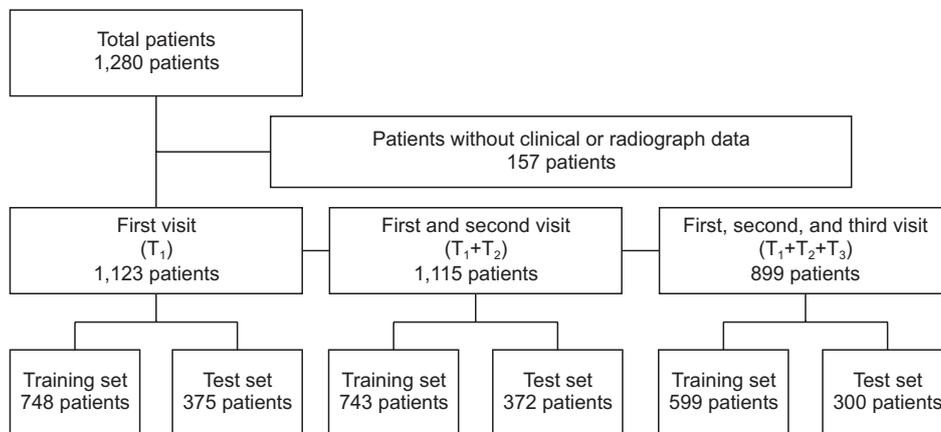


Figure 2. Flowchart of the study.

Table 2. Prediction performance evaluation according to time points and machine learning models

Prediction of radiographic progression ( $P_{n+1}$ ) with visit data ( $T_n$ )	LASSO and logistic regression				Random forest				XGBoost			
	Sensitivity (%)	Specificity (%)	Accuracy (%)	AUC	Sensitivity (%)	Specificity (%)	Accuracy (%)	AUC	Sensitivity (%)	Specificity (%)	Accuracy (%)	AUC
$P_2$ with $T_1$	68.25	68.31	68.3	0.7169	73.72	73.73	73.73	0.7959	70.99	70.84	70.88	0.7729
$P_3$ with $T_1+T_2$	66.18	66.3	66.27	0.6831	67.95	67.27	67.44	0.7467	66.21	66.3	66.28	0.7132
$P_4$ with $T_1+T_2+T_3$	61.39	60.03	60.4	0.6442	68.47	67.93	68.08	0.7348	66.8	67.94	67.63	0.7062

LASSO: least absolute shrinkage and selection operation, AUC: area under receiver operating characteristic curve.

the second visit, and 899 patients at the third visit were divided into training and test sets (Figure 2).

### Predicting radiographic progression with three time-point datasets

The radiographic progression was predicted using clinical data at the first, second, and third visits (Table 2). Among the machine learning models, the RF model exhibited the best performance, with higher mean sensitivity, mean specificity, mean accuracy, and mean AUC than those of the LASSO regression and XGBoost models. In the RF model,  $P_2$  with  $T_1$  dataset showed better performance compared to  $P_3$  with  $T_1+T_2$  dataset and  $P_4$  with  $T_1+T_2+T_3$  dataset.

The confusion matrix and ROC for the prediction of  $P_2$  with  $T_1$  dataset are shown in Figure 3A and 3B, respectively. In three-fold cross-validation, the mean sensitivity, specificity, and accuracy are 73.72%, 73.73%, and 73.73%, respectively. The mean AUC of three-fold cross-validation is 0.7959 (Supplementary Figures 1 and 2 show the confusion matrix and ROC of LASSO regression and XGBoost model for  $P_2$  with  $T_1$  dataset, Supplementary Figures 3~8 show the confusion matrix and ROC of three machine learning models in  $P_3$  with  $T_1+T_2$  dataset and  $P_4$  with  $T_1+T_2+T_3$  dataset).

### Importance of features for predicting radiographic progression

The variables in the first visit data contributing to radiographic progression prediction at the second visit using RF are listed in Figure 3C. The most important feature in three-fold cross-validation is the total mSASSS. The second and third most important features are age and ALP followed by CRP, cholesterol, ESR, hematocrit, and ALT. Drugs such as sulfasalazine and methotrexate, clinical features such as eye and peripheral involvement, sex, and HLA B27 contributed less to radiographic progression than laboratory findings. In the XGBoost model for  $P_2$  with  $T_1$ , mSASSS is the most important feature; however, drugs such as sulfasalazine and methotrexate also ranked high in feature importance (Supplementary Figure 2). In addition, feature importance was identified in the RF and XGBoost models in  $P_3$  with  $T_1+T_2$  (Supplementary Figures 4 and 5) and  $P_4$  with  $T_1+T_2+T_3$  (Supplementary Figures 7 and 8). Supplementary Table 1 shows the top and bottom 5 most important features of the RF and XGBoost models. For most models, mSASSS is the most important feature. In addition, variables related to baseline

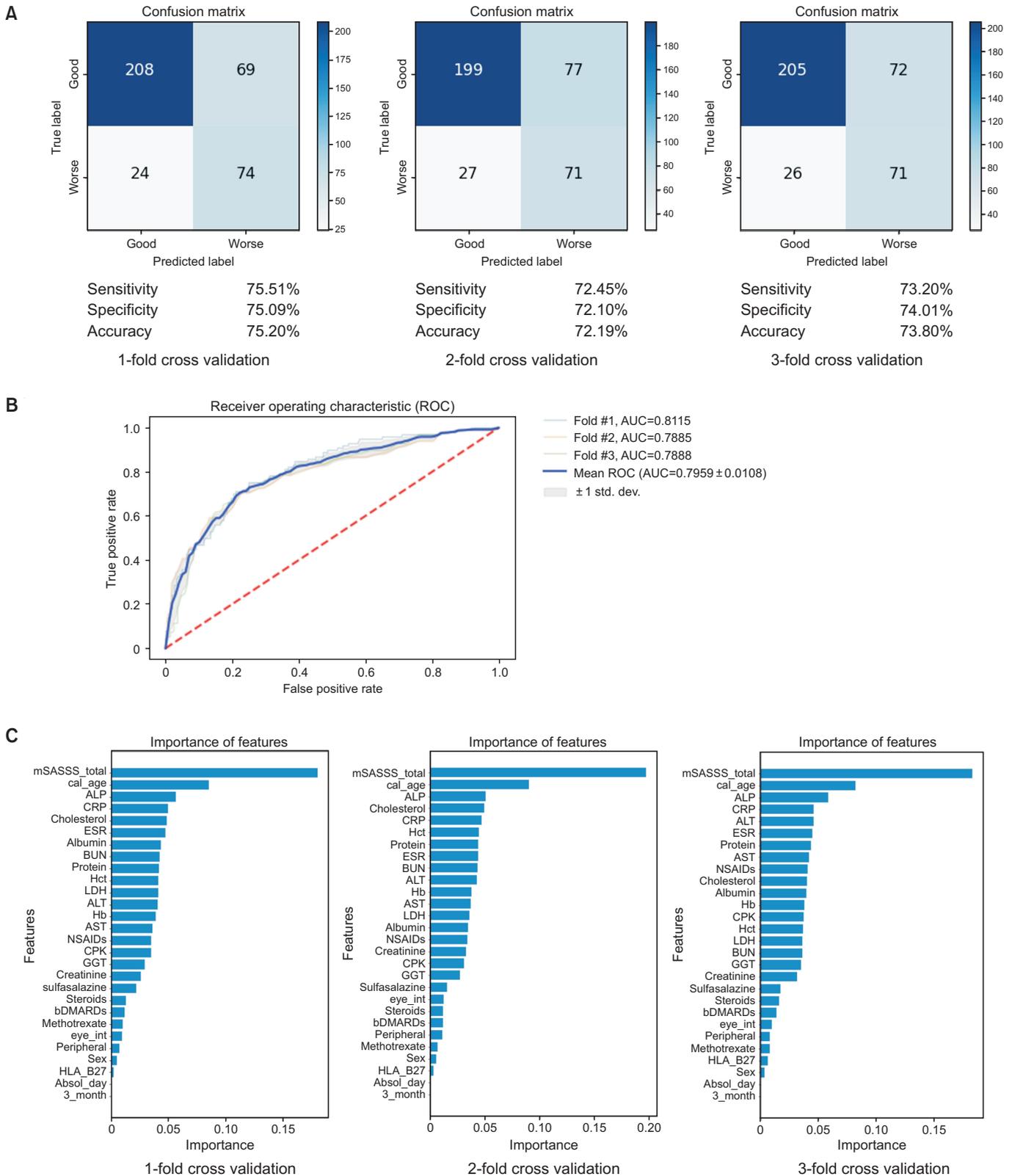
characteristics rank in the top 5.

## DISCUSSION

We developed a machine learning model that predicts radiographic progression using EMR data between January 2001 and December 2018. The RF model trained on data from the first visit predicted radiographic progression with an accuracy of 73.73% and an AUC of 0.7959, showing the best performance among the three models. Moreover, the accuracy and AUC decreased in the model trained with the second and third visit data. These results suggest that the data accumulated over an extended period did not increase the model performances, and the data from the first visit may contain important predictors for predicting radiographic progression in AS. Although the prediction model did not exhibit exceptionally high accuracy, this study is significant in identifying the data set among the three time points that predicts radiographic progression effectively and determining the essential features for prediction.

mSASSS, age, and CRP are ranked as highly important features, and their association with radiographic progression is well-known in statistical studies [15,22-25]. Interestingly, in our study, ALP ranked the highest in laboratory finding for predicting radiographic progression. ALP is produced in the liver, bone, and kidneys [26]. Bone and liver-specific isoforms of ALP form more than 90% of total serum ALP with a 1:1 ratio. In some studies, serum ALP activity was related to inflammatory markers in mineral metabolism [27,28]. In addition, serum ALP is associated with high disease activity, low bone mineral density, and high structural damage scores in patients with spondyloarthritis [29]. Therefore, radiographic progression may be associated with elevated serum ALP, particularly bone-specific ALP. In the future, statistical analysis will be conducted to prove the relationship between ALP and radiographic progression.

Statistical studies have linked radiographic progression to age, gender, inflammation, HLA B27, and smoking [12]. In this study, the baseline characteristics were important features in  $P_2$  and in predicting  $P_3$  and  $P_4$ . However, bDMARDs, such as tumor necrosis factor (TNF) inhibitors known to delay radiographic progression, did not belong to the top key features in five of the six models predicting radiographic progression. In this cohort, TNF inhibitors were used in patients initially refractory to treatment with NSAIDs and sulfasalazine. Because patients with long disease duration were included, bDMARDs



**Figure 3.** Prediction results with the random forest model. Confusion matrix (A), AUC (B), and importance of features in cross-validation (C). AUC: area under receiver operating characteristic curve, mSASSS: modified stoke ankylosing spondylitis spine score, ALP: alkaline phosphatase, CRP: C-reactive protein, ESR: erythrocyte sedimentation rate, BUN: blood urea nitrogen, Hct: hematocrit, LDH: lactate dehydrogenase, ALT: alanine transaminase, Hb: hemoglobin, AST: aspartate transaminase, NSAIDs: nonsteroidal anti-inflammatory drugs, CPK: creatine phosphokinase, GGT: gamma glutamyl peptidase, bDMARDs: biological disease-modifying antirheumatic drugs.

might have had little effect on radiographic progression.

Although machine learning models have recently been introduced to predict radiographic progression, disease activity, treatment response, and AS diagnosis [30-37], the performance of these models can vary due to differences in the type and quantity of data, hyperparameter tuning, and outcome settings. Walsh et al. [34,35,37] developed several models for AS diagnosis. In differentiating sacroiliitis, the developed model demonstrated an accuracy of 91.1% using text documents from the EMR [17]. Additionally, various algorithms applied to the same data showed an area under the receiver operating characteristic curve ranging from 0.86 to 0.96 to confirm axial spondyloarthritis [14]. For identifying AS, Deodhar et al. [33] developed a model using medical and pharmacy claim data, with a positive predictive value of 6.24%.

Joo et al. [38] predicted radiographic progression using machine learning on the training (n=253) and test sets (n=173). The balanced accuracy in the test set was above 65% in all models and 69.3% in RF, the highest of all models. In addition, the generalized linear model and support vector machine showed the best performance with an AUC of above 0.78. The outcomes of their study are similar to ours in predicting radiographic progression but with significant differences in detail. First, we examined machine learning-based prediction models for radiographic progression according to each visit using three time-point datasets containing EMR data accumulated over 18 years. Moreover, we used more time-series data and could identify clinical characteristics affecting radiographic progression at each time point. These results provide insight into the factors and timing that influence the prediction of radiographic progression in AS patients. In addition, the accuracy and AUC achieved in our study were higher. This difference in predictive power may be related to the difference in the amount of data and variables, such as limited features for bone marrow density and syndesmophyte score and additional laboratory findings.

We used time-series EMR data from the first, second, and third visits to predict radiographic progression at subsequent visits. Data from the first visit may be important clinical information related to radiographic progression. In addition, as treatment with NSAIDs started at the first visit, the disease activity index, such as the Bath AS Disease Activity Index, CRP, and ESR decreased subsequently. A decrease in the disease activity index, which leads to an increase in mSASSS [2-4], may have reduced the differences in important features between individuals. Thus,

the prediction performance may have deteriorated with datasets from the second and third visits.

Recurrent neural networks (RNNs) are also powerful models for learning and predicting temporal patterns and dependencies in data. We tried using RNN to study this dataset, but it did not train properly and was unsuitable for our problem setting, which involves irregular event sequences daily. As Che et al. [30] pointed out, irregular events pose a very challenging problem for RNNs, in terms of capturing temporal regularities. Moreover, our dataset lacks sequential data to perform RNN analysis effectively. Therefore, we organized data by time of patient visit in our dataset to predict the deterioration of the disease using a machine learning model that can better handle irregular event sequences. In the future, we will apply an appropriate deep learning model more suitable for predicting progression in this dataset.

The EMR data of AS patients accumulate over years or decades of follow-up. Radiographic progression with recurrent or chronic inflammatory status may be due to the delayed effects of clinical or environmental factors; for example, in AS, inflammation begins, ossifies, progresses to syndesmophyte, and is confirmed on radiographs. Although disease activity markers such as CRP or AS activity score, is an important predictor of radiographic progression [24,25], it need not be an absolute long-term factor determining radiographic progression. For example, the radiographic progression continues even when recurrent transient inflammations are actively controlled [31]. This evidence suggests that many important clinical factors influence radiographic progression. Unlike investigating numerous statistical associations, this study provides insight into the timing and factors important for predicting radiographic progression.

Several machine learning models using large datasets have been useful for diagnosing axial spondyloarthritis [32]. Those approaches can help in early diagnosis and reduce the social burden of diseases. Using a claim dataset, Deodhar et al. [33] suggested that machine learning models have a positive predictive value of 6.24% compared to the Assessment of SpondyloArthritis International Society classification criteria with a positive predictive value of 1.29%. In addition, machine learning models with EMR datasets have also shown good performance for early diagnosis of axial spondyloarthritis, with accuracies ranging from 82.6% to 91.8% [34-36]. It can be used for early diagnosis of AS by creating a machine learning model with image and text data because images such as radiographs are important in AS

diagnosis. The detection of sacroiliitis using X-ray, computed tomography, and magnetic resonance imaging using machine learning methods has been conducted recently with excellent performance in screening AS patients [37]. Therefore, developing a machine learning model for diagnosis by combining images, life-log, and clinical information is essential to improve diagnosis accuracy, which is worthy of future challenges for predicting radiographic progression in AS patients. Furthermore, an important task is assembling a representative and diverse dataset to meet the demands of high-performance machine learning models [39].

Despite the advantages, there are some limitations to our study. First, we applied three machine learning models to predict individual radiographic progression and identified the importance of features contributing to their prediction. Interpreting the importance of features is possible because previous statistical studies have shown the factors related to radiographic progression. Therefore, machine learning methods may complement statistical methods. However, additional statistical validation is needed to generalize important unknown features contributing to radiographic progression. Second, we used the EMR data from a single center. Validation using EMR data from various centers is required. Third, we used a machine learning model using EMR data at diagnosis and initial treatment. Therefore, this model can predict radiographic progression only when a patient first visits the hospital. While there exists a substantial correlation between disease activity and radiological alterations [25], the model falls short in accounting for the cumulative disease activity over time, primarily due to the absence of information on various aspects of disease activity spanning from the first visit to subsequent ones. In the future, developing a model that can predict radiographic progression at various time points will be necessary by advancing machine learning models. Fourth, there may be models using algorithms that are better than the machine learning models developed in this study. It is possible to try a better model using an artificial neural network, but it may become more difficult for clinical application owing to the limitations of the “black box” model. Fifth, it is important to note that smoking is a recognized factor associated with radiographic progression [15]. However, this study could not include smoking as a variable due to the absence of available information regarding smoking habits. Sixth, given the extensive 18-year duration of the data used in this study, it is imperative to consider the potential influence of changes in treatment proto-

cols and alterations in insurance coverage when interpreting the study's findings.

## CONCLUSION

Among the datasets, including for the first, second, and third visits, predicting the radiographic progression of the second visit using the first visit dataset resulted in the best performance, with the highest accuracy and AUC. Therefore, the clinical features of the first visit are likely to contain essential information for predicting radiographic progression. In terms of the importance of features, mSASSS, age, ALP, and CRP were ranked high. In addition to EMR data, various types of data, such as images and life-log, may be required to increase accuracy.

## SUPPLEMENTARY DATA

Supplementary data can be found with this article online at <https://doi.org/10.4078/jrd.2023.0056>

## FUNDING

This work was supported by the National Research Foundation of Korea (NRF) grant funded by the Korean government (MSIT) (No. NRF-2021R1C1C1009815) (BSK). The funder has played no role in the research. There was no additional external funding received for this study.

## ACKNOWLEDGMENTS

We would like to thank all members of Biomedical Engineering in Asan Medical Institute of Convergence Science and Technology.

## CONFLICT OF INTEREST

B.S.K. has been an editorial board member since May 2022, but has no role in the decision to publish this article. The authors declare that they have no competing interests.

## AUTHOR CONTRIBUTIONS

B.S.K., M.J., N.K., and T.H.K. contributed to the study conception and design. All authors contributed to data acquisition,

analysis, or interpretation. S.L. and K.B.J. scored the spinal radiographs independently. B.S.K., M.J., and N.K. were responsible for the statistical analyses. B.S.K. and M.J. drafted the manuscript and all coauthors were involved in critical revisions for maintenance of intellectual content. N.K. and T.H.K. provided administrative, technical, or material support. T.H.K. and N.K. had full access to all study data and takes responsibility for data integrity and data-analysis accuracy. All authors approved the final version to be submitted for publication.

## ORCID

Bon San Koo, <https://orcid.org/0000-0002-4212-2634>  
 Miso Jang, <https://orcid.org/0000-0003-4409-411X>  
 Ji Seon Oh, <https://orcid.org/0000-0002-0205-6492>  
 Keewon Shin, <https://orcid.org/0000-0002-5028-5716>  
 Seunghun Lee, <https://orcid.org/0000-0002-4348-7993>  
 Kyung Bin Joo, <https://orcid.org/0009-0004-1598-1670>  
 Namkug Kim, <https://orcid.org/0000-0002-3438-2217>  
 Tae-Hwan Kim, <https://orcid.org/0000-0002-3542-2276>

## REFERENCES

- Inman RD. Axial spondyloarthritis: current advances, future challenges. *J Rheum Dis* 2021;28:55-9.
- Brown MA, Li Z, Cao KL. Biomarker development for axial spondyloarthritis. *Nat Rev Rheumatol* 2020;16:448-63.
- Lorenzin M, Ometto F, Ortolan A, Felicetti M, Favero M, Doria A, et al. An update on serum biomarkers to assess axial spondyloarthritis and to guide treatment decision. *Ther Adv Musculoskelet Dis* 2020;12:1759720X20934277.
- Rademacher J, Tietz LM, Le L, Hartl A, Hermann KA, Sieper J, et al. Added value of biomarkers compared with clinical parameters for the prediction of radiographic spinal progression in axial spondyloarthritis. *Rheumatology (Oxford)* 2019;58:1556-64.
- Beam AL, Kohane IS. Big data and machine learning in health care. *JAMA* 2018;319:1317-8.
- Fontanella S, Cucco A, Custovic A. Machine learning in asthma research: moving toward a more integrated approach. *Expert Rev Respir Med* 2021;15:609-21.
- Beckmann JS, Lew D. Reconciling evidence-based medicine and precision medicine in the era of big data: challenges and opportunities. *Genome Med* 2016;8:134.
- Lee CH, Yoon HJ. Medical big data: promise and challenges. *Kidney Res Clin Pract* 2017;36:3-11.
- Rumsfeld JS, Joynt KE, Maddox TM. Big data analytics to improve cardiovascular care: promise and challenges. *Nat Rev Cardiol* 2016;13:350-9.
- van der Linden S, Valkenburg HA, Cats A. Evaluation of diagnostic criteria for ankylosing spondylitis. A proposal for modification of the New York criteria. *Arthritis Rheum* 1984;27:361-8.
- Creemers MC, Franssen MJ, van't Hof MA, Gribnau FW, van de Putte LB, van Riel PL. Assessment of outcome in ankylosing spondylitis: an extended radiographic scoring system. *Ann Rheum Dis* 2005;64:127-9.
- van der Heijde D, Braun J, Deodhar A, Baraliakos X, Landewé R, Richards HB, et al. Modified stoke ankylosing spondylitis spinal score as an outcome measure to assess the impact of treatment on structural progression in ankylosing spondylitis. *Rheumatology (Oxford)* 2019;58:388-400.
- Koo BS, Oh JS, Park SY, Shin JH, Ahn GY, Lee S, et al. Tumour necrosis factor inhibitors slow radiographic progression in patients with ankylosing spondylitis: 18-year real-world evidence. *Ann Rheum Dis* 2020;79:1327-32.
- Lee TH, Koo BS, Nam B, Oh JS, Park SY, Lee S, et al. Conventional disease-modifying antirheumatic drugs therapy may not slow spinal radiographic progression in ankylosing spondylitis: results from an 18-year longitudinal dataset. *Ther Adv Musculoskelet Dis* 2020;12:1759720X20975912.
- Haroon N, Inman RD, Learch TJ, Weisman MH, Lee M, Rahbar MH, et al. The impact of tumor necrosis factor inhibitors on radiographic progression in ankylosing spondylitis. *Arthritis Rheum* 2013;65:2645-54.
- Pedregosa F, Varoquaux G, Gramfort A, Michel V, Thirion B, Grisel O, et al. Scikit-learn: machine learning in Python. *J Mach Learn Res* 2011;12:2825-30.
- Breiman L. Random forests. *Mach Learn* 2001;45:5-32.
- Sheridan RP, Wang WM, Liaw A, Ma J, Gifford EM. Extreme gradient boosting as a method for quantitative structure-activity relationships. *J Chem Inf Model* 2016;56:2353-60. Erratum in: *J Chem Inf Model* 2020;60:1910.
- Kim JH. Estimating classification error rate: repeated cross-validation, repeated hold-out and bootstrap. *Comput Stat Data Anal* 2009;53:3735-45.
- Greenwell BM, Boehmke BC, McCarthy AJ. A simple and effective model-based variable importance measure. *ArXiv [Preprint]*. 2018 [cited 2021 Jun 15]. Available from: <https://doi.org/10.48550/arXiv.1805.04755>.
- Kuhn M, Johnson K. *Applied predictive modeling*. New York, Springer, 2013.
- Poddubnyy D, Haibel H, Listing J, Märker-Hermann E, Zeidler H, Braun J, et al. Baseline radiographic damage, elevated acute-phase reactant levels, and cigarette smoking status predict spinal radiographic progression in early axial spondylarthritis. *Arthritis Rheum* 2012;64:1388-98.
- Poddubnyy D, Protopopov M, Haibel H, Braun J, Rudwaleit M, Sieper J. High disease activity according to the Ankylosing Spondylitis Disease Activity Score is associated with accelerated radiographic spinal progression in patients with early axial spondyloarthritis: results from the German Spondyloarthritis Inception Cohort. *Ann Rheum Dis* 2016;75:2114-8.
- Poddubnyy DA, Rudwaleit M, Listing J, Braun J, Sieper J. Comparison of a high sensitivity and standard C reactive protein measurement in patients with ankylosing spondylitis and non-radiographic

- axial spondyloarthritis. *Ann Rheum Dis* 2010;69:1338-41.
25. Ramiro S, van der Heijde D, van Tubergen A, Stolwijk C, Dougados M, van den Bosch F, et al. Higher disease activity leads to more structural damage in the spine in ankylosing spondylitis: 12-year longitudinal data from the OASIS cohort. *Ann Rheum Dis* 2014;73:1455-61.
  26. Haarhaus M, Brandenburg V, Kalantar-Zadeh K, Stenvinkel P, Magnusson P. Alkaline phosphatase: a novel treatment target for cardiovascular disease in CKD. *Nat Rev Nephrol* 2017;13:429-42.
  27. Cheung BM, Ong KL, Cheung RV, Wong LY, Wat NM, Tam S, et al. Association between plasma alkaline phosphatase and C-reactive protein in Hong Kong Chinese. *Clin Chem Lab Med* 2008;46:523-7.
  28. Damera S, Raphael KL, Baird BC, Cheung AK, Greene T, Beddhu S. Serum alkaline phosphatase levels associate with elevated serum C-reactive protein in chronic kidney disease. *Kidney Int* 2011;79:228-33.
  29. Kang KY, Hong YS, Park SH, Ju JH. Increased serum alkaline phosphatase levels correlate with high disease activity and low bone mineral density in patients with axial spondyloarthritis. *Semin Arthritis Rheum* 2015;45:202-7.
  30. Che Z, Purushotham S, Cho K, Sontag D, Liu Y. Recurrent neural networks for multivariate time series with missing values. *Sci Rep* 2018;8:6085.
  31. Koo BS, Lee S, Oh JS, Park SY, Ahn GY, Shin JH, et al. Early control of C-reactive protein levels with non-biologics is associated with slow radiographic progression in radiographic axial spondyloarthritis. *Int J Rheum Dis* 2022;25:311-6.
  32. Walsh JA, Rozycki M, Yi E, Park Y. Application of machine learning in the diagnosis of axial spondyloarthritis. *Curr Opin Rheumatol* 2019;31:362-7.
  33. Deodhar A, Rozycki M, Garges C, Shukla O, Arndt T, Grabowsky T, et al. Use of machine learning techniques in the development and refinement of a predictive model for early diagnosis of ankylosing spondylitis. *Clin Rheumatol* 2020;39:975-82.
  34. Walsh JA, Pei S, Penmetsa G, Hansen JL, Cannon GW, Clegg DO, et al. Identification of axial spondyloarthritis patients in a large dataset: the development and validation of novel methods. *J Rheumatol* 2020;47:42-9.
  35. Walsh JA, Pei S, Penmetsa GK, Leng J, Cannon GW, Clegg DO, et al. Cohort identification of axial spondyloarthritis in a large healthcare dataset: current and future methods. *BMC Musculoskelet Disord* 2018;19:317.
  36. Walsh JA, Shao Y, Leng J, He T, Teng CC, Redd D, et al. Identifying axial spondyloarthritis in electronic medical records of US veterans. *Arthritis Care Res (Hoboken)* 2017;69:1414-20.
  37. Bressemer KK, Vahldiek JL, Adams L, Niehues SM, Haibel H, Rodriguez VR, et al. Deep learning for detection of radiographic sacroiliitis: achieving expert-level performance. *Arthritis Res Ther* 2021;23:106.
  38. Joo YB, Baek IW, Park YJ, Park KS, Kim KJ. Machine learning-based prediction of radiographic progression in patients with axial spondyloarthritis. *Clin Rheumatol* 2020;39:983-91.
  39. Rajkomar A, Dean J, Kohane I. Machine learning in medicine. *N Engl J Med* 2019;380:1347-58.