

Expert Review

Narrative Review of Machine Learning in Rheumatic and Musculoskeletal Diseases for Clinicians and Researchers: Biases, Goals, and Future Directions

Amanda E. Nelson¹  and Liubov Arbeeva² 

ABSTRACT. There has been rapid growth in the use of artificial intelligence (AI) analytics in medicine in recent years, including in rheumatic and musculoskeletal diseases (RMDs). Such methods represent a challenge to clinicians, patients, and researchers, given the “black box” nature of most algorithms, the unfamiliarity of the terms, and the lack of awareness of potential issues around these analyses. Therefore, this review aims to introduce this subject area in a way that is relevant and meaningful to clinicians and researchers. We hope to provide some insights into relevant strengths and limitations, reporting guidelines, as well as recent examples of such analyses in key areas, with a focus on lessons learned and future directions in diagnosis, phenotyping, prognosis, and precision medicine in RMDs.

Key Indexing Terms: arthritis, musculoskeletal system, rheumatic diseases

Artificial intelligence (AI; Figure 1) and its subcategory machine learning (ML) have rapidly gained traction as analytic methods in a variety of conditions including rheumatic and musculoskeletal diseases (RMDs).¹ These terms have seemingly taken over the medical literature in recent years, but often in a way that is not readily accessible to most clinicians or researchers. Beam and Kohane provided a very useful perspective on AI/ML in 2018 as part of a spectrum from fully human-guided analysis and decision making to fully automated network-based algorithms.² They sagely noted that AI/ML provides “...no guarantees of fairness, equitability, or even veracity.”²

In 2020, the European Alliance of Associations for Rheumatology (EULAR) endorsed principles relating to the use of big data (defined as large, complex, and/or multidimensional, from heterogeneous sources) in RMDs.³ These include an imperative to consider ethical issues and an overarching goal to use big

data to improve the lives of patients with RMDs. Key points focused on the need for harmonized standards and the FAIR principle (Findable, Accessible, Interoperable, and Reusable), open data platforms with privacy considerations and interdisciplinary collaboration, use of explicit reporting of methods, benchmarking of computational methods, and independent validation, along with interdisciplinary training in big data for clinicians and scientists from various backgrounds.³

A variety of recent reviews have focused on the use of ML in RMDs, including an overview of definitions and performance characteristics of ML, a set of representative clinical studies through early 2021,⁴ and a more technical overview of definitions, methods, classification procedures, prediction models, and algorithms⁵; these reviews noted that most datasets are not purpose-built and thus lack necessary sample size (SS) as well as novel features. Additionally, there are several reviews focused specifically on the role of ML in imaging, including in RMDs,⁶⁻¹⁰ so this large topic will not be reviewed here. Therefore, rather than providing a systematic or technical review, we refer the reader to these publications,⁴⁻¹⁰ and instead provide a narrative overview of recent work. The goal of this review is to serve as an introduction to the area of AI/ML for clinicians and researchers in RMDs who are new to this field (see also Hugel et al for a good introduction to types of AI/ML in rheumatology¹¹); to improve understanding around how incorporating these methods might benefit their work, which data types might be useful in AI/ML analyses (Figure 2), and how they might work with collaborators; and to provide examples of work in key areas. These key areas include (1) diagnosis, (2) phenotyping, (3) prognosis, (4) precision medicine, (5) limitations and biases, and (6) future directions.

Support in the form of grants or industrial support was provided by the National Institutes of Health/National Institute of Arthritis and Musculoskeletal Diseases (NIH/NLAMS; P30AR072580 and R21AR074685).

¹A.E. Nelson, MD, MSCR, Department of Medicine, Division of Rheumatology, Allergy, and Immunology, University of North Carolina at Chapel Hill; ²L. Arbeeva, MS, Thurston Arthritis Research Center, University of North Carolina at Chapel Hill, Chapel Hill, North Carolina, USA.

The authors declare no conflicts of interest relevant to this article.

Address correspondence to Dr. A.E. Nelson, 3300 Thurston Building, Campus Box #7280, Chapel Hill, NC 27599-7280, USA.

Email: aenelson@med.unc.edu.

Accepted for publication June 21, 2022.

Rheumatic and Musculoskeletal Diseases (RMDs): A diverse group of more than 200 diseases that often affect joints but can affect any organ, often caused by problems in the immune system, or from inflammation or infections of joints, and can result in significant disability.¹

Artificial Intelligence (AI): The development of computer systems that can perform tasks normally requiring human intelligence.

Machine Learning (ML): A discipline within AI where computer algorithms are developed to learn and make decisions based on data.

Figure 1. Definitions and abbreviations of key terms in this review.

DATA SOURCES
 Individual patients
 Observational cohorts
 Clinical cohorts
 Clinical trials
 Electronic health records
 National reimbursement databases
 Registries

DATA TYPES (standard)
 Clinical/demographic
 Patient reported outcomes
 Laboratory data
 Tissue analysis
 Imaging
 Medications
 Clinical notes
 Disease activity

DATA TYPES (specialized)
 Genetics
 Proteomics, other omics
 Nutrition
 Wearables

Figure 2. Examples of data sources and types in AI/ML approaches. AI: artificial intelligence; ML: machine learning.

AI/ML for diagnosis: Identifying the condition of interest in a patient or cohort

AI/ML methods can assist with a variety of diagnostic challenges, including in the clinical setting based on available lab and clinical data, identification of affected patients in the electronic health record (EHR), or optimal selection of clinical study participants.

One of the main interests for clinicians and their patients is the definition of the disease state. As many RMDs are rare diseases, by definition affecting fewer than 1 in 2000 individuals, this can be particularly challenging; however, ML holds specific promise to improve strategies and develop new drugs for the treatment of rare diseases. Data from genomic and multiomic approaches have provided new insights, as have other big data like gait assessments and imaging.¹² In a recent scoping review, most studies of rare diseases employing ML were focused on diagnosis or prognosis, and many suffered from small sample

sizes and a lack of external validation.¹³ Registries can increase sample sizes for even the rarest of conditions, although harmonization is needed for these to be useful. Open data sources have similar limitations and may have poorer reliability than carefully curated data sources. Possible solutions include enhancing SS with the incorporation of unlabeled non-case samples outside the rare disease of interest, artificially built samples, and transfer learning.¹²

Systemic lupus erythematosus (SLE) is an example of a rare disease and RMD that has been investigated using ML methods given the challenges of making this diagnosis in clinical practice. The SLE Risk Probability Index has been proposed to assist with diagnosis in the clinical setting, improving time to diagnosis and treatment in SLE.¹⁴ Clinical guidance was used to create 20 feature panels, each of which were submitted to random forest (RF) and least absolute shrinkage and selection operator (LASSO) penalized regression, resulting in 40 models trained on data from 2 SLE registries. The model with highest accuracy was evaluated in a validation cohort and converted into a scoring system, using a threshold of 7 to separate SLE vs other RMD in adults, and adjusted to 8 in a follow-up study in a pediatric SLE cohort.¹⁵ Although using a relatively small sample size and retrospective design, this work exemplifies the importance of internal and external validation in ML-based diagnostic algorithms. Another study used both structured and narrative data to identify patients with SLE in EHR data.¹⁶ They selected definite cases, probable cases, and definite noncases by chart review to determine the positive predictive value (PPV) of the algorithms and features in internal and external cohorts. The ML algorithm had a 92% PPV for definite/probable SLE in the internal cohort, and 94% in an external cohort, comparing favorably to the PPV of using 1 or 2 International Classification of Diseases, 9th/10th revision codes, which was cited as around 50%. The EHR phenotyping protocol is published and available for use in clinical and translational studies.¹⁷ The performance characteristics of previously published algorithms were also tested, demonstrating the importance of adjusting for portability (ie, their application in other systems); several challenges were identified, such as different medical billing practices, medication prescribing and reporting, and disease prevalence.¹⁶

AI/ML for phenotyping: Defining important subtypes of disease

Individuals with RMDs have variable courses, including rates of progression, transition to other conditions, and response to treatments.

Molecular phenotyping is an area of growing interest, given the substantial overlap in clinical features and the lack of specific diagnostic studies for many RMDs. Myositis is a good example, considering the multiple antibodies and clinical phenotypes within inflammatory myositis and our growing understanding of their effect on outcomes. Muscle biopsies were collected from 119 patients enrolled in several key myositis cohorts and 20 healthy controls. The myositis cohorts included those with myositis-specific autoantibodies, anti-synthetase syndrome, necrotizing myopathy, or inclusion body myositis.¹⁸ Ten different ML algorithms, including decision trees and RF, among others, were trained using transcriptomic data to determine disease-specific gene expression patterns. This allowed for accurate identification of subgroups in over 90% of muscle biopsies using the linear support vector machine model.¹⁸ Although the sample size was small, the use of biopsy data was one of the strong points of this study and demonstrated the usefulness of objective transcriptomics in the interpretation of tissue biopsies. These markers may be useful to tailor therapies to a specific molecular diagnosis in the future.

Juvenile-onset SLE is a rare RMD that is challenging to study and often reliant on small cohorts, particularly for subgrouping. The choice of ML algorithms that can potentially address difficulties associated with rare diseases is particularly important. Cross-validation is of value in settings without a readily available validation cohort. Robinson et al applied supervised ML approaches for classification (ie, discrimination of 67 SLE patients from 39 healthy controls) and selection of important variables, including immune cell profiles.¹⁹ These variables were further used in an unsupervised k-means clustering that identified 4 potentially important subgroups among patients with SLE.¹⁹ Limitations of this study included the small sample size, low number of Black patients, and imperfect outcome measures. However, such immune-based phenotyping may improve patient stratification for future clinical studies and may eventually inform clinical practice.

AI/ML for prognosis: Defining disease course for targeted intervention

To better identify those at risk and take appropriate action, it is important to know which individuals are most likely to worsen rapidly and which may experience improvement or resolution of their condition.

While clinical disease activity measures are available and validated in most RMDs, their application and use in clinical practice is inconsistent. ML can be used to estimate these values from available clinical information to identify patients with active disease in clinical datasets. An earlier study in this area used only structured data (eg, laboratory values, existing Clinical Disease Activity Index [CDAI] scores, and medications) from the EHRs of 2 distinct clinical settings to build a deep learning model to predict CDAI in patients with rheumatoid arthritis (RA).²⁰ More than 20 variables were significantly important for accuracy of the predictions. This paper demonstrates both strong statistical design and a detailed discussion of limitations associated with such data, including missing values, subsequent

biases, and the usefulness of these models in clinical practice. Another approach is to use unstructured data, such as clinical notes, to predict other quantifiable outcomes. Alves et al developed natural language processing algorithms followed by validation to estimate SLE Disease Activity Index categories in SLE from unstructured notes.²¹ This approach was validated to estimate CDAI scores in RA.²² Both groups were able to estimate disease activity with area under the curve (AUC) around 0.9 and correlation with true clinical scores around 0.7. The ability to estimate disease activity in the absence of clinician-entered scores would dramatically increase the data available for research use, providing large numbers of patients for outcomes research, or for metaanalyses, and potentially reducing disparities by provider or clinic. Such algorithms would still require sufficient input for estimation and may not be applicable to all settings (eg, handwritten notes, international/low resource settings). A combination of structured and unstructured data is likely optimal for the prediction of prognosis in clinical datasets.

It is important to identify patients who are most likely to progress or even to develop disease. One such circumstance is patients with undifferentiated arthritis, some of whom go on to develop RA and some of whom do not. A small study assessed the DNA methylome of patients with undifferentiated arthritis ($n = 72$), where about half remained stable and half developed RA after a year, and identified differential methylation between groups.²³ Both supervised and unsupervised methods were used along with internal and external validation. Distinct methylation patterns were seen among those who did develop RA, those who did not develop RA, as well as those in a separate group with RA at baseline, demonstrating the potential of methylation markers to sense early disease determinants in these patients.²³ Despite the small sample size, this work highlights the possibility of incorporating basic and clinical data for clinically relevant risk assessment.

Data from randomized clinical trials (RCTs) can be used to identify predictors of prognosis, as such studies include well-phenotyped individuals, balanced across treatment groups at baseline, who are followed over time. Pooling individual data across different trials, while appropriately addressing heterogeneity,²⁴ can increase sample size, but this data source provides a limited number of variables. One such study pooled data from nearly 1900 patients with psoriatic arthritis enrolled in 4 RCTs to determine subgroups of response trajectory to secukinumab therapy over 52 weeks.²⁵ They applied model-based clustering methods to identify 7 clusters of participants, where patients within a cluster had a common distribution of 206 baseline measures; this procedure was repeated on 200 different subsets to access cluster stability. The clusters, characterized according to longitudinal responses, were clinically interpretable, with features such as higher polyarticular disease burden, greater foot symptoms, more dactylitis, or more nail and skin involvement.²⁵ The overall population was skewed as a result of the RCT design, with a high proportion of active polyarticular disease compared to other subtypes (such as oligoarticular involvement). However, this type of work could be used to inform trial selection for specific therapeutics or dosing regimens in future RCTs. In another study, data from several RCTs were examined and a remission prediction

score for patients with RA treated with tocilizumab was developed and validated.²⁶ Importantly, this prediction rule was subsequently tested in registry real-world data with an extended set of variables in a follow-up paper,²⁷ finding that the RCT model could similarly predict discrimination in the registry data (with AUC ~0.7 to 0.8). Both studies are excellent examples of robust design and rigorous statistical analysis.

There has been a great deal of interest in predicting progression in osteoarthritis (OA), one of the most common RMDs. An objective endpoint of total joint replacement (TJR) can be used to overcome some of the challenges of subjective pain outcomes and discordance with imaging in OA, although this is complicated by issues of preference, practice variability, and access to care. Jamshidi et al used baseline data from the publicly available Osteoarthritis Initiative (OAI) dataset to predict TJR at 96 months.²⁸ Using a LASSO method to select features followed by multiple ML models, they could predict time to TJR with high accuracy (AUC 0.9). Given the nature of the OAI cohort, which included only people with or at risk for OA, these results are not generalizable to the general population, and only features known to be associated with OA were included in the dataset, so there was no opportunity to discover novel features, a challenge for many existing cohorts.

The IMI-APPROACH (Applied Public-Private Research enabling Osteoarthritis Clinical Headway) study used a novel selection method to identify individuals most likely to progress.²⁹ This group developed algorithms in 2 large OA cohort studies to best classify patients who could be considered “progressors” (and avoid selection of likely “nonprogressors”), potentially improving efficiency for future clinical studies. This procedure was subsequently used to preselect likely progressors (those with a high likelihood of developing joint space loss or pain) from existing OA cohorts,³⁰ resulting in 297 participants to be followed for 2 years. Their inclusion was decided using RF and other supervised ML models that provided the probability of progression based on structure and/or pain within the lifetime of the study.³¹ To improve the performance of RF, a single model was trained to assign pain progression and structure progression labels independently (multilabel classification), while a duo classifier was used for 2 independent models, each trained to predict a single label (pain or structure progression).³¹ As a purpose-built cohort designed for the application of ML methods, this work is an important step forward in OA.

AI/ML for precision medicine: Using data to guide therapy and avoid adverse events

Several recent publications in RMDs reflect the goals of precision medicine, which can be understood as the provision of the right treatment,³² at the right dose,³³ to the right person, at the right time,³⁴ while minimizing unnecessary testing, side effects and overuse issues, including opioid use and abuse,³⁵⁻³⁷ specifically opioid use around TJR,³⁸⁻⁴⁰ and to explore issues of inequity in classification.⁴¹

Prediction of clinical response among patients with RMD, and thus the ability to make an informed decision about optimal

treatment recommendations, has long been a goal of clinicians and researchers. Using 275 baseline variables from Pournara et al,²⁵ a separate analysis employed Bayesian elastic net, which is useful for a large number of potentially correlated patient characteristics, to determine predictors of 16-week outcome based on starting dose of secukinumab in psoriatic arthritis.³³ While still limited by RCT data and the need for validation, this work provides insight relevant to precision medicine in RMDs. Another study, in a small cohort of 39 women with RA starting anti-TNF therapy, researchers assessed differences in multiomics from peripheral blood mononuclear cells (PBMCs) among EULAR responders and nonresponders at 3 months,⁴² although ML methods were not fully integrated into this analysis.

A preliminary study aiming to predict the 6-month clinical response to adalimumab and etanercept was undertaken in 80 patients with RA enrolled in an observational cohort in the Netherlands as they started biologic treatment.³² The investigators obtained PBMCs prior to biologic therapy and performed genome-wide expression and DNA methylation assays, which demonstrated different signatures in those who eventually responded to therapy. RF models using these multiomics data had > 80% accuracy for prediction of response.³² Several internal cross-validation techniques were used, although the validation and training sets were from the same sample.⁴³ The key strength of this study is the incorporation of true multiomics data and integrated data analysis in the prediction models. Future work will benefit from larger samples with robust outcomes and truly independent external validation sets to avoid overfitting and to mitigate feature instability, which is often challenging in rare diseases. Another example is a study that used consortium data to develop an algorithm to predict methotrexate response in patients with early RA (n = 643).³⁴ An RF model was trained on UK patients (n = 336) and externally validated on independent, non-UK patients from Sweden and the Netherlands (n = 307). Overfitting and class imbalance were directly addressed; however, the sample included only White Europeans, so generalizability remains limited. The incorporation of genetic data in the prediction algorithm substantially improved prediction accuracy, supporting the feasibility of pharmacogenomic markers for precision medicine, although the overall response rate remained low.³⁴

We used 24 ML algorithms to select the optimal model and to develop individualized treatment rules based on RCT data from the Intensive Diet and Exercise for Arthritis (IDEA) trial.⁴⁴ IDEA randomized overweight or obese individuals with symptomatic knee OA to 3 groups: exercise alone, diet alone, or a combination of diet plus exercise.⁴⁵ Using data from 343 participants and multiple outcome RF and list-based models, subgroups of participants were identified who would have improved outcomes for weight loss and for IL-6 (an inflammatory cytokine) if they had been assigned according to the decision rule rather than to the diet plus exercise intervention using value functions.⁴⁴ This work highlights the use of RCT data from a nonpharmacologic trial, exploration of multiple features and outcomes, and multiple model evaluation, all of which could improve the design of future studies.

Limitations and biases in AI/ML

Here we discuss several key issues including (1) bioethics, (2) missing data, (3) model bias, and (4) translation.

Bioethics. A recent excellent piece on bioethics in big data and RMD research identified 4 main areas of potential concern: privacy, informed consent, impact on the medical profession, and justice.⁴⁶ First, privacy and confidentiality are a challenge when large datasets are linked, as the detailed information that results could increase the risk of reidentification even when the datasets themselves are deidentified or even fully public. These may not even be considered human subjects data, but they can still be used to extract sensitive information. The authors astutely recommend the use of an honest broker to maintain and distribute data, thus avoiding providing full access to any potentially interested entity (eg, private funders, industry). Second, the nature of these big data analytics means that future developments, potential uses, and consequences are not known at the time of data collection, making fully informed consent a challenge to participants⁴⁷ and investigators, as well as institutional review boards and ethics committees. Third is the potential effect on the medical profession; that is, if an algorithm makes a mistake that causes harm, who is responsible? Thus, ML analytics carry the potential to undermine the physician-patient relationship. This leads to the fourth area of concern, justice, reflected in the potential for these technologies to worsen the existing digital divide as well as local and global health disparities. The risk of security breaches and hacking are higher in areas with lower health literacy, greater corruption, or rapid technology expansion without appropriate oversight, further placing underserved populations at risk.⁴⁶

Missing data. Considerations around health equity in relation to ML and big data have recently gained more attention, including in the study of RMDs. It is essential to consider who is in the dataset, who is not, and why not, as well as the effect these missing data may have on results from an ML analysis. For example, missing data could represent inconsistent care, an issue that more often affects individuals of low socioeconomic status, those with mental health issues, or immigrant populations. The existence of multiple care instances in a single EHR is often required for diagnostic algorithms and thus may exclude these individuals. Such missing data are not random, leading to potentially erroneous inferences from models that assume random missingness.⁴⁸ Individuals of lower socioeconomic status may already receive suboptimal care; failure to recognize this could result in an algorithm that preferentially directs these patients to inadequate care.⁴⁸ A lack of health care is not equivalent to lower disease burden but could be interpreted as such by an ML algorithm lacking appropriate context. Use of proxies for health, such as mortality, readmission, or cost can introduce biases owing to unequal access to care, resulting in underestimated illness burden and, potentially, in further inequities in access to care.⁴⁹ Over-the-counter medications are often missing or incompletely reported in EHRs and national reimbursement databases, and more accurate prescription dispensation data may require linkage to pharmacy or other databases to get a more complete picture of what patients are taking.⁵⁰ EHRs often lack data on

social determinants that might improve the ability of the ML algorithm to identify such equity issues. Similarly, race/ethnicity and preferred language may be missing or incorrect, leading to misclassification.⁴⁸ Specific analyses focused on addressing these issues, including subgroup analysis, stratification, and validation in a representative cohort, should be considered.⁵⁰ Importantly, in addition to avoiding potential harm, attention to fairness can also help identify areas of greatest need and lead to improved equity.⁵¹

Model bias. Given the subjectivity of the model selection process, there is an obvious need for both clinical/provider and patient input in making these decisions.⁵² The inclusion of patient collaborators in RMD research, including when using big data/ML applications, is important and may help address some of these issues.⁵² In the authors' experience, most papers using big data or ML methods state, without evidence, that it is somehow not possible or not reasonable to involve patient collaborators because of the nature of the work.

There are of course many other potential sources of bias in ML models.⁵³ A systematic review of prediction models using supervised ML methods found that the vast majority of the approximately 150 studies reviewed were at high risk of bias for a few key reasons, including an inadequate number of events per predictor and overfitting—issues that have not improved in the literature over time.⁵⁴ Another study focused on biases in observational clinical studies in secondary databases, identified confounding, selection bias, and measurement bias as the most reported, and provided a detailed summary table⁵⁵ as well as guidance regarding potential ways to address these issues. That ML algorithms can pick up on noninformative features and incorrectly interpret them is well described, such as prioritizing studies marked as urgent or “stat,” or recognizing features indicative of portable vs departmental imaging.⁵⁰ Investigators may be concerned by sample size, resulting in lack of consideration of potentially important subgroups in the data that are smaller in number, thus affecting prediction for underrepresented or minority groups.⁵⁰

Temporal data drift is an uncommonly discussed limitation to the generalizability of ML algorithms that can have substantial implications.^{50,56} A systematic review focused on approaches to mitigate the effects of temporal shift found only 15 papers explicitly covering this topic in clinical areas,⁵⁷ although this phenomenon is better studied and appreciated in nonclinical work.⁵⁸ Temporal shifts can occur at the patient (demographics, referrals, new diseases), practice (trial or guideline results, practice patterns, drug/test availability, reimbursement policies), or administration (EHR modification, vendor, coding system and practices) level and can affect performance and reproducibility. Strategies to address this issue, as well as those to be developed in the future, would benefit from a rigorous benchmarking procedure to best characterize impact and solutions.⁵⁷

“All models are wrong, but some are useful.” This aphorism is often used to emphasize the importance of acknowledgment of limitations, assumptions, and potential biases relevant to the analyses being used, whether in ML or more traditional statistical methods. For the researcher new to the area, awareness of potential bias and limitations is important. Use of reporting guidelines

such as TRIPOD (Transparent Reporting of a multivariable prediction model for Individual Prognosis or Diagnosis),^{59,60} or critical appraisal tools like PROBAST (Prediction model Risk of Bias Assessment Tool)⁶¹ may be useful to avoid, address, and identify such potential biases, and are required by many peer-reviewed journals. In recognition of the specific issues around prediction models using AI and ML, extensions of these tools, TRIPOD-ML, TRIPOD-AI (reporting guideline), and PROBAST-AI (critical appraisal tool), are currently under development.⁶⁰ Other tools developed for fields such as cardiology⁶² and orthopedics,⁶³ or for clinical trials^{64,65} are also available.

Translation. There are a variety of challenges with the translation of AI/ML to clinical practice, many of which are directly related to the challenges mentioned above.⁵⁰ Selection of reliable outcomes is essential but can be challenging. In addition, the frequent (and understandable) use of retrospective studies to develop algorithms will result in better performance metrics compared to application to prospective, real-world data, making their implementation difficult and potentially unreliable.⁵⁰ It is essential that studies with the goal of eventual clinical adoption be rigorously performed, appropriately reported, and peer-reviewed; many such studies are published only as pre-prints.⁵⁰ The development of understandable and clinically relevant assessments of model performance, reflecting its practical importance is a key in the clinical realm. As noted above and throughout this review, it is difficult to compare algorithms because of methodologies, populations, sample distributions and characteristics, and differing performance metrics, again highlighting the need for independent test sets and large open datasets for validation and benchmarking.⁵⁰

Future directions for AI/ML in RMDs. A variety of other clinical uses for AI/ML are also in various stages of development, but space precludes extensive discussion, including digital health, smart technology, wearables, care algorithms, and monitoring of adherence.⁶⁶ Wearables are of particular interest given the potential for continuous monitoring.⁶⁷ These types of data could allow for automated alerts to patients or their physicians, direct patient feedback, and/or algorithm-based automatic interventions,⁶⁷ while providing an opportunity to increase access to care and potentially improve monitoring and outcomes.⁶⁶ An obvious limitation, in addition to cost and use of the technology itself, is the need for enhanced health and digital literacy of both patients and their care providers to allow for optimal use of such tools. So-called explainable ML has been another hot topic of late,⁶⁸ implying that ML algorithms, not always straightforwardly interpretable to humans as regression coefficients or heatmaps, undermine the creditability of these ML algorithms.⁶⁹ Therefore, explanation techniques are needed to make these black box approaches explainable and trustworthy,⁷⁰ particularly in the healthcare setting. Unfortunately, the currently available methods (eg, regression with understandable coefficients, heatmaps for imaging applications) do not imply accurate performance and may give false assurances, and thus, are better understood as tools for developers.⁴⁷ Any tool to be used in clinical care must undergo “robust assessments of the efficacy,

affordability, and scalability of AI in the context of digital health for rare connective tissue diseases...to avoid the detrimental waste of scarce resources.”⁶⁶

AI/ML techniques can inform all stages of drug development and repurposing, including identification of potential targets, validation of those targets, identification of biomarkers, and optimization of clinical trial endpoints. These methods can harness a variety of datatypes, incorporating information from images, text, wearables, assays, and complex omics data, which can be used in concert to objectively inform some of the previously trial-and-error steps in this complex process.⁷¹

Additional AI/ML applications have been developed in other fields that will likely appear soon in RMD research. For example, epigenetic biomarkers of aging have been studied in cardiovascular disease, Alzheimer’s disease, and various cancers,⁷² but not yet in RMDs. Epigenetic clocks, reflecting one’s biological age, were developed to study age-related diseases and excess mortality. Clocks based on age-related inflammation have been created using ML but have not yet been studied in RMDs.⁷³ Other such clocks have been developed using a variety of omics data, although frequently in isolation.⁷⁴⁻⁷⁶ In contrast, the simultaneous incorporation of multimodal data (eg, genetic, omics, images, psychosocial, and/or clinical data), which hold substantial promise, is challenging because of the need to integrate multiple data types, potentially from different studies and cohorts. To date, most studies with such data are relatively small and primarily focused on the multiomics aspect rather than integration across all data types.⁴² It will be important in the future to collect these types of multiomics data on larger and more representative samples and fully integrate the different data types into ML models. A few studies have incorporated such multimodal data,³⁴ and others are collecting it,^{30,77} but additional rigorously designed longitudinal studies will be needed to establish this knowledge base and allow for discovery and validation using existing and newly developed methodologies capable of handling this type of multimodal information.

Summary

The promise of ML for advances in RMD research and clinical care is enormous, although not yet fully realized. As exemplified by papers discussed in this review (Table), development and implementation of ML algorithms requires collaborative efforts from a variety of experts including those with analytic, programming, and subject area expertise working together to achieve robust results. Examples discussed in this review include a range of RMDs, data types, data sources, approaches, and outcomes, reflecting the breadth of AI’s potential while also considering its limitations. We mention examples of future directions, although these are nearly limitless as technologies evolve. RMD research stands to benefit greatly from such technologies given the challenge of studying these rare diseases with traditional methodologies, but care must be taken to mitigate rather than amplify potential disparities and other potential biases.

ACKNOWLEDGMENT

We would like to thank Dr. Yvonne Golightly for her insightful comments on the draft manuscript.

Table. Summary of studies included in this review, data types and methods used, and highlights and limitations of each.

Aim/Research Question	Data Types	Methods Used	What Does This Study Add?	Drawbacks
To develop an algorithm that can aid SLE diagnosis ¹⁴	Clinical/demographic	RF; LASSO LR	Developed and validated an accurate, clinician-friendly algorithm based on classical disease features for early SLE diagnosis and treatment to improve patient outcomes	Small SS; retrospective design; further validation in prospective studies needed
To generate ML-based algorithms to accurately identify patients with SLE from EHRs ¹⁶	EHRs (structured and narrative data)	Natural language processing; penalized LR	Generated algorithms were internally and externally validated and compared to existing rule-based algorithms; the EHR phenotyping protocol is published and available for use in clinical and translational studies ¹⁷	The authors emphasize that future work is needed to fully adjust for portability
To define unique gene expression profiles in muscle biopsies from patients with different types of myositis ¹⁸	Transcriptomics data	Phenotyping Decision trees; RF; linear SVM model; other classification algorithms	Demonstrated the usefulness of objective transcriptomics to tailor therapies to a specific molecular diagnosis in the future	Small SS
To characterize the immune cell profile of patients with β SLE and investigate links to the disease trajectory over time ¹⁹	Laboratory data	Supervised ML approaches for classification (balanced RF, sPLS-DA) and feature selection, followed by unsupervised k-means clustering	Identified 4 potentially important subgroups among patients with SLE; may eventually inform clinical practice and clinical studies	Small SS; low number of Black patients; imperfect outcome measures
To assess the ability of an artificial intelligence system to predict disease activity of patients with RA at their next clinical visit ²⁰	Medications; disease activity; clinical/demographic; laboratory data	Prognosis Recurrent neural networks (longitudinal deep learning model)	The findings suggest that building accurate models to forecast complex disease outcomes using EHRs is possible	Missing data; usefulness of these models in clinical practice is not clear
To validate an ML model to estimate SLEDAI score categories from clinical notes and to apply the model to a large, real-world dataset to generate estimated score categories ²¹	Clinical notes	Natural language processing	Developed natural language processing algorithms, followed by validation to estimate SLEDAI categories from unstructured notes	Notes without sufficient text were excluded; the model will need to be modified for different data sources
To validate an ML model ²¹ to estimate CDAI scores in RA using clinical notes ²²	Clinical notes	Natural language processing	External validation of previously developed algorithm	The authors was limited to US practices only
To investigate whether alterations in the DNA methylation profiles of immune cells can discriminate between disease subtypes ²³	Clinical/demographic; DNA methylation	LR; RF; SVM	Demonstrated the potential of methylation markers as early disease determinants	Small SS
Identify distinct clusters of patients with PsA based on their baseline features and explore clinical and therapeutic implications ²⁵	Clinical/demographic	Model-based clustering methods	Data were pooled from 4 RCTs (increased SS); identified 7 distinct clusters; some showed greater improvements if using higher dose of SEC	RCT population
To identify patients with RA most likely to achieve remission with TCZ monotherapy by developing and validating a prediction model and associated remission score ²⁶	Clinical/demographic; disease activity; laboratory data	LR	Data were pooled from 4 RCTs (increased SS); derived a remission prediction score; robust design and rigorous statistical analysis	Inconsistency in variables between trials; validation using RWD is needed; low disease activity was not considered

Aim/Research Question	Data Types	Methods Used	What Does This Study Add?	Drawbacks
External validation of the prediction score using RWD ²⁷	Clinical/demographic; disease activity; laboratory data	LR; RF	Remission prediction scores derived in RCTs discriminated patients in RWD about as well as in RCTs; discrimination was further improved by retraining models on RWD	Missing data; patients were from North America only; other DMARDs need to be considered in the future work
To build a model to predict risk and time to TKR in osteoarthritis ²⁸	Clinical/demographic; patient reported outcomes; imaging; medications	Cox regression; DeepSurv; RF; linear/kernel SVM; linear/neural multitask LR models	Developed a model using the OAI cohort to predict with high accuracy if and when a given knee would require TKR	Not generalizable to the general population; only features known to be associated with OA were included in the dataset
To develop and use a novel selection method to identify individuals most likely to develop OA progression ²⁹⁻³¹	Clinical/demographic; PROs; imaging; medications	LR; multinomial LR; kNN classifier; support vector classifier; RF	The procedure was developed and subsequently used to preselect likely progressors (within the lifetime of the study) from existing OA cohorts, resulting in 297 participants to be followed for 2 years	Weak preprocessing strategy; difficult to implement in clinical practice
Precision Medicine				
To determine predictors of 16-week outcome based on starting dose of SEC in PsA ³³	Clinical/demographic	Bayesian elastic net	Determined predictors of 16-week outcome based on starting dose of SEC in PsA; provided insight relevant to precision medicine	RCT data; need for validation
To detect biomarkers and expression signatures of treatment response to TNF inhibition ⁴²	Multomics from PBMCs	Linear models with regularization; RF; SVM with radial basis function kernel	Suggested new predictive models of anti-TNF treatment in patients with RA	ML methods were not fully integrated into this analysis; small SS
To predict response to anti-TNF prior to treatment in patients with RA, and understand mechanisms of response ³²	Multomics	RF	Incorporation of true multiomics data and integrated data analysis in the prediction models	Future work will benefit from larger SS, robust outcomes, and independent external validation
To test the ability of ML approaches with clinical and genomic biomarkers to predict MTX treatment response in patients with early RA ³⁴	Clinical/demographic; genetics	RF	Integration of clinical and genomic data for individualized prediction of response; the incorporation of genetic data in the prediction algorithm substantially improved prediction accuracy	Sample included only White Europeans; limited generalizability
To apply a precision medicine approach to maximize expected outcomes in a knee OA RCT ⁴⁴	Clinical/demographic; PROs; laboratory data; medications	24 ML algorithms; optimal model was selected to develop individualized treatment rules	Highlights the use of RCT data from a nonpharmacologic trial, exploration of multiple features and outcomes, and multiple model evaluation, all of which could improve the design of future studies	Participants with missing outcomes and covariates with a large proportion of missing data were excluded; data were from a single RCT

CDAI: Clinical Disease Activity Index; DeepSurv: deep feed-forward neural network; DMARD: disease-modifying antirheumatic drug; EHR: electronic health record; jSLE: juvenile-onset systemic lupus erythematosus; kNN: k-nearest neighbors; LASSO: least absolute shrinkage and selection operator; LR: logistic regression; ML: machine learning; MTX: methotrexate; OA: osteoarthritis; OAI: Osteoarthritis Initiative; PBMC: peripheral blood mononuclear cells; PRO: patient-reported outcome; PsA: psoriatic arthritis; RA: rheumatoid arthritis; RCT: randomized clinical trial; RF: random forest; RNA: ribonucleic acid; RWD: real-world data; SEC: secukinumab; SLE: systemic lupus erythematosus; SLEDAI: Systemic Lupus Erythematosus Disease Activity Index; sPLS-DA: sparse partial least squares discriminant analysis; SS: sample size; SVM: support vector machines; TCZ: tocilizumab; TKR: total knee replacement; TNF: tumor necrosis factor.

REFERENCES

- van der Heijde D, Daikh DI, Betteridge N, et al. Common language description of the term rheumatic and musculoskeletal diseases (RMDs) for use in communication with the lay public, healthcare providers and other stakeholders endorsed by the European League Against Rheumatism (EULAR) and the American College of Rheumatology (ACR). *Ann Rheum Dis* 2018;77:829-32.
- Beam AL, Kohane IS. Big data and machine learning in health care. *JAMA* 2018;319:1317-8.
- Gossec L, Kedra J, Servy H, et al. EULAR points to consider for the use of big data in rheumatic and musculoskeletal diseases. *Ann Rheum Dis* 2020;79:69-76.
- Jiang M, Li Y, Jiang C, Zhao L, Zhang X, Lipsky PE. Machine learning in rheumatic diseases. *Clin Rev Allergy Immunol* 2021;60:96-110.
- Kingsmore KM, Puglisi CE, Grammer AC, Lipsky PE. An introduction to machine learning and analysis of its use in rheumatic diseases. *Nat Rev Rheumatol* 2021;17:710-30.
- Pedoia V, Majumdar S, Link TM. Segmentation of joint and musculoskeletal tissue in the study of arthritis. *MAGMA* 2016;29:207-21.
- Stoel B. Use of artificial intelligence in imaging in rheumatology – current status and future perspectives. *RMD Open* 2020;6:e001063.
- Gutiérrez-Martínez J, Pineda C, Sandoval H, Bernal-González A. Computer-aided diagnosis in rheumatic diseases using ultrasound: an overview. *Clin Rheumatol* 2020;39:993-1005.
- Joseph GB, McCulloch CE, Sohn JH, Pedoia V, Majumdar S, Link TM. AI MSK clinical applications: cartilage and osteoarthritis. *Skeletal Radiol* 2022;51:331-43.
- Nelson AE. How feasible is the stratification of osteoarthritis phenotypes by means of artificial intelligence? *Expert Rev Precis Med Drug Dev* 2021;6:83-5.
- Hügler M, Omoumi P, van Laar JM, Boedecker J, Hügler T. Applied machine learning and artificial intelligence in rheumatology. *Rheumatol Adv Pract* 2020;4:rkaa005.
- Decherchi S, Pedrini E, Mordenti M, Cavalli A, Sangiorgi L. Opportunities and challenges for machine learning in rare diseases. *Front Med* 2021;8:747612.
- Schaefer J, Lehne M, Schepers J, Prasser F, Thun S. The use of machine learning in rare diseases: a scoping review. *Orphanet J Rare Dis* 2020;15:145.
- Adamichou C, Genitsaridi I, Nikolopoulos D, et al. Lupus or not? SLE Risk Probability Index (SLERPI): a simple, clinician-friendly machine learning-based model to assist the diagnosis of systemic lupus erythematosus. *Ann Rheum Dis* 2021;80:758-66.
- Batu ED, Kaya Akca U, Basaran O, Bilginer Y, Ozen S. Correspondence on 'Lupus or not? SLE Risk Probability Index (SLERPI): a simple, clinician-friendly machine-learning-based model to assist the diagnosis of systemic lupus erythematosus'. *Ann Rheum Dis* 2021 Apr 2 (Epub ahead of print).
- Jorge A, Castro VM, Barnado A, et al. Identifying lupus patients in electronic health records: Development and validation of machine learning algorithms and application of rule-based algorithms. *Semin Arthritis Rheum* 2019;49:84-90.
- Zhang Y, Cai T, Yu S, et al. High-throughput phenotyping with electronic medical record data using a common semi-supervised approach (PheCAP). *Nat Protoc* 2019;14:3426-44.
- Pinal-Fernandez I, Casal-Dominguez M, Derfoul A, et al. Machine learning algorithms reveal unique gene expression profiles in muscle biopsies from patients with different types of myositis. *Ann Rheum Dis* 2020;79:1234-42.
- Robinson GA, Peng J, Dönnnes P, et al. Disease-associated and patient-specific immune cell signatures in juvenile-onset systemic lupus erythematosus: patient stratification using a machine-learning approach. *Lancet Rheumatol* 2020;2:e485-96.
- Norgeot B, Glicksberg BS, Trupin L, et al. Assessment of a deep learning model based on electronic health record data to forecast clinical outcomes in patients with rheumatoid arthritis. *JAMA Netw Open* 2019;2:e190606.
- Alves P, Bandaria J, Leavy MB, et al. Validation of a machine learning approach to estimate Systemic Lupus Erythematosus Disease Activity Index score categories and application in a real-world dataset. *RMD Open* 2021;7:e001586.
- Spencer AK, Bandaria J, Leavy MB, et al. Validation of a machine learning approach to estimate Clinical Disease Activity Index Scores for rheumatoid arthritis. *RMD Open* 2021;7:e001781.
- de la Calle-Fabregat C, Niemantsverdriet E, Cañete JD, et al. Prediction of the progression of undifferentiated arthritis to rheumatoid arthritis using DNA methylation profiling. *Arthritis Rheumatol* 2021;73:2229-39.
- Cahan A, Cimino JJ. Improving precision medicine using individual patient data from trials. *CMAJ* 2017;189:E204-7.
- Pournara E, Kormaksson M, Nash P, et al. Clinically relevant patient clusters identified by machine learning from the clinical development programme of secukinumab in psoriatic arthritis. *RMD Open* 2021;7:e001845.
- Collins JE, Johansson FD, Gale S, et al. Predicting remission among patients with rheumatoid arthritis starting tocilizumab monotherapy: model derivation and remission score development. *ACR Open Rheumatol* 2020;2:65-73.
- Johansson FD, Collins JE, Yau V, et al. Predicting response to tocilizumab monotherapy in rheumatoid arthritis: a real-world data analysis using machine learning. *J Rheumatol* 2021;48:1364-70.
- Jamshidi A, Pelletier JP, Labbe A, Abram F, Martel-Pelletier J, Droit A. Machine learning-based individualized survival prediction model for total knee replacement in osteoarthritis: data from the Osteoarthritis Initiative. *Arthritis Care Res* 2021;73:1518-27.
- Widera P, Welsing PMJ, Ladel C, et al. Multi-classifier prediction of knee osteoarthritis progression from incomplete imbalanced longitudinal data. *Sci Rep* 2020;10:8427.
- van Helvoort EM, van Spil WE, Jansen MP, et al. Cohort profile: The Applied Public-Private Research enabling OsteoArthritis Clinical Headway (IMI-APPROACH) study: a 2-year, European, cohort study to describe, validate and predict phenotypes of osteoarthritis using clinical, imaging and biochemical markers. *BMJ Open* 2020;10:e035101.
- van Helvoort EM, Ladel C, Mastbergen S, et al. Baseline clinical characteristics of predicted structural and pain progressors in the IMI-APPROACH knee OA cohort. *RMD Open* 2021;7:e001759.
- Tao W, Concepcion AN, Vianen M, et al. Multiomics and machine learning accurately predict clinical response to adalimumab and etanercept therapy in patients with rheumatoid arthritis. *Arthritis Rheumatol* 2021;73:212-22.
- Gottlieb AB, Mease PJ, Kirkham B, et al. Secukinumab efficacy in psoriatic arthritis: machine learning and meta-analysis of four phase 3 trials. *J Clin Rheumatol* 2021;27:239-47.
- Myasoedova E, Athreya AP, Crowson CS, et al. Toward individualized prediction of response to methotrexate in early rheumatoid arthritis: a pharmacogenomics-driven machine learning approach. *Arthritis Care Res* 2022;74:879-88.
- Mullin S, Zola J, Lee R, et al. Longitudinal K-means approaches to clustering and analyzing EHR opioid use trajectories for clinical subtypes. *J Biomed Inform* 2021;122:103889.
- Dong X, Deng J, Rashidian S, et al. Identifying risk of opioid use disorder for patients taking opioid medications with deep learning. *J Am Med Inform Assoc* 2021;28:1683-93.
- Badger J, LaRose E, Mayer J, Bashiri F, Page D, Peissig P. Machine learning for phenotyping opioid overdose events. *J Biomed Inform* 2019;94:103185.

38. Klemt C, Harvey MJ, Robinson MG, Esposito JG, Yeo I, Kwon YM. Machine learning algorithms predict extended postoperative opioid use in primary total knee arthroplasty. *Knee Surg Sports Traumatol Arthrosc* 2022;30:2573-81.
39. Grazal CF, Anderson AB, Booth GJ, Geiger PG, Forsberg JA, Balazs GC. A machine-learning algorithm to predict the likelihood of prolonged opioid use following arthroscopic hip surgery. *Arthroscopy* 2022;38:839-47.
40. Lee S, Wei S, White V, Bain PA, Baker C, Li J. Classification of opioid usage through semi-supervised learning for total joint replacement patients. *IEEE J Biomed Health Inform* 2021; 25:189-200.
41. Thompson HM, Sharma B, Bhalla S, et al. Bias and fairness assessment of a natural language processing opioid misuse classifier: detection and mitigation of electronic health record data disadvantages across racial subgroups. *J Am Med Inform Assoc* 2021;28:2393-403.
42. Yoosuf N, Maciejewski M, Ziemek D, et al. Early prediction of clinical response to anti-TNF treatment using multi-omics and machine learning in rheumatoid arthritis. *Rheumatology* 2022;61:1680-9.
43. Plant D, Barton A. Machine learning in precision medicine: lessons to learn. *Nat Rev Rheumatol* 2021;17:5-6.
44. Jiang X, Nelson AE, Cleveland RJ, et al. Precision medicine approach to develop and internally validate optimal exercise and weight-loss treatments for overweight and obese adults with knee osteoarthritis: data from a single-center randomized trial. *Arthritis Care Res* 2021;73:693-701.
45. Messier SP, Mihalko SL, Legault C, et al. Effects of intensive diet and exercise on knee joint loads, inflammation, and clinical outcomes among overweight and obese adults with knee osteoarthritis: the IDEA randomized clinical trial. *JAMA* 2013;310:1263-73.
46. Manrique de Lara A, Pelaez-Ballestas I. Big data and data processing in rheumatology: bioethical perspectives. *Clin Rheumatol* 2020;39:1007-14.
47. Ghassemi M, Oakden-Rayner L, Beam AL. The false hope of current approaches to explainable artificial intelligence in health care. *Lancet Digit Health* 2021;3:e745-50.
48. Gianfrancesco MA, Tamang S, Yazdany J, Schmajuk G. Potential biases in machine learning algorithms using electronic health record data. *JAMA Intern Med* 2018;178:1544-7.
49. Obermeyer Z, Powers B, Vogeli C, Mullainathan S. Dissecting racial bias in an algorithm used to manage the health of populations. *Science* 2019;366:447-53.
50. Kelly CJ, Karthikesalingam A, Suleyman M, Corrado G, King D. Key challenges for delivering clinical impact with artificial intelligence. *BMC Med* 2019;17:195.
51. Rajkomar A, Hardt M, Howell MD, Corrado G, Chin MH. Ensuring fairness in machine learning to advance health equity. *Ann Intern Med* 2018;169:866-72.
52. Shoop-Worrall SJW, Cresswell K, Bolger I, et al. Nothing about us without us: involving patient collaborators for machine learning applications in rheumatology. *Ann Rheum Dis* 2021;80:1505-10.
53. Martínez-García M, Hernández-Lemus E. Data integration challenges for machine learning in precision medicine. *Front Med* 2021;8:784455.
54. Andaur Navarro CL, Damen JAA, Takada T, et al. Risk of bias in studies on prediction models developed using supervised machine learning techniques: systematic review. *BMJ* 2021;375:n2281.
55. Prada-Ramallal G, Takkouche B, Figueiras A. Bias in pharmacoepidemiologic studies using secondary health care databases: a scoping review. *BMC Med Res Methodol* 2019;19:53.
56. Finlayson SG, Subbaswamy A, Singh K, et al. The clinician and dataset shift in artificial intelligence. *N Engl J Med* 2021;385:283-6.
57. Guo LL, Pfohl SR, Fries J, et al. Systematic review of approaches to preserve machine learning performance in the presence of temporal dataset shift in clinical medicine. *Appl Clin Inform* 2021;12:808-15.
58. Lu J, Liu A, Dong F, Gu F, Gama J, Zhang G. Learning under concept drift: a review. *IEEE Transactions on Knowledge and Data Engineering* 2019;31:2346-63.
59. Collins GS, Reitsma JB, Altman DG, Moons KG. Transparent reporting of a multivariable prediction model for individual prognosis or diagnosis (TRIPOD): the TRIPOD statement. *BMJ* 2015;350:g7594.
60. Collins GS, Dhiman P, Andaur Navarro CL, et al. Protocol for development of a reporting guideline (TRIPOD-AI) and risk of bias tool (PROBAST-AI) for diagnostic and prognostic prediction model studies based on artificial intelligence. *BMJ Open* 2021;11:e048008.
61. Wolff RF, Moons KGM, Riley RD, et al. PROBAST: a tool to assess the risk of bias and applicability of prediction model studies. *Ann Intern Med* 2019;170:51-8.
62. Stevens LM, Mortazavi BJ, Deo RC, Curtis L, Kao DP. Recommendations for reporting machine learning analyses in clinical research. *Circ Cardiovasc Qual Outcomes* 2020;13:e006556.
63. Olczak J, Pavlopoulos J, Prijs J, et al. Presenting artificial intelligence, deep learning, and machine learning studies to clinicians and healthcare stakeholders: an introductory reference with a guideline and a Clinical AI Research (CAIR) checklist proposal. *Acta Orthop* 2021;92:513-25.
64. Liu X, Cruz Rivera S, Moher D, Calvert MJ, Denniston AK; SPIRIT-AI and CONSORT-AI Working Group. Reporting guidelines for clinical trial reports for interventions involving artificial intelligence: the CONSORT-AI extension. *Nat Med* 2020;26:1364-74.
65. Cruz Rivera S, Liu X, Chan AW, et al. Guidelines for clinical trial protocols for interventions involving artificial intelligence: the SPIRIT-AI extension. *Nat Med* 2020;26:1351-63.
66. Bergier H, Duron L, Sordet C, et al. Digital health, big data and smart technologies for the care of patients with systemic autoimmune diseases: where do we stand? *Autoimmun Rev* 2021;20:102864.
67. Gossec L, Guyard F, Leroy D, et al. Detection of flares by decrease in physical activity, collected using wearable activity trackers in rheumatoid arthritis or axial spondyloarthritis: an application of machine learning analyses in rheumatology. *Arthritis Care Res* 2019;71:1336-43.
68. Belle V, Papantonis I. Principles and practice of explainable machine learning. *Front Big Data* 2021;4:688969.
69. Babic B, Gerke S, Evgeniou T, Cohen IG. Beware explanations from AI in health care. *Science* 2021;373:284-6.
70. Herzog C. On the risk of confusing interpretability with explicability. *AI and Ethics* 2022;2:219-25.
71. Vamathevan J, Clark D, Czodrowski P, et al. Applications of machine learning in drug discovery and development. *Nat Rev Drug Discov* 2019;18:463-77.
72. Levine ME, Lu AT, Quach A, et al. An epigenetic biomarker of aging for lifespan and healthspan. *Aging* 2018;10:573-91.
73. Sayed N, Huang Y, Nguyen K, et al. An inflammatory aging clock (iAge) based on deep learning tracks multimorbidity, immunosenescence, frailty and cardiovascular aging. *Nat Aging* 2021;1:598-615.
74. Holly AC, Melzer D, Pilling LC, et al. Towards a gene expression biomarker set for human biological age. *Aging Cell* 2013;12:324-6.
75. Yu Z, Zhai G, Singmann P, et al. Human serum metabolic profiles are age dependent. *Aging Cell* 2012;11:960-7.
76. Hertel J, Friedrich N, Wittfeld K, et al. Measuring biological age via metabonomics: the metabolic age score. *J Proteome Res* 2016;15:400-10.
77. All of Us Research Program I, Denny JC, Rutter JL, et al. The "all of us" research program. *N Engl J Med* 2019;381:668-76.