Highlights

Understanding the role and adoption of artificial intelligence techniques in rheumatology research: an in-depth review of the literature

Alfredo Madrid-García,Beatriz Merino-Barbancho,Alejandro Rodríguez-González,Benjamín Fernández-Gutiérrez,Luis Rodríguez-Rodríguez,Ernestina Menasalvas-Ruiz

- The rheumatology research community is increasingly adopting novel AI techniques
- There is an upward trend in the number of articles that combine AI and rheumatology
- Rheumatic and musculoskeletal rare diseases are taking advantage of AI techniques
- Independent validation of the models should be promoted

Understanding the role and adoption of artificial intelligence techniques in rheumatology research: an in-depth review of the literature

Alfredo Madrid-García[a,b,*,1], Beatriz Merino-Barbancho[b], Alejandro Rodríguez-González[c], Benjamín Fernández-Gutiérrez[a], Luis Rodríguez-Rodríguez[a,2] and Ernestina Menasalvas-Ruiz[c,2]

[a]*Grupo de Patología Musculoesquelética. Hospital Clínico San Carlos, Prof. Martin Lagos s/n, Madrid, 28040, Spain*

[b]*Escuela Técnica Superior de Ingenieros de Telecomunicación. Universidad Politécnica de Madrid, Avenida Complutense, 30, Madrid, 28040, Spain*

[c]*Centro de Tecnología Biomédica. Universidad Politécnica de Madrid, Pozuelo de Alarcón, Madrid, 28223, Spain*

ARTICLE INFO

Keywords:
artificial intelligence
machine learning
real-world data
rheumatology
rheumatic and musculoskeletal diseases
electronic health record

ABSTRACT

The outstanding and upward trend in the number of published research related to rheumatic and musculoskeletal diseases, in which artificial intelligence plays a key role, has exhibited the interest of rheumatology researchers in using these techniques to answer their research questions. In this review, we analyse the original research articles that combine both worlds in a five-year period (2017-2021). In contrast to other published papers on the same topic, we first studied the review and recommendation articles that were published during that period, including up to October 2022, as well as the publication trends. Secondly, we review the published research articles and classify them into one of the following categories: disease classification, disease prediction, predictors identification, patient stratification and disease subtype identification, disease progression and activity, and treatment response. Thirdly, we provide a table with illustrative studies in which artificial intelligence techniques have played a central role in more than twenty rheumatic and musculoskeletal diseases. Finally, the findings of the research articles, in terms of disease and/or data science techniques employed are highlighted in a discussion. Therefore, the present review arises with the aim of characterising how researchers are employing data science techniques in the rheumatology medical field.

1. Introduction

1.1. Clinical and technical background

Rheumatic and Musculoskeletal Diseases (RMDs) are defined by the major scientific societies of rheumatology, European Alliance of Associations for Rheumatology (EULAR) and American College of Rheumatology (ACR), as a heterogeneous group of more than 200 diseases and syndromes present in all age segments and in both genders, affecting not only joints, bones, cartilage, tendons, ligaments, nerves, blood vessels, and muscles but also internal organs [1]. The aetiology and pathophysiology of RMDs can be variable. From a genetic, environmental, postural hygiene, and physical injury perspective, to immunological system disorders, such as inflammation derived from autoimmune responses, infections, or mechanical deterioration of tendons, muscles, and bones. This group of diseases is commonly characterised by chronicity, pain, fatigue, disability, motion dysfunction, and larger female and elder affectation; producing a negative impact on the life expectancy and Quality of Life (QoL) of patients. The economic burden associated with RMDs is not negligible and has recently been under the spotlight, as these diseases are

This work was supported by the Instituto de Salud Carlos III, Ministry of Health, Madrid, Spain [RD21/002/0001]. The sponsor or funding organization had no role in the design or conduct of this research. The journal's Fee was funded by the institution employing the senior author of the manuscript (Fundación Biomédica del Hospital Clínico San Carlos)

The authors declare there are no competing interests

*Corresponding author

alfredo.madrid.garcia@alumnos.upm.es (A. Madrid-García)

ORCID(s): 0000-0002-1591-0467 (A. Madrid-García); 0000-0001-5070-4178 (B. Merino-Barbancho); 0000-0001-8801-4762 (A. Rodríguez-González); 0000-0002-6126-8786 (B. Fernández-Gutiérrez); 0000-0002-2869-7861 (L. Rodríguez-Rodríguez); 0000-0002-5615-6798 (E. Menasalvas-Ruiz)

[1]First author

[2]Share senior authorship

responsible for loss of productivity costs and costs derived from sick leave and work disability [2]. Concisely, RMDs have a high overall prevalence, a significant economic burden, a deleterious impact on the patients QoL, and some particularities that hinder the patients' management, making them unique and complex.

From a data science perspective, RMDs also have their own particularities and challenges. To begin with, RMDs data are usually longitudinal, as a result of the long patient follow-up, which can range from weeks to decades. Therefore, new approaches that seek to take advantage of these data, such as Group-Based Multi-Trajectory Modeling (GBMT) analyses are emerging [3]. Moreover, RMDs data tend to be heterogeneous and multidimensional. Not only clinical and demographic data but also image, genomic, and -to a lesser extent- sensor data have been used to characterise the patient's disease, the disease progression or the treatment response and its effect. For instance, the disease progression can be studied with radiological progression measures obtained from medical images. Other data sources and types, such as Patient-Reported Outcomes Measures (PROMs) (e.g., Health-Related Quality of Life (HRQoL)) are not uncommon in rheumatology [4]. In addition, data from different medical specialities, such as orthopedy, ophthalmology, pulmonology, immunology, pharmacy, cardiology, or radiology, often complement the original RMDs data. In this scenario, the dimensionality of the data can increase significantly, especially with genomic data and Genome-Wide Association Studies (GWAS). The outpatient setting of most rheumatic clinics also has an impact on how data is collected. These data often fall under the definition of Real-World Data (RWD). Although RWD has been shown to be a valuable source of information, some of its implications cannot be neglected, such as its less structured nature or the occurrence of biases (i.e., selection bias or informed consent bias) which may require additional processing [5]. In this regard, approaches based on Natural Language Processing (NLP) and topic modelling have been proposed to characterise the evolution of rare diseases in RMDs clinical narratives [6].

The complexity of this data has led to the search for tools capable of modelling and capturing complex statistical interactions and patterns in the data. Researchers have found in Artificial Intelligence (AI) tools a suitable collection of techniques to extract knowledge from data. These tools have been applied to basic, clinical, and translational rheumatology research studies and to both autoimmune and not-autoimmune RMDs. Some of the supervised learning algorithms employed in rheumatology research studies for regression, classification, and inference are linear, logistic, Poisson regression; regularised linear models (i.e., Least Absolute Shrinkage and Selection Operator (Lasso), Ridge and elastic net); Decision Trees (DT); Support Vector Machines (SVM); Bayesian Models (BM); Naive Bayes (NB); K-Nearest Neighbors (KNN); Random Forest (RF); Neural Networks (NN) and boost-based algorithms such as Gradient Boosted Models (GBM) or AdaBoost. These algorithms have been used for a wide range of applications, among them, to predict response to some biological treatments (e.g., anti-Tumor Necrosis Factor (TNF)) [7], disease flare risk based on physical activity [8], and suicide risk in patients with fibromyalgia [9]. On its behalf, unsupervised learning algorithms have played a key role in dealing with high-dimensional data, such as gene expression data and biomarker identification. In this regard, Principal Component Analysis (PCA) has been found to be extremely useful for dimensionality reduction when identifying biomarkers [10], avoiding overfitting and speeding up training time, and t-Distributed Stochastic Neighbor Embedding (t-SNE) for visualisation [11]. Clustering algorithms have followed multiple strategies: conectivity-based clustering (e.g., hierarchical clustering), centroid-based clustering (e.g., k-means, fuzzy c-means), density-based clustering (e.g., DBSCAN) and probabilistic models (e.g., Gaussian Mixture Models (GMM)). Moreover, the ability of Deep Neural Networks (DNN) to capture complex patterns has propitiated its use in computer vision and texture analysis tasks such as Region of Interest (ROI) identification and segmentation in radiology images. In this regard, DL has been used to quantify the cartilage stage severity in osteoarthritis [12], radiological progression in Rheumatoid Arthritis (RA) [13] or lumbar spinal stenosis grading [14]. Furthermore, DL has also been employed satisfactorily with structure data from Electronic Health Record (EHR) to forecast clinical outcomes using multiple network architectures, including Convolutional Neural Networks (CNN), Recurrent Neural Networks (RNN) and Long-short Term Memory (LSTM) [15]. Free text and unstructured data from clinical notes have been analysed following a NLP approach, taking advantage of algorithms like Latent Dirichlet Allocation (LDA) for topic modelling, which has shown to be useful for disease classification into different meaningful subgroups [16]. Other NLP based-procedures such as *word2vec* have been employed for rheumatic diseases phenotyping [17]. More recently, novel approaches are reaching the RMDs world, such as Few-Shot Learning (FSL) [18].

The present review arises with the aim of characterising how rheumatology researchers are employing data science techniques and to study to what extent these techniques have been adopted by rheumatologists. In this review, we intentionally omit the description and implication of the different learning techniques and the most widely used algorithms. As these topics have already been addressed in past reviews and usually account for a greater part of the manuscript, we decided to prioritise the description of the different studies.

1.2. Publication trends

The promising early results of AI techniques have been a decisive step toward its adoption among the different rheumatology research groups. This has been reflected in a growing number of publications, in recent years, in high-impact rheumatology specialised journals. In fact, it has endorsed the necessity of EULAR to elaborate good-practise recommendations when dealing with big data [19]. When running the Medline query presented in Section 2.1, including the five-year period of the state-of-the-art review (that is, 2017 to 2021), the upward trend can be easily appreciated. The number of published articles has grown by almost 300% from 2017 to 2021. See Figure 1.

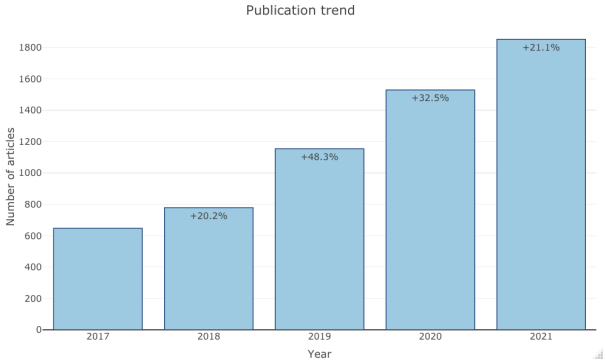

Figure 1: Publication trends in Medline when running the query presented in Section 2.1

1.3. Past reviews

Over the last few years different review and recommendation articles, in which AI and RMDs fields interact, have emerged.

In 2017, a review article addressed the ability of Machine Learning (ML) algorithms to discriminate between chronic pain patients (e.g., chronic pelvic pain, fibromyalgia, low back pain) and healthy controls [20].

In 2019, a systematic literature review informing EULAR recommendations for the use of big data and artificial intelligence in RMDs came to light [21], as well as, a compilation of studies that covered the feasibility and clinical utility of ML to stratify patients and predict treatment response [22]. A review covering the methods used in the analysis of rheumatic and musculoskeletal clinical data also saw the daylight in [23]. Throughout the year, other review articles that dealt with specific diseases were published: prediction models for Osteoarthritis (OA) [24], the role of ML and cardiovascular risk assessment in RA patients [25], detecting early RA [26], the application of ML in the Axial Spondyloarthritis (axSpA), and prediction of osteoporosis [27].

A year later, in 2020, the EULAR points to consider for the use of big data and artificial intelligence in RMDs were presented [19], along with two reviews of ML methods applied to rheumatic diseases [28, 29]. That same year, an introduction to the ethical issues of big data [30], as well as a review of the use of AI in imaging in rheumatology [31] were available.

In 2021, the review article titled *'An introduction to machine learning and analysis of its use in rheumatic diseases'* [32] provided a distinguished overview of the current situation of ML techniques applied to rheumatology, with a description of the most commonly employed algorithms, as well as, examples of ML applications in rheumatology. Another review, with a strong educational purpose, evaluated the relevance of data science to the field of rheumatology [33]. In addition, a scoping literature review of ML approaches to improve disease management of patients with RA was published at the end of the same year [34]. Particular interest was also given to the use of ML solutions for osteoporosis in [35]. The applicability of AI in the management of RMDs, with data collected from wearable activity trackers, was examined in [36].

In 2022, some authors provided an outline of ML applications in musculoskeletal histopathology [37], and reviewed the ML methods applied to OA research [38], with a special emphasis on Magnetic Resonance Imaging (MRI) [39]. Other authors addressed the use of three specific ML algorithms including logistic regression, in the diagnosis of rheumatic illnesses [40]. Another review, published in the middle of the year, discussed recent technologies and innovations that are expected to benefit clinical practice in the early 2030s, with regard to big data [41]. A narrative review of ML in RMDs for clinicians and researchers was published by Nelson AE et al. [42]. Furthermore, another

narrative review [43] covered the applications of ML in Systemic Sclerosis (SS). A review of AI and deep learning (DL) [44] attempted to highlight the relevance of these techniques in the near future of the field of rheumatology. Finally, a review explored the opportunities and challenges of using RWD focused mainly on electronic health record (EHR) data, to advance clinical research in rheumatology [5].

The increasingly frequent appearance of this kind of article (i.e., review and recommendation) gives an idea of the general adoption of AI and ML in RMDs. However, in the following section, particular interest is given to original research articles that support the convergence of RMDs and AI with concrete examples.

2. Materials and Methods

A literature search was conducted to identify publications related to RMDs in which data science techniques played a relevant role. Firstly, results from Medline, Scopus and Web of Science (WOS) were extracted. Lastly, specific searches were performed in the main rheumatology journals using their integrated search engines. The boolean operators AND and OR were used to streamline the procedure. The selected articles had to be indexed in PubMed (i.e., with a PubMed Identifier (PMID)) at the time of the search (that is, the index dates were used instead of the publication dates).

2.1. Search in Medline

The search consisted of two stages. First, articles published from January, 1st 2017 to June, 17th 2020 were retrieved. Then, an update was performed on February, 22th 2021, with posts published between the two previous dates.

The Medline search strategy included a combination of keywords and Medical Subject Headings (MeSH) terms. Due to the large number of keywords related to RMDs and AI and potential omissions, the search strategy did not specify, for example, a concrete type of disease or a concrete AI technique or algorithm. The keywords and MeSH terms from Table 1 were used to build the Medline query:

((Artificial Intelligence)[All Fields] OR (Artificial Intelligence[MeSH Terms]) OR (Big Data)[All Fields] OR (Big Data[MeSH Terms]) OR (Data Mining)[All Fields] OR (Data Mining[MeSH Terms]) OR (Machine Learning)[All Fields] OR (Supervised Learning)[All Fields] OR (Supervised Machine Learning[MeSH Terms]) OR (Unsupervised Learning)[All Fields] OR (Unsupervised Machine Learning[MeSH Terms]) OR (Deep Learning)[All Fields] OR (Deep Learning[MeSH Terms]))

AND

((Rheumatology) OR (Rheumatology[MeSH Terms]) OR (Rheumatic) OR (Musculoskeletal) OR (Musculoskeletal diseases[MeSH Terms]))

2.2. Search in Scopus

The Scopus query was restricted to the article title, keywords, and abstract. Only results from January, 1st 2017 to February, 22th 2021 indexed in Medline (i.e., with a PubMed identifier (PMID)) were included. The query performed in Scopus was:

TITLE-ABS-KEY(((artificial AND intelligence) OR (big AND data) OR (data AND mining) OR (machine AND learning) OR (supervised AND learning) OR (unsupervised AND learning) OR (deep AND learning))

AND

((rheumatology) OR (rheumatic) OR (musculoskeletal)))

2.3. Search in Web of Science

The Web of Science (WOS) search was similar to the Scopus search. The query performed was as follows:

TS=((("Artificial Intelligence") OR ("Big Data") OR ("Data Mining") OR ("Machine Learning") OR ("Supervised Learning") OR ("Unsupervised Learning") OR ("Deep Learning"))

AND

((Rheumatology) OR (Rheumatic) OR (Musculoskeletal)))

Table 1
Keywords and MeSH terms used in the Medline search

Domain	Keywords	MeSH terms
Artificial Intelligence	Artificial Intelligence	Artificial Intelligence
	Big Data	Big Data
	Data Mining	Data Mining
	Machine Learning	
	Supervised Learning	Supervised Machine Learning
	Unsupervised Learning	Unsupervised Machine Learning
	Deep Learning	Deep Learning
Rheumatic and Musculoskeletal Diseases	Rheumatology	Rheumatology
	Rheumatic Musculoskeletal	Musculoskeletal Diseases

2.4. Search in rheumatology journals

Articles published in Q1 and Q2 rheumatology journals (according to 2019 Journal Citation Reports) were retrieved, excluding those classified as 'Congress', 'Abstract' or 'Miscellaneous'. The decision to limit this search to Q1 and Q2 journals was made to ensure that the articles retrieved had a high impact. The journals included in this search were: *Nature Reviews Rheumatology, Annals of the Rheumatic Diseases, Arthritis & Rheumatology, Rheumatology, Therapeutic Advances in Musculoskeletal Disease, Osteoarthritis and Cartilage, Seminars in Arthritis and Rheumatism, Arthritis Research & Therapy, Arthritis Care & Research, Current Opinion in Rheumatology, Current Rheumatology Reports, Joint Bone Spine, Rheumatology and Therapy, Journal of Rheumatology, and Clinical and Experimental Rheumatology*.

The query used in the search engine of the different journals was:

"Machine Learning"

3. Literature review

The number of records identified through the database search steps described in the previous sections was 4,325. In Table 2 the initial number of articles retrieved with the different sources is shown. From this point on, different exclusion criteria were applied. Figure 2, shows the exclusion and inclusion criteria.

To sum up, we included unique articles, with a PMID, written in English, with no animal involvement. In addition, we excluded articles identified as *analysis, annotation, biography, case report, clinical trial, cohort profile, comment, correspondence, correspondence response, editorial, editorial comment, epilogue, erratum, guideline, letter, meta-analysis, methodology, news, news & analysis, opinion, overview, perspective, protocol, research highlight, research report, response, video article*.

We also removed articles not related to RMDs; articles covering the following topics: *medical surgical procedures such as arthroplasty or arthroscopy, biomechanics, cancer, muscle and bone malignancies, education, force simulation, gait, image generation, reconstruction, joint replacement, orthopaedics, radiology, rehabilitation and other non-farmacological interventions, robotics, simulation and surgery*; articles out of scope, not applying AI techniques or similar to previously identified ones.

The articles remaining after the elimination of duplicates (i.e., exclusion criteria 2) are shown in Supplementary Excel File 'Unique Articles'. Finally, the article title, the identifiers (Digital Object Identifier (DOI) and PMID), the journal, the publication year, the disease, the algorithms/techniques, the programming language, the number of patients, the validation type, the objective, and the authorship information are available in the Supplementary Excel File 'Included Articles'.

Table 2
Number of articles retrieved and with PMID in the different search engines

Source	Number of records identified	PMID
MedLine	2,071	2,071
Scopus	1,282	492
Web of Science	731	433
Rheumatology journals	241	241

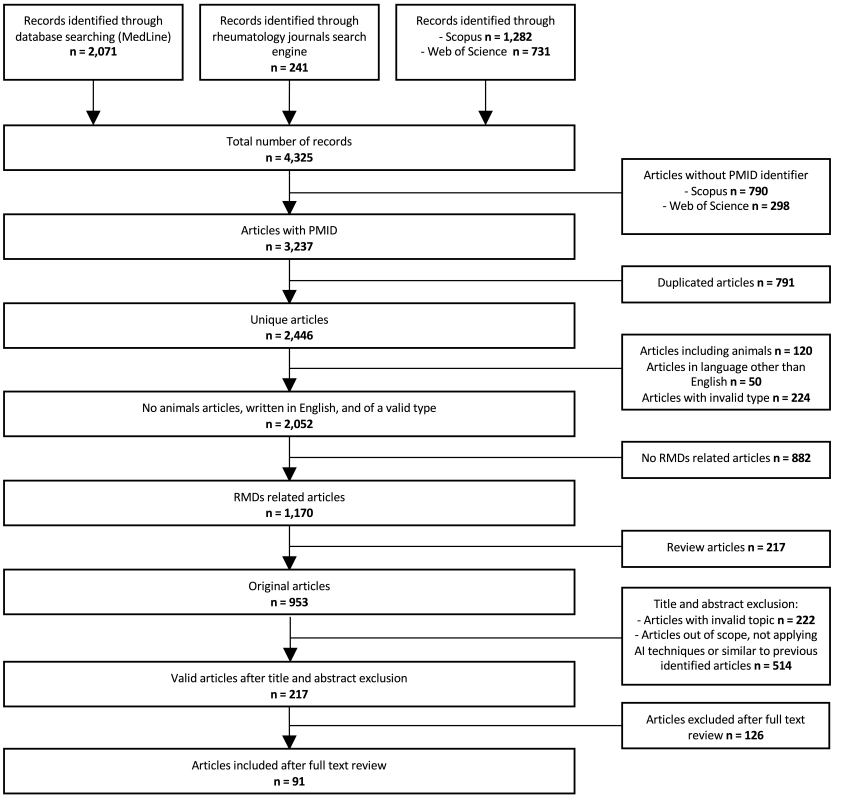

Figure 2: State of the art inclusion and exclusion criteria

3.1. Classification of topics and predictors

Hereafter, a review of the data science techniques that have been employed in rheumatology research over the last five years (2017-2021) is done. Different categories have been proposed to classify research articles. For instance, the scoping review presented in [34] suggested the following categories and subcategories to describe the ML studies applied in RA research:

- Diagnosis
  - Diagnosis based on EHR
  - Diagnosis based on biological samples
  - Diagnosis based on imaging and image recognition
  - Diagnosis based on sensors
- Monitoring
  - Monitoring the disease

  - Monitoring comorbidities

- Prediction
  - Prediction of response to treatment
  - Prediction of outcomes

The authors from the review [32], proposed the following classification:

- Patient classification
  - EHR and clinical data
  - Imaging and biometric data
  - Urinalysis, flowcytometry, and genomics
- Risk classification and outcome prediction
- Predicting treatment response or candidates for treatment
- Patient clustering to determine disease subtypes

As shown, there is no standard form to classify the different clinical research articles in which data science is involved. The proposal followed in this review article tries to achieve a balance between resolution and significance. For this work, the articles are categorised into six main topics:

1. **Disease prediction**: the disease prediction task can be seen as an individualisation of disease classification, in which the classification between healthy (i.e., controls) and unhealthy patients is the main objective. This distinction is usually based on a threshold cutoff point. In this section, the studies which attempt to predict the disease status rather than to classify different RMDs are detailed. Predictive models that address early diagnosis are an important branch of this topic, since a late diagnosis can be paired with increased flares and organ dysfunction, according to the therapeutic window of opportunity concept previously introduced. The lack of specific symptoms, diagnostic criteria, and validated biomarkers; as well as, the existence of diseases with a similar course, can hinder and difficult this early diagnosis, for some diseases, compromising the long-term well-being of patients. An important task on this topic is the development of algorithms capable of identifying patients in a EHR, facilitating the construction of specific datasets for epidemiological and observational studies. A fraction of the RMDs are considered to be *rare diseases*, and some of them have a worse prognosis and an increased risk of death. For this reason, some authors have put their efforts into developing pipelines that take advantage of statistical learning techniques capable of detecting patients, while avoiding manual and time-consuming chart reviews.
2. **Disease classification**: predictive models for disease classification have gained interest due to their ability to discriminate the pathology of a patient among diseases that share a similar course, symptoms, or early manifestations. For example, in this last scenario, the diagnosis of RA or OA may not be clear when starting from a *early inflammatory arthritis* status. This kind of task often relies on immune or genetic data.
   The relevance of these models lies in their ability to assist physicians in determining the disease of a patient. With the model results, the physicians will follow the most suitable therapeutic strategy for that concrete disease. The choice of inappropriate treatment, due to misdiagnosis, can jeopardise the remission status and maintain or worsen the pathological manifestations produced by the true underlying disease. Furthermore, a late diagnosis may displace the therapeutic window of opportunity [45], compromising future outcomes and patient recovery. Therefore, making the correct diagnosis is imperative.
3. **Predictors identification**: the identification of predictors associated with a dependent variable is the core of many investigations. Usually, this problem is tackled with linear and logistic regression models. DT, RF, and other highly interpretable algorithms have been used lately, along with regression models, to obtain a variable importance measure among the different candidate predictors. The identification of predictors has usually been the first step taken to address other research questions, as well as, to build predictive models. In the medical field, the identification of clinical variables that can be related to the development of a disease, a worse prognosis, or a better HRQoL is of particular importance.

4. **Patient stratification and disease subtype identification**: patient stratification and disease subtype identification have also been interesting areas for rheumatology researchers. The key idea of these studies is to stratify patients into meaningful subgroups that share similar characteristics (e.g., histological, molecular features), and are different from other subgroups of patients. Therefore, patients who belong to different subgroups can benefit from specific treatments and care, according to the different mechanisms involved in the pathogenesis of the disease. These studies are particularly relevant when the disease to be studied has significant individual variability in terms of clinical, laboratory, and immunological abnormalities. Unsupervised algorithms and clustering techniques are essential for this task.
   Different manifestations of the disease are also considered on this topic. For instance, patients who present a concrete pathology can be prone to the development of different associated pathologies, such as malignancies or cardiovascular events. In this context, the control group is usually formed by patients with the main pathology, but without the associated manifestation. Finally, graph-based techniques have been used to assess the strength of the association between the different predictors to generate prototypical variable profiles capable of discriminating between different phenotypes of the disease.
5. **Disease progression and activity**: measuring and quantifying the disease progression and activity often depends on the concrete underlying disease. In many cases, the anatomical damage is a good indicator of this progression, since it could serve as a footprint for persistent inflammatory activity. In RA or axSpA, radiological progression (i.e., erosion) is commonly employed to quantify this progression, while, in OA, the KL scale, which considers the marginal osteophytes, joint space narrowing, subchondral bone sclerosis, and altered shape of the bone, is well adopted. In this situation, radiological data are almost indispensable to quantify such a measure. In some other situations, the presence of flares can be an indicator of disease activity. Finally, in other scenarios, complex indices [46] are required to estimate the activity of the disease, such as SLEDAI.
6. **Treatment response**: the response variability to treatment among different patients, especially in complex treatments such as biological DMARDs (bDMARDs), is currently a promising research challenge. The polygenic response to some of these treatments requires large datasets from which meaningful statistical associations can be identified. Together with the polygenic response, other factors, such as the dosage, are involved in the fluctuating treatment response. Similarly, response to treatment may be quantified differently depending on the disease explored. For example, in RA the Treat to Target (T2T) approach is commonly used. Moreover, the time horizon of the prediction is variable. While some studies try to predict an outcome from a few months to years after starting the treatment, others try to predict the response before the initiation of the treatment. Data science techniques can be extremely helpful for modelling and describing such statistical associations. For instance, defining a rescue therapy or prescribing an alternative treatment before it is too late can improve the patient's well-being and control the progression of the disease.
   Different research avenues for treatment response can exist. Hence, researchers have addressed different scenarios such as non-responder patients (i.e., drug resistance) or Adverse Event (AE) reactions derived from a medication. The starting dose of treatment is also a research question addressed in treatment response predictive model studies. Finally, studies focused on the patient's perception of treatment are also included in this topic.

Sometimes, the boundary between these categories is fuzzy. Therefore, when the studies tackled more than one topic, they were assigned according to the primary research objective.

Although a common approach is to enrich the dataset with predictors of different natures (e.g., demographic, clinical, molecular biomarkers or radiological); a distinction between the types of the predictors and their role in the study was intended. As was the case with the classification of the research topic, it is sometimes unclear how to classify an article depending on the type of the variables involved, as there can be a mix of variable types. The following classification was proposed:

1. **Clinical and demographic**: clinical and demographic predictors, such as sex, age, diagnoses, treatments, comorbidities, and age of symptom onset are commonly used to build disease classification predictive models due to their relative easy accessibility from an EHR. Historically, traditional statistics models have benefited from these kinds of predictors, as they were relatively easy to obtain and manipulate. However, they are still an essential source of information for conducting epidemiological research with both traditional and modern statistical learning approaches.
2. **Molecular biomarkers**: changes in cellular and tissue metabolism and composition may appear as a consequence of chronic inflammatory disease. As a result, serum levels of certain metabolites or antibodies may

be increased. Differences in these metabolites and in their concentrations can constitute a useful fingerprint to differentiate diseases, acting as a biomarker signature.

Serum/urinary proteins, microRNAs, T-cell receptors, aminoacids, DNA methylation patterns, type I Interferon (IFN) signature, chemokines, cytokines and proinflammatory molecules may be useful fingerprints for identifying specific diseases and/or their prognosis. As a result, multiple immunological, histological, and genomic molecular biomarkers have been studied or are being investigated. However, some antibodies can be shared between different diseases such as rheumatoid factor (i.e., RA and primary Sjogren's syndrome (Primary Sjögren's Syndrome (pSS))). For this reason, additional and unambiguous biomarkers are desirable for identifying concrete pathologies. Blood serum, urine, and saliva samples have been collected in multiple studies for the identification of biomarkers.

A common challenge in studies that use these molecular signatures is to face the *high dimensionality problem*, this is, a scenario where the number of potential predictors is much higher than the number of observations (commonly known as $p \gg n$). Therefore, preprocessing steps like feature selection play a key role in studies in which this casuistry occurs.

3. **Medical images**: an image is understood as a matrix of pixels, each of them with an intensity value. When working with radiological images, DL is one of the most promising data science techniques due to its ability to capture complex interactions and patterns from the data. Computer vision for image segmentation, image registration, anatomical measurement, and pathology identification and detection are some of the capabilities DL can offer to rheumatology researchers who work with medical images. Computer vision has also been applied as an intermediate step in other research pipelines. For example, DL techniques have been used to segment parts of the body that could be used in a subsequent step of a classification model. The structure and topology of the network differ between studies. Parameters and hyperparameter optimisation, such as the number of hidden layers, the number of neural units, the penalisation and regularisation (e.g., data augmentation, dropout), the learning rate, the activation and loss functions, or the number of epochs, define the complexity of the networks. Therefore, the performance of the final model can be highly determined by the preprocessing steps, the network infrastructure and its parameters. Moreover, the use of previously pre-trained networks and transfer learning is also a well-extended option which facilitates the use and implementation of this promising technique, reducing the training time and the number of training examples.

   Depending on the methodology followed by the researcher, radiological information can be used alone or can complement other demographic and clinical characteristics in ensemble methods. DL is usually preferred when the image is used solely to train the model, in fact, already published results have demonstrated that algorithms trained exclusively on image data are robust enough to provide remarkable results. However, when relevant radiological information is extracted from the image after a feature selection process (i.e., a radiologist measure and quantification, the distance between structures, histogram features, etc.), and algorithms are trained with these data combined with other clinical and demographic parameters, ML classifiers can be used. Research studies combining DL and ML techniques for training and benchmarking algorithms have been published [47]. In these studies, the performance of the algorithms is usually compared to the radiologist's criterion.

   Due to the key role radiological images play in the diagnosis and staging of certain diseases such as knee OA or osteoporosis, many of the RMDs research studies employing DL focus on these diseases. However, research on other diseases can also be carried out with the aid of such techniques.

   Finally, the data source variability of medical imaging-based studies is high. Apart from well-extended image techniques such as Computed Tomography (CT), MRI, X-Rays (RX), Ultrasound (US); other less common image techniques such as optoacoustic [48], photomicrographs [49], smartphone photos [47], and thermograms [50] have also been employed in research studies.

The classification proposed could have also considered data from sensors, wearables and activity trackers (e.g., accelerometer and gyroscope), this is, sensor and signal data. However, for an easy understanding, the classification made is based on the purpose of the study rather than the data type.

To conclude this introduction, as an example of dataset enrichment with different data sources, authors in [51] combined 82, clinical, demographic, laboratory and image measures (i.e., ultrasound (US)) to predict intravenous immunoglobulin resistance in Kawasaki disease using different ML models.

3.2. Disease prediction

3.2.1. Clinical and demographic

Early diagnosis predictive models have been developed for Ankylosing Spondylitis (AS), after applying mutual information for feature selection [52]. In this study, the authors trained different algorithms such as SVM, RF or GBM, and compared their performance against more traditional methods like linear regression and clinically-based models. The purpose of this study was the early identification of AS patients based on medical and pharmacy claims history. Of 228,471 patients, without a history of AS, the ML model predicted that 1,923 patients would develop AS. However, only 120 of the predicted patients developed AS. A total of 1,242 out of 228,471 finally developed the disease. While the linear regression model accounted for a higher AUC, 0.71, compared to the ML model, 0.63, the False Positive Rate (FPR) and Positive Predictive Value (PPV) values were 4.67% and 2.55%, for the linear regression and 0.79% and 6.24% for the ML model.

In recent years, different research groups have been working to identify axSpA patients in large datasets with both structured and unstructured data. As the authors stated, identifying axSpA patients for observational studies is a challenge, partly due to the lack of codification for the different disease phenotypes. Historically, axSpA International Statistical Classification of Diseases and Related Health Problems (ICD) codes and manual chart review have been used for identification. Recently, three models to identify patients with axSpA were proposed by a US research group [53]. These models considered a different number of unique predictors from 16 to 49, including concepts extracted with NLP and demographics, laboratory, healthcare utilisation, ICD codes and comorbidity index variables. To assess the performance of the identification methods, the authors used the RF algorithm to quantify the axSpA risk score. The authors concluded that the most complete proposed algorithm, which included 49 variables, had the highest overall performance.

Previously, a research group from the UK [54] had applied unsupervised and supervised algorithms, also using concepts extracted after applying NLP, to accurately identify axSpA patients in an EHR from an enriched axSpA cohort. In this regard, the authors generated a list of candidate axSpA concepts with Surrogate Assisted Feature Extraction (SAFE) and selected the most informative concepts with Lasso after processing clinical reports with NLP. Furthermore, they tested an unsupervised implementation called Multimodal Automated Phenotyping (MAP) which combined information from three key domains (i.e., ICD codes, NLP concepts and healthcare utilisation). This last approach had the highest sensitivity and PPV combination, 0.78 and 0.80, respectively, with an AUC value of 0.927. Lastly, authors concluded that axSpA concepts could be accurately identified in EHR by incorporating narrative data.

In a similar manner, efforts have been made to identify patients with SS, using meaningful variables such as ICD codes, laboratory data and keywords [55]. Observational SS studies are usually limited to the sample size. With this in mind, the authors aimed to implement different approaches to broadly capture SS patients from an EHR. This time, the performance of rule-based, Classification and Regression Tree (CART) and RF algorithms was compared. Although the authors showed that ML based algorithms were not the highest performing ones, in terms of PPV and in comparison with rule-based algorithms, they were not as time-consuming to develop and validate as the first ones. In addition, the authors highlighted some potential advantages of ML algorithms: they do not require specific domain knowledge and can be automatically tuned to easily identify the optimal model parameters.

In lupus research, studies to identify patients in EHR and to predict risk probabilities have been carried out. For instance, authors from [17] presented a study in which they used different text classifiers techniques for identifying SLE patients while reviewing common and new NLP approaches (e.g., Bag-of-Words (BOWs), Concept Unique Identifiers (CUIs), *Word2vec*). On this basis, the authors showed a pipeline with three clearly separated pathways and two main split nodes, based on the way of transforming textual data into features and processing them. If the final word representation was based on word/CUIs frequency in each clinical note (i.e., data resulting from applying BOWs and CUIs), a RF classifier to extract the variable importance and to select features was applied; followed by NN, RF, NB and SVM, classifiers. On the other hand, if the word representation was based on vectors for each word presented in a document (e.g., *Word2vec*), bayesian inversion was conducted. When benchmarking results from the test set, the AUC scores ranged from 0.80 to 0.99. Although the novelty proposed method, bayesian inversion, was not the best-performing, authors concluded that it is promising for identifying patients, since it has fewer dependencies, lower testing, and is more scalable.

A model to produce SLE risk probabilities was recently presented [56]. In this study, authors combined clinical and serological features from three SLE classification criteria with non-criteria features to calculate SLE risk probabilities. After performing a correlation analysis, researchers created 20 panels of features, each panel was used to train two algorithms Lasso-LR and RF. The best-performing model, in terms of AUC, achieved a 0.98 AUC in a validation

cohort, contained 14 features and was based on Lasso logistic regression. To facilitate the adoption of the model into clinical practice, the authors used k-means clustering to detect unbiased risk probabilities partitions.

Studies that assess the temporal validity of predictive models, developed years ago, over time have been carried out [57]. With a seven-year difference, the authors compared the performance of a RA logistic regression algorithm, trained in 2010, against a new logistic regression model that included new ICD-10 codes and RA treatments. ICD codes, prescription, laboratory results, and NLP concepts from narrative data, stored in a new EMR, were used. The authors evaluated the performance of the ancient algorithm in the updated data, as well as, the performance of the updated algorithm that uses the same model coefficients as the first one, while incorporating new variables. Researchers finally concluded that the performance of the old algorithm was similar when validating with updated data.

Leiden and Erlangen researchers developed a workflow for building a ML algorithm capable of accurately identifying patients with RA from clinical narratives using NLP [58]. The aim of this research was to implement a broadly applicable workflow to equip centres with their own high-performing algorithm. Therefore, this research study was more oriented to the technical side.

A study for the early OA detection, using exclusively clinical and demographic data, has been conducted in [59]. In this study, the authors highlighted the relevance of building OA classification models when missing image data. For this study, data from 5,749 subjects were used and a DNN trained. The authors achieved an AUC value of 0.77, when using variables such as age, gender, household income, or physical activity.

The importance of topic modelling and NLP in RMDs was shown in the context of pseudogout disease [60]. As the authors reported, identifying pseudogout in large datasets can be challenging due to several factors: its incidence is not well-characterized and there exists a lack of specific billing codes. Almost ten million narrative notes from more than 50,000 patients were processed and Unified Medical Language System (UMLS) codes were applied to structure the data. Then, a filter based on billing codes and NLP concepts followed by random patient selection (n = 900) was performed. Subsequently, with the aid of a novelty topic modelling algorithm followed by penalised regression, sureLDA, a pseudogout propensity score was estimated and regression models were computed to predict the probability of pseudogout.

3.2.2. Molecular biomarkers

RF, Lasso, and Ridge regression were used for classifying healthy from SLE patients (n = 80) in an epigenetic context in which DNA methylation signatures and their relevance in patients' ethnicity were considered [61]. In this study, the authors explored epigenetic defects in B cell development patterns and took advantage of statistical learning methods to identify the most informative genes involved in the different epigenetic states. To this end, the algorithms were tested across ethnicity groups in an independent validation cohort.

RF, SVM and Artificial Neural Network (ANN) algorithms were trained to discriminate patients with Juvenile Idiopathic Arthritis (JIA) from healthy controls using deep immunophenotyping characteristics [62]. Up to 42 immunological parameters that could act as a disease signature were generated for 128 patients. After training the different models, the authors found that the RF discrimination ability was superior to other models, achieving a 0.89 AUC value. With this development, researchers showed that ML could be used together with the immunophenotyping technique to identify immune signatures that correlate with JIA disease.

Another JIA research article was recently published [63]. This time, authors pursued to build Genomic Risk Scores (GRSs) for diagnosis using Single-Nucleotide Polymorphism (SNP)s and Lasso penalised regression models. External validation was performed with two cohorts. A key idea of this study was the design of JIA subtype-specific GRSs for seven mutually exclusive categories of JIA.

3.2.3. Image

Certain anatomical structures can act as a disease's sign or symptom and, therefore used to diagnose a pathology. For instance, halo sign detection in Giant Cell Arteritis (GCA), is an example of these structures. In a recently published study [64], authors used 1,311 colour doppler US images from 137 patients to train a U-Net able to detect halo sign based on a pixel score metric, achieving a 0.83 AUC in the test set.

On the other hand, sacroiliac joint erosion is an early symptom of AS. A recently published study compared the performance of ML and DL classifiers in detecting erosion in a set of 681 sacroiliacs joint CT images from 53 patients. For the former, the authors extracted the features and built a set of predictors with Gray-level co-occurrence matrices (GLCM) and Local Binary Patterns (LBP) texture analysis techniques. Then, different classifiers (e.g., KNN, RF) were trained and compared. For the latter, DL, they used and modified a pre-trained model, InceptionV3, via transfer

learning. The authors concluded that DL outperformed, 0.97 AUC, ML algorithms and a radiologist with 9 years of experience in terms of sensitivity and specificity [65].

Following an ensemble method approach, investigators from Japan [66] achieved a 0.93 AUC in the diagnosis of osteoporosis of the proximal femoral, after training five well-known pre-trained networks (e.g., ResNet18, ResNet34, GoogleNet, EfficientNet b3, EfficientNet b4) with a set of 1,131 hip Dual-energy X-ray Absorptiometry (DXA) images plus four clinical covariates: age, sex, Body Mass Index (BMI) and history of hip fracture. A relevant contribution to highlight from this study is that structured data from patient records, in this case, clinical covariates, were useful to improve the DL network performance in images, thanks to ensemble models.

Finally, other studies with a strong computer vision component have been proposed. For example, central sarcopenia detection was the main goal of [67]. In this study, CT scans of 102 patients were segmented and analysed using a U-Net NN model. The performance of the system was evaluated using Dice coefficients.

3.3. Disease classification

3.3.1. Clinical and demographic

Yu SC et al. [68] built different classification models (i.e., SVM with different kernels, RF) to distinguish lupus lymphadenitis (n = 19) from Kikuchi disease (n = 81) using clinicopathological characteristics, in a case series of one hundred patients. Models were externally validated, achieving high sensitivity, 1, and specificity, 0.96. However, the validation cohort was small and extremely unbalanced, with only two cases of lupus lymphadenitis. This study also highlighted the relevance of AI techniques in the study of rare diseases and manifestations.

3.3.2. Molecular biomarkers

PCA and Lasso techniques were used for disease classification based on the combination of distinct serum protein biomarkers in RA-Interstitial Lung Disease (ILD) and RA-no ILD patients [69]. The serum levels of 45 proteins consisting of cytokines, chemokines, growth factors, and remodelling proteins were measured, and the authors identified seven biomarker signatures that effectively differentiated both diseases, achieving an Area Under the Curve (AUC) value of 0.93 (95% CI: 0.85-1).

In another study [70], authors developed a model capable to classify RA patients from controls, AUC 0.71 (95% Confidence Interval (CI): 0.58-0.84), Systemic Lupus Erythematosus (SLE) patients from controls AUC 0.80 (95% CI: 0.65-0.96) and RA from SLE patients AUC 0.63 (95% CI: 0.44-0.82) using clinical data and MicroRNA (miRNA)s biomarkers. For this purpose, RF and Lasso were employed. Whereas the former was used to capture a list of candidate miRNAs, the latter was used to select a final miRNA panel that maximised discrimination between diseases. Although the panel differentiated RA and SLE patients from controls, it was unable to differentiate properly between patients with RA and SLE.

Liu et al. [71] carried out a research for classifying RA, SLE and controls using T-cell receptors (TCRs) as biomarkers, training a RF model and achieving an AUC of 0.99.

Authors from [72], used RF, NB, multivariate logistic regression and hierarchical clustering to identify patients suffering from seronegative RA and Psoriatic Arthritis (PsA). They achieved an AUC value of 0.71 using demographic characteristics and serological concentrations of amino acids as predictors.

DNA methylation was used to classify patients with SLE (n = 347), pSS (n = 100) and controls (n = 400), employing RF. Four different models of disease status were developed: SLE/control, pSS/control, pSS/SLE and pSS with an specific type of antibodies and without them. Authors achieved an AUC value between 0.83 (pSS/SLE) and 0.96 (pSS/control) [73]. This study shows that obtaining a good performance for disease classification (lowest AUC, 0.83) is usually harder than for disease prediction.

Finally, the authors in [74], studied bacterial nucleic acids in synovial fluids from 58 OA and 125 RA patients, and built a classification model using SVM. They obtained a mean AUC value of 0.79.

3.3.3. Image

In [47] researchers used DNN and transfer learning, InceptionV3, for classifying photographs of hands into OA, RA, and PsA. Firstly, they developed a DL model to estimate the probability that an image had one of those conditions. Then, they combined this information with validated questionnaires and a single examination technique to determine the most likely diagnosis in a patient presenting hand arthritis. The number of participants was 280, and the algorithms employed were SVM, RF and Logistic Regression (LR). The accuracy oscillated between 0.78 and 0.97 when classifying OA and inflammatory arthritis, and between 0.90 and 0.95 when classifying RA and PsA.

3.4. Predictors identification

3.4.1. Clinical and demographic

Pain-associated arthritis predictors have been studied in [75]. Using the J48 DT algorithm, researchers predicted pain from 5,721 arthritis patients, regardless of the arthritic condition, with a 0.86 accuracy. From an initial set of 200 predictors including demographic, PROMs, laboratory and socio-behavioural characteristics, researchers built the final predictive model with just 12 variables. Of them, the physical and mental component summary score from the Short Form 12 (SF-12) were the most meaningful ones.

Inpatient gout flares predictors have been recently assessed in [76]. Up to 52 potential variables of five different domains were evaluated, including demographic, comorbidity, admission, disease history, and laboratory data. The researchers followed three different approaches: a clinical knowledge-driven model (logistic regression), a statistics-driven model (Lasso) and a DT model. The model validation was done with bootstrapping. Based on the C-statistics (C = 0.82) the first model was selected as the best-performing with just nine predictors. Almost half (4 out of 9) of the selected variables were chosen by the three different models such as *pre-admission urate>0.36 mmol/L*. This study reveals the importance of building an intuitive model for clinicians, and feasible to implement in a routine hospital setting.

The identification of SS disease worsening and death predictors was the aim of Becker et al. [77]. The SS disease worsening definition was agreed upon by an expert group who considered different clinical events, such as renal crisis, decreased forced vital capacity or death. A total of 42 variables were studied, including demographic (e.g., age, disease duration), laboratory parameters (e.g., Anti-Nuclear Antibody (ANA), Anti-Neutrophil Cytoplasmic Antibody (ANCA)); and other medical speciality domain variables (e.g., digital ulcer, synovitis, dyspnoea), and active disease. 228 out of 706 patients met the criteria for disease worsening. Lasso towards with multiple missing data imputation techniques were applied to find the eight most relevant predictors. Of them, five were strongly associated with disease progression (i.e., age, active digital ulcers, C-Reactive Protein (CRP), lung fibrosis, muscle weakness). The validation was assessed with bootstrap, achieving a C-index of 0.705.

3.4.2. Molecular biomarkers

Although part of the studies presented in the Disease classification section have addressed also the biomarkers identification task before building a classification model, the main purpose of other studies is to validate biomarkers rather than classify diseases. For example, investigators from [78] validated 17 novel urinary protein biomarkers of lupus nephritis using Lasso, RF, Bayesian Network (BN), PCA and graph-based clustering.

On its behalf, Riahi et al. [79] investigated the interactions of a SNP (i.e., ERAP1) in developing Beçhet's disease using a non-parametric data mining technique able to detect gene-gene or SNP-SNP interactions called Model-Based Multifactor Dimensionality Reduction (MB-MDR). Authors included 1524 patients, 748 cases and 776 controls. After applying MB-MDR, authors find plenty of synergistic and antagonistic significant interactions between ERAP1 polymorphisms and Behçet's disease development.

The predictive power of chemokines, cytokines, and biomarkers in saliva from pSS patients (n = 11) was evaluated using eight different classifiers (i.e., SVM, RF, NB, Gaussian process, AdaBoost, LR) [80]. From an initial set of 105 predictors and after applying five feature selection methods, 43 predictors remained. These predictors were grouped into a set of features. In a further step, hierarchical clustering and PCA validation were also performed. Once trained, the best-performing model was KNN with a 0.93 AUC value. This AUC was achieved with only two predictors: Interleukin (IL)-27 and Chemokine C-C Motif Ligand 4 (CCL4).

Blood serum samples were analysed by researchers in [81] to identify the correlation between vitamin D and ferritin with chronic neck pain using an ANN. The model achieved an 85% accuracy value.

Moreover, researchers in [82], also used serum to search for potential biomarkers in patients with Behçet's disease (n = 10), sarcoidosis (n = 17) and Vogt–Koyanagi–Harada disease (n = 13). After using PCA to discriminate the different biological samples, the authors used RF to output a variable importance measure. As a result, the researchers were able to identify the three miRNAs which best predicted each disease studied.

Eventually, one of the main objectives of the authors in [83], was to evaluate the predictive ability of inflammatory biomarkers along with other predictors (e.g., the severity of depression and anxiety) in patients with fibromyalgia. Sleep quality, perceived stress scale, and hospital anxiety were the variables that best predicted the widespread pain index.

3.5. Patient stratification and disease subtype identification

3.5.1. Clinical and demographic

Spielmann et al. [84] identified three clusters of connective tissue disease patients (n = 42) with anti-Ku antibodies and with similar clinical and biological features and prognosis, using hierarchical clustering on a set of 28 clinical features. The number of clusters was determined by minimising a partitioning criterion. Multiple Correspondence Analysis (MCA) was used for dimensionality reduction. From a data science perspective, the relevance of the study lies in the debate generated by the scientific community, on the suitability of the techniques employed, through letters to the publisher [85, 86, 87, 88].

A similar hierarchical clustering approach was carried out by Ogata et al. [89] to identify and clarify the characteristics of the subgroups of Antiphospholipid Syndrome (APS) patients with the poorest prognosis (n = 168). Three different clusters were identified, after visually analysing the dendrogram, by combining serological and clinical data. Although the clustering groups identified in this study were different compared to previously published results, the authors highlighted the existence of a cluster that accumulates cardiovascular risk events and arterial thrombosis events.

This unsupervised algorithm, hierarchical clustering, has also been applied to a dataset of 74 adult-onset Still's disease (AOSD) inpatients [90]. Three distinct fever patterns were characterised after computing the ideal number of clusters, k, with the Kellgren and Lawrence (KL) index. In view of the results obtained, after applying logistic regression to compare the prognosis of AOSD between the three groups, the authors showed that a higher temperature at the time of diagnosis was associated with a higher risk of AOSD-related mortality.

Attempts have been made to identify patients predisposed to the development of lymphomas associated with pSS in [91]. The dataset studied consisted of 449 pSS patients (76 of them with lymphoma) and 90 features including demographic, and laboratory measures (e.g., C3, C4, haemoglobin). Two algorithms were trained, RF and Extreme Gradient Boosting (XGBoost); and the prediction performance was assessed with cross-validation. While RF showed an AUC value of 0.83, XGBoost achieved a score of 0.88. C4 levels, the rheumatoid factor and the focus score at first biopsy were the predictors with the highest relative importance.

Another research group employed unsupervised ANN algorithms and graphs (i.e., semantic connectivity maps) to explore hidden trends and non-linear associations among clinical and serological pSS features and to predict lymphoma [92]. This research group was also responsible for studying the increased prevalence of cardiovascular events in pSS patients using, again, unsupervised ANN algorithms and agglomerative hierarchical clustering [93].

3.5.2. Molecular biomarkers

In [94], authors obtained two differentiated groups of paediatric SLE patients (n = 31), depending on the predominant component of the disease: autoimmune or autoinflammatory. K-means clustering was used for this purpose considering type I IFN score, Systemic Lupus Erythematosus Disease Activity Measure (SLEDAI)-2K and mean complement levels (C3 and C4 normalised values) variables.

On their behalf, the authors of [95] investigated how they could obtain meaningful clusters applying Non-Negative Matrix Factorisation (NMF) in a cohort of 173 patients with SS, using skin gene expression profiles. Using silhouettes score to determine the number of clusters, k, the authors satisfactorily achieved a four-cluster separation based on distinct SS activities.

The same unsupervised clustering technique, NMF, was used to identify joint involvement patterns that predicted the trajectory of the disease in JIA patients [96].

In addition, clinical and laboratory biomarkers from a cohort of 150 children with JIA were used to identify clusters [97] with the help of GMM. Visits at baseline and six months later were considered. With a feature selection process based on a variable contribution threshold, 191 features were embedded into three principal components which represented the 35% and 40% of variance. Using Bayesian Information Criteria (BIC) authors found three and five clusters. Later, the researchers tried to compare the results of the clusters with the JIA categories using circular plots. In light of the results, the authors concluded that the clusters did not match the JIA categories, suggesting that the pathobiological processes are shared between the different categories and fluctuate during the course of the disease.

Consensus clustering and k-means were the approaches chosen by the researchers to reveal three RA synovial gene expression subtypes using the top 500 most variable genes expressed in 45 RNA-seq samples [98]. The ideal number of clusters was visually confirmed according to the likelihood scores. After that, the authors used histology scores as modelling features in a standard SVM to predict the three RNA-seq subtypes.

3.5.3. Image

Gribbons et al. [99] included 1,068 patients in a study to identify groups of patients with similar patterns of Takayasu's Arteritis (TAK) and GCA large vessel vasculitis. By defining 11 arterial territories (e.g., carotid, subclavian, axillary, renal, mesenteric, and aorta) and combining catheter-based, magnetic resonance, computed tomographic angiography, ultrasonography, and Fluorodeoxyglucose Positron Emission Tomography (FDG-PET) images; patients were clustered based on disease within those arterial territories. Silhouettes and gap-statistic methods were used for determining the optimal number of clusters, k, and k-means was chosen as the clustering algorithm. The key aspect of this study was the sample size for such *rare diseases*, which facilitated the use of clustering techniques. The results of this study were proposed by the authors to be considered in future classification criteria for large vessel vasculitis. In another study [100], this same research group also included a DT model to predict k-means cluster assignment between these two diseases achieving a 87.6% accuracy in the replication cohort. In both studies, independent cohorts were used for validation.

Novelty detection models to screen for myopathies and find rare presentations of myopathic disease have been presented in [101]. In this study, the authors used 3,586 US images and tried different novelty detection algorithms: discriminative DL and generative methods. The best-performing approach resulted in a 0.72 AUC value.

The researchers in [102], fitted 15 classifiers to a cohort of 92 osteoporosis patients to predict fragility fractures from MRI data. Up to 6 different datasets with different features were employed. The average F1 score obtained was 0.63 for all features dataset.

Ultimately, in [103], the authors calculated features from the histogram of MRI images (e.g., kurtosis, skeweness, maximum pixel value) and trained three ML methods, including SVM, KNN and Multilayer Perceptron (MLP), for computer-aided classification of active inflammatory sacroiliitis, to aid in the Spondyloarthritis (SpA) classification. Some of the patients developed axSpA and some others OA, fibromyalgia, gout, or psychiatric disorders. The best classifier achieved an accuracy value of 0.80.

3.6. Disease progression and disease activity

3.6.1. Clinical and demographic

Topic modelling has been applied to characterise the temporal evolution of ANCA-Associated Vasculitis (AAV) in [6]. With a follow-up of seven years, more than 113,000 clinical notes from 660 patients were processed following a topic modelling approach. Temporal trends, before and after the treatment initiation date for a diagnosis of AAV were modelled with LDA finding 90 different topics that included diagnosis, treatments, comorbidities, and complications of AAV. The authors showed the suitability of this unsupervised method to provide unique information on the clinical course of a disease that could not be captured in the structured data from the EHR. As the researchers mentioned, identifying the topics is of special relevance in multiorgan diseases where structured data fields are unlikely to reflect the full extent of signs, symptoms, comorbidities and complications.

In [104], investigators developed a predictive model for SLE disease activity based on routinely available demographic, clinical, and laboratory data. With this model, they tried to overcome the current limitations of using a complex composite index such as SLEDAI-2K to measure disease activity. The authors used 16 pathological variables and a multinomial logistic regression approach that was compared with the performance of a NB model. To select the best-performing model from the space search (i.e., 2^n, where $n = 16$), an 0.82 AUC threshold was fixed. Models with up to 8 variables that did not include anti-dsDNA assay results were selected. Finally, researchers found that the multinomial approach overcomes the performance of NB, with 0.83 and 0.66 AUC values respectively.

One remarkable study on the application of DL techniques to RA patients was carried out by Norgeot et al. [15] In their research, the authors aimed to forecast RA disease activity, measured using Crohn's Disease Activity Index (CDAI), in future clinic visits using 45 structured variables from the EHR, including Disease-Modifying Antirheumatic Drugs (DMARDs), corticosteroids, CDAI score, Erythrocyte Sedimentation Rate (ESR), CRP, anti-Cyclic Citrullinated Peptides (CCP), rheumatoid factor and demographic variables. In doing this, the authors imitated time series forecasting studies, creating sliding time windows of a fixed interval to model the longitudinal nature of the study. Permutation Importance Scores (PISs), to measure the contribution of each independent variable, were calculated. Dense, time-distributed, convolutional, and recurrent layers topologies were tested. The best model achieved a 0.91 AUC value in the test set.

Another RA study conducted in Spain, focused on developing a RA mortality predictive model, trained (n = 1,461) and validated (n = 280) with Random Survival Forests (RSF), a supervised predictive algorithm [13]. In this study, nine demographic and clinical related variables, such as age at RA diagnosis, gender or presence of rheumatoid factor;

and collected during the two years after diagnosis, were included in the final model after a variable importance analysis according to their predictive ability. The authors identified three different mortality risk groups (low, intermediate, and high) using the predicted ensemble mortality. The prediction error in the validation cohort was 0.233.

Under the hypothesis that RA and axSpA flares are associated with physical activity, a French research group took advantage of wearable activity trackers and the NB algorithm to study potential associations between both concepts [8]. With 155 patients, 82 RA and 73 axSpA, and 1,339 weeks evaluated, the authors concluded that patient-reported flares were strongly linked to physical activity. To reach this result, different wearable data time levels of aggregation were considered, the binary variable flare/no flare was used as the dependent variable, and the performance of the models was evaluated using patient-reported flares as the gold standard, assessed every week.

3.6.2. Molecular biomarkers

In [105], authors used data from the Osteoarthritis Initiative to identify serum antibodies that could predict radiographic knee OA in asymptomatic individuals (measured with the KL scale) that will develop the disease before 96 months using multivariate logistic regression analysis. An AUC value of 0.76 (95% CI: 0.66-0.86) was achieved in the replication phase. Osteoarthritis Initiative (OAI) database was also used in [106] as an independent cohort to predict radiographic progression using peripheral blood leukocyte inflammatory gene expression (IL-1β, TNFα, and Cyclooxygenase-2 (COX-2)), using SVM models.

Radiographic progression in RA patients was evaluated in [107], using Lasso regression, in a model with clinical and gene covariates. The authors of this study also evaluated the treatment response to conventional synthetic DMARDs (csDMARDs) in early RA patients.

Clusters enriched in active SLE (quantified using SLEDAI) were calculated using hierarchical clustering in 140 SLE patients [108]. In this study, the bootstrap forest model was used to predict SLE activity and to identify potential predictors related to this activity. The results were validated performing multivariable logistic analyses.

Poppenberg et al. considered different ML algorithms (i.e., KNN, RF and SVM) to predict JIA disease activity from transcriptomes from peripheral blood mononuclear cells [109]. In this research, the authors defined active disease according to the presence of physical signs of synovitis in at least one joint. The identification of predictors (i.e., transcripts with the greatest predictive power) was performed with Lasso. Finally, the 35 genes identified were used as the input of four different predictive models. RF outperformed the rest of the models in the testing cohort, achieving a 0.94 AUC value.

Circulating protein biomarkers capable of distinguishing between active vasculitis and remission in GCA (n=60), TAK (n=29), Polyarteritis Nodosa (PAN) (n=26) and Eosinophilic Granulomatosis with Polyangiitis (EGPA) (n=37) patients have been identified in a study conducted by the Vasculitis Clinical Research Consortium [110]. In this study, 22 proteins potentially linked to vasculitis were measured from samples collected during active and remission periods. J48 algorithm was used to identify biomarkers capable of distinguishing between active and inactive GCA.

3.6.3. Images

Doppler ultrasound images have proven to be helpful when training CNN architectures. Scientists have recently demonstrated its viability to automate the classification of disease activity into four degrees for RA patients, performing similarly to a human expert [111, 112]. In [111], authors proposed CNN architectures, InceptionV3 and VGG-16, for automatically scoring the disease activity of RA patients, using images from the wrist and hand of 40 patients with early or longstanding disease, achieving an accuracy between 75%. A year later, the same group of researchers [112], used a dataset of 1,678 US images, and trained different cascaded CNN architectures, InceptionV3, achieving a four-class accuracy of 83.9% and beating their previous results.

US images have also been used to discriminate between low- and high-grade synovitis [49] in inflammatory arthritis patients. In the previous study, the authors explored the ability of CNN and transfer learning, ResNet34, for discriminating between both synovitis grades, in 150 photomicrographs of 12 patients, achieving perfect accuracy.

Moreover, smartphone pictures have been proposed to train SVM, and classify hand arthritis photographs into three stages: early, moderate, and late [113]. The accuracy obtained ranged between 0.77 (classification into the three stages) and 0.97 (binary classification, healthy/unhealthy).

Relevant quantitative measures of joint degeneration often require a thorough segmentation preprocessing step. Therefore, it is not surprising that many researchers have focused on this first step as the main goal of their studies. In the case of knee OA, different authors have applied multiple DL topologies such as CNN [114, 115], conditional Generative Adversarial Network (GAN) [116] and Holistically Nested Network (HNN) [117] for the segmentation

of knee joint tissues, including, among others, femoral cartilage, tibial cartilage, patella, patellar cartilage, meniscus, quadriceps and patellar tendon, or infrapatellar fat pad.

3.7. Treatment response

3.7.1. Clinical and demographic

A large proportion of treatment response studies have been carried out to assess the efficacy of new therapeutic lines in the RA population. The response to TNFi has been investigated in multiple studies. For example, in [118], penalised regression models were used to estimate changes in ESR and Swollen Joint Count (SJC), two Disease Activity Score 28 (DAS28) composites, between 3 and 6 months after treatment initiation. Clinical and genotypic scores covariates were used to build the predictive models. Nonetheless, the authors were unable to find strong predictors of TNFi response among alleles linked to the development of RA.

In another study, authors [119] predicted changes in disease activity scores 24 months after baseline assessment (i.e., ΔDAS28), and identified non-responders to anti-TNF treatments using different ML techniques (e.g., SVM, Ridge, RF, LR, and Gaussian Process Regression (GPR)). Demographic, clinical, and genetic features were included as predictors, although the last ones did not improve the prediction accuracy. The AUC value of the best model in the independent cohort was 0.62.

On its behalf, the authors in [120] based their research on determining how the transcriptomic and epigenetic profiles of immune cell types and whole peripheral blood mononuclear cells could help to predict the response to two different TNFi prior to treatment initiation, using RF models. After 6 months, the response to treatment was evaluated for a total of 80 patients. One of the most promising conclusions of this study was the discovery of divergent gene signatures between different TNFi, suggesting a potentially different mechanism of action between them.

Furthermore, studies to assess the patient's response to classical drugs, such as methotrexate are also of interest to researchers. A group of them obtained a 0.78 AUC value when training a penalised logistic regression, Ridge, model to predict respondent patients (DAS28-CRP) at month six [121]. This score was achieved using gene transcripts expression ratio between 4-weeks and pre-treatment.

Another recently published study, tried to compare the performance of ML algorithms (i.e., Lasso, RF, XGBoost) with logistic regression in the prediction of insufficient response to methotrexate, measured using DAS28-ESR [122]. Based on the AUC results, authors concluded that there was no benefit in using ML algorithms (AUC for XGBoost 0.77) over logistic regression (AUC 0.78).

In [123], the authors fed an ANN, among other ML algorithms like XGBoost, with clinical and laboratory data from almost 600 AS patients to predict early-TNF responders, obtaining a 0.783 AUC value with the ANN model. In addition, a feature importance analysis based on gradient descent was helpful to find that CRP and ESR were the most significant baseline characteristics for predicting early-TNF responders.

NLP techniques were used to identify arthralgia in clinical notes from Inflammatory Bowel Disease (IBW) patients, as a preliminary step to compare two different treatments, one of them apparently linked to an increased risk of arthralgia due to adverse events [124]. The importance of this study lies in the fact that thanks to narrative notes and NLP, authors were able to identify a potential adverse effect where coding was suboptimal.

Web scraping techniques have been used to extract data from social media networks in a variety of contexts. Treato, a deprecated data analytics service, was used in some of them. This service incorporated NLP processing pipelines, medical ontology mapping, classifiers, and sentiment analysis, among others. For example, in an attempt to evaluate the suitability of social media as a data source for drug safety, some authors studied patient-reported herpes zoster events associated with arthritis medication [125]. For this purpose, the authors used Treato to analyse and classify more than 785,000 posts mentioning inflammatory arthritis with a PPV of 0.91.

Another social media data web scrapping research that used Treato and LDA for topic modelling, have been carried out by Dzubur et al. to examine AS patients' knowledge, attitudes, and beliefs regarding biologic therapies [126]. 27,000 posts from more than 600 social media sites were studied. The investigators found 112 topics, 67 of them focused on discussions surrounding AS treatment, such as the side-effects of biological treatments, biological attributes (e.g., dose and frequency) and concerns (e.g., cancer risk, reproductive concerns).

RA patient's perception to 13 DMARDs was assessed in [127] using Treato. This time, the NLP task was oriented to identify medical concepts and to extract patients' self-reported descriptions of their experiences with various health conditions and medications to conduct sentiment analyses. The authors found that the ratio of patients with a positive sentiment to bDMARDs and targeted synthetic DMARDs (tsDMARDs) was higher than the ratio of patients with a

positive sentiment to csDMARDs. In addition, they showed that the efficacy and side effects were the most frequently discussed topics.

Researchers have studied the response to methotrexate monotherapy [128] and to TNFi [129] in JIA patients using DAS44/ESR-3 indices. Regarding the former, [128], treatment response models before and after administration (within three months) were built in a cohort of 362 patients. The algorithms proven were XGBoost, SVM, LR and RF. A median importance ranking with ensemble methods was also computed. A set of ten predictors before and six predictors after treatment administration was chosen by the XGBoost algorithm. The performance of the model in both scenarios, before administration and before and after administration, was 0.97 and 0.99 AUC. Regarding the latter, [129], treatment response models before administration, were built in a cohort of 87 patients. The algorithms proven were XGBoost, Gradient Boosting Decision Tree (GBDT), Extremely Random Trees (ET), LR and RF. XGBoost model achieved the best performance with a 0.79 AUC value and just four features.

The cardiovascular side effects of analgesics in 4,350 patients, extracted from the OAI dataset, were modelled by an XGBoost prediction model along with a risk feature identification [130]. Of 300 demographics, anthropometry, comorbidity, blood measures, and physical activity features, the authors found and described the 20 most informative ones. The model achieved a 0.92 AUC value.

The response to treatment has also been evaluated in PsA patients (n = 2148) by authors in [131]. In this original article, the efficacy of the starting dose of secukinumab, an IL-17A inhibitor, was evaluated thanks to the bayesian elastic net ML algorithm. More specifically, the study sought to investigate whether there were specific baseline clinical characteristics that could predict which patients could gain additional benefit from the secukinumab 300 mg dose. With a cohort of 2,148 patients and 275 predictors, different efficacy endpoints (e.g., ACR20/50, PASI 75/90, PASDAD, Health Assessment Questionnaire (HAQ)-DI) were analysed at week 16. Although there was no single predictor with enough discriminatory power, the authors found that there were common covariates for different endpoints, such as the presence of enthesitis at baseline. Furthermore, the authors also identified subpopulation groups that could benefit from the 300 mg dose over the 150 mg dose, such as patients treated without concomitant Methotrexate (MTX) or patients with psoriasis. The AUC scores ranged from 0.75 to 0.81 for the different endpoints.

Serious infections in RA patients under IL-6 inhibitor treatment have been studied [132]. More precisely, researchers from Japan, extracted data using text mining approaches, from a post-marketing AE-reporting database, to identify signs and symptoms before the development of serious infection (i.e., defined by the authors as those infections in which the patient attended the hospital). Once the signs and symptoms were extracted from clinical narratives 28 days before serious infection, a codification with MedDRA Preferred Terms, and a review to determine if they were already generally known as signs or symptoms of infection was done. As a result, the authors showed that more than 60% of patients with a confirmed date of serious infection diagnosis had signs or symptoms within 28 days before that diagnosis.

Response to intravenous immunoglobulin therapy in patients with Kawasaki disease has been studied in [133]. Researchers applied seven ML algorithms and obtained a 0.72 AUC with GBM. The feature importance was evaluated with SHapley Additive exPlanations (SHAP).

Reinforcement learning and sequential decision-making algorithms have been implemented to promote physical activity in patients with Chronic Back Pain (CBP) [134] in a smartphone application. A similar approach, based on smartphones, took advantage of MLP to improve self-management of chronic neck and back pain [135].

3.7.2. Molecular biomarkers

Liu et al. [7] developed a TNF blocker treatment response predictive model after evaluating quantitative changes in IgG galactosylation, alone and in combination with AS associated SNPs. Up to eight ML models were developed with SVM, 0.87 AUC, and Flexible Discriminant Analysis (FDA) 0.82 AUC as the best performing ones.

Treatment response to Glucocorticoids (GCs), commonly used as the first-line therapy in patients with AOSD, has been studied by a research group from China [136], with a SVM predictive model. The motivation of the investigators was to balance the side effects and the effectiveness of the treatment, considering clinical and laboratory features (i.e., four neutrophil extracellular traps proteins). With this in mind, the authors developed two SVM models. The first one, tried to assess whether the proteins could serve as biomarkers and the second to predict if the levels of these circulating proteins could predict treatment response in terms of resistance to low-dose GCs. To evaluate the second outcome, the authors categorised the treatment response into a binary variable with low and high GCs levels as the dependent variable. The AUC values for the first and the second outcomes were 0.88 and 0.91, respectively.

3.7.3. Image

DL and ML algorithms to assess treatment response have also been developed in the context of radiological images.

For instance, Chandrika et al. [137] presented an architecture to assess bisphosphonates response in 28 patients with Chronic Non-bacterial Osteitis (CNO). The number of included patients was 28 and the number of pairs of images 55. The proposed architecture consists of two components followed by an ensemble method, which classifies scans as "improved", "worse", or "stable". The first, used an InceptionV3 network to extract features, embeddings and representations, that were used in a linear logistic model to produce a probability score. The second component, used unsupervised clustering techniques to label the images and SVM to produce a probability score. Although the results were not remarkable (i.e., low specificity and accuracy), presumably due to class imbalance and low training examples, this study showed that rare RMDs research could also benefit from data science techniques.

3.8. Illustrative studies

Table 3 shows a subcollection of the revised research articles that exemplifies the wide variety of RMDs studies in which AI has been an essential tool. This table structure is inspired by [32]. With the articles presented in this table, more than 20 different diseases are covered, exhibiting the following:

- AI techniques can be used for multiple purposes: as the tools necessary to perform the primary statistical analysis or as complementary tools that help researchers achieve their main objectives. Supervised learning (classification, regression), unsupervised learning (clustering, topic modelling, dimension reduction, novelty detection, visualisation), reinforcement learning (recommendation systems), deep learning (computer vision) and other procedures and techniques, such as feature selection or transfer learning are being used in RMDs research. For instance, Lasso, and tree-based algorithms (e.g., RF, DT, XGBoost) are commonly used supervised techniques, while PCA and clustering algorithms (e.g., k-means, GMM) unsupervised.

- The sample size of the input data can range from a few patients to thousands. Large cohorts of patients are not indispensable to take advantage of AI techniques in RMDs research studies. In addition, the tnput data can come from multiple sources: clinical and demographic data, gene data, image data and wearable activity tracker and sensor data.

- The topics addressed with these techniques are varied, the algorithms used numerous, and the potential results promising. As an example, in *rare diseases*, AI techniques can also be useful and have a real impact, as they can be used to obtain new insights and findings from clinical notes.

- The 57% of the articles presented in Table 3 are published in rheumatology specialised journals.

Some of the findings listed above have also been discussed in other reviews, such as [28].

4. Discussion and conclusion

In this review, we have explored the clinical and technical background of RMDs that motivate the employment of AI techniques for research. Different contributing factors, such as the longitudinal, multidimensional and heterogeneous nature of data seem to facilitate the adoption of such techniques. We have also introduced the review articles published since 2017. These articles have addressed multiple topics, all of them with a strong rheumatology component: wearable activity trackers, bioethical perspectives, RWD, DL and so on. Then we performed a literature review considering four different sources. The articles retrieved in this review were classified into six thematic categories. For each category, we made a distinction depending on the main predictor type of the study. Furthermore, we provided a table with more than twenty illustrative studies that exemplify the wide variety of RMDs in which AI techniques have penetrated. We also mentioned the limitations of ML algorithms with a particular focus on rheumatology.

From the 91 articles finally included we could appreciate some interesting findings. For example, it seems that the RMDs that accounts for the majority of the research studies in which data science techniques are used more assiduously are: OA, osteoporosis, RA and SLE. It is hypothesised that the following factors may contribute to that fact:

- OA and RA are the most common rheumatic diseases [138]. In addition, traditionally, a considerable part of the research efforts and funding have focused on both diseases.

- As medical images can be used to diagnose and quantify the status and progression of a disease, especially in OA, computer vision and DL algorithms may have attracted researchers, with a special interest in those techniques.

On the other hand, data mining techniques have allowed studying and characterising *rare diseases* in a different way, obtaining valuable information and uncovering new patterns that probably would not have been discovered otherwise. For instance, unsupervised learning techniques, and more concretely clustering, have been decisive in characterising disease subgroups. Moreover, some research groups have recognised the potential of these techniques and have adopted them as relevant tools for knowledge extraction when studying the pathology of specific diseases. The Big Data Sjögren Project Consortium is an example of this [139].

AI adoption is growing yearly in rheumatology research, as shown by the trend in the number of publications retrieved in Medline in a five-year period. Not only common algorithms, but also the latest advances are being employed. After the study inclusion period of this review, new approaches have blossomed and have been applied to RMDs research, highlighting the interest that AI raises to researchers. From FSL approaches for early RA prediction to the employment of SHAP and GBMT to stratify patients with RA according to the trend of disease activity [3]. Regarding the data science languages employed, R and Python are the most widespread, but others such as Matlab, Weka, SPSS, Stata or JMP have also been used. However, further efforts must be made to validate the models in independent cohorts. The EULAR points to consider highlighted in 2019 the importance of this issue [19]: *conclusions drawn from big data need independent validation (in other datasets) to overcome current limitations and to assure scientific soundness*. As shown in Supplementary Excel File 'Included Articles', this point to consider is not fully addressed, although it seems to be gaining relevance.

4.1. Limitations of machine learning algorithms

Many technical and ethical limitations of ML algorithms are not exclusive to clinical research, but are also shared with other research fields. Focusing on the technical limitations, authors in [140] highlighted six main points that may affect the performance of a model, making a distinction between *bad algorithms* and *bad data*. Particularising the six points to the medical field:

1. **Insufficient training data**: autoimmune rheumatic diseases (e.g., Mixed Connective Tissue Disease (MCTD), polymyositis, SS, vasculitis) with low prevalence may be particularly affected by this problem. The reduced number of cases can hinder researchers from drawing valid conclusions. Therefore, multi-centre studies and database sharing may be proposed as efficient approaches that may mitigate this problem. Technical efforts have also been made with different methods such as FSL. For instance, this type of ML approach has recently been applied in MRI images from RA patients [18].
2. **Non-representative training data**: algorithms not trained with representative data are likely to fail with unseen population instances. This translates into the need to consider patients with different demographic and clinical characteristics (e.g., ethnicity, educational level, and so on). Otherwise, the model will not be robust enough to ensure generalisation to new cases. In the clinical field, subtle differences, such as different patient management procedures and drugs employed between centres, may hinder the algorithm generalisation capability. A broader discussion is held in [141].
3. **Poor-quality data**: in contrast to the data generated in a clinical trial context where patients are thoroughly selected and the recorded variables are perfectly defined; RWD is more complex, heterogeneous, noisy and less structured; with an increased risk of reaching inaccurate results and erroneous conclusions [142].
4. **Irrelevant features**: feature engineering and feature extraction processes should ensure that selected features are relevant to build statistical models. Sometimes, extracting clinical features is not straightforward, easily accessible (e.g., data protection regulation), or cost-efficient (e.g., genomic analysis). For example, in GWAS analysis there may be hundreds of thousands of gene variants and only a few of them are useful to predict the disease status of a patient. Moreover, the addition of unnecessary features may degrade the performance of the model, adding noise to the data.
5. **Overfitting**: the lack of generalisation capability to new cases can lead to useless models. Gathering more training data to minimise this problem is not always easy in clinical practice (especially for *rare diseases*), and reducing noise and outliers can be time-consuming, when possible. The *high dimensionality problem*, which characterises genetic studies, can be especially exacerbated by overfitting. Furthermore, in low-dimensional data, overfitting can have a non-negligible impact if the relationship between the outcome to the set of predictor variables is not strong [143]. Regularisation techniques (e.g., feature selection, less flexible models, constraints) constitute the most immediate solution for fighting overfitting. Underfitting should also be considered.

Other specific limitations [34, 144] that gain special relevance in the medical research field are:

- Strict and complex regulatory requirements, specially in the model implementation and adoption phases.
- Lack of standardisation: data acquisition (e.g., manually or digitally), and different medical procedures or treatment guidelines may complicate and slow down the use of multiple databases for training the algorithms; introducing heterogeneity and obstructing the preprocessing steps. Multicentre studies should consider the different variables, QoL questionnaires (e.g., SF-12, EuroQol5D (EQ-5D)), and the variable measurement employed (e.g., mmol/L, mg/dl) in each participant centre to have comparable data.
- External validation: the performance of the models should be evaluated on different validation cohorts to extract reliable conclusions. Otherwise, the generalisation ability of the model will probably remain unclear or perform poorly; however, this is not always possible. Cross-Validation (CV), bootstrapping or unsupervised techniques may be useful when the external validation is not guaranteed.
- Explainability: Most flexible statistical learning algorithms, such as NN are frequently considered "black-box" models. This raises ethical issues, such as how findings that may have an impact on a patient's health status can be applied without a clear perspective of why and how the model works the way it does. In general, when building prediction models, the researcher's objective is to obtain the smallest set of characteristics capable of predicting an outcome with the maximum predictive performance [145]. Easy, understandable, economic, and accessible predictors are usually preferred, since well-performing models are expected to be deployed on a large scale, regardless of the means or resources of the different centres or countries. When choosing and deploying the different algorithms, the researcher must consider a trade-off between interpretability and flexibility. Interpretable algorithms are preferred when the main goal is to find associations between a set of predictors and a dependent variable, as well as in inference studies. When the aim of the researcher is to build the best predictive or classification model, the chosen algorithm might be non-linear, and therefore, end with a hardly or unexplainable model but with good performance. This limitation has recently been discussed in [146, 147], but the debate remains open, as authors pointed out: *Black-box medical practice hinders clinicians from assessing the quality of model inputs and parameters. If clinicians cannot understand the decision-making, they might violate patients' rights to informed consent and autonomy*. Finally, some technical efforts have been made to improve the explainability, such as SHAP.

The European Commission (EC) has been working on the issues listed above for years [148], and it is about to publish the first-ever legal framework on AI [149]. Expanding on this topic, the first Ethics Guidelines for Trustworthy Artificial Intelligence [150] created by the EC have highlighted seven requirements for AI systems to be deemed trustworthy. How to apply them in the healthcare domain has been explained in [151]. The UK government has recently published the first National AI Strategy Action Plan [152]. Finally, other recently published articles have provided a detailed outline of more specific limitations and points to consider that apply to RMDs [21, 19, 32].

4.2. Limitations of the review

This review article has some limitations:

- The keywords used during the search in the different sources may omit some potential articles. For instance, common acronyms AI or ML were not used. In addition, the keywords employed are not exactly the same in the different sources. As this was a general overview of the state-of-the-art, searches for specific diseases were not done (e.g., replacing the terms *rheumatology, rheumatic and musculoskeletal* in the different queries or *low-back pain*). This may limit the number of articles retrieved. In addition, the rheumatology journals search may introduce some bias since this search was limited to Q1 and Q2 journals of a specific year. Furthermore, articles without a PMID were excluded. However, by combining four different data sources, we tried to reduce this shortcoming.
- We did not provide an introduction to the different learning methods, so the potentially interested audience in this review article may be shortened. As explained in the introduction section, this was made to maximise the discussion of articles. Nevertheless, we provided enough references for the people interested in deepening in the technical background. In addition, we employed *data science* and *artificial intelligence* terms indistinctly during this review. This is not entirely correct as subtle differences exist.

- Numerous studies addressing data mining techniques in RMDs research have been published from the state-of-the-art cutoff date (i.e., February, 22th 2021) to the date of this manuscript submission. From studies with the aim to distinguish PsA, seronegative, and seropositive RA patients based on hand MRI using an ANN [153], to studies that examine the validity of ML models in predicting GCA flares after GCs tapering [154]. However, the review presented here addresses the main topics in a detailed way.

- The classification proposed into six main topics, may not be suitable for capturing subtle differences between articles. In fact, establishing the topic of an article following this classification is sometimes arduous, as the frontier of the different topics is fuzzy. For instance, the *disease classification* and the *disease prediction* categories differ only in the presence of healthy and sick patients, rather than groups of patients with different pathologies. However, since the number of approaches in which *disease classification* is lower and less studied, we tried to give enough relevance to this particular case. We have also faced this issue when trying to assign articles to the *predictors identification* and *disease progression and activity* topics.

A. Appendix

Table 3: Examples of data mining techniques applied to different RMDs. Illustrative research articles

Publication	Disease	Aim	Sample size	Type of analysis	Input data type	Algorithm	Results (best model)
Kim MJ et al. (2018) [90]	AOSD	To study the association between fever pattern and clinical manifestations of AOSD	70	Clustering	Electronic medical reports	HC	Three clusters. Group 1 (n = 14), 2 (n = 35), 3 (n = 21). Clinical features and extent of organ involvement did not differ significantly between groups.
Ogata Y et al. (2021)[89]	APS	To identify the subgroup of patients with APS with the poorest prognosis and clarify the characteristics of that subgroup	168	Clustering	Medical reports and clinical and laboratory data	HC	Three clusters. A (n= 61), B (n = 56), C (n = 61). Cluster B showed significantly higher frequency of the events and higher mortality compared with the other clusters
Gribbons KB et al. (2020)[99]	Arteritis	To identify and validate, patterns of arterial disease in Takayasu's arteritis and giant cell arteritis	Two cohorts: 1 (n = 678) 2 (n = 390)	Clustering	Vacular imaging data. angiography, ultrasonography, or FDG-PET	K-means	Six clusters. Group 1 (n = 92), 2 (n = 150), 3 (n = 144), 4 (n = 166), 5 (n = 65), 6 (n = 61). The location of arterial disease and the nature of disease involvement should be considered in future classification criteria for large vessel vasculitis.
Deodhar A et al. (2020)[52]	AS	Early identification of ankylosing spondylitis	Cohort segment 1 Cases = 6325 Controls = 49198 Cohort segment 2 Cases = 1242 Controls = 227,229	Classification	Medical and pharmacy claims history of patients	Bagging, RF, ET, GB, lSVM, rbfSVM, EM	AUC = 0.63 Sensitivity = 0.10
Zhao SS et al. (2020) [54]	axSpA	To develop algorithms incorporating codified and narrative data to identify axSpA patients in EHR, and to compare them against traditional approaches	Cases = 127 Controls = 423	Classification	Electonic medical reports	LR, Lasso MAP	AUC = 0.93 Sensitivity = 0.78 Specificity = 0.94
Rabbi M et al. (2018)[134]	CBP	To find routine behaviors and generate physical activity recommendations that are similar to existing behaviors	10	Recommendation system	Sensor data	MAB	Recommendations from the RL system easier to adopt compared to the control (β_{int}=0.42, P<.001)
Riahi P et al. (2020)[79]	Behçet's	To investigate the interactions of ERAP1 SNPs using MB-MDR	Cases = 748 Controls = 776	Dimension reduction	Gene data	MDR	-
Spielmann L et al. (2019)[84]	Overlap Syndrome	To identify subgroups of anti-Ku-positive patients without an a priori diagnosis	42	Clustering	Clinical, demographic and serological	HC	Three clusters. Group 1 (n = 15), 2 (n = 18), 3 (n = 9). Clusters have distinct phenotype and outcomes, easily distinguishable at first evaluation by simple testing
Andrés-Rodríguez L et al. (2019)[83]	Fibromyalgia	To predict the fibromyalgia symptoms scale rating (Widespread Pain Index and Symptom Severity Scale scores) as well as the diagnosis	Cases = 748 Controls = 776	Regression and classification	Clinical, demographic, immune biomarkers data	NN	Fibromyalgia Symptoms rating subscales scores: R = 0.625 FM diagnoses: Sensitivity = 0.86 Specificity = 0.89
Tedeschi SK et al. (2020)[60]	Gout	To identify a cohort of definite/probable pseudogout cases from EHR data	900	Topic modelling	Electronic medical reports	sureLDA + regression	AUC = 0.70 Sensitivity = 0.42 Specificity = 0.98 PPV = 0.81
Rezaei E et al. (2020) [97]	JIA	To identify clusters in children with JIA and to determine cluster alignment with JIA categories	150	Clustering	Demographic, clinical, functional ability, quality of life and laboratory	GMM	Three and five clusters were identified for baseline visit and second visit. None of the clusters aligned with any JIA categories

Table 3 continued from previous page

Publication	Disease	Aim	Sample size	Type of analysis	Input data type	Algorithm	Results (best model)
Wang T et al. (2020)[51]	Kawasaki	To predict intravenous immunoglobulin resistance in Kawasaki disease patients	Training = 497 Test = 147	Classification	Demographic, clinical Sonography, laboratory	LR, DT RF, AdaBoost GB, lightGBM	AUC = 0.74 Sensitivity = 0.30 Specificity = 0.99
Burlina P et al. (2019)[101]	Myopathy	To find anomalies in US images via DL and to find rare presentations of myopathic disease	Cases = 35 Controls = 54	Novelty detection	Image data (ultrasound)	DCNN (VGG-16) GAN	AUC = 0.71
Eloqayli H et al. (2018)[81]	Neck pain	To determine if Ferritin and Vitamin D are modifiable risk factors with chronic neck pain	Cases = 54 Controls = 54	Classification	Clinical	FFNN	Accuracy = 0.85 Sensitivity = 0.87 Specificity = 0.83
Lim J et al. (2019)[59]	OA	To detect the occurrence of OA using patient's statistical data of medical utilisation and health behavior information	Cases = 1240 Controls = 4509	Classification	Clinical and demographic	DNN	Accuracy = 0.72 Sensitivity = 0.66 Specificity = 0.73
Ferizi U et al. (2019)[102]	Osteoporosis	To predict osteoporotic bone fractures and to highlight the imaging features and the anatomical regions that contribute most to prediction performance	Cases = 32 Controls = 60	Classification	Clinical and image data (Magnetic resonance)	PCA, LR LD, lSVM, qSVM, cSVM, DT, KNN Ensemble	F1 = 0.63 +- 0.04
Gottlieb AB et al. (2020)[131]	PsA	To predict which psoriatic arthritis patients may gain additional benefit from a starting dose of secukinumab 300 mg over 150 mg	2049	Classification	Clinical	BEN	AUC = 0.81
Norgeot B et al. (2019)[15]	RA	To prognosticate the state of disease activity of RA patients at their next clinical visit	820	Classification	Clinical and demographic variables	Dense, GRU LSTM, CNN	AUC = 0.74
Faleiros MC et al. (2020)[103]	Sacroiliitis	To evaluate the applicability of ML models and feature selection methods for the classification of active inflammatory sacroiliitis in MRI	Cases = 24 Controls = 32	Classification	Image data (Magnetic resonance)	SVM, KNN ANN	Accuracy = 0.80 Sensitivity = 1 Specificity = 0.67
Burns JE et al. (2020)[67]	Sarcopenia	To detect and analyse truncal musculature at multiple lumbar vertebral levels and muscle groupings on abdominal CT for central sarcopenia detection	102	Segmentation	Image data (Computer tomography)	CNN(U-Net)	Dice coefficients from 0.938 to 0.946
Baldini C et al. (2018)[92]	pSS	To compare ANN vs PCA in discriminating different pSS subsets	542	Classification	Clinical, demographic immune biomarkers data	ANN, PCA	Sensitivity = 0.96 Specificity = 0.96
Adamichou C et al. (2021)[56]	SLE	To construct a model that can accurately detect SLE against competing rheumatologic conditions	Discovery cohort: Cases = 401 Controls = 401 External cohort: Cases = 512 Controls = 143	Classification	Clinical, demographic, immune biomarkers data	RF Lasso + LR	AUC = 0.95
Jamian L et al. (2019)[55]	SS	To develop and validate algorithms that incorporate different variables of different nature to accurately identify SSc patients in the EHR	Training = 200 Validation = 100	Classification	Electronic medical reports	CART, RF	PPV = 0.84 Sensitivity = 0.92 F-score = 0.88
Venerito V et al. (2021)[49]	Synovitis	To accurately identify synovitis' grade according to Krenn's synovitis score	Patients: 6 PsA, 5 RA, 1 SpA Training: 90 images Validation: 30 images Test: 30 images	Classification	Photomicrographs	CNN (ResNet34)	AUC = 1
Asakage M et al. (2020)[82]	Uveitis	To identify biomarkers in three major forms of uveitis. Variable selection	Cases = 40 Controls = 11	Feature selection	Gene data	RF, HC, PCA	AUC Behçet's = 0.963 AUC Sarcoidosis = 0.847 AUC VKH = 0.821

Table 3 continued from previous page

Publication	Disease	Aim	Sample size	Type of analysis	Input data type	Algorithm	Results (best model)
Wang L et al. (2021)[6]	Vasculitis	To discover clinically-relevant themes across longitudinal notes to study AAV	660	Topic modelling	Electronic medical reports	Mallet, LDA	From 90 proposed topics they were grouped into 8 clinically relevant

ANN: Artificial Neural Network, AUC: Area Under the Curve, BEN: Bayesian Elastic Net, CART: Classification And Regression Trees, cSVM: cubic Support Vector Machines, CT: Computer Tomography, DCNN: Deep Convolutional Neural Network, DL: Deep Learning, DNN: Deep Neural Network, DT: Decision Trees, EHR: Electronic Health Record, EM: Ensemble Models, ET: Extra Trees, FFNN: Feed Forward Neural Network, GAN: Generative Adversarial Network, GB: Gradient Boosting, GBM: Gradient Boosting Modelling, GMM: Gaussian Mixture Models, GRU: Gated recurrent units, FDG-PET: Fluorodeoxyglucose Positron Emission Tomography, HC: Hierarchical clustering, KNN: K-Nearest-Neighbor, LDA: Latent Dirichlet Allocation, LD: Linear Discriminant, LSTM: Long-short Term Memory, lSVM: linear Support Vector Machines, MB-MDR: Model Based Multifactor Dimensionality Reduction, MAB: Multi Armed Bandit (Reinforcement learning), MAP: Multimodal Automated Phenotyping, MB-MDR: Model-Based Multifactor Dimensionality Reduction, ML: Machine Learning, MRI: Magnetic Resonance Imaging, NN: Neural Network, rbfSVM: radial basis function Support Vector Machines, PCA: Principal Component Analysis, qSVM: quadratic Support Vector Machines, RF: Random Forest, SNP: Single Nucleotide Polymorphism US: Ultrasound

CRediT authorship contribution statement

Alfredo Madrid-García: Conceptualization of this study, methodology, review, writing (original draft preparation). **Beatriz Merino-Barbancho:** Methodology, writing (original draft preparation). **Alejandro Rodríguez-González:** Conceptualization of this study. **Benjamín Fernández-Gutiérrez:** Conceptualization of this study. **Luis Rodríguez-Rodríguez:** Conceptualization of this study, methodology. **Ernestina Menasalvas-Ruiz:** Conceptualization of this study, methodology, writing (original draft preparation).

All of the authors were involved in the drafting and/or revising of the manuscript.

Acknowledgments: **Lydia Abásolo-Alcázar** for her continuous feedback.

References

[1] Désirée van der Heijde, David I Daikh, Neil Betteridge, Gerd R Burmester, Afton L Hassett, Eric L Matteson, Ronald van Vollenhoven, and Sharad Lakhanpal. 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). *Annals of the Rheumatic Diseases*, 77(6):829–832, 2018. ISSN 0003-4967. doi: 10.1136/annrheumdis-2017-212565. URL https://ard.bmj.com/content/77/6/829.

[2] Stephen Bevan. Economic impact of musculoskeletal disorders (MSDs) on work in Europe. *Best Practice & Research: Clinical Rheumatology*, 29(3):356–373, 2015. ISSN 1521-6942. doi: http://dx.doi.org/10.1016/j.berh.2015.08.002. URL https://www.clinicalkey.com/#!/content/1-s2.0-S1521694215000947.

[3] Bon San Koo, Seongho Eun, Kichul Shin, Seokchan Hong, Yong-Gil Kim, Chang-Keun Lee, Bin Yoo, and Ji Seon Oh. Differences in trajectory of disease activity according to biologic and targeted synthetic disease-modifying anti-rheumatic drug treatment in patients with rheumatoid arthritis. *Arthritis Research & Therapy*, 24:233, 10 2022. ISSN 1478-6362. doi: 10.1186/s13075-022-02918-3.

[4] Jutta Richter, Christina Kampling, and Matthias Schneider. *Electronic Patient-Reported Outcome Measures (ePROMs) in Rheumatology*, pages 371–388. Springer International Publishing, Cham, 2016. ISBN 978-3-319-32851-5. doi: 10.1007/978-3-319-32851-5_15. URL https://doi.org/10.1007/978-3-319-32851-5_15.

[5] Rachel Knevel and Katherine P Liao. From real-world electronic health record data to real-world results using artificial intelligence. *Annals of the Rheumatic Diseases*, 2022. ISSN 0003-4967. doi: 10.1136/ard-2022-222626. URL https://ard.bmj.com/content/early/2022/09/23/ard-2022-222626.

[6] Liqin Wang, Eli Miloslavsky, John H. Stone, Hyon K. Choi, Li Zhou, and Zachary S. Wallace. Topic modeling to characterize the natural history of anca-associated vasculitis from clinical notes: A proof of concept study. *Seminars in Arthritis and Rheumatism*, 51(1):150–157, 2021. ISSN 0049-0172. doi: https://doi.org/10.1016/j.semarthrit.2020.10.012. URL https://www.sciencedirect.com/science/article/pii/S0049017220303115.

[7] Jing Liu, Qi Zhu, Jing Han, Hui Zhang, Yuan Li, Yanyun Ma, Dongyi He, Jianxin Gu, Xiaodong Zhou, John D Reveille, Li Jin, Hejian Zou, Shifang Ren, and Jiucun Wang. IgG Galactosylation status combined with MYOM2-rs2294066 precisely predicts anti-TNF response in ankylosing spondylitis. *Molecular Medicine*, 25(1):25, 2019. ISSN 1528-3658. doi: 10.1186/s10020-019-0093-2. URL https://doi.org/10.1186/s10020-019-0093-2.

[8] Laure Gossec, Frédéric Guyard, Didier Leroy, Thomas Lafargue, Michel Seiler, Charlotte Jacquemin, Anna Molto, Jérémie Sellam, Violaine Foltz, Frédérique Gandjbakhch, Christophe Hudry, Stéphane Mitrovic, Bruno Fautrel, and Hervé Servy. 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 & Research*, 71(10):1336–1343, 2019. doi: https://doi.org/10.1002/acr.23768. URL https://onlinelibrary.wiley.com/doi/abs/10.1002/acr.23768.

[9] Lindsey C. McKernan, Matthew C. Lenert, Leslie J. Crofford, and Colin G. Walsh. Outpatient engagement and predicted risk of suicide attempts in fibromyalgia. *Arthritis Care & Research*, 71(9):1255–1263, 2019. doi: https://doi.org/10.1002/acr.23748. URL https://onlinelibrary.wiley.com/doi/abs/10.1002/acr.23748.

[10] Oliver S. Burren, Guillermo Reales, Limy Wong, John Bowes, James C. Lee, Anne Barton, Paul A. Lyons, Kenneth G. C. Smith, Wendy Thomson, Paul D. W. Kirk, and Chris Wallace. Genetic feature engineering enables characterisation of shared risk factors in immune-mediated diseases. *Genome Medicine*, 12:106, 12 2020. ISSN 1756-994X. doi: 10.1186/s13073-020-00797-4.

[11] Anas Z. Abidin, Botao Deng, Adora M. DSouza, Mahesh B. Nagarajan, Paola Coan, and Axel Wismüller. Deep transfer learning for characterizing chondrocyte patterns in phase contrast x-ray computed tomography images of the human patellar cartilage. *Computers in Biology and Medicine*, 95:24–33, 2018. ISSN 0010-4825. doi: https://doi.org/10.1016/j.compbiomed.2018.01.008. URL https://www.sciencedirect.com/science/article/pii/S0010482518300167.

[12] Valentina Pedoia, Berk Norman, Sarah N. Mehany, Matthew D. Bucknor, Thomas M. Link, and Sharmila Majumdar. 3d convolutional neural networks for detection and severity staging of meniscus and pfj cartilage morphological degenerative changes in osteoarthritis and anterior cruciate ligament subjects. *Journal of Magnetic Resonance Imaging*, 49(2):400–410, 2019. doi: https://doi.org/10.1002/jmri.26246. URL https://onlinelibrary.wiley.com/doi/abs/10.1002/jmri.26246.

[13] José M. Lezcano-Valverde, Fernando Salazar, Leticia León, Esther Toledano, Juan A. Jover, Benjamín Fernandez-Gutierrez, Eduardo Soudah, Isidoro González-Álvaro, Lydia Abasolo, and Luis Rodriguez-Rodriguez. Development and validation of a multivariate predictive model for rheumatoid arthritis mortality using a machine learning approach. *Scientific Reports*, 7(1):10189, 12 2017. ISSN 2045-2322. doi: 10.1038/s41598-017-10558-w. URL http://www.nature.com/articles/s41598-017-10558-w.

[14] Yuyu Ishimoto, Amir Jamaludin, Cyrus Cooper, Karen Walker-Bone, Hiroshi Yamada, Hiroshi Hashizume, Hiroyuki Oka, Sakae Tanaka, Noriko Yoshimura, Munehito Yoshida, Jill Urban, Timor Kadir, and Jeremy Fairbank. Could automated machine-learned mri grading aid

epidemiological studies of lumbar spinal stenosis? validation within the wakayama spine study. *BMC Musculoskeletal Disorders*, 21:158, 12 2020. ISSN 1471-2474. doi: 10.1186/s12891-020-3164-1.

[15] Beau Norgeot, Benjamin S. Glicksberg, Laura Trupin, Dmytro Lituiev, Milena Gianfrancesco, Boris Oskotsky, Gabriela Schmajuk, Jinoos Yazdany, and Atul J. Butte. Assessment of a Deep Learning Model Based on Electronic Health Record Data to Forecast Clinical Outcomes in Patients With Rheumatoid Arthritis. *JAMA Network Open*, 2(3):e190606–e190606, 03 2019. ISSN 2574-3805. doi: 10.1001/jamanetworkopen.2019.0606. URL https://doi.org/10.1001/jamanetworkopen.2019.0606.

[16] Yanshan Wang, Yiqing Zhao, Terry M. Therneau, Elizabeth J. Atkinson, Ahmad P. Tafti, Nan Zhang, Shreyasee Amin, Andrew H. Limper, Sundeep Khosla, and Hongfang Liu. Unsupervised machine learning for the discovery of latent disease clusters and patient subgroups using electronic health records. *Journal of Biomedical Informatics*, 102:103364, 2020. ISSN 1532-0464. doi: https://doi.org/10.1016/j.jbi.2019.103364. URL https://www.sciencedirect.com/science/article/pii/S1532046419302849.

[17] Clayton A. Turner, Alexander D. Jacobs, Cassios K. Marques, James C. Oates, Diane L. Kamen, Paul E. Anderson, and Jihad S. Obeid. Word2Vec inversion and traditional text classifiers for phenotyping lupus. *BMC Medical Informatics and Decision Making*, 17(1):126, 12 2017. ISSN 1472-6947. doi: 10.1186/s12911-017-0518-1. URL https://bmcmedinformdecismak.biomedcentral.com/articles/10.1186/s12911-017-0518-1.

[18] Yanli Li, Denis P. Shamonin, Tahereh Hassanzadeh, Monique Reijnierse, Annette H.M. van der Helm-van Mil, and Berend Stoel. A simple but effective training process for the few-shot prediction task of early rheumatoid arthritis from MRI. In *Medical Imaging with Deep Learning*, 2022. URL https://openreview.net/forum?id=8fk23e6ftYP.

[19] Laure Gossec, Joanna Kedra, Hervé Servy, Aridaman Pandit, Simon Stones, Francis Berenbaum, Axel Finckh, Xenofon Baraliakos, Tanja A Stamm, David Gomez-Cabrero, Christian Pristipino, Remy Choquet, Gerd R Burmester, and Timothy R D J Radstake. Eular points to consider for the use of big data in rheumatic and musculoskeletal diseases. *Annals of the Rheumatic Diseases*, 79(1):69–76, 2020. ISSN 0003-4967. doi: 10.1136/annrheumdis-2019-215694. URL https://ard.bmj.com/content/79/1/69.

[20] Jeff Boissoneault, Landrew Sevel, Janelle Letzen, Michael Robinson, and Roland Staud. Biomarkers for musculoskeletal pain conditions: Use of brain imaging and machine learning. *Current Rheumatology Reports*, 19:5, 1 2017. ISSN 1523-3774. doi: 10.1007/s11926-017-0629-9.

[21] Joanna Kedra, Timothy Radstake, Aridaman Pandit, Xenofon Baraliakos, Francis Berenbaum, Axel Finckh, Bruno Fautrel, Tanja A Stamm, David Gomez-Cabrero, Christian Pristipino, Remy Choquet, Hervé Servy, Simon Stones, Gerd Burmester, and Laure Gossec. Current status of use of big data and artificial intelligence in rmds: a systematic literature review informing eular recommendations. *RMD Open*, 5(2), 2019. doi: 10.1136/rmdopen-2019-001004. URL https://rmdopen.bmj.com/content/5/2/e001004.

[22] Aridaman Pandit and Timothy R. D. J. Radstake. Machine learning in rheumatology approaches the clinic. *Nature Reviews Rheumatology*, 16:69–70, 2 2020. ISSN 1759-4790. doi: 10.1038/s41584-019-0361-0.

[23] Tagkopoulos Ilias Kim Ki-Jo. Application of machine learning in rheumatic disease research. *Korean J Intern Med*, 34(4):708–722, 2019. doi: 10.3904/kjim.2018.349. URL http://www.kjim.org/journal/view.php?number=170155.

[24] Afshin Jamshidi, Jean-Pierre Pelletier, and Johanne Martel-Pelletier. Machine-learning-based patient-specific prediction models for knee osteoarthritis. *Nature Reviews Rheumatology*, 15(1):49–60, 1 2019. ISSN 1759-4790. doi: 10.1038/s41584-018-0130-5.

[25] Narendra N. Khanna, Ankush D. Jamthikar, Deep Gupta, Matteo Piga, Luca Saba, Carlo Carcassi, Argiris A. Giannopoulos, Andrew Nicolaides, John R. Laird, Harman S. Suri, Sophie Mavrogeni, A.D. Protogerou, Petros Sfikakis, George D. Kitas, and Jasjit S. Suri. Rheumatoid arthritis: Atherosclerosis imaging and cardiovascular risk assessment using machine and deep learning–based tissue characterization. *Current Atherosclerosis Reports*, 21:7, 2 2019. ISSN 1523-3804. doi: 10.1007/s11883-019-0766-x.

[26] Berend C. Stoel. Artificial intelligence in detecting early ra. *Seminars in Arthritis and Rheumatism*, 49(3, Supplement):S25–S28, 2019. ISSN 0049-0172. doi: https://doi.org/10.1016/j.semarthrit.2019.09.020. URL https://www.sciencedirect.com/science/article/pii/S0049017219306559. Advances in Targeted Therapies: Proceedings of the 2019 Meeting.

[27] Uran Ferizi, Stephen Honig, and Gregory Chang. Artificial intelligence, osteoporosis and fragility fractures. *Current Opinion in Rheumatology*, 31:368–375, 7 2019. ISSN 1040-8711. doi: 10.1097/BOR.0000000000000607.

[28] Mengdi Jiang, Yueting Li, Chendan Jiang, Lidan Zhao, Xuan Zhang, and Peter E Lipsky. Machine Learning in Rheumatic Diseases. *Clinical Reviews in Allergy & Immunology*, 60(1):96–110, 2021. ISSN 1559-0267. doi: 10.1007/s12016-020-08805-6. URL https://doi.org/10.1007/s12016-020-08805-6.

[29] Maria Hügle, Patrick Omoumi, Jacob M van Laar, Joschka Boedecker, and Thomas Hügle. Applied machine learning and artificial intelligence in rheumatology. *Rheumatology Advances in Practice*, 4(1), 02 2020. ISSN 2514-1775. doi: 10.1093/rap/rkaa005. URL https://doi.org/10.1093/rap/rkaa005. rkaa005.

[30] Amaranta Manrique de Lara and Ingris Peláez-Ballestas. Big data and data processing in rheumatology: bioethical perspectives. *Clinical Rheumatology*, 39:1007–1014, 2020. ISSN 1434-9949. doi: 10.1007/s10067-020-04969-w. URL https://doi.org/10.1007/s10067-020-04969-w.

[31] Berend Stoel. Use of artificial intelligence in imaging in rheumatology – current status and future perspectives. *RMD Open*, 6(1), 2020. doi: 10.1136/rmdopen-2019-001063. URL https://rmdopen.bmj.com/content/6/1/e001063.

[32] Kathryn M Kingsmore, Christopher E Puglisi, Amrie C Grammer, and Peter E Lipsky. An introduction to machine learning and analysis of its use in rheumatic diseases. *Nature Reviews Rheumatology*, 2021. ISSN 1759-4804. doi: 10.1038/s41584-021-00708-w. URL https://doi.org/10.1038/s41584-021-00708-w.

[33] David Soriano-Valdez, Ingris Pelaez-Ballestas, Amaranta Manrique de Lara, and Alfonso Gastelum-Strozzi. The basics of data, big data, and machine learning in clinical practice. *Clinical Rheumatology*, 40:11–23, 1 2021. ISSN 0770-3198. doi: 10.1007/s10067-020-05196-z.

[34] Joanna Kedra, Thomas Davergne, Ben Braithwaite, Hervé Servy, and Laure Gossec. Machine learning approaches to improve disease management of patients with rheumatoid arthritis: review and future directions. *Expert Review of Clinical Immunology*, 0(ja):null, 2021. doi: 10.1080/1744666X.2022.2017773. URL https://doi.org/10.1080/1744666X.2022.2017773. PMID: 34890271.

[35] Julien Smets, Enisa Shevroja, Thomas Hügle, William D Leslie, and Didier Hans. Machine learning solutions for osteoporosis—a review. *Journal of Bone and Mineral Research*, 36:833–851, 5 2021. ISSN 0884-0431. doi: 10.1002/jbmr.4292.

[36] Thomas Davergne, Joanna Kedra, and Laure Gossec. Wearable activity trackers and artificial intelligence in the management of rheumatic diseases. *Zeitschrift für Rheumatologie*, 80:928–935, 12 2021. ISSN 0340-1855. doi: 10.1007/s00393-021-01100-5.

[37] Maxwell A. Konnaris, Matthew Brendel, Mark Alan Fontana, Miguel Otero, Lionel B. Ivashkiv, Fei Wang, and Richard D. Bell. Computational pathology for musculoskeletal conditions using machine learning: advances, trends, and challenges. *Arthritis Research & Therapy*, 24:68, 12 2022. ISSN 1478-6362. doi: 10.1186/s13075-021-02716-3.

[38] Marie Binvignat, Valentina Pedoia, Atul J Butte, Karine Louati, David Klatzmann, Francis Berenbaum, Encarnita Mariotti-Ferrandiz, and Jérémie Sellam. Use of machine learning in osteoarthritis research: a systematic literature review. *RMD Open*, 8(1), 2022. doi: 10.1136/rmdopen-2021-001998. URL https://rmdopen.bmj.com/content/8/1/e001998.

[39] Francesco Calivà, Nikan K. Namiri, Maureen Dubreuil, Valentina Pedoia, Eugene Ozhinsky, and Sharmila Majumdar. Studying osteoarthritis with artificial intelligence applied to magnetic resonance imaging. *Nature Reviews Rheumatology*, 18:112–121, 2 2022. ISSN 1759-4790. doi: 10.1038/s41584-021-00719-7.

[40] Yuan Li and Linru Zhao. Application of machine learning in rheumatic immune diseases. *Journal of Healthcare Engineering*, 2022:1–9, 1 2022. ISSN 2040-2309. doi: 10.1155/2022/9273641.

[41] Diederik De Cock, Elena Myasoedova, Daniel Aletaha, and Paul Studenic. Big data analyses and individual health profiling in the arena of rheumatic and musculoskeletal diseases (rmds). *Therapeutic Advances in Musculoskeletal Disease*, 14:1759720X2211059, 1 2022. ISSN 1759-720X. doi: 10.1177/1759720X221105978. URL http://journals.sagepub.com/doi/10.1177/1759720X221105978.

[42] Amanda E. Nelson and Liubov Arbeeva. Narrative review of machine learning in rheumatic and musculoskeletal diseases for clinicians and researchers: biases, goals, and future directions. *The Journal of Rheumatology*, 2022. ISSN 0315-162X. doi: 10.3899/jrheum.220326. URL https://www.jrheum.org/content/early/2022/07/14/jrheum.220326.

[43] Francesco Bonomi, Silvia Peretti, Gemma Lepri, Vincenzo Venerito, Edda Russo, Cosimo Bruni, Florenzo Iannone, Sabina Tangaro, Amedeo Amedei, Serena Guiducci, Marco Matucci Cerinic, and Silvia Bellando Randone. The use and utility of machine learning in achieving precision medicine in systemic sclerosis: A narrative review. *Journal of Personalized Medicine*, 12(8), 2022. ISSN 2075-4426. doi: 10.3390/jpm12081198. URL https://www.mdpi.com/2075-4426/12/8/1198.

[44] Christopher McMaster, Alix Bird, David FL Liew, Russell R Buchanan, Claire E Owen, Wendy W Chapman, and Douglas EV Pires. Artificial intelligence and deep learning for rheumatologists: A primer and review of the literature. *Arthritis & Rheumatology*, 2022. doi: https://doi.org/10.1002/art.42296. URL https://onlinelibrary.wiley.com/doi/abs/10.1002/art.42296.

[45] Leonie E Burgers, Karim Raza, and Annette H van der Helm van Mil. Window of opportunity in rheumatoid arthritis – definitions and supporting evidence: from old to new perspectives. *RMD Open*, 5(1), 2019. doi: 10.1136/rmdopen-2018-000870. URL https://rmdopen.bmj.com/content/5/1/e000870.

[46] Sociedad Española de Reumatología. Activity indices, questionnaires and other measurement instruments in Rheumatology, 2021. URL https://www.ser.es/profesionales/que-hacemos/investigacion/herramientas/catalina/.

[47] Mark Reed, Timothy Le Souëf, and Elliot Rampono. A pilot study of a machine-learning tool to assist in the diagnosis of hand arthritis. *Internal Medicine Journal*, 52(6):959–967, 2022. doi: https://doi.org/10.1111/imj.15173. URL https://onlinelibrary.wiley.com/doi/abs/10.1111/imj.15173.

[48] Suhanyaa Nitkunanantharajah, Katja Haedicke, Tonia B. Moore, Joanne B. Manning, Graham Dinsdale, Michael Berks, Christopher Taylor, Mark R. Dickinson, Dominik Jüstel, Vasilis Ntziachristos, Ariane L. Herrick, and Andrea K. Murray. Three-dimensional optoacoustic imaging of nailfold capillaries in systemic sclerosis and its potential for disease differentiation using deep learning. *Scientific Reports*, 10(1): 16444, 2020. ISSN 2045-2322. doi: 10.1038/s41598-020-73319-2. URL https://doi.org/10.1038/s41598-020-73319-2.

[49] Vincenzo Venerito, Orazio Angelini, Gerardo Cazzato, Giuseppe Lopalco, Eugenio Maiorano, Antonietta Cimmino, and Florenzo Iannone. A convolutional neural network with transfer learning for automatic discrimination between low and high-grade synovitis: a pilot study. *Internal and Emergency Medicine*, 16:1457–1465, 2021. ISSN 1970-9366. doi: 10.1007/s11739-020-02583-x. URL https://doi.org/10.1007/s11739-020-02583-x.

[50] Shawli Bardhan and Mrinal Kanti Bhowmik. 2-Stage classification of knee joint thermograms for rheumatoid arthritis prediction in subclinical inflammation. *Australasian Physical & Engineering Sciences in Medicine*, 42(1):259–277, 2019. ISSN 1879-5447. doi: 10.1007/s13246-019-00726-9. URL https://doi.org/10.1007/s13246-019-00726-9.

[51] Tengyang Wang, Guanghua Liu, and Hongye Lin. A machine learning approach to predict intravenous immunoglobulin resistance in kawasaki disease patients: A study based on a southeast china population. *PLOS ONE*, 15(8):1–15, 08 2020. doi: 10.1371/journal.pone.0237321. URL https://doi.org/10.1371/journal.pone.0237321.

[52] Atul Deodhar, Martin Rozycki, Cody Garges, Oodaye Shukla, Theresa Arndt, Tara Grabowsky, and Yujin Park. Use of machine learning techniques in the development and refinement of a predictive model for early diagnosis of ankylosing spondylitis. *Clinical Rheumatology*, 39(4):975–982, 2020. ISSN 1434-9949. doi: 10.1007/s10067-019-04553-x. URL https://doi.org/10.1007/s10067-019-04553-x.

[53] Jessica A. Walsh, Shaobo Pei, Gopi K. Penmetsa, Rebecca S. Overbury, Daniel O. Clegg, and Brian C. Sauer. Identifying patients with axial spondyloarthritis in large datasets: Expanding possibilities for observational research. *The Journal of Rheumatology*, 2020. ISSN 0315-162X. doi: 10.3899/jrheum.200570. URL https://www.jrheum.org/content/early/2021/01/25/jrheum.200570.

[54] Sizheng Steven Zhao, Chuan Hong, Tianrun Cai, Chang Xu, Jie Huang, Joerg Ermann, Nicola J Goodson, Daniel H Solomon, Tianxi Cai, and Katherine P Liao. Incorporating natural language processing to improve classification of axial spondyloarthritis using electronic health records. *Rheumatology*, 59(5):1059–1065, 09 2019. ISSN 1462-0324. doi: 10.1093/rheumatology/kez375. URL https://doi.org/10.1093/rheumatology/kez375.

[55] Lia Jamian, Lee Wheless, Leslie J Crofford, and April Barnado. Rule-based and machine learning algorithms identify patients with systemic sclerosis accurately in the electronic health record. *Arthritis Research & Therapy*, 21(1):305, 2019. ISSN 1478-6362. doi: 10.1186/s13075-019-2092-7. URL https://doi.org/10.1186/s13075-019-2092-7.

[56] Christina Adamichou, Irini Genitsaridi, Dionysis Nikolopoulos, Myrto Nikoloudaki, Argyro Repa, Alessandra Bortoluzzi, Antonis Fanouriakis, Prodromos Sidiropoulos, Dimitrios T Boumpas, and George K Bertsias. Lupus or not? sle risk probability index (slerpi): a simple,

clinician-friendly machine learning-based model to assist the diagnosis of systemic lupus erythematosus. *Annals of the Rheumatic Diseases*, 80(6):758–766, 2021. ISSN 0003-4967. doi: 10.1136/annrheumdis-2020-219069. URL https://ard.bmj.com/content/80/6/758.

[57] Sicong Huang, Jie Huang, Tianrun Cai, Kumar P Dahal, Andrew Cagan, Zeling He, Jacklyn Stratton, Isaac Gorelik, Chuan Hong, Tianxi Cai, and Katherine P Liao. Impact of ICD10 and secular changes on electronic medical record rheumatoid arthritis algorithms. *Rheumatology*, 59 (12):3759–3766, 05 2020. ISSN 1462-0324. doi: 10.1093/rheumatology/keaa198. URL https://doi.org/10.1093/rheumatology/keaa198.

[58] Tjardo D Maarseveen, Timo Meinderink, Marcel J T Reinders, Johannes Knitza, Tom W J Huizinga, Arnd Kleyer, David Simon, Erik B van den Akker, and Rachel Knevel. Machine learning electronic health record identification of patients with rheumatoid arthritis: Algorithm pipeline development and validation study. *JMIR Medical Informatics*, 8:e23930, 11 2020. ISSN 2291-9694. doi: 10.2196/23930.

[59] Jihye Lim, Jungyoon Kim, and Songhee Cheon. A deep neural network-based method for early detection of osteoarthritis using statistical data. *International Journal of Environmental Research and Public Health*, 16(7), 2019. ISSN 1660-4601. doi: 10.3390/ijerph16071281. URL https://www.mdpi.com/1660-4601/16/7/1281.

[60] Sara K. Tedeschi, Tianrun Cai, Zeling He, Yuri Ahuja, Chuan Hong, Katherine A. Yates, Kumar Dahal, Chang Xu, Houchen Lyu, Kazuki Yoshida, Daniel H. Solomon, Tianxi Cai, and Katherine P. Liao. Classifying pseudogout using machine learning approaches with electronic health record data. *Arthritis Care & Research*, 73(3):442–448, 2021. doi: https://doi.org/10.1002/acr.24132. URL https://onlinelibrary.wiley.com/doi/abs/10.1002/acr.24132.

[61] Megan E. Breitbach, Ryne C. Ramaker, Kevin Roberts, Robert P. Kimberly, and Devin Absher. Population-specific patterns of epigenetic defects in the b cell lineage in patients with systemic lupus erythematosus. *Arthritis & Rheumatology*, 72(2):282–291, 2020. doi: https://doi.org/10.1002/art.41083. URL https://onlinelibrary.wiley.com/doi/abs/10.1002/art.41083.

[62] Erika Van Nieuwenhove, Vasiliki Lagou, Lien Van Eyck, James Dooley, Ulrich Bodenhofer, Carlos Roca, Marijne Vandebergh, An Goris, Stéphanie Humblet-Baron, Carine Wouters, and Adrian Liston. Machine learning identifies an immunological pattern associated with multiple juvenile idiopathic arthritis subtypes. *Annals of the Rheumatic Diseases*, 78(5):617–628, 2019. ISSN 0003-4967. doi: 10.1136/annrheumdis-2018-214354. URL https://ard.bmj.com/content/78/5/617.

[63] Rodrigo Cánovas, Joanna Cobb, Marta Brozynska, John Bowes, Yun R Li, Samantha Louise Smith, Hakon Hakonarson, Wendy Thomson, Justine A Ellis, Gad Abraham, Jane E Munro, and Michael Inouye. Genomic risk scores for juvenile idiopathic arthritis and its subtypes. *Annals of the Rheumatic Diseases*, 79(12):1572–1579, 2020. ISSN 0003-4967. doi: 10.1136/annrheumdis-2020-217421. URL https://ard.bmj.com/content/79/12/1572.

[64] Christophe Roncato, Lior Perez, Antoine Brochet-Guégan, Caroline Allix-Béguec, Alizée Raimbeau, Giovanni Gautier, Christian Agard, Gaëtan Ploton, Simon Moisselin, Fanny Lorcerie, Guillaume Denis, Bruno Gombert, Elisabeth Gervais, and Olivier Espitia. Colour Doppler ultrasound of temporal arteries for the diagnosis of giant cell arteritis: a multicentre deep learning study. *Clinical and experimental rheumatology*, 38 Suppl 1(2):120–125, 2020. ISSN 0392-856X. URL http://www.ncbi.nlm.nih.gov/pubmed/32441644.

[65] Riel Castro-Zunti, Eun Hae Park, Younhee Choi, Gong Yong Jin, and Seok bum Ko. Early detection of ankylosing spondylitis using texture features and statistical machine learning, and deep learning, with some patient age analysis. *Computerized Medical Imaging and Graphics*, 82:101718, 2020. ISSN 0895-6111. doi: https://doi.org/10.1016/j.compmedimag.2020.101718. URL https://www.sciencedirect.com/science/article/pii/S0895611120300215.

[66] Norio Yamamoto, Shintaro Sukegawa, Akira Kitamura, Ryosuke Goto, Tomoyuki Noda, Keisuke Nakano, Kiyofumi Takabatake, Hotaka Kawai, Hitoshi Nagatsuka, Keisuke Kawasaki, Yoshihiko Furuki, and Toshifumi Ozaki. Deep learning for osteoporosis classification using hip radiographs and patient clinical covariates. *Biomolecules*, 10(11), 2020. ISSN 2218-273X. doi: 10.3390/biom10111534. URL https://www.mdpi.com/2218-273X/10/11/1534.

[67] Joseph E. Burns, Jianhua Yao, Didier Chalhoub, Joseph J. Chen, and Ronald M. Summers. A machine learning algorithm to estimate sarcopenia on abdominal ct. *Academic Radiology*, 27(3):311–320, 2020. ISSN 1076-6332. doi: https://doi.org/10.1016/j.acra.2019.03.011. URL https://www.sciencedirect.com/science/article/pii/S1076633219301655.

[68] Shan-Chi Yu, Kung-Chao Chang, Hsuan Wang, Meng-Fang Li, Tsung-Lin Yang, Chun-Nan Chen, Chih-Jung Chen, Ko-Chin Chen, Chieh-Yu Shen, Po-Yen Kuo, Long-Wei Lin, Yueh-Min Lin, and Wei-Chou Lin. Distinguishing lupus lymphadenitis from Kikuchi disease based on clinicopathological features and C4d immunohistochemistry. *Rheumatology*, 60(3):1543–1552, 11 2020. ISSN 1462-0324. doi: 10.1093/rheumatology/keaa524. URL https://doi.org/10.1093/rheumatology/keaa524.

[69] Daniel J. Kass, Mehdi Nouraie, Marilyn K. Glassberg, Nitya Ramreddy, Karen Fernandez, Lisa Harlow, Yingze Zhang, Jean Chen, Gail S. Kerr, Andreas M. Reimold, Bryant R. England, Ted R. Mikuls, Kevin F. Gibson, Paul F. Dellaripa, Ivan O. Rosas, Chester V. Oddis, and Dana P. Ascherman. Comparative profiling of serum protein biomarkers in rheumatoid arthritis–associated interstitial lung disease and idiopathic pulmonary fibrosis. *Arthritis & Rheumatology*, 72(3):409–419, 2020. doi: https://doi.org/10.1002/art.41123. URL https://onlinelibrary.wiley.com/doi/abs/10.1002/art.41123.

[70] Michelle J. Ormseth, Joseph F. Solus, Quanhu Sheng, Fei Ye, Qiong Wu, Yan Guo, Annette M. Oeser, Ryan M. Allen, Kasey C. Vickers, and C. Michael Stein. Development and validation of a microrna panel to differentiate between patients with rheumatoid arthritis or systemic lupus erythematosus and controls. *The Journal of Rheumatology*, 2019. ISSN 0315-162X. doi: 10.3899/jrheum.181029. URL https://www.jrheum.org/content/early/2019/05/13/jrheum.181029.

[71] Xiao Liu, Wei Zhang, Ming Zhao, Longfei Fu, Limin Liu, Jinghua Wu, Shuangyan Luo, Longlong Wang, Zijun Wang, Liya Lin, Yan Liu, Shiyu Wang, Yang Yang, Lihua Luo, Juqing Jiang, Xie Wang, Yixin Tan, Tao Li, Bochen Zhu, Yi Zhao, Xiaofei Gao, Ziyun Wan, Cancan Huang, Mingyan Fang, Qianwen Li, Huanhuan Peng, Xiangping Liao, Jinwei Chen, Fen Li, Guanghui Ling, Hongjun Zhao, Hui Luo, Zhongyuan Xiang, Jieyue Liao, Yu Liu, Heng Yin, Hai Long, Haijing Wu, huanming Yang, Jian Wang, and Qianjin Lu. T cell receptor β repertoires as novel diagnostic markers for systemic lupus erythematosus and rheumatoid arthritis. *Annals of the Rheumatic Diseases*, 78 (8):1070–1078, 2019. ISSN 0003-4967. doi: 10.1136/annrheumdis-2019-215442. URL https://ard.bmj.com/content/78/8/1070.

[72] Margarida Souto-Carneiro, Lilla Tóth, Rouven Behnisch, Konstantin Urbach, Karel D Klika, Rui A Carvalho, and Hanns-Martin Lorenz. Differences in the serum metabolome and lipidome identify potential biomarkers for seronegative rheumatoid arthritis versus psoriatic

arthritis. *Annals of the Rheumatic Diseases*, 79(4):499–506, 2020. ISSN 0003-4967. doi: 10.1136/annrheumdis-2019-216374. URL https://ard.bmj.com/content/79/4/499.

[73] Juliana Imgenberg-Kreuz, Jonas Carlsson Almlöf, Dag Leonard, Christopher Sjöwall, Ann-Christine Syvänen, Lars Rönnblom, Johanna K. Sandling, and Gunnel Nordmark. Shared and unique patterns of dna methylation in systemic lupus erythematosus and primary sjögren's syndrome. *Frontiers in Immunology*, 10, 2019. ISSN 1664-3224. doi: 10.3389/fimmu.2019.01686. URL https://www.frontiersin.org/articles/10.3389/fimmu.2019.01686.

[74] Yan Zhao, Bin Chen, Shufeng Li, Lanxiu Yang, Dequan Zhu, Ye Wang, Haiying Wang, Tao Wang, Bin Shi, Zhongtao Gai, Jun Yang, Xueyuan Heng, Junjie Yang, and Lei Zhang. Detection and characterization of bacterial nucleic acids in culture-negative synovial tissue and fluid samples from rheumatoid arthritis or osteoarthritis patients. *Scientific Reports*, 8:14305, 12 2018. ISSN 2045-2322. doi: 10.1038/s41598-018-32675-w.

[75] Man Hung, Jerry Bounsanga, Fangzhou Liu, and Maren W. Voss. Profiling arthritis pain with a decision tree. *Pain Practice*, 18(5):568–579, 2018. doi: https://doi.org/10.1111/papr.12645. URL https://onlinelibrary.wiley.com/doi/abs/10.1111/papr.12645.

[76] Kanon Jatuworapruk, Rebecca Grainger, Nicola Dalbeth, and William J. Taylor. Development of a prediction model for inpatient gout flares in people with comorbid gout. *Annals of the Rheumatic Diseases*, 79(3):418–423, 2020. ISSN 0003-4967. doi: 10.1136/annrheumdis-2019-216277. URL https://ard.bmj.com/content/79/3/418.

[77] Mike Becker, Nicole Graf, Rafael Sauter, Yannick Allanore, John Curram, Christopher P Denton, Dinesh Khanna, Marco Matucci-Cerinic, Janethe de Oliveira Pena, Janet E Pope, and Oliver Distler. Predictors of disease worsening defined by progression of organ damage in diffuse systemic sclerosis: a european scleroderma trials and research (eustar) analysis. *Annals of the Rheumatic Diseases*, 78(9):1242–1248, 2019. ISSN 0003-4967. doi: 10.1136/annrheumdis-2019-215145. URL https://ard.bmj.com/content/78/9/1242.

[78] Kamala Vanarsa, Sanam Soomro, Ting Zhang, Briony Strachan, Claudia Pedroza, Malavika Nidhi, Pietro Cicalese, Christopher Gidley, Shobha Dasari, Shree Mohan, Nathan Thai, Van Thi Thanh Truong, Nicole Jordan, Ramesh Saxena, Chaim Putterman, Michelle Petri, and Chandra Mohan. Quantitative planar array screen of 1000 proteins uncovers novel urinary protein biomarkers of lupus nephritis. *Annals of the Rheumatic Diseases*, 79(10):1349–1361, 2020. ISSN 0003-4967. doi: 10.1136/annrheumdis-2019-216312. URL https://ard.bmj.com/content/79/10/1349.

[79] Parisa Riahi, Anoshirvan Kazemnejad, Shayan Mostafaei, Akira Meguro, Nobuhisa Mizuki, Amir Ashraf-Ganjouei, Ali Javinani, Seyedeh Tahereh Faezi, Farhad Shahram, and Mahdi Mahmoudi. Erap1 polymorphisms interactions and their association with behçet's disease susceptibly: Application of model-based multifactor dimension reduction algorithm (mb-mdr). *PLOS ONE*, 15(2):1–10, 02 2020. doi: 10.1371/journal.pone.0227997. URL https://doi.org/10.1371/journal.pone.0227997.

[80] M Paula Gomez Hernandez, Emily E Starman, Andrew B Davis, Miyuraj Harishchandra Hikkaduwa Withanage, Erliang Zeng, Scott M Lieberman, Kim A Brogden, and Emily A Lanzel. A distinguishing profile of chemokines, cytokines and biomarkers in the saliva of children with Sjögren's syndrome. *Rheumatology*, 01 2021. ISSN 1462-0324. doi: 10.1093/rheumatology/keab098. URL https://doi.org/10.1093/rheumatology/keab098. keab098.

[81] Haytham Eloqayli, Ali Al-Yousef, and Raid Jaradat. Vitamin d and ferritin correlation with chronic neck pain using standard statistics and a novel artificial neural network prediction model. *British Journal of Neurosurgery*, 32(2):172–176, 2018. doi: 10.1080/02688697.2018.1436691. URL https://doi.org/10.1080/02688697.2018.1436691. PMID: 29447493.

[82] Masaki Asakage, Yoshihiko Usui, Naoya Nezu, Hiroyuki Shimizu, Kinya Tsubota, Naoyuki Yamakawa, Masakatsu Takanashi, Masahiko Kuroda, and Hiroshi Goto. Comprehensive miRNA Analysis Using Serum From Patients With Noninfectious Uveitis. *Investigative Ophthalmology & Visual Science*, 61(11):4–4, 09 2020. ISSN 1552-5783. doi: 10.1167/iovs.61.11.4. URL https://doi.org/10.1167/iovs.61.11.4.

[83] Laura Andrés-Rodríguez, Xavier Borràs, Albert Feliu-Soler, Adrián Pérez-Aranda, Antoni Rozadilla-Sacanell, Belén Arranz, Jesús Montero-Marin, Javier García-Campayo, Natalia Angarita-Osorio, Michael Maes, and Juan V. Luciano. Machine learning to understand the immune-inflammatory pathways in fibromyalgia. *International Journal of Molecular Sciences*, 20(17), 2019. ISSN 1422-0067. doi: 10.3390/ijms20174231. URL https://www.mdpi.com/1422-0067/20/17/4231.

[84] Lionel Spielmann, Benoit Nespola, François Séverac, Emmanuel Andres, Romain Kessler, Aurélien Guffroy, Vincent Poindron, Thierry Martin, Bernard Geny, Jean Sibilia, and Alain Meyer. Anti-ku syndrome with elevated ck and anti-ku syndrome with anti-dsdna are two distinct entities with different outcomes. *Annals of the Rheumatic Diseases*, 78(8):1101–1106, 2019. ISSN 0003-4967. doi: 10.1136/annrheumdis-2018-214439. URL https://ard.bmj.com/content/78/8/1101.

[85] Alain Meyer, Lionel Spielmann, and François Séverac. On how to not misuse hierarchical clustering on principal components to define clinically meaningful patient subgroups. response to: 'on using machine learning algorithms to define clinical meaningful patient subgroups' by pinal-fernandez and mammen. *Annals of the Rheumatic Diseases*, 79(10):e129–e129, 2020. ISSN 0003-4967. doi: 10.1136/annrheumdis-2019-215868. URL https://ard.bmj.com/content/79/10/e129.

[86] Iago Pinal-Fernandez and Andrew Lee Mammen. On using machine learning algorithms to define clinically meaningful patient subgroups. *Annals of the Rheumatic Diseases*, 79(10):e128–e128, 2020. ISSN 0003-4967. doi: 10.1136/annrheumdis-2019-215852. URL https://ard.bmj.com/content/79/10/e128.

[87] Alain Meyer, Lionel Spielmann, and François Séverac. Response to 'augmented vs. artificial intelligence for stratification of patients with myositis' by mahler et al. *Annals of the Rheumatic Diseases*, 79(12):e163–e163, 2020. ISSN 0003-4967. doi: 10.1136/annrheumdis-2019-216014. URL https://ard.bmj.com/content/79/12/e163.

[88] Michael Mahler, Brenden Rossin, and Olga Kubassova. Augmented versus artificial intelligence for stratification of patients with myositis. *Annals of the Rheumatic Diseases*, 79(12):e162–e162, 2020. ISSN 0003-4967. doi: 10.1136/annrheumdis-2019-216000. URL https://ard.bmj.com/content/79/12/e162.

[89] Yusuke Ogata, Yuichiro Fujieda, Masanari Sugawara, Taiki Sato, Naoki Ohnishi, Michihito Kono, Masaru Kato, Kenji Oku, Olga Amengual, and Tatsuya Atsumi. Morbidity and mortality in antiphospholipid syndrome based on cluster analysis: a 10-year longitudinal cohort study. *Rheumatology*, 60(3):1331–1337, 09 2020. ISSN 1462-0324. doi: 10.1093/rheumatology/keaa542. URL https://doi.org/10.1093/

rheumatology/keaa542.

[90] Min Jung Kim, Eun Young Ahn, Woochang Hwang, Youngjo Lee, Eun Young Lee, Eun Bong Lee, Yeong Wook Song, and Jin Kyun Park. Association between fever pattern and clinical manifestations of adult-onset Still's disease: unbiased analysis using hierarchical clustering. *Clinical and experimental rheumatology*, 36(6 Suppl 115):74–79, 2018. ISSN 0392-856X. doi: 30582502. URL http://www.ncbi.nlm.nih.gov/pubmed/30582502.

[91] Vasileios C. Pezoulas, Themis P. Exarchos, Athanasios G. Tzioufas, Salvatore De Vita, and Dimitrios I. Fotiadis. Predicting lymphoma outcomes and risk factors in patients with primary sjögren's syndrome using gradient boosting tree ensembles. In *2019 41st Annual International Conference of the IEEE Engineering in Medicine and Biology Society (EMBC)*, pages 2165–2168, 2019. doi: 10.1109/EMBC.2019.8857557.

[92] C. Baldini, F. Ferro, N. Luciano, S. Bombardieri, and E. Grossi. Artificial neural networks help to identify disease subsets and to predict lymphoma in primary Sjögren's syndrome. *Clinical and Experimental Rheumatology*, 36:S137–S144, 2018. ISSN 1593098X.

[93] Elena Bartoloni, Chiara Baldini, Francesco Ferro, Alessia Alunno, Francesco Carubbi, Giacomo Cafaro, Stefano Bombardieri, Roberto Gerli, and Enzo Grossi. Application of artificial neural network analysis in the evaluation of cardiovascular risk in primary Sjögren's syndrome: A novel pathogenetic scenario? *Clinical and Experimental Rheumatology*, 37(3):S133–S139, 2019. ISSN 1593098X.

[94] Alessandra Tesser, Luciana Martins de Carvalho, Paula Sandrin-Garcia, Alessia Pin, Serena Pastore, Andrea Taddio, Luciana Rodrigues Roberti, Rosane Gomes de Paula Queiroz, Virginia Paes Leme Ferriani, Sergio Crovella, and Alberto Tommasini. Higher interferon score and normal complement levels may identify a distinct clinical subset in children with systemic lupus erythematosus. *Arthritis Research & Therapy*, 22(1):91, 2020. ISSN 1478-6362. doi: 10.1186/s13075-020-02161-8. URL https://doi.org/10.1186/s13075-020-02161-8.

[95] Su-Jin Moon, Jung Min Bae, Kyung-Su Park, Ilias Tagkopoulos, and Ki-Jo Kim. Compendium of skin molecular signatures identifies key pathological features associated with fibrosis in systemic sclerosis. *Annals of the Rheumatic Diseases*, 78(6):817–825, 2019. ISSN 0003-4967. doi: 10.1136/annrheumdis-2018-214778. URL https://ard.bmj.com/content/78/6/817.

[96] Simon W. M. Eng, Florence A. Aeschlimann, Mira van Veenendaal, Roberta A. Berard, Alan M. Rosenberg, Quaid Morris, Rae S. M. Yeung, and on behalf of the ReACCh-Out Research Consortium. Patterns of joint involvement in juvenile idiopathic arthritis and prediction of disease course: A prospective study with multilayer non-negative matrix factorization. *PLOS Medicine*, 16(2):1–22, 02 2019. doi: 10.1371/journal.pmed.1002750. URL https://doi.org/10.1371/journal.pmed.1002750.

[97] Elham Rezaei, Daniel Hogan, Brett Trost, Anthony J Kusalik, Gilles Boire, David A Cabral, Sarah Campillo, Gaëlle Chédeville, Anne-Laure Chetaille, Paul Dancey, Ciaran Duffy, Karen Watanabe Duffy, Simon W M Eng, John Gordon, Jaime Guzman, Kristin Houghton, Adam M Huber, Roman Jurencak, Bianca Lang, Ronald M Laxer, Kimberly Morishita, Kiem G Oen, Ross E Petty, Suzanne E Ramsey, Stephen W Scherer, Rosie Scuccimarri, Lynn Spiegel, Elizabeth Stringer, Regina M Taylor-Gjevre, Shirley M L Tse, Lori B Tucker, Stuart E Turvey, Susan Tupper, Richard F Wintle, Rae S M Yeung, Alan M Rosenberg, and for the BBOP Study Group. Associations of clinical and inflammatory biomarker clusters with juvenile idiopathic arthritis categories. *Rheumatology*, 59(5):1066–1075, 09 2019. ISSN 1462-0324. doi: 10.1093/rheumatology/kez382. URL https://doi.org/10.1093/rheumatology/kez382.

[98] Dana E. Orange, Phaedra Agius, Edward F. DiCarlo, Nicolas Robine, Heather Geiger, Jackie Szymonifka, Michael McNamara, Ryan Cummings, Kathleen M. Andersen, Serene Mirza, Mark Figgie, Lionel B. Ivashkiv, Alessandra B. Pernis, Caroline S. Jiang, Mayu O. Frank, Robert B. Darnell, Nithya Lingampali, William H. Robinson, Ellen Gravallese, the Accelerating Medicines Partnership in Rheumatoid Arthritis and Lupus Network , Vivian P. Bykerk, Susan M. Goodman, and Laura T. Donlin. Identification of three rheumatoid arthritis disease subtypes by machine learning integration of synovial histologic features and rna sequencing data. *Arthritis & Rheumatology*, 70(5): 690–701, 2018. doi: https://doi.org/10.1002/art.40428. URL https://onlinelibrary.wiley.com/doi/abs/10.1002/art.40428.

[99] K. Bates Gribbons, Cristina Ponte, Simon Carette, Anthea Craven, David Cuthbertson, Gary S. Hoffman, Nader A. Khalidi, Curry L. Koening, Carol A. Langford, Kathleen Maksimowicz-McKinnon, Carol A. McAlear, Paul A. Monach, Larry W. Moreland, Christian Pagnoux, Kaitlin A. Quinn, Joanna C. Robson, Philip Seo, Antoine G. Sreih, Ravi Suppiah, Kenneth J. Warrington, Steven R. Ytterberg, Raashid Luqmani, Richard Watts, Peter A. Merkel, and Peter C. Grayson. Patterns of arterial disease in takayasu arteritis and giant cell arteritis. *Arthritis Care & Research*, 72(11):1615–1624, 2020. doi: https://doi.org/10.1002/acr.24055. URL https://onlinelibrary.wiley.com/doi/abs/10.1002/acr.24055.

[100] Ruchika Goel, K Bates Gribbons, Simon Carette, David Cuthbertson, Gary S Hoffman, George Joseph, Nader A Khalidi, Curry L Koening, Sathish Kumar, Carol Langford, Kathleen Maksimowicz-McKinnon, Carol A McAlear, Paul A Monach, Larry W Moreland, Aswin Nair, Christian Pagnoux, Kaitlin A Quinn, Raheesh Ravindran, Philip Seo, Antoine G Sreih, Kenneth J Warrington, Steven R Ytterberg, Peter A Merkel, Debashish Danda, and Peter C Grayson. Derivation of an angiographically based classification system in Takayasu's arteritis: an observational study from India and North America. *Rheumatology*, 59(5):1118–1127, 10 2019. ISSN 1462-0324. doi: 10.1093/rheumatology/kez421. URL https://doi.org/10.1093/rheumatology/kez421.

[101] Philippe Burlina, Neil Joshi, Seth Billings, I-Jeng Wang, and Jemima Albayda. Deep embeddings for novelty detection in myopathy. *Computers in Biology and Medicine*, 105:46–53, 2019. ISSN 0010-4825. doi: https://doi.org/10.1016/j.compbiomed.2018.12.006. URL https://www.sciencedirect.com/science/article/pii/S0010482518304049.

[102] Uran Ferizi, Harrison Besser, Pirro Hysi, Joseph Jacobs, Chamith S. Rajapakse, Cheng Chen, Punam K. Saha, Stephen Honig, and Gregory Chang. Artificial intelligence applied to osteoporosis: A performance comparison of machine learning algorithms in predicting fragility fractures from mri data. *Journal of Magnetic Resonance Imaging*, 49(4):1029–1038, 2019. doi: https://doi.org/10.1002/jmri.26280. URL https://onlinelibrary.wiley.com/doi/abs/10.1002/jmri.26280.

[103] Matheus Calil Faleiros, Marcello Henrique Nogueira-Barbosa, Vitor Faeda Dalto, José Raniery Ferreira Júnior, Ariane Priscilla Magalhães Tenório, Rodrigo Luppino-Assad, Paulo Louzada-Junior, Rangaraj Mandayam Rangayyan, and Paulo Mazzoncini de Azevedo-Marques. Machine learning techniques for computer-aided classification of active inflammatory sacroiliitis in magnetic resonance imaging. *Advances in Rheumatology*, 60:25, 12 2020. ISSN 2523-3106. doi: 10.1186/s42358-020-00126-8.

[104] Alberta Hoi, Hieu T Nim, Rachel Koelmeyer, Ying Sun, Amy Kao, Oliver Gunther, and Eric Morand. Algorithm for calculating high disease activity in SLE. *Rheumatology*, 60(9):4291–4297, 01 2021. ISSN 1462-0324. doi: 10.1093/rheumatology/keab003. URL

https://doi.org/10.1093/rheumatology/keab003.

[105] Maria Camacho-Encina, Vanesa Balboa-Barreiro, Ignacio Rego-Perez, Florencia Picchi, Jennifer VanDuin, Ji Qiu, Manuel Fuentes, Natividad Oreiro, Joshua LaBaer, Cristina Ruiz-Romero, and Francisco J Blanco. Discovery of an autoantibody signature for the early diagnosis of knee osteoarthritis: data from the osteoarthritis initiative. *Annals of the Rheumatic Diseases*, 78(12):1699–1705, 2019. ISSN 0003-4967. doi: 10.1136/annrheumdis-2019-215325. URL https://ard.bmj.com/content/78/12/1699.

[106] Mukundan Attur, Svetlana Krasnokutsky, Hua Zhou, Jonathan Samuels, Gregory Chang, Jenny Bencardino, Pamela Rosenthal, Leon Rybak, Janet L Huebner, Virginia B Kraus, and Steven B Abramson. The combination of an inflammatory peripheral blood gene expression and imaging biomarkers enhance prediction of radiographic progression in knee osteoarthritis. *Arthritis Research & Therapy*, 22(1):208, 2020. ISSN 1478-6362. doi: 10.1186/s13075-020-02298-6. URL https://doi.org/10.1186/s13075-020-02298-6.

[107] Frances Humby, Myles Lewis, Nandhini Ramamoorthi, Jason A Hackney, Michael R Barnes, Michele Bombardieri, A. Francesca Setiadi, Stephen Kelly, Fabiola Bene, Maria DiCicco, Sudeh Riahi, Vidalba Rocher, Nora Ng, Ilias Lazarou, Rebecca Hands, Désirée van der Heijde, Robert B M Landewé, Annette van der Helm-van Mil, Alberto Cauli, Iain McInnes, Christopher Dominic Buckley, Ernest H Choy, Peter C Taylor, Michael J Townsend, and Costantino Pitzalis. Synovial cellular and molecular signatures stratify clinical response to csdmard therapy and predict radiographic progression in early rheumatoid arthritis patients. *Annals of the Rheumatic Diseases*, 78(6):761–772, 2019. ISSN 0003-4967. doi: 10.1136/annrheumdis-2018-214539. URL https://ard.bmj.com/content/78/6/761.

[108] François Chasset, Camillo Ribi, Marten Trendelenburg, Uyen Huynh-Do, Pascale Roux-Lombard, Delphine S Courvoisier, Carlo Chizzolini, and for the Swiss SLE Cohort Study (SSCS. Identification of highly active systemic lupus erythematosus by combined type I interferon and neutrophil gene scores vs classical serologic markers. *Rheumatology*, 59(11):3468–3478, 05 2020. ISSN 1462-0324. doi: 10.1093/rheumatology/keaa167. URL https://doi.org/10.1093/rheumatology/keaa167.

[109] Kerry E. Poppenberg, Kaiyu Jiang, Lu Li, Yijun Sun, Hui Meng, Carol A. Wallace, Teresa Hennon, and James N. Jarvis. The feasibility of developing biomarkers from peripheral blood mononuclear cell RNAseq data in children with juvenile idiopathic arthritis using machine learning approaches. *Arthritis Research & Therapy*, 21(1):230, 12 2019. ISSN 1478-6362. doi: 10.1186/s13075-019-2010-z. URL https://arthritis-research.biomedcentral.com/articles/10.1186/s13075-019-2010-z.

[110] Alicia Rodriguez-Pla, Roscoe L . Warner, David Cuthbertson, Simon Carette, Nader A. Khalidi, Curry L. Koening, Carol A. Langford, Carol A. McAlear, Larry W. Moreland, Christian Pagnoux, Philip Seo, Ulrich Specks, Antoine G. Sreih, Steven R. Ytterberg, Kent J. Johnson, Peter A. Merkel, and Paul A. Monach. Evaluation of potential serum biomarkers of disease activity in diverse forms of vasculitis. *The Journal of Rheumatology*, 2019. ISSN 0315-162X. doi: 10.3899/jrheum.190093. URL https://www.jrheum.org/content/early/2020/01/27/jrheum.190093.

[111] Jakob Kristian Holm Andersen, Jannik Skyttegaard Pedersen, Martin Sundahl Laursen, Kathrine Holtz, Jakob Grauslund, Thiusius Rajeeth Savarimuthu, and Søren Andreas Just. Neural networks for automatic scoring of arthritis disease activity on ultrasound images. *RMD Open*, 5(1), 2019. doi: 10.1136/rmdopen-2018-000891. URL https://rmdopen.bmj.com/content/5/1/e000891.

[112] Anders Bossel Holst Christensen, Søren Andreas Just, Jakob Kristian Holm Andersen, and Thiusius Rajeeth Savarimuthu. Applying cascaded convolutional neural network design further enhances automatic scoring of arthritis disease activity on ultrasound images from rheumatoid arthritis patients. *Annals of the Rheumatic Diseases*, 79(9):1189–1193, 2020. ISSN 0003-4967. doi: 10.1136/annrheumdis-2019-216636. URL https://ard.bmj.com/content/79/9/1189.

[113] Farhad Akhbardeh, Fartash Vasefi, Nick MacKinnon, Mohammad Amini, Alireza Akhbardeh, and Kouhyar Tavakolian. Classification and assessment of hand arthritis stage using support vector machine. In *2019 41st Annual International Conference of the IEEE Engineering in Medicine and Biology Society (EMBC)*, pages 4080–4083, 2019. doi: 10.1109/EMBC.2019.8857022.

[114] Zhaoye Zhou, Gengyan Zhao, Richard Kijowski, and Fang Liu. Deep convolutional neural network for segmentation of knee joint anatomy. *Magnetic Resonance in Medicine*, 80(6):2759–2770, 2018. doi: https://doi.org/10.1002/mrm.27229. URL https://onlinelibrary.wiley.com/doi/abs/10.1002/mrm.27229.

[115] Felix Eckstein, Akshay S. Chaudhari, David Fuerst, Martin Gaisberger, Jana Kemnitz, Christian F. Baumgartner, Ender Konukoglu, David J Hunter, and Wolfgang Wirth. A deep learning automated segmentation algorithm accurately detects differences in longitudinal cartilage thickness loss – data from the fnih biomarkers study of the osteoarthritis initiative. *Arthritis Care & Research*, 74(6):929–936, 2022. doi: https://doi.org/10.1002/acr.24539. URL https://onlinelibrary.wiley.com/doi/abs/10.1002/acr.24539.

[116] Sibaji Gaj, Mingrui Yang, Kunio Nakamura, and Xiaojuan Li. Automated cartilage and meniscus segmentation of knee mri with conditional generative adversarial networks. *Magnetic Resonance in Medicine*, 84(1):437–449, 2020. doi: https://doi.org/10.1002/mrm.28111. URL https://onlinelibrary.wiley.com/doi/abs/10.1002/mrm.28111.

[117] Ruida Cheng, Natalia A. Alexandridi, Richard M. Smith, Aricia Shen, William Gandler, Evan McCreedy, Matthew J. McAuliffe, and Frances T. Sheehan. Fully automated patellofemoral mri segmentation using holistically nested networks: Implications for evaluating patellofemoral osteoarthritis, pain, injury, pathology, and adolescent development. *Magnetic Resonance in Medicine*, 83(1):139–153, 2020. doi: https://doi.org/10.1002/mrm.27920. URL https://onlinelibrary.wiley.com/doi/abs/10.1002/mrm.27920.

[118] Athina Spiliopoulou, Marco Colombo, Darren Plant, Nisha Nair, Jing Cui, Marieke JH Coenen, Katsunori Ikari, Hisashi Yamanaka, Saedis Saevarsdottir, Leonid Padyukov, S Louis Bridges Jr., Robert P Kimberly, Yukinori Okada, Piet L CM van Riel, Gertjan Wolbink, Irene E van der Horst-Bruinsma, Niek de Vries, Paul P Tak, Koichiro Ohmura, Helena Canhão, Henk-Jan Guchelaar, Tom WJ Huizinga, Lindsey A Criswell, Soumya Raychaudhuri, Michael E Weinblatt, Anthony G Wilson, Xavier Mariette, John D Isaacs, Ann W Morgan, Costantino Pitzalis, Anne Barton, and Paul McKeigue. Association of response to tnf inhibitors in rheumatoid arthritis with quantitative trait loci for cd40 and cd39. *Annals of the Rheumatic Diseases*, 78(8):1055–1061, 2019. ISSN 0003-4967. doi: 10.1136/annrheumdis-2018-214877. URL https://ard.bmj.com/content/78/8/1055.

[119] Yuanfang Guan, Hongjiu Zhang, Daniel Quang, Ziyan Wang, Stephen C. J. Parker, Dimitrios A. Pappas, Joel M. Kremer, and Fan Zhu. Machine learning to predict anti–tumor necrosis factor drug responses of rheumatoid arthritis patients by integrating clinical and genetic markers. *Arthritis & Rheumatology*, 71(12):1987–1996, 2019. doi: https://doi.org/10.1002/art.41056. URL https://onlinelibrary.wiley.com/doi/abs/10.1002/art.41056.

[120] Weiyang Tao, Arno N. Concepcion, Marieke Vianen, Anne C. A. Marijnissen, Floris P. G. J. Lafeber, Timothy R. D. J. Radstake, and Aridaman Pandit. Multiomics and machine learning accurately predict clinical response to adalimumab and etanercept therapy in patients with rheumatoid arthritis. *Arthritis & Rheumatology*, 73(2):212–222, 2021. doi: https://doi.org/10.1002/art.41516. URL https://onlinelibrary.wiley.com/doi/abs/10.1002/art.41516.

[121] Darren Plant, Mateusz Maciejewski, Samantha Smith, Nisha Nair, the RAMS Study Group the Maximising Therapeutic Utility in Rheumatoid Arthritis Consortium, Kimme Hyrich, Daniel Ziemek, Anne Barton, and Suzanne Verstappen. Profiling of gene expression biomarkers as a classifier of methotrexate nonresponse in patients with rheumatoid arthritis. *Arthritis & Rheumatology*, 71(5):678–684, 2019. doi: https://doi.org/10.1002/art.40810. URL https://onlinelibrary.wiley.com/doi/abs/10.1002/art.40810.

[122] Helen R. Gosselt, Maxime M. A. Verhoeven, Maja Bulatović-Ćalasan, Paco M. Welsing, Maurits C. F. J. de Rotte, Johanna M. W. Hazes, Floris P. J. G. Lafeber, Mark Hoogendoorn, and Robert de Jonge. Complex machine-learning algorithms and multivariable logistic regression on par in the prediction of insufficient clinical response to methotrexate in rheumatoid arthritis. *Journal of Personalized Medicine*, 11(1), 2021. ISSN 2075-4426. doi: 10.3390/jpm11010044. URL https://www.mdpi.com/2075-4426/11/1/44.

[123] Seulkee Lee, Yeonghee Eun, Hyungjin Kim, Hoon-Suk Cha, Eun-Mi Koh, and Jaejoon Lee. Machine learning to predict early TNF inhibitor users in patients with ankylosing spondylitis. *Scientific Reports*, 10(1):20299, 2020. ISSN 2045-2322. doi: 10.1038/s41598-020-75352-7. URL https://doi.org/10.1038/s41598-020-75352-7.

[124] Tianrun Cai, Tzu-Chieh Lin, Allison Bond, Jie Huang, Gwendolyn Kane-Wanger, Andrew Cagan, Shawn N Murphy, Ashwin N Ananthakrishnan, and Katherine P Liao. The Association Between Arthralgia and Vedolizumab Using Natural Language Processing. *Inflammatory Bowel Diseases*, 24(10):2242–2246, 05 2018. ISSN 1078-0998. doi: 10.1093/ibd/izy127. URL https://doi.org/10.1093/ibd/izy127.

[125] Jeffrey R. Curtis, Lang Chen, Phillip Higginbotham, W. Benjamin Nowell, Ronit Gal-Levy, James Willig, Monika Safford, Joseph Coe, Kaitlin O'Hara, and Roee Sa'adon. Social media for arthritis-related comparative effectiveness and safety research and the impact of direct-to-consumer advertising. *Arthritis Research & Therapy*, 19(1):48, 12 2017. ISSN 1478-6362. doi: 10.1186/s13075-017-1251-y. URL http://arthritis-research.biomedcentral.com/articles/10.1186/s13075-017-1251-y.

[126] Eldin Dzubur, Carine Khalil, Christopher V. Almario, Benjamin Noah, Deeba Minhas, Mariko Ishimori, Corey Arnold, Yujin Park, Jonathan Kay, Michael H. Weisman, and Brennan M. R. Spiegel. Patient concerns and perceptions regarding biologic therapies in ankylosing spondylitis: Insights from a large-scale survey of social media platforms. *Arthritis Care & Research*, 71(2):323–330, 2019. doi: https://doi.org/10.1002/acr.23600. URL https://onlinelibrary.wiley.com/doi/abs/10.1002/acr.23600.

[127] Chanakya Sharma, Samuel Whittle, Pari Delir Haghighi, Frada Burstein, Roee Sa'adon, and Helen Isobel Keen. Mining social media data to investigate patient perceptions regarding dmard pharmacotherapy for rheumatoid arthritis. *Annals of the Rheumatic Diseases*, 79(11): 1432–1437, 2020. ISSN 0003-4967. doi: 10.1136/annrheumdis-2020-217333. URL https://ard.bmj.com/content/79/11/1432.

[128] Xiaolan Mo, Xiujuan Chen, Hongwei Li, Jiali Li, Fangling Zeng, Yilu Chen, Fan He, Song Zhang, Huixian Li, Liyan Pan, Ping Zeng, Ying Xie, Huiyi Li, Min Huang, Yanling He, Huiying Liang, and Huasong Zeng. Early and accurate prediction of clinical response to methotrexate treatment in juvenile idiopathic arthritis using machine learning. *Frontiers in Pharmacology*, 10:1155, 2019. ISSN 1663-9812. doi: 10.3389/fphar.2019.01155. URL https://www.frontiersin.org/article/10.3389/fphar.2019.01155.

[129] Xiaolan Mo, Xiujuan Chen, Chifong Ieong, Song Zhang, Huiyi Li, Jiali Li, Guohao Lin, Guangchao Sun, Fan He, Yanling He, Ying Xie, Ping Zeng, Yilu Chen, Huiying Liang, and Huasong Zeng. Early prediction of clinical response to etanercept treatment in juvenile idiopathic arthritis using machine learning. *Frontiers in Pharmacology*, 11:1164, 2020. ISSN 1663-9812. doi: 10.3389/fphar.2020.01164. URL https://www.frontiersin.org/article/10.3389/fphar.2020.01164.

[130] Liangliang Liu, Ying Yu, Zhihui Fei, Min Li, Fang-Xiang Wu, Hong-Dong Li, Yi Pan, and Jianxin Wang. An interpretable boosting model to predict side effects of analgesics for osteoarthritis. *BMC Systems Biology*, 12(S6):105, 11 2018. ISSN 1752-0509. doi: 10.1186/s12918-018-0624-4. URL https://bmcsystbiol.biomedcentral.com/articles/10.1186/s12918-018-0624-4.

[131] Alice B Gottlieb, Philip J Mease, Bruce Kirkham, Peter Nash, Alejandro C Balsa, Bernard Combe, Jürgen Rech, Xuan Zhu, David James, Ruvie Martin, Gregory Ligozio, Ken Abrams, and Luminita Pricop. Secukinumab Efficacy in Psoriatic Arthritis: Machine Learning and Meta-analysis of Four Phase 3 Trials. *JCR: Journal of Clinical Rheumatology*, 27(6), 2021. ISSN 1076-1608. URL https://journals.lww.com/jclinrheum/Fulltext/2021/09000/Secukinumab_Efficacy_in_Psoriatic_Arthritis_.4.aspx.

[132] Tatsuya Atsumi, Yoshiaki Ando, Shinichi Matsuda, Shiho Tomizawa, Riwa Tanaka, Nobuhiro Takagi, and Ayako Nakasone. Prodromal signs and symptoms of serious infections with tocilizumab treatment for rheumatoid arthritis: Text mining of the japanese postmarketing adverse event-reporting database. *Modern Rheumatology*, 28(3):435–443, 2018. doi: 10.1080/14397595.2017.1366007. URL https://doi.org/10.1080/14397595.2017.1366007. PMID: 28880689.

[133] Ryunosuke Goto, Ryo Inuzuka, Takahiro Shindo, Yoshiyuki Namai, Yoichiro Oda, Yutaka Harita, and Akira Oka. Relationship between post-IVIG IgG levels and clinical outcomes in Kawasaki disease patients: new insight into the mechanism of action of IVIG. *Clinical Rheumatology*, 39(12):3747–3755, 2020. ISSN 1434-9949. doi: 10.1007/s10067-020-05153-w. URL https://doi.org/10.1007/s10067-020-05153-w.

[134] Mashfiqui Rabbi, Min SH Aung, Geri Gay, M Cary Reid, and Tanzeem Choudhury. Feasibility and Acceptability of Mobile Phone–Based Auto-Personalized Physical Activity Recommendations for Chronic Pain Self-Management: Pilot Study on Adults. *Journal of Medical Internet Research*, 20(10):e10147, 10 2018. ISSN 1438-8871. doi: 10.2196/10147. URL http://www.jmir.org/2018/10/e10147/.

[135] Wai Leung Ambrose Lo, Di Lei, Le Li, Dong Feng Huang, and Kin-Fai Tong. The Perceived Benefits of an Artificial Intelligence–Embedded Mobile App Implementing Evidence-Based Guidelines for the Self-Management of Chronic Neck and Back Pain: Observational Study. *JMIR mHealth and uHealth*, 6(11):e198, 11 2018. ISSN 2291-5222. doi: 10.2196/mhealth.8127. URL http://mhealth.jmir.org/2018/11/e198/.

[136] Jinchao Jia, Mengyan Wang, Yuning Ma, Jialin Teng, Hui Shi, Honglei Liu, Yue Sun, Yutong Su, Jianfen Meng, Huihui Chi, Xia Chen, Xiaobing Cheng, Junna Ye, Tingting Liu, Zhihong Wang, Liyan Wan, Zhuochao Zhou, Fan Wang, Chengde Yang, and Qiongyi Hu. Circulating neutrophil extracellular traps signature for identifying organ involvement and response to glucocorticoid in adult-onset still's disease: A machine learning study. *Frontiers in Immunology*, 11:2784, 2020. ISSN 1664-3224. doi: 10.3389/fimmu.2020.563335. URL

https://www.frontiersin.org/article/10.3389/fimmu.2020.563335.

[137] Chandrika S Bhat, Mark Chopra, Savvas Andronikou, Suvadip Paul, Zach Wener-Fligner, Anna Merkoulovitch, Izidora Holjar-Erlic, Flavia Menegotto, Ewan Simpson, David Grier, and Athimalaipet V Ramanan. Artificial intelligence for interpretation of segments of whole body MRI in CNO: pilot study comparing radiologists versus machine learning algorithm. *Pediatric Rheumatology*, 18(1):47, 2020. ISSN 1546-0096. doi: 10.1186/s12969-020-00442-9. URL https://doi.org/10.1186/s12969-020-00442-9.

[138] O. Sangha. Epidemiology of rheumatic diseases. *Rheumatology*, 39:3–12, 12 2000. ISSN 1462-0324. doi: 10.1093/rheumatology/39.suppl_2.3. URL https://doi.org/10.1093/rheumatology/39.suppl_2.3.

[139] Nihan Acar-Denizli, Belchin Kostov, Manuel Ramos-Casals, and Sjögren Big Data Consortium. The big data sjögren consortium: a project for a new data science era. *Clinical and experimental rheumatology*, 37 Suppl 118:19–23, 2019. ISSN 0392-856X.

[140] Aurélien Géron. *Hands-on Machine Learning with Scikit-Learn, Keras, and TensorFlow: Concepts, Tools, and Techniques to Build Intelligent Systems*. O'Reilly Media, Sebastopol, CA, 2017. ISBN 978-1492032649.

[141] Joseph Futoma, Morgan Simons, Trishan Panch, Finale Doshi-Velez, and Leo Anthony Celi. The myth of generalisability in clinical research and machine learning in health care. *The Lancet Digital Health*, 2:e489–e492, 9 2020. ISSN 25897500. doi: 10.1016/S2589-7500(20)30186-2.

[142] Hun-Sung Kim, Suehyun Lee, and Ju Han Kim. Real-world evidence versus randomized controlled trial: Clinical research based on electronic medical records. *Journal of Korean Medical Science*, 33, 2018. ISSN 1011-8934. doi: 10.3346/jkms.2018.33.e213.

[143] Jyothi Subramanian and Richard Simon. Overfitting in prediction models – is it a problem only in high dimensions? *Contemporary Clinical Trials*, 36(2):636–641, 2013. ISSN 1551-7144. doi: https://doi.org/10.1016/j.cct.2013.06.011. URL https://www.sciencedirect.com/science/article/pii/S1551714413001031.

[144] Christopher J. Kelly, Alan Karthikesalingam, Mustafa Suleyman, Greg Corrado, and Dominic King. Key challenges for delivering clinical impact with artificial intelligence. *BMC Medicine*, 17(1):195, 12 2019. ISSN 1741-7015. doi: 10.1186/s12916-019-1426-2. URL https://bmcmedicine.biomedcentral.com/articles/10.1186/s12916-019-1426-2.

[145] Vincenzo Lagani, George Kortas, and Ioannis Tsamardinos. Biomarker signature identification in "omics" data with multi-class outcome. *Computational and Structural Biotechnology Journal*, 6(7):e201303004, 2013. ISSN 2001-0370. doi: https://doi.org/10.5936/csbj.201303004. URL https://www.sciencedirect.com/science/article/pii/S2001037014601136.

[146] Marzyeh Ghassemi, Luke Oakden-Rayner, and Andrew L Beam. The false hope of current approaches to explainable artificial intelligence in health care. *The Lancet Digital Health*, 3:e745–e750, 11 2021. ISSN 25897500. doi: 10.1016/S2589-7500(21)00208-9.

[147] Sandeep Reddy. Explainability and artificial intelligence in medicine. *The Lancet Digital Health*, 4:e214–e215, 4 2022. ISSN 25897500. doi: 10.1016/S2589-7500(22)00029-2.

[148] Gómez-González E and Gomez Gutierrez E. Artificial intelligence in medicine and healthcare: applications, availability and societal impact. Scientific analysis or review KJ-NA-30197-EN-N (online), European Union, Luxembourg (Luxembourg), 2020.

[149] European Comission. Regulatory framework proposal on artificial intelligence, 2022. URL https://digital-strategy.ec.europa.eu/en/policies/regulatory-framework-ai.

[150] European Commission, Content Directorate-General for Communications Networks, and Technology. *Ethics guidelines for trustworthy AI*. Publications Office, 2019. doi: doi/10.2759/346720.

[151] Roberto V. Zicari, James Brusseau, Stig Nikolaj Blomberg, Helle Collatz Christensen, Megan Coffee, Marianna B. Ganapini, Sara Gerke, Thomas Krendl Gilbert, Eleanore Hickman, Elisabeth Hildt, Sune Holm, Ulrich Kühne, Vince I. Madai, Walter Osika, Andy Spezzatti, Eberhard Schnebel, Jesmin Jahan Tithi, Dennis Vetter, Magnus Westerlund, Renee Wurth, Julia Amann, Vegard Antun, Valentina Beretta, Frédérick Bruneault, Erik Campano, Boris Düdder, Alessio Gallucci, Emmanuel Goffi, Christoffer Bjerre Haase, Thilo Hagendorff, Pedro Kringen, Florian Möslein, Davi Ottenheimer, Matiss Ozols, Laura Palazzani, Martin Petrin, Karin Tafur, Jim Tørresen, Holger Volland, and Georgios Kararigas. On assessing trustworthy ai in healthcare. machine learning as a supportive tool to recognize cardiac arrest in emergency calls. *Frontiers in Human Dynamics*, 3, 2021. ISSN 2673-2726. doi: 10.3389/fhumd.2021.673104. URL https://www.frontiersin.org/articles/10.3389/fhumd.2021.673104.

[152] Media Secretary of State for Digital, Culture and Sport. Establishing a pro-innovation approach to regulating ai, 2022. URL https://www.gov.uk/government/publications/establishing-a-pro-innovation-approach-to-regulating-ai.

[153] Lukas Folle, Sara Bayat, Arnd Kleyer, Filippo Fagni, Lorenz A Kapsner, Maja Schlereth, Timo Meinderink, Katharina Breininger, Koray Tascilar, Gerhard Krönke, Michael Uder, Michael Sticherling, Sebastian Bickelhaupt, Georg Schett, Andreas Maier, Frank Roemer, and David Simon. Advanced neural networks for classification of MRI in psoriatic arthritis, seronegative, and seropositive rheumatoid arthritis. *Rheumatology*, 03 2022. ISSN 1462-0324. doi: 10.1093/rheumatology/keac197. URL https://doi.org/10.1093/rheumatology/keac197. keac197.

[154] Vincenzo Venerito, Giacomo Emmi, Luca Cantarini, Pietro Leccese, Marco Fornaro, Claudia Fabiani, Nancy Lascaro, Laura Coladonato, Irene Mattioli, Giulia Righetti, Danilo Malandrino, Sabina Tangaro, Adalgisa Palermo, Maria Letizia Urban, Edoardo Conticini, Bruno Frediani, Florenzo Iannone, and Giuseppe Lopalco. Validity of machine learning in predicting giant cell arteritis flare after glucocorticoids tapering. *Frontiers in Immunology*, 13, 2022. ISSN 1664-3224. doi: 10.3389/fimmu.2022.860877. URL https://www.frontiersin.org/articles/10.3389/fimmu.2022.860877.