*Review*

# The Use and Utility of Machine Learning in Achieving Precision Medicine in Systemic Sclerosis: A Narrative Review

Francesco Bonomi [1,†], Silvia Peretti [1,†], Gemma Lepri [1], Vincenzo Venerito [2], Edda Russo [1], Cosimo Bruni [1,3], Florenzo Iannone [2], Sabina Tangaro [4], Amedeo Amedei [1], Serena Guiducci [1], Marco Matucci Cerinic [1,5] and Silvia Bellando Randone [1,*]

1   Department of Clinical and Experimental Medicine, University of Florence, 50134 Florence, Italy;
    francesco.bonomi@unifi.it (F.B.); silvia.peretti@unifi.it (S.P.); gemma.lepri@unifi.it (G.L.);
    edda.russo@unifi.it (E.R.); cosimo.bruni@unifi.it (C.B.); amedeo.amedei@unifi.it (A.A.);
    serena.guiducci@unifi.it (S.G.); marco.matuccicerinic@unifi.it (M.M.C.)
2   Rheumatology Unit, Department of Emergency and Organ Transplantations, University of Bari Aldo Moro,
    70121 Bari, Italy; vincenzo.venerito@uniba.it (V.V.); florenzo.iannone@uniba.it (F.I.)
3   Department of Rheumatology, University Hospital of Zurich, University of Zurich, 8006 Zurich, Switzerland
4   Department of Soil, Plant and Food Sciences, University of Bari Aldo Moro, Istituto Nazionale di Fisica
    Nucleare, Sezione di Bari, 70121 Bari, Italy; sabina.tangaro@uniba.it
5   Unit of Immunology, Rheumatology, Allergy and Rare Diseases (UnIRAR), IRCCS San Raffaele Hospital,
    20132 Milan, Italy
*   Correspondence: silvia.bellandorandone@unifi.it
†   These authors contributed equally to this work.

**Abstract:** Background: Systemic sclerosis (SSc) is a rare connective tissue disease that can affect different organs and has extremely heterogenous presentations. This complexity makes it difficult to perform an early diagnosis and a subsequent subclassification of the disease. This hinders a personalized approach in clinical practice. In this context, machine learning (ML), a branch of artificial intelligence (AI), is able to recognize relationships in data and predict outcomes. Methods: Here, we performed a narrative review concerning the application of ML in SSc to define the state of art and evaluate its role in a precision medicine context. Results: Currently, ML has been used to stratify SSc patients and identify those at high risk of severe complications. Additionally, ML may be useful in the early detection of organ involvement. Furthermore, ML might have a role in target therapy approach and in predicting drug response. Conclusion: Available evidence about the utility of ML in SSc is sparse but promising. Future improvements in this field could result in a big step toward precision medicine. Further research is needed to define ML application in clinical practice.

**Keywords:** systemic sclerosis; machine learning; artificial intelligence; precision medicine

## 1. Introduction

Systemic sclerosis (SSc) is a rare connective tissue disease characterized by autoimmune features, vasculopathy, and fibrosis [1]. Among rheumatic diseases, SSc has the highest mortality rate, due to the fact that treatment options do not address both the fibrotic and inflammatory disease features [2,3]. Skin fibrosis is a well-known hallmark of the disease and the extension of skin involvement has been proven to influence disease-associated mortality [4]. However, many different organs can be involved in the course of the disease, including heart, lungs, kidneys, and gastrointestinal tract. Pulmonary complications are the most common cause of death in SSc, with interstitial lung disease and pulmonary arterial hypertension being the most common manifestations, occurring in approximately one-third of cases and being associated with reduced survival [2].

Since SSc is an uncommon disease with multiple heterogeneous symptoms, early diagnosis and predicting the risk of disease progression pose a significant challenge for physicians [5,6]. In fact, SSc is characterized by a wide variability in both the clinical

phenotype, the genetic expression, the autoantibody pattern, and disease evolution, limiting the progress in research due to relevant patient heterogeneity and imperfect outcome measurements [7]. The disease is still classified according to skin fibrosis extension in a limited and a diffuse cutaneous subset, although recently it has become clearly evident that this subclassifying method may not be able, in clinical practice, to predict the risk of progression [8,9]. For this reason, we searched for new methods able to combine multiple variables with the final aim to provide a patients' outcome and risk stratification. Promising help could come from the use of artificial intelligence and specifically the machine learning model.

Machine learning (ML) is a branch of artificial intelligence (AI) that uses algorithms to recognize relationships in data and its use in the medical field is increasing [10]. An ML approach includes supervised and unsupervised model learning [11]: the former is constructed to predict known groups or values, while the latter is used to find associations and patterns out of raw data that results in groups of similar samples. As a general rule, ML algorithms try to discover underlying and unanticipated connections in data, helping to generate hypotheses [11].

Until now, ML has been used in various areas of bio-medicine for different purposes such as the analysis of medical images [12], the subclassification of patient cohorts [13], the prediction of drug response [14], and to guide personalized medicine [15]. Of interest, ML has also been used in gut microbiome research and has been useful to predict atherosclerotic cardiovascular disease [16] and to predict chemotherapy effectiveness and tolerability in the oncology setting [17]. In rheumatology, ML approaches have been studied so far to aid in the identification of new possible therapies in primary Sjogren's syndrome and systemic lupus erythematosus (SLE) [18], and to classify patients or predict disease outcomes based on genetic data [19]. When deployed on data from patients with inflammatory arthritis, supervised ML algorithms have proven effective in predicting treatment responses or for the automatic histopathological grading of synovitis [20,21].

To date, few studies have reported the use of ML in SSc as a possibility to overcome many current limitations in diagnosis and treatment. It is conceivable that SSc might need a multidimensional approach to build predictive models. The superiority of ML methods considering nonlinear relationships is crucial as it could refine our *modus cogitandi* when dealing with the complexity of SSc [22]. In fact, it could help physicians predict organ involvement and allow a customized treatment according to genetics, autoantibody profile, and organ involvement, and help in finding new biomarkers to tailor the follow-up [23]. The aim of this paper was to summarize the current evidence on the use of ML in SSc and evaluate its possible role in precision medicine through a narrative review of the literature.

## 2. Materials and Methods

We performed a nonsystematic (PRISMA protocol not followed) narrative literature review on PubMed with the following MeSH terms: [scleroderma, systemic AND machine learning]; [scleroderma, systemic AND machine intelligence]; [scleroderma, systemic AND artificial intelligence]; [scleroderma, systemic AND deep learning]. No time limit was put on the research and each paper's reference list was checked for additional studies.

## 3. Results

Our literature search retrieved articles published between 2015 and 2022. The data derived showed that ML has found possible applications in different fields, ranging from the subclassification of SSc patients to predictions of treatment response. A summary of the analyzed works is reported in Table 1.

**Table 1.** Summary of analyzed works regarding ML use in SSc.

| Authors | Year of Publishing | Journal | No. of Patients | Aim of ML Use |
|---|---|---|---|---|
| **Leeuwen N.M. van, et al. [24]** | 2021 | *RMD Open* | 248 | To predict risk of disease progression in order to develop a tailor-made follow-up |
| **Franks J.M., et al. [25]** | 2019 | *Arthritis Rheumatol.* | 102 | To identify specific molecular signatures from skin biopsies which can be related to disease outcome |
| **Xu X., et al. [26]** | 2020 | *PLoS ONE* | 221 | To identify molecular pathways from skin biopsies in order to obtain a finer SSc stratification |
| **Showalter K., et al. [27]** | 2021 | *Ann. Rheum. Dis.* | 26 | To identify molecular signatures able to predict the treatment response (improvers vs. nonimprovers) |
| **Murdaca G., et al. [28]** | 2021 | *Diagnostics* | 38 | To predict early pulmonary involvement in asymptomatic patients |
| **Andrade D.S.M., et al. [29]** | 2021 | *Biomed. Eng. OnLine* | 82 | To examine lung function data coming from respiratory oscillometry test |
| **Chassagnon G., et al. [30]** | 2020 | *Radiol. Artif. Intell.* | 208 | To quantify lung disease extension from HRCT images |
| **Taroni J.N., et al. [31]** | 2017 | *J. Invest. Dermatol.* | Meta-analysis (total 35) | To evaluate gene expressions on skin biopsies and predict response to different treatments |
| **Ebata S., et al. [32]** | 2022 | *Rheumatol. Oxf. Engl.* | 54 | To find possible predictors of favorable response to RTX |
| **Zamanian R.T., et al. [33]** | 2021 | *Am. J. Respir. Crit. Care Med.* | 57 | To evaluate RTX response in SSc-related PAH |
| **Franks J.M., et al. [34]** | 2020 | *Ann. Rheum. Dis.* | 63 | To evaluate stem cell response in severe SSc |
| **Schniering J., et al. [35]** | 2022 | *Eur. Respir. J.* | 118 | To identify homogeneous imaging-based ILD clusters through a radiomic analysis of lung CT in SSc patients |

### 3.1. ML in the Stratification of SSc Patients

The natural course of disease is extremely heterogeneous and the management of SSc patients remains a great challenge. Hence, the need for clinicians to stratify patients according to the risk of disease progression and organ involvement. The common goal of the scientific community is to develop a precision medicine model with a tailor-made follow-up able to identify progressors early and limit over-testing. Currently, there are no evidence-based guidelines regarding the intensity and the frequency of follow-up in SSc patients [23], except for the management of pulmonary arterial hypertension (PAH) [36].

In this framework, ML may integrate several data and it may be helpful for clinicians to identify high-risk patients, with few but promising studies available so far.

Van Leeuwen and colleagues [24] successfully created a prediction model able to identify nonprogressors. First, they defined progression as worsening in one or more organ systems and/or start of immunosuppressive therapy or death between two visits. Patients with none of these events were labeled as 'non progressors'. Consequently, they considered a group of patients with at least three assessments and with a complete organ evaluation available (such as: modified Rodnan Skin Score (mRSS), blood tests, 24 h ECG recording, echocardiography, pulmonary function test, high-resolution chest tomography, cardiopulmonary exercise test). A total of 90 variables derived from the assessments were approached through a machine learning model able to extract signs of progression. Patients were classified as high, intermediate, and low risk of progression and they were compared with real outcomes after a period of 5.4 years. The results of this study showed that the model perfectly recognized patients at low risk of progression with a correctness of 100%; in fact, none of the low-risk patients really experienced worsening of disease. By contrast, less than half of the patients identified by ML as being high-risk really progressed. These

results may have implications in clinical practice, as low-risk patients might be assessed with less intensive follow-up protocols, therefore reducing the burden of health care cost with minimal risks of missing organ progressors.

Other authors proposed a classification based on skin transcriptomic data, with the rationale that a deeper understanding of SSc pathogenesis could better define patient subgroups. The aim was to identify specific molecular signatures that can be related to the disease course. In this framework, ML was used to integrate genetic and clinical data with molecular phenotype with the final goal to predict disease outcome. Multiple studies have already identified four molecular subsets through gene expression in skin biopsies [37–40]: normal-like, limited, fibroproliferative, and inflammatory. These subsets combine a plethora of cytokines (such as IL-5, IL-6, IL-13, IL-18, PDGF, etc.) playing a role in SSc pathogenesis [41,42]. Of note, Frank et al. [25] identified a classifier unsupervised machine learning algorithm able to assign a single skin sample to a specific gene expression subset. Their study included 297 microarrays from 102 patients, presenting genes derived from all the expression subsets. The four subsets were all homogeneously represented, except for the limited subtype that was underrepresented. They found 245 genes expressed in the inflammatory subset, mostly related to immune system response, and 245 genes in the fibroproliferative one with high specificity (95.8% vs. 94.1%, respectively) and good sensitivity (83.3% vs. 89.7%, respectively). In contrast, the limited subset showed lower sensitivity, likely due to the limited size of the sample. This study is proof that a set of genes might identify a specific patient subset, with potential implications in prognosis and therapy. For example, the inflammatory subset is characterized by the activation of the immune system and seems to be associated with a better response to disease-modifying antirheumatic drugs (DMARDs) [38].

Xu and colleagues [26] recently identified 80 pathway signatures that could stratify patients into eight subtypes. They analyzed microarrays from 221 involved skin samples of 141 SSc patients at the time of diagnosis and 80 healthy controls. An ML-assisted model identified 80 pathways that could be summarized in five mean pathways: Metabolism-1, Metabolism-2, Immune-fibrosis, Immune Response-1, and Immune Response-2. These five pathway modules could differentiate the eight clusters of disease, built through different combinations of their gene expressions. Clusters 1 and 6 were characterized by metabolic pathways (Metabolism-1 and Metabolism-2) and they were similar to the control cohort, although cluster 1 showed more immune activation than cluster 6 and control cohort. Clusters 2 and 4 were characterized by inflammation and immune response activation with high levels of IL-17 and IL-22. Cluster 3 was characterized by an immune-fibrosis pathway. Clusters 5 and 8 exhibited a mixture of inflammatory, pro-fibrotic, and metabolism signatures, and lastly cluster 7 was identified as a separate normal-like cohort with low expression of Metabolism-1 and high expression of Metabolism-2. From the different clusters, the authors managed to underline different cells involved in pathogenesis, with clusters 1 and 2 enriched with myeloid cells and macrophages, cluster 3 enriched with fibroblast and endothelial cells, and cluster 4 with strong expression in all four types of cells. This study highlighted different molecular signatures and pathways, leading the way to a tailor-made treatment and a personalized assessment.

Another study [27] analyzed 48 forearm skin samples from 26 SSc patients treated with nilotinib or with belimumab, integrating molecular data with clinical data such as mRSS, blood tests, FVC, and patient-reported outcomes. The aim was to define a molecular signature at treatment initiation and 52 weeks after, in order to identify improving and nonimproving patients using an ML approach. Samples were evaluated by considering seven histologic features, among which CD34+ and aSMA appeared to be the most predictive elements. This study shows that samples with high aSMA and low CD34+ had higher inflammatory gene expression and higher mRSS. In contrast, low aSMA and high CD34+ were compatible with the normal-like subset. The machine learning model identified CD34+ as a predictor of fibroproliferative and normal-like subsets, whereas aSMA was predictive of the inflammatory subset.

### *3.2. ML Algorithms to Diagnose and Evaluate Lung Involvement*

The assessment of lung involvement in SSc currently includes pulmonary function tests (PFTs) and high-resolution computerized tomography (HRCT). Although the latter is the gold standard for ILD detection, there are no clear recommendations about its use and timing to repeat it [43]. Nowadays, clinicians perform HRCT when PFTs show a decline or respiratory symptoms worsen, with the result that some patients might be missed and some others over-tested [44]. In this context, ML algorithms might detect pulmonary involvement before any functional sign of deterioration occurs, with the benefit of improving survival, reducing health care costs, and relieving an unnecessary burden for patients, including radiation exposure.

Murdaca et al. [28] developed a risk-free screening test using ML algorithms, with the aim to predict early pulmonary involvement in asymptomatic patients. They collected clinical data including PFT from SSc patients: seven parameters were identified as the most performant in predicting the presence of lung involvement. Total Lung Capacity (TLC) was estimated to be the best marker, mostly if taken along with forced expiratory volume in the 1st second (FEV1), forced vital capacity (FVC), diffusion lung carbon monoxide (DLCO), and impedance pH monitoring.

Andrade et al. [29] studied the role of the Forced Oscillation Technique (FOT) to detect resistance and reactance in respiratory dynamics. This technique is considered complementary to spirometry [45] because it evaluates different respiratory parameters, with the advantage that it requires minimal cooperation since forced expiratory maneuvers are not needed. The authors proposed an ML approach able to examine lung function data coming from respiratory oscillometry tests.

The extension of lung involvement on CT images has been acknowledged as an independent predictor of mortality [46,47] regardless of the radiological pattern. Chassagnon and colleagues [30] conceived an automated deep-learning-based model to quantify lung involvement, overcoming the variability of a visual approach [48,49]. A machine learning algorithm was trained with reported CT images to determine the extension of the ILD and subsequently a group of patients, with at least two CT of the chest and PFT tests, were included in the analysis. They then compared assessments of ILD extension coming from ML with the ones reported by three qualified radiologists. This algorithm proved to be an accurate and reproducible tool able to quantify lung extension with great accuracy and showing a good correlation between lung involvement and PFT values, surely higher than the one reported for visual assessment.

ML is also rapidly emerging as a powerful tool for radiomic analysis. Radiomic analysis represents a method for the quantitative description of medical images able to describe the tissue in terms of its intensity, texture, and advanced statistical properties through computationally retrieved quantitative data derived from medical images. The added value of radiomic analysis lies in the ability to capture tissue phenotypes on different spatial scales ranging from the radiological/macroscopic to the molecular/microscopic levels [50]. Schniering et al. [35] used an ML approach for radiomic analysis of lung CT in SSc patients to identify homogeneous imaging-based ILD clusters. The analysis produced two distinct and stable patient clusters based on their radiomic profiles. The differences in clinical characteristics were substantial, with patients in cluster 2 (*n* = 31) having a significantly more impaired lung performance in the 6 min walk test and a higher frequency of pulmonary hypertension than patients in cluster 1 (*n* = 59). Cluster 2 was also significantly enriched for honeycombing as a radiological sign of more severe fibrotic lung remodeling [35].

### *3.3. Early Detection of PAH with ML*

PAH is a serious complication of SSc which can occur in 8–12% of patients, leading to death in 25–40% of cases [51–53]. The gold-standard diagnostic tool for this complication is right heart catheterization (RHC) [54]. The early diagnosis and treatment of milder forms of PAH can lead to better prognosis and increase survival. Nowadays, several tools are

available for screening of PAH such as echocardiography, PFT with DLCO, and blood biomarkers (N-terminal prohormone of brain-natriuretic peptide -NT pro-BNP, uric acid). The lack of specificity or sensitivity of the isolated evaluations is partially overcome through their combination into multidomain algorithms [55].

Many authors have shown interest in finding tools able to identify patients with high risk of PAH early, in the framework of precision medicine. Bauer et al. [56] proposed a proteomic screening panel that may incorporate the DETECT algorithm, a two-step noninvasive prediction score, with the potential to enhance specificity. They collected clinical data and serum samples from the DETECT cohort and identified a panel of eight proteins whose expression could discriminate PAH patients from non-PAH patients with the aid of ML. The identified proteins have different roles in the context of PAH: some of them play a role in pulmonary vascular remodeling (RAGE, MMP-2); others are indicative of cardiac dysfunction (NT pro-BNP and IGFBP-7), whereas others are involved in angiogenesis (Collagen IV, endostatin, IGFBP-2, neuropilin-1). Interestingly, this protein panel showed higher accuracy in identifying PAH-SSc patients than NT pro-BNP alone, suggesting that the ML algorithm may detect PAH before the onset of cardiac stress.

*3.4. ML and Tailored SSc Treatment*

The treatment of SSc is still a challenge since most currently available drugs target a single component of the disease pathogenesis (either the vasculopathy, the autoimmune, or fibrotic features). [57] For this reason, every SSc patient should receive a personalized targeted regimen according to the organs involved, the prevalent pathogenic pattern, and the genetic expression. ML can be a very promising tool in these cases because it could help in predicting the response to different drugs. At the moment, few studies have reported the use of an ML approach to personalize SSc treatment.

One study used an ML approach in a meta-analysis [31] to analyze the gene expression data from skin biopsies of patients with SSc before and after five different treatments (Mycophenolate Mofetil (MMF), Rituximab, abatacept, nilotinib and fresolimumab) and to evaluate the treatment response in the skin according to the variations in mRSS over time. Gene expression analysis before and after treatment showed that all therapies except for fresolimumab (an anti-TGFβ monoclonal antibody) modulated immune response-related hallmarks (i.e., IL6/JAK/STAT3 signaling and TNFA/NFKB signaling). The ML algorithm showed that patients who did not respond to fresolilumab had an increased expression of inflammatory genes in the skin biopsy at baseline, suggesting a potential benefit from an immunosuppressant such as MMF. MMF responders, in fact, showed low levels of these inflammatory genes (genes involved in lymphocyte aggregation and type I interferon production). This approach allowed researchers to understand whether patients who failed one therapy could benefit from a different one.

Ebata et al. [32] used ML to analyze the skin sclerosis response in a cohort of 54 patients who received either rituximab or placebo. The study was a double-blind, parallel-group comparison in patients with SSc and showed the efficacy of rituximab on skin sclerosis. An ML approach was used to analyze twenty-seven baseline factors (including sex, different SSc-related signs and symptoms, previous treatments, autoantibody patterns, peripheral CD19-positive cell counts, blood test parameters, and lung involvement) to identify subpopulations with different degree of response to rituximab based on mRSS change at 24 weeks. The algorithm showed that SSc patients with high CD19-positive cell counts ($\geq$57/$\mu$L) and high mRSS ($\geq$17) are expected to have greater improvement in mRSS with rituximab. Another interesting subgroup that showed good treatment outcomes was characterized by high CD19-positive cell counts, low mRSS (<17) and serum Surfactant Protein-D (SP-D) levels $\geq$151 ng/mL; this final variable may hence help in predicting the treatment effect.

An ML approach was also used to predict rituximab responsiveness in SSc-related PAH (SSc-PAH) [33]. Data from 57 SSc patients with PAH were collected in a multicenter, double-blinded, placebo-controlled study which randomized participants to receive two infusions 14 days apart of 1000 mg rituximab or a placebo. The ML algorithm showed that

low levels of rheumatoid factors IL-12 and IL-17 were sensitive and specific predictors of a better rituximab response, as measured by an improved 6-min walking distance.

ML also predicted responses to hematopoietic stem-cell transplants (HSCT) in 63 severe SSc cases by using data on gene expression from peripheral blood cells (PBCs) [34]. Patients were divided into two groups: one was treated with cyclophosphamide (CYC) while the other received HSCT. PBCs were collected at baseline and at follow-up at different times after transplant (8, 14, 20/26, 38, 44/48/54 months). Participants who completed any treatment protocol were stratified by intrinsic gene-expression subsets at baseline, evaluated for event-free survival (EFS), and analyzed for differentially expressed genes (DEGs). The ML algorithm showed that patients from the fibroproliferative subset receiving HSCT experienced a significant improvement in EFS with respect to the ones who were treated with CYC ($p$ = 0.0091). On the other hand, patients from the normal-like subset or the inflammatory subset did not show a significant difference in EFS regardless of treatment allocation. ML was therefore helpful in determining that just the fibroproliferative subset shows a long-term benefit from HSCT compared to CYC.

## 4. Discussion and Future Perspectives

This narrative review on the use of ML in SSc highlighted many advantages we could obtain from the use of AI in either the diagnostic or the therapeutic approach to the disease (Figure 1). Predicting the future of ML is difficult, as the time scale at which developments will occur is very hard to estimate [58].
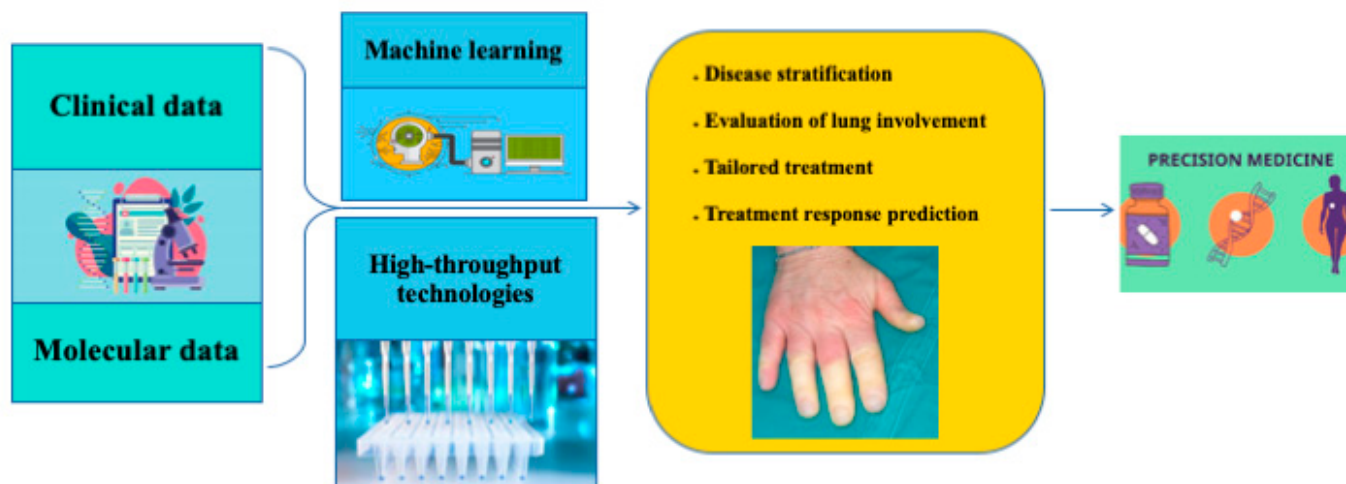


**Figure 1.** Overview of ML application in SSc to help in precision medicine.

To date, the main use of ML has been in oncology-targeted therapy to predict response to chemotherapies and personalize the treatment [14]. In rheumatology, ML has been mainly utilized in nonrare diseases, such as rheumatoid arthritis and SLE, to diagnose patients early at disease onset, identify the ones at higher risk of developing internal organ complications, and find the most effective personalized treatment. The application of ML in this field could provide useful information to guide treatment and predict drug response; however, at the moment, this promising field is largely unexplored [59]. The use of ML in SSc borrowed many concepts and applications from the abovementioned conditions and its future probably rests in their findings, as ML needs a big amount of data to find correlations and connections, making its application in rare diseases slower.

ML could become an extremely helpful tool in personalizing the follow-up of SSc patients, as it could quantify the individual risk of developing specific organ complications and subclassify patients according to their risk of disease progression or the target involved organ. In this regard, an ML approach is a potential game-changer, given its capability of handling both linear and complex nonlinear relationships between patient attributes that are hard to model with traditional statistical methods, especially in complex diseases

such as SSc. These modern approaches might allow physicians to both personalize the follow-up (as to which time interval to monitor tests and which exams to make according to the organ more at risk of involvement) and to target the therapy according to both the predicted clinical course and the drug to which each patient could respond (depending on genes expressed at biopsy, blood biomarkers, and instrumental examinations). ML provides promising tools for analyzing large amounts of data (clinical characteristics, laboratory results, treatments, and outcomes) and could allow us to discover new markers to determine the disease outcomes or drug responses. Additionally, ML use in comparative imaging could allow physicians to detect abnormalities that would otherwise be missed and could help to differentiate relevant from nonrelevant changes.

However, barriers to the adoption of ML in rheumatology do exist. One of the reasons emphasized in much of the literature is the role of complementary innovations in the successful adoption of ML and other information technology by companies. The performance of ML algorithms is also contingent on the quality of the data available. Thus, a second barrier to adoption is limited access to data. Medical data are often difficult to collect and difficult to access. Rheumatologists often resent the data-collection process when it interrupts their workflow, and the collected data are often incomplete [60].

Finally, another barrier is that ML and AI models are selected with the criterion of maximizing accuracy. This results in the selection of more complex models and often the loss of transparency of predictive outputs.

The present review's main limitation is in the nature of the method: being a narrative review, it is too subjective and lacks a systemic approach.

## 5. Conclusions

AI and ML have been used successfully in different medical fields, from oncology to rheumatology. Despite its application in SSc being still sparse, the data collected up until now are promising. Applications of ML are potentially going to increase with time, proving its usefulness in supporting physicians towards personalized medicine.

## References

1. Varga, J.; Trojanowska, M.; Kuwana, M. Pathogenesis of systemic sclerosis: Recent insights of molecular and cellular mechanisms and therapeutic opportunities. *J. Scleroderma Relat. Disord.* **2017**, *2*, 137–152. [CrossRef]
2. Elhai, M.; Meune, C.; Boubaya, M.; Avouac, J.; Hachulla, E.; Balbir-Gurman, A.; Riemekasten, G.; Airò, P.; Joven, B.; Vettori, S.; et al. Mapping and predicting mortality from systemic sclerosis. *Ann. Rheum. Dis.* **2017**, *76*, 1897–1905. [CrossRef] [PubMed]
3. Toledano, E.; Candelas, G.; Rosales, Z.; Martínez Prada, C.; León, L.; Abásolo, L.; Loza, E.; Carmona, L.; Tobías, A.; Jover, J.Á. A meta-analysis of mortality in rheumatic diseases. *Reumatol. Clín.* **2012**, *8*, 334–341. [CrossRef] [PubMed]
4. Shand, L.; Lunt, M.; Nihtyanova, S.; Hoseini, M.; Silman, A.; Black, C.M.; Denton, C.P. Relationship between change in skin score and disease outcome in diffuse cutaneous systemic sclerosis: Application of a latent linear trajectory model. *Arthritis Rheum.* **2007**, *56*, 2422–2431. [CrossRef]
5. Bellando-Randone, S.; Galdo, F.D.; Lepri, G.; Minier, T.; Huscher, D.; Furst, D.E.; Allanore, Y.; Distler, O.; Czirják, L.; Bruni, C.; et al. Progression of patients with Raynaud's phenomenon to systemic sclerosis: A five-year analysis of the European Scleroderma Trial and Research group multicentre, longitudinal registry study for Very Early Diagnosis of Systemic Sclerosis (VEDOSS). *Lancet Rheumatol.* **2021**, *3*, e834–e843. [CrossRef]
6. Akay, M.; Du, Y.; Sershen, C.L.; Wu, M.; Chen, T.Y.; Assassi, S.; Mohan, C.; Akay, Y.M. Deep Learning Classification of Systemic Sclerosis Skin Using the MobileNetV2 Model. *IEEE Open J. Eng. Med. Biol.* **2021**, *2*, 104–110. [CrossRef] [PubMed]

7. Gordon, J.K.; Domsic, R.T. Clinical Trial Design Issues in Systemic Sclerosis: An Update. *Curr. Rheumatol. Rep.* **2016**, *18*, 38. [CrossRef] [PubMed]

8. Leclair, V.; Hudson, M.; Proudman, S.M.; Stevens, W.M.; Fritzler, M.J.; Wang, M.; Nikpour, M.; Baron, M. Subsets in systemic sclerosis: One size does not fit all. *J. Scleroderma Relat. Disord.* **2016**, *1*, 298–306. [CrossRef]

9. LeRoy, E.C.; Black, C.; Fleischmajer, R.; Jablonska, S.; Krieg, T.; Medsger, T.A.; Rowell, N.; Wollheim, F. Scleroderma (systemic sclerosis): Classification, subsets and pathogenesis. *J. Rheumatol.* **1988**, *15*, 202–205.

10. Jordan, M.I.; Mitchell, T.M. Machine learning: Trends, perspectives, and prospects. *Science* **2015**, *349*, 255–260. [CrossRef]

11. Topol, E.J. High-performance medicine: The convergence of human and artificial intelligence. *Nat. Med.* **2019**, *25*, 44–56. [CrossRef]

12. de Bruijne, M. Machine learning approaches in medical image analysis: From detection to diagnosis. *Med. Image Anal.* **2016**, *33*, 94–97. [CrossRef]

13. Deeb, S.J.; Tyanova, S.; Hummel, M.; Schmidt-Supprian, M.; Cox, J.; Mann, M. Machine Learning-based Classification of Diffuse Large B-cell Lymphoma Patients by Their Protein Expression Profiles. *Mol. Cell. Proteom.* **2015**, *14*, 2947–2960. [CrossRef]

14. Ali, M.; Aittokallio, T. Machine learning and feature selection for drug response prediction in precision oncology applications. *Biophys. Rev.* **2019**, *11*, 31–39. [CrossRef] [PubMed]

15. Lou, B.; Doken, S.; Zhuang, T.; Wingerter, D.; Gidwani, M.; Mistry, N.; Ladic, L.; Kamen, A.; Abazeed, M.E. An image-based deep learning framework for individualizing radiotherapy dose. *Lancet Digit. Health* **2019**, *1*, e136–e147. [CrossRef]

16. Kato-Kogoe, N.; Sakaguchi, S.; Kamiya, K.; Omori, M.; Gu, Y.-H.; Ito, Y.; Nakamura, S.; Nakano, T.; Tamaki, J.; Ueno, T.; et al. Characterization of Salivary Microbiota in Patients with Atherosclerotic Cardiovascular Disease: A Case-Control Study. *J. Atheroscler. Thromb.* **2022**, *29*, 403–421. [CrossRef] [PubMed]

17. Cammarota, G.; Ianiro, G.; Ahern, A.; Carbone, C.; Temko, A.; Claesson, M.J.; Gasbarrini, A.; Tortora, G. Gut microbiome, big data and machine learning to promote precision medicine for cancer. *Nat. Rev. Gastroenterol. Hepatol.* **2020**, *17*, 635–648. [CrossRef]

18. Kingsmore, K.M.; Grammer, A.C.; Lipsky, P.E. Drug repurposing to improve treatment of rheumatic autoimmune inflammatory diseases. *Nat. Rev. Rheumatol.* **2020**, *16*, 32–52. [CrossRef]

19. Eyre, S.; Orozco, G.; Worthington, J. The genetics revolution in rheumatology: Large scale genomic arrays and genetic mapping. *Nat. Rev. Rheumatol.* **2017**, *13*, 421–432. [CrossRef]

20. Venerito, V.; Angelini, O.; Cazzato, G.; Lopalco, G.; Maiorano, E.; Cimmino, A.; Iannone, F. A convolutional neural network with transfer learning for automatic discrimination between low and high-grade synovitis: A pilot study. *Intern. Emerg. Med.* **2021**, *16*, 1457–1465. [CrossRef]

21. Venerito, V.; Angelini, O.; Fornaro, M.; Cacciapaglia, F.; Lopalco, G.; Iannone, F. A Machine Learning Approach for Predicting Sustained Remission in Rheumatoid Arthritis Patients on Biologic Agents. *J. Clin. Rheumatol.* **2021**, *28*, e334–e339. [CrossRef]

22. Venerito, V.; Lopalco, G.; Abbruzzese, A.; Colella, S.; Morrone, M.; Tangaro, S.; Iannone, F. A Machine Learning Approach to Predict Remission in Patients With Psoriatic Arthritis on Treatment With Secukinumab. *Front. Immunol.* **2022**, *13*, 917939. [CrossRef]

23. Lepri, G.; Hughes, M.; Bruni, C.; Cerinic, M.M.; Randone, S.B. Recent advances steer the future of systemic sclerosis toward precision medicine. *Clin. Rheumatol.* **2020**, *39*, 1–4. [CrossRef]

24. Van Leeuwen, N.M.; Maurits, M.; Liem, S.; Ciaffi, J.; Marsan, N.A.; Ninaber, M.; Allaart, C.; van Dongen, H.G.; Goekoop, R.; Huizinga, T.; et al. New risk model is able to identify patients with a low risk of progression in systemic sclerosis. *RMD Open* **2021**, *7*, e001524. [CrossRef]

25. Franks, J.M.; Martyanov, V.; Cai, G.; Wang, Y.; Li, Z.; Wood, T.A.; Whitfield, M.L. A Machine Learning Classifier for Assigning Individual Patients With Systemic Sclerosis to Intrinsic Molecular Subsets. *Arthritis Rheumatol.* **2019**, *71*, 1701–1710. [CrossRef]

26. Xu, X.; Ramanujam, M.; Visvanathan, S.; Assassi, S.; Liu, Z.; Li, L. Transcriptional insights into pathogenesis of cutaneous systemic sclerosis using pathway driven meta-analysis assisted by machine learning methods. *PLoS ONE* **2020**, *15*, e0242863. [CrossRef]

27. Showalter, K.; Spiera, R.; Magro, C.; Agius, P.; Martyanov, V.; Franks, J.M.; Sharma, R.; Geiger, H.; Wood, T.A.; Zhang, Y.; et al. Machine learning integration of scleroderma histology and gene expression identifies fibroblast polarisation as a hallmark of clinical severity and improvement. *Ann. Rheum. Dis.* **2021**, *80*, 228–237. [CrossRef] [PubMed]

28. Murdaca, G.; Caprioli, S.; Tonacci, A.; Billeci, L.; Greco, M.; Negrini, S.; Cittadini, G.; Zentilin, P.; Ventura Spagnolo, E.; Gangemi, S. A Machine Learning Application to Predict Early Lung Involvement in Scleroderma: A Feasibility Evaluation. *Diagnostics* **2021**, *11*, 1880. [CrossRef] [PubMed]

29. Andrade, D.S.M.; Ribeiro, L.M.; Lopes, A.J.; Amaral, J.L.M.; Melo, P.L. Machine learning associated with respiratory oscillometry: A computer-aided diagnosis system for the detection of respiratory abnormalities in systemic sclerosis. *BioMed. Eng. OnLine* **2021**, *20*, 31. [CrossRef] [PubMed]

30. Chassagnon, G.; Vakalopoulou, M.; Régent, A.; Zacharaki, E.I.; Aviram, G.; Martin, C.; Marini, R.; Bus, N.; Jerjir, N.; Mekinian, A.; et al. Deep Learning–based Approach for Automated Assessment of Interstitial Lung Disease in Systemic Sclerosis on CT Images. *Radiol. Artif. Intell.* **2020**, *2*, e190006. [CrossRef]

31. Taroni, J.N.; Martyanov, V.; Mahoney, J.M.; Whitfield, M.L. A Functional Genomic Meta-Analysis of Clinical Trials in Systemic Sclerosis: Toward Precision Medicine and Combination Therapy. *J. Investig. Dermatol.* **2017**, *137*, 1033–1041. [CrossRef] [PubMed]

32. Ebata, S.; Oba, K.; Kashiwabara, K.; Ueda, K.; Uemura, Y.; Watadani, T.; Fukasawa, T.; Miura, S.; Yoshizaki-Ogawa, A.; Yoshihide, A.; et al. Predictors of Rituximab Effect on Modified Rodnan Skin Score in Systemic Sclerosis: A machine learning analysis of the DESIRES trial. *Rheumatology* **2022**, keac023. [CrossRef] [PubMed]

33. Zamanian, R.T.; Badesch, D.; Chung, L.; Domsic, R.T.; Medsger, T.; Pinckney, A.; Keyes-Elstein, L.; D'Aveta, C.; Spychala, M.; White, R.J.; et al. Safety and Efficacy of B-Cell Depletion with Rituximab for the Treatment of Systemic Sclerosis-associated Pulmonary Arterial Hypertension: A Multicenter, Double-Blind, Randomized, Placebo-controlled Trial. *Am. J. Respir. Crit. Care Med.* **2021**, *204*, 209–221. [CrossRef] [PubMed]

34. Franks, J.M.; Martyanov, V.; Wang, Y.; Wood, T.A.; Pinckney, A.; Crofford, L.J.; Keyes-Elstein, L.; Furst, D.E.; Goldmuntz, E.; Mayes, M.D.; et al. Machine learning predicts stem cell transplant response in severe scleroderma. *Ann. Rheum. Dis.* **2020**, *79*, 1608–1615. [CrossRef] [PubMed]

35. Schniering, J.; Maciukiewicz, M.; Gabrys, H.S.; Brunner, M.; Blüthgen, C.; Meier, C.; Braga-Lagache, S.; Uldry, A.-C.; Heller, M.; Guckenberger, M.; et al. Computed tomography-based radiomics decodes prognostic and molecular differences in interstitial lung disease related to systemic sclerosis. *Eur. Respir. J.* **2022**, *59*, 2004503. [CrossRef] [PubMed]

36. Diab, N.; Hassoun, P.M. Pulmonary arterial hypertension: Screening challenges in systemic sclerosis and future directions. *Eur. Respir. J.* **2017**, *49*, 1700522. [CrossRef]

37. Milano, A.; Pendergrass, S.A.; Sargent, J.L.; George, L.K.; McCalmont, T.H.; Connolly, M.K.; Whitfield, M.L. Molecular Subsets in the Gene Expression Signatures of Scleroderma Skin. *PLoS ONE* **2008**, *3*, e2696. [CrossRef]

38. Pendergrass, S.A.; Lemaire, R.; Francis, I.P.; Mahoney, J.M.; Lafyatis, R.; Whitfield, M.L. Intrinsic gene expression subsets of diffuse cutaneous systemic sclerosis are stable in serial skin biopsies. *J. Investig. Dermatol.* **2012**, *132*, 1363–1373. [CrossRef]

39. Hinchcliff, M.; Huang, C.-C.; Wood, T.A.; Matthew Mahoney, J.; Martyanov, V.; Bhattacharyya, S.; Tamaki, Z.; Lee, J.; Carns, M.; Podlusky, S.; et al. Molecular signatures in skin associated with clinical improvement during mycophenolate treatment in systemic sclerosis. *J. Investig. Dermatol.* **2013**, *133*, 1979–1989. [CrossRef]

40. Mahoney, J.M.; Taroni, J.; Martyanov, V.; Wood, T.A.; Greene, C.S.; Pioli, P.A.; Hinchcliff, M.E.; Whitfield, M.L. Systems Level Analysis of Systemic Sclerosis Shows a Network of Immune and Profibrotic Pathways Connected with Genetic Polymorphisms. *PLOS Comput. Biol.* **2015**, *11*, e1004005. [CrossRef]

41. Bruni, C.; Guignabert, C.; Manetti, M.; Cerinic, M.M.; Humbert, M. The multifaceted problem of pulmonary arterial hypertension in systemic sclerosis. *Lancet Rheumatol.* **2021**, *3*, e149–e159. [CrossRef]

42. Raja, J.; Denton, C.P. Cytokines in the immunopathology of systemic sclerosis. *Semin. Immunopathol.* **2015**, *37*, 543–557. [CrossRef]

43. Hoffmann-Vold, A.-M.; Maher, T.M.; Philpot, E.E.; Ashrafzadeh, A.; Barake, R.; Barsotti, S.; Bruni, C.; Carducci, P.; Carreira, P.E.; Castellví, I.; et al. The identification and management of interstitial lung disease in systemic sclerosis: Evidence-based European consensus statements. *Lancet Rheumatol.* **2020**, *2*, e71–e83. [CrossRef]

44. Wallace, B.; Vummidi, D.; Khanna, D. Management of connective tissue diseases associated interstitial lung disease: A review of the published literature. *Curr. Opin. Rheumatol.* **2016**, *28*, 236–245. [CrossRef] [PubMed]

45. MacIntyre, N.R. The future of pulmonary function testing. *Respir. Care* **2012**, *57*, 154–161. [CrossRef]

46. Goh, N.S.L.; Desai, S.R.; Veeraraghavan, S.; Hansell, D.M.; Copley, S.J.; Maher, T.M.; Corte, T.J.; Sander, C.R.; Ratoff, J.; Devaraj, A.; et al. Interstitial lung disease in systemic sclerosis: A simple staging system. *Am. J. Respir. Crit. Care Med.* **2008**, *177*, 1248–1254. [CrossRef]

47. Moore, O.A.; Goh, N.; Corte, T.; Rouse, H.; Hennessy, O.; Thakkar, V.; Byron, J.; Sahhar, J.; Roddy, J.; Gabbay, E.; et al. Extent of disease on high-resolution computed tomography lung is a predictor of decline and mortality in systemic sclerosis-related interstitial lung disease. *Rheumatology* **2013**, *52*, 155–160. [CrossRef]

48. Collins, C.D.; Wells, A.U.; Hansell, D.M.; Morgan, R.A.; MacSweeney, J.E.; Du Bois, R.M.; Rubens, M.B. Observer variation in pattern type and extent of disease in fibrosing alveolitis on thin section computed tomography and chest radiography. *Clin. Radiol.* **1994**, *49*, 236–240. [CrossRef]

49. Sverzellati, N.; Devaraj, A.; Desai, S.R.; Quigley, M.; Wells, A.U.; Hansell, D.M. Method for minimizing observer variation for the quantitation of high-resolution computed tomographic signs of lung disease. *J. Comput. Assist. Tomogr.* **2011**, *35*, 596–601. [CrossRef]

50. van Timmeren, J.E.; Cester, D.; Tanadini-Lang, S.; Alkadhi, H.; Baessler, B. Radiomics in medical imaging—"How-to" guide and critical reflection. *Insights Imaging* **2020**, *11*, 91. [CrossRef]

51. Hao, Y.; Hudson, M.; Baron, M.; Carreira, P.; Stevens, W.; Rabusa, C.; Tatibouet, S.; Carmona, L.; Joven, B.E.; Huq, M.; et al. Early Mortality in a Multinational Systemic Sclerosis Inception Cohort. *Arthritis Rheumatol.* **2017**, *69*, 1067–1077. [CrossRef] [PubMed]

52. Mukerjee, D.; St George, D.; Coleiro, B.; Knight, C.; Denton, C.P.; Davar, J.; Black, C.M.; Coghlan, J.G. Prevalence and outcome in systemic sclerosis associated pulmonary arterial hypertension: Application of a registry approach. *Ann. Rheum. Dis.* **2003**, *62*, 1088–1093. [CrossRef] [PubMed]

53. Hachulla, E.; Carpentier, P.; Gressin, V.; Diot, E.; Allanore, Y.; Sibilia, J.; Launay, D.; Mouthon, L.; Jego, P.; Cabane, J.; et al. Risk factors for death and the 3-year survival of patients with systemic sclerosis: The French ItinérAIR-Sclérodermie study. *Rheumatology* **2009**, *48*, 304–308. [CrossRef]

54. Galiè, N.; Hoeper, M.M.; Humbert, M.; Torbicki, A.; Vachiery, J.-L.; Barbera, J.A.; Beghetti, M.; Corris, P.; Gaine, S.; Gibbs, J.S.; et al. Guidelines for the diagnosis and treatment of pulmonary hypertension: The Task Force for the Diagnosis and Treatment of Pulmonary Hypertension of the European Society of Cardiology (ESC) and the European Respiratory Society (ERS), endorsed by the International Society of Heart and Lung Transplantation (ISHLT). *Eur. Heart J.* **2009**, *30*, 2493–2537. [CrossRef] [PubMed]

55. Bruni, C.; De Luca, G.; Lazzaroni, M.-G.; Zanatta, E.; Lepri, G.; Airò, P.; Dagna, L.; Doria, A.; Matucci-Cerinic, M. Screening for pulmonary arterial hypertension in systemic sclerosis: A systematic literature review. *Eur. J. Intern. Med.* **2020**, *78*, 17–25. [CrossRef]

56. Bauer, Y.; de Bernard, S.; Hickey, P.; Ballard, K.; Cruz, J.; Cornelisse, P.; Chadha-Boreham, H.; Distler, O.; Rosenberg, D.; Doelberg, M.; et al. Identifying early pulmonary arterial hypertension biomarkers in systemic sclerosis: Machine learning on proteomics from the DETECT cohort. *Eur. Respir. J.* **2021**, *57*, 2002591. [CrossRef] [PubMed]

57. Connolly, M.K. Systemic sclerosis (scleroderma): Remaining challenges. *Ann. Transl. Med.* **2021**, *9*, 438. [CrossRef] [PubMed]

58. Stoel, B. Use of artificial intelligence in imaging in rheumatology—Current status and future perspectives. *RMD Open* **2020**, *6*, e001063. [CrossRef]

59. Bellando-Randone, S.; Russo, E.; Venerito, V.; Matucci-Cerinic, M.; Iannone, F.; Tangaro, S.; Amedei, A. Exploring the Oral Microbiome in Rheumatic Diseases, State of Art and Future Prospective in Personalized Medicine with an AI Approach. *J. Pers. Med.* **2021**, *11*, 625. [CrossRef]

60. Matuchansky, C. Deep medicine, artificial intelligence, and the practising clinician. *Lancet* **2019**, *394*, 736. [CrossRef]