International Journal of
Molecular Sciences

MDPI

Review

Application of Machine Learning Models in Systemic Lupus Erythematosus

Fulvia Ceccarelli *, Francesco Natalucci, Licia Picciariello, Claudia Ciancarella, Giulio Dolcini, Angelica Gattamelata, Cristiano Alessandri and Fabrizio Conti

Lupus Clinic, Rheumatology, Dipartimento di Scienze Cliniche Internistiche Anestesiologiche e Cardiovascolari, Sapienza Università di Roma, Viale del Policlinico 155, 00161 Rome, Italy
* Correspondence: fulviaceccarelli@gmail.com; Tel.: +39-06-4997-4631

Abstract: Systemic Lupus Erythematosus (SLE) is a systemic autoimmune disease and is extremely heterogeneous in terms of immunological features and clinical manifestations. This complexity could result in a delay in the diagnosis and treatment introduction, with impacts on long-term outcomes. In this view, the application of innovative tools, such as machine learning models (MLMs), could be useful. Thus, the purpose of the present review is to provide the reader with information about the possible application of artificial intelligence in SLE patients from a medical perspective. To summarize, several studies have applied MLMs in large cohorts in different disease-related fields. In particular, the majority of studies focused on diagnosis and pathogenesis, disease-related manifestations, in particular Lupus Nephritis, outcomes and treatment. Nonetheless, some studies focused on peculiar features, such as pregnancy and quality of life. The review of published data demonstrated the proposal of several models with good performance, suggesting the possible application of MLMs in the SLE scenario.

Keywords: Systemic Lupus Erythematosus; artificial intelligence; machine learning models

check for updates

Citation: Ceccarelli, F.; Natalucci, F.; Picciariello, L.; Ciancarella, C.; Dolcini, G.; Gattamelata, A.; Alessandri, C.; Conti, F. Application of Machine Learning Models in Systemic Lupus Erythematosus. *Int. J. Mol. Sci.* **2023**, *24*, 4514. https://doi.org/10.3390/ijms24054514

Academic Editor: Chi Chiu Mok

Received: 14 January 2023
Revised: 14 February 2023
Accepted: 22 February 2023
Published: 24 February 2023

Copyright: © 2023 by the authors. Licensee MDPI, Basel, Switzerland. This article is an open access article distributed under the terms and conditions of the Creative Commons Attribution (CC BY) license (https://creativecommons.org/licenses/by/4.0/).

1. Introduction

Systemic Lupus Erythematosus (SLE) is a systemic autoimmune disease, extremely heterogeneous in terms of immunological features and clinical manifestations (Figure 1A). Thus, this condition could potentially involve any organ and system, leading to different severity degrees and outcomes. Traditionally, it is possible to distinguish more severe disease, including renal and neurological manifestations, from mild/moderate disease, characterized by other manifestations such as skin and joint involvement [1]. This clinical complexity could result in a diagnostic delay, especially evident when the disease begins with rarer manifestations. Data from the literature reported an interval between the appearance of first symptom and diagnosis of about 70 months. Of note, it must be underlined that diagnostic delay, even when it amounts to only 6 months, could lead to severe organ involvement, high flare rates and chronic damage development. This derives certainly from the later introduction of appropriate treatment, together with prolonged treatment of glucocorticoids, widely recognized as the most relevant risk factors for chronic damage progression [2,3].

In this view, the purpose of the classification criteria is to facilitate and to anticipate SLE diagnosis and to allow the identification of homogeneous populations for clinical studies. The classification criteria proposed until now have been summarized in Figure 1B [4–6]. The latest EULAR/ACR criteria published in 2019 introduced important innovations. First of all, the presence of an entry criterion, represented by the ANA positivity, is necessary to apply these criteria, underlining the autoimmune pathogenesis of SLE. In addition, a weighted score system has been introduced, with different scores for different clinical and laboratory features. Accordingly, only patients reaching a score higher than 10 could be classified as affected by SLE. The application of these criteria leads to a sensitivity and

specificity of 98.0% and 96.4%, respectively in the derivation cohort, and of 96.1% and 93.4%, respectively in the validation cohort [6].

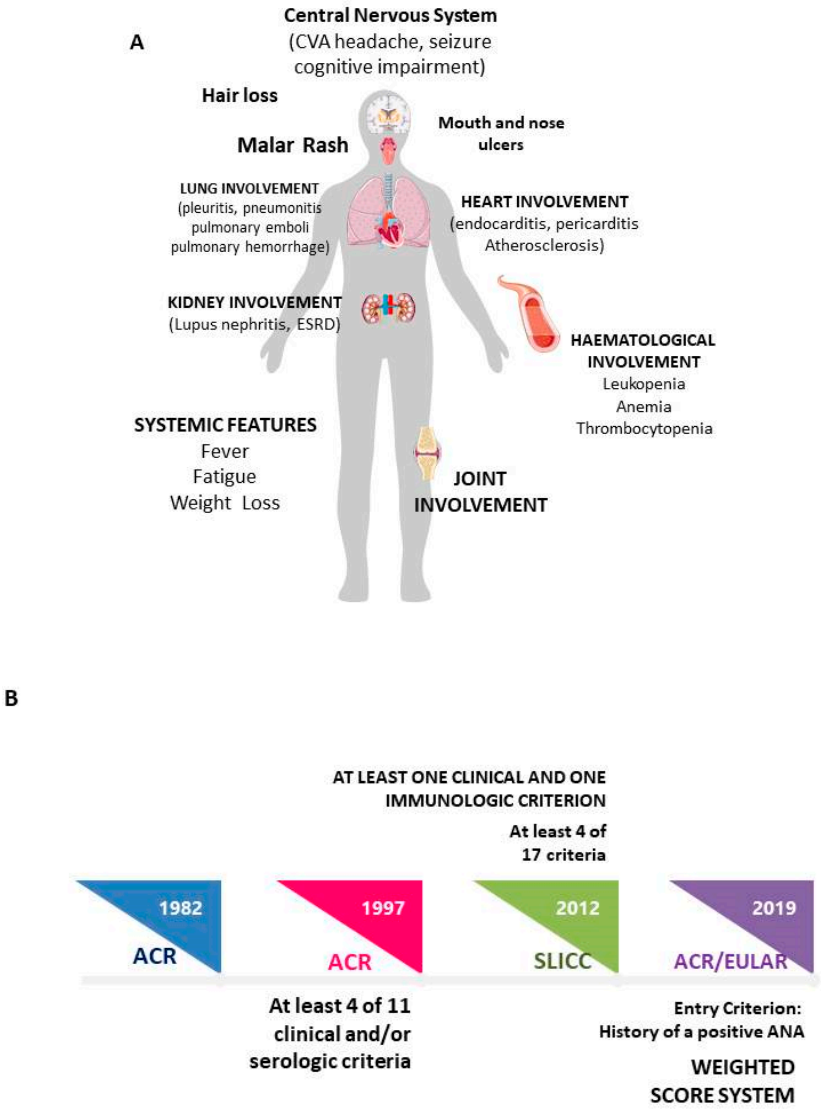


Figure 1. (**A**) Systemic Lupus Erythematosus: main clinical manifestations; (**B**) Classification criteria timeline.

From a pathogenic point of view, a multifactorial etiology has been widely demonstrated for SLE. More than one hundred genetic variants have been associated with disease susceptibility and phenotype. Then, the interplay between genetic background and different environmental factors leads to the activation of an aberrant autoimmune response with the production of several autoantibodies [1,7]. The production of autoantibodies has been described several years before the appearance of clinical manifestations, suggesting a stage of subclinical autoimmunity preceding the disease development [8,9].

SLE is traditionally characterized by a relapsing-remitting course, with the occurrence of disease flare. This evolution could result in the development of irreversible chronic damage, determined by disease activity itself and by the adverse events of treatment, in particular glucocorticoids [10,11]. The application of more appropriate therapeutic approaches, in particular the so-called treat-to-target, could significantly impact the course of the disease. In fact, a better control of disease activity, by reaching remission or a low disease activity state, could determine the reduction of chronic damage progression, with improvement in long-term outcome and survival [12]. In 2019, the latest recommendations have been

published, based on a comprehensive management of SLE patients. In fact, it is not only the need to treat the disease itself that has been underlined, but also comorbidities, and the need to educate patients to an appropriate lifestyle, with emphasis on sun protection, vaccination, exercise, and smoking cessation [13]. Concerning the pharmacological approach, the recommendations distinguished mild, moderate and severe manifestations to prescribe more appropriate treatment. Of note, the possibility to use a biological treatment, in particular belimumab, was introduced for the first time in the routine care of SLE patients [13]. In the view of disease complexity, several unmet needs are still present for the diagnosis and the management of SLE patients, suggesting the application of innovative tools, such as machine learning models (MLMs). Thus, the purpose of the present review is to provide the reader with information about the possible application of artificial intelligence (AI) in patients with SLE from a medical perspective.

Methods

A literature search was done in PubMed, accessed via the National Library of Medicine PubMed interface (http://www.ncbi.nlm.nih.gov/pubmed, accessed on 1 December 2022). Firstly, PubMed was searched using the term "systemic lupus erythematosus" OR "lupus" in combination with (AND) "machine learning models". Secondly, the same PubMed search was combined with other terms, such as "artificial intelligence" OR "classification" OR "clustering" OR "regression".

2. Machine Learning Models: General Concepts

In the last years, AI has generated increasing interest in the field of medical conditions, including rheumatic diseases. In particular MLMs, a subcategory of AI, have been widely applied for different purposes, such as diagnosis, identification of disease phenotypes, prognosis and precision medicine [14].

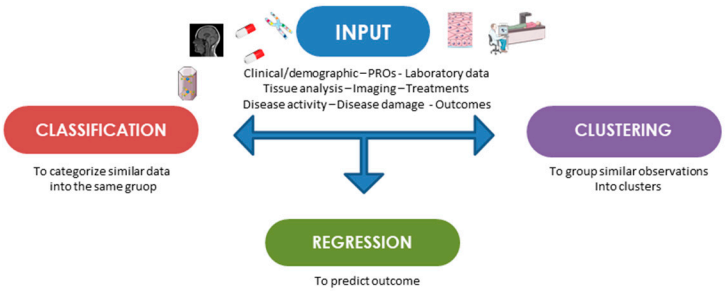
Differently from the statistical method, MLMs extract knowledge from input data. Indeed, if the statistical models aim at explaining specified or hypothesis-driven relationships, MLMs work to search underlying data connections and make decisions according to the newly discovered associations. Thus, MLMs extrapolate relationships unidentifiable with other statistical techniques that are more suitable to generate new hypotheses [15].

The ideal application of MLMs involves the use of so-called big data, deriving from electronic health records, imaging tools, genetics, and transcriptomic procedures. This could be very interesting in the evaluation of complex chronic conditions, such as rheumatic diseases, characterized by great heterogeneity in clinical and laboratory features, by overtime evolution and by the contribution of multiple factors in disease susceptibility and course [15]. Thus, MLMs could help in predicting disease outcome e treatment response, a challenge in diseases such as SLE, characterized by alternating clinical course and various severity degrees requiring different treatments.

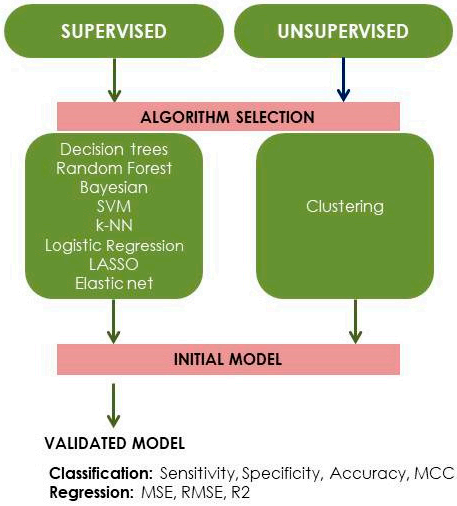
The aim of MLMs is the generation of a predictive model, potentially relevant in routine care for the following outcomes: classification, regression or clustering [16,17]. As reported in Figure 2A, different types of data could be used as input to create the MLMs. In particular, it is possible to use clinical/demographic information, laboratory data, results from patient-reported outcomes, data from tissues analysis or imaging tools, response to different treatments, and information about disease activity course or chronic damage development [14,16,17].

The use of medical data in MLMs frequently requires a process of adaptation, in particular they should be translated into a numerical format that could be processed by the AI. Furthermore, it should also be considered the possibility of missing data. Sometimes, a scaling process should be applied to transform existing features into a smaller set of variables [15,18]. Moreover, MLMs perform better when the number of input variables is optimized. Indeed, features selection is a dimensionality technique of reduction that is applied to identify the most appropriate variables to use as input into MLMs algorithm, as all measured variables might not provide information that is necessary for outcome

prediction. The features selection could be made by using different modalities, including filter, wrapper and embedded methods [19].



(**A**)



(**B**)

Figure 2. (**A**) Summary of input data used in machine learning models for medical studies, aiming at evaluating the following outcomes: classification, regression or clustering. (**B**) Summary of the different MLMs that could be applied in the supervised and unsupervised modalities.

In detail, different types of MLMs are available: to summarize, supervised and unsupervised algorithms can be differentiated according to the labeling of used variables. Thus, a supervised model is constructed to predict known values, whereas an unsupervised model works to predict unknown variables [20,21]. Figure 2B summarizes the different MLMs that could be applied in the supervised and unsupervised modalities [15,20,21]. Of note, the performance of supervised models could be improved by using two independent datasets: the training and the validation dataset. Furthermore, the performance could be assessed by applying different metrics, such as accuracy (the ratio of correct predictions to total predictions), sensitivity (the true positive rate) and specificity (the true negative rate). If the classification problem is binary, these values are often represented by using receiver operating characteristic (ROC) curves. Thus, the area under the ROC curve (AUC) represents the probability that the model can distinguish correct and incorrect outcomes. An AUC value of 1.0 indicates a perfect model performance, whereas a score of 0.5 indicates that the model's performance is comparable to random chance [22]. For regression analysis, other parameters could be used to assess the model performance: in particular, mean squared error, root mean squared error and the coefficient of determination [23].

Given the availability of different algorithms, it is essential to select the most appropriate MLMs according to the goal (classification, regression, clustering or reduction of dimensionality). Furthermore, the MLMs selection should be based on the available input data, and the comparison between multiple algorithms is recommended in order to identify the model with the greatest performance [24,25].

3. Machine Learning Models in SLE Cohorts

In the last years several studies have applied MLMs in SLE cohorts in different disease-related fields. In particular, the majority of studies focused on diagnosis and pathogenesis, disease-related manifestations, outcomes and treatment.

3.1. Diagnosis

Table 1 summarizes data about the studies applying MLMs for diagnostic purposes [26–41]. Overall, it is possible to identify three fields of application according to the data input considered in the studies. First of all, moving from the role exerted by the genetic background in disease development, more recent studies applied MLMs in this context [27–30]. Therefore, MLMs could be applied to select candidate genes able to identify SLE patients, suggesting the possibility to use these inputs as diagnostic biomarkers. Furthermore, other laboratory features have been considered as input for AI models, such as proteomic data deriving from serum, plasma or peripheral blood mononuclear cells (PBMCs) of SLE patients [26,41]. Already in 2009 Huang and colleagues proposed a Decision Trees model, including a panel of four proteins that resulted able to recognize SLE patients [41]. More recently, Li and colleagues, by using Random Forest model, demonstrated a good performance for a six-protein combination model (SLE versus healthy controls, AUC = 0.7; SLE versus rheumatoid arthritis, AUC = 0.815). The AUC increased up to 0.990 when considering the ability of a nine-protein combination in discriminating SLE patients with disease flare from patients with stable disease [26].

Table 1. Main data about the studies applying Machine Learning Models for diagnostic purposes.

Study	MLM	Input Data	Results
Li, 2022 [26]	Random Forest	PBMC proteomics	Six-protein combination (IFIT3, MX1, TOMM40, STAT1, STAT2, and OAS3) exhibited good performance for SLE disease diagnosis (AUC= 0.723 versus HC; AUC= 0.815 versus RA). Nine-protein combination (PHACTR2, GOT2, L-selectin, CMC4, MAP2K1, CMPK2, ECPAS, SRA1, and STAT2) showed a robust performance in assessing disease exacerbation (AUC = 0.990)
Zhong, 2022 [27]	LASSO Support Vector Machines	Differentially expressed genes (DEGs)	Selection of six candidate diagnostic biomarkers for SLE (ABCB1, EIF2AK2, HERC6, ID3, IFI27, and PLSCR1), with AUC from 0.96 to 0.913
Jiang, 2022 [28]	Logistic regression Random Forest XGBoost Support Vector Machines Artificial Neural Network	Genetic biomarkers from GSE65391 and GSE72509 datasets	IFI44 was determined to be the optimal diagnostic biomarker of SLE
Ma, 2022 [29]	Random Forest Support Vector Machines Artificial Neural Network	Genome-wide association studies	RF model AUC = 0.84 At the optimal cut-off, the RF predictor reached a sensitivity of 84% with a specificity of 68% in SLE classification.
Martorell-Marugán, 2022 [30]	XGBoost model	Biomarkers for gene expression and DNA methylation	The model is able to discriminate SLE from Sjogren Syndrome (gene expression MCC = 0.5791 ± 0.0409; methylation data MCC = 0.5546 ± 0.0484)

Table 1. *Cont.*

Study	MLM	Input Data	Results
Barnado, 2022 [31]	Random Forest XGBoost model	Electonic Health Data	PPV 74–77%
Matthiesen, 2021 [32]	Partial Least Square	Plasma lipidomes	SLE vs CVD (Sensitivity = 0.91, Specificity = 1) IS vs SLE (Sensitivity = 1, Specificity = 0.82)
Adamichou, 2020 [33]	LASSO Logistic Regression	Clinical/laboratory features according classification criteria EULAR/ACR	Accuracy = 94.8% for identifying SLE High sensitivity for early disease (93.8%), LN (97.9%), NPSLE (91.8%), SLE requiring immunosuppressives/biologics (96.4%). Development of a scoring system (>7, 94.2%) accuracy
Ceccarelli, 2021 [34]	ReliefF algorithm, Logistic Regression Support Vector Machines DT models	Clinical/laboratory features according classification criteria EULAR/ACR	At the ReliefF model, anti-dsDNA positivity, low C3/C4 serum levels and malar/maculopapular rash resulted the strongest predictor features. A good model's performance was obtained already when only the three highest scoring features were considered (AUC = 0.94)
Park, 2020 [35]	Cluster Analysis	Serum cyotkines	Cluster analysis revealed two distinct patient groups characterized by high levels of IL8, MIP1α and MIP1β (group 1) or of IL2, IL6, IL10, IL12, IFNγ and TNF (group 2). Active disease was more common in group 1 (55.7%) than in group 2 (34.8%). More patients in group 2 had renal involvement (42/115, 36.5%) than in group 1 (22/88, 25%).
Guthridge, 2020 [36]	K-means clustering Random Forest	Data from plasma, serum, RNA, clinical, laboratory features	Identification of 7 SLE clusters. Inflammation and interferon modules were elevated in Clusters 1 (moderately) and 4 (strongly), with decreased T-cell modules in Cluster 4. Active clinical features were similar across clusters. Clinical SLEDAI trended highest in Clusters 3 and 4, though Cluster 3 lacked strong interferon and inflammation signatures. Renal activity was more frequent in Cluster 4, and rare in Clusters 2, 5, and 7. Serology findings were lowest in Clusters 2 and 5.
Jorge, 2019 [37]	Logistic Regression	Registration data according classification criteria ACR/SLICC	PPV 90% for definite SLE PPV 92% for definite/probable SLE
Murray, 2019 [38]	Logistic Regression	Electronic health record	AUC=0.97 to automate identification of SLE patients
Turner, 2017 [39]	Artificial Neural Network Random Forest Naïve Bayes model Support Vector Machines Word2Vec	Electronic health record	ICD-9 accuracy 90.00% (AUC = 0.9) Shallow neural network with CUIs accuracy 92.10% (AUC = 0.970) Random forest with BOWs accuracy 95.25% (AUC = 0.994) Random forest with CUIs accuracy 95.00% (AUC = 0.979) Word2Vec inversion accuracy 90.03% (AUC = 0.905)
Dai, 2010 [40]	k-nearest neighbors	Serum peptidome patterns	Blinded verification of the classification model showed 91.7% sensitivity in active SLE, 83.3% sensitivity in stable SLE, and 86.7% specificity in normal controls.
Huang, 2009 [41]	Decision Trees	Serum proteomic	A panel of four potential protein biomarkers could accurately recognize 25 of 32 patients with SLE, 36 of 42 patients with other autoimmune diseases and 36 of 40 healthy people.

SLE: Systemic Lupus Erythematosus; RA: Rheumatoid Arthritis; AUC: Area Under Curve; RF: Random Forest; MCC: Matthews correlation coefficient; PPV: positive predictive value; CVD: cardiovascular disease; IS: ischemic stroke; LN: Lupus Nephritis; NPSLE: neuropsychiatric SLE.

In 2021, Matthiensen and colleagues for the first time applied MLMs to assess the diagnostic role of plasma lipidome, showing good sensitivity and specificity in distinguishing SLE from patients with cardiovascular disease and ischemic stroke [32].

In the remaining studies, the ability of MLMs for a diagnostic purpose has been tested by using electronic health data (EHD) or clinical/laboratory disease-related features, frequently as defined by the available classification criteria. Overall, the use of EHD as

input demonstrated a good performance of MLMs in identifying SLE patients in terms of AUC values (up to 0.97) [38,39].

The study published by Adamichou and colleagues in 2020 aimed at assessing the accuracy of 2019 EULAR/ACR criteria in SLE diagnosis by using a LASSO-LR model (ref). The inclusion as input of all the features included in the three classification criteria sets (ACR 1997, SLICC 2012, ACR/EULAR 2019) allowed to observe an accuracy for the most recent criteria of 94.8% in identifying SLE patients. In detail, a higher sensitivity was demonstrated for subjects with an early disease, for patients with Lupus Nephritis (LN) and neuropsychiatric SLE (NPSLE), and for patients treated by immunosuppressant drugs or biological agents. Furthermore, the authors were able to develop a predictive score (the so-called SLERPI score): for a score higher than seven, an accuracy of 94.2% was observed [33]. Our group employed different MLMs—in particular the ReliefF algorithm, Logistic Regression, nonlinear Support Vector Machines, and Decision Trees models—to identify the stronger predictors for SLE diagnosis. By enrolling SLE patients and control subjects with miscellaneous rheumatic diseases, relevant to the differential diagnosis, we obtained a good model's performance, already when only the three highest scoring features were considered (AUC = 0.94). Furthermore, anti-dsDNA positivity, low C3/C4 serum levels and malar/maculopapular rash resulted in the strongest predictor features for classifying a patient as having SLE [34].

Moreover, the application of cluster analysis could be used to identify subsets of patients by integrating clinical features, immunological profiles and molecular pathways. In this context, the study conducted by Guthridge and colleagues in 2020 used different parameters as input, by combining data from plasma, serum and RNA evaluation with clinical and immunological features. Indeed, the application of a cluster analysis allowed to identify different disease clusters in terms of molecular profile, such as expression of interferon, and disease activity, as assessed by SLEDAI-2k [36]. Similarly, Diaz-Gallo and colleagues in 2022 applied an unsupervised cluster analysis by identifying four SLE subgroups, different in terms of the autoantibody profile, HLA-DRB1 alleles, immunological and clinical features [42].

The same model allowed to differentiate SLE patients according to lymphocyte subsets. Indeed, the study conducted by Lu and colleagues identified four clusters (B high, CD4 high, CD8 high and NK high). These clusters differed in terms of clinical manifestations: in fact, the incidence of arthritis was significantly higher in B high cluster, while nephritis was more frequent in CD8 high and NK high clusters. Finally, CD4 high cluster showed SLEDAI-2k values significantly lower compared with the remaining three clusters [43]. In this view, cluster analysis could also differentiate SLE patients according to cytokine profile: as demonstrated by Reynold and colleagues, it is possible to identify three distinct groups of patients, characterized by higher levels of interferon-alpha and B lymphocyte stimulator (group 1), increased CXCL10 and CXCL13 (group 2) or low levels of cytokines (group 3). Furthermore, group 2 had significantly lower serum complement and higher anti-dsDNA antibodies with increased prevalence of arthritis [44].

3.2. Disease Features

The majority of the available studies focused on the application of MLMs on SLE cohorts with renal involvement, representing one of the most fearful disease-related manifestations, with possible progress into end-stage renal disease in 20% of patients and then requiring more aggressive treatment [45]. Table 2 summarizes data about these studies [46–52], the first of which was published in 2011 and focused on the probability of 3-year allograft survival after renal transplantation [46]. By considering different input data, such as previous and current treatments and data about transplantation and comorbidities, the authors applied different models, obtaining a good performance in terms of AUC (up to 0.74 when considering the logistic regression model) [46]. Only one recent study included, as input, gene expression datasets, downloaded from the GEO database. The application of LASSO and SVM-FRE models suggested the possible role as diagnostic biomarkers for the following genetic variants: C1QA (AUC = 0.741), C1QB (AUC = 0.758), MX1 (AUC = 0.865), RORC (AUC = 0.911), CD177 (AUC = 0.855), DEFA4 (AUC = 0.843), HERC5 (AUC = 0.880) [48]. In the remaining

studies on LN, the models used simultaneously clinical and demographic data, serum and urinary biomarkers, and histological features for diagnostic and predictive purposes [47,49–52]. Two studies published in 2022 demonstrated a good performance of MLMs in discriminating different histological classes. Indeed, Wang and colleagues proposed a model able to distinguish between ISN/RPS pure class V and classes III ± V or IV ± V, while Yang and colleagues observed good accuracy for mask R-CNN and LSTM models on recognizing different glomerular diseases based on slide images (AUC = 0.947) [48,52]. Furthermore, MLMs resulted able to predict a one-year response to treatment, the complete remission, or the risk of flare at 5 years follow-up [47,49,51].

Neurological involvement represents another complex SLE manifestation, with heterogeneous phenotype and lack of specific biomarkers. Thus, the differential diagnosis between SLE-related neurological symptoms and other confounder disorders is not always easy [53]. In this view, MLMs could facilitate clinicians in discriminating the real NPSLE from other pathological conditions. The main aspects of the studies published so far were summarized in Table 2 [54–57]. In detail, two studies applied MLMs in the analysis of the role of imaging techniques for diagnostic purposes. Thus, cluster analysis resulted able to discriminate five subsets of magnetic resonance characterized by the predominant involvement of different cerebral areas in terms of white matter hyperintensities distribution [55]. Moreover, the application of proton magnetic resonance spectroscopy was evaluated by SVM with feature selection. The authors proposed a diagnostic model with 94.9% of accuracy, which was able to identify patients with early NPSLE [56]. The study conducted by Barraclough and colleagues in 2022 focused on patients with cognitive impairment: the application of the network fusion model is able to discriminate patients with different performances in cognitive functions [57]. Gu and colleagues proposed a model by integrating the presence of anxiety and T-cells subsets evaluated by flow cytometry: the XGBoost model allowed to identify a significant difference in terms of T-cell subsets in patients with or without anxiety (AUC= 0.922) [54].

Two studies conducted by our research group focused on the application of MLMs in SLE-related joint involvement, one of the most frequent manifestation, potentially involving up to 90% of patients [58]. The first study published in 2018 applied logistic regression with the Forward Wrapper method in a cohort of patients with joint involvement evaluated from a laboratory and ultrasonographic point of view. We obtained a model with a good performance in identifying SLE patients with erosive arthritis (AUC = 0.806). Furthermore, at the feature selection, anti-carbamylated proteins antibodies (anti-CarP) resulted the most relevant factors for the presence of erosive arthritis [59]. Thus, an unsupervised hierarchical cluster analysis was applied to identify the aggregation of patients with and without erosive arthritis into different subgroups sharing common characteristics in terms of clinical and laboratory phenotypes. Our results demonstrated the identification of four main clusters: in particular, erosive arthritis was located in a cluster including renal and neuropsychiatric involvement, serositis, positivity for anti-CarP, anti-citrullinated protein antibodies, anti-Sm, anti-RNP, detectable levels of Dkk1 [60]. This could suggest the presence of a more aggressive disease phenotype, sharing a common pathogenic background [61].

Furthermore, MLMs have been also applied in the field of SLE comorbidity. In detail, Liu and colleagues in 2022 used AI to identify potential biomarkers for SLE patients with atherosclerosis (AS). By applying LASSO, SVM-RFE, and RF models, the authors identified five hub genes (specifically, SPI1, MMP9, C1QA, CX3CR1, and MNDA) with a high predictive performance in distinguishing subjects with and without AS (AUC ranging from 0.900 to 0.981) [62]. Wang and colleagues aimed at identifying the shared genes between SLE and metabolic syndrome (MetS): RF and LASSO algorithms were used to screen shared hub genes, and a diagnostic model was created by applying XG-Boost. Finally, the authors identified shared hub genes and constructed an effective diagnostic model in SLE and MetS. In detail, TNFSF13B and OAS1 had a positive correlation with cholesterol and xenobiotic metabolism. Both biomarkers and metabolic pathways were potentially linked to monocytes, providing novel insights into the disease pathogenesis [63].

Table 2. Data about studies applying Machine Learning Models in different disease-related features, in particular Lupus Nephritis, Neuropsychiatric SLE, joint involvement.

Study	Disease Feature	MLM	Input Data	Results
Tang, 2011 [46]	Lupus Nephritis	Classification trees Logistic Regression Artificial Neural Network	Demographic, clinical, laboratory data; treatment, data about transplantation, comorbidity	Model to predict the probability of 3-year allograft survival after renal transplantation. LR, AUC = 0.74 Classification trees, AUC = 0.70 95% CI: 0.67–0.72) ANN, AUC = 0.71
Chen, 2021 [47]	Lupus Nephritis	XGBoost SR-SPM	Clinical, laboratory and histological data	Development of a model to evaluate the risk of renal flare 5 years after remission. Good performance (XGBoost, C-index = 0.819) (SR-SPM, C-index = 0.746)
Wang, 2022 [48]	Lupus Nephritis	LASSO Support Vector Machine	LN gene expression datasets downloaded from the GEO database	Possible role as diagnostic biomarkers for C1QA (AUC = 0.741), C1QB (AUC = 0.758), MX1 (AUC = 0.865), RORC (AUC = 0.911), CD177 (AUC = 0.855), DEFA4 (AUC = 0.843), HERC5 (AUC = 0.880)
Stojanowski, 2022 [49]	Lupus Nephritis	Multi-layer perceptron	Demographic and laboratory features	Development of a predictive models for complete remission, (accuracy = 91.67%, AUC 0.9375)
Wang, 2022 [50]	Lupus Nephritis	HMFO Support Vector Machine	Demographic and laboratory features	Development of a model distinguishing between ISN/RPS pure class V and classes III $\pm$ V or IV $\pm$ V
Ayoub, 2022 [51]	Lupus Nephritis	Logistic Regression Random Forest Support Vector Machine Artificial Neural Network	Clinical data, urine biomarkers	Development of a model to predict 1-year treatment response (AUC = 0.7)
Yang, 2022 [52]	Lupus Nephritis	Mask R-CNN LSTM	Human kidney biopsy samples	Good accuracies (up to 0.940) on recognizing different glomerular diseases based on H&E whole slide images (AUC = 0.947)
Gu, 2021 [54]	NPSLE	LASSO Random Forest XGBoost	Clinical data, flow cytometry data on T-cell subsets, Self-Rating Anxiety/Depression Scale and Beck Depression Inventory	Identification of difference in T-cell subsets in SLE patients with or without anxiety Best performer XGBoost (AUC = 0.922)
Rumetshofer, 2022 [55]	NPSLE	Cluster analysis	White matter hyperintensities on MRI	Identification of five distinct clusters with predominant involvement of different areas.
Tan, 2022 [56]	NPSLE	Support Vector Machine Feature selection	Proton magnetic resonance spectroscopy	Diagnostic model with 94.9% accuracy, 91.3% sensitivity, 100% specificity and 0.87 cross-validation score
Barraclough, 2022 [57]	NPSLE	Network fusion	Cognitive assessment using the ACR Neuropsychological Battery (ACR-NB)	Identification of two subtypes with different performance in cognitive function ($p < 0.03$)
Ceccarelli, 2018 [59]	Joint involvement	Logistic Regression Forward Wrapper method Feature selection	Clinical and laboratory data, Ultrasound assessment	Good performance to identify patients with erosive arthritis (AUC = 0.806).
Ceccarelli, 2022 [60]	Joint involvement	Cluster analysis	Clinical and laboratory data, Ultrasound assessment	Identification of four clusters. Erosive arthritis was located in a cluster including renal and NPSLE.

LR: Logistic Regression; AUC: Area Under Curve; ANN: Artificial Neural Network; ISN/RPS: International Society of Nephrology (ISN)/Renal Pathology Society (RPS).

3.3. Disease Activity and Damage

The main outcome in the management of SLE patients is certainly the control of disease activity in order to prevent chronic damage development. The longitudinal assessment of disease activity allowed to identify different patterns: the so-called relapsing-remitting pattern has been prevalently associated with damage progression, due to the need to use glucocorticoids to treat disease relapse [12]. Several efforts have been made to develop tools able to properly measure disease activity, but the failure of the majority of randomized controlled trials enrolling SLE patients suggests that this field represents still an unmet need [64]. In this view, MLMs could play a potential role.

In 2018 the study published by Toro-Dominguez aimed at stratifying SLE patients in terms of disease activity according to gene expression. The application of cluster analysis allowed to identify three different clusters in pediatric and adult patients; furthermore, in one cluster the authors observed a significant correlation between neutrophils percentage and a lower disease activity, evaluated by SLEDAI [65]. Furthermore, by using a real-world dataset, MLMs resulted able to discriminate SLE patients with different SLEDAI values, when using a cut-off equal to five [66].

Other studies proposed the integration of clinical data with gene expression, also providing suggestions for pathogenic mechanisms implicated in determining disease activity. In this field, Kegerreis and colleagues proposed a Random Forest model with an accuracy equal to 83% in discriminating patients with active and inactive disease according to genetic profile [67]. More recently, rule-based machine learning models and rule networks were applied to develop gene networks to separate pediatric SLE patients according to a state of low and high disease activity. The authors proposed a model with a good performance (accuracy 81%) to distinguish different levels of disease activity. Furthermore, the application of unsupervised hierarchical clustering revealed additional subgroups characterized by the association between specific gene pathways and disease activity. In detail, the following genetic variants have been clustered: IFI35 and OTOF; KLRB1 encoding CD161; CKAP4 [68].

Interestingly, cluster analysis was applied to identify an association between risk flare and peripheral immunophenotypes, as assessed by flow cytometry. Thus, the so-called memory B-cells cluster showed a lower risk to develop disease flares compared with the non-memory B-cells group, including naïve B- and T-cells [69].

In 2017 our research group applied recurrent neural networks to predict chronic damage development, assessed by the SLICC Damage index (SDI) [70]. Thus, for the Recurrent Neural Network model we selected two groups of patients: patients with SDI = 0 at the baseline, developing damage during the follow-up, and those without damage during the whole follow-up. By using these data inputs, we could create a model with an AUC value equal to 0.77, able to predict damage development. A threshold value of 0.35 (sensitivity = 0.74, specificity = 0.76) seemed able to identify patients at risk to develop damage [71]. More recently, in the study conducted by Ahn and colleagues, cluster analysis allowed to identify three groups of patients according to the damage severity and mortality risk [72]. The relationship between damage clustering and mortality was previously evaluated by Pego-Regoisa et al. in a large Spanish SLE cohort. Overall, the authors identified three clusters according to the severity of damage, two of them showed a significantly higher mortality rate [73].

MLMs were recently used by our group to propose a new outcome in SLE field: the so-called Lupus comprehensive disease control (LupusCDC), including both the achievement of remission and the absence of damage progression [74]. The proposal of LupusCDC originates from the evidence that the control of disease activity is not always sufficient to stop the damage progression, due to the presence of other factors concurring with its development [11,75]. Thus, we applied SVM models and Decision Trees, followed by features ranking with the ReliefF algorithm. Our model, characterized by AUC value equal to 0.703, identified glucocorticoids, renal involvement and the use of immunosuppressant drugs as the most relevant factors concurring to the failure to achieve LupusCDC [74].

3.4. Treatment

In the last years the concept of precision medicine has been widely spread in the context of rheumatic conditions, including SLE. The heterogeneity of this disease, possible expression of different underlying pathogenic mechanisms, suggests the need for personalized treatment according to the most relevant manifestation [76]. In this context, MLMs could help the clinician in the treatment choice, by predicting drug response. However, very few data are available on this specific topic. In 2016 Kan and colleagues evaluated a large newly diagnosed SLE cohort by using cluster analysis: 10 treatment clusters were identified and the most common consisted of minimally treated patients (42.8%). In this cluster, hydroxychloroquine monotherapy, glucocorticoid monotherapy, and corticosteroid/hydroxychloroquine combination therapy were received by 34.0%, 11.2%, and 7.8% of patients, respectively [77]. More recently, Carter and colleagues observed that response to RTX in non-European SLE patients was lowest in an interferon-low, neutrophil-high cluster and highest in a cluster with high expression across all signatures ($p < 0.001$) [78]. Wang et al. applied MLMs to predict the effect of sirolimus on disease activity in 103 SLE patients. The so-called Emax model was selected for MLMs, where the evaluation indicator was the change rate of SLEDAI from the baseline value. The authors concluded that in order to achieve a better therapeutic effect (80% Emax, plateau), maintaining a concentration of 8–10 ng/mL sirolimus for at least 6–12 months was necessary [79]. Finally, recently MyPROSLE, an omic-based analytical workflow for measuring the molecular portrait of individual patients to support clinicians in their therapeutic decisions has been proposed. This is a machine learning-based classification model aiming at assessing the association between dysregulation in immunological response, clinical manifestations, prognosis, flare and remission events and response to Tabalumab. The model MyPROSLE allowed to molecularly summarize patients in 206 gene-modules, clustered into nine main lupus signatures. Preliminary results suggest that the dysregulation of certain gene-modules is strongly associated with specific clinical manifestations, the occurrence of relapses or the presence of long-term remission and drug response. Thus, the authors suggest the possible use of this model to predict clinical outcomes, including treatment response [80].

3.5. Pregnancy

SLE mostly affects women of childbearing age and as widely demonstrated, it could be associated with unfavorable pregnancy outcomes. Furthermore, fetal complications, in particular, fetal death and neonatal lupus syndrome could develop in SLE. Finally, disease flare during and after pregnancy is a common complication, with a prevalence ranging from 35% to 70% of patients [81]. In the last years, the pre-gestational counseling and the multidisciplinary approach adopted in the daily clinical practice allowed SLE patients to experience even more uncomplicated pregnancies [82]. Nonetheless, it is very important to select factors able to identify patients at risk of maternal–fetal complications.

In this context, the possible role of MLMs have been evaluated by two recent studies. Thus, Deng and colleagues applied Random Forest, support vector machine-recursive feature elimination and least absolute shrinkage with selection operator to identify genetic biomarkers for adverse pregnancy outcomes. The model identified three feature genes, specifically SEZ6, NRAD1, and LPAR4. Among these, SEZ6 showed the highest in-sample predictive performance, with an AUC value equal to 0.753 [83]. Moreover, Fazzari and colleagues confirmed, by using MLMs, the role of antihypertensive medication use, low platelets, SLE disease activity and lupus anticoagulant positivity as risk factors for adverse events during pregnancy. In detail, the authors evaluated a large SLE cohort by applying different models, in particular Logistic regression with stepwise selection, LASSO, Random Forest, neural network, Support Vector Machines, gradient boosting and SuperLearner. The best performance in terms of AUC was observed for LASSO model (AUC = 0.78) [84].

3.6. Other Possible Application

The increasing interest in the possible role of MLMs in SLE cohorts was demonstrated by the application of these tools in other disease-related fields.

Jorge and colleagues applied decision tree, Random Forest, naïve Bayes and logistic regression to predict hospitalization in SLE patients. By analyzing 1996 patients, 4.6% of them were hospitalized in the most recent year of follow-up, the authors demonstrated a good performance for Random Forest model (AUC = 0.751) in predicting hospitalization. Furthermore, anti-dsDNA positivity, low C3 levels, blood cell counts, and increased inflammatory biomarkers, as well as age and albumin, represented the most relevant risk factors for hospitalization [85].

Finally, Margiotta and colleagues used MLMs in the evaluation of quality of life in SLE patients by using cluster analysis. This approach allowed to distinguish different patterns of quality of life, characterized by the prominent involvement of mental or physical components, as assessed by Short-Form 36 (SF-36) [86].

Furthermore, the same MLM was able to identify different clusters related to sleep disorders in SLE subjects, by integrating data deriving from the Pittsburgh Sleep Quality Index and those from SF-36 and anxiety scores [87].

4. Conclusions

In conclusion, the present review focused on the possible application of MLMs in the SLE scenario. In particular, MLMs have been applied in the field of diagnosis, pathogenic mechanisms, definition of different disease features and courses, and finally treatment response. As demonstrated by our literature revision, several models have been proposed, revealing good performance in terms of accuracy and AUC.

These results suggest several possible future applications for MLMs. Among these, the application of specific models could help the physicians to identify patients at risk to develop more aggressive disease phenotypes, and thus could guide in the choice of a more appropriate treatments. Nonetheless, MLMs could be used to predict different phenomena, including the response to treatment, thus finding a place in the so-called precision medicine.

However, the application of MLMs in a real-life context finds some obstacles and may still be anticipatory. First of all, they require studies for internal and external validation, secondly it is mandatory to test the MLMs reliability and reproducibility. In the SLE scenario, the studies published so far are characterized by some limitations, such as the sample size of the analyzed cohorts and the lack of replication studies. These aspects certainly do not allow the use of these models in a real-life context.

Author Contributions: F.C. (Fulvia Ceccarelli) and F.C. (Fabrizio Conti) have designed the review, coordinated the bibliographic research, written and revised the final text. F.N., L.P., C.C., A.G., G.D., C.A. conducted the bibliographical research and wrote the first version of the text. All authors have read and agreed to the published version of the manuscript.

Funding: This research received no external funding.

Conflicts of Interest: The authors declare no conflict of interest.

References

1. Dörner, T.; Furie, R. Novel paradigms in systemic lupus erythematosus. *Lancet* **2019**, *393*, 2344–2358. [CrossRef]
2. Kent, T.; Davidson, A.; Newman, D.; Buck, G.; D'Cruz, D. Burden of illness in systemic lupus erythematosus: Results from a UK patient and carer online survey. *Lupus* **2017**, *26*, 1095–1100. [CrossRef]
3. Al Sawah, S.; Zhang, X.; Zhu, B.; Magder, L.S.; Foster, S.A.; Iikuni, N.; Petri, M. Effect of corticosteroid use by dose on the risk of developing organ damage over time in systemic lupus erythematosus-the Hopkins Lupus Cohort. *Lupus Sci. Med.* **2015**, *2*, e000066. [CrossRef] [PubMed]
4. Hochberg, M.C. Updating the American College of Rheumatology revised criteria for the classification of systemic lupus erythematosus. *Arthritis Rheum.* **1997**, *40*, 1725. [CrossRef]

5. Petri, M.; Orbai, A.M.; Alarcón, G.S.; Gordon, C.; Merrill, J.T.; Fortin, P.R.; Bruce, I.N.; Isenberg, D.; Wallace, D.J.; Nived, O.; et al. Derivation and validation of the Systemic Lupus International Collaborating Clinics classification criteria for systemic lupus erythematosus. *Arthritis Rheum.* **2012**, *64*, 2677–2686. [CrossRef] [PubMed]
6. Aringer, M.; Costenbader, K.; Daikh, D.; Brinks, R.; Mosca, M.; Ramsey-Goldman, R.; Smolen, J.S.; Wofsy, D.; Boumpas, D.T.; Kamen, D.L.; et al. 2019 European League Against Rheumatism/American College of Rheumatology Classification Criteria for Systemic Lupus Erythematosus. *Arthritis Rheumatol.* **2019**, *71*, 1400–1412. [CrossRef] [PubMed]
7. Ceccarelli, F.; Perricone, C.; Borgiani, P.; Ciccacci, C.; Rufini, S.; Cipriano, E.; Alessandri, C.; Spinelli, F.R.; Sili Scavalli, A.; Novelli, G.; et al. Genetic Factors in Systemic Lupus Erythematosus: Contribution to Disease Phenotype. *J. Immunol. Res.* **2015**, *2015*, 745647. [CrossRef]
8. Arbuckle, M.R.; McClain, M.T.; Rubertone, M.V.; Scofield, R.H.; Dennis, G.J.; James, J.A.; Harley, J.B. Development of autoantibodies before the clinical onset of systemic lupus erythematosus. *N. Eng. J. Med.* **2003**, *349*, 1526–1533. [CrossRef]
9. Ceccarelli, F.; Natalucci, F.; Olivieri, G.; Pirone, C.; Picciariello, L.; Orefice, V.; Truglia, S.; Spinelli, F.R.; Alessandri, C.; Chistolini, A.; et al. Development of Systemic Autoimmune Diseases in Healthy Subjects Persistently Positive for Antiphospholipid Antibodies: Long-Term Follow-Up Study. *Biomolecules* **2022**, *12*, 1088. [CrossRef]
10. Arnaud, L.; Tektonidou, M.G. Long-term outcomes in systemic lupus erythematosus: Trends over time and major contributors. *Rheumatology* **2020**, *59*, v29–v38. [CrossRef]
11. Conti, F.; Ceccarelli, F.; Perricone, C.; Leccese, I.; Massaro, L.; Pacucci, V.A.; Truglia, S.; Miranda, F.; Spinelli, F.R.; Alessandri, C.; et al. The chronic damage in systemic lupus erythematosus is driven by flares, glucocorticoids and antiphospholipid antibodies: Results from a monocentric cohort. *Lupus* **2016**, *25*, 719–726. [CrossRef]
12. Fanouriakis, A.; Tziolos, N.; Bertsias, G.; Boumpas, D.T. Update on the diagnosis and management of systemic lupus erythematosus. *Ann. Rheum. Dis.* **2021**, *80*, 14–25. [CrossRef]
13. Fanouriakis, A.; Kostopoulou, M.; Alunno, A.; Aringer, M.; Bajema, I.; Boletis, J.N.; Cervera, R.; Doria, A.; Gordon, C.; Govoni, M.; et al. 2019 update of the EULAR recommendations for the management of systemic lupus erythematosus. *Ann. Rheum. Dis.* **2019**, *78*, 736–745. [CrossRef]
14. Nelson, A.E.; Arbeeva, L. Narrative Review of Machine Learning in Rheumatic and Musculoskeletal Diseases for Clinicians and Researchers: Biases, Goals, and Future Directions. *J. Rheumatol.* **2022**, *49*, 1191–1200. [CrossRef]
15. Kingsmore, K.M.; Puglisi, C.E.; Grammar, A.C.; Lipsky, P.E. An introduction to machine learning and analysis of its use in rheumatic diseases. *Nat. Rev. Rheumatol.* **2021**, *17*, 710–730. [CrossRef]
16. Kohavi, R.; Provost, F. Glossary of terms. Machine Learning—Special Issue on Applications of Machine Learning and the Knowledge Discovery Process. *Mach. Learn.* **1998**, *30*, 271–274.
17. Zhu, X.; Goldberg, A.B. Introduction to Semi-Supervised Learning. *Synth. Lect. Artif. Intell. Mach. Learn.* **2009**, *3*, 1–130.
18. Lever, J.; Krzywinski, M.; Altman, N. Principal component analysis. *Nat. Methods* **2017**, *14*, 641–642. [CrossRef]
19. Guyon, I.; Elisseeff, A. An introduction to variable and feature selection. *J. Mach. Learn. Res.* **2003**, *3*, 27–46.
20. Krogh, A. What are artificial neural networks? *Nat. Biotechnol.* **2008**, *26*, 195–197. [CrossRef]
21. Cross, S.S.; Harrison, R.F.; Kennedy, R.L. Introduction to neural networks. *Lancet* **1995**, *346*, 1075–1079. [CrossRef]
22. Kumar, R.; Indrayan, A. Receiver operating characteristic (ROC) curve for medical researchers. *Indian Pediatr.* **2011**, *48*, 277–287. [CrossRef]
23. Hodson, T.O. Root-mean-square error (RMSE) or mean absolute error (MAE): When to use them or not. *Geosci. Model Dev.* **2022**, *15*, 5481–5487. [CrossRef]
24. Bastanlar, Y.; Özuysal, M. Introduction to machine learning. *Methods Mol. Biol.* **2014**, *1107*, 105–128.
25. Libbrecht, M.; Noble, W. Machine learning applications in genetics and genomics. *Nat. Rev. Genet.* **2015**, *16*, 321–332. [CrossRef]
26. Li, Y.; Ma, C.; Liao, S.; Qi, S.; Meng, S.; Cai, W.; Dai, W.; Cao, R.; Dong, X.; Krämer, B.K.; et al. Combined proteomics and single cell RNA-sequencing analysis to identify biomarkers of disease diagnosis and disease exacerbation for systemic lupus erythematosus. *Front Immunol.* **2022**, *13*, 969509. [CrossRef]
27. Zhong, Y.; Zhang, W.; Hong, X.; Zeng, Z.; Chen, Y.; Liao, S.; Cai, W.; Xu, Y.; Wang, G.; Liu, D.; et al. Screening Biomarkers for Systemic Lupus Erythematosus Based on Machine Learning and Exploring Their Expression Correlations With the Ratios of Various Immune Cells. *Front Immunol.* **2022**, *13*, 873787. [CrossRef]
28. Jiang, Z.; Shao, M.; Dai, X.; Pan, Z.; Liu, D. Identification of Diagnostic Biomarkers in Systemic Lupus Erythematosus Based on Bioinformatics Analysis and Machine Learning. *Front Genet.* **2022**, *13*, 865559. [CrossRef]
29. Ma, W.; Lau, Y.L.; Yang, W.; Wang, Y.F. Random forests algorithm boosts genetic risk prediction of systemic lupus erythematosus. *Front Genet.* **2022**, *13*, 902793. [CrossRef]
30. Martorell-Marugán, J.; Chierici, M.; Jurman, G.; Alarcón-Riquelme, M.E.; Carmona-Sáez, P. Differential diagnosis of systemic lupus erythematosus and Sjögren's syndrome using machine learning and multi-omics data. *Comput. Biol Med.* **2022**, *152*, 106373. [CrossRef]
31. Barnado, A.; Eudy, A.M.; Blaske, A.; Wheless, L.; Kirchoff, K.; Oates, J.C.; Clowse, M.E.B. Developing and Validating Methods to Assemble Systemic Lupus Erythematosus Births in the Electronic Health Record. *Arthritis Care Res.* **2022**, *74*, 849–857. [CrossRef]
32. Matthiesen, R.; Lauber, C.; Sampaio, J.L.; Domingues, N.; Alves, L.; Gerl, M.J.; Almeida, M.S.; Rodrigues, G.; Araújo Gonçalves, P.; Ferreira, J.; et al. Shotgun mass spectrometry-based lipid profiling identifies and distinguishes between chronic inflammatory diseases. *EBioMedicine* **2021**, *70*, 103504. [CrossRef]

33. Adamichou, C.; Genitsaridi, I.; Nikolopoulos, D.; Nikoloudaki, M.; Repa, A.; Bortoluzzi, A.; Fanouriakis, A.; Sidiropoulos, P.; Boumpas, D.T.; Bertsias, G.K. Lupus or not? SLE Risk Probability Index (SLERPI): A simple, clinician-friendly machine learning-based model to assist the diagnosis of systemic lupus erythematosus. *Ann. Rheum. Dis.* **2021**, *80*, 758–766. [CrossRef]
34. Ceccarelli, F.; Lapucci, M.; Olivieri, G.; Sortino, A.; Natalucci, F.; Spinelli, F.R.; Alessandri, C.; Sciandrone, M.; Conti, F. Can machine learning models support physicians in systemic lupus erythematosus diagnosis? Results from a monocentric cohort. *Jt. Bone Spine* **2022**, *89*, 105292. [CrossRef]
35. Park, J.; Jang, W.; Park, H.S.; Park, K.H.; Kwok, S.K.; Park, S.H.; Oh, E.J. Cytokine clusters as potential diagnostic markers of disease activity and renal involvement in systemic lupus erythematosus. *J. Int. Med. Res.* **2020**, *48*, 300060520926882. [CrossRef]
36. Guthridge, J.M.; Lu, R.; Tran, L.T.; Arriens, C.; Aberle, T.; Kamp, S.; Munroe, M.E.; Dominguez, N.; Gross, T.; DeJager, W.; et al. Adults with systemic lupus exhibit distinct molecular phenotypes in a cross-sectional study. *EClinicalMedicine* **2020**, *20*, 100291. [CrossRef]
37. Jorge, A.; Castro, V.M.; Barnado, A.; Gainer, V.; Hong, C.; Cai, T.; Cai, T.; Carroll, R.; Denny, J.C.; Crofford, L.; et al. Identifying lupus patients in electronic health records: Development and validation of machine learning algorithms and application of rule-based algorithms. *Semin. Arthritis Rheum.* **2019**, *49*, 84–90. [CrossRef]
38. Murray, S.G.; Avati, A.; Schmajuk, G.; Yazdany, J. Automated and flexible identification of complex disease: Building a model for systemic lupus erythematosus using noisy labeling. *J. Am. Med. Inform. Assoc.* **2019**, *26*, 61–65. [CrossRef]
39. Turner, C.A.; Jacobs, A.D.; Marques, C.K.; Oates, J.C.; Kamen, D.L.; Anderson, P.E.; Obeid, J.S. Word2Vec inversion and traditional text classifiers for phenotyping lupus. *BMC Med. Inform. Decis. Mak.* **2017**, *17*, 126. [CrossRef]
40. Dai, Y.; Hu, C.; Wang, L.; Huang, Y.; Zhang, L.; Xiao, X.; Tan, Y. Serum peptidome patterns of human systemic lupus erythematosus based on magnetic bead separation and MALDI-TOF mass spectrometry analysis. *Scand. J. Rheumatol.* **2010**, *3*, 240–246. [CrossRef]
41. Huang, Z.; Shi, Y.; Cai, B.; Wang, L.; Wu, Y.; Ying, B.; Qin, L.; Hu, C.; Li, Y. MALDI-TOF MS combined with magnetic beads for detecting serum protein biomarkers and establishment of boosting decision tree model for diagnosis of systemic lupus erythematosus. *Rheumatology* **2009**, *48*, 626–631. [CrossRef]
42. Diaz-Gallo, L.M.; Oke, V.; Lundström, E.; Elvin, K.; Ling Wu, Y.; Eketjäll, S.; Zickert, A.; Gustafsson, J.T.; Jönsen, A.; Leonard, D.; et al. Four Systemic Lupus Erythematosus Subgroups, Defined by Autoantibodies Status, Differ Regarding HLA-DRB1 Genotype Associations and Immunological and Clinical Manifestations. *ACR Open Rheumatol.* **2022**, *4*, 27–39. [CrossRef]
43. Lu, Z.; Li, W.; Tang, Y.; Da, Z.; Li, X. Lymphocyte subset clustering analysis in treatment-naive patients with systemic lupus erythematosus. *Clin. Rheumatol.* **2021**, *40*, 1835–1842. [CrossRef]
44. Reynolds, J.A.; McCarthy, E.M.; Haque, S.; Ngamjanyaporn, P.; Sergeant, J.C.; Lee, E.; Lee, E.; Kilfeather, S.A.; Parker, B.; Bruce, I.N. Cytokine profiling in active and quiescent SLE reveals distinct patient subpopulations. *Arthritis Res. Ther.* **2018**, *20*, 173. [CrossRef]
45. Anders, H.J.; Saxena, R.; Zhao, M.H.; Parodis, I.; Salmon, J.E.; Mohan, C. Lupus nephritis. *Nat. Rev. Dis. Primers.* **2020**, *6*, 7. [CrossRef] [PubMed]
46. Tang, H.; Poynton, M.R.; Hurdle, J.F.; Baird, B.C.; Koford, J.K.; Goldfarb-Rumyantzev, A.S. Predicting three-year kidney graft survival in recipients with systemic lupus erythematosus. *ASAIO J.* **2011**, *57*, 300–309. [CrossRef]
47. Chen, Y.; Huang, S.; Chen, T.; Liang, D.; Yang, J.; Zeng, C.; Li, X.; Xie, G.; Liu, Z. Machine Learning for Prediction and Risk Stratification of Lupus Nephritis Renal Flare. *Am. J. Nephrol.* **2021**, *52*, 152–160. [CrossRef]
48. Wang, L.; Yang, Z.; Yu, H.; Lin, W.; Wu, R.; Yang, H.; Yang, K. Predicting diagnostic gene expression profiles associated with immune infiltration in patients with lupus nephritis. *Front. Immunol.* **2022**, *13*, 839197. [CrossRef]
49. Stojanowski, J.; Konieczny, A.; Rydzyńska, K.; Kasenberg, I.; Mikołajczak, A.; Gołębiowski, T.; Krajewska, M.; Kusztal, M. Artificial neural network—An effective tool for predicting the lupus nephritis outcome. *BMC Nephrol.* **2022**, *23*, 381. [CrossRef]
50. Wang, M.; Liang, Y.; Hu, Z.; Chen, S.; Shi, B.; Heidari, A.A.; Zhang, Q.; Chen, H.; Chen, X. Lupus nephritis diagnosis using enhanced moth flame algorithm with support vector machines. *Comput. Biol. Med.* **2022**, *145*, 105435. [CrossRef]
51. Ayoub, I.; Wolf, B.J.; Geng, L.; Song, H.; Khatiwada, A.; Tsao, B.P.; Oates, J.C.; Rovin, B.H. Prediction models of treatment response in lupus nephritis. *Kidney Int.* **2022**, *101*, 379–389. [CrossRef]
52. Yang, C.K.; Lee, C.Y.; Wang, H.S.; Huang, S.C.; Liang, P.I.; Chen, J.S.; Kuo, C.F.; Tu, K.H.; Yeh, C.Y.; Chen, T.D. Glomerular disease classification and lesion identification by machine learning. *Biomed. J.* **2022**, *4*, 675–685. [CrossRef]
53. Carrión-Barberà, I.; Salman-Monte, T.C.; Vílchez-Oya, F.; Monfort, J. Neuropsychiatric involvement in systemic lupus erythematosus: A review. *Autoimmun Rev.* **2021**, *20*, 102780. [CrossRef]
54. Gu, X.X.; Jin, Y.; Fu, T.; Zhang, X.M.; Li, T.; Yang, Y.; Li, R.; Zhou, W.; Guo, J.X.; Zhao, R.; et al. Relevant Characteristics Analysis Using Natural Language Processing and Machine Learning Based on Phenotypes and T-Cell Subsets in Systemic Lupus Erythematosus Patients with Anxiety. *Front. Psychiatry* **2021**, *12*, 793505. [CrossRef] [PubMed]
55. Rumetshofer, T.; Inglese, F.; de Bresser, J.; Mannfolk, P.; Strandberg, O.; Jönsen, A.; Bengtsson, A.; Nilsson, M.; Knutsson, L.; Lätt, J.; et al. Tract-based white matter hyperintensity patterns in patients with systemic lupus erythematosus using an unsupervised machine learning approach. *Sci. Rep.* **2022**, *12*, 21376. [CrossRef]
56. Tan, G.; Huang, B.; Cui, Z.; Dou, H.; Zheng, S.; Zhou, T. A noise-immune reinforcement learning method for early diagnosis of neuropsychiatric systemic lupus erythematosus. *Math. Biosci. Eng.* **2022**, *19*, 2219–2239. [CrossRef]

57. Barraclough, M.; Erdman, L.; Diaz-Martinez, J.P.; Knight, A.; Bingham, K.; Su, J.; Kakvan, M.; Muñoz Grajales, C.; Tartaglia, M.C.; Ruttan, L.; et al. Systemic lupus erythematosus phenotypes formed from machine learning with a specific focus on cognitive impairment. *Rheumatology* **2022**, *17*, keac653. [CrossRef]
58. Ceccarelli, F.; Perricone, C.; Cipriano, E.; Massaro, L.; Natalucci, F.; Capalbo, G.; Leccese, I.; Bogdanos, D.; Spinelli, F.R.; Alessandri, C.; et al. Joint involvement in systemic lupus erythematosus: From pathogenesis to clinical assessment. *Semin. Arthritis Rheum.* **2017**, *47*, 53–64. [CrossRef]
59. Ceccarelli, F.; Sciandrone, M.; Perricone, C.; Galvan, G.; Cipriano, E.; Galligari, A.; Levato, T.; Colasanti, T.; Massaro, L.; Natalucci, F.; et al. Biomarkers of erosive arthritis in systemic lupus erythematosus: Application of machine learning models. *PLoS ONE* **2018**, *13*, e0207926. [CrossRef]
60. Ceccarelli, F.; Natalucci, F.; Pirone, C.; Olivieri, G.; Colasanti, T.; Picciariello, L.; Spinelli, F.R.; Alessandri, C.; Conti, F. Erosive arthritis in systemic lupus erythematosus: Application of cluster analysis. *Clin. Exp. Rheumatol.* **2022**, *40*, 2175–2178. [CrossRef]
61. Ceccarelli, F.; Natalucci, F.; Olivieri, G.; Perricone, C.; Pirone, C.; Spinelli, F.R.; Alessandri, C.; Conti, F. Erosive arthritis in systemic lupus erythematosus: Not only Rhupus. *Lupus* **2021**, *30*, 2029–2041. [CrossRef]
62. Liu, C.; Zhou, Y.; Zhou, Y.; Tang, X.; Tang, L.; Wang, J. Identification of crucial genes for predicting the risk of atherosclerosis with system lupus erythematosus based on comprehensive bioinformatics analysis and machine learning. *Comput. Biol. Med.* **2023**, *152*, 106388. [CrossRef] [PubMed]
63. Wang, Y.; Huang, Z.; Xiao, Y.; Wan, W.; Yang, X. The shared biomarkers and pathways of systemic lupus erythematosus and metabolic syndrome analyzed by bioinformatics combining machine learning algorithm and single-cell sequencing analysis. *Front. Immunol.* **2022**, *13*, 1015882. [CrossRef] [PubMed]
64. Isenberg, D.A.; Merrill, J.T. Why, why, why de-lupus (does so badly in clinical trials). *Expert Rev. Clin. Immunol.* **2016**, *12*, 95–98. [CrossRef] [PubMed]
65. Toro-Domínguez, D.; Martorell-Marugán, J.; Goldman, D.; Petri, M.; Carmona-Sáez, P.; Alarcón-Riquelme, M.E. Stratification of Systemic Lupus Erythematosus Patients Into Three Groups of Disease Activity Progression According to Longitudinal Gene Expression. *Arthritis Rheumatol.* **2018**, *70*, 2025–2035. [CrossRef]
66. Alves, P.; Bandaria, J.; Leavy, M.B.; Gliklich, B.; Boussios, C.; Su, Z.; Curhan, G. Validation of a machine learning approach to estimate Systemic Lupus Erythematosus Disease Activity Index score categories and application in a real-world dataset. *RMD Open* **2021**, *7*, e001586. [CrossRef] [PubMed]
67. Kegerreis, B.; Catalina, M.D.; Bachali, P.; Geraci, N.S.; Labonte, A.C.; Zeng, C.; Stearrett, N.; Crandall, K.A.; Lipsky, P.E.; Grammar, A.C. Machine learning approaches to predict lupus disease activity from gene expression data. *Sci. Rep.* **2019**, *9*, 9617. [CrossRef]
68. Yones, S.A.; Annett, A.; Stoll, P.; Diamanti, K.; Holmfeldt, L.; Barrenäs, C.F.; Meadows, J.R.S.; Komorowski, J. Interpretable machine learning identifies paediatric Systemic Lupus Erythematosus subtypes based on gene expression data. *Sci. Rep.* **2022**, *1*, 7433. [CrossRef]
69. Zheng, J.; Zhu, L.; Ju, B.; Zhang, J.; Luo, J.; Wang, Y.; Lv, X.; Pu, D.; He, L.; Wang, J. Peripheral immunophenotypes associated with the flare in the systemic lupus erythematosus patients with low disease activity state. *Clin. Immunol.* **2022**, *245*, 109166. [CrossRef]
70. Gladman, D.D.; Urowitz, M.B.; Goldsmith, C.H.; Fortin, P.; Ginzler, E.; Gordon, C.; Hanly, J.G.; Isenberg, D.A.; Kalunian, K.; Nived, O.; et al. The reliability of the Systemic Lupus International Collaborating Clinics/American College of Rheumatology Damage Index in patients with systemic lupus erythematosus. *Arthritis Rheum.* **1997**, *40*, 809–813. [CrossRef]
71. Ceccarelli, F.; Sciandrone, M.; Perricone, C.; Galvan, G.; Morelli, F.; Vicente, L.N.; Leccese, I.; Massaro, L.; Cipriano, E.; Spinelli, F.R.; et al. Prediction of chronic damage in systemic lupus erythematosus by using machine-learning models. *PLoS ONE* **2017**, *3*, e0174200. [CrossRef]
72. Ahn, G.Y.; Lee, J.; Won, S.; Ha, E.; Kim, H.; Nam, B.; Kim, J.S.; Kang, J.; Kim, J.H.; Song, G.G.; et al. Identifying damage clusters in patients with systemic lupus erythematosus. *Int. J. Rheum. Dis.* **2020**, *23*, 84–91. [CrossRef]
73. Pego-Reigosa, J.M.; Lois-Iglesias, A.; Rúa-Figueroa, Í.; Galindo, M.; Calvo-Alén, J.; de Uña-Álvarez, J.; Balboa-Barreiro, V.; Ibáñez Ruan, J.; Olivé, A.; Rodríguez-Gómez, M.; et al. Relationship between damage clustering and mortality in systemic lupus erythematosus in early and late stages of the disease: Cluster analyses in a large cohort from the Spanish Society of Rheumatology Lupus Registry. *Rheumatology* **2016**, *55*, 1243–1250. [CrossRef]
74. Ceccarelli, F.; Olivieri, G.; Sortino, A.; Dominici, L.; Arefayne, F.; Celia, A.I.; Cipriano, E.; Garufi, C.; Lapucci, M.; Mancuso, S.; et al. Comprehensive disease control in systemic lupus erythematosus. *Semin. Arthritis Rheum.* **2021**, *51*, 404–408. [CrossRef]
75. Ceccarelli, F.; Olivieri, G.; Pirone, C.; Ciccacci, C.; Picciariello, L.; Natalucci, F.; Perricone, C.; Spinelli, F.R.; Alessandri, C.; Borgiani, P.; et al. The Impacts of the Clinical and Genetic Factors on Chronic Damage in Caucasian Systemic Lupus Erythematosus Patients. *J. Clin. Med.* **2022**, *11*, 3368. [CrossRef]
76. Lever, E.; Alves, M.R.; Isenberg, D.A. Towards Precision Medicine in Systemic Lupus Erythematosus. *Pharmgenomics Pers. Med.* **2020**, *13*, 39–49. [CrossRef]
77. Kan, H.; Nagar, S.; Patel, J.; Wallace, D.J.; Molta, C.; Chang, D.J. Longitudinal Treatment Patterns and Associated Outcomes in Patients with Newly Diagnosed Systemic Lupus Erythematosus. *Clin. Ther.* **2016**, *38*, 610–624. [CrossRef] [PubMed]
78. Carter, L.M.; Alase, A.; Wigston, Z.; Psarras, A.; Burska, A.; Sutton, E.; Md Yusof, M.Y.; Reynolds, J.A.; Masterplans Consortium; McHugh, N.; et al. Gene expression and autoantibody analysis reveals distinct ancestry-specific profiles associated with response to rituximab in refractory systemic lupus erythematosus. *Arthritis Rheumatol.* 2022, *Epub ahead of print*. [CrossRef] [PubMed]

79. Wang, D.D.; Li, Y.F.; Zhang, C.; He, S.M.; Chen, X. Predicting the effect of sirolimus on disease activity in patients with systemic lupus erythematosus using machine learning. *J. Clin. Pharm. Ther.* **2022**, *47*, 1845–1850. [CrossRef] [PubMed]
80. Toro-Domínguez, D.; Martorell-Marugán, J.; Martinez-Bueno, M.; López-Domínguez, R.; Carnero-Montoro, E.; Barturen, G.; Goldman, D.; Petri, M.; Carmona-Sáez, P.; Alarcón-Riquelme, M.E. Scoring personalized molecular portraits identify Systemic Lupus Erythematosus subtypes and predict individualized drug responses, symptomatology and disease progression. *Brief Bioinform.* **2022**, *23*, bbac332. [CrossRef] [PubMed]
81. Kwok, L.-W.; Tam, L.-S.; Zhu, T.; Leung, Y.-Y.; Li, E. Predictors of maternal and fetal outcomes in pregnancies of patients with systemic lupus erythematosus. *Lupus* **2011**, *20*, 829–836. [CrossRef] [PubMed]
82. Andreoli, L.; Bertsias, G.K.; Agmon-Levin, N.; Brown, S.; Cervera, R.; Costedoat-Chalumeau, N.; Doria, A.; Fischer-Betz, R.; Forger, F.; Moraes-Fontes, M.F.; et al. EULAR recommendations for women's health and the management of family planning, assisted reproduction, pregnancy and menopause in patients with systemic lupus erythematosus and/or antiphospholipid syndrome. *Ann. Rheum. Dis.* **2017**, *76*, 476–485. [CrossRef] [PubMed]
83. Deng, Y.; Zhou, Y.; Shi, J.; Yang, J.; Huang, H.; Zhang, M.; Wang, S.; Ma, Q.; Liu, Y.; Li, B.; et al. Potential genetic biomarkers predict adverse pregnancy outcome during early and mid-pregnancy in women with systemic lupus erythematosus. *Front. Endocrinol* **2022**, *13*, 957010. [CrossRef] [PubMed]
84. Fazzari, M.J.; Guerra, M.M.; Salmon, J.; Kim, M.Y. Adverse pregnancy outcomes in women with systemic lupus erythematosus: Can we improve predictions with machine learning? *Lupus Sci. Med.* **2022**, *1*, e000769. [CrossRef]
85. Jorge, A.M.; Smith, D.; Wu, Z.; Chowdhury, T.; Costenbader, K.; Zhang, Y.; Choi, H.K.; Feldman, C.H.; Zhao, Y. Exploration of machine learning methods to predict systemic lupus erythematosus hospitalizations. *Lupus* **2022**, *31*, 1296–1305. [CrossRef] [PubMed]
86. Margiotta, D.P.E.; Fasano, S.; Basta, F.; Pierro, L.; Riccardi, A.; Navarini, L.; Valentini, G.; Afeltra, A. Clinical features of patients with systemic lupus erythematosus according to health-related quality of life, entity of pain, fatigue and depression: A cluster analysis. *Clin. Exp. Rheumatol.* **2019**, *37*, 535–539. [PubMed]
87. Margiotta, D.P.E.; Laudisio, A.; Navarini, L.; Basta, F.; Mazzuca, C.; Angeletti, S.; Ciccozzi, M.; Incalzi, R.A.; Afeltra, A. Pattern of sleep dysfunction in systemic lupus erythematosus: A cluster analysis. *Clin. Rheumatol.* **2019**, *38*, 1561–1570. [CrossRef]

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