CLINICAL SCIENCE

Lupus or not? SLE Risk Probability Index (SLERPI): a simple, clinician-friendly machine learning-based model to assist the diagnosis of systemic lupus erythematosus

Christina Adamichou,[1] Irini Genitsaridi,[1] Dionysis Nikolopoulos [ID],[2] Myrto Nikoloudaki,[1] Argyro Repa,[1] Alessandra Bortoluzzi,[3] Antonis Fanouriakis [ID],[2,4] Prodromos Sidiropoulos [ID],[1,5] Dimitrios T Boumpas [ID],[2,6] George K Bertsias [ID][1,5]

Correspondence to
Dr George K Bertsias, Rheumatology, Clinical Immunology and Allergy, University of Crete School of Medicine, Heraklion 700 13, Greece; gbertsias@uoc.gr

CA and IG contributed equally.

ABSTRACT

Objectives Diagnostic reasoning in systemic lupus erythematosus (SLE) is a complex process reflecting the probability of disease at a given timepoint against competing diagnoses. We applied machine learning in well-characterised patient data sets to develop an algorithm that can aid SLE diagnosis.

Methods From a discovery cohort of randomly selected 802 adults with SLE or control rheumatologic diseases, clinically selected panels of deconvoluted classification criteria and non-criteria features were analysed. Feature selection and model construction were done with Random Forests and Least Absolute Shrinkage and Selection Operator-logistic regression (LASSO-LR). The best model in 10-fold cross-validation was tested in a validation cohort (512 SLE, 143 disease controls).

Results A novel LASSO-LR model had the best performance and included 14 variably weighed features with thrombocytopenia/haemolytic anaemia, malar/maculopapular rash, proteinuria, low C3 and C4, antinuclear antibodies (ANA) and immunologic disorder being the strongest SLE predictors. Our model produced SLE risk probabilities (depending on the combination of features) correlating positively with disease severity and organ damage, and allowing the unbiased classification of a validation cohort into diagnostic certainty levels (unlikely, possible, likely, definitive SLE) based on the likelihood of SLE against other diagnoses. Operating the model as binary (lupus/not-lupus), we noted excellent accuracy (94.8%) for identifying SLE, and high sensitivity for early disease (93.8%), nephritis (97.9%), neuropsychiatric (91.8%) and severe lupus requiring immunosuppressives/biologics (96.4%). This was converted into a scoring system, whereby a score >7 has 94.2% accuracy.

Conclusions We have developed and validated an accurate, clinician-friendly algorithm based on classical disease features for early SLE diagnosis and treatment to improve patient outcomes.

Key messages

What is already known about this subject?
► Systemic lupus erythematosus (SLE) diagnosis often poses significant challenges especially at early stages and formal diagnostic criteria are currently missing.

What does this study add?
► By the use of machine learning (Least Absolute Shrinkage and Selection Operator-logistic regression) training of well-defined features of SLE, including features that are not part of the classification criteria, derived from a large discovery cohort, we have developed a novel statistical model for SLE diagnosis.
► The new model, including 14 variably weighed, standard clinical and serological features, can produce individualised SLE risk probabilities enabling the classification of a validation cohort into unlikely, possible, likely and definite SLE.
► When treated as binary (ie, SLE or not SLE), the model shows excellent combination of sensitivity and specificity for SLE (including early and severe forms of the disease) against competing rheumatologic diseases.
► The logistic regression model can be converted into a simple scoring system for both clinical and serological features, with an operational cut-off score of 7.

How might this impact on clinical practice?
► Pending further validation in prospective studies, the new diagnostic model (SLE Risk Probability Index) can assist the early diagnosis and treatment of patients with SLE to improve disease outcomes.

INTRODUCTION

Diagnosis of systemic lupus erythematosus (SLE) can be challenging and delayed by several months or years,[1–3] resulting in increased patient uncertainty, referrals and healthcare utilisation.[4] Delays in diagnosis and treatment initiation have been linked to increased flares and organ dysfunction.[4–6]

SLE diagnosis often relies on the acumen of physicians and is typically elicited by the presence of 'high-yield' features or multiple, although less-specific findings. Due to absence of diagnostic criteria, classification criteria, developed to facilitate the inclusion of homogenous disease populations in clinical

BMJ

studies,[7] are commonly used as a diagnostic aid. The Systemic Lupus International Collaborating Clinics (SLICC)[8] and European League Against Rheumatism/American College of Rheumatology (EULAR/ACR)[9 10] criteria enable the earlier classification of increased number of patients.[11] Moreover, the EULAR/ACR 2019 criteria achieve the highest combination of sensitivity and specificity.[9–11] Improved classification has not remedied the challenge for diagnosis especially at early stages.[11 12]

Artificial intelligence tools based on machine learning (ML) are increasingly used to manage difficult medical tasks. Such models can be trained from different kinds of medical or biological data.[13 14] ML has been used for the molecular classification of inflammatory myositis[15] and rheumatoid arthritis,[16] for predicting mortality,[17] response to biological agents[18] and disease activity,[19] whereas less effort has been directed towards diagnosis.[20 21] Building robust computational models that avoid excess complexity represents an important challenge.[14 22]

Herein, we applied ML on panels of clinical features aiming to construct a model that can accurately detect SLE against competing rheumatologic conditions. We used a discovery cohort of patients with SLE or control diseases to train two standard ML algorithms, namely, the Random Forests (RF) the Least Absolute Shrinkage and Selection Operator (LASSO) followed by Logistic Regression (LR). RF is a non-linear method with high complexity and thus, less explainable, whereas the LR is a linear method supporting simpler, more clinically interpretable results.[23] The best model selected by internal cross-validation (CV) was further evaluated in an independent validation cohort. Through this process, we developed a novel, simple Least Absolute Shrinkage and Selection Operator-logistic regression (LASSO-LR) model of variably weighed, standard clinical features that can produce individualised SLE risk probabilities alike clinical diagnostic reasoning. Our model had excellent accuracy for SLE, including early and severe forms of the disease, therefore it could represent a useful clinical tool.

METHODS
Discovery and validation cohorts
We used data from the Rheumatology Clinics at the University Hospital of Heraklion and the 'Attikon' University Hospital, Athens. Both centres have established SLE registries and use homogenised, structured forms for collecting clinical characteristics (including classification criteria), use of treatments and disease outcomes.[11 24–26] We included patients diagnosed during 01/2005-06/2019 with SLE or miscellaneous control rheumatological diseases that are relevant to the differential diagnosis of lupus (online supplemental table S1) by consultant rheumatologists with ≥5 years clinical practice. A randomly selected discovery cohort of 401 patients with SLE and 401 controls ere used to construct, train and compare the ML models. The balanced (1:1) ratio of SLE and controls helps to minimise any predictive modeling biases. An external validation cohort of consecutively registered 512 patients with SLE and 143 controls was used to provide an unbiased estimate of the diagnostic accuracy of the best model. The study was approved by the local ethics committees.

Variables and data set preparation
For each patient, demographics, rheumatological disease and date of diagnosis, date of earliest reported occurrence of each of the items from the three classification criteria (ACR 1997,[27] SLICC 2012,[8] EULAR/ACR 2019[9 10]) and date of last follow-up visit/assessment were extracted. Attribution of the criteria items

to SLE or not was arbitrated by rheumatologists (DTB, GKB, AF) using the EULAR/ACR attribution rule.[9 10] We used criteria items both in their original version and after deconvolution into subitems (eg, 'maculopapular rash' subitem from the EULAR/ACR 2019 'acute cutaneous lupus' criterion). In addition, we monitored a predefined list of non-criteria features (online supplemental table S2). Missing data were eliminated through vigorous charts review and quality control.

Disease subsets and outcomes
Early SLE was defined as duration less than 24 months since diagnosis. Lupus nephritis was determined according to kidney histological findings suggestive of lupus in a patient with compatible clinical and/or serological findings. Neuropsychiatric lupus was diagnosed through multidisciplinary approach[28] and ascertained by the Italian Study Group attribution model.[29] The British Isles Lupus Assessment Group (BILAG) glossary[30] was used to classify the severity of manifestations as previously detailed.[11 26] Use of immunosuppressive/biologic treatments and the physician global assessment of disease severity were also collected. The date of each item of the SLICC/ACR damage index (SDI)[31] was monitored.

Feature selection, model construction and evaluation
We followed two approaches for developing a predictive model for SLE. First, we combined each one of the three classification criteria with additional, non-redundant features from the other two criteria sets and with non-criteria features; second, we developed a de novo model based on clinical variables selected from the three classification criteria and non-criteria features. Univariable LR (online supplemental table S3) was performed in the discovery cohort to determine the association of each individual feature with SLE and correlation analysis (online supplemental table S4) to detect collinearity between features/predictors and assist clinicians in the construction of feature panels. Clinicians (GKB,CA) created 20 panels of features with the aim to introduce alternative feature versions. Each panel was submitted into two ML algorithms for feature selection, namely, RF and LASSO, the latter followed by LR (figure 1). Details are provided in the online supplemental methods.

We performed a 10-fold stratified CV) process (division of the dataset into 10 folds of near-equal size without resubstitution) to construct and compare the 40 multivariable models for their predictive capability. Each fold (10%) was used as a test data set to determine the model performance, while the remaining nine folds (90%) were used as the training data set for the model construction. We evaluated the following metrics (averaged from the 10 CV test data sets): sensitivity, specificity, accuracy and area under the receiver operating characteristic curve (AUC-ROC). The model with the highest accuracy was selected as the best to undergo evaluation in the validation cohort.

Statistical analysis
The Kruskal-Wallis analysis of variance was used to compare means and the χ^2 test to compare proportions. To convert the LASSO-LR model into scoring system, regression coefficients were divided by the smallest coefficient followed by rounding to the nearest 0.5 value. Statistical analyses were performed using the R software (V.3.5.1) and SPSS (V.25.0). Feature selection and ranking, model construction, evaluation and validation were developed in MATLAB V.9.2.
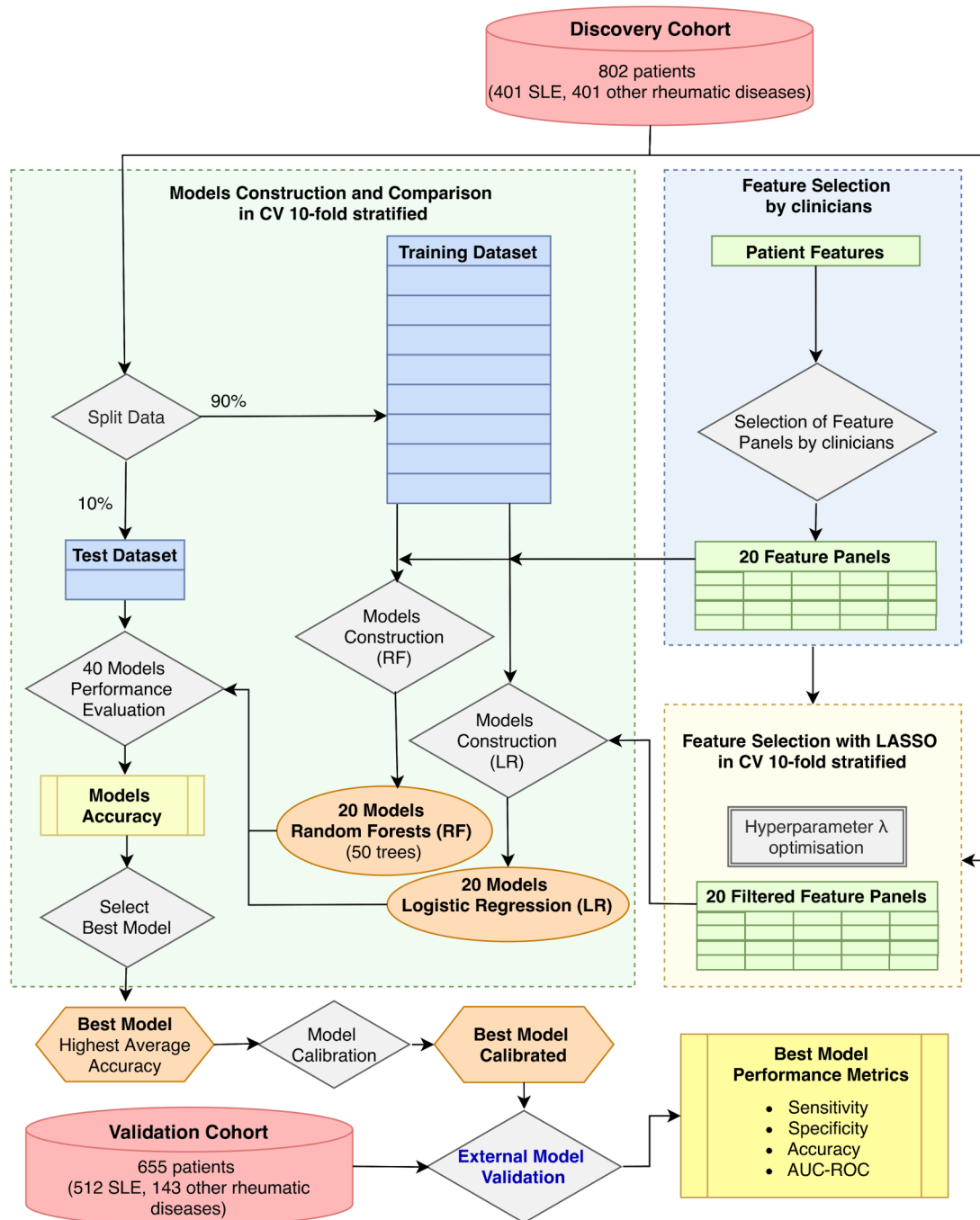
Figure 1 Schematic overview of the methodology for developing a machine learning-based diagnostic model for SLE. We used a *discovery cohort* of randomly selected 802 adults with SLE or control rheumatologic diseases (1:1 ratio) to prepare 20 clinically selected panels of classification criteria items (both in their original version and deconvoluted into subitems in the case of composite items) and non-criteria features. Two machine learning methods were applied for feature selection and model construction for each panel: (A) Random Forests (RF) and (B) Least Absolute Shrinkage and Selection Operator (LASSO) followed by logistic regression (LASSO-LR). The best model (highest accuracy the in 10-fold cross-validation process) was further tested in an independent dataset of 512 patients with systemic lupus erythematosus (SLE) and 143 disease controls (*validation cohort*). AUC, area under the curve; CV, cross-validation; ROC, receiver operating curve.

RESULTS
Combination of the classification criteria with additional features yields modest improvements in diagnostic accuracy for SLE

Classification criteria comprising different collections of clinical and immunological features classify patients with SLE in routine practice with high sensitivity and specificity.[11][32–38] We examined whether their combination with additional, non-redundant

features can improve their performance. We used a balanced discovery cohort of 802 patients with clinically diagnosed SLE or control diseases to fit two ML algorithms, RF and LASSO-LR (figure 1). Combinatory models of the criteria with additional features showed increased accuracy for SLE (by 0.38%–3.11% in the 10-fold CV runs) over the original versions of the criteria (online supplemental table S5). The greatest improvement was observed for ACR 1997-based models where LASSO feature

selection identified alopecia, hypocomplementinemia, maculopapular rash and interstitial lung disease (ILD) as additional predictors for SLE. Modeling the EULAR/ACR 2019 classification score together with antinuclear antibodies (ANA) (treated as additional feature rather than as *entry criterion*), ILD and livedo reticularis showed enhanced diagnostic performance. These results suggest that certain modifications that could improve—albeit modestly—the accuracy of the classification criteria for SLE.

A de novo-constructed LR model has superior performance for SLE diagnosis

We next sought to develop a novel statistical algorithm by integrating individual items from the classification criteria and additional non-criteria manifestations. Feature selection was performed either embedded in RF or prior to the model construction phase, with LR based on LASSO-LR. An important difference between these two methods is that if several highly correlated variables are predictive, LASSO may select one or a few while RF may use all of them. The best model in the discovery cohort 10-fold CV runs was a LASSO-LR model of 14 clinical parameters (hereafter referred to as 'SLERPI': SLE Risk Probability Index) (online supplemental table S5).

The model parameters included features from all three sets of classification criteria and ILD as a single non-criteria feature. Autoimmune thrombocytopenia or haemolytic anaemia, malar or maculopapular rash, low C3 and C4, proteinuria (all defined according to the EULAR/ACR 2019 criteria), ANA and the ACR 1997 immunological disorder (modified to include anti-β2-glycoprotein antibodies) had the strongest positive association with SLE (figure 2A, online supplemental figure S1). Using a validation cohort of 512 clinically diagnosed patients with SLE and 143 disease controls to confirm our model, we noted excellent ability to discriminate true positive (SLE) versus false positive (control) cases with an area under the ROC curve (AUC) of 0.981 (figure 2B).

The new LR model enables SLE risk stratification into distinct diagnostic certainty levels

To determine how our model could be used in clinical practice, we applied the LR equation to generate SLE risk probabilities ranging 0%–100%, depending on the combination of features/predictors. We reasoned that different ranges of probabilities correspond to varying diagnostic certainty levels alike clinical thinking. For this, we calculated the SLE risk probabilities for all patients in the discovery cohort followed by unsupervised k-means clustering to detect unbiased risk probabilities partitions. Following merging of the closely related clusters C and D (online supplemental figure S2), we obtained four groups of increasing risk probability bins (0%–14%, 15%–43%, 44%–86%, 87%–100%).

Next, we used the validation cohort to determine the proportion of actual SLE and control patients captured within each predicted SLE risk group (figure 3A). Results were averaged from randomly generated, non-overlapping patient subsets (seven subsets each containing 73 or 74 patients with SLE, two subsets containing 71 and 72 disease controls). We confirmed the high discriminating capacity of our model as the majority of control (80%) and patients with SLE (82%) were in the lowest (0%–14%) and highest (87%–100%) risk groups, respectively. Concordantly, accuracy was highest in the two extreme risk groups but dropped in the intermediate ones. Thus, about 21% and 71% of the validation cohort patients in the 15%–43% and 44%–86% risk groups, respectively, had clinical SLE (figure 3B).

Using the upper limit probabilities of the risk groups as diagnostic thresholds (>14%, >43%, >86%), the average positive likelihood ratios (LR) for SLE were 5.0, 13.8 and 58.4, respectively, corresponding to moderate, large and very large increases in the likelihood of SLE (figure 3C). The >14% threshold had a negative LR 0.017, suggesting it can be used to exclude SLE against competing diseases with relatively high certainty. Taken together, we can assign the groups 'definitive SLE', 'likely SLE', 'possible/cannot rule out SLE' and 'unlikely SLE' to our model probability bins 87%–100%, 44%–86%, 15%–43% and 0%–14%, respectively. To put these data into clinical context, figure 3D illustrates matrices of predicted SLE risk probabilities based on combinations of various features.

Next, we examined the criterion validity of our model by determining its predictive ability against disease-relevant outcomes in the validation SLE cohort. Patients' risk probabilities correlated positively with increasing disease severity (p<0.0001) and organ damage (p=0.0019) (figure 3E), reflecting increased disease burden. Likewise, patients with SLE with low predicted risk probabilities (0%–14%, 15%–43%) had milder forms of the disease due to lower prevalence of British Isles Lupus Assessment Group (BILAG) A manifestations and organ damage (online supplemental figure S3A-B).

The SLERPI has high accuracy for detecting SLE including patients with early disease and severe disease requiring potent treatment

In addition to continuous risk prediction, binary outcome models (disease of interest is *present or absent*) are most helpful in decision-making. We used the discovery cohort for the unbiased definition of the model probability cut-off to separate SLE versus other rheumatological diseases. Based on the maximal Youden's statistics, the 50% risk probability threshold was chosen (online supplemental figure S4A-C). At this threshold, the SLERPI demonstrated high sensitivity (95.1%), specificity (93.7%) and accuracy (94.8%, corrected to 93.9% based on an expected 3:17 ratio of SLE: controls in real-life setting) in the total validation cohort (figure 4A, (online supplemental figure S5A). When tested against the control subset with undifferentiated connective tissue disease (n=56), the model specificity was 91.1%. We further determined the model discriminative ability in disease subsets of clinical relevance such as early SLE, lupus nephritis, neuropsychiatric SLE(NPSLE) and severe disease necessitating potent immunosuppressive or biological treatment. The model yielded very high rates of correct predictions within the aforementioned patient groups (figure 4B, online supplemental figure S5B).

Finally, to facilitate its implementation in daily practice, we converted our model into a simple scoring system (table 1). The scoring system-generated SLE probabilities showed high correlation with the risk probabilities produced by the original LR model (r^2 0.996) in the validation cohort. When operated at a threshold of >7 (out of maximum score 30.5), the sensitivity, specificity and accuracy were estimated at 94.2%, 94.4% and 94.2%, respectively, suggesting comparable performance with the original model.

DISCUSSION

Herein, we have developed and validated a simple, clinically applicable model to assist SLE diagnosis through ML training of well-characterised data from two large discovery and validation patient cohorts. Our model comprising 14 classical, variably weighted features, enables continuous risk prediction for
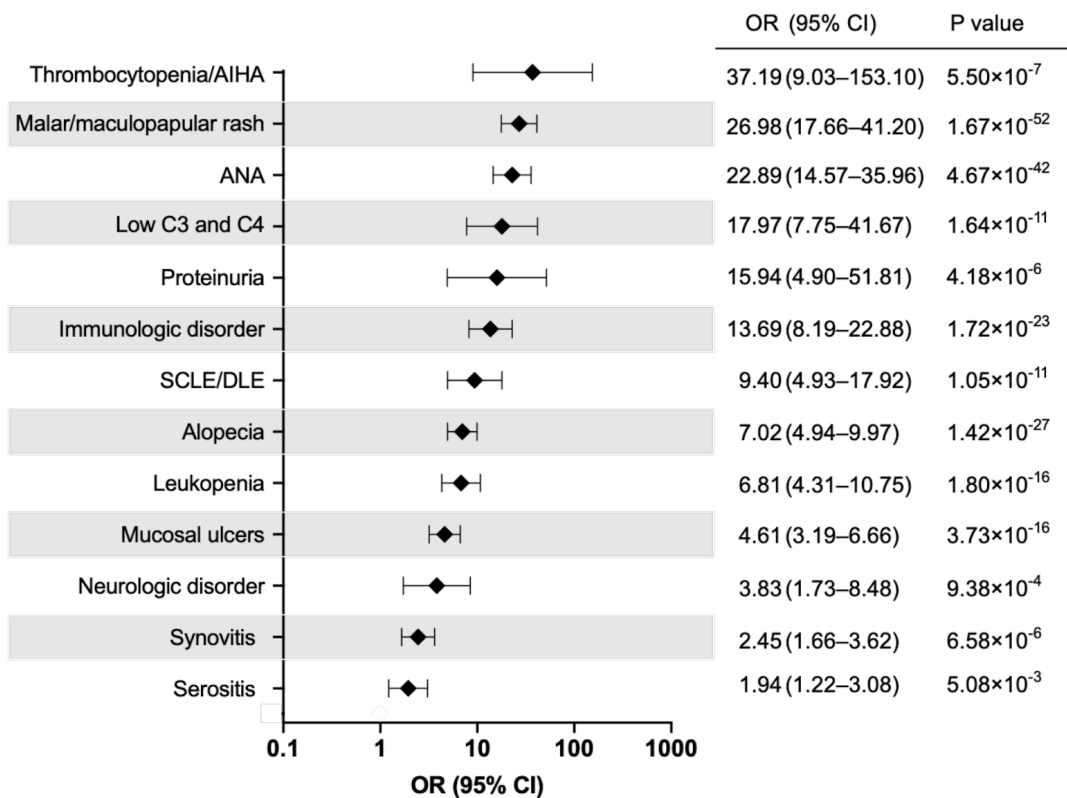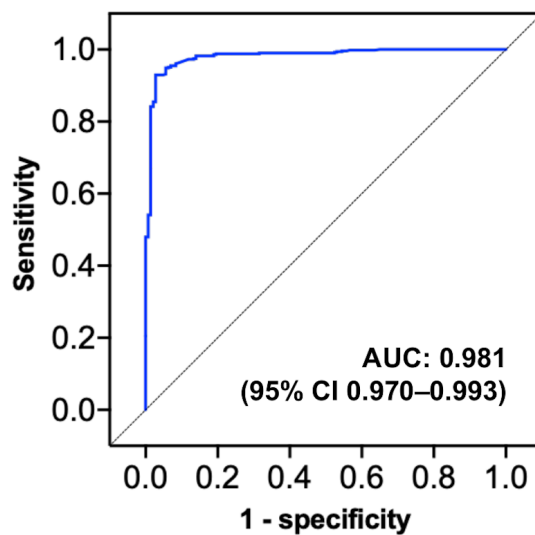
Figure 2 A Least Absolute Shrinkage and Selection Operator-logistic regression (LASSO-LR) model shows high discriminating capacity for SLE against competing rheumatological diseases. (A) A LASSO-LR model comprising of 14 clinical and serological features showed the highest accuracy for SLE in the 10-fold cross-validation runs from the *discovery cohort*. The plot illustrates the features associated with increased likelihood for SLE as compared with control rheumatological diseases along with the corresponding effect sizes (OR; 95% CI, p value). All model parameters are treated as dichotomous (ie, present=1, absent=0) in the LR equation as follows: *F(x)=Intercept + (1.80×mucosal ulcers [1]) + (2.96×synovitis[1]) + (1.83×serositis[1]) + (3.66×immunologic disorder[2]) + (4.42×antinuclear antibodies (ANA)[3]) + (2.13×alopecia[4]) + (2.17×neurologic disorder[4]) + (4.25×malar and/or maculopapular rash[3]) + (2.58×subacute cutaneous lupus erythematosus (SCLE) and/or discoid lupus erythematosus (DLE)[3]) + (1.82×leucopenia[3]) + (6.46×thrombocytopenia and/or autoimmune haemolytic anaemia (AIHA)[3]) + (6.63×low C3 and C4[3]) – (1.45×interstitial lung disease (ILD)* [5]); [1]defined according to the ACR 1997 classification criteria, [2]defined according to the ACR 1997 criteria modified to include also positive anti-β2 glycoprotein IgG or IgM antibodies, [3]defined according to the EULAR/ACR 2019 classification criteria, [4]defined according to the SLICC 2012 classification criteria, [5]see online supplemental table S2) for definition. (B) The LASSO-LR model presented in (A) was further evaluated in an external (validation) cohort of patients with 512 patients with SLE and 143 disease controls. The graph represents the receiver operating curve with a calculated area under the curve of 0.981 indicating an excellent capacity of the model to discriminate SLE versus disease controls.
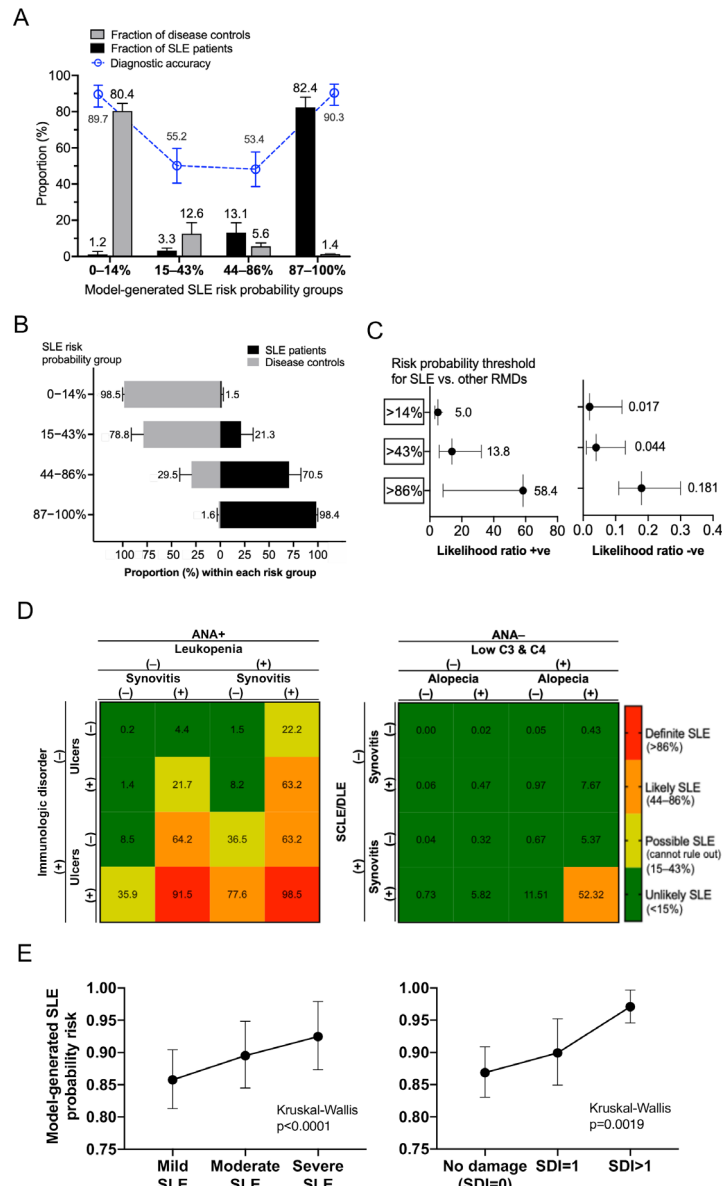
Figure 3 The Least Absolute Shrinkage and Selection Operator-logistic regression (LASSO-LR) model can generate SLE risk probabilities, which correspond to distinct diagnostic certainty levels and correlate with disease outcomes. (A) Bar plot representation of the fraction of patients with SLE patients and disease controls (*validation cohort*) according to increasing bins of predicted SLE *risk probabilities* (0%–14%, 15%–43%, 44%–86%, 87%–100%) calculated by the LASSO-LR model shown in figure 2. Superimposed are the diagnostic accuracies (blue-coloured) corresponding to the rates of correct classification of disease controls against patients with SLE in the lower two probability bins (0%–14%, 15%–43%), and of patients with SLE against disease controls in the higher two probability bins (44%–86%, 87%–100%). Results are averages (±SD for patient fractions or 95% CI for the accuracy metric) calculated from randomly generated, non-overlapping subsets of patients with SLE (seven subsets each containing 73 or 74 patients) and disease controls (two subsets containing 71 and 72 patients) from the validation cohort. The majority of control (average 80%) and SLE (average 82%) patients belong to the lowest (0%–14%) and the highest (87%–100%) *risk probability* groups, respectively. In accordance, accuracy was highest in these two extreme risk groups but dropped in the intermediate ones (15%–43%, 44%–86%). (B) Bar plot representation of the relative proportion of SLE and disease controls (*validation cohort*) within each SLE *risk probability* bin (0%–14%, 15%–43%, 44%–86%, 87%–100%). Calculations were made from the non-overlapping subsets of patients with SLE and disease controls as outlined in (A). (C) Positive- and negative-likelihood ratios (LRs) (mean, 95% CI) for the diagnosis of SLE against control diagnoses, according to different SLE *risk probability* thresholds (>14%,>43%,>86%) applied to the *discovery cohort*. Calculations were made from the non-overlapping subsets of patients with SLE and disease controls as outlined in (A). The >14% threshold had an average LR+5.0 and LR−0.017, which correspond to a *moderate increase* when tested *positive* and a *large decrease* when tested *negative* in the likelihood for SLE, respectively. (D) Matrices of SLE *risk probabilities* based on different combinations of features included in the LASSO-LR diagnostic model. In each scenario, the calculated probability fits to one of the four SLE risk groups corresponding to varying diagnostic certainty levels (*unlikely SLE*: 0%–14%, *possible/cannot rule out SLE*: 15%–43%, *likely SLE*: 44%–86%, *definite SLE*: 87%–100%). (E) Dot plot analysis of the model-generated SLE risk probabilities according to the severity of disease manifestations (defined based on the BILAG system) and organ damage (SLICC/ACR Damage Index (SDI)). Data were generated from the *validation cohort* patients with SLE (n=512) and are presented as mean (95% CI). The Kruskal-Wallis (non-parametric) analysis of variance test was performed and two-tailed p values are shown. ANA, antinuclear antibodies; RMDs, rheumatic diseases; SCLE, subacute cutaneous lupus erythematosus; SDI, SLICC/ACRdamage index; SLE, systemic lupus erythematosus.

A

		Actual cases	
		SLE	**Control diseases**
Predicted cases	**SLE**	487	9
	Control diseases	25	134

Metric	Value (95% CI)
Sensitivity	0.951 (0.928-0.968)
Specificity	0.937 (0.880-0.969)
Accuracy	0.948 (0.928-0.964)
Likelihood ratio +ve	15.11 (8.03-28.45)
Likelihood ratio –ve	0.052 (0.036-0.076)

B

Disease subgroups	Sensitivity (95% CI)
Early SLE (n=276)	0.938 (0.903-0.964)
Lupus nephritis (n=48)	0.979 (0.891-0.996)
Neuropsychiatric lupus (n=61)	0.918 (0.822-0.965)
Hematological lupus (n=74)	0.986 (0.927-0.998)
Severe SLE (n=110)	0.964 (0.910-0.986)

Figure 4 The new diagnostic model has high accuracy for systemic lupus erythematosus (SLE) including early and severe disease requiring immunosuppressive or biologic treatment. (A) Confusion matrix of the actual versus predicted cases of patients with SLE (n=512) and disease controls (n=143) in the *validation cohort*. The LASSO-LR diagnostic model was operated as binary (SLE or not-SLE) by setting the SLE *risk probability* threshold at ≥50%. Based on the number of true-positive, true-negative, false-positive and false-negative cases, sensitivity, specificity, accuracy, positive- and negative-likelihood ratios are estimated as metrics of the model diagnostic performance. (B) Sensitivity of the LASSO-LR model (operated as binary) for the detection of clinically relevant subsets of SLE including early disease, lupus nephritis, neuropsychiatric lupus, haematological lupus and severe lupus requiring potent immunosuppressive and/or biologic treatment.

clinical SLE, thus resembling clinical reasoning, while attaining a combination of high sensitivity and specificity against alternative rheumatologic diseases. When used as a dichotomous algorithm (SLE-or-not), the SLERPI exhibits high accuracy for SLE, including early and severe/organ-threatening disease forms.

In clinical practice, physicians can elicit the diagnosis of SLE even in the presence of a few high-yield manifestations such as typical malar rash in an individual with anti-DNA autoantibodies.[1 39] Such decisions reflect a form of human intelligence that develops through clinical experience even with a limited number of patients. Conversely, computational intelligence tools require training on large comprehensive data sets to produce valid results.[14] We used a discovery sample of well-characterised SLE and control patients for unbiased selection of features that contribute most to clinical SLE diagnosis. Patients with SLE with relatively early disease (median duration 4.2 years) and irrespective of the severity of manifestations were included, as compared with developing classification criteria, which typically rely on cases with long-standing disease.

Thrombocytopenia/autoimmune haemolytic anaemia (AIHA), malar rash, proteinuria, ANA, immunological disorder (anti-DNA, anti-Sm, anti-phospholipid antibodies) and combined C3 and C4 hypocomplementemia were the strongest predictors against competing rheumatological diseases. These results are in line with the variably weighted items introduced in the EULAR/ACR 2019 classification criteria, where, for example, thrombocytopenia/AIHA is scored higher than leucopenia and malar rash higher than other rashes.[9 10 40]

SLE displays marked phenotypic heterogeneity ranging from systemic to organ-limited/dominant forms. Clinical and immunological features may accrue sequentially in time, thus reflecting an evolving process.[41 42] Indeed, various terms have been used to describe different patient profiles such as 'definitive SLE', 'probable SLE', 'possible SLE', 'lupus-like' or 'incomplete lupus'. Our model calculates risk probabilities, which correlate with certainty levels for the presence of SLE versus competing rheumatological diseases. Based on unsupervised clustering, we hereby propose a risk probability-based stratification of patients with suspected SLE into 'unlikely', 'possible' (cannot rule out), 'likely' and 'definitive' SLE, depending on the type and number of features. This approach resembles diagnostic reasoning especially when encountering a patient for the first time.[43 44] Our model can be used not only to exclude (when risk probability is <14%) or confirm (when risk probability exceeds 86%) SLE but also to alert physicians to identify and monitor patients with intermediate probabilities. Similar approaches have been used in other complex diseases.[45]

By operating our model as binary, we achieved very high rates of sensitivity, specificity and accuracy assessed in a validation cohort. Our model can identify SLE under different clinical scenarios such as: (a) lupus autoantibodies concurring with a single clinical feature from a major organ (eg, thrombocytopenia/AIHA), (b) multiple clinical but no immunological features, (c) limited or non-specific serological features (eg, ANA) concurring with high-yield clinical manifestations (eg, malar rash). We noted excellent performance within patient subgroups with early disease, biopsy-proven lupus nephritis, neuropsychiatric disease and severe disease necessitating potent immunosuppressive or biologic therapies.

Table 1 A simple scoring system version of the SLE Risk Probability Index*

Feature	Score
Malar rash or maculopapular rash†	3
Subacute cutaneous lupus erythematosus or discoid lupus erythematosus†	2
Alopecia‡	1.5
Mucosal ulcers§	1
Arthritis§	2
Serositis§	1.5
Leucopenia<4000/µL (at least once)†	1.5
Thrombocytopenia or autoimmune haemolytic anaemia†	4.5
Neurological disorder‡	1.5
Proteinuria>500 mg/24 hours†	4.5
ANA†	3
Low C3 and C4†	2
Immunological disorder (any of: anti-DNA, anti-Sm, anti-phospholipid antibodies)¶	2.5
Interstitial lung disease**	−1
SLE if total score >7††	

*Apply the model in individuals with clinical suspicion for SLE. Each feature is counted if present (ever) and if not explained by other cause (eg, drug effects, infections, malignant disorders, alternative more likely disease).
†Defined as in Aringer et al.[9 10]
‡Defined as in Petri et al.[8]
§Defined as in Hochberg.[27]
¶Defined as in Hochberg[27] modified to include also positive anti-β2 glycoprotein IgG or IgM or IgA antibodies.
**Radiologic features of lung disease suggesting inflammation and fibrosis of the alveoli, distal airways and septal interstitium of the lung, as observed with a high-resolution CT scan of the chest.
††When operated at a threshold (sum of individual scores) of >7 (out of a maximum value 30.5), the sensitivity, specificity and accuracy rates are 94.2%, 94.4% and 94.2%, respectively.
ANA, antinuclear antibodies; SLE, systemic lupus erythematosus.

ML-based tools are increasingly used to simulate human 'medical reasoning' and effectively handle complex tasks. Such models can be trained from many different kinds of medical or biological data. Our data sets included well-defined features derived from the three classification criteria, and also non-criteria features often considered by physicians in cases of suspected SLE. ILD was a feature alienating the probability of SLE while favouring alternative rheumatological disease. Integration of additional clinical, laboratory or biological (eg, transcriptome) variables could lead to the development of even more robust models.[46 47] The fact that our model comprises 14 classical, easily retrieved clinical variables facilitates its clinical implementation.

Additional studies should prospectively evaluate and independently validate the proposed model to establish its clinical utility and effect on a variety of patient and healthcare outcomes. Notwithstanding, our analysis might provide useful insights towards the possible future development of formal SLE diagnostic criteria, a currently unmet need.[39] To this end, establishing a firm diagnosis and treatment plan still remains at the judgement of experienced physicians.

Our study is limited by its retrospective design and data extraction from medical records; accordingly, some clinical information may have been missed or underestimated. Nonetheless, both centres maintain detailed patient registries and use structured forms for collecting clinical data, which helps to reduce possible information/data completeness bias. Developing a model for early diagnosis should ideally be based on cohorts with very early disease and before the appearance of adverse outcomes, however, it can be challenging to recruit large numbers of such cases.[3 48] Although we used two state-of-the art ML approaches, a number of other sophisticated algorithms of higher complexity exist (eg, deep neural networks[49]). Our model also requires validation in additional cohorts of diverse population characteristics (eg, non-Caucasians), including infectious disease controls.

Conclusively, we have developed and evaluated a new, simple and interpretable model for the detection of SLE based on common clinical and serological features. Our model provides risk predictions that correlate with clinical endpoints and support patient probabilistic disease classification of potential clinical relevance. Pending further confirmation of its performance, the SLERPI could assist the early diagnosis and treatment of SLE, including early and severe forms, to improve patient outcomes.

Author affiliations
[1]Rheumatology, Clinical Immunology and Allergy, University of Crete School of Medicine, Heraklion, Crete, Greece
[2]Rheumatology and Clinical Immunology Unit, 4th Department of Internal Medicine, Attikon University Hospital, National and Kapodistrian University of Athens, Athens, Greece
[3]Section of Rheumatology, Department of Medical Sciences, Azienda Ospedaliero Universitaria di Ferrara Arcispedale Sant'Anna, Cona, Emilia-Romagna, Italy
[4]Rheumatology, "Asklepieion" General Hospital, Athens, Greece
[5]Institute of Molecular Biology and Biotechnology, Foundation of Research and Technology—Hellas, Heraklion, Crete, Greece
[6]Laboratory of Immune Regulation and Tolerance, Autoimmunity and Inflammation, Biomedical Research Foundation of the Academy of Athens, Athens, Attica, Greece

Twitter Dimitrios T Boumpas @none and George K Bertsias @george_bertsias

ORCID iDs
Dionysis Nikolopoulos http://orcid.org/0000-0002-9894-6966
Antonis Fanouriakis http://orcid.org/0000-0003-2696-031X
Prodromos Sidiropoulos http://orcid.org/0000-0001-9607-0326
Dimitrios T Boumpas http://orcid.org/0000-0002-9812-4671
George K Bertsias http://orcid.org/0000-0001-5299-1406

REFERENCES

1 Doria A, Zen M, Canova M, *et al*. Sle diagnosis and treatment: when early is early. *Autoimmun Rev* 2010;10:55–60.

2 Feng X, Zou Y, Pan W, *et al*. Associations of clinical features and prognosis with age at disease onset in patients with systemic lupus erythematosus. *Lupus* 2014;23:327–34.

3 Morgan C, Bland AR, Maker C, *et al*. Individuals living with lupus: findings from the lupus UK members survey 2014. *Lupus* 2018;27:681–7.

4 Oglesby A, Korves C, Laliberté F, *et al*. Impact of early versus late systemic lupus erythematosus diagnosis on clinical and economic outcomes. *Appl Health Econ Health Policy* 2014;12:179–90.

5 Faurschou M, Dreyer L, Kamper A-L, *et al*. Long-Term mortality and renal outcome in a cohort of 100 patients with lupus nephritis. *Arthritis Care Res* 2010;62:873–80.

6 Gergianaki I, Bertsias G. Systemic lupus erythematosus in primary care: an update and practical messages for the general practitioner. *Front Med* 2018;5:161.

7 Aggarwal R, Ringold S, Khanna D, *et al*. Distinctions between diagnostic and classification criteria? *Arthritis Care Res* 2015;67:891–7.

8 Petri M, Orbai A-M, Alarcón GS, *et al*. Derivation and validation of the systemic lupus international collaborating clinics classification criteria for systemic lupus erythematosus. *Arthritis Rheum* 2012;64:2677–86.

9 Aringer M, Costenbader K, Daikh D, *et al*. European League against Rheumatism/American College of rheumatology classification criteria for systemic lupus erythematosus. *Arthritis Rheumatol* 2019;2019:1400–12.

10 Aringer M, Costenbader K, Daikh D, *et al*. European League against Rheumatism/American College of rheumatology classification criteria for systemic lupus erythematosus. *Ann Rheum Dis* 2019;2019:1151–9.

11 Adamichou C, Nikolopoulos D, Genitsaridi I, *et al*. In an early SLE cohort the ACR-1997, SLICC-2012 and EULAR/ACR-2019 criteria classify non-overlapping groups of patients: use of all three criteria ensures optimal capture for clinical studies while their modification earlier classification and treatment. *Ann Rheum Dis* 2020;79:232–41.

12 Pons-Estel GJ, Ugarte-Gil MF, Harvey GB, *et al*. Applying the 2019 EULAR/ACR lupus criteria to patients from an established cohort: a Latin American perspective. *RMD Open* 2020;6:e001097.

13 Pelaccia T, Forestier G, Wemmert C. Deconstructing the diagnostic reasoning of human versus artificial intelligence. *CMAJ* 2019;191:E1332–5.

14 Rajkomar A, Dean J, Kohane I. Machine learning in medicine. *N Engl J Med* 2019;380:1347–58.

15 Pinal-Fernandez I, Casal-Dominguez M, Derfoul A, *et al*. Machine learning algorithms reveal unique gene expression profiles in muscle biopsies from patients with different types of myositis. *Ann Rheum Dis* 2020;79:1234–42.

16 Orange DE, Agius P, DiCarlo EF, *et al*. Identification of three rheumatoid arthritis disease subtypes by machine learning integration of synovial histologic features and RNA sequencing data. *Arthritis Rheumatol* 2018;70:690–701.

17 Lezcano-Valverde JM, Salazar F, León L, *et al*. Development and validation of a multivariate predictive model for rheumatoid arthritis mortality using a machine learning approach. *Sci Rep* 2017;7:10189.

18 Guan Y, Zhang H, Quang D, *et al*. Machine learning to predict anti-tumor necrosis factor drug responses of rheumatoid arthritis patients by integrating clinical and genetic markers. *Arthritis Rheumatol* 2019;71:1987–96.

19 Kegerreis B, Catalina MD, Bachali P, *et al*. Machine learning approaches to predict lupus disease activity from gene expression data. *Sci Rep* 2019;9:9617.

20 Huang Z, Shi Y, Cai B, *et al*. 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;48:626–31.

21 Walsh JA, Rozycki M, Yi E, *et al*. Application of machine learning in the diagnosis of axial spondyloarthritis. *Curr Opin Rheumatol* 2019;31:362–7.

22 Watson DS, Krutzinna J, Bruce IN, *et al*. Clinical applications of machine learning algorithms: beyond the black box. *BMJ* 2019;364:l886.

23 Gao C, Sun H, Wang T, *et al*. Model-Based and model-free machine learning techniques for diagnostic prediction and classification of clinical outcomes in Parkinson's disease. *Sci Rep* 2018;8:7129.

24 Gergianaki I, Fanouriakis A, Repa A, *et al*. Epidemiology and burden of systemic lupus erythematosus in a southern European population: data from the community-based lupus Registry of Crete, Greece. *Ann Rheum Dis* 2017;76:1992–2000.

25 Nikolopoulos D, Kostopoulou M, Pieta A, *et al*. Evolving phenotype of systemic lupus erythematosus in Caucasians: low incidence of lupus nephritis, high burden of neuropsychiatric disease and increased rates of late-onset lupus in the 'Attikon' cohort. *Lupus* 2020;29:514–22.

26 Nikolopoulos DS, Kostopoulou M, Pieta A, *et al*. Transition to severe phenotype in systemic lupus erythematosus initially presenting with non-severe disease: implications for the management of early disease. *Lupus Sci Med* 2020;7:e000394.

27 Hochberg MC. Updating the American College of rheumatology revised criteria for the classification of systemic lupus erythematosus. *Arthritis Rheum* 1997;40:40.

28 Papadaki E, Fanouriakis A, Kavroulakis E, *et al*. Neuropsychiatric lupus or not? cerebral hypoperfusion by perfusion-weighted MRI in normal-appearing white matter in primary neuropsychiatric lupus erythematosus. *Ann Rheum Dis* 2018;77:441–8.

29 Bortoluzzi A, Scirè CA, Bombardieri S, *et al*. Development and validation of a new algorithm for Attribution of neuropsychiatric events in systemic lupus erythematosus. *Rheumatology* 2015;54:891–8.

30 Isenberg DA, Rahman A, Allen E, *et al*. BILAG 2004. development and initial validation of an updated version of the British Isles lupus assessment group's disease activity index for patients with systemic lupus erythematosus. *Rheumatology* 2005;44:902–6.

31 Gladman D, Ginzler E, Goldsmith C, *et al*. The development and initial validation of the systemic lupus international collaborating Clinics/American College of rheumatology damage index for systemic lupus erythematosus. *Arthritis Rheum* 1996;39:363–9.

32 Aberle T, Bourn RL, Chen H, *et al*. Use of SLICC criteria in a large, diverse lupus registry enables SLE classification of a subset of ACR-designated subjects with incomplete lupus. *Lupus Sci Med* 2017;4:e000176.

33 Amezcua-Guerra LM, Higuera-Ortiz V, Arteaga-García U, *et al*. Performance of the 2012 systemic lupus international collaborating clinics and the 1997 American College of rheumatology classification criteria for systemic lupus erythematosus in a real-life scenario. *Arthritis Care Res* 2015;67:437–41.

34 Hartman EAR, van Royen-Kerkhof A, Jacobs JWG, *et al*. Performance of the 2012 systemic lupus international collaborating clinics classification criteria versus the 1997 American College of rheumatology classification criteria in adult and juvenile systemic lupus erythematosus. A systematic review and meta-analysis. *Autoimmun Rev* 2018;17:316–22.

35 Ighe A, Dahlström Örjan, Skogh T, *et al*. Application of the 2012 systemic lupus international collaborating clinics classification criteria to patients in a regional Swedish systemic lupus erythematosus register. *Arthritis Res Ther* 2015;17:3.

36 Inês L, Silva C, Galindo M, *et al*. Classification of systemic lupus erythematosus: systemic lupus international collaborating clinics versus American College of rheumatology criteria. A comparative study of 2,055 patients from a real-life, International systemic lupus erythematosus cohort. *Arthritis Care Res* 2015;67:1180–5.

37 Pons-Estel GJ, Wojdyla D, McGwin G, *et al*. The American College of rheumatology and the systemic lupus international collaborating clinics classification criteria for systemic lupus erythematosus in two multiethnic cohorts: a commentary. *Lupus* 2014;23:3–9.

38 Sacre K, Delaval L, Dossier A, *et al*. New 2019 SLE EULAR/ACR classification criteria are valid for identifying patients with SLE among patients admitted for pericardial effusion. *Ann Rheum Dis* 2019. doi:10.1136/annrheumdis-2019-216712. [Epub ahead of print: 06 Dec 2019].

39 Bertsias GK, Pamfil C, Fanouriakis A, *et al*. Diagnostic criteria for systemic lupus erythematosus: has the time come? *Nat Rev Rheumatol* 2013;9:687–94.

40 Tedeschi SK, Johnson SR, Boumpas DT, *et al*. Multicriteria decision analysis process to develop new classification criteria for systemic lupus erythematosus. *Ann Rheum Dis* 2019;78:634–40.

41 Lambers WM, Westra J, Jonkman MF, *et al*. Incomplete systemic lupus erythematosus: what remains after application of American College of rheumatology and systemic lupus international collaborating clinics criteria? *Arthritis Care Res* 2020;72:607–14.

42 Mosca M, Tani C, Carli L, *et al*. Undifferentiated CTD: a wide spectrum of autoimmune diseases. *Best Pract Res Clin Rheumatol* 2012;26:73–7.

43 Donner-Banzhoff N. Solving the diagnostic challenge: a patient-centered approach. *Ann Fam Med* 2018;16:353–8.

44 Soltani A, Moayyeri A. Deterministic versus evidence-based attitude towards clinical diagnosis. *J Eval Clin Pract* 2007;13:533–7.

45 Ntranos A, Lublin F, Criteria D. Diagnostic criteria, classification and treatment goals in multiple sclerosis: the Chronicles of time and space. *Curr Neurol Neurosci Rep* 2016;16:90.

46 Alarcón-Riquelme ME. New attempts to define and clarify lupus. *Curr Rheumatol Rep* 2019;21:11.

47 Panousis NI, Bertsias GK, Ongen H, *et al*. Combined genetic and transcriptome analysis of patients with SLE: distinct, targetable signatures for susceptibility and severity. *Ann Rheum Dis* 2019;78:1079–89.

48 Piga M, Floris A, Sebastiani GD, *et al*. Risk factors of damage in early diagnosed systemic lupus erythematosus: results of the Italian multicentre early lupus project inception cohort. *Rheumatology* 2020;59:2272–81.

49 Tomašev N, Glorot X, Rae JW, *et al*. A clinically applicable approach to continuous prediction of future acute kidney injury. *Nature* 2019;572:116–9.