Original Research Article

Construction and evaluation of glucocorticoid dose prediction model based on genetic and clinical characteristics of patients with systemic lupus erythematosus

Xin Luo[1,2]* 🆔, Jinjun Zhao[3]*, Danfeng Zou[4], Xiaoning Luo[1,2], Meida Fan[3], Hongling Hu[5], Ping Zheng[1,2], Yilei Li[1,2], Renfei Xia[6] and Liqian Mo[1,2]

Abstract

Currently, no glucocorticoid dose prediction model is available for clinical practice. This study aimed to utilise machine learning techniques to develop and validate personalised dosage models. Participants were patients with SLE who were registered at Nanfang Hospital and received prednisone. Univariate analysis was used to confirm the feature variables. Subsequently, the random forest (RF) algorithm was utilised to interpolate the absent values of the feature variables. Finally, we assessed the prediction capabilities of 11 machine learning and deep-learning algorithms (Logistic, SVM, RF, Adaboost, Bagging, XGBoost, LightGBM, CatBoost, MLP, and TabNet). Finally, a confusion matrix was used to validate the three regimens. In total, 129 patients met the inclusion criteria. The XGBoost algorithm was selected as the preferred method because of its superior performance, achieving an accuracy of 0.81. The factors exhibiting the highest correlation with the prednisone dose were CYP3A4 (rs4646437), albumin (ALB), haemoglobin (HGB), anti-double-stranded DNA antibodies (Anti-dsDNA), erythrocyte sedimentation rate (ESR), age, and HLA-DQA1 (rs2187668). Based on validation, the precision and recall rates for low-dose prednisone ($\geq$5 mg but <7.5 mg/d) were 100% and 40% respectively. Similarly, for medium-dose prednisone ($\geq$7.5 mg but <30 mg/d), the accuracy and recall rates were 88% and 88%, and for high-dose prednisone ($\geq$30 mg but $\leq$100 mg/d), the accuracy and recall rates were 62% and 100% respectively. A robust machine learning model was developed to accurately predict prednisone dosage by integrating the identified genetic and clinical factors.

[1]Department of Pharmacy, Nanfang Hospital, Southern Medical University, Guangzhou, China
[2]Clinical Pharmacy Center, Nanfang Hospital, Southern Medical University, Guangzhou, China
[3]Department of Rheumatology and Immunology, Nanfang Hospital, Southern Medical University, Guangzhou, China
[4]Overseas Patient Center, Nanfang Hospital, Southern Medical University, Guangzhou, China
[5]Department of Trauma and Joint Surgery, Shunde Hospital, Southern Medical University, Foshan, China
[6]Department of Transplantation, Nanfang Hospital, Southern Medical University, Guangzhou, China

*Xin Luo and Jinjun Zhao contributed equally to this work.

Corresponding authors:
Liqian Mo, Department of Pharmacy, Nanfang Hospital, Southern Medical University, Guangzhou, Guangdong, China.
Email: moliqian@smu.edu.cn

Renfei Xia, Department of Transplantation, Nanfang Hospital, Southern Medical University, Guangzhou, Guangdong, China.
Email: xiarf@smu.edu.cn

Introduction

Systemic lupus erythematosus (SLE) is an autoimmune inflammatory connective tissue disease affecting multiple organs. The global incidence of SLE is estimated to range from 1.5 to 11 per 100,000 individuals per year.[1] Glucocorticoids (GCs) continue to serve as a fundamental component in the management of SLE,[2] Despite the availability of immunosuppressive agents,[3] up to one third of patients never discontinue GCs.[4] GCs have pros and cons. GCs remain statistically linked to the occurrence of severe infections and delayed cardiovascular morbidity in SLE.[5,6] In addition, GCs exert deleterious effects on bone resorption[7] and even induce bone necrosis.[8] GCs are an independent predictor of decreased quality of life in SLE.[9] The harmful effects of glucocorticoids on chronic organ damage have recently been validated in recent studies.[10] Excessive immune activation caused by GCs may result in long-term inflammation. GCs encourage the growth of harmful T helper cells and lead to tissue damage by causing inflammation in the nerves and intestine.[11] The absence of GC therapy may be associated with the absence of chronic harm accumulation.[12] Based on the uncertainty of benefits versus risks, the optimal management of GCs in SLE is recognised as a significant contemporary challenge.[13]

The efficacy of a drug response is influenced by various factors, including pharmacogenomics and patient-specific attributes such as sex, age, and concurrent medication. Pharmacogenomics plays a pivotal role in personalised medicine.[14] Therefore, it is imperative to further investigate and validate the synthesis of these potential factors and their application in devising prednisone dose adjustment strategies. Machine learning is a promising avenue for integrating and leveraging data from patients[15] and exhibits superior generalisation and accuracy compared with traditional models.[16–18] Notably, recent advancements in algorithmic principles have led to the development of more sophisticated algorithms such as eXtreme Gradient Boosting (XGBoost), light gradient boosting machine (LightGBM), Categorical Boosting (CatBoost), Gradient Boosting Decision Tree (GBDT), and TabNet, which have received significant recognition within the algorithmic community.[19–23] Recently, a surge in the application of machine- and deep-learning methodologies in practical applications has been observed. A notable example is the development of a predictive model for tacrolimus blood concentration in patients with autoimmune disorders.[24] In addition, CatBoost was used to develop a pioneering model for sodium valproate dose prediction.[25] Numerous studies have demonstrated the advantages of machine-learning algorithms over traditional statistical techniques.

Despite the widespread utilisation of GCs, very few small-scale clinical trials[26–29] and observational studies[30–33] have been conducted to ascertain the optimal form, dosage, or administration regimen. Currently, there is no GCs dose prediction model that can provide guidance for clinical practice. In the Systemic Lupus International Collaborating Clinics Inception Cohort,[2] the use of GCs was primarily based on physicians' clinical practices rather than individual patient characteristics. To optimise the management of GCs, our study aimed to employ machine learning techniques to construct personalised dosage models for GCs.

Materials and methods

Study design and participants

This prospective observational study was conducted at NanFang Hospital, affiliated with Southern Medical University between 2017 and 2023, and enrolled adult patients with SLE diagnosed based on the 1997 American College of Rheumatology classification criteria. Every 3 months, the daily prednisone dosage and laboratory parameters were recorded, and SLE disease activity was assessed. During the one-year follow-up period, rheumatologists made adjustments to the GC therapy based on the daily prednisone dose and disease remission status. The study protocol was approved by the Ethics Committee of Nanfang Hospital, Southern Medical University (NFEC-2021-320) and all patients provided written informed consent.

The inclusion criteria for this study were: (1) initial doses of ≥5 mg/day at the baseline visit; (2) ability to complete 1 year of follow-up. The exclusion criteria were as follows: (1) initiation of GCs prior to study enrolment, (2) irregular medication, (3) intricate and severe coexisting medical conditions and a projected lifespan shorter than the duration of the study. Patients who met the inclusion criteria and none of the exclusion criteria underwent polygenic drug
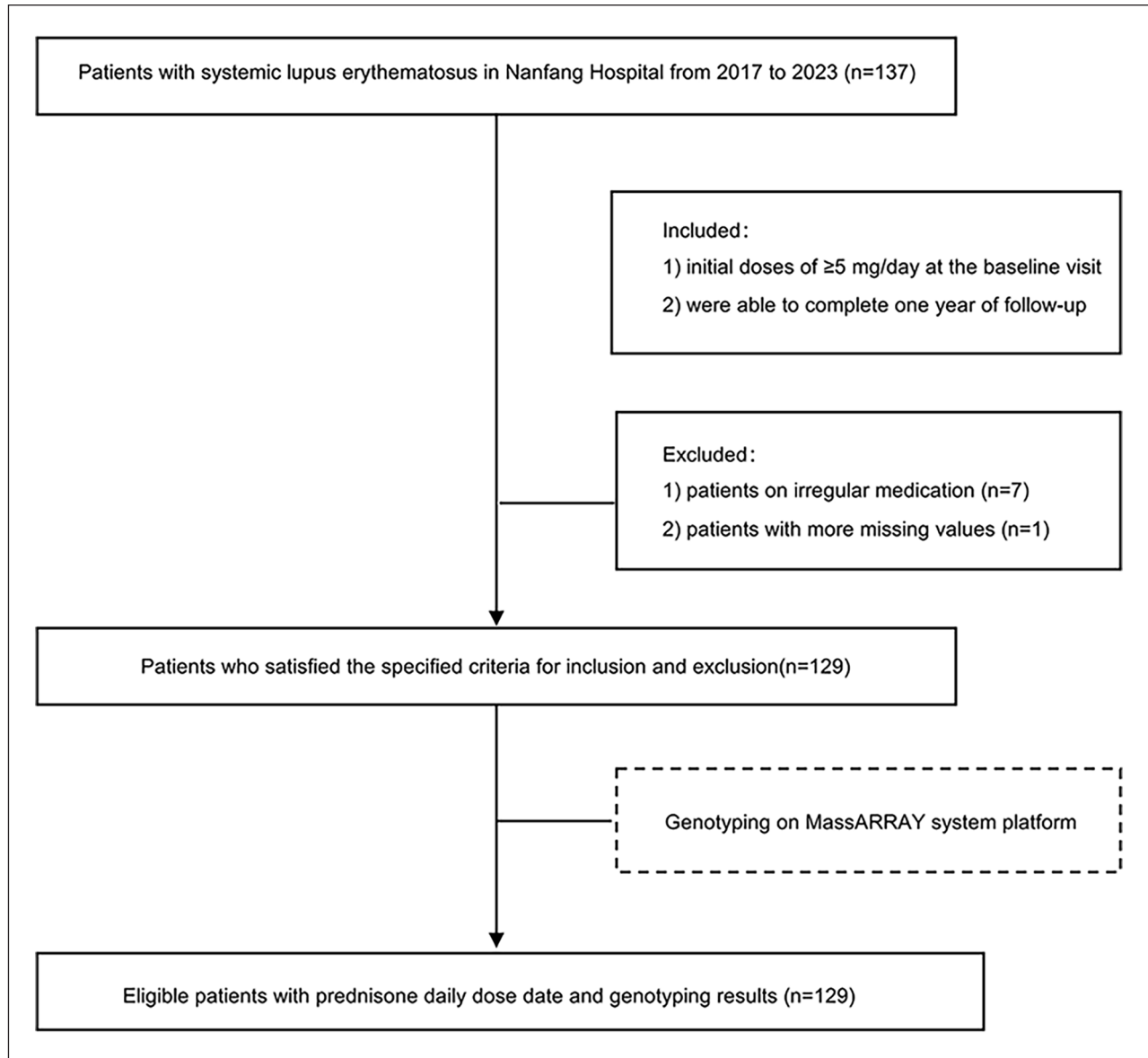
Figure 1. Enrolment of patients.

testing. They were then closely monitored and followed up for 1 year, during which relevant indicators of efficacy and safety were systematically collected and documented. Finally, 129 patients who received daily doses of prednisone were included in this study. The patient enrolment process is illustrated in Figure 1.

Patients and clinical assessment

The demographic characteristics, inspection information and combination of medication were recorded at baseline. Further, information collected included erythrocyte sedimentation rate (ESR), C-reactive protein (CRP), anti-nuclear antibodies (ANA), anti-double-stranded DNA antibodies (Anti-dsDNA), Sm antibodies (Anti-Sm), RNP antibodies (Anti-RNP), alanine aminotransferase (ALT), albumin (ALB), serum creatinine (Scr), uric acid (UA), white blood cell count (WBC), blood platelet count (PLT), haemoglobin (HGB), Ca, K, and glucose (Glu), total cholesterol (TC), triglyceride (TG), low density lipoprotein (LDL), high density lipoprotein (HDL). Data on disease activity according to the SLE disease activity index (SLEDAI) were collected. When SLEDAI $\leqslant 4$, the prednisone dose is appropriate and the treatment is effective.

GCs use

The average daily oral prednisone doses between visits were recorded. The oral average daily prednisone doses were categorised into three categories based on clinical expertise[31,34]: low (≥5 mg but <7.5 mg/d), medium (≥7.5 mg but <30 mg/d), and high (≥30 mg but ≤100 mg/d).

Genotyping

Drug-related target genes were identified by conducting a comprehensive search of the PharmGKB, Clinical Pharmacogenetics Implementation Consortium (CPIC), and PubMed databases to identify medications commonly used for the treatment of SLE. A total of 45 single nucleotide polymorphisms (SNPs) in the target genes (Supplemental Table S1) were assessed. The primer design was completed and validated by Simcere Diagnostics. Venous blood samples were collected from the participants and genomic DNA was isolated. The amplified samples were analysed using a computer-based assay specifically using the MassARRAY system platform. The specimens were collected on a microchip and analysed. Specimens were obtained from the microchip and subsequently subjected to scanning. The resulting scan data were processed using Typer4.0 software. Time-flight mass spectrometry was used to determine the mutation status of individual target genes by quantifying the nucleotide count at a specific mutation site. A peak was classified as a mutation if it was in close proximity to the base of the mass spectrum peak.

Variable selection and model establishment

The modelling process is illustrated in Figure 2. To ensure that the study design for the three-category problem was robust and to enhance the reliability of the results, we selected Cohen's *w* to calculate the required sample size. Classification variables such as sex, medication combination, and genetic type were encoded using heat as the sole method. To address extreme imbalances, classification variables with a single category comprising more than 90% of the total should be removed. In addition, variables with more than 20% missing data were excluded. To ascertain the pertinent factors associated with prednisone dosage, feature screening was performed using univariate differential analysis. It

is imperative to train both linear and nonlinear models to ensure data precision. Next, the implementation of single-factor feature screening was performed. It was necessary to interpolate the missing values of the seven variables to estimate the dosage for patients diagnosed with SLE. The random forest method (RF) proved effective in managing the complex relationships among numerous characteristics, thus facilitating the interpolation of missing values using RF. Statistical analysis of the modelling data after interpolation is presented in Supplemental Table S2.

The dataset was divided into a training cohort and a validation cohort in an 8:2 ratio, with the objective variable 'prednisone daily dose' and seven selected variables serving as covariates. Ten algorithms were constructed to predict the prednisone dose in patients with SLE. The training set was further divided into six segments, with five segments assigned for training and the sixth segment designated for verification. This process was repeated six times with each iteration involving the selection of a unique verification set. Finally, the performance of the model was evaluated by computing the mean of six validation outcomes. The selection grid search technique was used to identify the most favourable combination within the model's parameter space, resulting in the highest performance. Six-fold cross-validation was performed to evaluate the model performance and ascertain the optimal parameter combination for each scenario. The specifics of the model and validation optimisation parameters are presented in Supplemental Tables S3 and S4.

Statistical analysis

Patient characteristics are reported as median (IQR), depending on the degree of similarity to a normal distribution. Categorical variables were analysed, and the results included both absolute and relative frequencies. The Mann–Whitney *U* test was used to perform univariate analysis of continuous variables. The chi-squared test was used to assess categorical data when the sample size was ≥5, whereas Fisher's exact test was used when the sample size was <5. A confusion matrix was generated using Sklearn software. The machine learning and deep learning algorithms were executed on a Windows 10 operating system utilising an Intel® Core™ i5-10400F CPU @ 2.90 GHz with 12 CPUs and 512 GB of RAM.
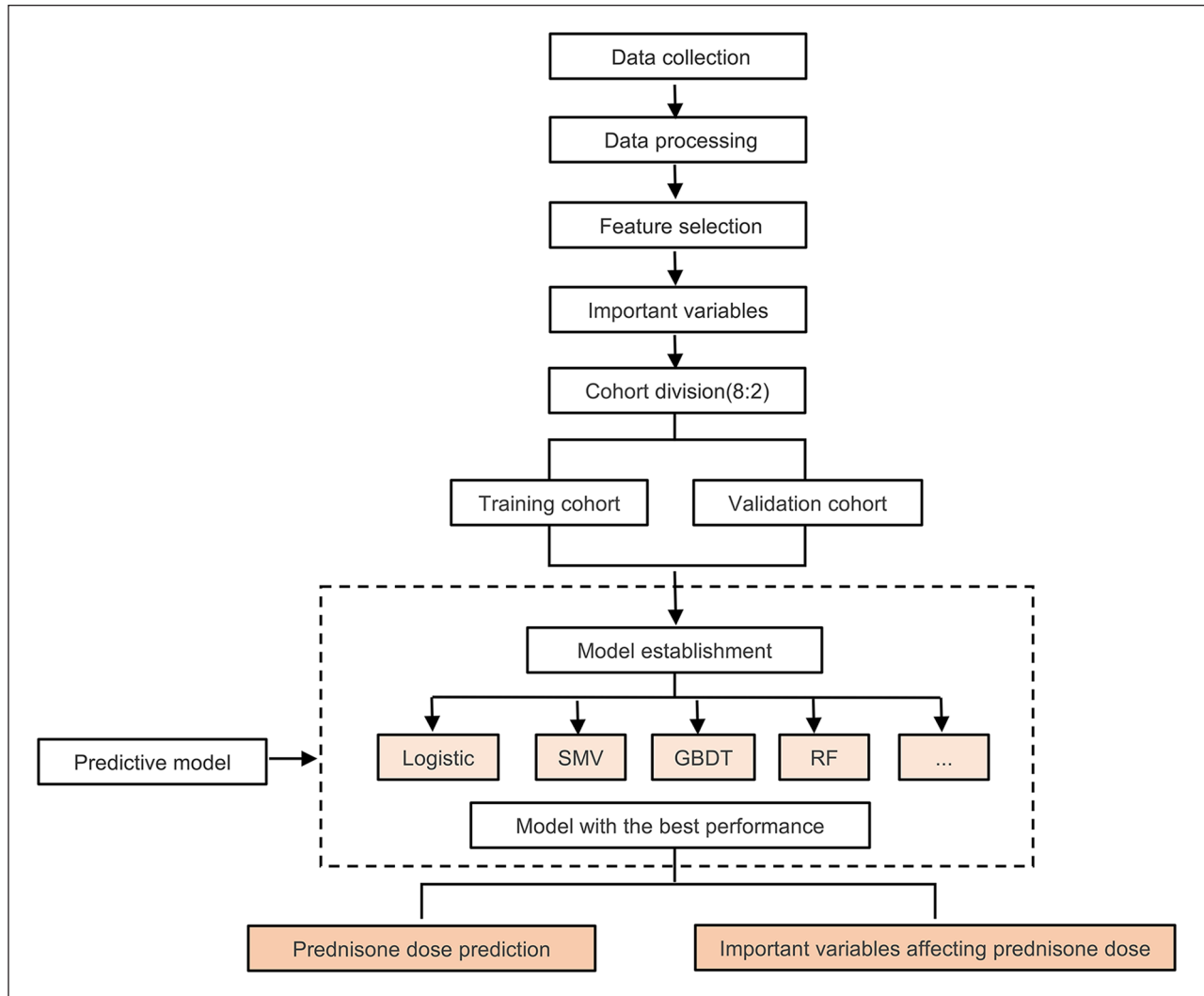
Figure 2. Workflow of data processing and model establishment.

Data analysis was performed using Python 3.8.8 and IBM SPSS Statistics 22. The significance threshold was defined as $P < 0.05$ (two-sided).

Results

Baseline information

In total, 129 patients diagnosed with SLE were enrolled in this study. Baseline characteristics and genotyping information are shown in Table 1 and Supplemental Table S5. The median age was 35 years (IQR 18–71), and 89.15% were female. Hydroxychloroquine (HCQ) was the most commonly used combination medication (80.62%). In addition, each patient was administered calcium supplements and proton pump inhibitors. In addition, the dataset provides information on the missing percentages for each variable. Regarding genotyping, only five SNPs (SLC19A1 rs1051266, ATIC rs4673993, ABCB1 rs2032582, CYP2D6 rs1065852, and ABCB1 rs1045642) had missing data with a miss rate of <2%. There remaining SNPs had no missing data.

Changes in daily prednisone dose and disease activity

After 12 months, among these patients, 13 (10.08%) received a daily prednisone dose of 5–7.4 mg, 69 (53.49%) received 7.5–29 mg, and 47 (36.43%) received 30–100 mg (Figure 3a). According to SLEDAI (Figure 3b), 129 patients had SLEDAI ≤ 4 and had stable disease control as assessed by a rheumatologist.

Table I. Demographic clinical characteristics.

Variables	Median (IQR) or n (%)	
Demographic information		
Age (y, median (IQR))	35 (18–71)	
Gender (n (%))		
Male	14 (10.85)	
Female	115 (89.15)	
Inspection information		
ESR (+)	24 (18.6)	
CRP (+)	12 (9.3)	
ANA (+)	88 (68.2)	
Anti-dsDNA (+)	95 (73.54)	
Anti-Sm (+)	46 (35.65)	
Anti-RNP (+)	34 (26.35)	
ALT (U/L, median (IQR))	15.50 (1.30–257.00)	
ALB (g/L, median (IQR))	35.50 (18.00–64.50)	
Scr (μmol/L, median (IQR))	61.00 (35.00–207.00)	
UA (μmol/L, median (IQR))	328.00 (130.00–764.00)	
WBC (×10^9/L, median (IQR))	5.80 (1.56–39.51)	
PLT (×10^9/L, median (IQR))	214.00 (1.00–501.00)	
HGB (g/L, median (IQR))	112.00 (51.00–163.00)	
Ca (mmol/L, median (IQR))	2.20 (1.78–224.00)	
K (mmol/L, median (IQR))	3.75 (2.63–5.66)	
Glu (median (IQR))	4.48 (1.32–8.59)	
TC (mmol/L, median (IQR))	4.20 (1.78–9.28)	
TG (mmol/L, median (IQR))	1.20 (0.41–6.41)	
LDL (mmol/L, median (IQR))	2.38 (0.95–5.69)	
HDL (mmol/L, median (IQR))	1.22 (0.40–2.75)	
Combination of medication		
Cyc	0	112 (86.82)
	1	17 (13.18)
MMF	0	90 (69.77)
	1	39 (30.23)
HCQ	0	25 (19.38)
	1	104 (80.62)
Belimumab	0	110 (85.27)
	1	19 (14.73)
Total glucosides of paeony	0	95 (73.64)
	1	34 (26.36)
Cordyceps militaris powder capsule	0	115 (89.15)
	1	14 (10.85)
Calcium supplements	1	129 (100)
Proton pump inhibitors	1	129 (100)

The use of medication corresponds to '1', non-use of medicines corresponds to '0'. Continuous variables were presented as 'mean (interquartile range, IQR)', while categorical variables were presented as 'frequency (percentage, %)'.
ESR: erythrocyte sedimentation rate; CRP: C-reactive protein; ANA: anti-nuclear antibodies; Anti-dsDNA: anti-double-stranded DNA antibodies; Anti-Sm: Sm antibodies; Anti-RNP: RNP antibodies; ALT: alanine aminotransferase; ALB: albumin; Scr: serum creatinine; UA: uric acid; WBC: white blood cell count; PLT: blood platelet count; HGB: hemoglobin; Glu: glucose; TC: total cholesterol; TG: triglyceride; LDL: low density lipoprotein; HDL: high-density lipoprotein; SLEDAI: systemic lupus erythematosus disease activity index; Cyc: cyclophosphamide; MMF: mycophenolate mofetil; HCQ: hydroxychloroquine.

Univariate analysis

Statistically significant associations ($P < 0.05$) between the univariate and prednisone daily dose variables were identified (Supplemental Table S6). Following a screening procedure, the seven variables in Table 2 were deemed significant and included in the subsequent processing and modelling stages.

Model performance

Ten algorithms were employed to establish the models, each of which was fitted with the optimal parameters for the training set. The performance of each model was assessed using various evaluation indicators, resulting in four model results: true negative (TN), false negative (FN), true positive (TP), and false positive (FP). Accuracy is frequently used to assess the overall performance of a model by predicting the proportion of correctly classified samples. The accuracy specifically pertains to the ratio of truly positive samples to all samples predicted to be positive. The recall denotes the accurate proportion of all positive samples predicted by the model. The F1 score is the harmonic mean of the accuracy and recall. The formulae are as follows:

$$Accuracy = (TP + TN) / (TP + FP + TN + FN)$$

$$Precision = TP / (TP + FP)$$

$$Recall = TP / (TP + FN)$$

$$F1 - score = (2 \times Precision \times Recall) / (Precision + Recall)$$

Table 3 presents the prediction performances of the nine models. XGBoost had Precision=1, Recall=0.4, and F1_score=0.57 for predicting the low-dose prednisone (≥ 5 mg but < 7.5 mg/d); Precision=0.88, Recall=0.88, and F1_score=0.88 for predicting the medium-dose prednisone (≥ 7.5 mg but < 30 mg/d); Precision=0.62, Recall=1, and F1_score=0.77 for predicting the high-dose prednisone (≥ 30 mg but ≤ 100 mg/d); and Accuracy=0.81 for the whole XGBoost model, surpassing other algorithms and exhibiting superior overall performance. Consequently, XGBoost was used to predict the daily dosage of prednisone and to evaluate the influence of various factors on dose prediction.
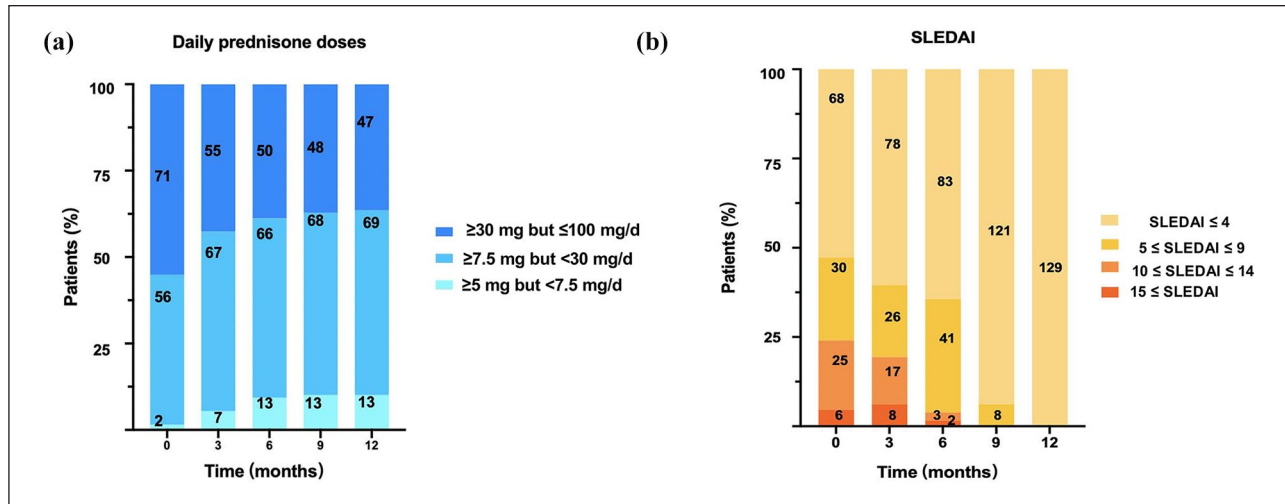
Figure 3. Changes in daily prednisone dose and disease activity. (a) Percentage of patients on daily prednisone dose at each visit. (b) Percentage of patients on SLEDAI at each visit. Data labels (numbers) refer to absolute values.

Table 2. List of variables significantly related to prednisone dose.

Variate	Statistic	P-value
Age	4.404	0.014
ESR	3.228	0.046
Anti-dsDNA	6.401	0.003
ALB	9.850	<0.001
HGB	6.746	0.002
HLA-DQA1 (rs2187668)	11.174	0.025
CYP3A4 (rs4646437)	10.719	0.030

ALB: albumin; HGB: hemoglobin; Anti-dsDNA: anti-double-stranded DNA antibodies; ESR: erythrocyte sedimentation rate. The Mann–Whitney U test was utilized to do univariate analysis on continuous variables. The chi-squared test was used to assess categorical data. A difference is considered statistically significant when the $P < 0.05$.

Therefore, the significance scores of the seven selected variables were calculated and ranked using XGBoost (Table 4 and Figure 4). Notably, the importance score of CYP3A4 rs4646437 (0.311) was remarkably higher than that of the other variables, followed by ALB (0.182) and HGB (0.154). The four subsequent variables, Anti-dsDNA, ESR, Age, HLA-DQA1 (rs2187668), had importance scores of 0.104, 0.097, 0.097, and 0.057, respectively, indicating their significance in the model. A higher importance score signified a greater impact of the variable on the prediction of daily prednisone dose.

The validation cohort comprised 26 patients, among whom five patients received a low daily dose (≥5 mg but <7.5 mg/d), 16 patients received a medium dose (≥7.5 mg but <30 mg/d), and five patients were prescribed a high dose (≥30 mg but ≤100 mg/d). The recommended prednisone dose for patients was determined using a confusion matrix derived from the XGBoost prediction model (Figure 5). The model exhibited a high level of accuracy in recommending a low daily dose regimen for two patients, achieving a precision rate of 100% and a recall rate of 40%. Similarly, the model accurately recommended a medium-dose regimen for 16 patients, with only two patients receiving an incorrect dose. This resulted in precision and recall rates of 88% and 88%, respectively. Furthermore, the model accurately recommended a high-dose regimen for eight patients; however, three patients were prescribed an incorrect dose. The precision rate for this category was 62% and the recall rate reached 100%.

Discussion

In this study, we developed a predictive model to determine the daily dose of prednisone. This involved evaluating three specific dose regimens (≥5 mg but <7.5 mg/d, ≥7.5 mg but <30 mg/d, and ≥30 mg but ≤100 mg/d). To achieve this, we used a leading-edge deep learning technique that yielded a prediction model with a commendable performance level (accuracy=0.81). Subsequently, we ranked the variables that exhibited a strong correlation with daily prednisone dose using an importance score. The importance is ranked as follows: CYP3A4 (rs4646437) > ALB > HGB > anti-dsDNA > ESR > HLA-DQA1 (rs2187668). Finally, a confusion matrix was used for model validation, which revealed that the low-dose prednisone regimen

Table 3. Prediction performance of different algorithms.

Dose regime[a]		Precision	Recall	F1-score	AUC	Accuracy	Sensitivity	Specificity
Logistic	0	0.00	0.00	0.00	0.50	0.69	0.00	1.00
	1	0.71	0.94	0.81	0.67		0.94	0.33
	2	0.60	0.60	0.60	0.75		0.60	0.88
SVM	0	0.00	0.00	0.00	0.50	0.73	0.00	1.00
	1	0.74	0.88	0.80	0.69		0.88	0.50
	2	0.71	1.00	0.83	0.95		1.00	0.88
RF	0	0.00	0.00	0.00	0.50	0.77	0.00	1.00
	1	0.79	0.94	0.86	0.77		0.94	0.56
	2	0.71	1.00	0.83	0.95		1.00	0.88
AdaBoost	0	0.00	0.00	0.00	0.50	0.73	0.00	1.00
	1	0.74	0.88	0.80	0.69		0.88	0.50
	2	0.71	1.00	0.83	0.95		1.00	0.88
Bagging	0	0.00	0.00	0.00	0.50	0.77	0.00	1.00
	1	0.83	0.94	0.88	0.82		0.94	0.62
	2	0.62	1.00	0.77	0.93		1.00	0.83
XGBoost	0	1.00	0.40	0.57	0.70	0.81	0.40	1.00
	1	0.88	0.88	0.88	0.84		0.88	0.78
	2	0.62	1.00	0.77	0.93		1.00	0.84
LightGBM	0	0.00	0.00	0.00	0.50	0.77	0.00	1.00
	1	0.79	0.94	0.86	0.77		0.94	0.56
	2	0.71	1.00	0.83	0.95		1.00	0.88
CatBoost	0	0.00	0.00	0.00	0.50	0.69	0.00	1.00
	1	0.71	0.94	0.81	0.67		0.94	0.33
	2	0.60	0.60	0.60	0.75		0.60	0.88
MLP	0	0.00	0.00	0.00	0.50	0.73	0.00	1.00
	1	0.76	1.00	0.86	0.75		1.00	0.38
	2	0.60	0.60	0.60	0.75		0.60	0.89
TabNet	0	0.50	0.40	0.44	0.55	0.73	0.40	0.89
	1	0.92	0.75	0.83	0.49		0.75	0.88
	2	0.56	1.00	0.71	0.90		1.00	0.78

[a]Dose Regimen of the daily prednisone dose of ≥5 mg but <7.5 mg/d corresponds to '0', regimen of the daily prednison dose of ≥7.5 mg but <30 mg/d corresponds to '1' and regimen of the daily prednison dose of ≥30 but ≤100 mg/d corresponds to '2'.
SVM: support vector machines; RF: random forest; Adaboost: adaptive boosting; XGBoost: eXtreme gradient boosting; LightGBM: light gradient boosting machine; CatBoost: categorical boosting; MLP: multi-layer perceptron.

Table 4. Variable importance score.

Variable	Importance score
CYP3A4 (rs4646437)	0.311
ALB	0.182
HGB	0.154
Anti-dsDNA	0.104
ESR	0.097
Age	0.097
HLA-DQA1 (rs2187668)	0.057

ALB: albumin; HGB: hemoglobin; Anti-dsDNA: anti-double-stranded DNA antibodies; ESR: erythrocyte sedimentation rate.

exhibited 100% precision and 40% recall rate. Similarly, the medium-dose regimen demonstrated a precision of 88% and recall rate of 88%, whereas the high-dose regimen had a precision of 62% and recall rate of 100%. Our findings indicate that the XGBoost model successfully integrated individual features, enabling their conversion into a strategy for optimising the daily dosage of prednisone in the treatment of patients with SLE. The XGBoost algorithm is a nonparametric estimation ensemble algorithm constructed on decision trees. The influence of the correlation between independent variables on the model is negligible. In cases where the sample size and dataset dimensions are substantial, the XGBoost algorithm exhibits superior performance compared to the logistic regression algorithm. When considering empirical evidence, machine learning and deep learning methodologies possess indisputable advantages over conventional modelling techniques. These
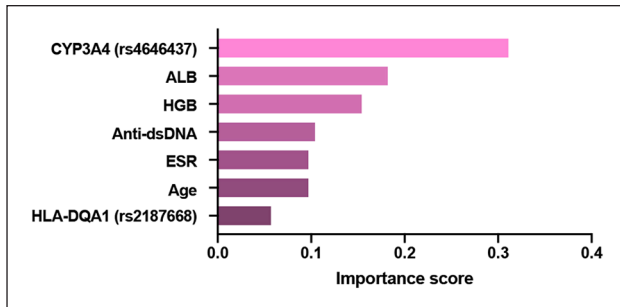
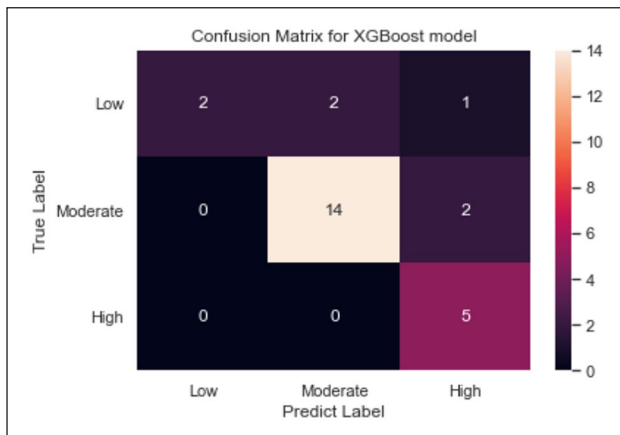**Figure 4.** Ranking of variable importance.

Figure 5. Confusing matrix diagram.

advantages include the ability of machine-learning and deep-learning models to handle intricate, high-dimensional, and interactive variables that traditional models cannot accommodate.

Numerous factors have been documented to influence the expression of metabolic enzymes and transporters, thereby affecting the pharmacokinetics of GCs. Cytochromes play a crucial role in the detoxification and maintenance of steroidal compounds. A significant proportion (78%) of hepatic clearance drugs undergo oxidative metabolism mediated by cytochrome P450 in families 1, 2, and 3. Among these, CYP3A4/5 (37%) exhibit the highest contribution.[35] Notably, CYP3A4 is involved in the 6f-hydroxylation of dexamethasone, contributing to the negative feedback loop of dexamethasone autoregulation.[36] Furthermore, prednisone stimulates CYP3A4.[37] The activity of the CYP3A4 enzyme is enhanced by the regulation of CAR receptor protein production, which is involved in transcription,[38] resulting in an improvement in the metabolic activity of CYP3A4 substrates such as drugs used in combination with prednisone. Schirmer et al.[39] discovered that the T > C variation in the CYP3A4 (rs4646437) gene, located on intron 7, influences the expression level of CYP3A4 protein in the liver and the activity of the enzyme. According to our findings, which are supported by previous research, the SNP in CYP3A4 (rs4646437) is the principal determinant of daily prednisone dosage.

An additional genetic variation, specifically HLA-DQA1 (rs2187668), was identified. The HLA-DQA1 genotype (rs2187668) may play a role in the development of cutaneous lupus erythematosus by abnormally affecting antigen presentation.[40] Bullich et al.[41] provide empirical support for the involvement of HLA-DQA1 (rs2187668) in predicting the efficacy of immunosuppressant drugs. In conjunction with our findings, it has been postulated that alterations in these two SNPs could potentially affect prednisone metabolism and therapeutic outcomes. Our study represents a pioneering effort to incorporate a patient's genetic composition into the management of GC dosing, thereby facilitating the precise administration of GCs, akin to the accurate delivery of immunosuppressive medications such as tacrolimus.

SLE exhibits a distinctly higher prevalence in particular sex and age cohorts, with women in their reproductive years having an elevated vulnerability.[42] We acknowledge that age is a crucial predictive factor in the dosage model. Furthermore, our findings suggest that specific biochemical markers, namely ALB, HGB, anti-dsDNA, and ESR, have considerable significance in the developed model for predicting the daily prednisone dosage. In particular, ALB correlates with the likelihood of mortality from all causes and infections in individuals with severe SLE-related illnesses.[43] Patients with SLE demonstrate a significant decrease in ALB levels compared to healthy individuals. Furthermore, a negative correlation was observed between disease activity and serum ALB levels.[44] Altered ESR has been consistently observed in acute SLE cases.[45] The measurement of anti-dsDNA is a critical factor in evaluating SLE disease activity, particularly in cases involving lupus nephritis.[46] In a personalised medication model for prednisone, ALB, HGB, anti-dsDNA, and ESR serve as notable predictors that can assist clinicians in appropriately adjusting prednisone dosage regimens.

In addition to the factors mentioned above, prognostic indicators and relapse variables include 24-hour proteinuria, urine sediment, and complement C3 and C4 levels, as well as clinical symptoms such as lupus nephritis, neurologic issues, vasculitis, haematological abnormalities, and refractory hypertension. However, insufficient clinical data on these variables hindered our ability to conduct further enquiries. It is crucial to include relevant markers as clinical variables in future investigations to estimate GC dosage.

Patients differ in age, sex, underlying health status, and genetic background, all of which can influence disease activity and treatment response. For example, young patients may tolerate the disease better and show relatively mild symptoms owing to better physical status, even when some indicators are abnormal; however, the condition of older patients or patients with other underlying diseases, even if the same index is abnormal, may be more serious. Patients with SLE usually require long-term drug therapy with glucocorticoids and immunosuppressants. These drugs are effective in managing the disease but may also affect the metrics in the model, leading to a change in the relationship between the indicators and disease severity.

The developed model will help physicians make preliminary assessments of patient sensitivity to glucocorticoids. Kandane-Rathnayake et al.[47] discovered that the dose of GCs was significantly higher in disease with higher activity. Higher doses of GC treatment often indicate low responsiveness, which means that these patients face adverse effects from high-dose GCs; therefore, physicians must consider whether to adjust the treatment strategy and use biological agents instead of GCs in these patients.

This study has some limitations. First, it only included participants of Han Chinese descent. Variations in GC pharmacokinetics according to race may result in a reduced predictive power of our model in patients of other ethnicities. We used a limited sample size from a single medical centre, highlighting the need for a larger samples encompassing several medical centres in further studies. Incorporating additional data will enhance the performance of the model. In addition, the exclusion of partial information on combination medications was attributed to an inadequate number of samples in our prediction model. Consequently, the inclusion of additional ethnic groups and combination

regimens is necessary to enhance discerning factors and broaden the applicability of the model.

Conclusion

We developed a GC dose prediction model using machine learning and deep learning methodologies, and subsequently evaluated multiple algorithms. XGBoost exhibited superior performance as a dose prediction model. To facilitate implementation in clinical settings, the model requires further research based on empirical data and potential expansion to incorporate novel variables. This study contributes to a more precise model for the management of GC dosage in patients with SLE.

Author contributions

All authors participated in the study's design, data collection, analysis, and critical revision for key intellectual content. XL wrote the manuscript and contributed to data curation. JJZ contributed to Methodology and review. MDF and XNL contributed to data curation and investigation. HLH contributed to data analysis. PZ and YLL contributed to project administration. RFX and LQM contributed to study design. All authors read and approved the final manuscript. Additionally, all authors gave their approval for the final version to be published.

Declaration of conflicting interests

The author(s) declared no potential conflicts of interest with respect to the research, authorship, and/or publication of this article.

Ethicals approval

The studies involving human participants were reviewed and approved by the Ethics Committee of NanFang Hospital, Southern Medical University (NFEC-2021-320).

The patients provided their written informed consent to participate in this study.

Informed consent

Written informed consent was obtained from all subjects before the study*.

ORCID iD

Xin Luo (iD) https://orcid.org/0009-0004-0575-6429

Data availability

The original contributions presented in the study are included in the article/supplementary material, further inquiries can be directed to the corresponding authors.

Supplemental material

Supplemental material for this article is available online.

References

1. Barber MRW, Drenkard C, Falasinnu T, et al. (2021) Global epidemiology of systemic lupus erythematosus. *Nature Reviews Rheumatology* 17(9): 515–532.
2. Little J, Parker B, Lunt M, et al. (2018) Glucocorticoid use and factors associated with variability in this use in the Systemic Lupus International Collaborating Clinics Inception Cohort. *Rheumatology (Oxford)* 57(4): 677–687.
3. Ruiz-Irastorza G and Bertsias G (2020) Treating systemic lupus erythematosus in the 21st century: New drugs and new perspectives on old drugs. *Rheumatology (Oxford)* 59(Suppl5): v69–v81.
4. Felten R, Scher F, Sibilia J, et al. (2019) Advances in the treatment of systemic lupus erythematosus: From back to the future, to the future and beyond. *Joint Bone Spine* 86(4): 429–436.
5. Jung JY, Yoon D, Choi Y, et al. (2019) Associated clinical factors for serious infections in patients with systemic lupus erythematosus. *Scientific Reports* 9(1): 9704.
6. Kostopoulou M, Nikolopoulos D, Parodis I, et al. (2020) Cardiovascular disease in systemic lupus erythematosus: Recent data on epidemiology, risk factors and prevention. *Current Vascular Pharmacology* 18(6): 549–565.
7. Ton FN, Gunawardene SC, Lee H, et al. (2005) Effects of low-dose prednisone on bone metabolism. *Journal of Bone and Mineral Research* 20(3): 464–470.
8. Kaneko K, Chen H, Kaufman M, et al. (2021) Glucocorticoid-induced osteonecrosis in systemic lupus erythematosus patients. *Clinical and Translational Medicine* 11(10): e526.
9. Piga M, Floris A, Sebastiani GD, et al. (2020) Risk factors of damage in early diagnosed systemic lupus erythematosus: Results of the Italian multicentre Early Lupus Project inception cohort. *Rheumatology (Oxford)* 59(9): 2272–2281.
10. Ugarte-Gil MF, Mak A, Leong J, et al. (2021) Impact of glucocorticoids on the incidence of lupus-related major organ damage: A systematic literature review and meta-regression analysis of longitudinal observational studies. *Lupus Science & Medicine* 8(1): 1–10.
11. Shimba A and Ikuta K (2020) Control of immunity by glucocorticoids in health and disease. *Seminars in Immunopathology* 42(6): 669–680.
12. Sada KE, Katayama Y, Asano Y, et al. (2022) Association of one-point glucocorticoid-free status with chronic damage and disease duration in systemic lupus erythematosus: A cross-sectional study. *Lupus Science & Medicine* 9(1): e000772.
13. Felten R, Sagez F, Gavand PE, et al. (2019) 10 most important contemporary challenges in the management of SLE. *Lupus Science & Medicine* 6(1): e000303.
14. Meyer UA, Zanger UM, et al. (2013) Omics and drug response. *Annual Review of Pharmacology and Toxicology* 53: 475–502.
15. Li QY, Tang BH, Wu YE, et al. (2023) Machine learning: A new approach for dose individualization. *Clinical Pharmacology & Therapeutics* 115(4): 727–744.
16. Mo X, Chen X, Li H, et al. (2019) Early and accurate prediction of clinical response to methotrexate treatment in juvenile idiopathic arthritis using machine learning. *Frontiers in Pharmacology* 10: 1155.
17. Lee HC, Yoon SB, Yang SM, et al. (2018) Prediction of acute kidney injury after liver transplantation: Machine learning approaches vs. logistic regression model. *Journal of Clinical Medicine* 7(11): 428.
18. Kruppa J, Ziegler A and Konig IR (2012) Risk estimation and risk prediction using machine-learning methods. *Human Genetics* 131(10): 1639–1654.
19. Luo Y (2022) Evaluating the state of the art in missing data imputation for clinical data. *Briefings in Bioinformatics* 23(1): bbab489.
20. Zhang J, Mucs D, Norinder U, et al. (2019) LightGBM: An effective and scalable algorithm for prediction of chemical toxicity – application to the Tox21 and mutagenicity data sets. *Journal of Chemical Information and Modeling* 59(10): 4150–4158.
21. Mitani T, Doi S, Yokota S, et al. (2020) Highly accurate and explainable detection of specimen mix-up using a machine learning model. *Clinical Chemistry and Laboratory Medicine* 58(3): 375–383.
22. Saber R, Henault D, Messaoudi N, et al. (2023) Radiomics using computed tomography to predict CD73 expression and prognosis of colorectal cancer

liver metastases. *Journal of Translational Medicine* 21(1): 507.

23. Hancock JT and Khoshgoftaar TM (2020) CatBoost for big data: An interdisciplinary review. *Journal of Big Data* 7(1): 94.

24. Zhang Q, Tian X, Chen G, et al. (2022) A Prediction model for tacrolimus daily dose in kidney transplant recipients with machine learning and deep learning techniques. *Frontiers in Medicine (Lausanne)* 9: 813117.

25. Zheng P, Yu Z, Mo L, et al. (2022) An individualized medication model of sodium valproate for patients with bipolar disorder based on machine learning and deep learning techniques. *Frontiers in Pharmacology* 13: 890221.

26. Edwards JC, Snaith ML and Isenberg DA (1987) A double blind controlled trial of methylprednisolone infusions in systemic lupus erythematosus using individualised outcome assessment. *Annals of the Rheumatic Diseases* 46: 773–776.

27. Adriana Danowski LM and Petri M (2006) Flares in lupus- outcome assessment trial (FLOAT), a comparison between oral 2006 methylprednisolone and intramuscular triamcinolone. *The Journal of Rheumatology* 33(1): 57–66.

28. Zeher M, Doria A, Lan J, et al. (2011) Efficacy and safety of enteric coated. *Lupus* (20): 1484–1493.

29. Wang CB, Zhang YJ, Zhao MM, et al. (2023) Dosage optimization of tacrolimus based on the glucocorticoid dose and pharmacogenetics in adult patients with systemic lupus erythematosus. *International Immunopharmacology* 124(Pt A): 110866.

30. Ruiz-Irastorza G, Garcia M, Espinosa G, et al. (2016) First month prednisone dose predicts prednisone burden during the following 11 months: An observational study from the RELES cohort. *Lupus Science & Medicine* 3(1): e000153.

31. Ruiz-Arruza I, Barbosa C, Ugarte A, et al. (2015) Comparison of high versus low-medium prednisone doses for the treatment of systemic lupus erythematosus patients with high activity at diagnosis. *Autoimmunity Reviews* 14(10): 875–879.

32. Ruiz-Irastorza G, Danza A, Perales I, et al. (2014) Prednisone in lupus nephritis: How much is enough? *Autoimmunity Reviews* 13(2): 206–214.

33. Houssiau FA and Isenberg D (2013) Towards treating lupus nephritis without oral steroids: A dream-come-true? *Annals of the Rheumatic Diseases* 72(8): 1271–1272.

34. Floris A, Chessa E, Sebastiani GD, et al. (2022) Glucocorticoid tapering and associated outcome in patients with newly diagnosed systemic lupus erythematosus: The real-world GULP prospective observational study. *RMD Open* 8(2): e002701.

35. Zanger UM, Turpeinen M, Klein K, et al. (2008) Functional pharmacogenetics/genomics of human cytochromes P450 involved in drug biotransformation. *Analytical and Bioanalytical Chemistry* 392(6): 1093–1108.

36. Hejazian SM, Zununi Vahed S, Moghaddas Sani H, et al. (2020) Steroid-resistant nephrotic syndrome: Pharmacogenetics and epigenetic points and views. *Expert Review of Clinical Pharmacology* 13(2): 147–156.

37. Quattrochi LC and Guzelian PS (2001) Cyp3A regulation: From pharmacology to nuclear receptors. *Drug Metabolism and Disposition* 29(5): 615–622.

38. Pascussi JM, Gerbal-Chaloin S, Fabre JM, et al. (2000) Dexamethasone enhances constitutive androstane receptor expression in human hepatocytes: Consequences on cytochrome P450 *Gene Regulation* 58: 1441–1450.

39. Schirmer M and Rosenberger A (2007) Sex-dependent genetic markers of CYP3A4 expression and activity in human liver microsomes. *Pharmacogenomics* 8(5): 443–453.

40. Kunz M, Konig IR, Schillert A, et al. (2015) Genome-wide association study identifies new susceptibility loci for cutaneous lupus erythematosus. *Experimental Dermatology* 24(7): 510–515.

41. Bullich G, Ballarín J, Oliver A, et al. (2014) HLA-DQA1 and PLA2R1 polymorphisms and risk of idiopathic membranous nephropathy. *Clinical Journal of the American Society of Nephrology* 9(2): 335–343.

42. Kaul A, Gordon C, Crow MK, et al. (2016) Systemic lupus erythematosus. *Nature Reviews Disease Primers* 2: 16039.

43. He S, Tang C, Yu J, et al. (2021) Combining C reactive protein and serum albumin to predict 90-day mortality in systemic lupus erythematosus with serious community-acquired infections. *Lupus Science & Medicine* 8(1): e000505.

44. Sahebari M, Abrishami-Moghaddam M, Moezzi A, et al. (2014) Association between serum trace element concentrations and the disease activity of systemic lupus erythematosus. *Lupus* 23(8): 793–801.

45. Aringer M (2020) Inflammatory markers in systemic lupus erythematosus. *Journal of Autoimmunity* 110: 102374.

46. Lou H, Ling GS and Cao X (2022) Autoantibodies in systemic lupus erythematosus: From immunopathology to therapeutic target. *Journal of Autoimmunity* 132: 102861.

47. Kandane-Rathnayake R, Louthrenoo W, Luo SF, et al. (2022) Patterns of medication use in systemic lupus erythematosus: A multicenter cohort study. *Arthritis Care and Research (Hoboken)* 74(12): 2033–2041.