An accurate predictive model for depressive symptoms in patients with primary Sjögren's disease based on machine learning algorithms

L. Yang[1], Y. Tong[2,3], Y. Jin[2,3], X. Wang[1], Y. Yan[2,3], W. Lu[2,3], D. Su[1], J. Zou[3,4], K. Huang[3,4]

[1]Department of Rheumatology and Immunology, Nanjing First Hospital, Nanjing Medical University, Nanjing; [2]School of Basic Medicine and Clinical Pharmacy, China Pharmaceutical University, Nanjing; [3]Department of Clinical Pharmacology Lab, Nanjing First Hospital, Nanjing Medical University, Nanjing; [4]Department of Pharmacy, Nanjing First Hospital, China Pharmaceutical University, Nanjing, China.

Abstract

Objective

To develop and validate a machine-learning (ML) model that flags primary Sjögren's disease (pSjD) patients at high risk of depressive symptoms for earlier clinical attention.

Methods

We retrospectively studied 147 pSjD patients (Nanjing First Hospital, 2019-2022). Depressive symptoms were screened with Patient Health Questionnaire-9 (PHQ-9); PHQ-9 ≥5 was the primary endpoint. Missing data were handled by multiple imputation. Data were split 70/30 for training/testing. After univariate screening and LASSO selection, eight ML algorithms (e.g. logistic regression, support vector machine (SVM), tree/boosting methods) were trained with stratified 10-fold cross-validation. Performance was summarised by AUROC/AUPRC, accuracy, precision/recall, Brier score, and calibration; SHAP provided model explainability.

Results

Four routinely available predictors were retained: fatigue frequency, sleep duration, lymphocyte count, and anti-Ro52 status. Across repeated cross-validation, SVM showed the best overall discrimination (mean AUROC≈0.90) with strong precision and accuracy. In the held-out test set, SVM maintained high performance (AUROC=0.929; AUPRC=0.959; Brier=0.106). SHAP confirmed predictor importance, indicating higher risk with shorter sleep, lower lymphocyte counts, greater fatigue frequency, and anti-Ro52 positivity.

Conclusion

This study presents the first ML-based model for predicting depressive symptoms in pSjD patients, highlighting the significance of immuno-inflammatory and clinical factors in depression pathogenesis. The SVM model offers a robust, non-invasive tool for early identification of high-risk individuals, enabling timely and personalised interventions. However, this single-centre, retrospective design with a modest sample limits generalisability; therefore, independent multi-centre validation is required before clinical use.

Key words

primary Sjögren's disease, depression, machine learning, predictive model, anti-Ro52 antibody

*Leilei Yang, MM**
*Yulan Tong, MSc**
*Yuzhan Jin, MSc**
XiaoqinWang, MM
Yuqing Yan, MSc
Wei Lu, MSc
Dinglei Su, MD
Jianjun Zou, PhD
Kaizong Huang, PhD

**Contributed equally.*

Please address correspondence to:
Dinglei Su
Department of Rheumatology and Immunology,
Nanjing First Hospital,
Nanjing Medical University,
no. 68 Changle Road,
210006 Nanjing (Jiangsu), China.
E-mail: sudinglei@njmu.edu.cn

Jianjun Zou
Department of Clinical Pharmacology, (postal address as above)
E-mail: zoujianjun100@126.com

Kaizong Huang
Department of Clinical Pharmacology, (postal address as above)
E-mail: kzhuang@nju.edu.cn

Received on: August 14, 2025; accepted in revised form on November 13, 2025.

Funding: this work was supported by Nanjing Medical Science and Technology Development Fund (ZKX24029) and National Natural Science Foundation of China (82173899).

Competing interests: none declared.

Introduction

Primary Sjögren's disease (pSjD) is a chronic autoimmune disorder that primarily targets the salivary and lacrimal glands, leading to marked ocular and oral dryness (1, 2). Beyond glandular dysfunction, patients frequently exhibit extra-glandular manifestations, including neuropsychiatric symptoms such as depression (3). Estimates of the prevalence of depression in pSjD vary significantly, ranging from 32% to 45.8%, depending on the diagnostic methods used (4). The COVID-19 pandemic has significantly exacerbated the incidence of depression (5), with a recent study indicating that the rate among pSjD patients has escalated to 57.9% (6). Coexisting pSjD and depression complicates diagnosis and treatment, undermines quality of life and adherence, and underscores the need for tools that identify high-risk patients early to enable timely intervention.

The mechanisms linking pSjD and depression are multifactorial. Proposed drivers include brain parenchymal changes, dysregulation of cytokine networks, and heightened autoimmune inflammation (4). Aberrant immunity may sustain chronic inflammation and perturb neurotransmission and neural circuits implicated in mood regulation. pSjD may also trigger stress responses that contribute to hypothalamic-pituitary-adrenal (HPA) axis dysfunction, reinforcing depressive phenotypes (7). Immune dysregulation in pSjD, marked by polyclonal B cell activation and various autoantibodies, is also implicated in the neurobiology of depression (8).

Several factors predispose pSjD patients to depression, including clinical manifestations, dysfunction of the cytokine regulatory network, and autoantibody production. A single-centre cross-sectional study in China indicated that disease activity and symptoms of dry mouth and eyes are risk factors for depression in pSjD patients (9). Recent study has shown that IL-1β and its related molecules are associated with fatigue in patients with pSjD, and these cytokines may also contribute to depression (10). Our recent study identified anti-Ro52 antibodies as risk factors and sleep duration as a protective factor for depression in pSjD patients (6). Nevertheless, the heterogeneous, network-level pathogenesis and the multitude of candidate predictors make early, accurate risk stratification challenging.

While questionnaire scales such as the Patient Health Questionnaire-9 (PHQ-9) are clinically useful, reliance on a single subjective instrument limits predictive accuracy for pSjD-related depression amid complex biology and cofounders (11). A more robust approach should integrate clinical and laboratory data to reduce bias and capture multidimensional patterns associated with depressive outcomes. Machine learning (ML)-by leveraging larger feature spaces and flexible algorithms-can detect nonlinear associations and interactions that may be overlooked by conventional methods (12, 13). As a result, ML is increasingly utilised in personalised and precision medicine, especially in the development of major disease risk prediction models (14, 15). Notably, an eXtreme Gradient Boosting (XGB) model accurately predicted postpartum depression; early interventions targeted to high-risk individuals identified by the model reduced incident cases *versus* usual care, demonstrating the clinical value of ML-enabled screening and early action (16). In this study, we aimed to develop an accurate predictive model using ML techniques to predict depressive symptoms related to pSjD. A wide range of parameters considered relevant for the onset of depression in pSjD patients, including key demographic data, clinical manifestations, levels of inflammatory cytokines, and the presence of autoantibodies, were incorporated into this model. The successful establishment of this model would enable the early identification and management of pSjD-associated depressive symptoms.

Materials and methods

Study design and population

This investigation is a single-centre retrospective analysis intended as a proof-of-concept, hypothesis-generating study. The data for this study were provided by the Department of Rheumatology and Immunology at Nanjing First Hospital, covering the period from August 2019 to June 2022. The study

was approved by the Ethics Committee of Nanjing First Hospital (Reference: KY20240603-KS-02), who waived the informed consent requirement given the retrospective design with pseudonymised data and minimal risk to participants.

The criteria for patient eligibility included: 1. being over 18 years old; 2. adhering to the 2016 American College of Rheumatology/European League Against Rheumatism classification criteria for pSjD (17); 3. filling out the PHQ-9 questionnaire during the medical history collection on the first day of admission. Additionally, patients with a history of severe mental illness, malignant tumours, those who had used drugs to treat mental disorders, and pregnant or lactating women were excluded. Given the modest sample size and geographic scope, all findings should be interpreted as preliminary and subject to confirmation in independent, multicentre cohorts.

Assessment of depression

The PHQ-9, a nine-item instrument for primary and general healthcare settings (16, 18, 19), was administered at day-1 initial assessment to standardise timing and mirror real-world triage. Consistent with its screening purpose, we prespecified screen-positive depressive symptoms as PHQ-9 ≥5, aligned with the original severity bands (5/10/15/20 denoting mild/moderate/moderately severe/severe symptom burden (18-20).

Covariates

We evaluated: 1. demographics (age, sex, marital status, education); 2. medical history (overlap CTDs; thyroid disease, diabetes, other chronic conditions; fibromyalgia); 3. haematology (WBC, haemoglobin, platelets, neutrophils, lymphocytes); 4. inflammatory markers (IL-2, IL-4, IL-6, IL-10, TNF-α, ESR, CRP); 5. autoantibodies (anti-Ro52, aPL, Rib-P, SSA, SSB, RNP/Sm, CENP-B); and 6. medications [corticosteroids (prednisone-equivalent), immunosuppressants (*e.g.* cyclophosphamide, mycophenolate mofetil, cyclosporine A, tacrolimus, azathioprine), and csDMARDs (*e.g.* methotrexate, hydroxychloroquine, iguratimod).

Data structuralising and pre-processing

The model for detecting depression was set up and validated through a division of the cohort data into training and testing samples, in a 7:3 random split. Except for the heavily missing variables that were deleted, other missing values within the data were imputed with Multiple Imputation by Chained Equations (MICE) (21, 22). We summarised variable-wise missingness and patterns (Supplementary Table S1, Supplementary Fig. S1); overall 5.82%; higher for T-SPOT 48.3%, cytokines ~32%, IFN-γ 31.3%, IL-6 29.3%; core clinical <1%. We imputed using MICE (m=5-10; PMM for continuous; logistic/multinomial for categorical) and pooled estimates via Rubin's rules.

The development and validation of the depression detection model proceeded through multiple phases. Initially, a univariate analysis was performed on the training set to identify variables that showed significant differences between patients with and without depressive symptoms, facilitating initial feature selection. Then, the training and testing sets were separately standardised to avoid poor model performance due to uneven data distribution. Lastly, the LASSO technique was applied to the initially selected variables for ML modelling.

ML modelling

In this section, depression ML models were developed with the Python Scikit-learn package 1.2.2. The input features used included those selected by LASSO. Guided by the dataset's characteristics, we adopted a systematic, diversified algorithm panel comprising Logistic Regression (LR), Random Forest Classifier (RFC), Support Vector Machine (SVM), K-Nearest Neighbors (KNN), Multilayer Perceptron (MLP), and three state-of-the-art gradient-boosting methods: LightGBM, XGBoost, and CatBoost. LR was included for its suitability to small samples; RFC for modelling non-linear relationships; SVM for robustness in high-dimensional spaces; KNN as a simple, instance-based comparator; MLP as a representative neural network; and LightGBM/XGBoost/CatBoost as leading tree-boosting frameworks. This breadth balances methodological completeness with the interpretability and robustness required in clinical contexts. Given the modest sample size, all models incorporated appropriate regularisation and repeated cross-validation to mitigate overfitting and enhance the reliability and accuracy of performance estimates.

The hyper-parameters for the ML models were optimised through 10-fold cross-validation and grid search. After determining the optimal hyper-parameters for each model, we proceeded to train each model with these optimal settings using randomly selected training data. Subsequently, each model underwent testing with the testing data.

To identify the optimal model, we assessed and compared eight models based on several metrics, including accuracy, specificity, sensitivity, precision, recall, area under the receiver operating characteristic (AUROC), area under the precision-recall curve (AUPRC), Brier score, positive predictive value (PPV), and negative predictive value (NPV). To ensure robust model evaluation, this study employed repeated ten-fold cross-validation with 10 repetitions, yielding 100 performance estimates. All performance metrics are reported as the mean and its 95% confidence interval, calculated based on the t-distribution. This approach enhances the reliability of the performance evaluation by increasing the number of resampling iterations and provides a quantifiable measure of estimation uncertainty.

Internal validation and uncertainty quantification

Models were tuned using repeated stratified 10-fold cross-validation. To assess stability and quantify uncertainty, we performed 1,000 patient-level, stratified bootstrap resamples, re-fitting the models within each resample to obtain optimism-corrected estimates for AUROC/AUPRC, Brier score, calibration slope/intercept, accuracy, sensitivity/specificity, PPV/NPV. We report 95% confidence intervals using percentile bootstrap. Because threshold-dependent metrics are sensitive to base-rate shifts, we emphasise discrimination

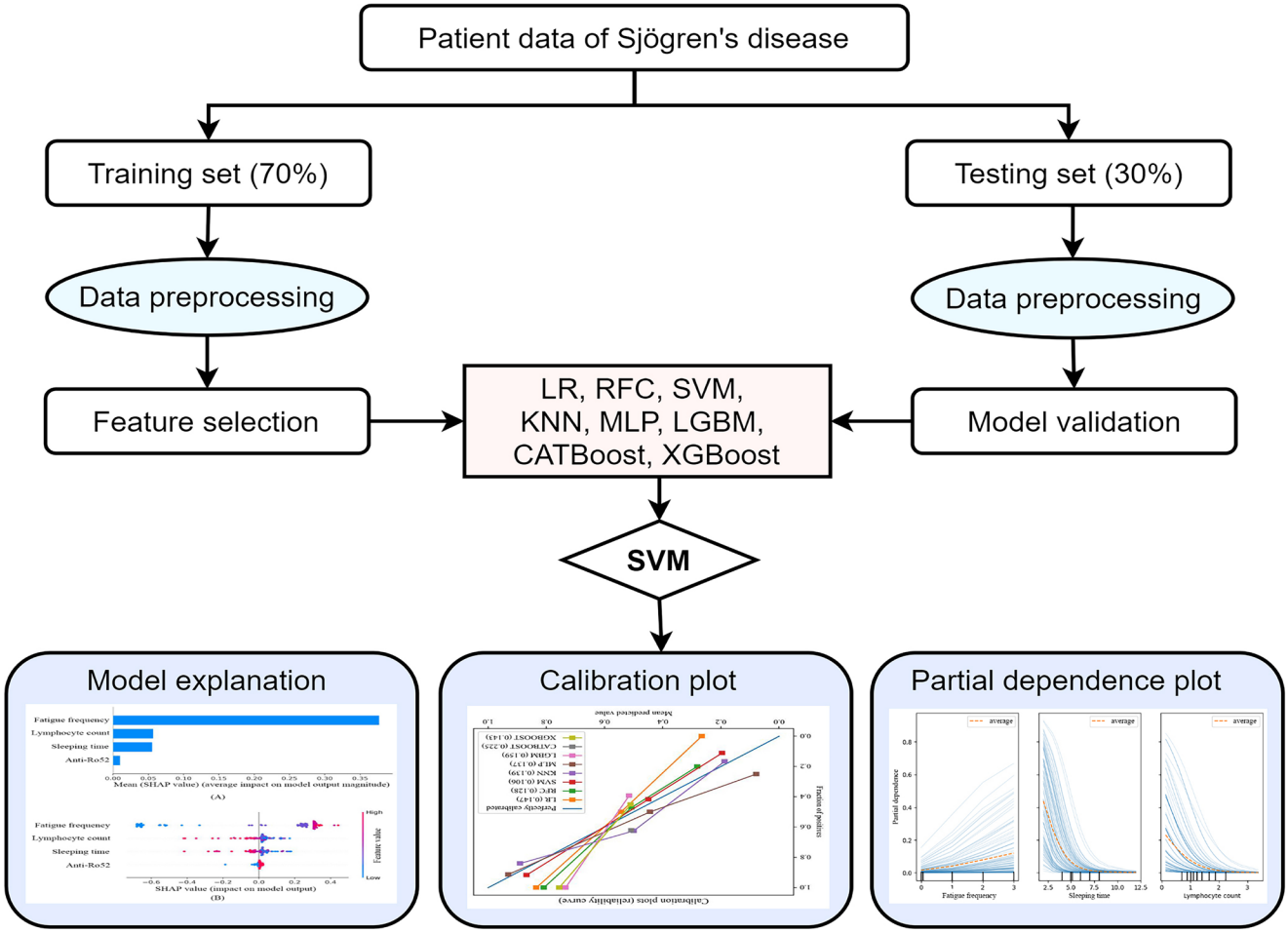

Fig. 1. Flow chart of the study.
LR: logistic regression; SVM: support vector machine; MLP: multilayer perceptron; KNN: K-nearest neighbour; RFC: random forest classification; LGBM: light gradient boosting machine; CatBoost: categorical boosting; XGBoost: extreme gradient boosting.

and calibration as primary performance summaries.

SHAP explanation

Many ML models are accurate yet opaque ('black boxes'), limiting adoption in high-stakes settings such as medical diagnosis. To enhance transparency, we employed Shapley Additive exPlanations (SHAP), a widely used *post-hoc* interpretability approach (23). Grounded in game theory, SHAP attributes each prediction to features via Shapley values, quantifying their marginal contributions and directions (24). In this study, SHAP v0.44.0 was used to visualise global and instance-level effects, thereby elucidating the model's decision process.

Other statistical methods

Apart from the Scikit-learn package, we also utilised R software (v. 4.3.2). Shapiro-Wilk normality tests were conducted for every continuous variable. Continuous variables with a normal distribution were reported as mean ± standard deviation (SD) while continuous variables with a skewed distribution were reported using the median (interquartile range (IQR)). T-tests or Mann-Whitney U-tests were employed to assess the discrepancies between two groups. Categorical variables were reported in percentages and analysed with the chi-square tests. Prior to modelling, one-hot encoding was applied. Any significance test was conducted with a two-tailed, maintaining a significance cut-off point at $p<0.05$.

Results

Baseline characteristics

Following the inclusion and exclusion criteria, 147 patients with pSjD were enrolled. The flow chart is displayed in Figure 1 and the baseline demographic, clinical and laboratory parameters is shown in Table I and Supplementary Table S2. The dataset was predominantly comprised of female patients, accounting for 93.1% of the records, while males made up 6.8%. In the training data, average ages between the participants with depressive symptoms (59.4±11.4 years) and those without depressive symptoms (57.7±10.0 years) showed no statistically significant difference. The two groups were comparable in most disease-related features, including EULAR Sjögren's Syndrome Disease Activity Index (ESSDAI) score and the prevalence of extra-glandular manifestations. However, patients with depressive symptoms reported significantly shorter sleep duration (median 5.0 *vs*. 6.0 hours, $p=0.01$) and a markedly higher burden of fatigue ($p<0.001$). Notably, 20.6% of the depression group

Table I. Comparison of baseline demographic, clinical and laboratory parameters between patients with or without depressive symptoms in the training set.

	No depression n=39	Depression n=63	*p*-value
Gender			0.154
Female	39 (100%)	58 (92.1%)	
Male	0 (0.0%)	5 (7.9%)	
Age (years)	57.7 (10.0)	59.4 (11.4)	0.437
Sleep time (hours)	6.0 [5.8;7.0]	5.0 [5.0;6.0]	0.01
Fatigue			<0.001
None	26 (66.7%)	5 (7.9%)	
Less than one week in 2 weeks	11 (28.2%)	22 (34.9%)	
More than one week in 2 weeks	2 (5.1%)	23 (36.5%)	
Almost every day	0 (0.0%)	13 (20.6%)	
Pain VAS	1.36 (2.2)	1.08 (1.9)	0.513
Sicca symptom			1.000
No	3 (7.7%)	5 (7.9%)	
Yes	36 (92.3%)	58 (92.1%)	
Salivary gland biopsy			0.823
No	25 (64.1%)	39 (61.9%)	
Yes	14 (35.9%)	24 (38.1%)	
ESSDAI	4.0 [2.0;7.0]	4.0 [2.0;6.0]	0.975
Extra-glandular involvement			
Pulmonary			0.329
No	32 (82.1%)	45 (71.4%)	
Yes	7 (17.9%)	18 (28.6%)	
Musculoskeletal			
No	32 (82.1%)	53 (84.1%)	0.785
Yes	7 (17.9%)	10 (15.9%)	
Renal			
No	36 (92.3%)	57 (90.5%)	1.000
Yes	3 (7.7%)	6 (9.5%)	
Haematological			
No	33 (84.6%)	50 (79.4%)	
Yes	6 (15.4%)	13 (20.6%)	
WBC count ($\times 10^9$/L)	5.7 [4.7;6.7]	5.2 [4.1;6.7]	0.252
Haemoglobin (g/L)	121 [110;134]	119 [108;124]	0.279
Platelet count ($\times 10^9$/L)	201 [143;246]	175 [146;218]	0.495
Neutrophil count ($\times 10^9$/L)	3.2 [2.3;4.6]	3.5 [2.2;5.0]	0.885
Lymphocyte count ($\times 10^9$/L)	1.6 [1.1;2.1]	1.1 [0.9;1.5]	0.002
Anti-Ro52			<0.001
(-)	24 (61.5%)	15 (23.8%)	
(+)	15 (38.5%)	48 (76.2%)	
Anti-SSA			0.829
(-)	19 (48.7%)	28 (44.4%)	
(+)	20 (51.3%)	35 (55.6%)	
Anti-SSB:			0.202
(-)	34 (87.2%)	47 (74.6%)	
(+)	5 (12.8%)	16 (25.4%)	
ESR (mm/h)	27.0 [15.0;53.0]	40.0 [17.5;74.5]	0.228
CRP (mg/L)	3.3 [2.0;7.4]	3.9 [2.3;10.4]	0.240
Complement C3 (g/L)	0.80 (0.19)	0.76 (0.21)	0.337
Complement C4 (g/L)	0.19 [0.15;0.22]	0.19 [0.15;0.24]	0.896
IgG (g/L)	13.1 [10.8;18.4]	14.6 [11.7;18.7]	0.430
Therapy at inclusion			
Prednisone equivalent (mg/day)	0.0 [0.0;7.5]	5.0 [0.0;10.0]	0.171
Immunosuppressants			0.273
No	36 (92.3%)	52 (82.5%)	
Yes	3 (7.7%)	11 (17.5%)	
Conventional anti-rheumatic drugs			0.829
No	12 (30.8%)	22 (34.9%)	
Yes	27 (69.2%)	41 (65.1%)	

Immunosuppressants included cyclophosphamide, mycophenolate mofetil, cyclosporine A, tacrolimus, and azathioprine; conventional synthetic disease-modifying antirheumatic drugs (csDMARDs) included methotrexate, hydroxychloroquine, and iguratimod.
VAS: visual analogue scale; ESSDAI: EULAR Sjögren's Syndrome Disease Activity Index; WBC: white blood cell count; ESR: erythrocyte sedimentation rate; CRP: C-reactive protein; IgG: immunoglobulin G.

experienced fatigue almost every day, compared to none in the non-depression group. Serologically, patients with depressive symptoms exhibited a significantly lower lymphocyte count (median 1.1 *vs*. 1.6 $\times 10^9$/L, p=0.002) and a strikingly higher positivity rate for anti-Ro52 antibody (76.2% *vs*. 38.5%, p<0.001). This serological marker demonstrated the most substantial difference between the cohorts. Other laboratory parameters, including inflammatory markers (ESR, CRP), complement levels, and IgG were not significantly different. Treatment patterns at inclusion, including the use of glucocorticoids (prednisone equivalent), immunosuppressants, and csDMARDs, were similar between the two groups.

Model performance

Following LASSO regression analysis revealed that the model comprised the following predictors: presence of anti-Ro52 antibodies, sleep time, fatigue frequency, and lymphocyte count.

All evaluation metrics specified in the Methods section were computed to compare models, with the performance of each model on the training and testing sets detailed in Table II and III. The SVM model was the most effective, with an average AUROC of 0.921, a precision of 0.904, and an accuracy of 0.878. In the test set, the SVM model also performed well, achieving an AUROC of 0.907, AUPRC of 0.940, sensitivity of 0.887, specificity of 0.774, PPV of 0.873, NPV of 0.834, accuracy of 0.843, precision of 0.873, recall of 0.887, and a Brier score of 0.106.

Bootstrap resampling produced wider 95% CIs for threshold-dependent metrics and overlapping intervals across several learners, reflecting expected variance in a modest sample. The parsimonious four-predictor logistic model showed AUROC 0.909 (95% CI 0.817–0.978), comparable to SVM and tree-based models, with favourable calibration; we therefore present it as the primary reference, retaining other models as benchmarks. Missingness and patterns are detailed in Supplementary Table S1 and Supplementary Figure S1. The ROC and PRC for the models in the test set are depicted in Figure 2. To

Table II. Predictive performance of machine learning models based on the repeated cross-validation in training set.

	AUROC (95%CI)	AUPRC (95%CI)	Sensitivity (95%CI)	Specificity (95%CI)	PPV (95%CI)	NPV (95%CI)	Accuracy (95%CI)	Precision (95%CI)	Recall (95%CI)
LR	0.913 (0.895-0.932)	0.944 (0.929-0.958)	0.907 (0.886-0.929)	0.720 (0.676-0.764)	0.851 (0.829-0.873)	0.708 (0.689-0.728)	0.835 (0.814-0.856)	0.851 (0.829-0.873)	0.907 (0.886-0.929)
RFC	0.887 (0.867-0.907)	0.944 (0.929-0.958)	0.892 (0.869-0.914)	0.640 (0.607-0.672)	0.848 (0.824-0.872)	0.692 (0.671-0.713)	0.823 (0.800-0.845)	0.848 (0.824-0.872)	0.892 (0869-0.914)
SVM	0.907 (0.889-0.925)	0.940 (0.926-0.954)	0.887 (0.863-0.910)	0.774 (0.737-0.812)	0.873 (0.853-0.893)	0.834 (0.802-0.867)	0.843 (0.824-0.863)	0.873 (0.853-0893)	0.887 (0.863-0.910)
KNN	0.807 (0.781-0.838)	0.940 (0.926-0.954)	0.825 (0.796-0.853)	0.663 (0.635-0.692)	0.857 (0.834-0.880)	0.656 (0.635-0.676)	0.795 (0.774-0.817)	0.857 (0.834-0.880)	0.825 (0.796-0.853)
MLP	0.882 (0.863-0.901)	0.936 (0.925-0.947)	0.822 (0.793-0.851)	0.777 (0.735-0.820)	0.872 (0.848-0.895)	0.757 (0.721-0.794)	0.803 (0.781-0.825)	0.872 (0.848-0.895)	0.822 (0.793-0.851)
LGBM	0.880 (0.859-0.900)	0.922 (0.907-0.938)	0.925 (0.905-0.946)	0.528 (0.493-0.563)	0.778 (0.756-0.801)	0.672 (0.648-0.696)	0.779 (0.755-0.802)	0.778 (0.756-0.801)	0.925 (0.905-0.946)
CATBoost	0.878 (0.857-0.899)	0.924 (0.909-0.939)	0.889 (0.865-0.912)	0.642 (0.610-0.674)	0.850 (0.826-0.873)	0.692 (0.670-0.713)	0.822 (0.800-0.845)	0.850 (0.826-0.873)	0.889 (0.865-0.912)
XGBoost	0.861 (0.840-0.882)	0.926 (0.914-0.939)	0.920 (0.900-0.941)	0.610 (0.578-0.642)	0.830 (0.807-0.852)	0.703 (0.682-0.725)	0.824 (0.801-0.846)	0.830 (0.807-0.852)	0.920 (0.900-0.941)

AUROC: area under the receiver operating characteristic; CI: confidence interval; AUPRC: area under the precision-recall curve; PPV: positive predictive value; NPV: negative predictive value; LR: logistic regression; RFC: random forest classifier; SVM: support vector machine; KNN: K-nearest neighbour; MLP: multilayer perceptron; LGBM: light gradient boosting machine; CatBoost: categorical boosting; XGBoost: extreme gradient boosting.

Table III. Predictive performance of machine learning models based on the testing set.

	AUROC	AUPRC	Sensitivity	Specificity	PPV	NPV	Accuracy	Precision	Recall
LR	0.882	0.937	0.750	0.882	0.913	0.682	0.800	0.913	0.750
RFC	0.922	0.960	0.821	0.941	0.958	0.762	0.867	0.958	0.821
SVM	0.929	0.959	0.786	0.882	0.917	0.714	0.822	0.917	0.786
KNN	0.803	0.865	0.679	0.765	0.826	0.591	0.711	0.826	0.679
MLP	0.897	0.943	0.643	1.000	1.000	0.630	0.778	1.000	0.643
LGBM	0.916	0.954	0.857	0.882	0.923	0.789	0.867	0.923	0.857
CATBoost	0.935	0.969	0.714	1.000	1.000	0.680	0.822	1.000	0.714
XGBoost	0.917	0.961	0.893	0.882	0.926	0.833	0.889	0.926	0.893

AUROC: area under the receiver operating characteristic; AUPRC: area under the precision-recall curve; PPV: positive predictive value; NPV: negative predictive value; LR: logistic regression; RFC: random forest classifier; SVM: support vector machine; KNN: K-Nearest Neighbour; MLP: multilayer perceptron; LGBM: light gradient boosting machine; CatBoost: categorical boosting; XGBoost: extreme gradient boosting.

compare the predicted probabilities of our model with the actual dataset labels, a calibration plot was illustrated in Figure 3, highlighting the Brier score (lower is better).

Model explanation

ML methods excel at identifying complex non-linear relationships among variables. However, these models often lack transparency, which complicates the interpretation of their predictions (25). Consequently, we employed SHapley Additive exPlanations (SHAP) to dissect the top-performing ML models, particularly SVM. As illustrated in Figure 4A, SHAP assigns an importance value to each feature for every prediction. In predicting depressive symptoms in patients with pSjD, fatigue frequency is identified as the most critical predictor, followed by lymphocyte count, sleep duration, and anti-Ro52 antibodies. The colours of the dots in Figure 4B represent the values of the variables, clarifying how variable values affect ML predictions. Dots on the left indicate that the feature reduces the predicted likelihood of depression, while dots on the right suggest a positive impact. Notably, a lower lymphocyte count and shorter sleep duration (indicated by blue dots) significantly increase the predicted probability of depression. Conversely, higher fatigue frequency and the presence of anti-Ro52 antibodies (indicated by red dots) contribute to an increased probability of depressive symptoms.

To further elucidate, Figure 5 includes partial dependency plots that demonstrate the relationship between the values of each continuous variable and their corresponding SHAP values. Figure 5 highlights a negative correlation between the probability of developing depressive symptoms and both sleep duration and lymphocyte count. Conversely, a positive correlation with fatigue frequency is evident. Moreover, the figure shows that when sleep duration is less than five hours and lymphocyte count is below $1.4 \times 10^{\wedge 9}$/L, the risk of depressive symptoms increases significantly.

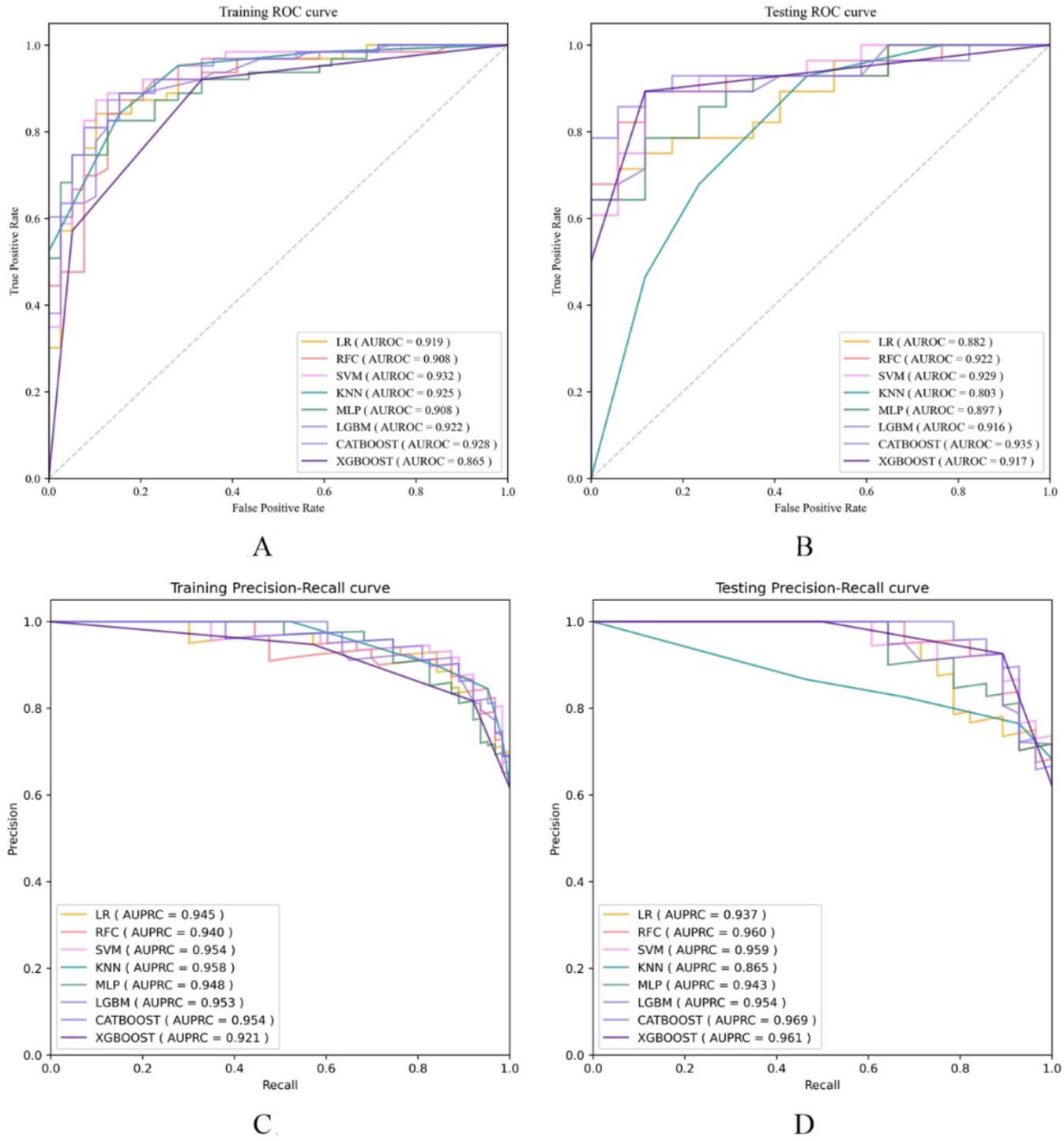

Fig. 2. Receiver operating characteristic curves and precision-recall curve of the different models. (**A**) receiver operating characteristic curves of training set, (**B**) receiver operating characteristic curves of testing set, (**C**) precision-recall curve of training set, (**D**) precision-recall curve of testing set.
AUROC: area under the receiver operating characteristic; AUPRC: area under the precision-recall Curve; LR: logistic regression; SVM: support vector machine; MLP: multilayer perceptron; KNN: K-nearest neighbour; RFC: random forest classification; LGBM: light gradient boosting machine; CatBoost: categorical boosting; XGBoost: extreme gradient boosting.

Discussion

To our knowledge, this is the first study to apply ML to develop a predictive model for pSjD-associated depressive symptoms. Frequent fatigue, lower lymphocyte count, shorter sleep duration, and anti-Ro52 positivity were independently associated with higher risk. The SVM achieved superior performance (average AUROC 0.921, precision 0.904, accuracy 0.878), supporting its utility for early identification and timely intervention in clinical practice. Chronic fatigue, pain, and depression commonly co-occur in pSjD and interact to impair quality of life; whereas pain is often framed as somatic, fatigue and depression are typically viewed as

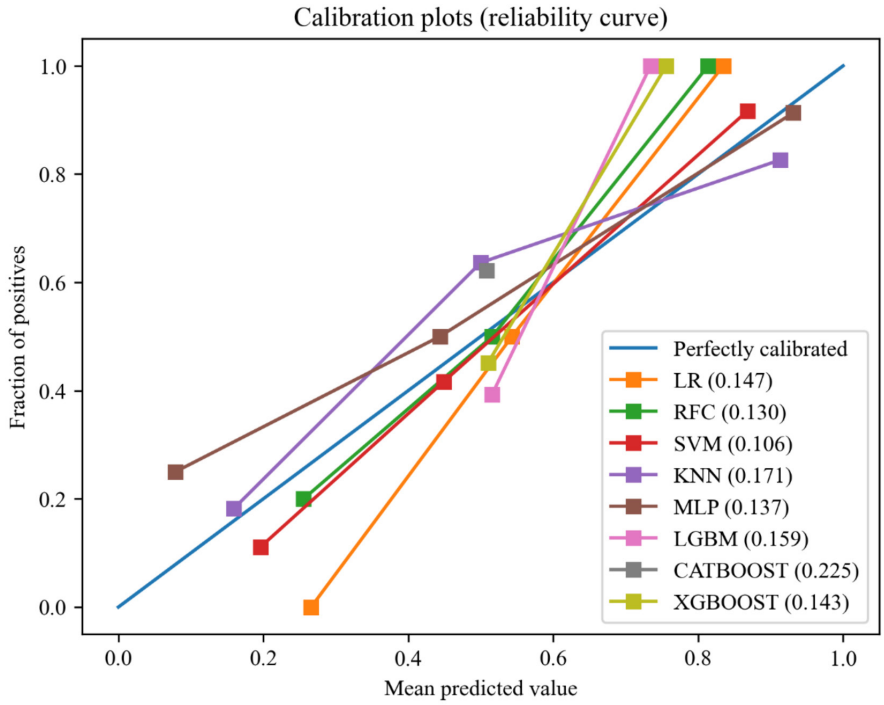

Fig. 3. Calibration plot of machine learning models.
LR: logistic regression; SVM: support vector machine; MLP: multilayer perceptron; KNN: K-nearest neighbour; RFC: random forest classification; LGBM: light gradient boosting machine; CatBoost: categorical boosting; XGBoost: extreme gradient boosting.

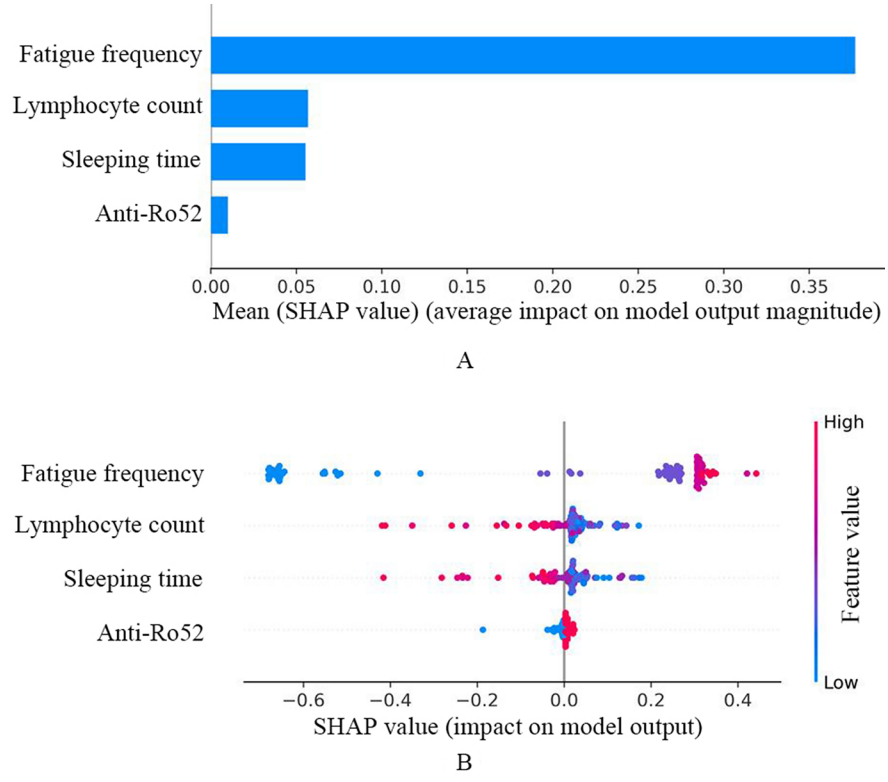

Fig. 4. SHAP plot for optimal SVM model.

psychological (26). Depression correlates strongly with sleep disturbance in pSjD, with poorer sleep and lower quality of life linked to higher depression, suggesting psychoneuroimmunological crosstalk (27). Poor sleep also associates with greater fatigue; notably, about one-third of patients report poor sleep even without depression, indicating sleep-symptom effects independent of mood (28). Proposed mechanisms include hypothalamic-pituitary-adrenal (HPA) axis dysregulation and immune perturbation that contribute to fatigue and sleep disruption in both pSjD and depression (25, 29). Systemic inflammation, reflected by elevated pro-inflammatory cytokines, may alter sleep architecture and mood, creating a self-reinforcing cycle (30); imbalances between pro- and anti-inflammatory cytokines can impair neuronal function, promoting pain, fatigue, and depressive symptoms (31).

Lymphopenia in pSjD reflects heightened systemic inflammation rather than immunosuppression. pSjD features B-cell hyperactivation and autoantibody production (*e.g.* anti-SSA/SSB), with lymphocyte infiltration of exocrine and extra-glandular tissues (32, 33). Reduced circulating lymphocytes, especially $CD4^+$ T cells, may indicate exhaustion or tissue-targeted trafficking and coincide with increased IL-6, TNF-α, and IFN-γ (34). These cytokines can disrupt blood-brain barrier integrity, access the central nervous system, and impair neuroplasticity in mood-regulatory regions (*e.g.* hippocampus, prefrontal cortex), thereby linking immune activation, neural circuitry, and depressive symptoms in pSjD (35).

Intriguingly, this study, characterised by the application of ML techniques for analysis, identified anti-Ro52 antibodies as a risk factor for depression accompanying pSjD. This result is consistent with the findings of our prior study, which employed traditional analytical methods (6). The anti-Ro52 antibody, prevalent in various connective tissue diseases (CTDs) and other autoimmune disorders, has received significant attention recently (36). It is well recognized that anti-Ro52 is associated with disease activity, leukopenia, and interstitial lung disease (ILD) among CTD patients (37). The link between anti-Ro52 and depression in patients with pSjD remains unclear. Animal studies have shown Ro52 expression in various regions of the mouse brain, such as the hippocampus, cerebral cortex, and cerebellum. Moreover, anti-

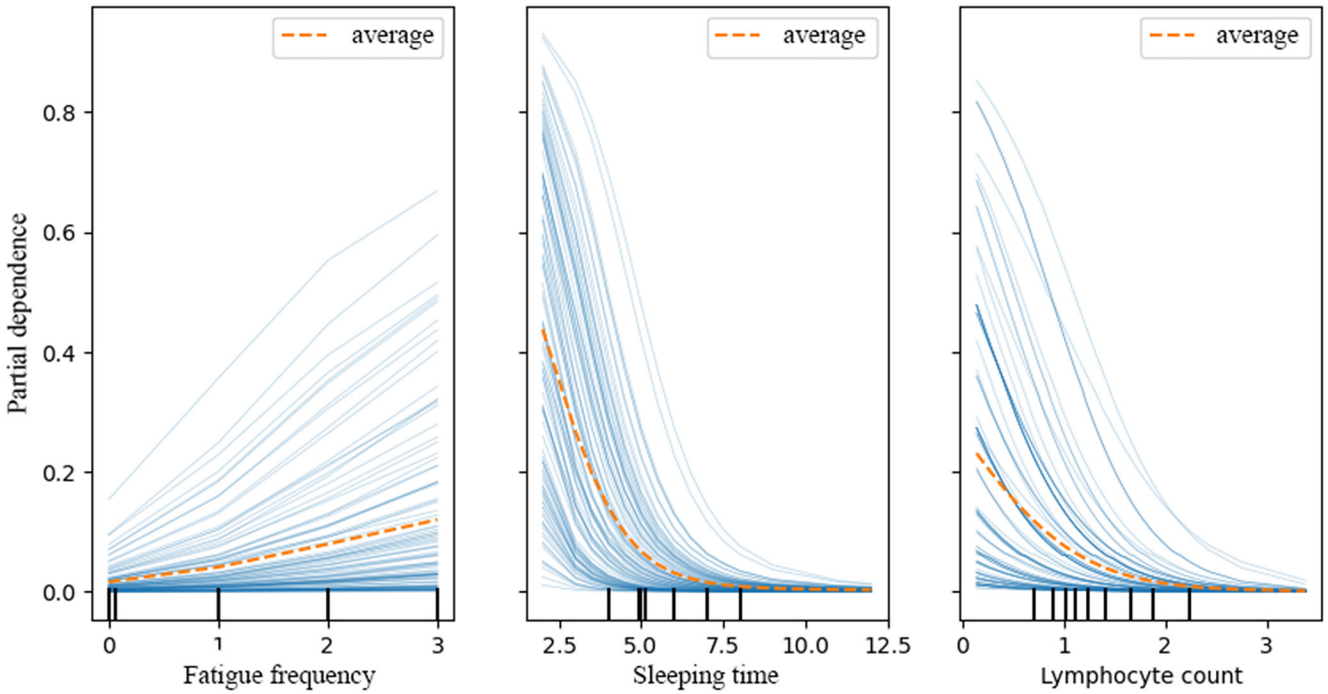

Fig. 5. Features' partial dependency plots.

Ro52 was detected in the cerebrospinal fluid of an SS patient with symptoms of cerebellar degeneration (38), suggesting a potential role for anti-Ro52 in cerebellar degeneration among SS patients, which may contribute to the onset of depressive symptoms. However, the observed relationship between anti-Ro52 and depressive symptoms should be interpreted cautiously as an association rather than causation. Temporality cannot be established in this cross-sectional design, and residual confounding (*e.g.* overall disease activity, treatment exposure, fatigue severity, sleep disturbance) may partially account for the signal. Therefore, the role of anti-Ro52 in pSjD co-occurring with depressive symptoms warrants further prospective and longitudinal validation.

ML models are increasingly utilized in medical diagnosis and prognosis across diverse settings. This study employed eight analytical models: LR, SVM, MLP, KNN, RFC, LGBM, CatBoost, and XGBoost, with the SVM model showing superior performance by achieving optimal average AUROC, precision, and accuracy. SVM relies extensively on vector spaces, transformations, and geometric concepts to determine the optimal division in multidimensional space (39). This feature potentially makes it particularly well-suited for complex medical data characterised by high-dimensional features (40). The efficacy of an SVM model in accurately classifying patients hinges on its ability to define a hyperplane that optimally separates various classes within the dataset. The bootstrap analysis intentionally exposes variability and optimism in small samples; consequently, intervals widen and apparent rank differences attenuate. This transparency reduces the risk of over-claiming performance and supports our decision to prioritise a low-dimensional logistic model with comparable discrimination and better deployability. Internal resampling improves credibility but cannot replace external validation, which remains essential for transportability.

There are no documented models for predicting depression in pSjD patients. Currently, only one study analyses variables linked to patient disability using logistic regression. The study found that patient disability was associated with depression, fatigue, and particularly lack of stamina (41). Our study also identified frequent fatigue and short sleep duration as high-risk factors for developing depressive symptoms. Additionally, anti-Ro52 antibodies and lymphocyte count were identified as risk factors. Further prospective research is required to confirm the therapeutic implications of these factors for depression. Nevertheless, physicians should carefully consider these elements when devising treatment regimens for pSjD patients. With external validation for clinical utility, this model could serve as a potent tool for delivering personalised care and interventions for patients with pSjD comorbid with depressive symptoms.

Our study has important limitations. This retrospective, single-centre study with a modest, geographically constrained sample limits external validity and transportability. Although we mitigated overfitting with a parsimonious four-feature model, repeated stratified cross-validation, and 1,000-bootstrap optimism correction, these internal checks do not replace external validation; the model is therefore hypothesis-generating and requires evaluation in larger, prospectively assembled multicentre cohorts before any clinical implementation. The context of depression assessment also matters: PHQ-9 was used as a screen on hospital day-1 to reflect clinical triage, and scores can be elevated by acute stressors, including hospitalisation itself (42). Thus, screen-positivity may reflect a mixture

of major depression and situational distress, underscoring the role of PHQ-9 as an initial flag rather than a diagnostic endpoint. Additionally, we did not capture all potential predictors (*e.g.* coping, social support, mental health literacy), and information bias, especially recall bias in fatigue frequency and sleep duration, with known gaps between perceived and actual sleep, may persist. Future studies should incorporate broader psychosocial measures, objective sleep metrics, and rigorous external validation.

In conclusion, this study used ML to develop an accurate predictive model for predicting the onset of depressive symptoms in patients with pSjD, indicating that factors like frequent fatigue, lymphopenia, reduced sleep, and anti-Ro52 antibodies increase the likelihood of developing depressive symptoms in pSjD patients. This ML model could serve as a foundational tool for elucidating the complex relationship between pSjD and depressive symptoms. Insights from this model could potentially improve prediction and treatment strategies for pSjD patients, ultimately enhancing their quality of life.

References

1. RAMOS-CASALS M, BAER AN, BRITO-ZERÓN MDP *et al*.: 2023 International Rome consensus for the nomenclature of Sjögren disease. *Nat Rev Rheumatol* 2025; 21(7): 426-37. https://doi.org/10.1038/s41584-025-01268-z
2. PRICE EJ, BENJAMIN S, BOMBARDIERI M *et al*.: British Society for Rheumatology guideline on management of adult and juvenile onset Sjögren disease. *Rheumatology* (Oxford) 2025; 64(2): 409-39. https://doi.org/10.1093/rheumatology/keae152
3. NEGRINI S, EMMI G, GRECO M *et al*.: Sjögren's syndrome: a systemic autoimmune disease. *Clin Exp Med* 2022; 22(1): 9-25. https://doi.org/10.1007/s10238-021-00728-6
4. HUANG T, LI Y, LUO Y *et al*.: Research progress on the pathogenesis and quality of life of patients with primary Sjögren's syndrome complicated by depression. *Clin Exp Rheumatol* 2022; 40(3): 647-54. https://doi.org/10.55563/clinexprheumatol/nspi47
5. COVID-19 MENTAL DISORDERS COLLABORATORS: Global prevalence and burden of depressive and anxiety disorders in 204 countries and territories in 2020 due to the COVID-19 pandemic. *Lancet* 2021; 398(10312): 1700-12. https://doi.org/10.1016/s0140-6736(21)02143-7
6. YANG L, WANG X, KANG H *et al*.: Association of anti-Ro52 antibody with depression and anxiety in patients with connective tissue diseases: an observational, single-centre, cross-sectional study. *Clin Exp Rheumatol* 2024; 42(7): 1350-58. https://doi.org/10.55563/clinexprheumatol/be9n92
7. MONTERO-LÓPEZ E, PERALTA-RAMíREZ MI, ORTEGO-CENTENO N *et al*.: Does stress response axis activation differ between patients with autoimmune disease and healthy people? *Stress Health* 2024; 40(4): e3392. https://doi.org/10.1002/smi.3392
8. DU W, HAN M, ZHU X *et al*.: The multiple roles of BcCells in the pathogenesis of Sjögren's syndrome. *Front Immunol* 2021; 12:684999. https://doi.org/10.3389/fimmu.2021.684999
9. CUI Y, XIA L, LI L *et al*.: Anxiety and depression in primary Sjögren's syndrome: a cross-sectional study. *BMC Psychiatry* 2018; 18(1): 131. https://doi.org/10.1186/s12888-018-1715-x
10. BÅRDSEN K, BREDE C, KVIVIK I *et al*.: Interleukin-1-related activity and hypocretin-1 in cerebrospinal fluid contribute to fatigue in primary Sjögren's syndrome. *J Neuroinflammation* 2019; 16(1): 102. https://doi.org/10.1186/s12974-019-1502-8
11. CUI Y, LI L, YIN R *et al*.: Depression in primary Sjögren's syndrome: a systematic review and meta-analysis. *Psychol Health Med* 2018; 23(2): 198-209. https://doi.org/10.1080/13548506.2017.1339895
12. RAJKOMAR A, DEAN J, KOHANE I: Machine Learning in Medicine. *N Engl J Med* 2019; 380(14): 1347-58. https://doi.org/10.1056/nejmra1814259
13. GREENER JG, KANDATHIL SM, MOFFAT L, JONES DT: A guide to machine learning for biologists. *Nat Rev Mol Cell Biol* 2022; 23(1): 40-55. https://doi.org/10.1038/s41580-021-00407-0
14. ZHONG M, ZHANG H, YU C, JIANG J, DUAN X: Application of machine learning in predicting the risk of postpartum depression: a systematic review. *J Affect Disord* 2022; 318: 364-79. https://doi.org/10.1016/j.jad.2022.08.070
15. SAJJADIAN M, LAM RW, MILEV R *et al*.: Machine learning in the prediction of depression treatment outcomes: a systematic review and meta-analysis. *Psychol Med* 2021; 51(16): 2742-51. https://doi.org/10.1017/S0033291721003871
16. COSTANTINI L, PASQUARELLA C, ODONE A *et al*.: Screening for depression in primary care with Patient Health Questionnaire-9 (PHQ-9): a systematic review. *J Affect Disord* 2021; 279: 473-83. https://doi.org/10.1016/j.jad.2020.09.131
17. SHIBOSKI CH, SHIBOSKI SC, SEROR R *et al*.: 2016 American College of Rheumatology/European League Against Rheumatism Classification Criteria for Primary Sjögren's Syndrome: a consensus and data-driven methodology involving three international patient cohorts. *Arthritis Rheumatol* 2017; 69(1): 35-45. https://doi.org/10.1002/art.39859
18. LEVIS B, BENEDETTI A, THOMBS BD: Accuracy of Patient Health Questionnaire-9 (PHQ-9) for screening to detect major depression: individual participant data meta-analysis. *BMJ* 2019; 365: l1476. https://doi.org/10.1136/bmj.l1476
19. KROENKE K, SPITZER RL, WILLIAMS JB: The PHQ-9: validity of a brief depression severity measure. *J Gen Intern Med* 2001; 16(9): 606-13. https://doi.org/10.1046/j.1525-1497.2001.016009606.x
20. HUA H, YANG K, WEI Z *et al*.: Depression and anxiety symptoms and associated factors among community residents in Shenzhen, China: a cross-sectional study. *BMC Public Health* 2025; 25(1). https://doi.org/10.1186/s12889-025-23241-x
21. WHITE IR, ROYSTON P, WOOD AM: Multiple imputation using chained equations: Issues and guidance for practice. *Stat Med* 2011; 30(4): 377-99. https://doi.org/10.1002/sim.4067
22. BEESLEY LJ, BONDARENKO I, ELLIOT MR *et al*.: Multiple imputation with missing data indicators. *Stat Methods Med Res* 2021; 30(12): 2685-700. https://doi.org/10.1177/09622802211047346
23. PONCE-BOBADILLA AV, SCHMITT V, MAIER CS, MENSING S, STODTMANN S: Practical guide to SHAP analysis: explaining supervised machine learning model predictions in drug development. *Clin Transl Sci* 2024; 17(11): e70056. https://doi.org/10.1111/cts.70056
24. SHAPLEY L: A Value for n-Person Games. Contributions to the Theory of Games II. 1953; 307-317. *In*: HAROLD WK (Ed.): Classics in Game Theory. Princeton: Princeton University Press, 1997, 69-79.
25. MENKE A: The HPA axis as target for depression. *Curr Neuropharmacol* 2024; 22(5): 904-15. https://doi.org/10.2174/1570159x21666230811141557
26. OMDAL R, MELLGREN SI, NORHEIM KB: Pain and fatigue in primary Sjögren's syndrome. *Rheumatology* (Oxford) 2021; 60(7): 3099-106. https://doi.org/10.1093/rheumatology/kez027
27. GOULABCHAND R, CASTILLE E, NAVUCET S *et al*.: The interplay between cognition, depression, anxiety, and sleep in primary Sjögren's syndrome patients. *Sci Rep* 2022; 12(1): 13176. https://doi.org/10.1038/s41598-022-17354-1
28. CHUNG SW, HUR J, HA YJ *et al*.: Impact of sleep quality on clinical features of primary Sjögren's syndrome. *Korean J Intern Med* 2019; 34(5): 1154-64. https://doi.org/10.3904/kjim.2017.158
29. MAVRAGANI CP, FRAGOULIS GE, MOUTSOPOULOS HM: Endocrine alterations in primary Sjögren's syndrome: an overview. *J Autoimmun* 2012; 39(4): 354-8. https://doi.org/10.1016/j.jaut.2012.05.011
30. LEE YT, WU CY, CHOU IJ, LIU CH: Sjögren's syndrome related autoimmune encephalopathy presenting as depression and involuntary agitation in a 15-year-old girl: a case report. *Front Immunol* 2023; 14: 1145948. https://doi.org/10.3389/fimmu.2023.1145948
31. GRYGIEL-GORNIAK B, LIMPHAIBOOL N, PUSZCZEWICZ M: Cytokine secretion and the risk of depression development in patients with connective tissue diseases. *Psychiatry Clin Neurosci* 2019; 73(6): 302-16. https://doi.org/10.1111/pcn.12826
32. SU C, WANG W, CHENG F, ZHAO F, ZHENG SG: The role of B cells in Sjögren's syndrome and their impact on the nervous system. *Autoimmunity Reviews* 2025; 24(8). https://doi.org/10.1016/j.autrev.2025.103852

33. LIAO J, YU X, HUANG Z *et al*.: Chemokines and lymphocyte homing in Sjögren's syndrome. *Frontiers in Immunology* 2024; 15. https://doi.org/10.3389/fimmu.2024.1345381
34. BALDINI C, FULVIO G, LA ROCCA G, FERRO F: Update on the pathophysiology and treatment of primary Sjögren syndrome. *Nat Rev Rheumatol* 2024; 20(8): 473-91. https://doi.org/10.1038/s41584-024-01135-3
35. BEUREL E, TOUPS M, NEMEROFF CB: The bidirectional relationship of depression and inflammation: double trouble. *Neuron* 2020; 107(2): 234-56. https://doi.org/10.1016/j.neuron.2020.06.002
36. DECKER P, MOULINET T, PONTILLE F *et al*.: An updated review of anti-Ro52 (TRIM21) antibodies impact in connective tissue diseases clinical management. *Autoimmun Rev* 2022; 21(3): 103013. https://doi.org/10.1016/j.autrev.2021.103013
37. CHAN EKL: Anti-Ro52 autoantibody is common in systemic autoimmune rheumatic diseases and correlating with worse outcome when associated with interstitial lung disease in systemic sclerosis and autoimmune myositis. *Clin Rev Allergy Immunol* 2022; 63(2): 178-93. https://doi.org/10.1007/s12016-021-08911-z
38. TETSUKA S, SUZUKI T, OGAWA T, HASHIMOTO R, KATO H: Anti-Ro/SSA antibodies may be responsible for cerebellar degeneration in Sjögren's syndrome. *J Clin Med Res* 2021; 13(2): 113-20. https://doi.org/10.14740/jocmr4429
39. BRERETON RG, LLOYD GR: Support vector machines for classification and regression. *Analyst* 2010; 135(2): 230-67. https://doi.org/10.1039/b918972f
40. GUO CY, CHOU YC: A novel machine learning strategy for model selections - Stepwise Support Vector Machine (StepSVM). *PLoS One* 2020; 15(8): e0238384. https://doi.org/10.1371/journal.pone.0238384
41. WESTHOFF G, DÖRNER T, ZINK A: Fatigue and depression predict physician visits and work disability in women with primary Sjögren's syndrome: results from a cohort study. *Rheumatology* (Oxford) 2012; 51(2): 262-69. https://doi.org/10.1093/rheumatology/ker208
42. ELSHAFEI A, JEMUTAI J, TICKELL KD *et al*.: Childhood Acute Illness and Nutrition (CHAIN) Network Investigators. Depressive symptoms among mothers with hospitalized children in South Asia and sub-Saharan Africa. *J Affect Disord* 2023; 323: 368-77. https:// doi.org/10.1016/j.jad.2022.10.051