



Using Machine Learning Model to Predict Anxiety in Systemic Lupus Erythematosus Patients

Mohammed Omari^{1*}, Ibtissam El Harch², Noura Qarmiche¹, Hind Bourkhime¹, Nouhaila Charef¹, Soumaia Elghazi¹, Samira El Fakir², Nada Otmani¹

¹Medical Informatics and Data Science Unit, Laboratory of Epidemiology, Clinical Research and Community Health, Faculty of Medicine, Pharmacy and Dentistry of Fez, Sidi Mohamed Ben Abdellah University, Fez, Morocco

²Laboratory of Epidemiology, Clinical Research and Community Health, Faculty of Medicine, Pharmacy and Dentistry of Fez, Sidi Mohamed Ben Abdellah University, Fez, Morocco

Email: *mohammed.omari@usmba.ac.ma, elibtissam1@gmail.com, noura.qarmiche@usmba.ac.ma, hbourkhime@gmail.com, nouhaila.charef@usmba.ac.ma, soumaia.elghazi@usmba.ac.ma, samira.elfakir@usmba.ac.ma, nada.otmani@usmba.ac.ma

How to cite this paper: Omari, M., El Harch, I., Qarmiche, N., Bourkhime, H., Charef, N., Elghazi, S., El Fakir, S. and Otmani, N. (2025) Using Machine Learning Model to Predict Anxiety in Systemic Lupus Erythematosus Patients. *Open Access Library Journal*, 12: e13039.

<https://doi.org/10.4236/oalib.1113039>

Received: February 4, 2025

Accepted: April 13, 2025

Published: April 16, 2025

Copyright © 2025 by author(s) and Open Access Library Inc.

This work is licensed under the Creative Commons Attribution International License (CC BY 4.0).

<http://creativecommons.org/licenses/by/4.0/>



Open Access

Abstract

Introduction: Systemic Lupus Erythematosus (SLE) is a chronic autoimmune disease that affects multiple organs and significantly impacts quality of life, particularly in women, with a global incidence of 5.14 cases per 100,000 person-years. Many SLE patients experience psychiatric complications, such as anxiety, with 55.4% affected in Morocco. To improve patient management and early detection of anxiety, a study proposes developing a machine learning algorithm to analyze patient data and identify predictive patterns. **Method:** A cross-sectional study conducted at the Hassan II University Hospital examined adults with systemic lupus erythematosus (SLE) between 2020 and 2023, excluding those with psychiatric disorders or communication problems. Data was collected through questionnaires assessing socio-demographic and clinical factors, including anxiety levels, using the Moroccan HADS. Machine learning techniques were used to analyse the data and predict anxiety. **Results:** The study found a significant positive association between anxiety and factors such as age, age of diagnosis, symptom duration, marital status, pre-existing conditions and respiratory problems. Model performance varied, with logistic regression achieving the highest accuracy (0.67) and recall (0.90), while random forest had the lowest accuracy (0.57) and recall (0.50). Precision scores ranged from 0.54 for SVM to 0.60 for logistic regression and decision tree, with F1 scores between 0.53 and 0.72. **Conclusion:** The study highlights the need to predict and treat anxiety in lupus patients, using logistic regression as an effective. Addressing anxiety through targeted care strategies can significantly improve the quality of life and overall well-being of lupus patients.

Subject Areas

Global Health, Machine Learning

Keywords

Systemic Lupus Erythematosus (SLE), Anxiety, Machine Learning, Prediction

1. Introduction

Systemic Lupus Erythematosus (SLE) is a chronic autoimmune disease that can affect various organs of the body, such as the skin, articulations, kidneys, heart and brain. Symptoms of SLE can vary considerably from person to person, ranging from mild to severe, and can have a significant impact on patient's quality of life [1].

Although lupus is considered to be a relatively uncommon disease, it cannot be classified as a rare condition. Worldwide, the incidence of systemic lupus erythematosus (SLE) is estimated at 5.14 cases per 100,000 person-years, representing around 0.40 million new diagnoses each year. Women are particularly affected, with an incidence of 8.82 cases per 100,000 person-years and around 0.34 million new cases annually, compared with 1.53 cases per 100,000 person-years in men, or 0.06 million cases. The worldwide prevalence of SLE is estimated at 43.7 cases per 100,000 people, affecting around 3.41 million people, of whom 3.04 million are women. Variation in the incidence and prevalence of lupus is also observed across regions, with the highest rates reported in Poland, the United States and Barbados [2].

Several complications can be present in lupus erythematosus patients, they can be organic, or psychiatric like depression and anxiety with an important impact on the quality of life [3], In Morocco, the prevalence of anxiety in SLE patients is estimated at 55.4% (CI 95%: 45.8% - 65%) [4]. In order to improve the quality of life of these patients, It is therefore essential to be able to predict and anticipate the development of anxiety in SLE patients. Such prediction would improve patient management and help prevent this serious complication.

From this perspective, the use of Machine Learning is proving to be a promising approach for predicting anxiety in SLE patients. In fact, machine learning has revolutionized the medical field, offering advanced tools for predicting diseases and their complications [5]. Thanks to machine learning algorithms, it is possible to analyze large quantities of patient data, in order to identify patterns and predict the occurrence or evolution of a disease [6]. These predictive models enable the early detection of disease, the development of personalized treatment plans and the prevention of potential complications. By integrating machine learning into the medical field, healthcare delivery is transformed and patient care is improved. In the specific case of systemic lupus erythematosus, the use of machine learning to predict anxiety in patients is showing promise. Algorithms could accurately

predict the development of anxiety in patients suffering from this disease. At present, there are no published studies on the prediction of anxiety in SLE patients.

The objective of our study is to develop a machine learning algorithm able to predict anxiety in SLE patients.

2. Material and Method

A cross-sectional study was conducted in the internal medicine department of the Hassan II University Hospital in Fez in 2023.

All patients aged 18 and over, diagnosed with SLE according to ACR criteria [7], and admitted to the internal medicine department at CHU HASSAN II FES between 2020 and 2023 were included in the study.

Patients with a history of depression or anxiety, patients with psychiatric disorders including schizophrenia or bipolar states, and patients with comprehension or communication difficulties were excluded.

2.1. Data Collection

All data were collected during face-to-face interviews using a questionnaire containing information relating to their socio-demographic characteristics (age, gender, marital status, education level and employment status) and their past history. Information on disease characteristics, such as duration, different manifestations, autoantibody status (anti-DNA, anti-nuclear) and types of treatment used, was obtained by interviewing patients and examining their medical records. Anxiety was measured using the Moroccan version of the Hospital anxiety and depression scale (HADS) [8].

2.1.1. Anxiety Measure (HADS Scale)

The dependent variable is the level of anxiety measured by the Hospital Anxiety and Depression Scale (HADS).

The HADS was developed by Zigmond and Snaith in 1983 to screen non-psychiatric inpatients for anxiety disorders and depressive syndromes, but was later validated for outpatient use. It is a self-assessment scale that identifies anxiety and depressive disorders. It comprises 14 items rated from 0 to 3. About 7 questions are related to anxiety (total A) and 7 others concern the depressive dimension (total D). For each item, the response is rated from 0 to 3 on a scale depending on the intensity of the symptom over the past week. The range of possible scores for each subscale thus extends from 0 to 21, with higher scores corresponding to the presence of more severe symptoms. For each subscale (anxiety and depression), threshold values have been determined: a score between 0 and 7 is considered normal, while a score of 8 or more indicates a significant disorder [9] [10].

2.1.2. Explanatory Variables

The explanatory variables studied are essentially sociodemographic characteristics, clinical features of lupus (duration of disease, severity, etc.), and types of treatment used. **Table 1** shows the description of these attributes and their values.

2.2. Data Pre-Processing

Microsoft Excel was used in this study for data entry and organization. Python (version 3.6.4), preprocessing and prediction algorithm development. The main software package used for the project was scikit-learn in Jupyter notebook.

Table 1. Characteristics of the sample.

Variables	classes	Total Eff. (%)	Anxiety		P-value
			Yes Eff. (%)	No Eff. (%)	
Total			57 (55.9)	45 (44.1)	
Sex					
	Men	8 (07.2)	3 (37.5)	5 (62.5)	0.300
	Women	94 (92.8)	54 (57.4)	40 (42.6)	
Age (year) mean \pm SD		41.64 \pm 13.75	45.65 \pm 13.73	36.56 \pm 12.11	0.010
Statut matrimonial					
	Single	27 (26.5)	8 (26.6)	19 (70.4)	0.006
	Divorced	14 (13.7)	9 (64.3)	5 (35.7)	
	Married	61 (59.8)	40 (65.6)	21 (34.4)	
Monthly Income					
	<2000 DH	70 (68.6)	43 (61.4)	27 (38.6)	0.100
	>2000 DH	32 (31.4)	14 (43.8)	18 (56.3)	
Geographical origin (populated areas)					
	Rural	32 (31.4)	16 (50.0)	16 (50.0)	0.42
	Urbain	70 (68.6)	41 (58.6)	29 (41.4)	
Number of previous medical conditions					
	0	33 (32.4)	14 (42.4)	19 (57.6)	0.037
	1	35 (34.3)	17 (48.6)	18 (51.4)	
	3	24 (23.5)	17 (70.8)	7 (29.2)	
	4	8 (7.8)	7 (87.5)	1 (12.5)	
	5	2 (2.0)	2 (100)	0 (0.0)	
Dermatological manifestations					
	Yes	30 (29.4)	16 (53.3)	14 (46.7)	0.740
	No	70 (70.6)	41 (56.9)	31 (43.1)	
Rheumatological manifestations					
	No	36 (35.29)	18 (50.0)	18 (50.0)	0.380
	Yes	66 (64.71)	39 (59.1)	27 (40.9)	
Neuropsychological manifestations					
	No	92 (90.2)	50 (54.3)	42 (45.7)	0.510
	Yes	10 (09.8)	7 (70.0)	3 (30.0)	
Respiratory symptoms					

Continued

	No	81 (79.4)	41 (50.6)	40 (49.4)	0.035
	Yes	21 (20.6)	16 (76.2)	5 (23.8)	
Ophthalmic manifestations					
	No	77 (75.5)	42 (54.5)	35 (45.5)	0.630
	Yes	25 (24.5)	15 (60.0)	10 (40.0)	
Facial damage					
	No	53 (52.0)	28 (52.8)	25 (47.2)	0.520
	Yes	49 (48.0)	29 (59.2)	20 (40.8)	
hematologic diseases					
	No	74 (72.5)	43 (58.1)	31 (41.9)	0.460
	Yes	28 (27.5)	14 (50.0)	14 (50.0)	

Quantitative data were normalized using sklearn.preprocessing's StandardScaler function. Categorical variables were encoded using the LabelEncoder function in the sklearn.preprocessing module.

There were no missing values or outliers in our database.

In this study, two feature selection techniques were used: Recursive Feature Elimination with Cross-Validation (RFECV) and Random Forest Importance.

The Sklearn train_test_split function was used to randomly divide the data into two subsets: a training set (80%) and a test set (20%).

2.3. Model Development

2.3.1. Algorithms Used

In our analysis, we employed four well-established supervised machine learning algorithms: Support Vector Machine (SVM), which is effective for high-dimensional spaces and works well for both linear and non-linear classification by finding the optimal hyperplane that separates different classes; Logistic Regression, a statistical method used for binary classification that predicts the probability of an outcome based on input features, making it interpretable and efficient for large datasets; Decision Tree, a versatile model that splits data into subsets based on feature values, making it easy to visualize and understand the decision-making process while capturing non-linear relationships; and Random Forest, an ensemble method that combines multiple decision trees to enhance predictive accuracy and control overfitting, leveraging the wisdom of the crowd to produce more robust and reliable predictions.

2.3.2. Performance Evaluation

Several measures were used to evaluate the performance of the machine learning algorithms in this study, namely accuracy, precision, recall and F-score. These metrics serve as standard measures for evaluating the effectiveness of classification models and are defined by the following equations: Accuracy (Number of correct predictions/Total number of predictions), Precision ($TP/(TP + FP)$), Recall ($TP/(TP$

+ FN)), and F1-score ($2 * (\text{Precision} * \text{Recall}) / (\text{Precision} + \text{Recall})$).

2.4. Ethical Considerations

Participants' anonymity and confidentiality were guaranteed. All participants gave written informed consent, and the study was approved by the Ethics Committee of the Hassan II University Hospital in Fez.

3. Results

In this study, a significant positive correlation was found between anxiety and several variables, including age, age at the time of diagnosis, duration of symptom progression, marital status, number of pre-existing conditions, and respiratory manifestations. Concurrently, a high negative correlation was observed between anxiety and educational level, sex, occupation, and monthly income. as shown in

Figure 1.

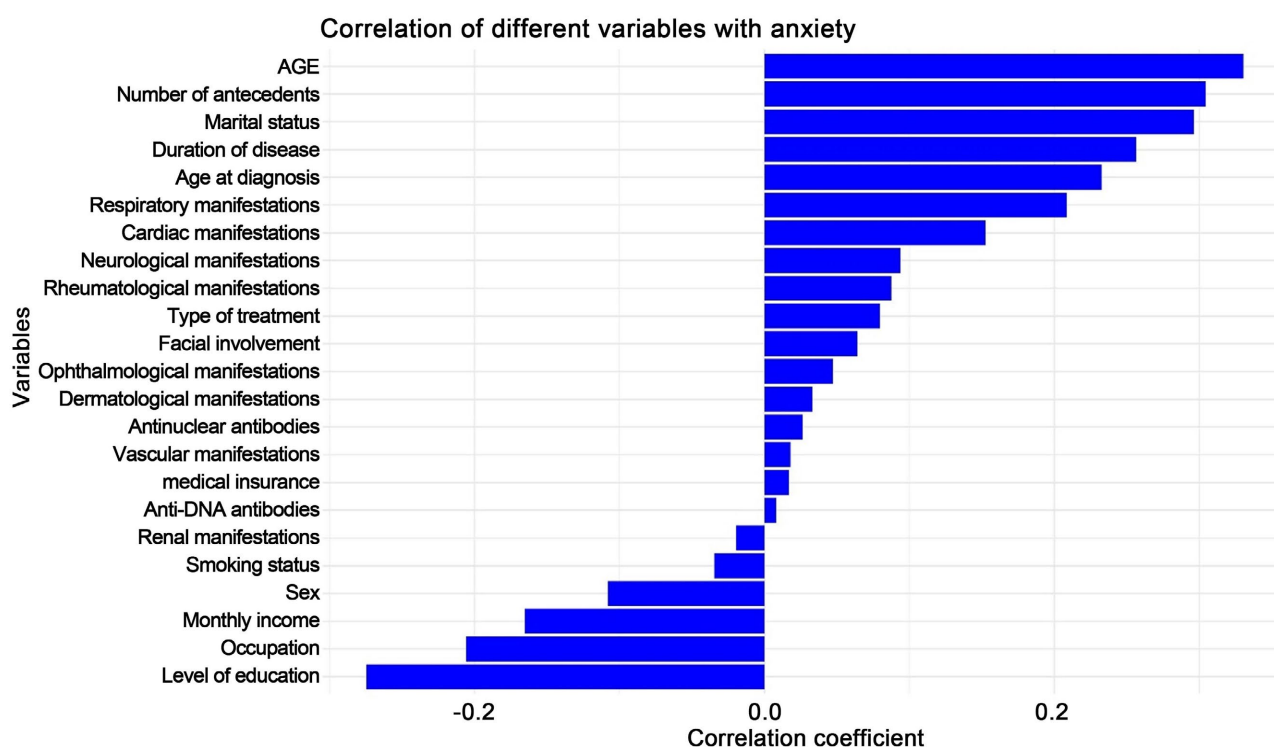


Figure 1. Correlation analysis between anxiety and socioeconomic factors.

Table 2. Comparison of model performance metrics: Accuracy, Precision, Recall, and F1 Score.

Method	Precision	Recall	F1-score	Accuracy
SVM	0.54	0.70	0.61	0.57
logistic Regression	0.60	0.9	0.72	0.67
Decision tree	0.60	0.60	0.60	0.62
Random forest	0.56	0.50	0.53	0.57

The model performance evaluation reveal variable measures of accuracy, ranging from 0.57 for random forest to 0.67 for logistic regression. In terms of precision, values ranged from 0.54 for SVM to 0.60 for logistic regression and decision tree. Recall ranged from 0.50 for random forest to 0.90 for logistic regression, while F1 score oscillated between 0.53 for random forest and 0.72 for logistic regression. as shown in **Table 2**.

4. Discussion

The aim of this work is to predict anxiety in lupus patients using machine learning, which could have a significant impact on the management of this autoimmune disease. The results of our study revealed important information about the performance of the different machine learning models used.

Anxiety was present in 55.9% of patients studied, indicating that it is a common problem in people with lupus. This underlines the importance of detecting and predicting this condition in these patients, so that appropriate interventions can be put in place to help them.

The results of the model performance evaluation showed accuracy rates ranging from 0.57 for random forest to 0.67 for logistic regression. These results suggest that logistic regression was the best-performing model in this study. These results demonstrate that machine learning models can be successfully used to predict anxiety in lupus patients.

It is interesting to compare these results with other similar studies that have looked at the prediction or detection of anxiety in the general population or in populations with other chronic illnesses. For example, a study on the prediction of anxiety in cancer survivors [11] showed accuracy rates ranging from 0.69 to 0.70 for different prediction models. These results are not far from those of the current study. Other studies have found accuracy values between 0.72 and 0.85 [12] and between 0.71 and 0.97 [13],

Furthermore, the accuracy values in this study ranged from 0.57 to 0.66, indicating a certain stability in the results obtained. However, it is important to note that recall rates varied between 0.50 and 0.90 and F1 scores between 0.53 and 0.72. These values show some fluctuation in model performance, which may be attributed to the complexity of predicting anxiety in these patients.

The results of Priya *et al.* [14] show variable model performance, with Random Forest achieving 71.4% accuracy for anxiety, which is lower than our best performance with logistic regression. This may indicate that the characteristics of the data and the study setting have a significant impact on the performance of the models.

On the other hand, the results of Wei *et al.* [15] suggest that the Random Forest and Multilayer Perceptron models have similar net benefits, highlighting the need to test multiple approaches for optimal results. The performance of Random Forest in our study could be improved by better feature selection, as suggested by the research of Zhou *et al.* [16] who achieved a high F1 score with a Random Forest

model in an emotional classification framework.

Finally, our results highlight the complexity of predicting anxiety, as shown by the analysis of Mahalingam *et al.* [17] where several algorithms were tested to predict anxiety, with different results depending on the algorithm used. This highlights the importance of choosing the right model according to the specificities of the data and the clinical characteristics of the populations studied.

In this study, data collection was carried out in an internal medicine department of the CHU, which is recognized as a reference center for the management of lupus in the region. This feature suggests that the sample used in the study is representative of the lupus patient, which reinforces the reliability and generalizability of the results obtained.

In addition, a validated and reliable scale was used to measure anxiety in patients. This ensures that the measures collected are accurate and consistent, enabling a relevant assessment of the influence of lupus on patients' anxiety.

Another strength of this discussion lies in the use of different measures to assess the performance of models employed in predicting anxiety in lupus patients. This multi-measurement approach provides a comprehensive and rigorous view of model effectiveness, and strengthens the validity of the results obtained.

However, there are a number of limitations to this study. First, the sample size used for the study is small. A larger sample size would have provided more robust results.

It is important to highlight that the proposed model was not tested by external validation, which implies that the results obtained may be specific to the sample studied and may not be applicable to other lupus patient populations. External validation would have strengthened the reliability and validity of the study results.

To confirm the conclusions of the study, it would be necessary to conduct research with larger samples. Increasing the sample size would improve the reliability of the conclusions and predictions of anxiety in these patients.

5. Conclusion

This study highlights the crucial importance of predicting and treating anxiety in lupus patients. Indeed, the results underline the effectiveness of logistic regression as a predictive model, while acknowledging that there are opportunities for improvement in this approach. Anxiety can have a significant impact on the daily lives of these patients, affecting their ability to manage their disease and maintain social interactions. By adopting these predictive models and implementing appropriate treatment strategies, healthcare professionals will be better prepared to offer targeted care and adequate support to these patients. This could result in a better quality of life and increased well-being for people with lupus, enabling them to lead more fulfilling lives despite the challenges posed by their condition.

Conflicts of Interest

The authors declare no conflicts of interest.

References

- [1] Smith, P.P. and Gordon, C. (2010) Systemic Lupus Erythematosus: Clinical Presentations. *Autoimmunity Reviews*, **10**, 43-45. <https://doi.org/10.1016/j.autrev.2010.08.016>
- [2] Tian, J., Zhang, D., Yao, X., Huang, Y. and Lu, Q. (2023) Global Epidemiology of Systemic Lupus Erythematosus: A Comprehensive Systematic Analysis and Modeling Study. *Annals of the Rheumatic Diseases*, **82**, 351-356. <https://doi.org/10.1136/ard-2022-223035>
- [3] Meszaros, Z.S., Perl, A. and Faraone, S.V. (2012) Psychiatric Symptoms in Systemic Lupus Erythematosus: A Systematic Review. *The Journal of Clinical Psychiatry*, **73**, 993-1001. <https://doi.org/10.4088/jcp.11r07425>
- [4] Harch, I.E., Benmaamar, S., Oubelkacem, N., Jennane, R., Diagne, B.J., Maiouak, M., et al. (2022) Prevalence and Associated Factors with Anxiety and Depression in Patients with Systemic Lupus Erythematosus in a Moroccan Region. *Open Access Library Journal*, **9**, 1-14. <https://doi.org/10.4236/oalib.1108394>
- [5] Deo, R.C. (2015) Machine Learning in Medicine. *Circulation*, **132**, 1920-1930. <https://doi.org/10.1161/circulationaha.115.001593>
- [6] Handelman, G.S., Kok, H.K., Chandra, R.V., Razavi, A.H., Lee, M.J. and Asadi, H. (2018) eDoctor: Machine Learning and the Future of Medicine. *Journal of Internal Medicine*, **284**, 603-619. <https://doi.org/10.1111/joim.12822>
- [7] Aringer, M., Costenbader, K.H., Daikh, D.I., Brinks, R., Mosca, M., Ramsey-Goldman, R., et al. (2019) 2019 EULAR/ACR Classification Criteria for Systemic Lupus Erythematosus. *Arthritis & Rheumatology*, **71**, 1400-1412.
- [8] Bendahhou, K., Serhir, Z., Ibrahim Khalil, A., Radallah, D., Amegrissi, S., Battas, O., et al. (2017) Validation de la version dialectale Marocaine de l'échelle «HADS». *Revue d'Épidémiologie et de Santé Publique*, **65**, S53. <https://doi.org/10.1016/j.respe.2017.03.016>
- [9] Botega, N.J., Bio, M.R., Zomignani, M.A., Garcia Jr, C. and Pereira, W.A.B. (1995) Transtornos do humor em enfermaria de clínica médica e validação de escala de medida (HAD) de ansiedade e depressão. *Revista de Saúde Pública*, **29**, 359-363. <https://doi.org/10.1590/s0034-89101995000500004>
- [10] Zigmond, A.S. and Snaith, R.P. (1983) The Hospital Anxiety and Depression Scale. *Acta Psychiatrica Scandinavica*, **67**, 361-370. <https://doi.org/10.1111/j.1600-0447.1983.tb09716.x>
- [11] Yan, R., Wang, J., Yang, X. and Yu, J. (2020) Prediction of Comorbid Anxiety and Depression Using Machine Learning Models in Cancer Survivors. <https://doi.org/10.21203/rs.3.rs-32449/v1>
- [12] Sau, A. and Bhakta, I. (2017) Predicting Anxiety and Depression in Elderly Patients Using Machine Learning Technology. *Healthcare Technology Letters*, **4**, 238-243. <https://doi.org/10.1049/htl.2016.0096>
- [13] Ahmed, A., Sultana, R., Ullas, M.T.R., Begom, M., Rahi, M.M.I. and Alam, M.A. (2020) A Machine Learning Approach to Detect Depression and Anxiety Using Supervised Learning. 2020 *IEEE Asia-Pacific Conference on Computer Science and Data Engineering (CSDE)*, Gold Coast, 16-18 December 2020, 1-6. <https://doi.org/10.1109/csde50874.2020.9411642>
- [14] Priya, A., Garg, S. and Tigga, N.P. (2020) Predicting Anxiety, Depression and Stress in Modern Life Using Machine Learning Algorithms. *Procedia Computer Science*, **167**, 1258-1267. <https://doi.org/10.1016/j.procs.2020.03.442>

- [15] Wei, Z., Wang, X., Ren, L., Liu, C., Liu, C., Cao, M., *et al.* (2023) Using Machine Learning Approach to Predict Depression and Anxiety among Patients with Epilepsy in China: A Cross-Sectional Study. *Journal of Affective Disorders*, **336**, 1-8. <https://doi.org/10.1016/j.jad.2023.05.043>
- [16] Zhou, Y., Han, W., Yao, X., Xue, J., Li, Z. and Li, Y. (2023) Developing a Machine Learning Model for Detecting Depression, Anxiety, and Apathy in Older Adults with Mild Cognitive Impairment Using Speech and Facial Expressions: A Cross-Sectional Observational Study. *International Journal of Nursing Studies*, **146**, Article ID: 104562. <https://doi.org/10.1016/j.ijnurstu.2023.104562>
- [17] Mahalingam, M., Jammal, M., Hoteit, R., Ayna, D., Romani, M., Hijazi, S., *et al.* (2023) A Machine Learning Study to Predict Anxiety on Campuses in Lebanon. In: *Studies in Health Technology and Informatics*, IOS Press, 85-88. <https://doi.org/10.3233/shti230430>