Machine Learning Classification of Neuropsychiatric Systemic Lupus Erythematosus patients using resting-state fMRI functional connectivity

N. J. Simos
Department of Electrical and Computer Engineering
Technical University of Crete
Chania, Greece
Computational Bio-Medicine Laboratory
Foundation for Research and Technology – Hellas
Heraklion, Greece
nsimos@isc.tuc.gr

E. Kavroulakis
Department of Radiology
Medical School, University of Crete
Heraklion, Greece

G. C. Manikis
Computational Bio-Medicine Laboratory
Foundation for Research and Technology – Hellas
Heraklion, Greece
gmanikis@ics.forth.gr

G. Bertsias
Department of Rheumatology
Medical School, University of Crete
Heraklion, Greece
Insitute of Molecular Biology and Biotechnology
Foundation for Research and Technology – Hellas
Heraklion, Greece
gbertsias@uoc.gr

E. Papadaki
Department of Radiology
Medical School, University of Crete
Heraklion, Greece
f.papadaki@med.uoc.gr

K. Marias
Computational Bio-Medicine Laboratory
Foundation for Research and Technology – Hellas
Heraklion, Greece
Department of Electrical and Computer Engineering
Hellenic Mediterranean University
Heraklion, Greece
kmarias@ics.forth.gr

***Abstract*— In this study we explored the robustness of machine learning algorithms for the classification of Neuropsychiatric systemic lupus erythematosus (NPSLE) patients and healthy controls using resting-state fMRI functional connectivity matrices. NPSLE, which is driven by systemic autoimmune inflammation in the context of lupus, involves a wide range of focal and diffuse central and peripheral nervous system symptoms and poses significant diagnostic challenges. Machine learning applications on clinical data may enhance the existing workflow for NPSLE classification as there is no established method of applying neuroimaging data to the diagnosis of NPSLE. Feature selection methods were applied prior to the classification process in order to perform the classification process on a lower dimension feature space. The Connectivity Matrix used consisted of pairwise regional functional associations of the fMRI signals (ROI to ROI correlations) within each of three predetermined brain networks in 41 NPSLE patients and 31 healthy control subjects. Support Vector Machines (SVM) was utilized in the final model. Results were evaluated using a nested cross validation methodology to prevent overfitting, and enhance generalization. Regions of Interest (ROI's) that contributed most in the final model were: Right Inferior Temporal, Thalamus, Left Angular Gyrus, Right Precuneus, Left Primary Motor Cortex, SMA, Left and Right Primary Motor Cortex. With a final F1 score of up to 77%, the results demonstrate the potential for the future implementation of similar methods in the diagnosis of NPSLE.**

Keywords—Neuropsychiatric Systemic Lupus Erythematosus - NPSLE, resting-state fMRI, functional connectivity, resting-state networks, Machine Learning, Supervised Learning, Feature Selection, Support Vector Machines, Recursive Feature Elimination, Nested Cross Validation

I. Introduction

In recent years machine learning techniques (ML) have often been used for a wide range of classification tasks using resting-state functional connectivity (FC) fMRI data [1]. Resting-state fMRI has received notable attention lately for its great potential of providing biomarkers useful in disease diagnosis. It is a non-invasive examination method based on the blood-oxygen-level-dependent signals. What differentiates resting-state fMRI from other common fMRI modalities is that the subjects are not performing any specific task during the examination, mental or otherwise. FC can be considered as a transformation of the original recorded fMRI time series, quantified by simple metrics such as correlation, covariance and mutual information of time series computed between different brain regions. Systemic lupus erythematosus (SLE) is a chronic, multisystem, autoimmune disease. Depending on the type of manifestations included and the method used for evaluation, the frequency of neuropsychiatric (NP) manifestations varies widely across studies. Because of the lack of standardization techniques as well as NP manifestations often being due to secondary causes, the diagnosis of NPSLE may be challenging. Over the last decades, several protocols have attempted to overcome this issue [2].

Methods using ML algorithms based on FC data have shown promising results in the diagnosis of other diseases such as Multiple Sclerosis (MS) [1] as well as autism classification [3]. However, research conducted using such algorithms on lupus and specifically NPSLE is almost nonexistent. Furthermore, the technique used for the calculation of the FC matrices is similar to both node and voxel-based connectivity methods implemented in software packages such as CONN. It was implemented for this study and details are presented below in the methods section.

Support Vector Machines (SVM) [4] is one of the most robust and frequently used ML algorithms for classification tasks. It is often applied to a wide variety of biomedical classification problems including EEG classification, cancer identification and seizure prediction. Particularly, SVM was

978-1-7281-3868-8/19/$31.00 ©2019 IEEE

proven an effective classifier in studies using rs-fMRI connectivity data [5].

Classification using SVM entails assigning all samples to two disjoint hyperplanes in a manner that maximizes the margin between the two hyperplanes. By maximizing the margin of the two hyperplanes, the "between-class" distance is also maximized by selecting the best fitted hyperplanes. This in turn maximizes distinction ability between the two classes leading to better classification ability.

In the present study we present and evaluate a method for the effective classification of NPSLE patients from healthy control subjects. In the data used, a set of brain networks were selected a priori and the FC matrices produced comprised the functional connections of all the regions of interest (ROI's) within these networks.

II. Materials And Methods

A. MRI acquisition

The data in this study was acquired during the period 2015-2016 from 41 patients suffering from Neuropsychiatric systemic lupus erythematosus (NPSLE) and 31 healthy control subjects (HC). NPSLE diagnosis was based on physician judgement, following multidisciplinary approach and considering patient age, European League Against Rheumatism (EULAR) risk factors for NPSLE (anti-phospholipid antibodies, prior neuropsychiatric manifestation, generalized disease activity, findings of conventional MRI imaging and other diagnostic procedures) [6]. Resting state fMRI data was recorded in the MRI Unit, University Hospital of Heraklion. The hospital review board approved the study and the procedure was thoroughly explained to all patients and volunteers, who signed informed consent before undergoing MRI.

Table 1. Demographic and clinical characteristics of NPSLE patients and controls.

	Healthy participants (n=31)	**NPSLE patients** (n=41)	**p-value**
Female	22 (71.0%)	38 (92.7%)	.02
Age, mean (SD in years)	45.8 (15.0)	45.0 (12.4)	.8
Disease duration, mean (SD in years)	--	5.9 (5.7)	--
SLEDAI, mean (SD)	--	5.1 (4.3)	--
SLICC/ACR score, mean (SD)	--	0.3 (0.5)	--

NPSLE, Neuropsychiatric SLE; SLEDAI, Systemic Lupus Erythematosus Disease Activity Index; SLICC, Systemic Lupus International Collaborating Clinics; ACR, American College of Rheumatology.

Brain MRI examinations were performed on a clinical, upgraded 1.5T Siemens Vision/Sonata scanner (Erlangen, Germany), Gradient strength: 45 mT/m, Gradient slew rate: 200 mT/m/ms and a standard four channel head array coil (minimum voxel dimensions: 70 μm X 70 μm X 300 μm). Resting-state functional MRI (rs-fMRI) was derived from a T2*-weighted, fat-saturated 2D-FID-EPI sequence with repetition time (TR) 3500 ms, echo time (TE) 50 ms, field of view (FOV) 192 x 192 x 108 (x, y, z), and acquisition voxel size 3 x 3 x 3 mm. Whole brain scans consist of 36 transverse slices with 3.0-mm slice thickness and no interslice gap. Each BOLD time series consists of 80 dynamic volumes. The fMRI images were smoothed, normalized and co-registered to the MNI space. This preprocessing was completed using spm (Statistical Parametric Mapping). The time series were detrended and zero-mean corrected using MATLAB prior to the calculations detailed below.

B. Connectivity Value Calculation

Three main brain networks, that have been often used in similar studies for classification of patients suffering from diseases with central nervous system manifestations [1] were selected namely, Default Mode Network (containing 7 ROI's), Frontoparietal Network (13 ROI's) and Sensorimotor Network (3 ROI's). Pairwise associations were computed for each pair of voxels (time series) within each network, using Pearson correlation coefficients. The individual voxel pairwise correlation values between ROI's were averaged in order to produce one correlation value for every pair of ROIs' within each network. This manner of calculating connectivity is similar to other node-based methods as the final matrix contains ROI-to-ROI connectivity values with the difference that the individual voxel time series for each ROI are not averaged in order to produce an average ROI time series. There are also similarities with various voxel-based methods but the final information is reduced. This procedure ensured a relatively small number of initial features. The final number of pairwise correlation values that were used for classification were 102.

C. Feature Selection Methods

The dimensionality of the data required considerable reduction in order for the classification algorithms to perform optimally [7]. The curse of dimensionality was an issue in our feature set as the original features were 102 with only 41(NPSLE) + 31(HC) subjects (samples). Optimally, the number of features used should be smaller than the number of subjects. This problem was also obvious in the preliminary testing conducted without the application of a selection method. The classification results using the entirety of the connectivity derived features were very poor.

The methods tested for supervised feature selection were: SelectFromModel, Recursive Feature Elimination (RFE) and Lasso using Python's package scikit-learn [8], and others. These were evaluated using different estimators for the extraction of feature importances. The best results were produced by RFE.

Recursive Feature Elimination (RFE) is a recursive feature selection method. An estimator such as linear kernel SVC capable of producing feature importance metrics is utilized. Starting with a large feature set, RFE "eliminates" the features with the lowest importance based on the feature importance metric computed on the train set. This is repeated recursively until the desired number of features is met. The "rate" of the reduction performed by RFE, representing the number of features removed/eliminated in each iteration is user defined and was set to 1 feature per

iteration. By doing this, the method, although more computationally demanding, is theoretically more "thorough" or exhaustive as opposed to removing multiple features.

Due to the supervised nature of the selection method used, the feature selection process was conducted on the train data only without using information from the test data in order to avoid overfitting the model. By performing selection in this manner, biasing the classification model is avoided. The reduced feature set derived from the implemented selection method was used as an input in the classification model.

D. Classification Methods

Several classification methods were evaluated for their binary classification robustness on this dataset. Some examples are Random Forest, k-nearest neighbor, Naive Bayes and SVM.

In the present study, SVM (SVC) was implemented using scikit-learn available in Python. The linear kernel was selected for the SVC classifier. A linear kernel was considered more desirable given that it provided satisfactory classification results as it is more generalizable due to its simplicity. All the parameters of the SVM model were left to default values for all the experiments discussed in this paper, including the SVC estimator used for classification as well as the SVC estimator used for feature selection with RFE. Specifically, as mentioned above, a linear kernel was utilized with the penalty parameter of the error term C equal to 1.

E. Classification Performance Metrics

Several indices were employed for the estimation of model classification performance. Apart from the accuracy score, which is calculated as the ratio of correctly classified samples by the total number of samples, showing the overall accuracy of the model, other metrics of classification performance were also used [9].

$$ACC = \frac{tp+tn}{tp+tn+fp+fn} \quad (1)$$

Precision (positive predictive value) is calculated as:

$$PPV = \frac{tp}{tp+fp} \quad (2)$$

Precision can be interpreted as the ability of the classifier not to label as positive a sample that is negative.

The sensitivity or Recall (true positive rate) of a test is its ability to determine the patient cases correctly. This is also obvious through its calculation:

$$TPR = \frac{tp}{tp+fn} \quad (3)$$

Specificity (true positive rate) represents the ability of a method to determine the negative (patient) cases correctly. In simpler terms, a higher value of specificity will allow for less false positives, or healthy subjects classified as patients.

$$TNR = \frac{tn}{tn+fp} \quad (4)$$

The F1 score metric can be thought of as a combination of Recall and Precision as it is calculated using these two metrics:

$$F1 = 2 \times \frac{PPV \times TPR}{PPV+TPR} = \frac{2tp}{tp+fp+fn} \quad (5)$$

**tp: true positive, tn: true negative, fp: false positive, fn: false negative*

As the harmonic average of precision and recall, F1 is often used in Machine Learning applications as it provides a combined metric. For these reasons the final model was selected with the F1 score in high regard.

Additionally, the AUC-ROC score metric was computed for the classification model. The ROC (Receiver operating characteristic) curve is created by plotting the true positive rate (TPR) or Recall against the false positive rate (FPR) or fall-out. The AUC (area under the curve) is equal to the probability that the classifier will "favor" a randomly chosen positive instance (patient) higher than a negative one (healthy). This metric was mostly used for model comparison and selection purposes.

F. Cross Validation Methodology

Cross validation (CV) is an effective estimation method of a model's prediction error. It is also a useful tool for model selection. The nested cross-validation method implemented was used for the evaluation of the model's performance in various stages as well as the final model selection.

A large number of Machine Learning studies in a variety of subjects including biomedical classification studies utilize a cross-validation method such as leave-one-out cross-validation (LOOCV) [3] or k-fold cross-validation [1]. In this case, feature selection, model selection, possibly, estimator hyper-parameter tuning and testing are all contained within the bounds of the iterations of a single CV method. This technique has a few downsides. Apart from often producing overly optimistic classification results, there also entails a risk for model overfitting. Additional to these issues, that can be avoided, the generalization of the final model can be slightly limited and the results falsely overoptimistic.

Nested cross-validation was preferred as opposed to a "simple" implementation of leave-one-out cross-validation (LOOCV) or k-fold cross-validation for a variety of reasons. Firstly, in cases such as ours where supervised feature selection methods are used for the reduction of the initial number of features, the selection of the final feature set is an important task. Of similarly high importance is the (unbiased/external) evaluation of the final model, with the final feature set on new test samples. In order to achieve this goal, a CV method such as k-fold is used internally on one portion of the data for model selection while the rest of the data is kept aside for the final model testing and evaluation. This is in turn repeated many times so as to eliminate random or cherry-picked train-test combinations.

The external data split is a simple 80/20 train-test split that randomly shuffles the data in each iteration. The internal CV (5-fold) is then performed on the larger portion (80%). In each iteration of the internal CV different combinations of the initial features are potentially selected by the feature selection method. This information is saved across each iteration in the form of indices/names of the selected features. Next, the feature "appearances" are counted and the features that appeared more across the

internal cross-validation iterations (subsets) are used in the final model.

In this manner the final features are selected by a combination of feature importance (due to better classification ability) and appearance in a larger number of the internal CV subsets. Finally, the model is trained with the entire 80% of the data using the aforementioned selected features and tested on the remaining 20% of the data that has been left untouched. The final prediction performance is calculated as the mean and standard deviation of the performance metrics of the 20% external test set, across several runs. The methodology analyzed above (Fig. 1) is visually depicted in the following diagram.

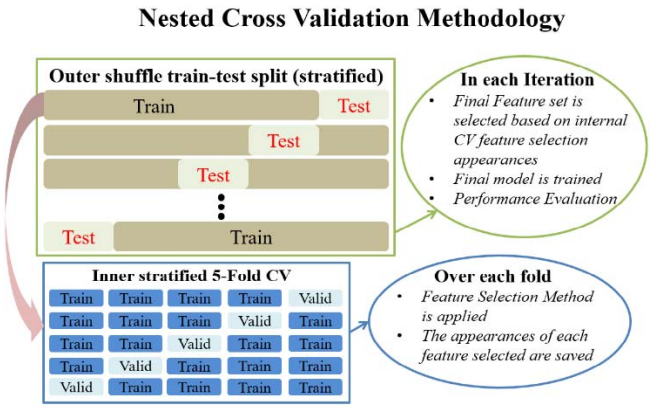

FIG. I. CROSS VALIDATION METHODOLOGY DIAGRAM

In this manner overfitting due to train-test set overlap, supervised selection on the entire data set, or model selection using the entire data set are avoided. Also, due to the reasons mentioned, creating an overoptimistic model is kept to a minimum.

III. RESULTS

The results presented are classification metrics computed using the Nested Cross Validation technique described above. Specifically, an external 80/20 train-test split applied to the entire sample set and an internal stratified 5-fold applied on the 80% of the split data. After extensive test runs, 1000 iterations were found to be adequate for the model's results to stabilize. The only variables changing across iterations of the nested cross-validation algorithm is the randomness of the external data split as well as the internal k-fold split. In each run a different subset of the entire sample set is chosen producing a slightly different result. The final results are the mean and standard deviation metrics across these iterations making them representative of the final model's classification ability.

The classifier used as specified in the methods section was SVM with a linear kernel. A hyperparameter search on the internal (nested) subset to avoid overfitting the model was evaluated. The 'optimal' SVM hyperparameters did not change the outcome of the model significantly so the parameters were left default.

The feature selection method used to select the features able to produce the greatest classification outcome was Recursive feature elimination (RFE) using a linear kernel SVC estimator. A few numbers of features to reduce (input of RFE) were tested with very similar results, 12 selected features were used as less features were preferred due to the relatively low sample size. Results were very similar, for selected feature numbers 9 through 28.

TABLE II. MEAN AND STANDARD DEVIATION OF CLASSIFICATION RESULTS

Metrics	Mean ± Standard Deviation
Accuracy (%)	75 ± 11
Precision (%)	82 ± 11
Sensitivity (%)	74 ± 15
Specificity (%)	75 ± 17
F1 Score (%)	77 ± 11
ROC-AUC (%)	75 ± 11

IV. DISCUSSION

A. Discussion of Results

The cross validated classification results obtained from the proposed method were quite conservative and not over-optimistic (Table II). This is in part due to the very strict cross validation methodology and the avoidance of model overfitting to the available data. The proposed classification and validation method attempt to ensure that the final results are indicative of real-world performance with new test subjects. The final model's classification metric scores indicate its ability to effectively distinguish between the two classes. Furthermore, the mean Recall (Sensitivity) score that is of highest importance in such clinical studies in conjunction with the F1 score (which contains Recall) were the two main metrics on which final model selection was based upon. A higher value of Recall (and by extension F1 score) lowers the probability of classifying a patient as healthy which is clearly not desirable.

The nested cross-validation method produced 12 final features after several selections in 1000 iterations. The pairwise ROI to ROI correlations corresponding to these features are given in Table III for each individual network.

TABLE III. ROI TO ROI FUNCTIONAL CONNECTIVITY CORRELATION NAMES, BY NETWORK

Connection 1	Right Inferior Temporal–Thalamus DMN
Connection 2	Left Angular Gyrus – Right Precuneus FPN
Connection 3	Left Primary Motor Cortex–SMA SMN
Connection 4	Left Primary Motor Cortex–Right Primary Motor Cortex SMN
Connection 5	Left Inferior Temporal–Thalamus DMN
Connection 6	Right Angular Gyrus–Right Inferior Parietal lobule FPN
Connection 7	Right Dorsolateral Prefrontal–Right Inferior Parietal lobule FPN
Connection 8	Left Angular Gyrus–Left Inferior Parietal lobule FPN
Connection 9	Left Thalamus–Right Thalamus FPN
Connection 10	Midcingulate Cortex–Right Angular Gyrus FPN
Connection 11	Left Inferior Parietal lobule–Right Thalamus FPN

Connection 12	Left Precuneus–Right Precuneus FPN

ROI correlations of the Frontoparietal Network comprised the majority of ROIs that contributed to the final features. Frontoparietal ROIs were also the most ROIs in number and as a result, most correlation values were derived from the Frontoparietal Network. Regarding Sensorimotor Network ROIs, only 3 ROIs were used from this brain network producing three correlation values. From these three functional connectivity correlations, two contributed as features through the final selection procedure, thus pointing out/denoting the importance of this network in this classification problem. This is also in line with a study for Multiple Sclerosis (MS) classification [1] sensorimotor-I was the highest contributing network.

The most prominent connections between ROIs of the 12 final features are depicted in Figure II. The most "important" of the 12 features were "chosen" by the selection algorithm in mostly all iterations of the external cross validation method.

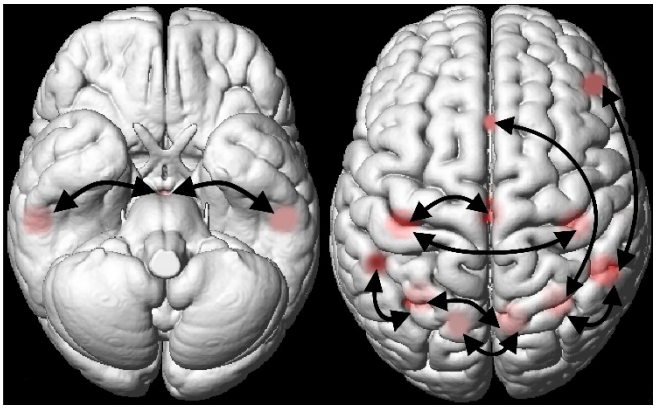

FIG. II. ANATOMIC LOCATIONS OF ROIS AND SELECTED FUNCTIONAL CONNECTIVITY INDICES ON A MODEL BRAIN.

One of the limitations of this study is the relatively small number of subjects (71 in total) leading to increased standard deviation in the performance metrics, as reported in Table II. Additionally, the nested cross validation methodology implemented would greatly benefit from a larger sample size possibly producing higher and even more reliable classification metric scores.

B. Conclusion and Future Work

The clinical question posed in this study was the distinction between NPSLE patients and age-matched healthy participants based exclusively on resting-state fMRI data. The results indicate the existence of biomarkers in the Networks studied, able to distinguish NPSLE patients from healthy subjects with considerable accuracy. In conclusion, the proposed method could potentially provide physicians with useful insight into the diagnosis of NPSLE in the future. As mentioned above, there is considerable research similar to the work presented here concerning diseases such as Multiple Sclerosis [10]. The findings presented here are of importance as there is effectively no research examining the ability of resting-state fMRI functional connectivity to differentiate NPSLE patients from healthy subjects. Additionally, these findings demonstrate the importance of resting-state fMRI in conjunction with machine learning algorithms and especially SVM, as a method for disease classification.

As a continuation of the work presented here, where ML algorithms were tested for their robustness on this classification problem, we would also aim to test the effectiveness of Neural Network techniques on this data set. Artificial Neural Networks are often are used in studies with several thousand samples with very high classification accuracies. The relatively small sample size is the most significant reason that deterred us from the evaluation of Neural Network techniques in the present study. However, we plan to exploit other alternatives such as transfer learning adaptations as well other deep learning methodologies designed for smaller datasets.

Furthermore, the proposed method could be potentially effective in discriminating between NPSLE patients with symptoms attributed to underlying SLE (primary NPSLE), from those with symptoms secondary to treatment, other systemic causes or concomitant psychiatric disorders (secondary NPSLE), presenting a significant diagnostic problem in clinical practice. With currently used methods this is often quite challenging. Healthcare professionals and patients could benefit from an improved pipeline for diagnosis. Provided such data is available this would be a possible follow up study. Finally, the incorporation of additional features derived from fMRI and possibly other imaging modalities as well as other clinical data for each patient could potentially increase the classification strength of the proposed method.

ACKNOWLEDGMENT

We would like to thank Professor Michalis Zervakis (Technical University of Crete) for his guidance and helpful insights on the technical aspects of this paper. Also, Dr. Vassilis Tsiaras for his contribution in the earlier stages of this study.

REFERENCES

[1] V. Saccà, A. Sarica, F. Novellino, S. Barone, T. Tallarico, E. Filippelli, A. Granata, C. Chiriaco, R. Bruno Bossio, P. Valentino, and A. Quattrone, "Evaluation of machine learning algorithms performance for the prediction of early multiple sclerosis from resting-state FMRI connectivity data," Brain Imaging Behav, vol. 13, no. 4, pp. 1103-1114, Aug, 2019. J. Clerk Maxwell, A Treatise on Electricity and Magnetism, 3rd ed., vol. 2. Oxford: Clarendon, 1892, pp.68–73.

[2] J. F. Vivaldo, J. C. de Amorim, P. R. Julio, R. J. de Oliveira, and S. Appenzeller, "Definition of NPSLE: Does the ACR Nomenclature Still Hold?," Front Med (Lausanne), vol. 5, pp. 138, 2018. K. Elissa, "Title of paper if known," unpublished.

[3] T. Price, C. Y. Wee, W. Gao, and D. Shen, "Multiple-network classification of childhood autism using functional connectivity dynamics," Med Image Comput Comput Assist Interv, vol. 17, no. Pt 3, pp. 177-84, 2014.

[4] C. Cortes, and V. Vapnik, "Support-Vector Networks," Machine Learning, vol. 20, no. 3, pp. 273-297, September 01, 1995. I. Guyon, Andr, #233, and Elisseeff, "An introduction to variable and feature selection," J. Mach. Learn. Res., vol. 3, pp. 1157-1182, 2003.

[5] A. Kazeminejad, and R. C. Sotero, "Topological Properties of Resting-State fMRI Functional Networks Improve Machine Learning-Based Autism Classification," Front Neurosci, vol. 12, pp. 1018, 2018.

[6] E. Papadaki, A. Fanouriakis, E. Kavroulakis, D. Karageorgou, P. Sidiropoulos, G. Bertsias, P. Simos, and D. T. Boumpas, "Neuropsychiatric lupus or not? Cerebral hypoperfusion by perfusion-weighted MRI in normal-appearing white matter in primary

neuropsychiatric lupus erythematosus," Ann Rheum Dis, vol. 77, pp. 441-448, 2018.

[7] I. Guyon, Andr, #233, and Elisseeff, "An introduction to variable and feature selection," J. Mach. Learn. Res., vol. 3, pp. 1157-1182, 2003.

[8] F. Pedregosa, G. Varoquaux, A. Gramfort, V. Michel, B. Thirion, O. Grisel, M. Blondel, P. Prettenhofer, R. Weiss, V. Dubourg, J. Vanderplas, A. Passos, D. Cournapeau, M. Brucher, M. Perrot and E. Duchesnay, "Scikit-learn: Machine Learning in Python," J. Mach. Learn. Res., vol. 12, pp. 2825-2830,2011.

[9] A. Baratloo, M. Hosseini, A. Negida, and G. El Ashal, "Part 1: Simple Definition and Calculation of Accuracy, Sensitivity and Specificity," Emerg (Tehran), vol. 3, no. 2, pp. 48-9, 2015.

[10] M. Zurita, C. Montalba, T. Labbé, J. P. Cruz, J. Dalboni da Rocha, C. Tejos, E. Ciampi, C. Cárcamo, R. Sitaram, and S. Uribe, "Characterization of relapsing-remitting multiple sclerosis patients using support vector machine classifications of functional and diffusion MRI data," Neuroimage Clin, vol. 20, pp. 724-730, 2018.