HAL open science

A machine learning approach to explore cognitive signatures in patients with temporo-mesial epilepsy

Elise Roger, L. Torlay, J. Gardette, C. Mosca, S. Banjac, L. Minotti, P. Kahane, M. Baciu

To cite this version:

Elise Roger, L. Torlay, J. Gardette, C. Mosca, S. Banjac, et al.. A machine learning approach to explore cognitive signatures in patients with temporo-mesial epilepsy. Neuropsychologia, 2020, 142, pp.107455. 10.1016/j.neuropsychologia.2020.107455 . hal-02895506

HAL Id: hal-02895506

https://hal.science/hal-02895506v1

Submitted on 16 Jul 2020

HAL is a multi-disciplinary open access archive for the deposit and dissemination of scientific research documents, whether they are published or not. The documents may come from teaching and research institutions in France or abroad, or from public or private research centers.

L'archive ouverte pluridisciplinaire **HAL**, est destinée au dépôt et à la diffusion de documents scientifiques de niveau recherche, publiés ou non, émanant des établissements d'enseignement et de recherche français ou étrangers, des laboratoires publics ou privés.

Distributed under a Creative Commons Attribution 4.0 International License

Neuropsychologia 142 (2020) 107455

Contents lists available at ScienceDirect

Neuropsychologia

journal homepage: http://www.elsevier.com/locate/neuropsychologia

A machine learning approach to explore cognitive signatures in patients with temporo-mesial epilepsy

E. Roger [a], L. Torlay [a], J. Gardette [a], C. Mosca [b], S. Banjac [a], L. Minotti [b], P. Kahane [b], M. Baciu [a,*]

[a] Univ. Grenoble Alpes, CNRS LPNC UMR 5105, 38000, Grenoble, France
[b] Univ. Grenoble Alpes, Grenoble Institute of Neuroscience 'Synchronisation et modulation des réseaux neuronaux dans l'épilepsie' & Neurology Department, 38000, Grenoble, France

ARTICLE INFO

Keywords:
Neuropsychology
Language
Memory
Epilepsy
SVM
XGBoost
Partial dependence plot

ABSTRACT

We aimed to identify cognitive signatures (phenotypes) of patients suffering from mesial temporal lobe epilepsy (mTLE) with respect to their epilepsy lateralization (left or right), through the use of SVM (Support Vector Machine) and XGBoost (eXtreme Gradient Boosting) machine learning (ML) algorithms. Specifically, we explored the ability of the two algorithms to identify the most significant scores (features, in ML terms) that segregate the left from the right mTLE patients. We had two versions of our dataset which consisted of neuropsychological test scores: a "reduced and working" version (n = 46 patients) without any missing data, and another one "original" (n = 57) with missing data but useful for testing the robustness of results obtained with the working dataset. The emphasis was placed on a precautionary machine learning (ML) approach for classification, with reproducible and generalizable results. The effects of several clinical medical variables were also studied. We obtained excellent predictive classification performances (>75%) of left and right mTLE with both versions of the dataset. The most segregating features were four language and memory tests, with a remarkable stability close to 100%. Thus, these cognitive tests appear to be highly relevant for neuropsychological assessment of patients. Moreover, clinical variables such as structural asymmetry between hippocampal gyri, the age of patients and the number of anti-epileptic drugs, influenced the cognitive phenotype. This exploratory study represents an in-depth analysis of cognitive scores and allows observing interesting interactions between language and memory performance. We discuss implications of these findings in terms of clinical and theoretical applications and perspectives in the field of neuropsychology.

1. Introduction

Cognitive impairment recently became an integral part of the definition and classification of epilepsies adopted by the International League Against Epilepsy (ILAE; Fisher et al., 2005). Cognitive deficits are common in epilepsy (up to 70% of patients) and reported very early in some cases (i.e. in newly diagnosed epilepsy), even before the introduction of the antiepileptic therapy (Witt and Helmstaedter, 2012, 2015). However, it is challenging to assume causal relationship between seizures and cognitive phenotypes. All the specific characteristics related to the causes and/or the consequences of the epileptic pathology could indeed be at the origin of the cognitive difficulties (i.e. lesions, atypical configuration of brain networks, abnormal inter-ictal activity, psychiatric co-morbities; Dinkelacker et al., 2016). In terms of cognitive symptoms, the focal subtypes of epilepsies are more frequently associated with specific and restricted cognitive deficits than the generalized forms (Brissart and Maillard, 2018) and the observed impairments are generally mild to moderate (for a review see Baciu and Perrone-Bertolotti, 2015). This suggests a continuous cerebral reorganization through time, depending on neuroplasticity phenomena taking over the impaired cognitive function(s) (i.e. chronic plasticity; (Berg and Scheffer, 2011).

Temporal lobe epilepsy (TLE) accounts over three-quarters of focal epilepsy cases in adults (Jaimes-Bautista et al., 2015), probably because the temporal lobe is the highest epileptogenic region of the human brain (Ladino et al., 2014). The underlying dysfunction (epileptogenic zone, EZ) is frequently located in temporal mesial structures (mTLE; Burianová et al., 2017). In terms of putative cause, hippocampal scleroses (HS) are very commonly observed in association with epilepsy (about 80% of cases; Tatum, 2012). In addition, this is the form of focal epilepsy

* Corresponding author. LPNC UMR CNRS 5105, Université Grenoble Alpes, BP 47 38040, Grenoble Cedex 09, France.
E-mail address: monica.baciu@univ-grenoble-alpes.fr (M. Baciu).

https://doi.org/10.1016/j.neuropsychologia.2020.107455
Received 15 November 2019; Received in revised form 25 February 2020; Accepted 27 March 2020
Available online 6 April 2020
0028-3932/© 2020 The Authors. Published by Elsevier Ltd. This is an open access article under the CC BY license (http://creativecommons.org/licenses/by/4.0/).

presenting the highest rate of drug-resistance and recurrent seizures, in this case, continue to induce brain damage.

Deficits of language and memory are frequently reported in the literature on TLE (Alessio et al., 2013; Jaimes-Bautista et al., 2015; McAndrews and Cohn, 2012; Metternich et al., 2014). In terms of neuroanatomical basis, temporal regions support language and memory networks, which could explain why these cognitive functions are more likely to be impaired in TLE (Mayeux et al., 1980). Nevertheless, the temporal lobe is also involved in other cognitive processes, liable to induce a variety of cognitive deficits (executive functioning, social cognition or even face recognition; Bora and Meletti, 2016; Lomlomdjian et al., 2017). One explanation may be found in the recent proposition of Genon et al. (2018) that have introduced an interesting conceptualization of the functional hippocampal specialization as a polyhedron, with as many facets as the various functions in which the hippocampus may be involved. Taking this stand, the neuropsychological deficits associated with mTLE are potentially multiple.

Next to their multiple and potentially disabling nature, cognitive impairments can worsen over time, sometimes aggravated by the anti-epileptic drugs, and they can have a negative impact on the quality of life (Witt et al., 2013). It has indeed been shown that cognitive deficits alter the quality of life in a similar way to other factors such as the frequency and severity of seizures, psychiatric co-morbidities, adverse drug reactions, or even as more social factors, namely the professional exclusion (Taylor et al., 2011). Therefore, it seems essential to know how to diagnose and identify cognitive profiles to monitor and propose a remediation if needed. Interestingly, Witt & Helmstaedter (2012) have shown that in TLE patients the objective language and memory deficits (48%) assessed during the neuropsychological evaluation (NPE) were more frequently observed than subjective complaints (25%). A similar pattern was observed for the executive functions (EF). The authors emphasize the underestimation of cognitive deficits when based only on subjective complaints and advocate thus the need for objective screening (Brissart and Maillard, 2018). A main issue in this context is the high variability of tools used for the NPE. Indeed, Vogt et al. (2017) highlighted a significant variability, reporting 186 different tests identified in 26 European Hospital Centers. Moreover, not a lot of information were provided by the clinicians on the validity and sensitivity of tools used to diagnose cognitive impairments in epilepsy (Vogt et al., 2017).

Based on these observations, the practical goal of our study is to estimate the psychometric properties of the main tests used in the traditional NPE of epileptic patients with mesio-temporal epilepsy. To this end, we have used a machine learning approach, a powerful tool to assess the sensitivity, the reliability and the predictive validity of the NPE. Machine learning (ML) refers to computational sophisticated algorithms used to emulate human intelligence and decision-making by learning from the environment (El Naqa and Murphy, 2015). This approach has been increasingly used in the past few years in the field of neuroscience and cognition. Significant amount of research highlighted the efficiency of ML for differential diagnosis of patient population (Salvatore et al., 2014), or even prediction of drug treatment consequences (Chekroud et al., 2016; and Munsell et al., 2015 in the case of epilepsy). Several studies also focused on the identification of different cognitive subtypes, especially in the case of schizophrenia (Gould et al., 2014). However, to our knowledge, very few studies used the ML approach to evaluate the effectiveness of NPE in pre-surgical evaluation of mTLE patients, focusing on an in-depth study of their cognitive phenotypes.

Concretely, we applied a supervised ML using both SVM (Support Vector Machine) and XGBoost (eXtreme Gradient Boosting). In the first step, we applied a binary classification and a feature selection. This allows characterizing specific cognitive signatures of mTLE patients. As the epilepsy lateralization has a major impact on the cerebral reorganization (Besson et al., 2014), mTLE patients should be considered as two separate groups according to the lateralization of epilepsy (left: L-mTLE and right: R-mTLE). We have therefore carried out classifications of these two groups of patients. Results allowed estimating NPE efficiency to classify patients (*binary classification performance*) and finding the most relevant cognitive scores (*feature selection*) for the classification. In the second step, we applied partial dependence analyses (PDP) to determine the predictive profile of cognitive scores and their interactions in the classification (*model interpretation*). Overall, this ML procedure provides, as a proof of concept, extensive and comprehensive identification and examination of cognitive profiles in mTLE patients.

2. Materials

2.1. Patients

Fifty-seven drug-resistant patients with unilateral mTLE according to the ILAE committee report (Wieser for the ILAE Commission on Neurosurgery of Epilepsy, 2004) have been included. All patients underwent pre-surgical examination including clinical (neurological), neuropsychological and speech assessment, as well as electrical (video-EEG recordings) and brain structure (MRI) evaluations. Pre-surgical evaluation allowed identifying the hemispheric and regional localization of the epileptogenic zone (EZ). According to it, patients were separated in two groups, left (L-mTLE) and right (R-mTLE). Only 46 patients had no missing NPE data, 27 L-mTLE and 19 R-mTLE patients. Thus, we used two versions of our dataset: a "reduced and working" version (D': n = 46 patients) without any missing data, and another one "original" (D: n = 57, 24% more patients than in D'; 5% of missing values). The rationale for using the "original" dataset with the missing data was to test the robustness of results.

2.2. Neuropsychological assessment

All the NPE were carried out by a neuropsychologist and a speech therapist from the Epilepsy Unit of the Neurology Department. It consisted of the evaluation of several cognitive domains assessed with standardized tests: (a) general cognitive level (IQ) assessment composed of verbal comprehension index (VCI) and perceptual reasoning index (PRI) (WAIS IV, Wechsler, 2011); (b) language assessment composed of naming (DO80 test, French equivalent of the Boston Naming Task; Deloche and Hannequin, 1997) and verbal fluency (phonemic and semantic fluency; Godefroy & GREFEX, 2008); (c) memory assessment composed of auditory memory index (AMI) and visual memory index (VMI) of the Wechsler Memory Scale (WMS IV; Wechsler, 2012); (d) assessment of executive functions including processing speed and mental flexibility (TMT: Trail Making Test B-A; Godefroy & GREFEX, 2008), as well as mental inhibition of irrelevant responses (Stroop; Stroop, 1935). For all the tests mentioned above, the raw scores were standardized according to the patient's age. Except for the indexes from the WAIS-V (VCI and PRI), the raw performances were also corrected with respect to gender and sociocultural level. The standardization and normalization of scores was performed by the neuropsychologist with respect to the norms provided in respective manuals (Appendix S1). These scores were then expressed in terms of standard deviation from the norm (z scores). In total, 9 cognitive tests (IQ, VCI, PRI, DO80, verbal fluency, AMI, VMI, TMT and Stroop) have been used as features, to perform ML analyses. All NPE information is provided for each patient in the supplementary material (Table S1; Appendix S1).

3. Methods

3.1. Machine learning approach

The first objective of this exploratory study was to assess the ability of NPE to predict the lateralization of epileptogenic zone in our population of patients (i.e. in other words, to classify categories of patients).

The second objective was to identify the most discriminating scores for classification and determine their interactions. To this end, we performed several ML workflows including binary classification and feature selection. In practical terms, two parallel analyses have been conducted on the two versions of the dataset (i.e., D and D') using two different algorithms: (a) a classical Support Vector Machine (SVM) algorithm (Cortes and Vapnik, 1995) with a Radial Basis Function (RBF); and (b) a state-of-the-art XGBoost algorithm (Chen and Guestrin, 2016), previously and successfully used by our team (Torlay et al., 2017). We decided to use two different algorithms in particular to deal with missing values. A possible solution for dealing with missing values would have been to use an imputation method resulting in the computation of artificial data. However, the use of imputed data remains a matter of debate (Jakobsen et al., 2017). Therefore, we opted for XGBoost, an algorithm that deals with missing values. Both SVM RBF and XGBoost were applied on the reduced and working dataset (D′), but only XGBoost was applied to analyze the original version of the dataset (D) that includes 5% of missing values. Our multi-algorithm approach also allowed to: (i) see if the results depend on the type of algorithms used; and (ii) perform supplementary analyses useful to verify the robustness of the results obtained in a different version of the dataset (D, including 24% more patients).

Finally, to get an insight into the relationships between neuropsychological scores (features), we used Partial Dependence Plots (PDP). Indeed, ML algorithms have often been criticized to be black boxes compared to more simple and directly interpretable modeling approaches as a linear regression. Namely, PDP can show a marginal effect that one or two features can have on the predicted outcome of a machine-learning model (Friedman, 2001). Taking that into account, PDP could be useful supplement to our analyses by providing an interpretation of our models. More concretely, a partial dependence plot can show whether the relationship between the target and a feature is linear, monotonic or more complex (Molnar, 2018) allowing to draw conclusions according to the observed pattern. Fig. 1 represents a global overview of our procedure.

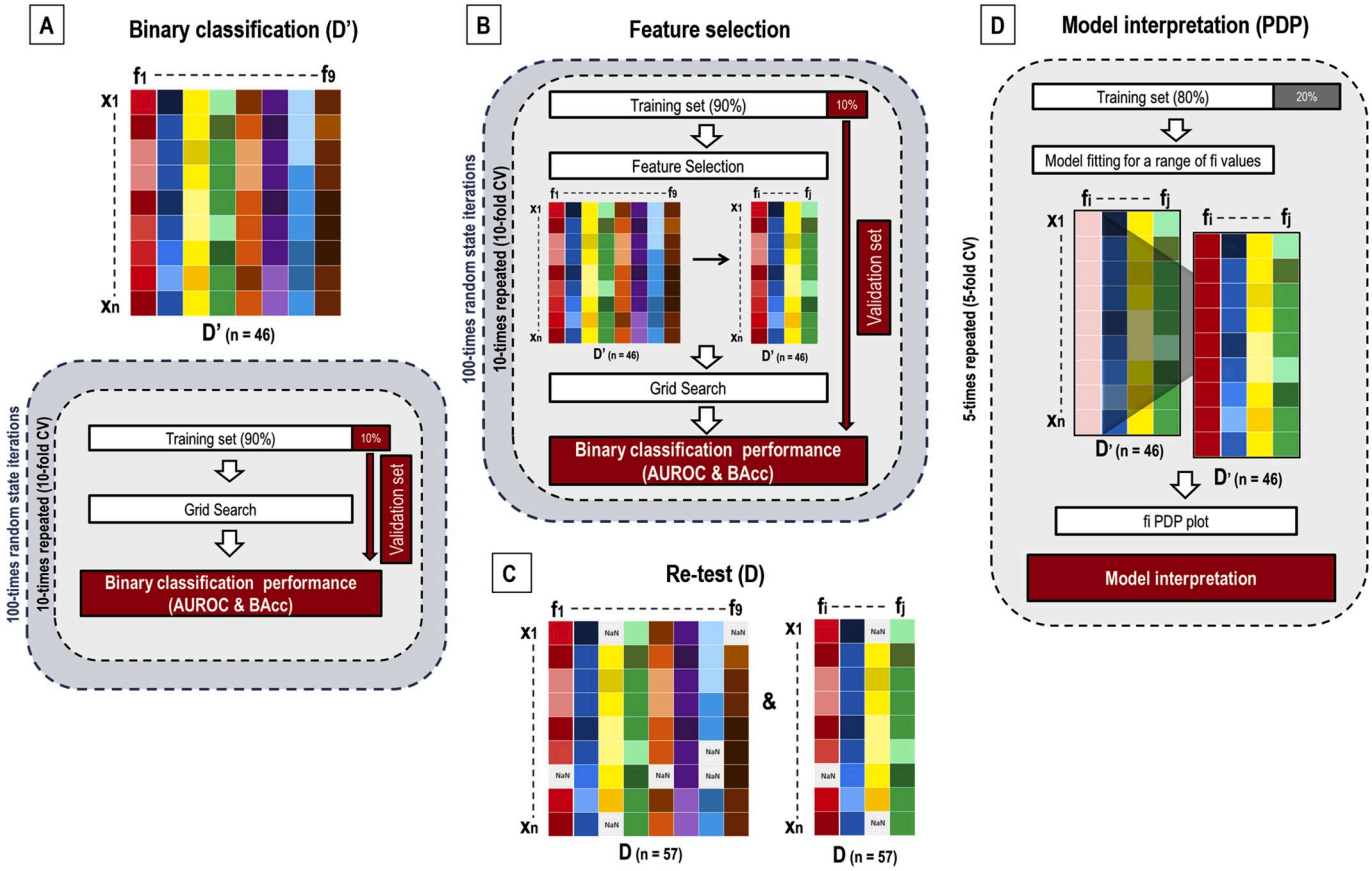


Fig. 1. General overview of the different steps of ML analyses. A. Workflow of the supervised binary classification applied between L-mTLE and R-mTLE. The classification was made using XGBoost and SVM on the dataset D′ composed of the 9 features of interest. The estimated performance (AUROC, BAcc) indicates the importance of these 9 features on the prediction. B. Workflow of the feature selection step. L2 logistic-regressions were used to select the most contributed features at each iteration. In this way we identify the most contributing and stable features of the classification but also if they are sufficient to separate on average our two mTLE populations. The selection stability was assessed using two metrics (frequency and phi: Φ) computed on the 1000 iterations, giving an idea of the robustness of the results obtained by the feature selection. Once the feature selected we have redone a binary classification on the reduced and working version of the dataset (D′) restricted to the selected features. C. Re-test step of the ML results obtained with D'. We used here the original version of the dataset D (24% more patients but 5% of missing values) to assess the classification performance of the 9 features of interest as well as on the restricted dataset composed of the same features selected in the previous step of feature selection (step B). Since this dataset is composed of missing values, only XGBoost was used to compute the classification performances. This step – although made on a not completely independent dataset – allows estimating the robustness and the generalization of the results obtained in the two first steps (A and B). D. Workflow of the partial dependence plot and model interpretation. The goal of the PDP is to observe how the selected features values change the prediction or interact with each other. The PDP was made on the dataset D′, restricted to the selected feature. The values of a chosen feature F are changed in small steps between the min and max values of F; whereas all the other features are kept unchanged. In this way we could estimate the effect of F on the classification (PDP 1 dimension). The same procedure could be done on 2 features in order to estimate the interaction between these two feature, on the prediction (PDP 2 dimensions). This method is efficient to go into the details of the classification and make interpretations about the model, as find some cut-offs above or below which the prediction increases in accuracy, for instance.

3.2. Binary classification

Many assumptions used in a learning algorithm (such as the Radial Basis Function kernel of Support Vector Machines (SVM-RBF) or the L2 regularizers of linear models (used in our feature selection workflow) assume that features are standardized. If not, the estimator may be unable to learn correctly. The reduced and working version of the dataset (D′) was hence standardized. Since XGBoost is not based on these assumptions, we did not preprocess the original version of the dataset (D). The goal was to predict the lateralization of mTLE patients (left or right) based on our neuropsychological measures of interest. In other words, we aimed to train a model (supervised learning) to assign correctly a patient to one of two classes, left or right (binary classification) based on a series of features. The algorithm uses the labeled data as the training set and its prediction performance is subsequently measured by using the unlabeled data as the validation set. A special attention has been paid on the generalization ability of the machine learning workflow. We used a classical 10-fold cross-validation (CV) scheme repeated 100 times and an inner CV in each training fold to do a grid search for hyper-parameters. Those CV were stratified, i.e., samples were randomly chosen in order to get always the same ratio of left and right epileptic patients in folds. To quantify the quality of predictions, we chose two widely used performance measurements: the Area Under the Curve of the Receiving Operator Curve (AUROC) and the balanced accuracy (BAcc). The AUROC of a classifier is equivalent to the probability that it will rank a randomly chosen positive instance higher than a randomly chosen negative instance. An AUROC score of a perfect model is of 100%, while a random classification score is of 50%. The BAcc is defined as the average recall obtained on each class (mean of the true positive and negative rates) in order to deal with imbalanced datasets. The error rate can be directly appreciating as: 1 - BAcc. Fig. 1 schematizes the binary classification procedure we used.

3.3. Feature selection and stability

While using a ML approach, we may mix relevant and irrelevant features to approximate the function between the input and the output, here, the lateralization of epilepsy. A feature selection step could help reducing the model to only relevant neuropsychological scores. To this end we used penalized linear models, often used to get sparse solutions since they offer solutions with fewer non-zero coefficients. We tried both L1 and L2 penalty logistic regression and the L2-norm was the sparsest approach. More precisely, we used it in each training set among 10 × 100 with the default threshold implemented in the scikit-learn library (v. 0.21.2; Pedregosa et al., 2011), i.e. mean of the features importance. Selected features were then used to train the algorithm with a grid search before measuring the performance with the held-out fold. To sum up, the feature selection was repeated 1000 times to get a good estimate of stability and performance. For reproducibility reasons, the measurement of stability is very important. We computed two types of stability indicators: the selected features' frequency and the stability metric $\widehat{\Phi}$ introduced by Nogueira et al. (2018) allowing rigorous algorithms comparisons. The Fig. 1 panel B represents a schematic illustration of our feature selection approach.

3.4. Algorithms

XGBoost belongs to the well-known decision trees family. They are invariant under scaling and they are robust to the effects of outliers. XGBoost is an optimized distributed gradient boosting library designed to be highly efficient. It provides a parallel tree boosting in a fast and accurate way. Several hyper-parameters (learning rate, maximum depth, gamma, minimum child weight and colsample bytree) were not fixed at default values but were optimized in each training set by a classical exhaustive grid search. The other algorithm, the SVM, is a versatile algorithm that constructs a hyper-plane with the largest margin separating the samples of any class. This algorithm is less time-consuming than XGBoost but may be less-performing and is unable to deal with missing values. The two hyper-parameters of the algorithm (C and gamma) were also classically optimized by grid search. The code used for the ML analyses is provided in the supplementary material (Appendix S2).

3.5. Model interpretation (PDP analyses)

Model-agnostic methods are not specific to ML (Ribeiro et al., 2016) but allow the interpretation of any model. This type of method offers the possibility to explore how state-of-the-art algorithms work rather than to be limited to directly interpretable but less-performing models like regressions or simple decision trees. We have chosen to use Partial Dependence Plot (PDP) in order to explore the way in which the values of selected features change the prediction and how these features interact. In practice, the PDP builds the model by averaging features with the exception of a chosen feature F and measures changes in prediction for different values of F (Fig. 1, Panel D). Taking into account that the working version of the dataset (D') was restricted to the selected features, we randomly performed a 5-fold cross validation with a mean AUROC of 95%, by using the XGBoost algorithm. We obtained five illustrations based on each training fold. Each illustration contained four 1D (one-dimensional)-plots (one by feature) and six 2D (bi-dimensional) plots (one by pair). By using the cross validation, we limited the risk of over interpretations and provides a more reliable overview of the dataset structure.

3.6. Modulatory factors

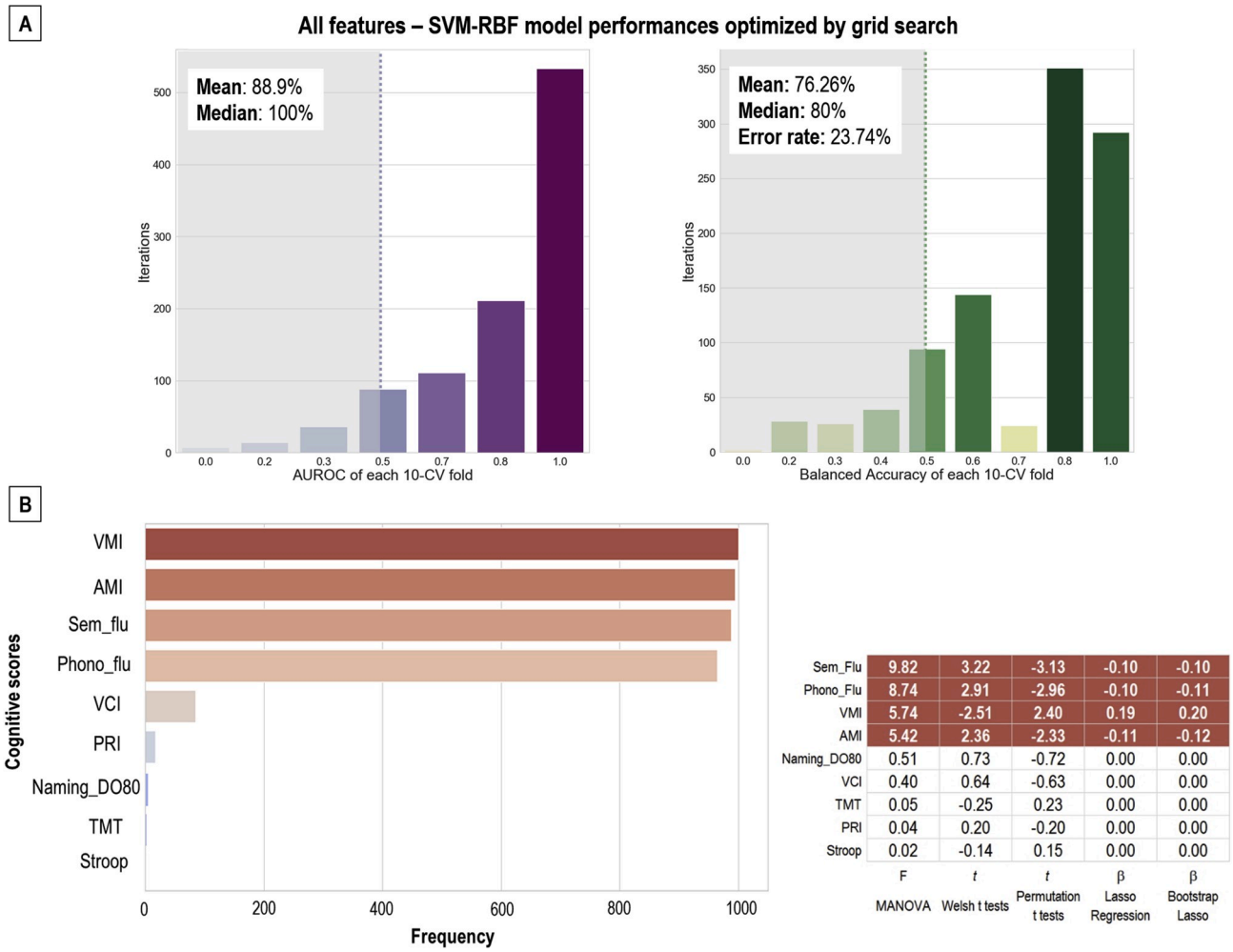
We further applied classical statistical analyses to assess the impact of clinical variables on the ML results obtained by the feature selection. Comparisons between groups of patients were assessed by the means of *t*-test. We also applied multiple regressions between each continuous clinical factor and selected cognitive features resulting from the feature selection step. All results were considered as significant at a threshold of $p < .05$.

4. Results

4.1. A. binary classification and feature selection

When using the reduced and working version of the dataset (D′) and all the nine neuropsychological features, we get an average $AUROC_{XGBoost} = 88.2\%$ and $AUROC_{SVM-RBF} = 88.9\%$ and an average $BAcc_{XGBoost} = 77.39\%$ and $BAcc_{SVM-RBF} = 76.26\%$ (cf. Fig. 2, Panel A for an example of the performances distributions obtained using SVM; Appendix S3 for all the distributions). This high level of performance clearly shows the ability of the complete NPE to predict epilepsy lateralization. The feature selection approach (cf. Fig. 2, Panel B) shows a remarkable quality of stability, $\widehat{\Phi} = 93.2\%$ with $k = 4$ features selected on average and a very good performance level ($AUROC_{XGBoost} = 89.7\%$ and $AUROC_{SVM-RBF} = 85.9\%$; $BAcc_{XGBoost} = 76.08\%$ and $BAcc_{SVM-RBF} = 76.41\%$). The four selected features were language and memory scores with respective frequencies of: VMI = 100%, AMI = 99.5%, Semantic Fluency = 98.8% and Phonological Fluency = 96.4%. We obtained an excellent level of prediction when we measured again the performance with the same 100-times repeated 10-CV on the 4 selected features, we get $AUROC_{XGBoost} = 90.2\%$ and $AUROC_{SVM-RBF} = 86\%$; $BAcc_{XGBoost} = 77.70\%$ and $BAcc_{SVM-RBF} = 77.68\%$.

In the re-test step, when using the original extended version of the dataset (D; 24% more patients), the performance remained very good despite the number of missing values: $AUROC_{XGBoost} = 82\%$ with all scores and 84% with the selected ones ($BAcc_{XGBoost} = 75.16\%$ and



	F MANOVA	t Welsh t tests	t Permutation t tests	β Lasso Regression	β Bootstrap Lasso
Sem_Flu	9.82	3.22	-3.13	-0.10	-0.10
Phono_Flu	8.74	2.91	-2.96	-0.10	-0.11
VMI	5.74	-2.51	2.40	0.19	0.20
AMI	5.42	2.36	-2.33	-0.11	-0.12
Naming_DO80	0.51	0.73	-0.72	0.00	0.00
VCI	0.40	0.64	-0.63	0.00	0.00
TMT	0.05	-0.25	0.23	0.00	0.00
PRI	0.04	0.20	-0.20	0.00	0.00
Stroop	0.02	-0.14	0.15	0.00	0.00

Fig. 2. Illustrations of the classification performances and the result of feature selection. A. Barplot illustrating the distribution of performance rates (AUROC and BAcc) obtained with SVM-RBF, for each iteration. B. At left: barplot of the frequency at which the 9 features of interest are selected from the feature selection step and through the 1000 iterations (at left). Four features were almost always selected: VMI, AMI, Sem_flu and Phono_flu; and have therefore a strong impact on the classification between mTLE patients. In contrast, the other features were almost never (or never) selected. At right: summary table of the comparisons with the results obtained using "traditional" statistical analyses. Significant results at $p < .05$ are highlighted in red. Overall, the results are similar to those from ML. **Note**: the assumption of normality has not been fully respected, which constitutes a limitation on the use of "classical" statistical analyses (see Appendix S3 for the complete statistical tables). (For interpretation of the references to colour in this figure legend, the reader is referred to the Web version of this article.)

$BAcc_{XGBoost} = 74.04\%$, respectively).

4.1.1. PDP and model interpretation

Regarding the PDP approach, we noted a clear threshold effect for AMI, VMI and Semantic fluency 1D plots (Fig. 3, Panel A). We thus examined interactions on 2D plots including these three features (Fig. 3, Panel B). Figs. S1–2 in the supplementary material show the different PDP obtained in the 5-fold cross-validation and for all 1D and 2D PDP.

In terms of model interpretation, the typical cognitive pattern of L-mTLE is represented by a poor auditory memory index (AMI; lower than norm −0.5 SD), and poor scores of semantic fluency (lower than norm −1 SD) in combination with high visual memory index (VMI; greater than norm −0.5 SD) (Fig. 3, panel A). The typical profile of R-mTLE is represented by AMI and semantic fluency score greater than norm (−0.5 SD) associated with a VMI score lower than norm (−1 SD) (Fig. 3, panel B). In general, despite areas of uncertainty close to ± 0.5 SD (gray area), the profiles become clearer as we move away from this cut-off points. Cut-offs associated with certainty levels can thus be identified.

4.1.2. Modulations by clinical data

On average, patient groups were matched regarding their demographical and clinical data. They did not differ significantly in terms of age ($t(55) = -0.27$, $p = .8$); handedness ($t(55) = -0.51$, $p = .6$); educational level ($t(55) = -0.58$, $p = .6$); hippocampal asymmetry ($t(55) = -0.49$, $p = .6$); age of seizures onset ($t(55) = -0.44$, $p = .7$); duration of epilepsy ($t(55) = 0.13$, $p = .9$); seizure frequency ($t(55) = 0.59$, $p = .5$); and number of antiepileptic drugs ($t(55) = 0.89$, $p = .4$). Only two regressions between the clinical factors and the 4 selected scores from the feature selection step were significant at $p < .05$. We observed a weak but significant effect of the HS asymmetry on AMI ($F(52) = 12.45$, $p < .001$, $R^2 = 0.19$) and VMI ($F(52) = 11.95$, $p < .001$, $R^2 = 0.19$) scores. More precisely, the higher the asymmetry between the two hippocampal gyri, the lower the AMI and VMI scores. We also found a significant effect of the patients' age on Phono_flu ($F(52) = 9.82$, $p < .001$, $R^2 = 0.29$) and of the Nb_AEDS on the Phono_flu ($F(52) = 9.8$, $p < .001$, $R^2 = 0.28$) as well. Overall, the higher the patient's age and the number of AEDs, the lower the phonological fluency scores. We did not observe significant effect of the other variables (handedness, duration of epilepsy, frequency of seizures) on the 4 selected scores. Fig. 4 illustrates how clinical factors influence the cognitive scores.

5. Discussion

Neuropsychological evaluation (NPE) represents an essential tool in

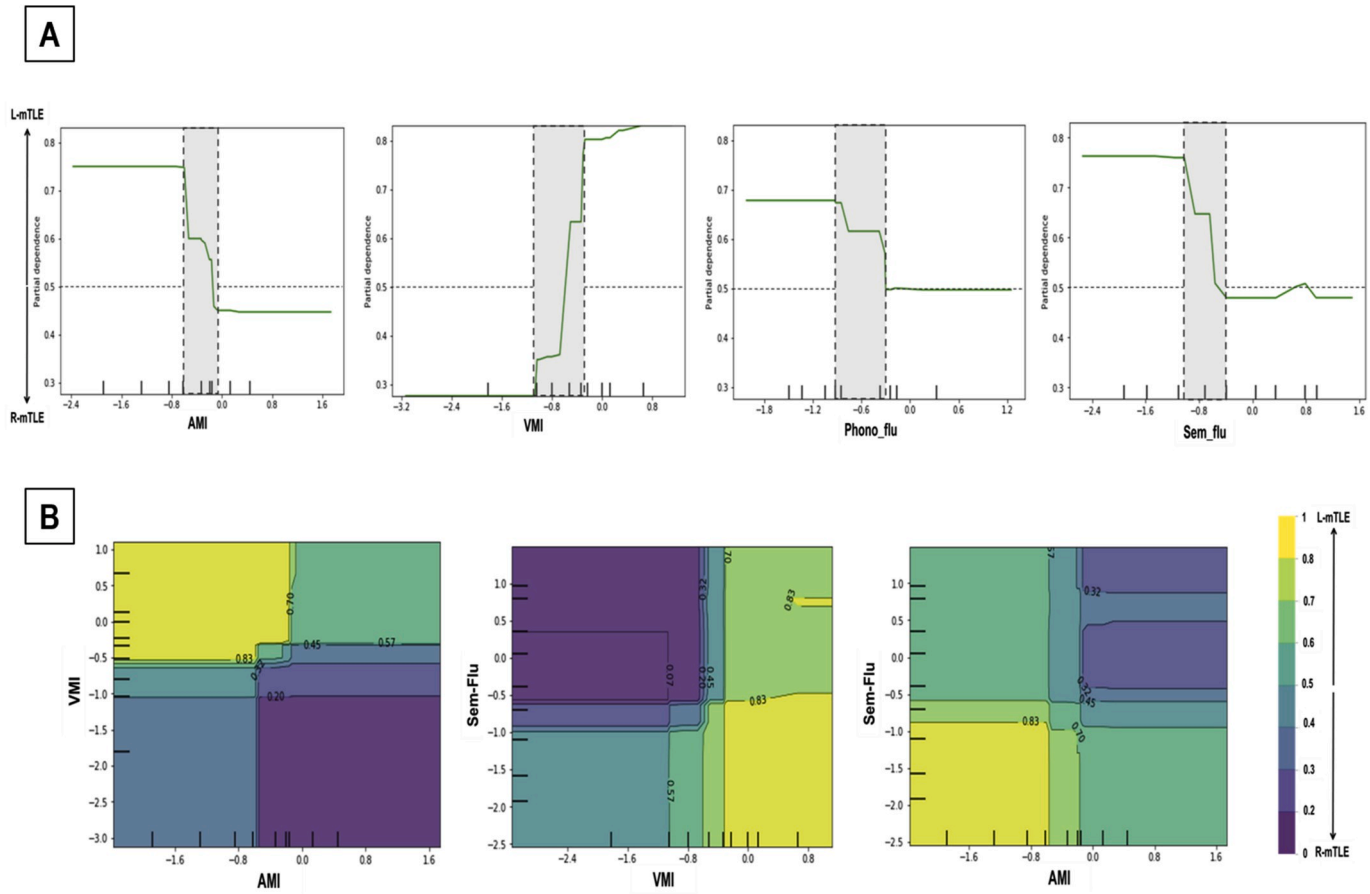


Fig. 3. The PDP observed for one CV-fold. **A.** Illustrates 1 dimension (1D) PDP. They show how the modifications of the values of each feature (taken separately whereas the others stay unchanged) affect the classification. The closer you get to 0 on the y-axis, the more likely you are to be an R-mTLE patient; the closer you are to 1, the more likely you are to be an L-mTLE patient. To take the example of a 'clear' feature, VMI, the more negative the values, the clearer the classification as R-mTLE becomes. Conversely, the more positive the values (above 0), the better the probability of being correctly classified L-mTLE when the patient is actually L-mTLE. The class jump for this feature is quite clear and the gray area of uncertainty is limited. To take the example of a slightly less clear-cut feature, Phono_flu, beyond the uncertainty zone the values are less clear-cut between 1 and 0 (y-axis). **B.** Illustrates 2 dimensions (2D) PDP. They show how the features interact and the influence of these interactions on the prediction. Different combinations of two features can be computed. The clearest combinations are represented here: VMI/AMI; Sem_Flu/VMI; and Sem_Flu/AMI. The more yellow the surface area, the more likely the patient is to be an L-mTLE; the more purple the area, the more likely the patient belongs to the R-mTLE group. The patterns are distinct and opposite. For example, for the VMI/AMI combination, L-mTLE patients tend to have poor AMI scores but good VMI scores. Regarding the R-mTLE, the pattern is reversed. See the supplementary materials (Fig. 1S and 2S) for the 5 CV-fold 1D and 2D PDP, giving a descriptive idea of the robustness of the results. (For interpretation of the references to colour in this figure legend, the reader is referred to the Web version of this article.)

the clinical care of epileptic patients. For instance, NPE can serve in practice for estimating epilepsy outcomes and the influence of pharmacological treatments on the behavior (Elger et al., 2004). In addition, when surgery is considered, the NPE could provide a valuable picture of the patients' cognitive landscape. As an indicator of the cognitive status before neurosurgical invasive procedures, the NPE supports the detection, location and lateralization of brain dysfunctions, helps the post-operative monitoring and guides cognitive remediation if needed. The NPE could also be an essential element of the pre-surgical planning and has been used for a long time to detect, locate and lateralize brain dysfunctions. Given its crucial clinical role, it is essential to conduct research aiming to provide indications that can assist and guide the neuropsychologist's practices.

In this perspective of evidence-based neuropsychology, the worth of the available neuropsychological instruments should be emphasized. Namely, the estimation of their validity, specificity and sensibility within the population of patients of interest is crucial. Previous research using traditional statistical procedures have already investigated the quality of the NPE in separating mTLE patients based on the presumed location of their epilepsies. However, the discriminatory power of the NPE in localizing and lateralizing the dysfunctional epileptogenic areas in the brain has not been clearly established. Some studies have indeed highlighted a limited role with a modest lateralization value (Dupont et al., 2002; Kim et al., 2004; Loring et al., 2008). Other studies, have demonstrated the utility of NPE to answer this specific question (e.g. Keary et al., 2007). Specifically, certain memory scores (WMS III auditory and visual memory index), language (Boston Naming Task and a reading task) and executive performance seemed to highly participate in the prediction of lateralization in TLE surgical candidates. Nevertheless, these studies are based on traditional statistical methods that allow testing only a few combinations of restricted models and are based on strong assumptions that limit their use. In addition, one of the major limitations of such studies is the generalizability of the results (Keary et al., 2007). ML approaches by contrast, estimate the predictive power and give an insight of the results' stability, allowing to address these issues. An interesting recent ML study conducted by Frank et al. (2018) had also provided some answers regarding the prediction of the seizure focus in TLE patients. By using different ML algorithms on a dataset restricted to language and memory tests (without IQ and executive functioning evaluations) they consistently found better-than-chance

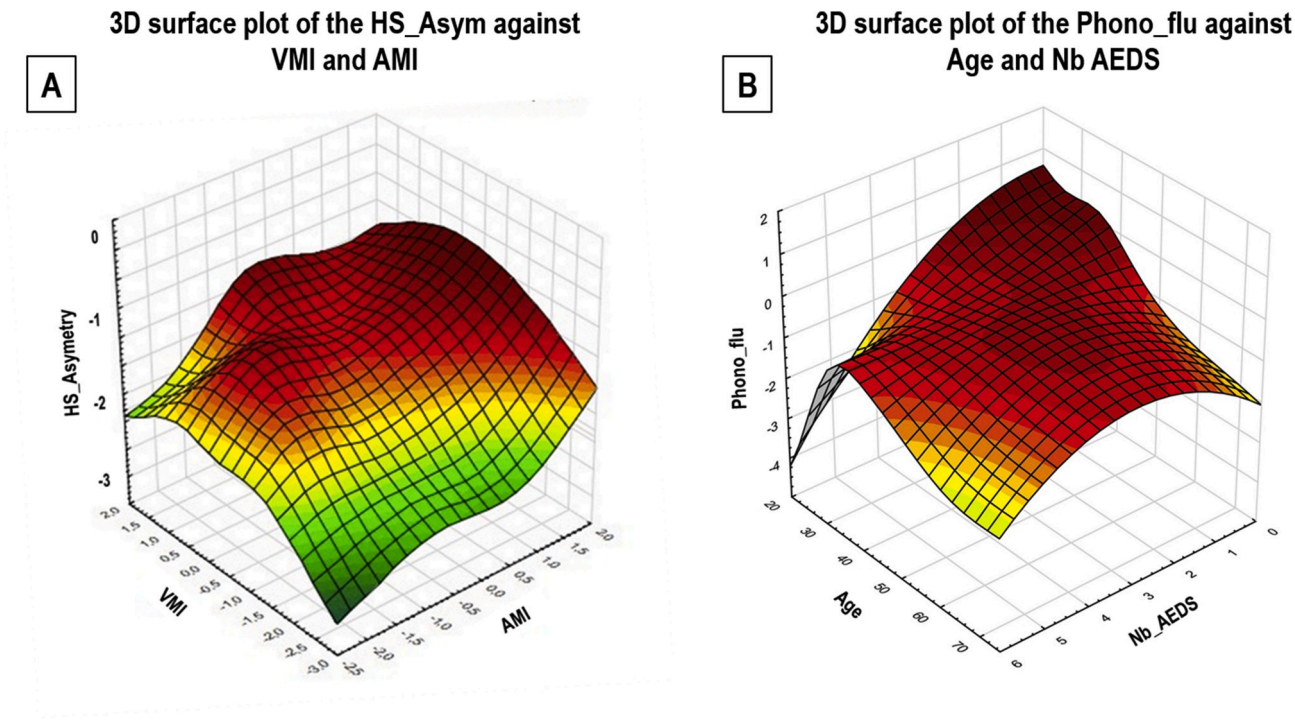


Fig. 4. 3D Surface plots of the modulatory effect observed between clinical variables and cognitive scores. **A.** We observed a significant effect of the hippocampal asymmetry (HS_asymetry) on the auditory memory index (AMI: $F(52) = 12.45$, $p < .001$, $R^2 = 0.19$) and on the visual memory index (VMI: $F(52) = 11.95$, $p < .001$, $R^2 = 0.19$) scores. On the z-axis is the z score of the asymmetry (the more negative the score and the greater the asymmetry between the two hippocampi). On the x-axis and y-axis are respectively the z scores of AMI and VMI. Overall, the higher the asymmetry between the two hippocampi, the lower the AMI and VMI scores. **B.** We found a significant effect of the patients' age on the phonological fluency (Phono_flu: $F(52) = 9.82$, $p < .001$, $R^2 = 0.29$) and of the number of antiepileptic drugs (Nb_AEDS) on the phonological fluency as well (Phono_flu: $F(52) = 9.8$, $p < .001$, $R^2 = 0.28$). On the z-axis the z scores of the phonological fluency test. On the x-axis is the number of antiepileptic drugs (taken daily) and on the y-axis is the age of patients (in number of years). Overall, the higher the patient's age and the number of AEDs, the lower the phonological fluency scores. Note: Below −1.5 standard deviations the cognitive scores can be considered pathological.

classification rates, suggesting a clinical utility of the NPE both to localize and lateralize the epilepsy (Frank et al., 2018). In the same way, the results of our exploratory ML analyses confirm and provide strong evidence about the validity and sensibility of the entire NPE to discriminate L-mTLE than R-mTLE patients, with a very good rate of performance (greater than 80% regardless of the algorithm used) and a very suitable stability.

Neuropsychological practices in epilepsy are variable and the selection of appropriate neuropsychological tests is difficult. The question of whether to approach the patient individually by eclectic test selection or through the use of a standard test battery represents a genuine problem (Helmstaedter and Witt, 2012). However, the identification of the most relevant cognitive tests for a given sub-population can lead to a tailored neuropsychological evaluation. For pragmatic reasons, the objective choice of these tests can indeed assist in determining which of them should be preferred to assess and to interpret in customized way cognitive profiles. The feature selection analysis we conducted for this purpose clearly supports here that among all the cognitive scores studied, language (phonological and semantic fluency) as well as memory performance (auditory memory index: AMI; and visual memory index: VMI) were the best predictors of the discrimination between L-mTLE and R-mTLE patients (see the feature selection results in Fig. 2).

Evidences from functional neuroimaging studies point in the same direction and highlight massive disruptions of the language and memory functioning in TLE (Dinkelacker et al., 2016; Pravatà et al., 2011; Roger et al., 2018 for a review). More precisely, previous fMRI studies have shown that L-mTLE patients are more likely to present an atypical brain organization of language (dominant in the right hemisphere or bilateral) than both patients with R-mTLE and healthy controls (Thivard et al., 2005). The incidence of atypical patterns is more than twice as much as in controls (4–6% versus 33% for L-mTLE; Adcock et al., 2003). In addition, some patients present subtler reorganizations within the dominant hemisphere for language. The latter tend to be underestimated given the level of precision required to estimate them accurately (Baciu and Perrone-Bertolotti, 2015). Other studies have demonstrated similar atypical patterns regarding the cerebral functioning of memory (Haag and Bonelli, 2013). These unusual patterns of brain processing previously identified in patients could constitute a possible origin of the behavioral disruption we observed for these cognitive functions in particular.

Next to the memory deficits usually reported in mTLE patients (Bell et al., 2011; Brissart and Maillard, 2018; Hoppe et al., 2007; Tramoni-Negre et al., 2017); one of the most frequently described language impairments of mTLE is the ability to name an object, including the well-known "tip of the tongue" phenomenon. Naming impairment would be more important in patients with L-mTLE than those with R-mTLE (Stemmer and Whitaker, 2008). Surprisingly, in our patient dataset, the naming score (DO80) was not considered as an important feature to differentiate groups of patients (feature selection results, Fig. 2). One possibility could be that naming deficits are underestimated or not systematically estimated in R-mTLE populations. The semantic network is indeed bilaterally represented and distributed across hemispheres (Cousin et al., 2006; Martin and Chao, 2001) and this vast network could therefore be similarly disrupted in R-mTLE cases. Some recent studies seem to show that naming deficits are only slightly more important in L-mTLE (between 40 and 55% of patients) than R-mTLE (36%; Bartha-Doering and Trinka, 2014) patients, which could be coherent with the hypothesis of an involvement of both hemispheres in semantic process. Another explanation could be that the DO80 naming task is not sufficiently sensitive and/or recently standardized to objectify clear differences between patients. These explanations do not necessarily compete with each other and are probably complementary.

Most of previous work focusing on prediction of epilepsy lateralization, including ML studies, has not specifically highlighted the interactions that may exist between the scores. In addition to the clinical interest of this approach for a better comprehension of the neuropsychological profile of the patient, there is also a fundamental interest in understanding cognitive functioning. By going further in the classification process, PDP analyses show how the different values that can be taken by a feature of interest (here the different possible z scores) affect and modulate the prediction. Technically, we can identify thresholds (cut-off points), beyond which the rates of prediction become sufficiently good and stable. In other words, the PDP analyses gave us an idea of the z scores above or below which we can classify patients with the greatest possible certainty (see Fig. 3, Panel A). PDP ML analysis also gives an idea of how features interact with each other, which is not entirely the case in traditional statistical analyses (Appendix S3). Using the PDP we observed very diverse combinations of features between the R-mTLE and L-mTLE patients, resulting in different interactive cognitive profiles (Fig. 3, Panel B). Contemporary cutting-edge studies go beyond the historically described modular framework of cognition and propose that there is in fact a vast "cognitive network" (Garcia-Ramos et al., 2016; Kellermann et al., 2016) with strong links between functions. Our PDP results described interactions between language and memory scores in the prediction of the hemispheric lateralization of seizures. The language and memory functions would indeed be strongly and directly interrelated to such an extent that Duff and Brown-Schmidt (2012) talk about a "language-and-memory interface". In terms of cerebral substrates and according to the same research team, the hippocampus would be the mediator of these language-and-memory interactions. Different pathways and paralleled distributed subsystems could interact closely. Namely, the phonological dorsal and the semantic ventral pathways, as well as a posterior parietal and hippocampus sub-circuit that is assumed to serve as a mediator between general language representations and other cognitive systems such as the episodic memory (Vandenberghe et al., 2013). The disruption of these communication streams, due to recurrent and refractory epileptic seizures originating in mesio-temporal structures and in the hippocampus in particular, could consequently have a double impact on both language and memory.

One limitation of our study is the size of our dataset, especially after sub-grouping, which could be problematic in ML. However, we have paid a special attention to the generalization of the results as well as to limiting the risk of overfitting. Namely, we used a multi-algorithm approach with a cross-validation and a re-test procedure performed in an extended version of the dataset. This "re-test" procedure on a second version of the dataset (D), including an additional quarter of patients, is not optimal since the sample used is not completely independent. However, as "*the use of different tests results in different outcomes, which cannot be directly compared*" (Helmstaedter and Witt, 2012), collecting a completely independent sample of patients presenting a one-sided diagnosis of mTLE and with the same cognitive evaluation would be a real challenge. We therefore proposed – for information purposes – this complementary and auxiliary analysis as an indicator of the robustness of the classification performances when adding additional patients.

As mentioned above, we paid a particular attention to the homogeneity of our sample. Some factors could influence the relations observed between the cognitive signatures and the location of the epilepsy. There were no significant differences between our two groups of patients on demographic data such as age, manual laterality or education level on average (Table S1). Similarly, the patients included in this study were clinically matched (no differences in clinical data such as duration of epilepsy, frequency of seizures, number of antiepileptic drugs, and hippocampal asymmetry). Nevertheless, these factors can have a transversal impact on cognition, independently of the patient groups. For example, TLE patients with hippocampal sclerosis (HS) have been found to have worse naming performance than those without HS and the volume of left hippocampus has been found to significantly predict verbal fluency and naming ability (Alessio et al., 2006). We have also found a significant and negative modulation of the hippocampal asymmetry on the cognitive data, but mainly on memory scores (AMI and VMI as well, Fig. 4). In addition, the number of antiepileptic drugs (AEDs) included in therapy as well as the patients' age have been reported to be significant predictors of language and executive functioning (Wang et al., 2011; see also Rudzinski and Meador, 2013). In line with these observations we have found a significant modulating effect of these factors on phonological fluency specifically (Fig. 4). The apparent susceptibility of frontal areas to the aging process (MacPherson et al., 2002) as well as to the influence of anti-epileptic drugs (Hamed, 2009) may explain phonological fluency difficulties probably resulting from executive functioning weaknesses. Other factors such as the severity of the disease including the duration (or the age of seizures onset) as well as the chronicity of the epilepsy (i.e. the seizures frequency) have previously been found to be predictors of poorer performance (Oyegbile et al., 2004; Rudzinski and Meador, 2013; Wang et al., 2011). However, none of these factors were significantly related to language and memory scores in our study. The neuropsychological tests included in this study are those that are typically used in the NPE of epileptic patients. However, as mentioned above for the DO80 naming test, some tests may not be sufficiently sensitive for some patients especially when the difficulties are not severe, which may cause some prediction errors. The use of newly developed standardized tests may be desirable. Finally, neuroimaging methods such as MRI volumetric study (Duchesne et al., 2006), resting state fMRI (Chiang et al., 2015) or combination of PET scan, structural MRI and DTI (Pustina et al., 2015) for example, could also help in predicting the lateralization of the seizure foci in TLE. It seems reasonable to believe that future studies will develop algorithms able to combine in an optimal way the results of all of these techniques (including the NPE) which could allow making highly reliable predictions, even on the most difficult cases.

6. Conclusion

The NPE is efficient to help clinicians in predicting the location/lateralization of the EZ. The cognitive tests used in this study are overall very sensitive and relevant in the discrimination of the two populations of mTLE patients. We observed different cognitive profiles according to the epilepsy location. Language (semantic and phonological fluency) and memory (WMS IV auditory and visual memory index) scores were the best predictors regardless the version of the patients' dataset used (i.e. the reduced or original version). Interestingly, some cut-off points have been identified beyond which the prediction increases with greater certainty. Finally, we found complex and interesting interactions between language and memory scores.

Some projects, as the European project E-PILEPSY, have the objective to harmonize practices and set standards for the NPE in epilepsy surgery (Vogt et al., 2017). The use of machine learning systems allows precisely a thorough study of the psychometric values of the cognitive tests classically used. In this perspective, our machine-learning experimentation-based study can provide direct guidance on relevant tests that should be used in the cognitive assessment of mTLE patients. Identifying the most relevant tests for the cognitive evaluation of mTLE patients provides the support for preoperative clinical practice (help in the hemispheric lateralization of the EZ, more tailored assessments for patients). This may help in reaching a decision if a neurosurgery is indeed necessary and possible. As "*the decision is more important than the incision*" (Senders et al., 2018), knowing the relevant indicators in the preoperative assessment is crucial to assist in neurosurgical decision-making process. ML analysis - associated with a cautious approach – therefore stands out as a privileged tool for the medicine of tomorrow as well. Moreover, ML analyses could provide details on the cognitive performance in the context of a pathological condition as epilepsy that, from a more theoretical point of view, allows moving beyond the modular vision of cognition towards a more interactive vision of an entangled cognitive functioning. In the light of that, the

perspectives and implications of the present study are multiple.

Funding

This work has been funded by the French program "AAP GENERIQUE 2017" run by the "Agence Nationale de la Recherche", grant "REORG" [grant number ANR-17-CE28-0015-01]; and by NeuroCoG IDEX UGA in the framework of the "Investissements d'avenir" program [grant number ANR-15-IDEX-02].

Ethical statement

Patients provided written informed consent to participate in the study, which was approved by the local ethics committee (CPP: 09-CHUG-14/ANSM (ID RCB) 2009-A00632-55).

Declaration of competing interest

The authors report no conflicts of interest.

CRediT authorship contribution statement

E. Roger: Writing - original draft, Writing - review & editing, Methodology, Formal analysis, Visualization, Validation. **L. Torlay:** Data curation, Methodology, Formal analysis, Visualization, Software, Validation. **J. Gardette:** Conceptualization, Investigation. **C. Mosca:** Investigation, Resources. **S. Banjac:** Writing - review & editing. **L. Minotti:** Resources. **P. Kahane:** Resources. **M. Baciu:** Conceptualization, Methodology, Writing - original draft, Writing - review & editing, Supervision, Project administration.

Appendix A. Supplementary data

Supplementary data to this article can be found online at https://doi.org/10.1016/j.neuropsychologia.2020.107455.

References

Adcock, J.E., Wise, R.G., Oxbury, J.M., Oxbury, S.M., Matthews, P.M., 2003. Quantitative fMRI assessment of the differences in lateralization of language-related brain activation in patients with temporal lobe epilepsy. Neuroimage 18 (2), 423–438. https://doi.org/10.1016/S1053-8119(02)00013-7.

Alessio, A., Bonilha, L., Rorden, C., Kobayashi, E., Min, L.L., Damasceno, B.P., Cendes, F., 2006. Memory and language impairments and their relationships to hippocampal and perirhinal cortex damage in patients with medial temporal lobe epilepsy. Epilepsy Behav. 8 (3), 593–600. https://doi.org/10.1016/j.yebeh.2006.01.007.

Alessio, A., Pereira, F.R., Sercheli, M.S., Rondina, J.M., Ozelo, H.B., Bilevicius, E., Pedro, T., Covolan, R.J., Damasceno, B.P., Cendes, F., 2013. Brain plasticity for verbal and visual memories in patients with mesial temporal lobe epilepsy and hippocampal sclerosis : an fMRI study. Hum. Brain Mapp. 34 (1), 186–199.

Baciu, M., Perrone-Bertolotti, M., 2015. What do patients with epilepsy tell us about language dynamics? A review of fMRI studies. Rev. Neurosci. 26 (3), 323–341. https://doi.org/10.1515/revneuro-2014-0074.

Bartha-Doering, L., Trinka, E., 2014. The interictal language profile in adult epilepsy. Epilepsia 55 (10), 1512–1525. https://doi.org/10.1111/epi.12743.

Bell, B., Lin, J.J., Seidenberg, M., Hermann, B., 2011. The neurobiology of cognitive disorders in temporal lobe epilepsy. Nat. Rev. Neurol. 7, 154.

Berg, A.T., Scheffer, I.E., 2011. New concepts in classification of the epilepsies : entering the 21st century. Epilepsia 52 (6), 1058–1062. https://doi.org/10.1111/j.1528-1167.2011.03101.x.

Besson, P., Dinkelacker, V., Valabregue, R., Thivard, L., Leclerc, X., Baulac, M., Sammler, D., Colliot, O., Lehéricy, S., Samson, S., Dupont, S., 2014. Structural connectivity differences in left and right temporal lobe epilepsy. Neuroimage 100 (Suppl. C), 135–144. https://doi.org/10.1016/j.neuroimage.2014.04.071.

Bora, E., Meletti, S., 2016. Social cognition in temporal lobe epilepsy : a systematic review and meta-analysis. Epilepsy Behav. 60, 50–57. https://doi.org/10.1016/j.yebeh.2016.04.024.

Brissart, H., Maillard, L., 2018. Neuropsychologie des épilepsies de l'adulte : Approche clinique et pratique. De Boeck Supérieur.

Burianová, H., Faizo, N.L., Gray, M., Hocking, J., Galloway, G., Reutens, D., 2017. Altered functional connectivity in mesial temporal lobe epilepsy. Epilepsy Res. 137, 45–52. https://doi.org/10.1016/j.eplepsyres.2017.09.001.

Chekroud, A.M., Zotti, R.J., Shehzad, Z., Gueorguieva, R., Johnson, M.K., Trivedi, M.H., Cannon, T.D., Krystal, J.H., Corlett, P.R., 2016. Cross-trial prediction of treatment outcome in depression : a machine learning approach. The Lancet Psychiatry 3 (3), 243–250. https://doi.org/10.1016/S2215-0366(15)00471-X.

Chen, T., Guestrin, C., 2016. XGBoost : a scalable tree boosting system. In: Proceedings of the 22Nd ACM SIGKDD International Conference on Knowledge Discovery and Data Mining, pp. 785–794. https://doi.org/10.1145/2939672.2939785.

Chiang, S., Levin, H.S., Haneef, Z., 2015. Computer-automated focus lateralization of temporal lobe epilepsy using fMRI. J. Magn. Reson. Imag. 41 (6), 1689–1694. https://doi.org/10.1002/jmri.24696.

Cortes, C., Vapnik, V., 1995. Support-vector networks. Mach. Learn. 20 (3), 273–297. https://doi.org/10.1007/BF00994018.

Cousin, E., Peyrin, C., Baciu, M., 2006. Hemispheric predominance assessment of phonology and semantics : a divided visual field experiment. Brain Cognit. 61 (3), 298–304. https://doi.org/10.1016/j.bandc.2006.02.005.

Deloche, G., Hannequin, D., 1997. *Test de dénomination orale d'images : DO 80*. Éd. du Centre de psychologie appliquée.

Dinkelacker, V., Dupont, S., Samson, S., 2016. The new approach to classification of focal epilepsies : epileptic discharge and disconnectivity in relation to cognition. Epilepsy Behav.: E&B 64 (Pt B), 322–328. https://doi.org/10.1016/j.yebeh.2016.08.028.

Duchesne, S., Bernasconi, N., Bernasconi, A., Collins, D.L., 2006. MR-based neurological disease classification methodology : application to lateralization of seizure focus in temporal lobe epilepsy. Neuroimage 29 (2), 557–566. https://doi.org/10.1016/j.neuroimage.2005.07.052.

Duff, M.C., Brown-Schmidt, S., 2012. The hippocampus and the flexible use and processing of language. Front. Hum. Neurosci. 6 https://doi.org/10.3389/fnhum.2012.00069.

Dupont, S., Samson, Y., Moortele, P.-F. V. de, Samson, S., Poline, J.-B., Hasboun, D., Bihan, D.L., Baulac, M., 2002. Bilateral hemispheric alteration of memory processes in right medial temporal lobe epilepsy. J. Neurol. Neurosurg. Psychiatr. 73 (5), 478–485. https://doi.org/10.1136/jnnp.73.5.478.

El Naqa, I., Murphy, M.J., 2015. What is machine learning? In: El Naqa, I., Li, R., Murphy, M.J. (Eds.), Machine Learning in Radiation Oncology : Theory and Applications. Springer International Publishing, pp. 3–11. https://doi.org/10.1007/978-3-319-18305-3_1.

Elger, C.E., Helmstaedter, C., Kurthen, M., 2004. Chronic epilepsy and cognition. Lancet Neurol. 3 (11), 663–672. https://doi.org/10.1016/S1474-4422(04)00906-8.

Fisher, R.S., van Emde Boas, W., Blume, W., Elger, C., Genton, P., Lee, P., Engel, J., 2005. Epileptic seizures and epilepsy : definitions proposed by the international League against epilepsy (ILAE) and the international bureau for epilepsy (IBE). Epilepsia 46 (4), 470–472. https://doi.org/10.1111/j.0013-9580.2005.66104.x.

Frank, B., Hurley, L., Scott, T.M., Olsen, P., Dugan, P., Barr, W.B., 2018. Machine learning as a new paradigm for characterizing localization and lateralization of neuropsychological test data in temporal lobe epilepsy. Epilepsy Behav. 86, 58–65. https://doi.org/10.1016/j.yebeh.2018.07.006.

Friedman, J.H., 2001. Greedy function approximation : a gradient boosting machine. Ann. Stat. 29 (5), 1189–1232. https://doi.org/10.1214/aos/1013203451.

Garcia-Ramos, C., Lin, J.J., Kellermann, T.S., Bonilha, L., Prabhakaran, V., Hermann, B. P., 2016. Graph theory and cognition : a complementary avenue for examining neuropsychological status in epilepsy. Epilepsy Behav. 64, 329–335. https://doi.org/10.1016/j.yebeh.2016.02.032.

Genon, S., Reid, A., Langner, R., Amunts, K., Eickhoff, S.B., 2018. How to characterize the function of a brain region. Trends Cognit. Sci. 22 (4), 350–364. https://doi.org/10.1016/j.tics.2018.01.010.

Godefroy, O., GREFEX, 2008. Fonctions exécutives et pathologies neurologiques et psychiatriques : Evaluation en pratique clinique. De Boeck Université.

Gould, I.C., Shepherd, A.M., Laurens, K.R., Cairns, M.J., Carr, V.J., Green, M.J., 2014. Multivariate neuroanatomical classification of cognitive subtypes in schizophrenia : a support vector machine learning approach. Neuroimage: Clinical 6, 229–236. https://doi.org/10.1016/j.nicl.2014.09.009.

Haag, A., Bonelli, S., 2013. Clinical application of language and memory fMRI in epilepsy. Epileptologie 30, 101–108.

Hamed, S.A., 2009. The aspects and mechanisms of cognitive alterations in Epilepsy : the role of antiepileptic medications. CNS Neurosci. Ther. 15 (2), 134–156. https://doi.org/10.1111/j.1755-5949.2008.00062.x.

Helmstaedter, C., Witt, J.-A., 2012. Clinical neuropsychology in epilepsy. In: Handbook of Clinical Neurology, vol. 107. Elsevier, pp. 437–459. https://doi.org/10.1016/B978-0-444-52898-8.00036-7.

Hoppe, C., Elger, C.E., Helmstaedter, C., 2007. Long-term memory impairment in patients with focal epilepsy. Epilepsia 48 (s9), 26–29.

Jaimes-Bautista, A.G., Rodríguez-Camacho, M., Martínez-Juárez, I.E., Rodríguez-Agudelo, Y., 2015. Semantic processing impairment in patients with temporal lobe epilepsy. Epilepsy Res. Treat. 2015, 1–8. https://doi.org/10.1155/2015/746745.

Jakobsen, J.C., Gluud, C., Wetterslev, J., Winkel, P., 2017. When and how should multiple imputation be used for handling missing data in randomised clinical trials – a practical guide with flowcharts. BMC Med. Res. Methodol. 17 (1), 162. https://doi.org/10.1186/s12874-017-0442-1.

Keary, T.A., Frazier, T.W., Busch, R.M., Kubu, C.S., Iampietro, M., 2007. Multivariate neuropsychological prediction of seizure lateralization in temporal epilepsy surgical cases. Epilepsia 48 (8), 1438–1446. https://doi.org/10.1111/j.1528-1167.2007.01098.x.

Kellermann, T.S., Bonilha, L., Eskandari, R., Garcia-Ramos, C., Lin, J.J., Hermann, B.P., 2016. Mapping the neuropsychological profile of temporal lobe epilepsy using cognitive network topology and graph theory. Epilepsy Behav. 63, 9–16. https://doi.org/10.1016/j.yebeh.2016.07.030.

Kim, H., Yi, S., Son, E.I., Kim, J., 2004. Lateralization of epileptic foci by neuropsychological testing in mesial temporal lobe epilepsy. Neuropsychology 18 (1), 141–151. https://doi.org/10.1037/0894-4105.18.1.141.

Ladino, L.D., Moien-Afshari, F., Téllez-Zenteno, J.F., 2014. A comprehensive review of temporal lobe epilepsy. In: Neurological Disorders : Clinical Methods. https://www.iconceptpress.com/book/neurological-disorders–clinical-methods/11000113/1306001000/.

Lomlomdjian, C., Múnera, C.P., Low, D.M., Terpiluk, V., Solís, P., Abusamra, V., Kochen, S., 2017. The right hemisphere's contribution to discourse processing : a study in temporal lobe epilepsy. Brain Lang. 171, 31–41. https://doi.org/10.1016/j.bandl.2017.04.001.

Loring, D.W., Strauss, E., Hermann, B.P., Barr, W.B., Perrine, K., Trenerry, M.R., Chelune, G., Westerveld, M., Lee, G.P., Meador, K.J., Bowden, S.C., 2008. Differential neuropsychological test sensitivity to left temporal lobe epilepsy. J. Int. Neuropsychol. Soc. 14 (3), 394–400. https://doi.org/10.1017/S1355617708080582.

MacPherson, S.E., Phillips, L.H., Della Sala, S., 2002. Age, executive function and social decision making : a dorsolateral prefrontal theory of cognitive aging. Psychol. Aging 17 (4), 598–609. https://doi.org/10.1037/0882-7974.17.4.598.

Martin, A., Chao, L.L., 2001. Semantic memory and the brain : structure and processes. Curr. Opin. Neurobiol. 11 (2), 194–201. https://doi.org/10.1016/S0959-4388(00)00196-3.

Mayeux, R., Brandt, J., Rosen, J., Benson, D.F., 1980. Interictal memory and language impairment in temporal lobe epilepsy. Neurology 30 (2). https://doi.org/10.1212/WNL.30.2.120, 120-120.

McAndrews, M.P., Cohn, M., 2012. Neuropsychology in temporal lobe Epilepsy : influences from cognitive neuroscience and functional neuroimaging. Epilepsy Res. Treat. https://doi.org/10.1155/2012/925238 [Research article].

Metternich, B., Buschmann, F., Wagner, K., Schulze-Bonhage, A., Kriston, L., 2014. Verbal fluency in focal Epilepsy : a systematic review and meta-analysis. Neuropsychol. Rev. 24 (2), 200–218. https://doi.org/10.1007/s11065-014-9255-8.

Molnar, C., 2018. *Interpretable Machine Learning*. Leanpub. https://leanpub.com/interpretable-machine-learning.

Munsell, B.C., Wee, C.-Y., Keller, S.S., Weber, B., Elger, C., da Silva, L.A.T., Nesland, T., Styner, M., Shen, D., Bonilha, L., 2015. Evaluation of machine learning algorithms for treatment outcome prediction in patients with epilepsy based on structural connectome data. Neuroimage 118, 219–230. https://doi.org/10.1016/j.neuroimage.2015.06.008.

Nogueira, S., Sechidis, K., Brown, G., 2018. On the Stability of Feature Selection Algorithms.

Oyegbile, T.O., Dow, C., Jones, J., Bell, B., Rutecki, P., Sheth, R., Seidenberg, M., Hermann, B.P., 2004. The nature and course of neuropsychological morbidity in chronic temporal lobe epilepsy. Neurology 62 (10), 1736–1742.

Pedregosa, F., Varoquaux, G., Gramfort, A., Michel, V., Thirion, B., Grisel, O., Blondel, M., Prettenhofer, P., Weiss, R., Dubourg, V., Vanderplas, J., Passos, A., Cournapeau, D., Brucher, M., Perrot, M., Duchesnay, É., 2011. Scikit-learn : machine learning in Python. J. Mach. Learn. Res. 12 (Oct), 2825–2830.

Pravatà, E., Sestieri, C., Mantini, D., Briganti, C., Colicchio, G., Marra, C., Colosimo, C., Tartaro, A., Romani, G.L., Caulo, M., 2011. Functional connectivity MR imaging of the language network in patients with drug-resistant epilepsy. Am. J. Neuroradiol. 32 (3), 532–540. https://doi.org/10.3174/ajnr.A2311.

Pustina, D., Avants, B., Sperling, M., Gorniak, R., He, X., Doucet, G., Barnett, P., Mintzer, S., Sharan, A., Tracy, J., 2015. Predicting the laterality of temporal lobe epilepsy from PET, MRI, and DTI : a multimodal study. Neuroimage: Clinical 9, 20–31. https://doi.org/10.1016/j.nicl.2015.07.010.

Ribeiro, M.T., Singh, S., Guestrin, C., 2016. Model-Agnostic Interpretability of Machine Learning. https://arxiv.org/abs/1606.05386v1.

Roger, E., Petit, L., Perrone-Bertolotti, M., Job, A.-S., Minotti, L., Kahane, P., Baciu, M., 2018. The link between structural connectivity and neurocognition illustrated by focal epilepsy. Epileptic Disord Epileptic Disord 20, 88–98. https://doi.org/10.1684/epd.2018.0958.

Rudzinski, L.A., Meador, K.J., 2013. Epilepsy and neuropsychological comorbidities. CONTINUUM: Lifelong Learn. Neurol. 19 (3), 682. https://doi.org/10.1212/01.CON.0000431382.06438.cd.

Salvatore, C., Cerasa, A., Castiglioni, I., Gallivanone, F., Augimeri, A., Lopez, M., Arabia, G., Morelli, M., Gilardi, M.C., Quattrone, A., 2014. Machine learning on brain MRI data for differential diagnosis of Parkinson's disease and Progressive Supranuclear Palsy. J. Neurosci. Methods 222, 230–237. https://doi.org/10.1016/j.jneumeth.2013.11.016.

Senders, J.T., Staples, P.C., Karhade, A.V., Zaki, M.M., Gormley, W.B., Broekman, M.L.D., Smith, T.R., Arnaout, O., 2018. Machine learning and neurosurgical outcome Prediction : a systematic review. World Neurosurgery 109, 476–486. https://doi.org/10.1016/j.wneu.2017.09.149 e1.

Stemmer, B., Whitaker, H.A., 2008. Handbook of the Neuroscience of Language. Academic Press.

Stroop, J.R., 1935. Studies of interference in serial verbal reactions. J. Exp. Psychol. 18 (6), 643–662. https://doi.org/10.1037/h0054651.

Tatum, W.O.I., 2012. Mesial temporal lobe epilepsy. J. Clin. Neurophysiol. 29 (5), 356. https://doi.org/10.1097/WNP.0b013e31826b3ab7.

Taylor, R.S., Sander, J.W., Taylor, R.J., Baker, G.A., 2011. Predictors of health-related quality of life and costs in adults with epilepsy : a systematic review. Epilepsia 52 (12), 2168–2180. https://doi.org/10.1111/j.1528-1167.2011.03213.x.

Thivard, L., Hombrouck, J., Tézenas du Montcel, S., Delmaire, C., Cohen, L., Samson, S., Dupont, S., Chiras, J., Baulac, M., Lehéricy, S., 2005. Productive and perceptive language reorganization in temporal lobe epilepsy. Neuroimage 24 (3), 841–851. https://doi.org/10.1016/j.neuroimage.2004.10.001.

Torlay, L., Perrone-Bertolotti, M., Thomas, E., Baciu, M., 2017. Machine learning–XGBoost analysis of language networks to classify patients with epilepsy. Brain Informatics 4 (3), 159–169. https://doi.org/10.1007/s40708-017-0065-7.

Tramoni-Negre, E., Lambert, I., Bartolomei, F., Felician, O., 2017. Long-term memory deficits in temporal lobe epilepsy. Rev. Neurol. 173 (7), 490–497. https://doi.org/10.1016/j.neurol.2017.06.011.

Vandenberghe, R., Wang, Y., Nelissen, N., Vandenbulcke, M., Dhollander, T., Sunaert, S., Dupont, P., 2013. The associative-semantic network for words and pictures : effective connectivity and graph analysis. Brain Lang. 127 (2), 264–272. https://doi.org/10.1016/j.bandl.2012.09.005.

Vogt, V.L., Äikiä, M., Barrio, A. del, Boon, P., Borbély, C., Bran, E., Braun, K., Carette, E., Clark, M., Cross, J.H., Dimova, P., Fabo, D., Foroglou, N., Francione, S., Gersamia, A., Gil-Nagel, A., Guekht, A., Harrison, S., Hecimovic, H., et al., 2017. Current standards of neuropsychological assessment in epilepsy surgery centers across Europe. Epilepsia 58 (3), 343–355. https://doi.org/10.1111/epi.13646.

Wang, W.-H., Liou, H.-H., Chen, C.-C., Chiu, M.-J., Chen, T.-F., Cheng, T.-W., Hua, M.-S., 2011. Neuropsychological performance and seizure-related risk factors in patients with temporal lobe epilepsy : a retrospective cross-sectional study. Epilepsy Behav. 22 (4), 728–734. https://doi.org/10.1016/j.yebeh.2011.08.038.

Wechsler, D., 2011. Wechsler Adult Intelligence Scale–Fourth Edition (WAIS–IV). ECPA.

Wechsler, D., 2012. Test MEM-IV échelle clinique de mémoire de WECHSLER- quatrième édition- Psychologie clinique. ECPA. https://www.ecpa.fr/psychologie-clinique/test.asp?id=1987.

Wieser for the ILAE Commission on Neurosurgery of Epilepsy, 2004. Mesial temporal lobe epilepsy with hippocampal sclerosis. Epilepsia 45 (6), 695–714. https://doi.org/10.1111/j.0013-9580.2004.09004.x.

Witt, J.-A., Elger, C.E., Helmstaedter, C., 2013. Which drug-induced side effects would be tolerated in the prospect of seizure control? Epilepsy Behav. 29 (1), 141–143. https://doi.org/10.1016/j.yebeh.2013.07.013.

Witt, J.-A., Helmstaedter, C., 2012. Should cognition be screened in new-onset epilepsies? A study in 247 untreated patients. J. Neurol. 259 (8), 1727–1731. https://doi.org/10.1007/s00415-012-6526-2.

Witt, J.-A., Helmstaedter, C., 2015. Cognition in the early stages of adult epilepsy. Seizure 26, 65–68. https://doi.org/10.1016/j.seizure.2015.01.018.