Efficiency ComEfficiency Comparison of Machine Learning Algorithms for EEG Interpretation parison of Machine Learning Algorithms for EEG Interpretation

Xia Han, Frédéric Amiel, Xun Zhang, Kunni Wei, Cong Yan, Wenjun Hu, Zefeng Wang

Efficiency Comparison of Machine Learning Algorithms for EEG Interpretation

Xia HAN[†], Frédéric AMIEL[†], Xun ZHANG[†], Kunni WEI[§], Cong YAN[§], Wenjun HU[*], Zefeng WANG[*]
Email: xia.han@ext.isep.fr, frederic.amiel@isep.fr, xun.zhang@isep.fr, iannya@bucm.edu.cn,
yancong@bucm.edu.cn, wenjun.hu@zjhu.edu.cn, zefeng.wang@zjhu.edu.cn
[†]LISIT-ECoS, Institut Supérieur D'électronique de Paris, Paris, France
[§]The School of Life Sciences, Beijing University of Chinese Medicine, Beijing, China
[*]The School of Information Engineering, Huzhou university, Huzhou, China

***Abstract*—This paper intends to use a small protocol to detect stroke disease on a patient by using signals provided by only three EEG probes. To achieve this objective, we compare the performances in terms of accuracy and time of six machine learning (ML) algorithms (Random Forest, Logistic Regression, Support Vector Machine, K-Nearest Neighbor, Decision Tree and CatBoost) during a process of EEG-based classification pathology. We use a database of EEG recording signals collected by three electrodes, established by Beijing University of Chinese Medicine and carried out on subjects healthy or affected by strokes when they are exposed to the vision of planes of five different colors. The subjects are known to be healthy or affected by strokes. The records are used to train each algorithm for 70% of the population, and the performances are estimated on the remaining 30%. Then the process is repeated one hundred times when changing the set used for training and the set used to test. We then consider a statistic on the results obtained using each method for comparison. Our results show that the SVM algorithm is the most efficient in terms of the accuracy of the results, and can detect stoke disease with a reliability of 70%.**

***Keywords*—Machine Learning, EEG, stroke EEG classification, algorithm performance**

I. Introduction

Machine learning (ML) technology allows a model to be established after a phase of learning from a data-set. Then the model is used to analyze and classify the following data. ML is adaptable to many applications in different fields such as finance, marketing, industrial and medical field. In medical field, ML often use Bio-electrical signal to detect or diagnosis the disease like epilepsy, depression and stroke. Bio-electrical signal is an electrical signal closely related to the state of health generated by cells or tissues.It contains hidden information of biological activities and can be used to reflect part of the current physiological state. The commonly used bio-electrical signal are ECG (Electrocardiogram), EMG (Electromyography) and EOG (Electro-Oculogram) and EEG (Electroencephalogram) [1]. EEG is one of the most commonly used bio-electrical signals. It records the electrical signals spontaneously generated during the transmission of information between neuron cells in the brain [2]. It is mostly used in the diagnosis of epilepsy [3], the detection of mental diseases, and also in the diagnosis of organic or functional brain lesions such as stroke [4] .

Random Forest (RF) is applied as a classifier in an EEG-based automatic sleep stage identification system [5] and it's also used to analyse and classify EEG data for human mental state [6]. Logistic Regression (LR) was regarded as a method to identify state of epileptic seizure or non-epileptic by using EEG signals in paper [7]. Support Vector Machine (SVM) is often used in biomedical field, it was exploited to classify eye events from EEG signal in [8] and for a time-frequency image of EEG seizure classifier in [9]. Raja [10] realised an EEG-based emotion classification system using SVM and K-nearest Neighbor (KNN). And in research [11], KNN is used in a motor imagery EEG recognition system. An automated detection of driver fatigue based on EEG signals uses Decision tree (DT) [12]. CatBoost has been used to analyses and classify the EEG signal for the identification of construction hazard-relatedin a context of a virtual reality device and achieved an accuracy result as 95.1% in [13]. And it was also used in an automatic detection of abnormal EEG signals task and get the binary classification accuracy of 87.68% [14].

EEG has been proven to be useful in the diagnosis of stroke [15], [16]. Stroke is an acute cerebrovascular disease, which is a group of diseases that cause brain tissue damage due to sudden rupture of blood vessels in the brain or blockage of blood vessels that prevents blood from flowing into the brain. As reported in [17], stroke occurs more in 2019 than in 2016 [18], showing a rising trend of incidence. Considering its suddenness, the diagnosis of stroke must be made as soon as possible for a timely treatment. However, there is a wide variety of clinical manifestations, which makes the diagnosis of stroke difficult despite the help of EEG. SVM, DT and RF were used to classify ischemic stroke and hemorrhagic stroke with the multi-feature fusion method in [19]. LR was used to help the prediction of post-stroke epilepsy in [20] and prove that EEG can predict the epilepsy after stroke. SVM and KNN were used in the research of the discrimination of stroke-related mild cognitive impairment and vascular dementia [21]. CatBoost with the statistical feature decomposition by wavelet package decomposition was used for the binary detection of abnormal EEG signal [22].

Notwithstanding the stat of art articles have employed ML algorithms in EEG-based applications related to stroke, we still would like to compare the performance of the mentioned

ML algorithms and to find the most suitable algorithm for EEG classification for the further research on stroke. Neural network permits to reach good performances in the case of neural disorders [23] but they require large data base to be trained and are more complex to tune. As we use the BUCM data base which is small at this moment (only 30 data sets), we don't have enough measurement to train a neural network efficiently. Neural Networks trained with small dataset exhibit unstable behavior in performances [24]. Also, the BUCM data base is specific as it uses a non-invasive system with only three electrodes to capture EEG signals. In this study we focus on classical ML algorithms which are directly interpretable. Therefore, in this paper, a comparison of accuracy and the execute time of these six ML algorithms is proposed. The rest of the paper is organized as follows: Section II presents the methods used for comparison including the using data-set, the pre-processing and feature extraction as well as classification methods. Section III shows the results and discussions of stroke classification. Section IV a conclusion and issues to be resolved in the future.

II. METHODS

A. Data-Set

The data-set used in this study is private and comes from the school of life sciences of Beijing University of Chinese Medicine. The experimental protocol is as follow: for each subject, five colors (black, green, red, white and yellow) are presented sequentially front of the subject on full screen (23.8 inches with a refresh rate of 75 Hz) (Figure 1) during 10 seconds. The actual data base contains EEG signals from 30 subjects, 15 of them are diagnosed with stroke and 15 are healthy. There are two types of stroke: ischemic and hemorrhagic, but they are not distinguished in this data-set. The records were made using an OpenBCI device, which samples three channels Fp1, Fp2 and Fpz with a sampling frequency of 250 Hz, according to the 10-20 international standard system. As a result, the records contain 2500 samples over three channels for each color. In this study, we considered different colors as different use cases.

B. Pre-processing method

EEG signal is very weak and is easily influenced by the environment (EMG noise) during the acquisition. In order to clean the signal, we apply high pass filter with a cut-off frequency at 0.1 Hz and a low pass filter at 30 Hz. We choose FIR structure.

C. Feature extraction methods

EEG signal changes with the emotions, the mentality, states during the test. The collection of EEG signals is a long-term process which provides EEG large amount data points. We will extract main characteristics into fewer parameters, and we call it the feature extraction. In this study, the Shannon Entropy was used to feature the signal. Shannon entropy also known as information entropy [25], reflects the relationship between the amount of information and its uncertainty, who measures the probability density based on the probability distribution of amplitude values. For a random variable S ($S = s_1, s_2, ..., s_n$), the Shannon entropy S_{en} is defined as follow:

$$\begin{aligned} S_{en} &= \sum_{i=1}^{n} P(s_i) log_a \frac{1}{P(s_i)} \\ &= -\sum_{i=1}^{n} P(s_i) log_a P(s_i) \end{aligned} \tag{1}$$

Where $P(s_i)$ is the probability that the random variable S is equal to s_i, i represents data point and a is stands for logarithmic base. In this experiment, we claculate the Shannon's entropy value for each second referenced by previous second. Since there are 10 seconds per color to analyze, we obtain 9 Shannon entropy values of each color. It is coded by a floating point with six decimal places.

D. Classification of stroke EEG signal

In order to avoid over fitting, the learning model and to extract as much information as possible from the data-set, the cross-validation method was applied 100 times on our different records. The *StratifiedShuffleSplit()* function permits to establish different sets (training / check) from our initial records. We use 70% of the total data of both diseased and healthy subjects for training and the remainder 30% for the test. Regarding the classification algorithms, we used *RandomForestClassifier, SVC, LogisticRegression, KNeighborsClassifier, DecisionTreeClassifier* from the *sklearn* package in Python to implement RF (Random Forest), LR (Logitic Regression), SVM (Support Vector Machine), KNN (K-Nearest Neighbor, where k=3 in this study) and DT (Decision Tree), respectively. Concerning the CatBoost methode, we use the *CatBoostClassifier*.

The state-of-art articles [26]–[28] present the performance of the algorithms in different ways, but in our study, the comparison is based on the average of their accuracies (obtained by hundreds of executions) and on their standard deviation and their execution times. A combination of those results gives our evaluation index. The whole experimental flow is showed in figure1.

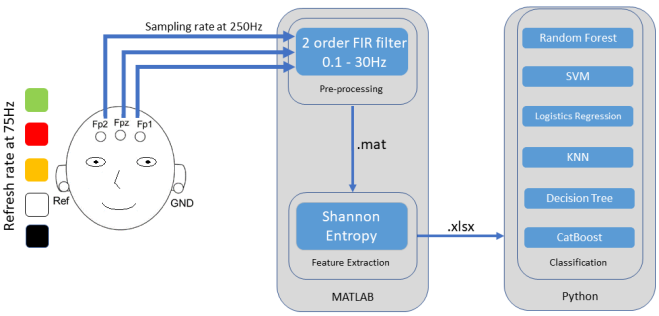

Fig. 1. Experimental flow chart

III. Result and Discussion

A. Mean accuracy of classification

Each algorithm gives a classification, in our case Healthy (H) or Stroke (S) associated with a score. Then we apply the algorithms for each plane of color which gives 100 scores (%H, %S) for each color (then we got 500 different scores) and we do it for our three channels. During 100 times training and test, we record the accuracy and the corresponding execute time of each classification. Then the mean accuracy and execute time of each case were calculated. To discuss the result of each channel, we organize the presentation with a "box plot" from the minimum value to the maximum one, in the following figures2, 3 and 4.

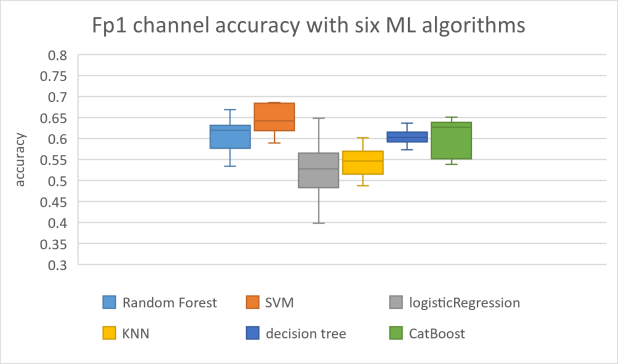

Fig. 2. Average accuracy of each ML algorithm in Fp1 channel

Figure 2 presents the results obtained by each of algorithms and for 5 different colors for the Fp1 channel only. We can see that SVM has the best classification accuracy at 68.56% and that decision tree shows more stable performance. The result of the logistic regression is more diverse.

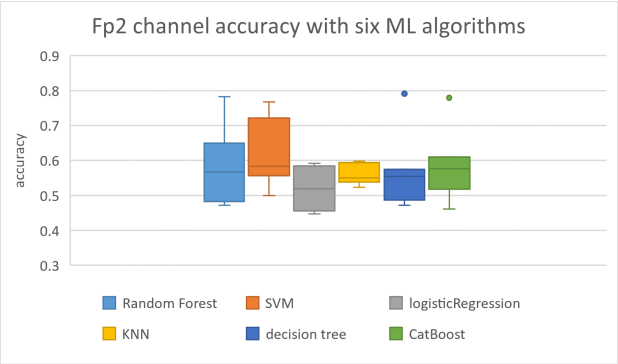

Fig. 3. Average accuracy of each ML algorithm in Fp2 channel

The result of average accuracy of 100 runs for five colors of channel Fp2 for the six ML algorithms is shown in Figure 3. We can see that both random forest and SVM have the best results for the classification task and that the distribution of KNN results is more concentrated than others. Moreover, the accuracy results of random forest, SVM and logistic regression seem worse than in Fp1. We also see that with the decision tree and CatBoost, we have a very good result but far for the majority scores displayed in the "box plot".

Figure 4 shows the results of the average of the results for the six ML algorithms with their execution times and for the five colors using the samples captured by the Fpz channel. SVM's algorithm achieves the get the highest ranking result with a maximum score of 71.89%.

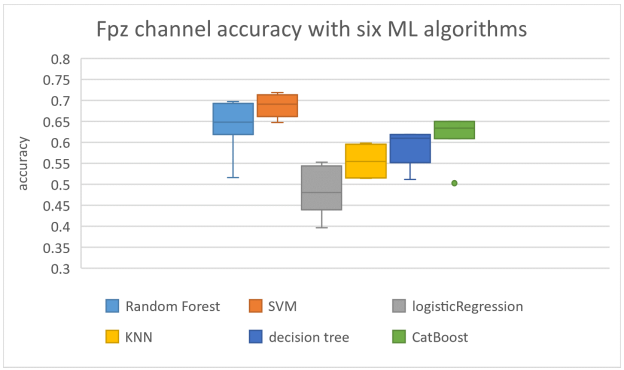

Fig. 4. Average accuracy of each ML algorithm in Fpz channel

We can see logistic regression gets bad scores, moreover for Fp2 its results are equivalent to a random classifier. We also see that the data sampled by Fpz obtain the best score at 68.67%, and those sampled by Fp2 are the worst. As the electrodes are quite close, a suspicion of trouble concerning connection is made. This observation suggests to create and to add a system to reject some measurements in the future.

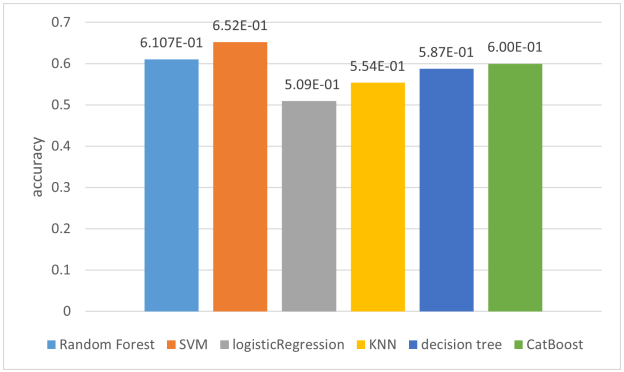

Fig. 5. Mean accuracy of each ML algorithm in three channel

In order to remove the effect of bad measurements on a probe which would be due to a possible bad contact, we consider the mean accuracy for the three electrodes and we present the obtained scores of the five colors for each algorithm in Figure 5. SVM has by far the best classification with a score of 65.22%.

Although the score is an important criterion, the dispersion of the results is also critical for the use of the results of the algorithms. Figure 6 shows the standard deviation of the classification accuracy of all channels (Fp1, Fp2, Fpz) for the six ML algorithms. In this figure, we can see that KNN has the

smallest standard deviation, and is therefore the most faithful. SVM also shows a good results considering this criterion.

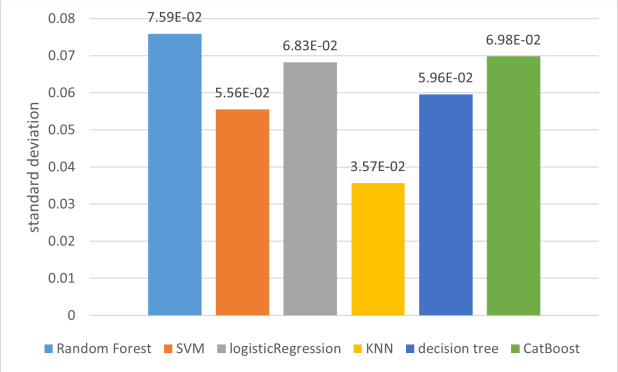

Fig. 6. Accuracy standard deviation of each ML algorithm

B. *Execution time of classification*

We programmed all the algorithms in python on a DELL laptop based on an Intel i7-1075 processor clocked at 2.6 GHz with 6 physical cores (12 logical cores). Table I gives the execution time for 100 executions of the algorithms. The entropy calculation from the sampled data is pre-processed before and is not included in these measurements. The decision tree is obviously the fastest algorithm, and CatBoost is the longest with a thousand times slower execution time.

TABLE I
MEAN EXECUTION TIME (UNIT: S) AND STANDARD DEVIATION

Algo	RF	LR	SVM	KNN	DT	CatBoost
Mean	9.82E-02	7.76E-04	2.88E-03	1.26E-03	6.34E-04	6.94E-01
SD	1.49E-03	1.16E-05	1.27E-04	1.95E-05	9.74E-06	2.47E-02

IV. CONCLUSION

The results presented in this study show that the used protocol: presenting colors on a screen front of a subject and collecting measurement by only three EEG probes permits to detect suspicion of stroke with an accuracy of 65%. Concerning the algorithms, SVM seems the more appropriate as it gives the best results among the different combination of electrodes and colors. However, KNN is also a serious candidate and gets more stability. Our study covers only 30 subjects and we have yet to increase this test sample, even if we were able to vary the conditions for establishing the training set and the test set. Further research is underway to verify the results of the experiment on more subjects and to analyze the variability of results depending on the choice on training sets / test sets. Possibly we may need to qualify each set of data coming from the channels before processing

ACKNOWLEDGMENT

This work was supported by China Scholarship Council under Grant 202208070074.

REFERENCES

[1] Pal A, Gautam A K, Singh Y N. Evaluation of bioelectric signals for human recognition[J]. Procedia Computer Science, 2015, 48: 746-752.

[2] Elul R. The genesis of the EEG[J]. International review of neurobiology, 1972, 15: 227-272.

[3] Smith S J M. EEG in the diagnosis, classification, and management of patients with pilepsy[J]. Journal of Neurology, Neurosurgery & Psychiatry, 2005, 76(suppl 2): ii2-ii7.

[4] Jordan K G. Emergency EEG and continuous EEG monitoring in acute ischemic stroke[J]. Journal of clinical neurophysiology, 2004, 21(5): 341-352.

[5] Luay Fraiwan, Khaldon Lweesy, Natheer Khasawneh, Heinrich Wenz, Hartmut Dickhaus, Automated sleep stage identification system based on time–frequency analysis of a single EEG channel and random forest classifier, Computer Methods and Programs in Biomedicine, Volume 108, Issue 1, 2012, Pages 10-19, ISSN 0169-2607, Https://doi.org/10.1016/j.cmpb.2011.11.005.

[6] Damodar Reddy Edla, Kunal Mangalorekar, Gauri Dhavalikar, Shubham Dodia, Classification of EEG data for human mental state analysis using Random Forest Classifier, Procedia Computer Science, Volume 132, 2018, Pages 1523-1532, ISSN 1877-0509, https://doi.org/10.1016/j.procs.2018.05.116.

[7] Abdulhamit Subasi, Ergun Erçelebi, Classification of EEG signals using neural network and logistic regression, Computer Methods and Programs in Biomedicine, Volume 78, Issue 2, 2005, Pages 87-99, ISSN 0169-2607, https://doi.org/10.1016/j.cmpb.2004.10.009.

[8] Singla R, Chambayil B, Khosla A, et al. Comparison of SVM and ANN for classification of eye events in EEG[J]. Journal of Biomedical Science and Engineering, 2011, 4(1): 62.

[9] Fu K, Qu J, Chai Y, et al. Classification of seizure based on the time-frequency image of EEG signals using HHT and SVM[J]. Biomedical Signal Processing and Control, 2014, 13: 15-22.

[10] R. M. Mehmood and H. J. Lee, "Emotion classification of EEG brain signal using SVM and KNN," 2015 IEEE International Conference on Multimedia & Expo Workshops (ICMEW), Turin, Italy, 2015, pp. 1-5, doi: 10.1109/ICMEW.2015.7169786.

[11] Tang X, Wang T, Du Y, et al. Motor imagery EEG recognition with KNN-based smooth auto-encoder[J]. Artificial intelligence in medicine, 2019, 101: 101747.

[12] Hu J, Min J. Automated detection of driver fatigue based on EEG signals using gradient boosting decision tree model[J]. Cognitive neurodynamics, 2018, 12: 431-440.

[13] JungHo Jeon, Hubo Cai, Classification of construction hazard-related perceptions using: Wearable electroencephalogram and virtual reality, Automation in Construction, Volume 132, 2021, 103975, ISSN 0926-5805, https://doi.org/10.1016/j.autcon.2021.103975.

[14] Hezam Albaqami, Ghulam Mubashar Hassan, Abdulhamit Subasi, Amitava Datta, Automatic detection of abnormal EEG signals using wavelet feature extraction and gradient boosting decision tree, Biomedical Signal Processing and Control, Volume 70, 2021, 102957, ISSN 1746-8094, https://doi.org/10.1016/j.bspc.2021.102957.

[15] Finnigan S, van Putten M J A M. EEG in ischaemic stroke: quantitative EEG can uniquely inform (sub-) acute prognoses and clinical management[J]. Clinical neurophysiology, 2013, 124(1): 10-19.

[16] Qureshi A A, Zhang C, Zheng R, et al. Ischemic stroke detection using EEG signals[C]//CASCON. 2018: 301-308.

[17] Owolabi MO, Thrift AG, Martins S, Johnson W, Pandian J, Abd-Allah F, Varghese C, Mahal A, Yaria J, Phan HT, Roth G, Gall SL, Beare R, Phan TG, Mikulik R, Norrving B, Feigin VL; Stroke Experts Collaboration Group. The state of stroke services across the globe: Report of World Stroke Organization-World Health Organization surveys. Int J Stroke. 2021 Oct;16(8):889-901. doi: 10.1177/17474930211019568. Epub 2021 May 27. PMID: 33988062; PMCID: PMC8800855.

[18] Global Health Estimates.Geneva: World Health Organization; 2012. Available from : http://www.who.int/healthinfo/global_burden_disease/en/

[19] Li, F., Fan, Y., Zhang, X. et al. Multi-Feature Fusion Method Based on EEG Signal and its Application in Stroke Classification. J Med Syst 44, 39 (2020). https://doi.org/10.1007/s10916-019-1517-9

[20] Bentes C, Martins H, Peralta A R, et al. Early EEG predicts poststroke epilepsy[J]. Epilepsia open, 2018, 3(2): 203-212.

[21] Al-Qazzaz, N.K., Ali, S.H.B.M., Ahmad, S.A. et al. Discrimination of stroke-related mild cognitive impairment and vascular dementia using EEG signal analysis. Med Biol Eng Comput 56, 137–157 (2018). https://doi.org/10.1007/s11517-017-1734-7

[22] Hezam Albaqami, Ghulam Mubashar Hassan, Abdulhamit Subasi, Amitava Datta, Automatic detection of abnormal EEG signals using wavelet feature extraction and gradient boosting decision tree, Biomedical Signal Processing and Control, Volume 70, 2021, 102957, ISSN 1746-8094, https://doi.org/10.1016/j.bspc.2021.102957.

[23] Hung C Y, Chen W C, Lai P T, et al. Comparing deep neural network and other machine learning algorithms for stroke prediction in a large-scale population-based electronic medical claims database[C]//2017 39th Annual International Conference of the IEEE Engineering in Medicine and Biology Society (EMBC). IEEE, 2017: 3110-3113.

[24] Shaikhina T, Khovanova N A. Handling limited datasets with neural networks in medical applications: A small-data approach[J]. Artificial intelligence in medicine, 2017, 75: 51-63.

[25] EEG signal processing and feature extraction[M]. Singapore: Springer Singapore, 2019.

[26] Merghadi A, Yunus A P, Dou J, et al. Machine learning methods for landslide susceptibility studies: A comparative overview of algorithm performance[J]. Earth-Science Reviews, 2020, 207: 103225.

[27] Vabalas A, Gowen E, Poliakoff E, et al. Machine learning algorithm validation with a limited sample size[J]. PloS one, 2019, 14(11): e0224365.

[28] Liu X, Faes L, Kale A U, et al. A comparison of deep learning performance against health-care professionals in detecting diseases from medical imaging: a systematic review and meta-analysis[J]. The lancet digital health, 2019, 1(6): e271-e297.