

Detection of Epilepsy Cases in Newborns

Gérard Behou N'Guessan^{1*}, Kouassi Saha Bernard², Coulibaly Tiékoura³, Diarra Bassira¹

¹Research and Digital Expertise Unit (UREN), Virtual University of Côte d'Ivoire (UVCI), Abidjan, Ivory Coast

²Higher Normal School of Abidjan, Abidjan, Ivory Coast

³Lastic, ESATIC, Abidjan, Ivory Coast

Email: *behou.nguessan@uvci.edu.ci

How to cite this paper: N'Guessan, G.B., Bernard, K.S., Tiékoura, C. and Bassira, D. (2023) Detection of Epilepsy Cases in Newborns. *Engineering*, 15, 134-142.
<https://doi.org/10.4236/eng.2023.152011>

Received: January 15, 2023

Accepted: February 25, 2023

Published: February 28, 2023

Copyright © 2023 by author(s) and Scientific Research Publishing Inc.
This work is licensed under the Creative Commons Attribution International License (CC BY 4.0).

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



Open Access

Abstract

Epilepsy is a very common worldwide neurological disorder that can affect a person's quality of life at any age. People with epilepsy typically have recurrent seizures that can lead to injury or in some cases even death. Curing epilepsy requires risky surgery. If not, the patient may be subjected to a long drug treatment associated with lifestyle advice without guarantee of total recovery. However, regardless of the type of treatment performed, late treatment necessarily creates psychological instability in the patient. It is therefore important to be able to diagnose the disease as early as possible if we desire that the patient does not suffer from its consequences on their mental health. That is why the study aims to propose a model for detecting epilepsy in order to be able to identify it as early as possible, especially in newborns. The objective of the article is to propose a model for detecting epilepsy using data from electroencephalogram signals from 10 newborns. This model developed using the extra trees classifier technique offers the possibility of predicting epilepsy in infants with an accuracy of around 99.4%.

Keywords

Neonatal Epilepsy, Electroencephalogram Signal, Supervised Classification, Random Forest, Extratrees, Gradient Boosting Tree

1. Introduction

Epilepsy is one of the most widespread and devastating neurological diseases in the world [1]. The hallmark of epilepsy is a set of abrupt and chronic changes in neuronal electrical activity causing recurrent seizures called "epileptic seizures" [2]. These misdiagnosed or ignored attacks can go unnoticed and intensify in the long term. Diagnosis of epileptic seizures is usually made by electroencephalography [3]. Electroencephalography is a technique that analyzes electrical signals

from the brain. Those electrical signals can be represented as graphs called electroencephalograms (EEG). The EEG interpretation permits to detect the possible epilepsy presence. Nevertheless, for the patient to be fully cured without the disease sequelae in their mental health, it is necessary that the diagnosis be made at the disease onset. So, simpler and easier-to-use methods are needed for early diagnosis of the disease on a large scale. However, the visual interpretation of the EEG carried out by the epileptologist not only takes time, but can prove to be inefficient because of the disease visual identification mechanism complexity. Beyond these problems, the fact that seizures can be overestimated when the patient moves; which leads to noises and muscular artifacts disturbing the signal. Considering these circumstances, we wonder: How can we automatically detect epileptic seizures in newborns?

To address the aforementioned concern, we set the goal of proposing a machine learning model capable of efficiently identifying the presence or absence of epilepsy in the patient by essentially using electrical signals from the brain. Therefore, we have organized that article around three stages. The first part presents work on epilepsy; the second section deals with the methodology used. The third one presents the results of our work in which we obtained a random forest model for detecting neonatal epileptic seizures that is 99.4% correct.

2. State of the Art

The word epilepsy comes from the Greek “epilēpsia” [4] which designates a sudden attack, a sudden stop. That disease is considered a demonic disease thus preventing any scientific progress [5]. The negative prejudices surrounding the disease were obstacles to the proper care of patients, their development and social integration. It was at the beginning of the eighteenth century that Samuel Auguste David Tissot had looked into that very disease from a scientific approach in his book “Traité de l'épilepsie” in 1770 [6]. ESAMBERT, B. and ALLONNEAU-ROBERTIE, E. proved that 20% of children with epilepsy showed a significant backwardness in certain learning areas, in particular reading and sport [7]. According to the work of Vibha Patel *et al.* [8] [9], the method of detecting epileptic seizures can be classified into four categories. It is a method based on the traditional machine learning approach, the deep learning approach, signal processing approach and hybrid approach. Artur Gramacki and Jarosław Gramacki, on the other hand, focus on the deep learning method. Although the said approach presents a very good performance, but it is all the same non-parametric and requires a large number of data for its configuration [10]. P. Geethanjali on the other hand, asserts that no method has so far shown both high sensitivity and zero false alarms to obtain reliable results [11]. P. Fergus, D. Hignett, A. Hussain, D. Al-Jumeily, K. Abdel-Aziz [12] [13] automatic detection of epileptic seizures using EEG [14] [15]. Using the aforementioned tool, the authors were able to extract the characteristics of a person living with epilepsy via their scalp. C. Donos, M. Dümpelmann, A. Schulze-Bonhage [16] also used the intracranial

EEG tool to detect seizures early. The authors use the same characteristics which gave them an average sensitivity of 93.84%. Likewise Jerry Yu [17] used the EEG to classify epileptic seizures. The method enabled them to classify the disease as epilepsy with an accuracy of 84.2%.

However, rare are the studies which are interested in the issue of epilepsy in newborns. Hence the interest of proposing a model capable of evaluating the presence or absence of that disease in children under one year of age.

3. Methodological Approach

As part of the study we used the Electrophysiological technique. That technique is based on the EEG (Electroencephalography) method and uses electrodes placed on the surface of the head to collect recordings from the cerebral cortex via the EEGLab software [15]. The data comes from epileptic children and listed in **Figure 1** below. We were able to obtain ten (10) EEG files that were preprocessed from the EEGlab software, labeled using the annotations of the three (3) experts, then concatenated. The new database consists of 1048544 seconds of annotated lines of recordings including the variables that are the EEG signal sensors placed on the head of infants.

The input variables in table 1 above are sensors of the EEG signals which are: Fp2, F4, C4, P4, O2, Fp1, F3, C3, P3, O1, F8, T4, T6, F7, T3, T5, Fz, Cz, Pz as well as the ECG/EKG and Resp-effort which collect the amplitudes of the cerebral, cardiac and respiratory activity of the patients. The designations of the sensors represent the positions at which they are placed on the patient's skull such as: F stands for Frontal, P stands for Parietal, T stands for Temporal, O stands for Occipital, C: Central and z represented the middle line. The even numbers 2, 4, 6, 8 represent the right hemisphere while the odd numbers 1, 3, 5, 7 represent the left hemisphere. The ECG/EKG corresponds to the electrocardiogram and Resp-effort corresponds to the infant's breathing effort. Our response variable y constitutes the signal state at that second (epileptics or non-epileptics).

4. Modelization

Through the present article, we evaluate the entropy measure and the Gini index in the detection of epilepsy disease in newborns. The Gini index allows to measure

Time	EEG Fp1-Ref	EEG Fp2-Ref	EEG F3-Ref	EEG F4-Ref	EEG F7-Ref	EEG F8-Ref	EEG Fz-Ref	EEG C3-Ref	EEG C4-Ref	EEG Cz-Ref	EEG T3-Ref	EEG T5-Ref	EEG T4-Ref	EEG T6-Ref	EEG P3-Ref	EEG P4-Ref	EEG				
1.0	-15.1	-10.5	-0.3	-4.5	-3.2	-5.5	-6.2	-0.2	0.3	0.2	3.7	-0.3	-4.8	-4.0	-1.8	3.4	0.0	1.2	4.2	-1.8	30
1.0	-14.7	-8.5	-0.7	-3.3	-4.2	-4.2	-5.0	0.8	0.5	0.5	3.7	-1.3	-7.9	-1.0	-1.8	4.4	0.0	1.3	4.7	0.5	30
1.1	-14.4	-8.4	-1.2	-3.3	-5.2	-4.8	-5.0	1.3	0.5	-0.2	3.0	1.5	-11.2	-0.8	-1.8	3.5	-0.5	0.7	4.9	2.2	30
1.1	-14.6	-9.9	0.2	-2.8	-3.2	-4.5	-4.2	1.8	1.0	0.2	3.2	1.2	-6.2	2.0	-1.0	5.5	0.8	1.8	3.9	1.2	30
1.1	-15.1	-8.0	-0.3	-5.0	-3.3	-5.0	-4.8	2.2	-0.3	0.0	4.0	4.5	-5.7	-5.2	-1.2	3.5	0.7	2.7	3.0	0.8	30
1.2	-13.9	-8.7	0.0	-4.0	-3.7	-5.0	-4.7	1.8	-0.5	-0.2	4.5	3.5	-4.3	-6.0	-1.8	4.0	0.2	1.8	3.5	1.5	30
1.2	-14.7	-6.4	0.2	-3.5	-3.3	-4.2	-4.2	1.3	0.2	0.0	4.7	-1.7	-5.7	0.0	-2.2	5.5	0.0	0.8	4.5	-1.0	30
1.3	-18.9	-6.7	-0.2	-3.3	-4.7	-4.3	-4.8	0.8	-0.7	-0.3	3.2	4.0	-10.7	-1.0	-2.0	4.5	-0.5	1.3	5.2	-5.3	30
1.3	-11.0	-6.9	0.2	-3.8	-4.3	-4.3	-4.3	2.0	-0.2	-0.3	6.2	-3.2	-8.9	2.2	-1.2	5.0	0.0	1.7	4.9	-4.8	30
1.3	-11.7	-7.5	-0.3	-4.3	-2.7	-4.2	-4.0	1.2	1.2	0.3	5.7	-1.2	-4.3	-1.2	-1.3	4.4	0.8	2.2	3.2	0.5	29
1.4	-12.4	-9.7	0.8	-4.3	-3.0	-5.0	-4.5	0.7	-0.3	-0.3	4.5	0.2	-4.8	-1.2	-1.8	4.2	0.0	1.5	3.4	2.8	29
1.4	-12.7	-7.5	0.5	-3.3	-3.0	-4.2	-3.5	1.7	0.5	-0.2	3.4	1.5	-6.9	0.3	-2.0	4.0	0.7	3.2	6.0	0.8	29
1.4	-12.2	-9.0	0.8	-3.0	-3.8	-4.3	-3.8	1.5	0.8	0.3	3.5	2.0	-10.9	3.4	-1.7	4.5	0.2	1.8	5.2	0.3	29
1.5	-11.7	-8.7	1.3	-3.0	-3.5	-4.2	-3.8	1.8	1.0	0.0	4.9	-1.2	-9.0	1.2	-0.5	4.2	0.0	2.0	4.4	4.2	29
1.5	-11.9	-7.2	1.3	-3.5	-2.0	-4.3	-3.2	1.7	0.5	0.5	5.2	0.3	-5.7	-3.8	-1.5	4.4	0.5	1.7	2.7	8.2	29
1.6	-12.0	-9.7	1.3	-3.5	-1.3	-5.0	-3.7	1.5	-0.3	0.3	6.2	1.2	-4.3	-3.0	-1.3	4.0	0.7	2.2	2.5	7.9	29
1.6	-13.0	-9.5	2.0	-3.5	-1.8	-4.5	-3.3	1.8	-0.5	0.5	5.9	3.4	-7.0	-3.0	-1.0	4.4	0.7	3.2	5.0	5.0	29

Figure 1. Text file resulting from the extraction of the selected components.

the membership frequency of a random element in a subset. While entropy is a measure of disorder in a data set and it is used to choose the value to maximize information gain. So, let us consider the explanatory variables vector of the epileptic children as follows: $x = (X^{(1)}, \dots, X^{(p)})$ of the explanatory variables is with values in $[0,1]^p$. Let Y us set the random variable of children with epilepsy to be predicted. Consider D_n the set of n -samples of children with epilepsy as follows: $D_n = \{(X_1, Y_1), \dots, (X_n, Y_n)\}$ independent and identically distributed random variables with the same distribution as the couple (X, Y) . Let M us pose a collection of random trees and the estimator associated with these M trees. For the $i^{\text{ème}}$ forest tree, the predicted value at a new x point is noted: $m_n(X, \phi_i, D_n)$ with ϕ_1, \dots, ϕ_M they are independent random variables, variable ϕ distributed and independent of D_n . Thus, the random forest estimator, which results from the aggregation of the different trees is given by the following formula:

$$m_{M,n}(x, \phi_1, \dots, \phi_M, D_n) = \frac{1}{M} \sum_{i=1}^M m_n(X, \phi_i, D_n)$$

Random Forest Algorithm

Input:

D_n : learning sample of children with epilepsy ;
 M : number of trees formed from this training set ;
 C : a given tree ;
 m_n : an estimator from the dataset D_n
 \emptyset : the independent and distributed random variable.

Output:

The random forest

Beginning

For $i = 1 \rightarrow M$ **do**

Beginning

$M^i \leftarrow \text{BootstrapSample}(M)$ //training sets
 $M^i \leftarrow (M^i, D_n)$ //Selection of random subspaces
 Tanque $i < do$ M^i

Beginning

// Selection of the most informative variable K using the Gini index or entropy

$C^i \leftarrow (M^i)$ //construction of the random drill tree
 $i \leftarrow i + 1$;

End ;

$m_{M,n}(x, \emptyset_1, \dots, \emptyset_M, D_n) \leftarrow m_{M,n}(x, \emptyset_1, \dots, \emptyset_M, D_n) \cup \{C^i\}$;

End ;

Return $m_{M,n}(x, \emptyset_1, \dots, \emptyset_M, D_n)$;

End.

5. Simulation and Results

The programming of the neonatal epileptic seizure detection algorithm was done using the python programming language in the Jupyter programming interface with the Scikit-Learn library. The simulation was carried out on the data in **Figure 2** below.

This database contains 1,048,544 observations and 23 variables that we will have subdivided into two different sets. The first is for training 70% and another

set for validation 30%. All training of the models contains the largest number of data which is 733,981 therefore the validation set contains the remaining 314,563 data. After sampling, we trained four (4) models. During the learning we vary the hyperparameters of the models as shown in **Figure 3** below:

Out[2]:

	Time	EEG Fp1-Ref	EEG Fp2-Ref	EEG F3-Ref	EEG F4-Ref	EEG F7-Ref	EEG F8-Ref	EEG Fz-Ref	EEG C3-Ref	EEG C4-Ref	...	EEG T4-Ref	EEG T6-Ref	EEG P3-Ref	EEG P4-Ref	EEG Pz-Ref	EEG O1-Ref	EEG O2-Ref	ECG EKG	Resp Effort	y	
0	1.0	-15.1	-10.5	-0.3	-4.5	-3.2	-5.5	-6.2	-0.2	0.3	...	-4.8	-4.0	-1.8	3.4	0.0	1.2	4.2	-1.8	30.8	0	
1	1.0	-14.7	-8.5	-0.7	-3.3	-4.2	-4.2	-5.0	0.8	0.5	...	-7.9	-1.0	-1.8	4.4	0.0	1.3	4.7	0.5	30.8	0	
2	1.1	-14.4	-8.4	-1.2	-3.3	-5.2	-4.8	-5.0	1.3	0.5	...	-11.2	-0.8	-1.8	3.5	-0.5	0.7	4.9	2.2	30.5	0	
3	1.1	-14.6	-9.9	0.2	-2.8	-3.2	-4.5	-4.2	1.8	1.0	...	-6.2	2.0	-1.0	5.5	0.8	1.8	3.9	1.2	30.5	0	
4	1.1	-15.1	-8.0	-0.3	-5.0	-3.3	-5.0	-4.8	2.2	-0.3	...	-5.7	-5.2	-1.2	3.5	0.7	2.7	3.0	0.8	30.5	0	
...
1048539	2039.2	21.2	12.7	0.9	24.3	2.1	6.7	2.6	3.1	13.9	...	12.8	18.5	13.3	11.4	11.4	8.1	2.9	3.7	5.5	1	
1048540	2039.3	20.9	11.8	-1.2	21.2	2.8	6.3	2.8	3.5	13.9	...	9.3	18.6	13.3	11.8	0.9	7.9	3.1	5.2	5.5	1	
1048541	2039.3	18.2	9.9	-3.2	22.0	2.4	6.0	2.1	3.4	13.4	...	10.7	18.2	13.1	12.1	3.7	7.8	3.4	4.6	4.9	1	
1048542	2039.3	16.3	8.4	-4.9	25.5	2.1	6.0	1.5	2.8	13.4	...	15.4	17.2	13.1	12.2	13.6	7.9	3.7	1.7	4.6	1	
1048543	2039.4	15.4	7.9	-7.2	28.1	2.3	5.8	1.2	2.6	13.6	...	16.6	16.8	13.3	11.9	14.3	8.4	3.7	-0.8	5.0	1	

1048544 rows x 23 columns

Figure 2. Database of children with epilepsy.

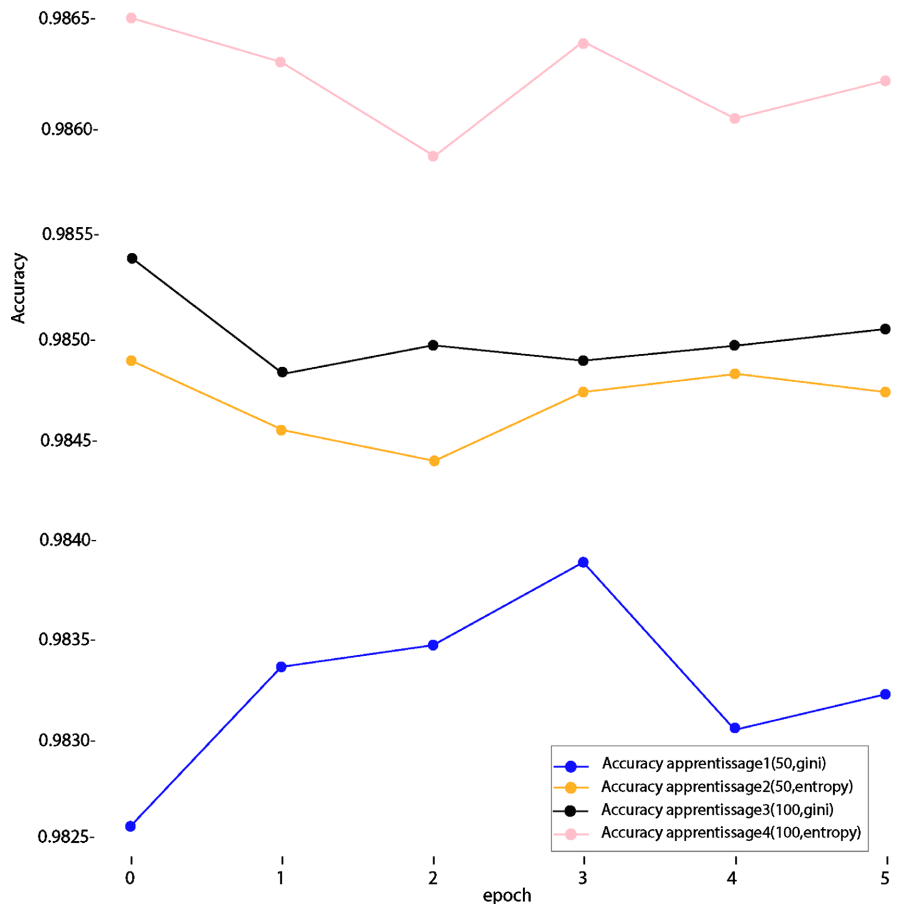


Figure 3. Model results.

We observe in **Figure 3**, the results of our models on the training sample to which we applied a validation 5-fold cross validation assessed by accuracy metric.

Following the training of our machine learning model of supervised classification using our learning sample, we observe that the optimized models obtained have very high performance. Despite the extreme differences in EEG readings between patients having a seizure and not having one. Moreover, by observing the accuracy index, we can attest that they distinguish very well between our two classes which are: presence of epileptic seizure/absence of epileptic seizure. However, we make the following observation:

The variant accuracy of the random forest model using entropy and a number of trees; 100 is 0.986; which is higher than the accuracy of other optimized models. We can then conclude that our best model is the random forest model (100; entropy). We then applied our validation data to our best model and obtained the following performance metrics as results:

```

Matrix de confusion:
[[277706  208]
 [ 4113 32536]]
Rapport de Classification :
      precision    recall  f1-score   support

     0       0.99      1.00      0.99    277914
     1       0.99      0.89      0.94    36649

   accuracy                0.99    314563
  macro avg              0.99      0.94    0.97    314563
 weighted avg              0.99      0.99    0.99    314563

Accuracy : 0.9862634829906887
precision : 0.9936476911800636

```

An accuracy of 0.986; a precision of 0.994 and a recall/specificity of 0.994.

6. Results Interpretation

Our random forests model scores very high on the accuracy, recall, and precision metrics with the test set. With an accuracy of 0.986; an accuracy of 0.994 and a recall of 0.994 in the test, our random forests classifier is a very powerful model.

To corroborate it, we observe the predictions of the confusion matrix in **Figure 4** below, we have:

277,681 true negatives versus 188 false negatives;

32,464 true positives versus 4230 false positives.

It gives us an error of about 1.4% on the validation dataset and means that the model excels at the difference between the two (2) classes (positive and negative samples).

We have created a machine learning model of neonatal epileptic seizure detection based on the Random forests classifier algorithm that can predict whether patients are having a seizure or not through EEG readings. It is correct to 99.4% to predict positive classes in the test set. The following **Figure 5** shows us the performance of the learning accuracy and the validation accuracy:

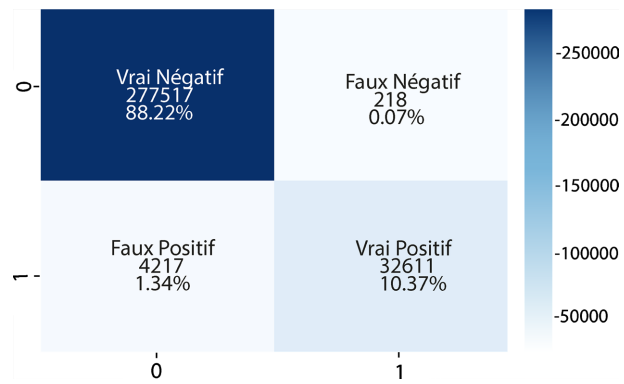


Figure 4. Confusion matrix.

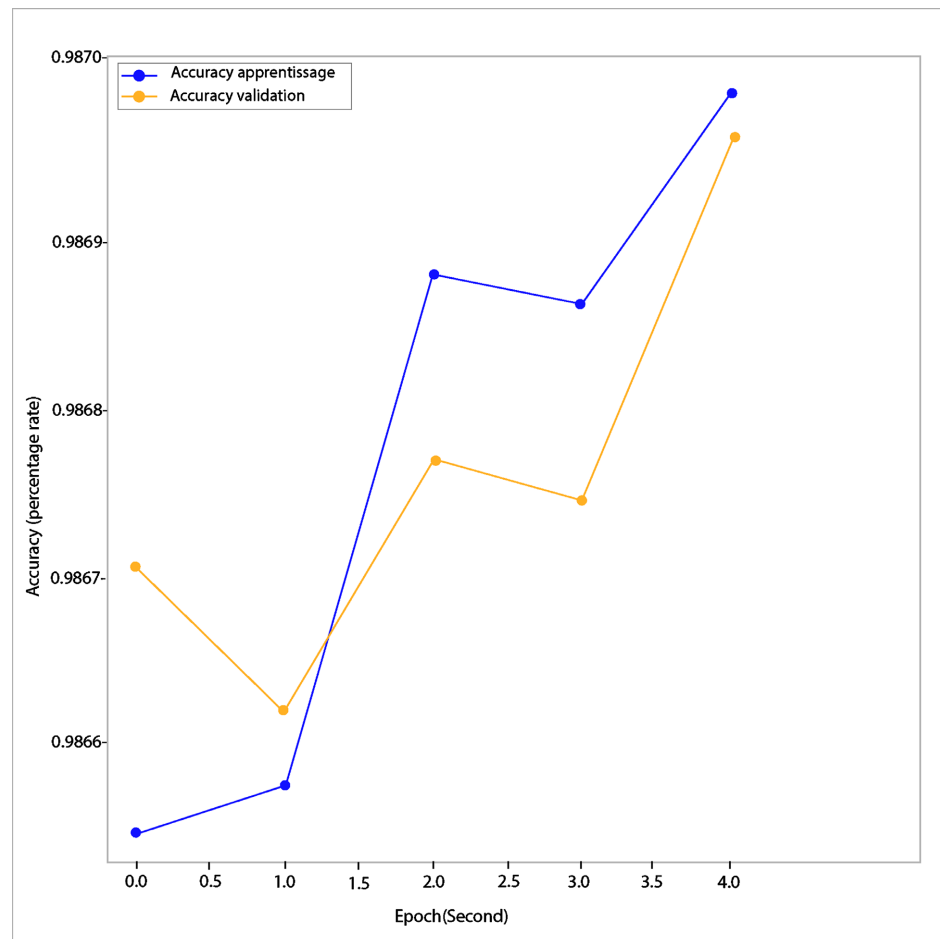


Figure 5. Representation of model performance.

We observed that the two curves are globally increasing. And it means that the model presents a large ability to learn the data characteristics in order to make effective predictions. Also, the small distance between the peaks of the two Accuracy curves demonstrates that the model is neither overfitting nor under-fitting. The model has good generalizability to other data.

The more the model will be used on new data, the more its accuracy will be increased and we will observe a low rate of false prediction.

7. Conclusion

The article presented a model to detect neonatal epileptic seizures. It is a supervised machine learning model combining the extra trees classifier with 5-fold cross validation for the implementation of a very efficient model. The present work allowed us to develop an automatic epilepsy crisis detection model from the electroencephalogram signals amplitudes in human newborns. To realize that work, the EEGlab software was used to preprocess and segment the EEG signals of 10 newborns. And the python programming language allowed us to implement the model. The resulting model, based on the extra trees approach, is able to predict the existence or absence of epilepsy with an accuracy of 99.4%. The use of the said model will greatly reduce the waiting time for the results of examinations and thus improve the time taken to take care of patients.

Acknowledgements

The authors would like to thank the Virtual University of Côte d'Ivoire, the Ivorian Ministry of Health and Dr. ACHIEPO Odilon Yapo Melaine for their contribution to the success of this research work.

Conflicts of Interest

The authors declare no conflicts of interest regarding the publication of this paper.

References

- [1] Kouyate, F. (2022) Antiepileptic Drugs and Reproductive Health about 50 Pregnant Women at the CHU du Point G. Doctoral Thesis, USTTB, Mali, Bamako.
- [2] Schoenen, J. (2022) # Migraine-TOO.
- [3] Bar, C. and Milh, M. (2022) Contribution of Genetics in the Management of Epilepsies of the Child. Improvement in Pediatrics.
- [4] Bennoun, I. (2013) The Psychotherapist's Experience of Children with Epilepsy in a SESSAD. *Contrast*, **38**, 121-141. <https://doi.org/10.3917/cont.038.0121>
- [5] Polepole, F.M., Mukuku, O., Murhula, A.C., *et al.* (2022) Epilepsy Knowledge Assessment among Epileptic Patients in Goma, Democratic Republic of Congo. *Journal of Medicine, Public Health and Policy Research*, **2**, 84-88.
- [6] Tissot, S.A.A.D. (1770) Treatise on Epilepsy. Chez Antoine Chapuis, Lausanne.
- [7] Esambert, B. and Allonneau-Roubertie, E. (2013) Newsletter from the French Foundation for Epilepsy Research.
- [8] Shah, F.P. and Patel, V. (2016) A Review on Feature Selection and Feature Extraction for Text Classification. 2016 *IEEE International Conference on Wireless Communications, Signal Processing and Networking (WiSPNET)*, Chennai, 23-25 March 2016, 2264-2268. <https://doi.org/10.1109/WiSPNET.2016.7566545>
- [9] Patel, V., Tailor, J. and Ganatra, A. (2021) Essentials of Predicting Epileptic Seizures Based on EEG Using Machine Learning: A Review. *Open Biomedical Engineering*, **15**, 90-104. <https://doi.org/10.2174/1874120702115010090>
- [10] Gramacki, A. and Gramacki, J. (2022) A Deep Learning Framework for Epileptic

- seizure Detection Based on Neonatal EEG Signals. *Scientific Reports*, **12**, Article No. 13010. <https://doi.org/10.1038/s41598-022-15830-2>
- [11] Sharmila, A. and Geethanjali, P. (2016) DWT Based Detection of Epileptic Seizure from EEG Signals Using Naive Bayes and k-NN Classifiers. *IEEE Access*, **4**, 7716-7727. <https://doi.org/10.1109/ACCESS.2016.2585661>
- [12] Geethanjali, P. (2015) Fundamentals of Brain Signals and Its Medical Application Using Data Analysis Techniques. In: *Computational Intelligence for Big Data Analysis*, Springer, Cham, 181-197. https://doi.org/10.1007/978-3-319-16598-1_8
- [13] Fergus, P., Hignett, D., Hussain, A., *et al.* (2015) Automatic Epileptic Seizure Detection Using Scalp EEG and Advanced Artificial Intelligence Techniques. *BioMed Research International*, **2015**, Article ID: 986736. <https://doi.org/10.1155/2015/986736>
- [14] Ali Issaka, M. (2016) Analysis of EEG Signals with Parametric Approach for Automatic Detection of Epilepsy Crisis. HAL, Vol. 2016.
- [15] Caparos, M. (2006) Automatic Analysis of Temporal Lobe Epileptic Seizures from Surface EEGs. Doctoral Thesis, National Polytechnic Institute of Lorraine, Lorraine.
- [16] Donos, C., Dümpelmann, M. and Schulze-Bonhage, A. (2015) Early Seizure Detection Algorithm Based on Intracranial EEG and Random Forest Classification. *International Journal of Neural Systems*, **25**, Article ID: 1550023. <https://doi.org/10.1142/S0129065715500239>
- [17] Yu, J. (2019) Epileptic Seizure Classification ML Algorithms.