Random convolution kernels applied to severity assessment of post-anoxic hypoxic-ischemic encephalopathy based on EEG signal values

1st Vitor dos Santos Silva
École Centrale de Lille
Villeneuve d'Ascq, France

2nd Sarah Ben Othman
École Centrale de Lille
Villeneuve d'Ascq, France

3rd Laure Lacan
CHRU Lille
Lille, France

4th Julien De Jonckheere
CHRU Lille
Lille, France

5th Slim Hammadi
École Centrale de Lille
Villeneuve d'Ascq, France

***Abstract*—Perinatal anoxia, which is the deprivation of oxygen during the birthing process, can arise due to various factors and, if prompt intervention is not administered, can lead to hypoxic-ischemic encephalopathy, a leading cause of neonatal death. However, this can be averted if quickly identified, by submitting the newborn to controlled hypothermia. This work aims to assess the level of severity of the encephalopathy and the need of hypothermia through time series classification. Random convolution kernels were applied to 105 EEG signal values, and were used to generate the inputs for a ridge classifier using the ROCKET algorithm. The final accuracy of 80% was evaluated on 45 EEGs, which is similar to methods using feature extraction. One potential factor contributing to a lower accuracy could be the manual attribution of classes, which was executed by a few doctors and does not follow objective and quantitative criteria. This could be improved upon by employing many doctors for diagnostic agreement in order to mitigate personal bias.**

***Index Terms*—EEG, neonatal, perinatal anoxia, ROCKET, artificial intelligence**

I. Introduction

Lack of oxygen in the brain during the birthing process, known as perinatal anoxia, if left untreated beyond the first 6 hours of life, may potentially lead to hypoxic-ischemic encephalopathy (HIE), a leading cause of neonatal mortality [1]. However, recent research, notably the study conducted by [2], demonstrated that the severity of the injury caused by perinatal anoxia can be evaluated through quantitave analysis of EEGs recorded during the first few hours after birth, which can be divided in three levels according to the French classification system for HIE [3]. Nevertheless, the interpretation of EEGs typically requires the expertise of a specialized neurologist, which may not be readily available in every hospital. Thus, the necessity of automated analysis arises.

There are a number of ways to make use of EEGs in the literature, the most common being feature extraction, followed by using raw signals [4]. Building upon the previous study, conducted by [2], which calculated metrics from the EEGs and fed them into an SVM model to obtain the predictions, this work shifts the focus, and exploits the raw signals from the EEGs themselves. To make use of these raw signals, they are transformed by employing random convolution kernels (ROCKET), a method developed by [5], which has demonstrated being efficient in scenarios with a small amount of data and low computational power.

II. Methods

A. Dataset

The data used in this study, which corresponds to the same dataset utilized in [2], was obtained at the hospital "CHU de Lille", upon the completion of all necessary ethical procedures. The dataset comprises of 150 EEGs, with 59 classified as level 1, 48 classified as level 2 and 43 classified as level 3, according to the aforementioned classification system.

Each EEG file contains signal recording of eight electrodes, namely F4, F3, C3, C4, T3, T4, O1 and O2. The recordings were captured at a sampling rate of 256Hz and varied in duration. In order to normalize it, the procedure was to divide the signal into windows. Different window lengths were experimented with, and it was determined that 3-minute windows provided the best results. Subsequently, the EEG files were divided into train and test sets for model development and evaluation.

Provided the fact that each EEG file produces a different number of windows (due to the variable length of the recording), the following systematic procedure was adopted to avoid data leakage:

1) Divide all EEG files into three stacks, of severity 1, 2 and 3
2) For each stack:
    a) Organize stack in a descending order by length of recording
    b) Pop the first item and add it to the train or test pool, whichever drives the train/test proportion closer to the desired value.
    c) Repeat until current stack is empty.
3) Merge train pool for all the severity values. Do the same with test pool.
4) Balance each pool to have the same amount of each severity in it.

Following this procedure, the train set consisted of 105 EEG files, which correspond to a total of 1473 time windows (491 windows per severity group). Similarly, the test set consisted of 45 EEG files, yielding a total of 636 time windows (212 windows per severity group).

B. Model

Differently from prior approaches that computed specific signal metrics, such as averages or amplitudes, this work takes a different approach, by directly inputting the raw EEG signal into the model. The time series data is then transformed by the ROCKET method, introduced by [5], which generates random convolution kernels. These kernels are convolved with the time series, resulting in new feature vectors. These generated features were used to train a ridge classifier, as recommended by [5]. To accommodate the multi-electrode nature of the EEG data, a separate model was trained for each electrode, yielding 8 models. The responses of these models are aggregated to provide one final result for each file.

III. Results

A. Results on individual time series

To assess the accuracy of each model, the results obtained are summarized in Table I:

TABLE I: Accuracy for each electrode

Electrode	Accuracy
F3	58.81%
F4	59.52%
T3	66.43%
C3	66.67%
C4	58.34%
T4	63.45%
O1	68.81%
O2	67.86%
Mean	**63.73%**

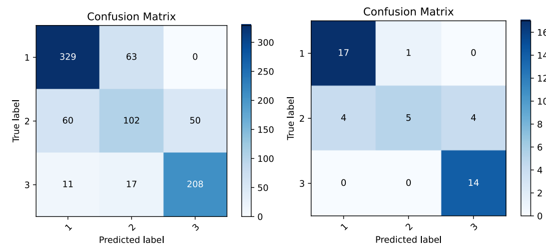
However, the goal of training multiple models lies in the aggregation of their responses. By applying the mode of the responses for each time window, the accuracy increases to 76.07%. The confusion matrix for the individual time series is displayed in Figure 1a.

B. Results on files

However, the primary objective is to evaluate the overall level of an entire EEG file, rather than the individual windows. Therefore, a histogram is calculated considering all the windows within each file, and the most prevalent class is considered the final prediction for the file. The accuracy evaluated for the 45 test files is 80%. The confusion matrix can be seen in Figure 1b, and an example of the results produced by the model can be seen on Figure 2.

IV. Conclusion and discussion

Based on the results achieved, we can verify that predicting the level of severity for individual windows and combining the responses from each electrode results in notable benefits, and increases the accuracy by approximately 12%. Additionally,



(a) Ridge classifier (b) Test files

Fig. 1: Confusion matrices

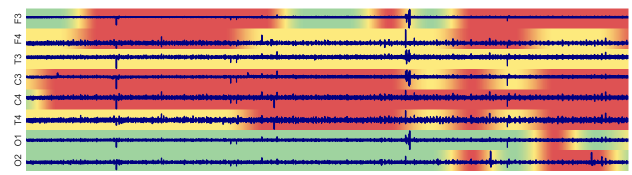


Fig. 2: Example of output from model

when evaluating the response based on the histogram generated for the files, there is further improvement of around 4% in accuracy. The accuracy is the same as the one obtained by [2], however, the recall is higher for classes 1 and 3, and the precision is higher for class 2. The F1 score for class 3 is also higher. Furthermore, dividing the raw signal into windows renders the model output more explainable, which is vital in a health environment, and the processing of raw signal data diminishes the need of third-party software for metrics calculation.

It is worth noting that all the errors in prediction are concentrated within the second column or second row of the confusion matrix, indicating a particular issue with the second class of severity. This issue has been previously addressed by [2], suggesting that one of the causes may be the manual evaluation performed by doctors. This may be counter measured by obtaining the diagnostic of different doctors for the dataset, as an attempt to reduce personal bias.

References

[1] V. Pierrat, N. Haouari, A. Liska, D. Thomas, D. Subtil, P. Truffert *et al.*, "Prevalence, causes, and outcome at 2 years of age of newborn encephalopathy: population based study," *Archives of Disease in Childhood-Fetal and Neonatal Edition*, vol. 90, no. 3, pp. F257–FF261, 2005.

[2] L. Lacan, "Analyse quantitative et automatisée des eeg néonataux post-anoxiques: développement d'un outil clinique d'aide au diagnostic précoce de l'encéphalopathie anoxo-ischémique néonatale," Ph.D. dissertation, Université de Lille, 2021.

[3] M. Chevallier, A. Ego, C. Cans, T. Debillon, and French Society of Neonatology, "Adherence to hypothermia guidelines: a french multicenter study of fullterm neonates," *PLoS One*, vol. 8, no. 12, p. e83742, Dec. 2013.

[4] A. Craik, Y. He, and J. L. Contreras-Vidal, "Deep learning for electroencephalogram (eeg) classification tasks: a review," *Journal of neural engineering*, vol. 16, no. 3, p. 031001, 2019.

[5] A. Dempster, F. Petitjean, and G. I. Webb, "Rocket: exceptionally fast and accurate time series classification using random convolutional kernels," *Data Mining and Knowledge Discovery*, vol. 34, no. 5, pp. 1454–1495, 2020.