



HAL
open science

Multi-label neonatal EEG segmentation enriched with semi-synthetic data

Vivien Kraus, Guillaume Dollé, Nathalie Bednarek, Gauthier Loron, Jonathan Beck, François Rousseau, Nicolas Passat

► **To cite this version:**

Vivien Kraus, Guillaume Dollé, Nathalie Bednarek, Gauthier Loron, Jonathan Beck, et al.. Multi-label neonatal EEG segmentation enriched with semi-synthetic data. International Symposium on Biomedical Imaging (ISBI), 2025, Huston, United States. hal-04863223

HAL Id: hal-04863223

<https://hal.science/hal-04863223v1>

Submitted on 7 Jan 2025

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

MULTI-LABEL NEONATAL EEG SEGMENTATION ENRICHED WITH SEMI-SYNTHETIC DATA

V. Kraus¹, G. Dollé², N. Bednarek^{1,3}, G. Loron^{1,3}, J. Beck³, F. Rousseau⁴, N. Passat¹

¹ Université de Reims Champagne Ardenne, CRESTIC, Reims, France

² Université de Reims Champagne Ardenne, CNRS, LMR, UMR 9008, Reims, France

³ Service de médecine néonatale et réanimation pédiatrique, CHU de Reims, France

⁴ IMT Atlantique, LaTIM INSERM U1101, 29238 Brest, France

ABSTRACT

Electroencephalography (EEG) is widely used in routine clinical practice. In particular, it is of the utmost importance for monitoring newborns suffering from acute neonatal encephalopathy. In such a pathological context, continuous acquisitions during the first days of life must be analyzed, which represents an insurmountable workload for clinicians. Machine learning approaches are a way of helping human experts in their work. More specifically, deep learning models have the ability to segment specific physiological and pathological patterns. The main obstacle to the development of these models is their correct training from a sufficient number of annotations. Indeed, building annotated EEG datasets involves a time cost that is generally unacceptable. In this paper, a deep learning model dedicated to the segmentation of multiple patterns in newborn EEG is proposed. We study the impact of training this model either with only human-defined annotations on the real EEG, or with augmented datasets that also include synthetic patterns embedded in the real signal, endowed with their annotations. The mixed use of semi-synthetic and real data leads to insights about future development of effective AI-based assistive tools for newborn EEG assessment.

Index Terms— Electroencephalography, Deep-learning, Synthetic patterns, Data augmentation, Newborn brain

1. INTRODUCTION

Assessment of brain function with Electroencephalography (EEG) is routine practice in the management of newborns with acute neonatal encephalopathy. Electrophysiology provides critical diagnostic, severity and prognostic information in infants exposed to a deleterious cascade of brain lesions following anoxic-ischemic insult. EEG monitoring should be offered continuously to these newborns for at least the first four days of life. Reliable information on brain function is needed as early as the first six hours of life in order to provide neuroprotection with 72 hours of total-body, controlled, and mild hypothermia. During therapeutic hypothermia, brain function should be continuously monitored.

However, the amount of data recorded by EEG over several days is overwhelming. Expert electrophysiologists available to evaluate these traces are scarce and cannot interpret the day-to-day record as it should be. Numerical tools may help interpret these traces by computing quantitative metrics. The continuous evolution of these quantitative metrics over the monitored period may greatly reduce the time and efforts spent by electrophysiologist and intensivist in managing these babies.

Deep Learning (DL) provides efficient solutions for EEG analysis in various clinical applications (see Sect. 2). Beyond the development of DL models designed with respect to the technical constraints and properties of EEG acquisition, the main challenge, in the case of pattern detection and segmentation tasks in EEG is related to the training of these models. Indeed, these patterns are often specific to a given clinical application. As a consequence, most often, there does not exist datasets endowed with relevant annotations. Despite the availability of EEG annotation tools [1, 2, 3], the construction of annotated datasets remains a tedious and time consuming effort, leading to small corpus of annotations compared to the requirements of model training. A way to tackle this issue is to enrich the set of manual annotations with computer-defined ones. To this end, a strategy may consist to build synthetic patterns, associated to annotations by construction. For the sake of coherence, such synthetic patterns may be embedded in real EEG.

Considering this approach, our contributions are twofold. First, we propose a DL architecture, based on U-Net [4], designed to process EEG acquired in routine, under the technical and clinical constraints of acquisitions in the case of neonatal encephalopathy. In particular, this architecture is designed for multilabel segmentation, i.e. the simultaneous analysis of various kinds of distinct patterns. Second, we investigate the impact of training this model with human-defined annotations, obtained in real conditions (leading to scarce, incomplete) sets of annotated datasets, vs. training with additional semi-synthetic data, composed of real EEG enriched with synthetic patterns, endowed with associated annotations.

This article is organized as follows. In Sect. 2, we present related work on DL models for EEG, neonatal EEG analysis and EEG data augmentation. In Sect. 3, we describe the patient data and their annotations, and we present our strategy for creating semi-synthetic data, obtained by enriching real EEG with synthetic patterns. In Sect. 4, we describe our DL model, adapted from U-Net for multilabel pattern segmentation from neonatal EEG. In Sect. 5 we propose experimental results. In Sect. 6 we summarize our contributions and discuss perspectives.

2. RELATED WORKS

In the medical setting, EEG signal processing is used for sleep stage classification or to diagnose conditions such as epilepsy [5, 6]. It is also used with newborn patients, as a tool to assess the post-menstrual age [7] or study the outcome of prematurity [8]. For term patients, EEG is also useful to diagnose rare conditions, and especially Hypoxic Ischemic Encephalopathy (HIE) [9]. While simpler tools such as amplitude-EEG [10] can already present relevant infor-

mation about the patient, the full signal remains useful for specialized tasks. The neonatal seizure detection is a difficult task that often requires the full EEG [11, 12, 13, 14, 15]. The additional complexity of neonatal sleep stage classification [16] emphasizes the need to detect bursts and inter-burst periods in the signal [17]. For rare conditions, specific EEG patterns may be present in the signal. When these patterns are not well understood, EEG processing may also consist of classifying them [9].

The EEG biosignal is often pre-processed with appropriate filtering in a referenced setting. Complexity metrics are computed over loosely overlapping windows of the signal, and the low-frequency evolution of these metrics helps perform the detection task. For the detection of physiological and pathological figures of HIE, this setting is unfortunately not precise enough. EEG processing is also sometimes pre-processed with a normalization step [18], however this normalization is not suitable for neonatal EEG as it may blur the difference between bursts and inter-bursts. The signal amplitude also depends on the patient’s age, which correlates with other relevant features of the signal related to brain maturation.

Unlike raw sensor data, processing aggregated features can be done efficiently by a machine at the patient’s bed, with post-processing improving the results [12]. Deep learning architectures are being tried [16], such as LSTM [14], Graph Neural Networks with attention [5]. If temporal insights are requested, Convolutional Neural Networks may process the signal as a time series [19].

3. REAL VS. SEMI-SYNTHETIC DATA

3.1. Patient data

The patient data comes from the Lytonepal cohort, where the EEGs were recorded in the first 6 hours of life. Nine of these EEGs have been reviewed by an expert who has annotated some occurrences of Front Sharp Transients, Anterior Slow Dysrhythmia and Spikes. The expert read the signal in a standard bipolar referenced setting.

The clinician was asked to record the onset and duration of the observed figures, as well as the derivations that supported the observation. The location of these figures are thus expressed as bipolar references and not raw electrodes.

During the annotation phase, many instances were found to be ambiguous. Since exhaustivity would not be feasible, the clinician was asked to skip ambiguous sections of the signal, and annotate as many occurrences as possible in the less ambiguous sections.

3.2. Semi-synthetic data

Since the annotated data is scarce, more data has been generated on the signal. The raw signal has been split into non-overlapping 10-second windows, and random instances of a diversity of the most precisely described physiological and pathological figures have been added to it at random. The signal is linearly attenuated in the first and last seconds of the window, so that all channels artificially start and end at 0 μV . Non-overlapping windows ensure that no validation input has previously been seen during learning, even in part. The 9 annotated signals plus 10 additional non-annotated signals have been chosen as support for synthetic data. This limit has been chosen so that learning can be achieved with 32 GB of memory.

Since the signal has been selected from the first 6 hours of life of the patient, it contains many burst and inter-burst periods. In order to add a synthetic figure to the signal, the first step is to decide whether the segment is active enough. The average absolute amplitude over the 10 seconds is computed. Channels that are always 0

(due to missing electrodes) are not considered in the following. If the channel with the largest amplitude has an amplitude 5 times as high as the channel with the least amplitude, then the window is rejected on suspicion of a high-voltage artifact. Otherwise, the window is rejected if the average amplitude over the non-zero channels is less than 3.6 μV , the average amplitude of the 10% most active windows.

For each selected active windows, each of the 5 selected figure types are independently selected with a 25% chance. If a figure is selected, it is added to the signal, and its time support is recorded.

The selected figure types are:

- Slow Continuous (physiological)
- Front Sharp Transient (physiological)
- Anterior Slow Dysrhythmia (physiological)
- Positive Temporal Spike (unknown)
- Rolandic Positive Spike (type B) (unknown)

These figures have been selected for the precision of their description in [20], in terms of location, amplitude, duration or frequency, and smoothness of the spikes, for term neonates. Instances of these figures are meant to represent some of the knowledge of the clinicians and need not be perfectly realistic.

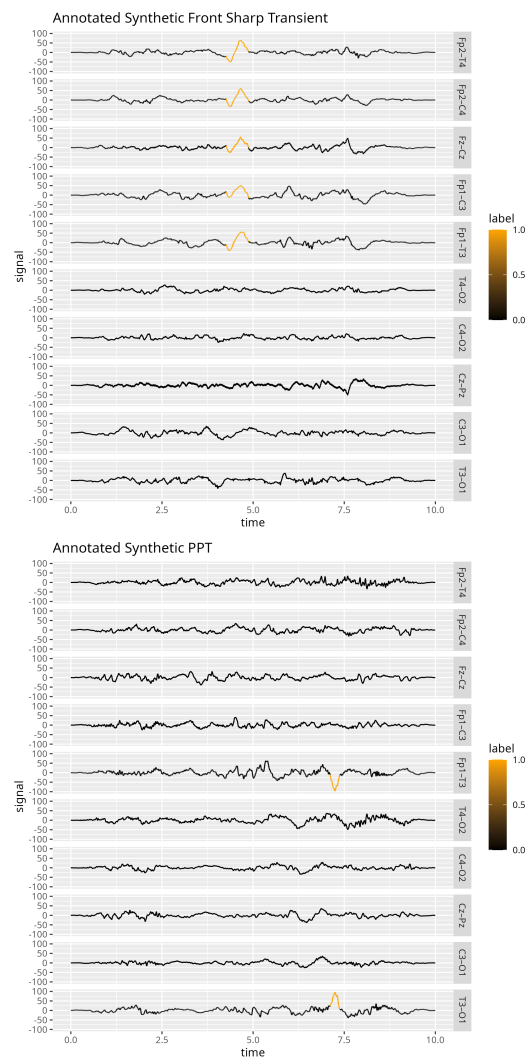


Fig. 1: Instances of synthetic Front Sharp Transient and Positive Temporal Spike

For each instance of a figure, the temporal onset, the spatial location, and the shape parameters are sampled uniformly in the prescribed range, except for the amplitude which favors the lowest bound of the range.

The figure detector is expected to understand some basic physiological facts about the EEG signal:

- some figures can only be characterized in some regions of the brain (frontal, parietal, temporal, occipital);
- figures occurring on a specific electrode may be visible in all bipolar references involving the specific electrode;
- some figures may occur on left or right electrodes with similar characteristics;
- the characterization of figures depends on observed amplitude and frequency or duration.

4. MULTI-CHANNEL MULTI-TARGET SEGMENTATION OF EEG

The learning task aims to simultaneously learn the segmentation of the 5 synthetic labels, and of the 3 real labels. The input signals are the 10-second windows (sampled at 256 Hz) with added synthetic figures, and the 8 labels are the binary presence or absence of each figure at a given time point in a given channel. As an exception, due to the ambiguous problem of real annotations, learning of each of the 3 real labels is disabled for that whole input if the 10-second window contains no instance of any of the figure types of interest.

More specifically, the input is a tensor of dimension $N \times 2560 \times 10$, where N is the size of the training set. Out of the 6320 unique 10-second signal windows, 4424 have been selected for training and the 1896 remaining serve as validation. Each instance of the training set is the 2560 observations of the 10 channels. The output of the network is a tensor of dimension $N \times 2560 \times 80$. For each instance, for each time point, for each channel and for each figure (synthetic or real), a binary score indicates whether a figure of this type is present at this point on this channel.

The architecture that was chosen is a one-dimensional CNN arranged in a U-Net [4] shape (Fig. 2). It features simple layers:

- 1-D convolutions: the input is padded so that output and input tensors have the same shape;
- ReLU non-linearity layers;
- temporal maximum pooling in the downward branch, and temporal upsampling (in which every observation is repeated twice).

It is optimized with respect to the binary cross-entropy across all time points and all labels on all channels. The optimizer is the simple gradient descent algorithm with no moment term. At each epoch, after each training instance is considered (the order is changed at each epoch), the network weights are updated.

5. EXPERIMENTS AND RESULTS

Two parameters of the network are tuned based on the F-score on the validation dataset. First, the learning rate is tuned so as to maximize the final F-score (computed as the average of the F-score in the last 50 epochs) of the synthetic label yielding the minimal F-score, considering the maximum number of synthetic data. Then, the number of generated synthetic annotations is reduced from the maximum that can be added with our proposed approach down to 0. In the latter case, the synthetic data cannot influence the learning of the real labels, so this case shows the effect of learning the real labels directly. In the former case, the real labels are learnt along with the synthetic labels, but the profusion of synthetic labels may distort the signal so much that the real labels cannot be learnt meaningfully.

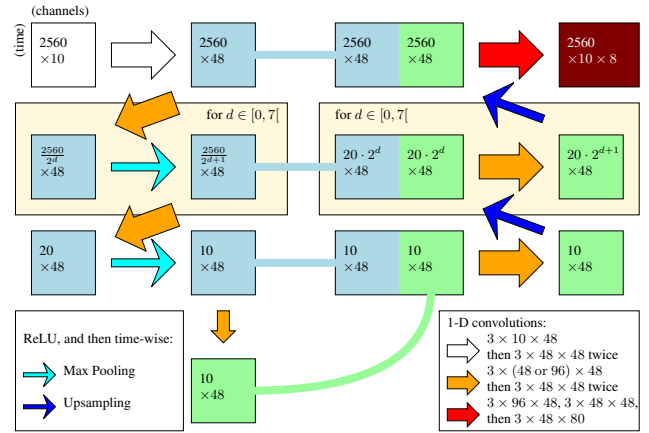


Fig. 2: Simple CNN-based U-net for simultaneous learning of 5 synthetic and 3 real labels

The value for the learning rate is 10, for a final F-score of the Synthetic Anterior Slow Dysrhythmia of 4% (which is the label with the worse F-score). The worst synthetic F-score drops to 2.8% (respectively 2.7%) for a learning rate of 20 (respectively 5).

Due to the small number of annotated Anterior Slow Dysrhythmia, the performance of the model (Fig. 3) is poor, regardless of the amount of synthetic data introduced. Adding synthetic data does not have a great impact on the performance for the Front Sharp Transient. It has an effect on the performance for the detection of Spikes though. In any case, we observe for all three real labels that the performance is not maximal with 0% synthetic data, nor 100% synthetic data. Learning the real labels is made easier by introducing a relatively small amount of synthetic data.

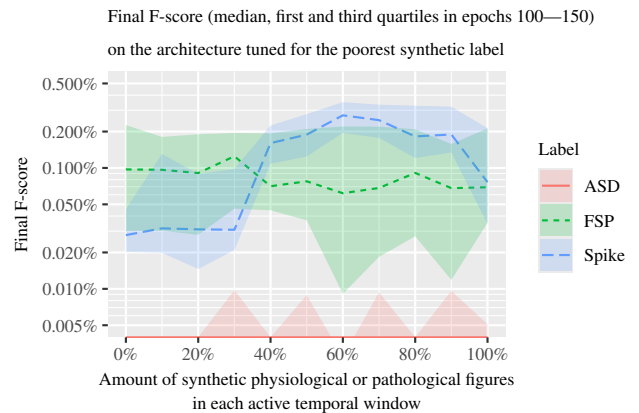


Fig. 3: Varying the synthetic data frequency

The shape of the Spike is easier to model, explaining the large effect of synthetic data, and it is also more visible even with an added synthetic wave, explaining why the effect lasts with a large amount of synthetic data. In contrast, the Front Sharp Transient is a complex wave susceptible to being deformed by other synthetic data, which challenges our understanding of its shape.

6. CONCLUSION

In this work, we presented a way to generate synthetic annotations on EEG signals to help learn the detection of physiological or pathological figures of the neonatal EEG. The synthetic labels and the real labels are learnt simultaneously. Adding some synthetic data improves the detection F-score, especially for the Spikes which are easy to model, and less so for the Front Sharp Transient which is more complex. If too much synthetic data is generated, the detection of the real labels is made more difficult by the synthetic wave to the signal, especially for figures with a longer time support.

Further development of the simultaneous segmentation of real and synthetic data should make the synthetic data more physiologically correct, in particular for the Front Sharp Transient. The effect of each of the 5 synthetic annotation types on each of the 3 real labels needs to be studied more closely.

7. COMPLIANCE WITH ETHICAL STANDARDS

This study was performed in line with the principles of the Declaration of Helsinki. Approval was granted by the French *Commission Nationale de l'Informatique et des Libertés* (CNIL) with number DR-2015-136 and by the *Comité Consultatif sur le Traitement de l'Information en matière de Recherche dans le domaine de la Santé* (CCTIRS) the 20 November 2014, and by the *Comité de Protection des Personnes* (CPP) Sud Est V the 9 July 2014.

8. ACKNOWLEDGMENTS

This work was supported by the French *Agence Nationale de la Recherche* (grants ANR-19-CHIA-0015, ANR-22-CE45-0034) and by the American Memorial Hospital Foundation.

9. REFERENCES

- [1] V Kraus, G Dollé, M Alloux, G Loron, N Bednarek, J Beck, F Rousseau, and N Passat, "SLAM - A thin-client for interoperable annotation and biomedical signal handling," *SoftwareX*, vol. 27, pp. 101795, 2024.
- [2] R Weiler, M Diachenko, E L Juarez-Martinez, A-E Avramiea, P Bloem, and K Linkenkaer-Hansen, "Robin's Viewer: Using deep-learning predictions to assist EEG annotation," *Front Neuroinform*, vol. 16, 2023.
- [3] M Beier, T Penzel, and D Krefting, "A performant web-based visualization, assessment, and collaboration tool for multidimensional biosignals," *Front Neuroinform*, vol. 13, 2019.
- [4] O Ronneberger, P Fischer, and T Brox, "U-Net: Convolutional networks for biomedical image segmentation," in *MICCAI*, 2015, pp. 234–241.
- [5] S Mazurek, R Blanco, J Falcó-Roget, and A Crimi, "Explainable graph neural networks for EEG classification and seizure detection in epileptic patients," in *ISBI*, 2024, pp. 1–5.
- [6] M McCumber, K Tyner, S Das, W C Stacey, G C Smith, M Alfatlawi, and S V Gliske, "Seizure onset localization from ictal intracranial EEG data using online dynamic mode decomposition," in *ISBI*, 2023, pp. 1–4.
- [7] M Gschwandtner, M M Hartmann, L Oberdorfer, F Fürbass, K Klebermaß-Schrehof, T Werther, N J Stevenson, G Gritsch, H Perko, A Berger, T Kluge, and V Giordano, "Deep learning for estimation of functional brain maturation from EEG of premature neonates," in *EMBC*, 2020, pp. 104–107.
- [8] Y Hajjar, M El-Sayed, A E S Al Hajjar, and B Daya, "Correlation analysis between EEG parameters to enhance the performance of intelligent predictive models for the neonatal newborn sick effects," in *ICIST*, 2019, pp. 1–5.
- [9] B M Murphy, *Data mining and machine learning techniques for neonatal EEG event recall*, Ph.D. thesis, University College Cork, Republic of Ireland, 2019.
- [10] Z A Vesoulis, P G Gamble, S Jain, N M El Ters, S M Liao, and A M Mathur, "WU-NEAT: A clinically validated, open-source MATLAB toolbox for limited-channel neonatal EEG analysis," *Comput Methods Programs Biomed*, vol. 196, pp. 105716, 2020.
- [11] A Hossein Ansari, P J Cherian, A Caicedo Dorado, K Jansen, A Dereymaeker, L De Wispelaere, C Dielman, J Vervisch, P Govaert, M De Vos, G Naulaers, and S Van Huffel, "Weighted performance metrics for automatic neonatal seizure detection using multiscored EEG data," *IEEE J Biomed Health Informatics*, vol. 22, pp. 1114–1123, 2018.
- [12] N J Stevenson, K T Tapani, and S Vanhatalo, "Hybrid neonatal EEG seizure detection algorithms achieve the benchmark of visual interpretation of the human expert," in *EMBC*, 2019, pp. 5991–5994.
- [13] K T Tapani, S Vanhatalo, and N J Stevenson, "Time-varying EEG correlations improve automated neonatal seizure detection," *Int J Neural Syst*, vol. 29, pp. 1–15, 2019.
- [14] M U Abbasi, A Rashad, A Basalamah, and M Tariq, "Detection of epilepsy seizures in neo-natal EEG using LSTM architecture," *IEEE Access*, vol. 7, pp. 179074–179085, 2019.
- [15] A O'Shea, G Lightbody, G B Boylan, and A Temko, "Neonatal seizure detection from raw multi-channel EEG using a fully convolutional architecture," *Neural Networks*, vol. 123, pp. 12–25, 2020.
- [16] S F Abbasi, J Ahmad, A Tahir, M Awais, C Chen, M Irfan, H A Siddiq, A B Waqas, X Long, B Yin, S Akbarzadeh, C Lu, L Wang, and W Chen, "EEG-based neonatal sleep-wake classification using multilayer perceptron neural network," *IEEE Access*, vol. 8, pp. 183025–183034, 2020.
- [17] S A Raurale, G B Boylan, G Lightbody, and J M O'Toole, "Identifying tracé alternant activity in neonatal EEG using an inter-burst detection approach," in *EMBC*, 2020, pp. 5984–5987.
- [18] Y Wang, Z Huang, B McCane, and P Neo, "Emotionet: A 3-d convolutional neural network for EEG-based emotion recognition," in *IJCNN*, 2018, pp. 1–7.
- [19] S Zhang, Z Pei, H Mou, W Yang, Q Li, and X Wu, "Visual explanations of deep convolutional neural network for EEG brain fingerprint," in *ISBI*, 2024, pp. 1–5.
- [20] M D Lamblin, M André, M J Challamel, L Curzi-Dascalova, A M d'Allest, E . De Giovanni, F Moussalli-Salefranque, Y Navelet, P Plouin, M F Radvanyi-Bouvet, D Samson-Dollfus, and M F Vecchierini-Blineau, "Électroencéphalographie du nouveau-né prématuré et à terme. Aspects maturatifs et glossaire," *Clinical Neurophysiology*, vol. 29, pp. 123–219, 1999.