Computer Methods and Programs in Biomedicine 285 (2026) 109552

Contents lists available at ScienceDirect

Computer Methods and Programs in Biomedicine

journal homepage: www.sciencedirect.com/journal/computer-methods-and-programs-in-biomedicine

Forecasting and real-time detection of neonatal seizures: A machine learning perspective

Tamara Skoric [a,*], Marija Djermanovic [b,c], John M. O'Toole [d,e], Maarten De Vos [f], Slobodan Spasojevic [b,c]

[a] University of Novi Sad, Faculty of Technical Sciences, Trg Dositeja Obradovica 6, Novi Sad, 21000, Serbia
[b] University of Novi Sad, Faculty of Medicine , Hajduk Veljkova 3, Novi Sad, 21000, Serbia
[c] Institute for Child and Youth Health Care of Vojvodina, Hajduk Veljkova 10, Novi Sad, 21000, Serbia
[d] CergenX Ltd, 23 Green Park, Orwell Road, Rathgar, D06 YX20, Dublin, Ireland
[e] INFANT Research Centre, University College Cork, Wilton, T12 YE02, Cork, Ireland
[f] KU Leuven, Department of Electrical Engineering, Department of Development and Regeneration, Kasteelpark Arenberg 10, Leuven, 3001, Belgium

ARTICLE INFO

Keywords:
Neonatal seizure forecasting
Neonatal seizure detection
Single-channel EEG
Machine-learning model
Simultaneous EEG/ECG features extraction

ABSTRACT

Background and Objectives: Recognition and treatment of neonatal seizures, primarily diagnosed using EEG, are essential to protect the developing brain, yet only 11% of seizures are treated within 1 hour of onset. Developing seizure forecasting and detection systems could help address this issue in intensive care settings with limited neurologist availability.

Methods: We propose an ML architecture with dual functionality: (1) forecasting seizures within short prediction intervals (5, 10, and 15 min), and (2) detecting seizures in real time, using explainable entropy-, singular value decomposition-, power spectrum-, and statistical moment-based features with an AdaBoost classifier. The same feature set supports both functionalities, with forecast reliability improving when features are derived from longer EEG intervals. The model was trained and tested on a single-channel (P3–P4, n = 82) dataset and a multi-channel public dataset (n = 79).

Results: The proposed ML model was able to forecast seizures at a 15-min prediction interval (86.6%/92.4% sensitivity/specificity; Matthews correlation coefficient, MCC = 0.39), with better performance at the shorter 5-min prediction interval (95.6%/88.9% sensitivity/specificity; MCC = 0.55) for the single-channel EEG seizure group. Short-range forecasting outperformed the pioneering study on the public multi-channel EEG dataset (MCC = 0.48). The proposed ML model achieved over a 30% improvement in MCC and Pearson's coefficient on the single-channel dataset compared to the state-of-the-art deep-learning architecture (*ConvNeXt*).

Conclusion: EEG segmentation and appropriate feature selection enhance ML model performance for seizure detection in small datasets. Validation of feature-based ML models is required on both single and multi-channel EEG, given that quantitative features can be sensitive to different bipolar channel pairs. Although the method has outperformed the pioneering studies in seizure forecasting, further improvement in performance and more extensive validation are needed for clinical adoption.

1. Introduction

Neonatal seizures are the most common clinical sign of neurological dysfunction in newborns and require prompt treatment to prevent long-term damage. The main cause of seizures is a hypoxic-ischaemic insult leading to hypoxic-ischaemic encephalopathy (HIE), followed by transient metabolic disturbances, central nervous system infection, stroke, and genetic disorders [1]. Unlike seizures in adults and children, clinical manifestations in neonates are often absent or subtle, so visual electroencephalography (EEG) interpretation remains the gold standard for seizure detection and forecasting and relies heavily on expert availability. A European multicentre study showed that only 11% of seizures are treated within 1 hour of onset [2]. To address the limited availability of experts and the time-consuming nature of EEG interpretation,

* Corresponding author.
E-mail addresses: ceranic@uns.ac.rs (T. Skoric), marija.djermanovic@mf.uns.ac.rs (M. Djermanovic), jotoole@cergenx.com (J.M. O'Toole), maarten.devos@kuleuven.be (M. De Vos), slobodan.spasojevic@mf.uns.ac.rs (S. Spasojevic).

https://doi.org/10.1016/j.cmpb.2026.109552
Received 8 November 2025; Received in revised form 10 July 2026; Accepted 16 July 2026
Available online 16 July 2026
0169-2607/© 2026 Elsevier B.V. All rights are reserved, including those for text and data mining, AI training, and similar technologies.

machine learning (ML) algorithms have been developed to enable automated neonatal seizure detection, as part of a growing trend towards developing ML models as support tools in neonatal intensive care units (NICU) [3]. The integration of ML into NICU workflows is hampered by more complex processes such as collecting sufficiently large and diverse datasets to represent expected population heterogeneity, addressing potential bias, and its potential negative impact on outcome and clinical evaluation through rigorous validation [4]. Additionally, the lack of ground-truth data annotation is a limitation due to the variations in EEG interpretation among top experts.

A pioneering study on seizure detection used a heuristic approach for seizure detection (validated on 26 infants) [5], while subsequent research has focused on developing algorithms based on ML models [6–9], ranging from simple Support Vector Machine (SVM) classifiers to deep-learning convolutional neural network (CNN) classifiers. These approaches were also validated on small datasets, ranging from 17 to 79 infants, which is partly due to the limited global incidence of HIE, estimated at 8.5 per 1000 live births [10]. Recently, a CNN model achieving expert-level neonate seizure detection was developed using a private dataset consisting of 50.000 annotated EEG hours from 202 neonates. Broad access to such databases is mainly limited by legal regulations regarding dataset sharing between centres and countries. To date, only two datasets of neonatal EEG are publicly available: a dataset of 79 infants with seizure annotations (available in [11–13]) and a dataset of 53 infants with grades of the severity of background EEG abnormalities (available in [14]).

Although a recent study [15] has shown that expert-level performance has been achieved for seizure detection on unseen data, qualifying this *ConvNeXt* model for further clinical validation, the development of ML models for seizure forecasting remains at an early stage. Unlike seizure forecasting in pediatric and adult cohorts [16–18], the development of analogous models for neonates remains constrained by the absence or subtle clinical manifestations of seizures. A pioneering study dedicated to short-term neonatal seizure forecasting using an ML model was based on convolutional long short-term memory (LSTM) neural networks [19]. Recently, a gradient boosting-based ML model (Neo-GB) focusing only on forecasting the initial onset of seizures was proposed [20], supported by clinical variables and quantitative features (time and frequency) extracted from segmented EEG. Both studies [19, 20] are based on classification of the preictal state (the period before the seizure) without clinical annotation of the preictal state, which is ill-defined for neonates. Study [19] defined the preictal state between 6 min and 1 min before seizure, while [20] estimated that the most effective prognosis of neonatal seizures is up to 16 min from the initial seizure onset. The authors in [19] stated that a possible bias in the short-term forecasting model is due to the selection of 5-minute preictal window. To avoid this potential bias, we avoided preictal annotation; data annotation was based solely on seizure annotation for both forecasting and seizure detection.

The motivation for developing ML models for short-term seizure forecasting and real-time seizure detection is to contribute to the optimization of resource allocation for monitoring in the NICU and to improve and facilitate clinical decision-making in the treatment of newborns. Exploring the possibility of forecasting seizures at shorter intervals can be the initial step toward developing an ML model that estimates the probability of a seizure occurrence over longer time intervals, e.g., in the next few hours. Development of such a system could help improve the triage of scarce EEG resources toward neonates at higher risk for seizures and enable earlier clinical intervention for neonatal seizures when they occur.

Our primary goal was to develop an ML method with dual functionality for seizure detection and forecasting: (1) forecasting neonatal seizures within short prediction intervals (5, 10, 15 min), and (2) detecting seizures in real time with high performance given a limited EEG dataset. As part of this goal, we hypothesized that careful feature selection would support the development of an accurate and robust seizure detection and forecasting framework. Given the challenge of developing ML models with small datasets [15], and motivated by [23–26], we examined the possibility of improving performance through simultaneous feature extraction from EEG and electrocardiography (ECG). The European Reference Network for Rare and Complex Epilepsies (EpiCARE) study from 2025 [27] reported that two-thirds of surveyed European centers still use amplitude-integrated EEG (aEEG) with single- or dual-channel EEG for monitoring high-risk infants. Given the widespread clinical use of aEEG with 1- or 2-channel EEG, we examined the sensitivity of the feature-based ML model to EEG device configuration, i.e., the use of single-channel versus multi-channel EEG, to ensure its robustness and applicability in real clinical conditions.

2. Data: collection and preprocessing for machine learning model

The proposed methods are trained and tested on the following datasets: a private single-channel EEG dataset collected from 82 infants at the Institute for Child and Youth Health Care of Vojvodina, Novi Sad, Serbia (ethical approval 4794–1, 1/11/2023), and an open-access multi-channel EEG dataset collected from 79 infants at the Helsinki University Hospital, Finland (ethical approval HUS/2347/2016, 15/12/2016). For the sake of comparison, following [19] we used an open-access multi--channel EEG database collected from 53 infants at Cork University Maternity Hospital, Ireland (ethical approval from the Cork Research Ethics Committee). This dataset was included to augment the seizure-free group, as it was originally collected for training ML models to assess the severity of background abnormalities. All datasets are small in a machine-learning context, consisting of fewer than 200 infants, with a variety of gestational ages (late preterm and full term) and diagnoses. To date, the largest private dataset (50.000 h of per-channel annotated data) used to develop an ML model for seizure detection includes 202 infants, collected at Cork University Maternity Hospital over several decades of research [15], but remains at the upper limit of what is consider a small dataset in terms of the number of infants.

For the sake of simplicity, we refer to the datasets according to their geographic locations: the Novi Sad, Helsinki, Cork (open-access dataset of neonatal EEG recordings graded for the severity of background abnormalities), and Cork-private (private dataset with seizure annotations) datasets. The seizure detection model was trained and evaluated using two datasets: a single-channel P3–P4 Novi Sad dataset (n = 82) and a multi-channel Helsinki dataset (n = 79). Although the P3–P4 channel could be derived from the multi-channel Helsinki recordings through re-montaging, this was not performed due to the lack of channel-specific annotations; only global annotations are available for the Helsinki dataset and a full montage was used during seizure review. For seizure forecasting, the proposed ML model was evaluated again on the single-channel P3–P4 Novi Sad dataset (n = 82), the multichannel Helsinki dataset (n = 79), and the multi-channel Cork dataset (n = 53). The inclusion of the Cork dataset was motivated by the need to ensure a fair comparison with state-of-the-art ML forecasting models, which used the Cork dataset to increase the non-seizure group in [19]. Only two records in the Cork dataset had short seizure events (localized to channels C3 and C4), whereas all other records are seizure-free. Consequently, the applicability of this dataset for machine learning–based seizure detection is limited. The Cork-private dataset was not accessible for use in this study. All procedures were performed in compliance with relevant laws and institutional guidelines and were approved by the appropriate institutional committee(s).

Seizure events were annotated by a physician experienced in neonatal EEG for the Novi Sad dataset and by three EEG experts for a global annotation across all channels in the Helsinki dataset. For this Helsinki dataset, we used the majority consensus of the three reviewers as the annotation. To avoid potential bias, we did not used preictal annotation due to the lack of a precise definition [21,22], for the seizure prediction task. Segments were classified as seizure segments if ≥ 10 s of

the segment was annotated as seizure (including a disjoint seizure event), otherwise they were classified as non-seizure segments. If at least one seizure segment occurred within a predefined prediction interval, the period was classified as the period with a predicted seizure occurrence. No distinction was made between periods containing one or more neonate seizure segments.

For the Novi Sad database, the first 6 h of high-quality single-channel EEG (scalp impedance <10 kΩ) were used. We opted for non-overlapping 20 s segments, in line with the pioneering study on short-range forecasting [19]. Raw EEG was band-pass filtered between 0.5–30 Hz.

We compared datasets by clinical and demographic characteristics to assess their diversity. In the Supplementary Material, Table S1 summarizes the demographic and clinical characteristics of the Novi Sad, Helsinki and Cork databases. The Novi Sad database included more infants and had a longer total duration of EEG recordings (492 h) compared to the Helsinki and Cork databases. The Novi Sad and Cork databases included only infants with hypoxic-ischaemic encephalopathy (HIE), with different grades of severity, while the Helsinki dataset predominantly consisted of infants with HIE but also included birth asphyxia, stroke and other aetiologies.

Table 1 shows the dataset characteristics for the Novi Sad, Helsinki and Cork datasets. Helsinki and Cork datasets used multi-channel EEG, partly recorded using EEG system from the same manufacturers, while the Novi Sad dataset uses a single-channel EEG device (with electrode locations in the parietal region, P3 and P4). The databases differed in terms of seizure-to- non-seizure imbalance ratios. The observed differences in class imbalance are likely influenced by both methodological and physiological factors. The use of single-channel EEG in the Novi Sad dataset may limit seizure detection compared with multi-channel recordings based on the 10–20 EEG system [28]. Furthermore, variations in recording timing may also play a role, as seizure incidence is known to be higher within the first 24 h of life in neonates. Datasets were split into training and testing sets, considering inter-subject stratification (subjects included in the training dataset were not part of the test dataset). In this study, multiple data-splitting strategies were employed, including an 80:20 train–test split, leave-one-subject-out validation, and 10-fold cross-validation, depending on the objective of the analysis. We explicitly indicated, below each reported result, the data-splitting technique used and the number of infants included in the test set.

Table 1
Characteristics of neonatal EEG databases for seizure detection.

	Novi Sad	Helsinki	Cork
Number of infants	82	79	53
Total duration, h	492 h	112h	169 h
Sampling frequency	200 Hz	256 Hz	256/200 Hz
Annotated seizure events	353	450	3
Class imbalance	15:1	5:1	N/A
Type of database	Private	Public open-access	Public open-access
EEG device	Single-channel	Multi-channel	Multi-channel
EEG system / manufacturer	Olympic Medical CFM Viewer 6000 device (Natus Medical, USA)	NicoletOne, Natus, USA (amplifier) EEG caps (ANT-Neuro, Germany)	NicoletOne ICU Monitor (Natus, Middleton, WI, USA), Neurofax EEG-1200 (Nihon Kohden, Tokyo, Japan)
EEG montage	P3-P4	Fp2-F4, F4-C4, C4-P4, P4-O2, Fp1-F3, F3-C3, C3-P3, P3-O1, Fp2-F8, F8-T4, T4-T6, T6-O2, Fp1-F7, F7-T3, T3-T5, T5-O1, Fz-Cz, Cz-Pz	F4-C4, F3-C3, C4-T4, C3-T3, C4-Cz, Cz-C3, C4-O2, C3-O1
Number of experts performing EEG annotations	1	3	2

3. Proposed ML model

The proposed ML method has dual functionality: (1) forecasting seizures within short prediction intervals (5, 10, and 15 min), and (2) detecting seizures in real time. Both approaches share a common processing pipeline consisting of: (1) segmentation of EEG signals into non-overlapping 20-s segments, (2) preprocessing by band-pass filtering (0.5 to 30 Hz), (3) feature extraction using the same feature set (Table 2), and (4) classification using AdaBoost.M1.

For seizure detection, each 20-s EEG segment is independently classified as seizure or non-seizure. A post-processing step is applied to account for temporal dependencies between consecutive EEG segments in the detection results. For seizure forecasting, three consecutive 20-s EEG segments (1-min sliding window) were used to predict seizure occurrence within a predefined prediction interval (5, 10, and 15 min).

The same AdaBoost.M1 algorithm was used for both seizure detection and seizure forecasting, although separate models were trained for each task. We opted for AdaBoost.M1 as an ensemble learning method that combines multiple weak learners into a strong classifier through an iterative reweighting procedure. In this study, shallow decision trees with a maximum depth of three were used as weak (base) classifiers, and a total of 300 boosting iterations were performed. At each iteration, a weak classifier is trained on the weighted training data, and the weights of previously misclassified EEG segments are increased, forcing subsequent classifiers to focus on difficult-to-classify samples. The final prediction is obtained as a weighted combination of all weak classifiers. By progressively emphasizing misclassified samples, AdaBoost improves classification performance and generalization compared with a single decision tree. The selected parameters were empirically chosen to provide a balance between classification performance, model complexity, and robustness, while reducing the risk of overfitting.

Fig. 1 presents the flowchart of the proposed ML models (a) seizure forecasting within a prediction interval and (b) seizure detection. The proposed ML model uses features extracted from EEG time series, which were grouped into four categories: (1) singular value–based features, (2) binarized entropy–based features, (3) spectral power band features, and

Table 2
Feature list of proposed ML model.

1. Maximum singular value of 20-s EEG segments	11. Power in the delta frequency range of 20-s EEG segments.
2. Norm difference between the vectors s_s and s_{BL}. ($\|l_s\|_2$)	**12. Power in the theta frequency range of 20-s EEG segments.**
3. Standard deviation of s_s	**13. Power in the alpha frequency range of 20-s EEG segments.**
4. Absolute difference between the sums of s_s and sum s_{BL}.	14. Power in the beta frequency range of 20-s EEG segments.
5. Norm difference between vectors $BinEn_s$ and $BinEn_{BL}$	15. Maximum PDF value f_s^{max} for each subsegment.
6. Sum of $BinEn_s$.	16. Maximum PDF value f^{max} for 20-s EEG segment.
7. Sum of skewness values of all subsegments, γ_s.	17. Bandwidth of the PDF estimated from a 20-s EEG segment.
8. Standard deviation of skewness values of all subsegments, γ_s.	18. Minimum EEG value (x-axis) of the estimated PDF from a 20-s EEG segment.
9. Norm difference between the vectors γ_s and γ_{BL}, $\|l_\gamma\|_2$	19. Maximum EEG value (x-axis) of the estimated PDF from a 20-s EEG segment.
10. Sum of standard deviation values of all subsegments, σ_s.	20. Difference between the maximum and minimum EEG values (x-axis) of the estimated PDF from a 20-s EEG segment.
	21. *BinEn* of 20-s ECG segments

Note: Bold features indicate features with higher importance. *SVD*-based features are highlighted in light blue; *BinEn*-based features are highlighted in light pink; power-estimation-based features are highlighted in gray; and statistical-moment-based features are highlighted in orange.

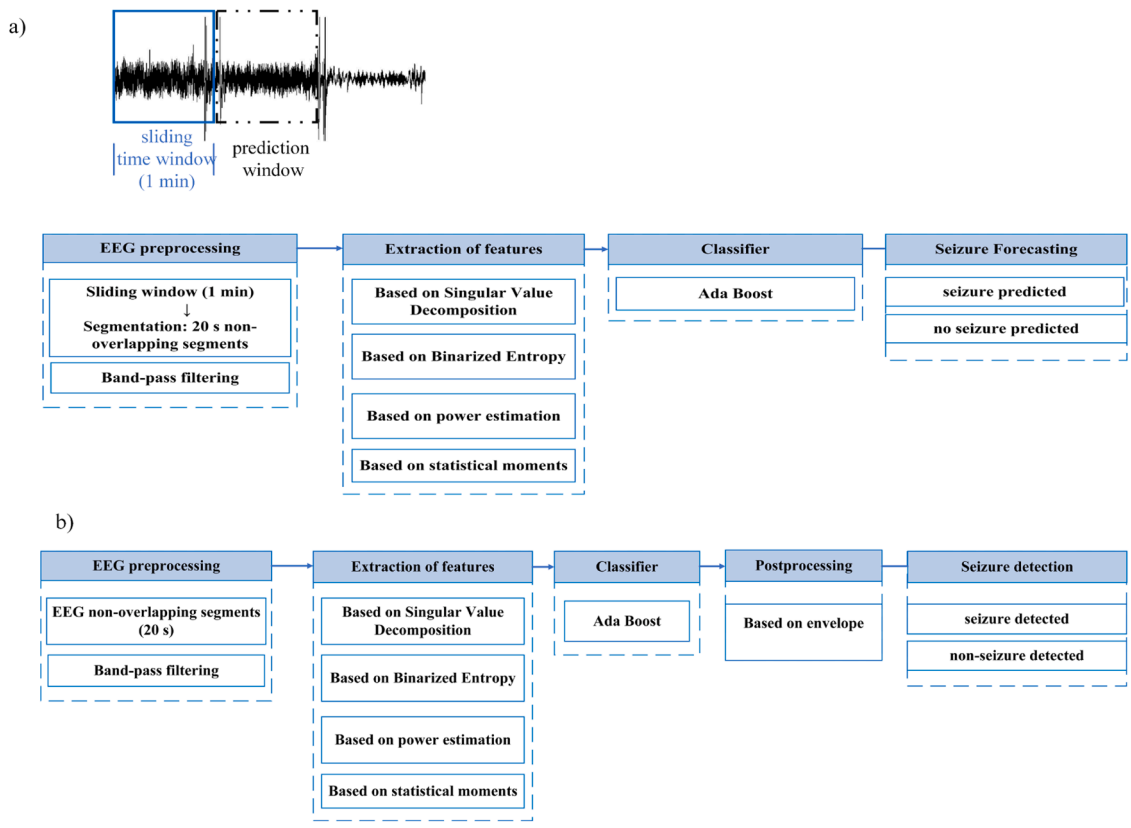

Fig. 1. Flowchart of the proposed ML model. (a) Seizure forecasting within the prediction interval (the period during which the ML model forecasting seizure occurrence). A sliding time window of 1-min (three consecutive 20-s EEG segments) was used. (b) Flowchart of the proposed ML model for real-time seizure detection. Detection was based on 20-s EEG segments. The feature set and classifier were the same those used in (a) and (b).

(4) statistical moment–based features. Feature selection was performed using a wrapper-based iterative forward-selection procedure within each training fold of the cross-validation process to avoid data leakage. The procedure began by selecting the individual feature that achieved the highest MCC. Subsequently, at each iteration, every remaining candidate feature was individually added to the current subset and evaluated using the AdaBoost classifier. Classification performance was assessed using the Matthews correlation coefficient (MCC), computed exclusively on the training data of the corresponding fold. The feature yielding the largest MCC improvement was retained in the subset. The initial pool of 90 candidate features included features derived from multiple extraction approaches, including statistical, spectral, entropy-based (using different estimation methods), time-frequency, and nonlinear analyses. The forward-selection procedure identified a final subset of 21 features that provided the best trade-off between classification performance and implementation complexity.

Feature selection was guided by hardware implementation constraints, with the aim of reducing the need for long-term EEG or ECG data storage requirements. Due to hardware-related constraints, heart rate variability (HRV) analysis from ECG signals was not considered, as reliable HRV estimation requires longer recording durations than the 20-s ECG segments used in this study. Statistical features, including mean, standard deviation, skewness, and kurtosis computed from 20-s ECG segments, as well as EEG–ECG cross-entropy features derived from 20-s EEG and ECG segments, were also evaluated during the initial feature selection process. However, they did not improve model performance and were therefore excluded from the final feature set. Instead, the analysis was limited to computationally efficient features, such as statistical moments, power spectral density and entropy-based features. Table 2 lists 20 EEG features and one ECG feature. The motivation for selecting these features is detailed below.

4. Feature extraction for the ML model

4.1. Singular value decomposition features

Singular value decomposition (SVD) has been shown to be useful in identifying HIE infants at risk of seizures [29] and for seizure forecasting in pediatric epilepsy [16]. We hypothesize that sudden changes in singular values, reflecting energy distribution, occur during seizures. Supplementary Figure S1 shows the maximum singular value observed in EEG segments, where there is a noticeable increase in the singular value during the seizure period. SVD was applied to single-channel EEG, organized as the matrix $\boldsymbol{A}^{m}_{M\times L}$, where M represents the number of EEG channels (M=1), L represents the number of samples, and m denotes the mth subsegment of observed 20-s EEG segmented single-channel EEG.

The singular value decomposition of the matrix $\boldsymbol{A}^{m}_{M\times L}$ is given [30]:

$$\boldsymbol{A}^{m}_{M\times L} = \mathbf{U}\cdot \boldsymbol{S}^{m}\cdot \mathbf{V}^{\mathrm{T}} = \sum_{i=1}^{r} s_i^m \cdot \mathbf{U}_i \cdot \mathbf{V}_i^{\mathrm{T}},\ r \le \min(M,L) \quad (1)$$

Here T denotes transpose, r denotes rank, and

$\mathbf{U}$ – orthogonal matrix of size MxM with columns $\mathbf{U}_i$ –the eigenvectors of the product $\mathbf{A}\cdot\mathbf{A}^{\mathbf{T}}$;

$\mathbf{V}$ – orthogonal matrix of size LxL with columns $\mathbf{V}_i$–the eigenvectors of the product $\mathbf{A}^{\mathrm{T}}\cdot\mathbf{A}$;

$\boldsymbol{S}^{\boldsymbol{m}}$ – matrix of size MxL, comprising of singular values s_i^m of the mth subsegment, arranged in descending order, i = 1, …, r, on the main diagonal, while the remaining off-diagonal elements are zero. Each 20-s segment is divided into four subsegments duration of 5 s, m = {1, 2, 3, 4}.

We formed vector $s_s = \left[s_1^1,\ldots,s_1^m\right]$, where m = {1, 2, 3, 4}, consisting of the largest singular value s_1^m of each 5-s sub-segment of the observed 20-s EEG segment. Baseline EEG segments were defined as

non-seizure segments extracted from seizure-free recordings and are hereafter denoted as BL. s_{BL} was estimated as the average of the s_s vectors calculated of the selected BL. We hypothesize that the norm between the vector s_s computed from the observed EEG segment and s_{BL} can improve classification performance:

$$\| l_s \|_2 = \sqrt{\sum_{i=1}^{m} \left| s_s^i - s_{BL} \right|^2} \quad (2)$$

In Table 2, *SVD*-based features are highlighted in light blue color.

4.1.1. Binarized sample entropy features

Entropy measures the irregularity of time series and has proven effective for seizure forecasting in adult, pediatric, and neonatal EEG [19,31,32]. Approximate entropy (*ApEn*) and Sample entropy (*SampEn*) are widely used in biomedical signal analysis [33]. Although stationarity was not originally a prerequisite for reliable entropy estimation, the literature points out that stationarity conditions are desirable [34]. Binarized entropy (*BinEn*), developed for computational efficiency and to enhance the stationarity of the time series [35], operates on differentially coded data and follows the *ApEn/SampEn* framework. The non-stationary nature of the EEG is further mitigated by EEG segmentation.

The motivation for selecting *BinEn* as a feature was its statistical significance between HIE infants with and without seizures. Supplementary Figure S2a compares *BinEn, ApEn, SampEn* values obtained from EEG recordings from HIE infants with and without seizures (Novi Sad).

The binarized sample entropy was estimated for each mth EEG subsegment to form the vector $BinEn_s = \left[BinEn_s^1, \ldots, BinEn_s^m \right]$. Each 20-s segment is divided into four 5-s subsegments, $m = \{1, 2, 3, 4\}$. Baseline $BinEn_{BL}$ was estimated as an average $BinEn_s$ from selected non-seizure 20-s EEG segments. Similar to the SVD feature in (2), we hypothesize that the norm between vector $BinEn_s$ computed from the observed EEG and $BinEn_{BL}$ can improve the discrimination between seizure and non-seizure EEG segments.

$$\| l_{BinEn} \|_2 = \sqrt{\sum_{i=1}^{m} \left| BinEn_S^i - BinEn_{BL} \right|^2} \quad (3)$$

Following [36,37], which showed that heart rate variability (HRV) entropy is a useful discriminative feature between groups of neonates with and without seizures, we examined the discriminative potential of *BinEn* computed from 20-s ECG segments. Figure S2b shows a statistically significant difference in ECG *BinEn* values between seizure and non-seizure segments (Helsinki dataset). All tests were two-sided, and a p-value <0.05 was considered statistically significant.

4.1.2. Power estimation and statistical moments features

Motivated by its reported importance in predicting the first neonatal seizure [20], we evaluated spectral power features computed from EEG segments. Table 2 lists features based on statistical moments and spectral power in four frequency bands (delta, theta, alpha, and beta).

The feature notation is described below:

- A vector containing the maximum values of the probability density functions (PDFs) $f_s^{max} = \left[f_s^1, \ldots, f_s^m \right]$, was formed, where f_s^m denotes the maximum value of the PDF estimated from the mth subsegment ($m = 1, 2, 3, 4$), each with a duration of 5 s.
- The vector $\gamma_s = \left[\gamma_s^1, \ldots, \gamma_s^m \right]$, $m = \{1, 2, 3, 4\}$ was formed from the skewness values of the mth subsegments, . γ_{BL} was calculated as the mean of the γ_s vectors obtained from all non-seizure 20-s EEG segments extracted from the selected baseline recordings (BL) of seizure-free neonates.

The norm between γ_s and γ_{BL} was calculated as follows:

$$\| l_\gamma \|_2 = \sqrt{\sum_{i=1}^{m} \left| \gamma_s^i - \gamma_{BL} \right|^2} \quad (4)$$

Features in multi-channel EEG signals are averaged across all channels.

Feature redundancy analysis was performed following [38], using Pearson correlation between features within each subject. The averaged correlation matrix is shown in Figure S3 in the Supplementary Materials. Most of the feature pairs show low or medium correlation (< 0.7) and no large clusters of highly correlated features were observed. Higher correlations are observed between feature pairs 8–9, 15–16 and 19–20, reflecting their inherent mathematical relationships rather than unintended redundancy (Table 2). Considering the expected high correlation due to mathematical relations, all features were retained in the proposed ML model. This design choice reflects a balance between statistical independence and domain relevance, which is particularly important for neonatal seizure forecasting and detection.

The authors in [39] noted that for neonatal EEG electrode placement, the spatial power spectrum decays rapidly until about 0.5–0.8 cycles per centimeter, and the amplitude of oscillatory events diminishes within 4–6 cm to the level of the global background activity. Higher frequencies (12–20 Hz) show the most rapid spatial decline in amplitude. The Neo-BG model for forecasting the onset of the first seizure includes two different pairs of electrodes (Fp1-T3 and C4-O2) from the multi-channel EEG with differences in the theta power features [20]. Realistic variations in skull conductivity and source depths can introduce orders-of-magnitude differences in the spatial frequency of scalp EEG.

Fig. 2a compares EEG energy across channels with different locations and electrode distances. The open dataset Cork [14] was used, excluding two seizure-positive infants, to avoid cases where some electrodes recorded seizures while others did not. Although the signal energy of bipolar channels with larger inter-electrode distances was within the range of the signal energy observed in the channels with shorter inter-electrode distances, electrode distances may affect features 2, 4, 5, 9, and 11–14 (Table 2). The Helsinki dataset was re-montaged to the P3–P4 bipolar derivation to assess whether the features differed significantly from those extracted from multichannel pairs. Fig. 2b illustrates these differences, with green indicating statistically significant differences for all features except feature 8.

4.2. Post-processing step

A post-processing step was introduced to generate a temporal prediction vector to account for the segment-independent analysis. Supplementary Figure S4 illustrates seizure probabilities (black) and the fitted envelope (green), which corrects misclassification of individual 20-s segments when the surrounding segments are correctly classified. Final decisions were based on this envelope, with probabilities above 0.6 reliably indicating seizure segments.

4.3. Experimental design and validation strategy

To ensure both a fair comparison with the state-of-art literature and a clinically meaningful evaluation, we employed different data splitting and validation strategies depending on the experimental objective. For clarity, Table 3 summarizes the tasks (forecasting/detection), validation strategies and used datasets.

5. Results

The dual functionality of the proposed models is evaluated in two subsections. Subsection 4.1 reports the results of short-term seizure forecasting, assessing the models' performance in forecasting seizures within prediction intervals of 5, 10, 15, and 30 min. Section 4.2 presents the results of real-time seizure detection, highlighting the models'

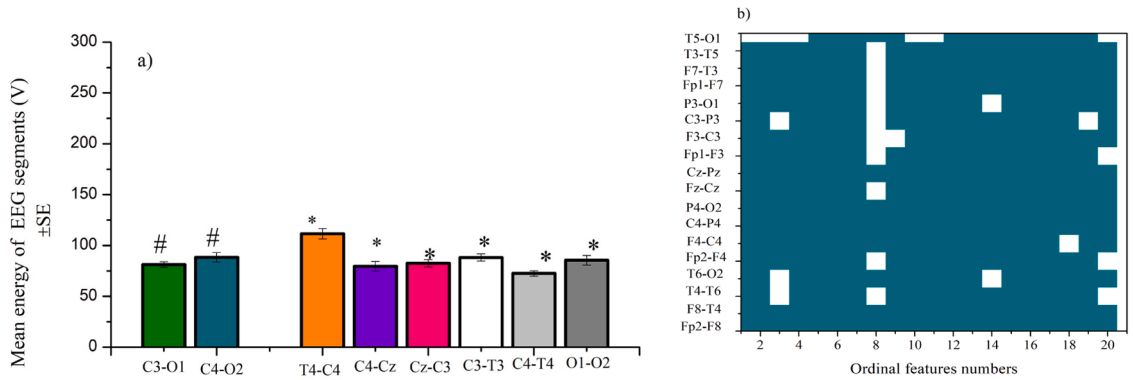

Fig. 2. (a) Mean energy of segmented EEG (± SE) in the Cork database. No differences were found between the bipolar channels with larger (#) and smaller (*) inter-electrode distances. (b) Statistical significance between features extracted from the single-channel P3-P4 derivation and the multichannel pairs. Features are numbered according to Table 2. Green indicates statistically significant differences, while white indicates no statistically significant differences between the features. The Wilcoxon signed-rank test was used.

Table 3
Summary of datasets, objectives (seizure detection and seizure forecasting), and corresponding validation strategies used in this study.

Dataset	Task	Notes	Validation strategy	Results presented in
FORECASTING				
Novi Sad	Seizure forecasting	single-channel	Leave-one-out cross-validation per subject with seizure	Tabel 4 Tabel S2 Tabel S3
Helsinki	Seizure forecasting	muti-channel	10-fold cross-validation	Tabel 5
Cork	Seizure forecasting	muti-channel, only non-seizure group	10-fold cross-validation	Tabel 5
DETECTION				
Novi Sad	Seizure detection	single-channel	80:20 split	Tabel 6
Novi Sad	Seizure detection	single-channel	Leave-one-out cross-validation per subject with seizure	Tabel 7 Figure S6
Novi Sad	Seizure detection	single-channel	10-fold cross-validation across subjects	Tabel S4
Helsinki	Seizure detection	muti-channel	80:20 train–test split	Tabel 8
Helsinki	Seizure detection	muti-channel	Leave-one-out cross-validation per subject with seizure	Figure S10

effectiveness in detecting seizures on single-channel P3-P4 EEG.

We tested the robustness of the proposed ML model on different datasets that differ in channel configurations (single-channel and multi-channel), vendors, database heterogeneity, and institutions with different clinical protocols. To compare the quality of the EEG recordings in the two datasets, we assessed the length of EEG recording pauses (electrode impedance checks) and the presence of large-amplitude artefacts (amplitudes $>$ 500 μV) [11]. In the Novi Sad dataset, all infants exhibited $<$1% of the EEG recording duration with amplitudes exceeding 500 μV (Figure S5a, Supplementary Materials), whereas in the Helsinki dataset, only six infants showed durations exceeding 2%. The prevalence of EEG recording pauses was $<$0.5% of the EEG recording duration in the Novi Sad dataset (Figure S5b, Supplementary Materials) and exceeded 1% of the EEG recording duration in 15.19% of infants in the Helsinki dataset. This difference may be a consequence of the use of multi-channel EEG, which likely require more impedance checks.

The evaluation metrics followed state-of-the-art recommendations for imbalanced datasets such as neonatal seizures [40]. Although AUROC is emphasized as the most commonly used metric for seizure detection [8,41–42], it is sensitive to class imbalance. Although Area Under the Precision-Recall Curve (AUPRC) is widely preceived as a better choice than AUROC for class-imbalanced data, both theoretical and empirical evidence has shown that this assumption is incorrect [43]. Following the guidelines of [15], we employed transparent metrics, including the MCC, Cohen's κ, false detections per hour (FD/h), and correlation with seizure burden. Metric computation was performed using the SPEED Python package [15,44].

Independent held-out validation, although considered good practice [40], was not feasible because the features extracted from the single-channel P3-P4 derivation differed from those extracted from the multichannel electrode pairs of the 10–20 EEG system. Features were extracted from the training dataset of the Novi Sad database using the single-channel P3-P4 derivation, which differs in measurement locations compared to the Helsinki dataset. Although the static dependence between features extracted from P3-P4 and other bipolar channels more commonly used in multi-channel EEGs was demonstrated in Fig. 2, we do not exclude the possibility that the validation of the ML model was further limited by the differences in the hardware implementation of EEG devices from different technology generations and clinical centers. For example, held-out validation of ML models for grading neonatal EEG on unseen datasets recorded with different multi-channel EEG devices across European centers has shown a significant drop in the performance of the validated ML models [45].

Held-out validation using only the P3–P4 channel from the Helsinki dataset, as well as the examination of the influence of channel selection on EEG recordings acquired with the same device, was limited by the lack of annotations for individual channels, as global annotations do not require seizures on every channel [15].

5.1. Performance analysis: forecasting seizures within the prediction interval

Table 4 shows the performance on subjects with seizure from the Novi Sad database for different prediction intervals (5, 10, 15 and 30 min) using a 1-min time sliding window (marked in Fig. 1). Performance increases as the prediction interval decreases. The higher F1-score indicates that the combination of complementary handcrafted EEG features and the AdaBoost ensemble learning strategy achieves a better balance between positive predictive value (PPV) and sensitivity, resulting in improved overall forecasting performance.

We examined the impact of the EEG sliding window duration on performance for the same prediction interval. Supplementary Table S2 shows the median performance values for subjects with seizures from the Novi Sad database for different time window durations (1, 3, and 10 min). The values shown in bold in Supplementary Table S2 indicate

Table 4
Subject-level seizure forecasting performance for different prediction intervals. Classification was performed without incorporating clinical data.

TD (min)	Prediction interval (min)	Sensitivity	Specificity	NPV	PPV	FPR	F1	MCC
1	5	.956 (.942 −.963)	.889 (.857-.936)	.368 (.286 −.517)	.997 (.996-.997)	.111 (.066-.137)	.976 (.968 −.980)	.551 (.475-.629)
1	10	.912 (.903 −.936)	.897 (.865-.938)	.250 (.197- 361)	.997 (.996-.997)	.102 (.063-.134)	.952 (.942-.966)	.443 (.389-.521)
1	15	.866 (.833 −.910)	.924 (.886-.954)	.210 (.178 −.328)	.996 (.995-.997)	.075 (.046-.112)	.926 (.907 −.952)	.387 (.348 −.460)
1	30	.787 (.733 −.831)	.889 (.842-.926)	.178 (.117 −.222)	.995 (.993-.996)	.111 (.074.159)	.879 (.846-.906)	.320 (.219-.362)

Note: Data are reported as the median (95% confidence interval). Novi Sad dataset. TD- time window duration. Observed EEG window length: 1 min. Leave-one-out cross-validation per subject with seizure was performed. Abbreviations: PPV, positive predictive value; NPV, negative predictive value; FPR, false-positive rate; F1, F1-score; MCC, Matthews correlation coefficient.

performance superior to that obtained with the 1-minute sliding window. Extending the length of the EEG sliding window did not improve all performance metrics and did not result in any significant changes.

Similar to the work in [20], we examined whether including additional clinical parameters alongside existing EEG features could improve the forecasting performance of the models. We expanded the features list to include Apgar scores, delivery mode, assisted ventilation, resuscitation, cord pH, base deficit, $stHCO_3$, lactate, and glycaemia for each infant. Median values (IQR) of the listed clinical parameters are shown in Table S1. We analyzed the influence of clinical parameters on the forecasting performance in prediction intervals of 1 and 10 min. In Supplementary Table S3, cases in which the features list was expanded with for clinical parameters are indicated this with a 'yes'. The inclusion of clinical parameters in the feature set did not significantly affect model performance, as reflected by similar median values and interquartile ranges (IQRs).

Fig. 3 shows the averaged probability of seizure risk for the non-seizure group from the Novi Sad datasets. Light green line marks the upper and lower limit of the confidence interval, while the dark green line marks the average probability risk.

The results in Table 5 show that the proposed method outperformed the approach reported in the pioneering study for short-range seizure forecasting [19]. The ConvLSTM model supported the extraction of 220 temporal features from 20-second non-overlapping EEG segments and used annotated data with two classes: the preictal period (period before seizure onset) and the interictal period (non-seizure) period. In [19], the ConvLSTM model was tested jointly on the Helsinki (79 infants) and open-access Cork datasets (53 infants). To ensure a fair comparison, we opted for patient-wise 10-fold cross-validation and the same choice of metrics (AUROC, AUPRC, F1 score, and MCC) as used in [19].

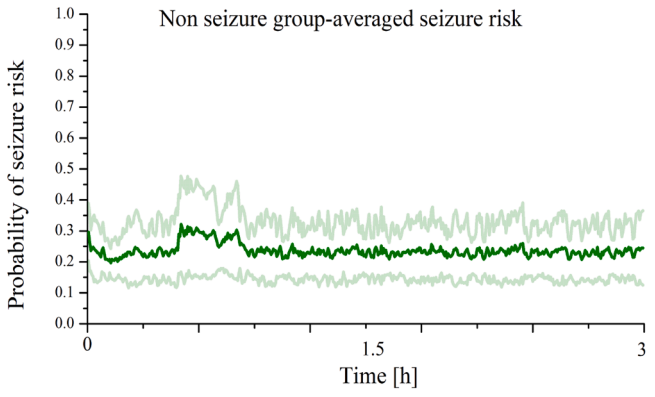

Fig. 3. The group-averaged seizure risk curves and 95% confidence intervals for the non-seizure group from Novi Sad dataset. The dark green line represents the average seizure risk, while the light green lines indicate the upper and lower bounds of the 95% confidence interval. Only the first 3 h of the 6 h recording are shown, as the remaining segment does not alter the overall trend.

Table 5
Seizure forecasting performance comparison between the proposed method with the existing ConvLSTM method [19] on the combined Helsinki and Cork datasets. Data are presented as mean with standard error (SE) across 10-fold cross-validation. The prediction interval was 10 min.

Mean ± SE	Proposed method	ConvLSTM [19]
AUROC	**.770 ± .082**	.678±.041
AUPRC	**.307 ± .170**	.218±.006
MCC	**.478 ± .166**	.255±.054
F1	**.966 ± .0083**	.334±.050

Note: Patient-wise 10-fold cross-validation was performed, resulting in approximately 7 infants from the Helsinki dataset and 5 infants from the Cork dataset per test fold. Abbreviations: AUROC, area under the receiver operating characteristic curve; AUPRC, area under the precision-recall curve; MCC, Matthews correlation coefficient; F1, F1-score.

5.2. Performance analysis: seizure detection

Table 6 compares the seizure detection performance of the proposed model on the Novi Sad dataset (82 infants, 492 h of EEG) with ConvNeXt models tested on the same dataset. The ConvNeXt architecture proposed in [15] is based on a modern convolutional neural network, and the extra-large (XL) model trained on the large Cork-private dataset achieved expert-level seizure detection. We opted for this architecture since it is designed on a per-channel basis, which allows validation on the single-channel Novi Sad dataset. The ConvNeXt network was designed to support testing with different capacity configurations. All sizes of the

Table 6
Seizure detection performance on the Novi Sad dataset. All methods were trained and tested on this dataset using an 80:20 split.

	Proposed method-AdaBoost	ConvNeXt Nano [15]	ConvNeXt Small [15]	CNN model [46]
AUROC	.785	.825	.871	.81
AP/ AP50	**.438/.441**	.426/0.288	.435/0.363	.188/ .130
Pearson's coefficient	**.635**	.446	.462	.29
MCC	**.635**	.438	.456	.207
Cohen's k	**.635**	.433	.452	.20
Sensitivity/ Specificity (%)	**57.9/99.1**	42.4/96.2	44.7/ 96.1	24.7/96
PPV/NPV (%)	**72.7/98.2**	56.3/93.5	57.2/93.7	24/96
FD/h	**.777**	2.060	1.620	13.235
Seizure Burden	**.853**	.626	.614	.20

Note: An 80:20 train–test split was performed at the subject level, with 16 infants included in the test set. Abbreviations: AUROC, area under the receiver operator characteristic curve; AP, average precision; AP50, average precision with recall > 50%; MCC, Matthews correlation coefficient; PPV, positive predictive value; NPV, negative predictive value; FD/h, false detections per hour.

ConvNext model (Nano, Small, Medium, Large and XL) were tested. Table 6 shows the results of the most successful architectures (Nano and Small). Additionally, we tested the CNN architecture proposed in [46], which was originally designed for a small multichannel EEG dataset (Helsinki), on the single-channel Novi Sad dataset. The CNN model showed lower performance overall in comparison to ConvNeXt and the proposed method, albeit with a comparable AUROC. All models were trained and tested on an 80:20 random split of this Novi Sad dataset.

Although [45] demonstrated that deep learning methods exhibit greater robustness than feature-based machine learning models when applied to data acquired using equipment of different ages and technological generations, the CNN showed a drop in performance on the single-channel Novi Sad dataset compared with the published results on the Helsinki dataset [46]. In contrast, the ConvNeXt model achieved strong performance on this dataset, which may be attributed to its improved robustness. One possible explanation is that the CNN architecture was not originally designed for single-channel EEG, leading to reduced performance.

For our dataset, we found no advantage of the deep learning models: the proposed ML model outperformed both ConvNeXt models (Nano and Small) and the CNN model on the Novi Sad dataset. To provide context for these results, the proposed ML method achieved comparable performance to a large ConvNeXt model trained on a significantly larger private dataset in [15], although the comparision here is limited by the different test sets. Bold values in Table 6 indicate superior performance, with the Small ConvNeXt generally outperforming the Nano version.

The proposed ML model was then trained and tested using leave-one-subject-out cross-validation on the same (Novi Sad) dataset. Results are presented in Table 7, showing similar performance to that obtained with the 80:20 train-test split in Table 6.

In the Supplementary Materials, the reported confusion matrix (Figure S6) is computed by pooling predictions across all subjects in the leave-one-out cross-validation performed per subject with seizure evaluation (segment-level aggregation). In contrast, the performance metrics reported in Table 7 are calculated on a per-subject basis and summarized using the median, resulting in subject-level aggregation. Due to strong class imbalance and inter-subject variability, these two evaluation strategies provide complementary but not identical representations of model performance.

For comparison purposes, additional ML classifiers, including XGBoost and Random Forest, were implemented using the same feature set and evaluation protocol as the proposed approach. Performance values for the best-performing optimized parameter setting are shown in Table S4. AdaBoost.M1 slightly outperformed the other classifiers. Figure S7 shows that BL selection (affecting features 2, 4, 5, and 9 in Table 2) does not affect MCC and Cohen's kappa.

Following the established practice of evaluation of robustness under varying data quality conditions (detailed in [47]), we performed an objective assessment of the EEG quality based on impedance variability (continuous impedance measurements were available for the Novi Sad dataset); the spectral power ratio between the frequency ranges where neonatal activity is expected (0.5 to 10 Hz) and those predominantly associated with artefacts (10 to 30 Hz), as well as atypically high EEG amplitudes. The recordings were divided into three EEG quality groups: high, medium, and low quality. The proposed ML model showed robustness across all metrics, except for a slight change in sensitivity and PPV between the medium- and high-quality groups (Figure S8 in the Supplementary Materials).

We evaluated the performance of the proposed method on a held-out validation set, the Helsinki dataset. Since the Helsinki dataset was annotated globally across all EEG channels, the global prediction was determined by taking the maximum value among the per-channel model predictions. Because some channels may not contain seizure activity, as also observed in [15], classification was performed for each individual channel.

Table 8 presents a comparison of the proposed method's performance on the Helsinki dataset, considering EEG features alone versus the joint utilization of EEG and ECG features. Performance was improved by adding short-duration ECG features. Compared to a state-of-the-art method evaluated on this dataset, which uses an XL ConvNeXt model trained on over 50k annotated hours of EEG [15], the combined EEG+ECG model performs well and is only slightly worse than the state-of-the-art method, with a 10% decrease in MCC, for example. However, there are some differences: in this study we take the majority consensus annotation, whereas in [15] the unanimous consensus annotation is tested; in this study, the resolution of the prediction mask is 1/20 Hz, in [15] it is 4 Hz. Given the large differences in training set sizes, with 50k hours of EEG in [15] compared with our training set of 492 h, we consider the proposed method a valid research prototype suitable for the goals of this study. A comparable analysis of ECG influence could only be performed on the Helsinki dataset, as synchronized ECG recordings were not available for the Cork and Novi Sad datasets.

Supplementary Figure S9 shows permutation-based feature importance for the Novi Sad and Helsinki datasets. In both, the most influential feature was the norm difference between the skewness of the current segment and BL (feature 9, Table 2). To evaluate whether only the six highest-ranked features (highlighted in Table 2) were sufficient, an additional leave-one-subject-out cross-validation experiment was performed on the Novi Sad dataset using the selected features. Although the classifier retained moderate predictive performance, the median MCC decreased from 0.634 (95% CI: 0.539–0.702) to 0.546 (95% CI: 0.460–0.625), while the false detection rate increased markedly from

Table 7
Seizure detection performance of the proposed ML method using leave-one-subject-out cross-validation on the Novi Sad dataset. Data are presented as median (95% confidence interval).

	Subject with seizure-level performances for seizure detection
AUROC	.750 (.697 −.787)
AP/ AP50	.494/.499 (.408/.431 519/.544)
Pearson's coefficient	.634 (.539 −.704)
MCC	.634 (.539 −.702)
Cohen's k	.624 (.564-.669)
Sensitivity/Specificity (%)	50/99.8 (40.35/57.1–99.7/99.9)
PPV/NPV (%)	87.25/98.75 (81.1/97.7 −94.7/99.25)
FD/h	.168 (.0-.336)
Seizure Burden	.836 (.423-.924)

Abbreviations: AUROC, area under the receiver operator characteristic curve; AP, average precision; AP50, average precision with recall > 50%; MCC, Matthews correlation coefficient; PPV, positive predictive value; NPV, negative predictive value; FD/h, false detections per hour.

Table 8
Seizure detection performance of the proposed method using EEG features only (20 features) and EEG features augmented with *BinEn* from 20-s ECG segments (21 features). Tested on the Helsinki dataset.

	only EEG features	EEG +ECG features
AUROC	.753	**.799**
AP/ AP50	.466/.466	**.571/.576**
Pearson's r	.599	**.668**
MCC	.599	**.668**
Cohen's k	.586	**.668**
Sensitivity/Specificity (%)	52.8/97.8	**62.4/97.3**
PPV/NPV (%)	77.2/93.7	**81.7/93.0**
FD/h	**.083**	.124
Seizure Burden, r	**.794**	**.843**

Note: An 80:20 train–test split was performed at the subject level, with 15 infants included in the test set. Abbreviations: AUROC, area under the receiver operator characteristic curve; AP, average precision; AP50, average precision with recall > 50%; MCC, Matthews correlation coefficient; PPV, positive predictive value; NPV, negative predictive value; FD/h, false detections per hour.

0.168/h (95% CI: 0.000–0.336) to 0.839/h (95% CI: 0.503–1.342). These findings indicate that the remaining features provide complementary information that improves model robustness and reduces false detections. The Helsinki dataset also showed a notable ECG-derived feature contribution, with spectral power features strongly affecting performance, especially in the multichannel data, aligning with recent findings [19].

The proposed method shows applicability to single-channel EEG (Novi Sad dataset) and multi-channel EEG (Helsinki dataset). To investigate robustness in a missing channel scenario, we randomly simulated data loss of 10%, 25%, 50% and 100% (corresponding to dropping a whole channel) by randomly removing different numbers of channels from the Helsinki dataset. Since only global annotations were available, we investigated whether there was sufficient information remaining in the data to recover the global annotation. The results are shown in Figure S10. MCC decreased by <1% when randomly missing 6 channels (even with 100% data loss per selected channel); however, completely missing 16 out of 18 channels decreased the MCC value by approximately 30%.

The proposed ML model is still at the level of a research prototype, but it has potential for real-time applicability in clinical settings with low latency and computational requirements. The time required for feature extraction from a 20-s EEG segment is 1.0824 s, and the time required for classification by the overall ML model, it is 7.0519 s. The processing was performed on a standard consumer-grade CPU (11th Gen Intel® Core™ i7–1165G7@2.80 GHz), without specialized hardware requirements. The proposed ML model is validated on single-channel EEG (Novi Sad dataset) and multi-channel EEG datasets (Helsinki dataset), so it is not dependent on montage configuration. Furthermore, the proposed approach operates on short segments (20 s EEG) and does not require long-term storage of EEG or ECG data (feature selection was based on this prerequisite), which makes it suitable for deployment in resource-constrained environments and supports real-time applications [48]. Although implementation-level evaluation on dedicated edge devices or clinical-grade computers, including memory footprint analysis and continuous streaming validation, was beyond the scope of this study, the computational characteristics of the proposed framework indicate its potential for future real-time clinical deployment. An illustration of the monitoring output from the proposed method for the Novi Sad dataset is given in Figure S11.

6. Discussion and conclusion

We have developed an ML model based on the AdaBoost M1 classifier, supported by 21 adequately selected features. The proposed ML model architecture has a dual functionality: (1) real-time seizure forecasting within prediction intervals (5, 10, and 15 min), and (2) real-time detection of seizures. The selected features enable effective ML model performance for forecasting and detection of neonatal seizures in real time, even with small datasets. The feature set consists of 20 features extracted from the segmented EEG and 1 feature extracted from the segmented ECG if available. Features are estimated based on singular values, binarized entropy, power estimation in frequency bands, and statistical moments of EEG/ECG segments. Although entropy estimation has proven to be a useful feature in seizure detection and forecasting for adults and children, among the different methods (including spectral entropy, Shannon entropy [49], and approximate entropy [50]) used in the literature, we opted for binarized entropy. In addition to its computationally efficient implementation, an important consideration for real-time monitoring, the features were able to separate the seizure and non-seizure groups. HRV (a time-series estimate based on the time interval between the R peaks of ECG) has been used in the detection of neonatal seizures [36]. However, HRV estimation would not be reliable due to the small number of detected local maxima in a 20-s ECG segment; therefore, we opted for feature extraction directly from ECG segments. Additionally, we found a statistically significant difference in *BinEn* values between seizure and non-seizure ECG segments across datasets with various diagnoses.

We have validated our hypothesis that independent classification of segmented EEG and adequate feature selection can achieve high performance for seizure detection on small datasets. The proposed method achieved a high performance and outperformed a deep learning approach using a *ConvNeXt* model architecture on Novi Sad dataset. No additional artefact removal was performed except bandpass filtering applied to EEG segments, without using automated artefact detection algorithms or manual annotation and removal of artefacts by experts. Other studies have shown that artefact removal can improve model performance [51], however, the results presented in this study may better reflect realistic real-world clinical conditions. Although short-term forecasting of neonatal seizures showed competitive performance compared with current state-of-the-art methods, it remains in the early stages of development, and further effort is required for performance improvement. The goal of seizure forecasting is to predict seizures hours ahead of time, which remains an unrealised objective. We avoided potential bias in the model due to imprecise preictal annotation, as previously discussed by other studies [19], by solely using annotation on based on expert idenitification of clearly defined seizure events. The proposed method was validated using both single-channel EEG (P3-P4) and a multi-channel EEG, which is considered good practice for feature-based ML models.

Recent studies have emphasized that AI systems intended for clinical use should remain robust under heterogeneous acquisition conditions and varying signal quality rather than relying solely on performance achieved within a single dataset [47]. Although previous work has primarily addressed this challenge in medical imaging [47], similar variability is encountered in neonatal EEG monitoring due to differences in recording hardware and signal quality [45]. In this study, the proposed framework was evaluated on three independent neonatal EEG datasets acquired using different EEG monitoring systems and clinical protocols. The Novi Sad dataset was recorded using a single-channel neonatal EEG monitor based on an earlier-generation hardware platform and amplifier architecture, whereas the Helsinki and Cork datasets were acquired using multichannel clinical EEG systems with different amplifier characteristics, electrode configurations, and acquisition protocols. Evaluating the proposed framework across these heterogeneous recording conditions provides insight into its robustness under realistic variability encountered in neonatal clinical practice. Additional analyses involving ECG feature integration and simulated channel data loss further demonstrated the robustness of the proposed approach under realistic signal degradation scenarios.

Unlike end-to-end deep learning methods, the proposed framework relies on interpretable handcrafted EEG descriptors, including entropy-based measures, spectral features, statistical moments, and SVD-derived characteristics. This design philosophy is consistent with recent biomarker-oriented machine learning approaches that prioritize reproducibility and clinical transparency over latent feature representations [38].

Although clinical deployment was outside the scope of this study, the proposed framework was intentionally designed using computationally lightweight handcrafted features and an AdaBoost classifier. Similar to the deployment-oriented design adopted in SignCraft [48], which emphasizes efficient inference and practical implementation in real-world environments, the proposed architecture was developed with computational efficiency and scalability in mind. Consequently, the framework has the potential to support future real-time deployment scenarios in which EEG data are acquired at the bedside or using portable monitoring systems, segmented into fixed-length windows, and processed either locally on edge devices (e.g., bedside monitors or tablets) or on cloud-based servers. The resulting outputs could provide clinicians with seizure detection and seizure forecasting capabilities with minimal latency. Compared with computationally intensive deep-learning approaches, the proposed pipeline is expected to require fewer

computational resources during inference, making it a promising candidate for deployment in portable and resource-constrained neonatal monitoring environments. Future work will include implementation-level evaluation, including edge-device profiling, memory footprint analysis, and continuous streaming validation to further assess the translational potential of the proposed framework in real-world neonatal monitoring environments.

This study is not without limitations. Given the observed statistically significant differences between features extracted from the single-channel P3–P4 derivation and those obtained from individual channels of multichannel EEG recordings, a primary limitation of the proposed method is that the feature set derived from a single-channel EEG system (P3-P4) may not be directly transferable for training models using multichannel EEG systems or different EEG devices.

Although the datasets were relatively large compared with typical neonatal EEG studies, increasing in the diversity of gestational and postnatal ages, as well as including a wider range of pathologies, would help develop more generalized methods. Overcoming this limitation is possible by continued data collection and store data in specialized centers.

An ML model for forecasting and real-time detection of neonatal seizures can contribute to the optimization of monitoring resources in the NICU and improve and facilitate decision-making in the treatment of newborns. Achieving strong performance for neonatal seizure detection supports further clinical evaluation, while seizure forecasting remains in the initial development phase, with additional research required before potential clinical translation.

Software availability

The open-source code for the proposed ML model is stored in a public GitHub repository:

EEG-AdaBoost-20F

https://github.com/tamaraceranic/EEG-AdaBoost-for-Seizure-Detection-and-Forecasting-Using-20-Features.

Ethics declaration

Written informed consent to take part in the study and to publish the article has been obtained from all participants or their legal representatives. The privacy rights of participants have been observed.

This study was performed in compliance with relevant laws, regulatory frameworks and guidelines where the research took place. This study was approved by the Institute of Child and Youth Health Care of Vojvodina, Serbia. (Approval No 4794-1)

Funding

This research did not receive any specific grant from funding agencies in the public, commercial, or not-for-profit sectors.

CRediT authorship contribution statement

Tamara Skoric: Writing – review & editing, Writing – original draft, Visualization, Validation, Supervision, Software, Project administration, Methodology, Investigation, Conceptualization. **Marija Djermanovic:** Writing – review & editing, Writing – original draft, Formal analysis, Data curation. **John M. O'Toole:** Writing – review & editing, Writing – original draft, Conceptualization. **Maarten De Vos:** Writing – review & editing, Writing – original draft, Conceptualization. **Slobodan Spasojevic:** Writing – review & editing, Writing – original draft, Supervision, Formal analysis, Data curation.

Declaration of competing interest

The authors declare that they have no known competing financial interests or personal relationships that could have appeared to influence the work reported in this paper.

Acknowledgement

This study was supported through participation in the EU-funded COST Actions projects CA20124 - "Maximising the impact of multidisciplinary research in early diagnosis of neonatal brain injury" and CA24148 – "EEG101: fundamentals of open & rigorous EEG Science". The authors would like to thank all families for their willingness to participate in studies and clinical teams for database collection. This research was conducted in accordance with institutional and ethical guidelines. No experiments involving human participants or identifiable personal data were conducted.

The authors declare that there are no competing interests.

The use of ChatGPT (OpenAI) was limited to improving linguistic clarity, grammar, and readability. All intellectual contributions, methodological design, and interpretations remain solely the responsibility of the author. Full control and authorship of the final manuscript are retained by the authors.

Supplementary materials

Supplementary material associated with this article can be found, in the online version, at doi:10.1016/j.cmpb.2026.109552.

References

[1] H. Tekgul, et al., The current etiologic profile and neurodevelopmental outcome of seizures in term newborn infants, Pediatrics. 117 (2006) 1270–1280, https://doi.org/10.1542/peds.2005-1178.

[2] J.M. Rennie, et al., Characterisation of neonatal seizures and their treatment using continuous EEG monitoring: a multicentre experience, Arch. Dis. Child Fetal. Neonatal. Ed. 104 (2019) F493–F501, https://doi.org/10.1136/archdischild-2018-315624.

[3] T.C. Kwok, C. Henry, S. Saffaran, et al., Application and potential of artificial intelligence in neonatal medicine, Semin. Fetal. Neonatal. Med. 27 (2022) 101346, https://doi.org/10.1016/j.siny.2022.101346.

[4] P. Kulakowski, A. Alarcón, T. Skoric, et al., A pathway for integrating artificial intelligence for neonatal neuromonitoring: from code to crib side, J. Pediatr. 285 (2025) 114708, https://doi.org/10.1016/j.jpeds.2025.114708.

[5] W. Deburchgraeve, P.J. Cherian, M. De Vos, et al., Automated neonatal seizure detection mimicking a human observer reading EEG, Clin. Neurophysiol. 119 (2008) 2447–2454, https://doi.org/10.1016/j.clinph.2008.07.281.

[6] A. Temko, E. Thomas, W. Marnane, G. Lightbody, GB. Boylan, Performance assessment for EEG-based neonatal seizure detectors, Clin. Neurophysiol. 122 (2011) 474–482, https://doi.org/10.1016/j.clinph.2010.06.035.

[7] K.T. Tapani, S. Vanhatalo, NJ. Stevenson, Time-varying EEG correlations improve automated neonatal seizure detection, Int. J. Neural Syst. 29 (2019) 1850030, https://doi.org/10.1142/S0129065718500302.

[8] A.H. Ansari, P.J. Cherian, A. Caicedo, et al., Neonatal seizure detection using deep convolutional neural networks, Int. J. Neural Syst. 29 (2019) 1850011, https://doi.org/10.1142/S0129065718500119.

[9] A. O'Shea, G. Lightbody, G.B. Boylan, A. Temko, Neonatal seizure detection from raw multi-channel EEG using a fully convolutional architecture, Neural Netw. 123 (2020) 12–25, https://doi.org/10.1016/j.neunet.2019.11.023.

[10] A.C. Lee, N. Kozuki, L. Blencowe, T. Vos, A. Bahalim, G.L. Darmstadt, S. Niermeyer, M. Ellis, N.J. Robertson, S. Cousens, JE. Lawn, Intrapartum-related neonatal encephalopathy incidence and impairment at regional and global levels for 2010 with trends from 1990, Pediatr. Res. 74 (Suppl 1) (2013) 50–72.

[11] N. Stevenson, K. Tapani, L. Lauronen, et al., A dataset of neonatal EEG recordings with seizure annotations, Sci. Data 6 (2019) 190039. https://zenodo.org/records/2547147.

[12] K.T. Tapani, S. Vanhatalo, N.J. Stevenson, et al., Time-varying EEG correlations improve automated neonatal seizure detection, Int. J. Neural Syst. 29 (2019) 1850030, https://doi.org/10.1142/S0129065718500302.

[13] N.J. Stevenson, K.T. Tapani, L. Lauronen, et al., The effect of reducing EEG electrode number on the visual interpretation of the human expert for neonatal seizure detection, Clin. Neurophysiol. 128 (2017) e418, https://doi.org/10.1016/j.clinph.2017.09.020.

[14] J.M. O'Toole, S.R. Mathieson, S.A. Raurale, et al., Neonatal EEG graded for severity of background abnormalities in hypoxic-ischaemic encephalopathy, Sci. Data 10 (2023) 129, https://doi.org/10.1038/s41597-023-020028. https://zenodo.org/records/7477575#.ZH8o89LMKic.

[15] R. Hogan, S.R. Mathieson, A. Luca, et al., Scaling convolutional neural networks achieves expert level seizure detection in neonatal EEG, NPJ. Digit. Med. 8 (2025) 17, https://doi.org/10.1038/s41746-024-01416-x.

[16] T. Skoric, D. Bajic, Epileptic seizures prediction of pediatric subjects in health-care pervasive environments, In: IEEE Int Conf Commun; (2023) 1690–1694.
[17] T. Proix, et al., Forecasting seizure risk in adults with focal epilepsy: a development and validation study, Lancet Neurol. 20 (2021) 127–135, https://doi.org/10.1016/S1474-4422(20)30396-3.
[18] A. Pandey, S.K. Singh, S.S. Udmale, KK. Shukla, An intelligent optimized deep learning model to achieve early prediction of epileptic seizures, Biomed. Signal. Process. Control 84 (2023) 104798, https://doi.org/10.1016/j.bspc.2023.104798.
[19] J. Kim, E. Amorim, V.R. Rao, H.C. Glass, D. Bernardo, Short-horizon neonatal seizure prediction using EEG-based deep learning, PLOS. Digit. Health 4 (2025) e0000890, https://doi.org/10.1371/journal.pdig.0000890.
[20] D. Bernardo, J. Kim, M-C Cornet, et al., Machine learning for forecasting initial seizure onset in neonatal hypoxic–ischemic encephalopathy, Epilepsia 66 (2025) 89–103, https://doi.org/10.1111/epi.18163.
[21] M. Bandarabadi, J. Rasekhi, C.A. Teixeira, et al., On the proper selection of preictal period for seizure prediction, Epilepsy Behav. 46 (2015) 158–166, https://doi.org/10.1016/j.yebeh.2015.03.010.
[22] W. Stacey, M. Le Van Quyen, F. Mormann, A. Schulze-Bonhage, What is the present-day EEG evidence for a preictal state? Epilepsy Res. 97 (2011) 243–251, https://doi.org/10.1016/j.eplepsyres.2011.07.012.
[23] R. Statello, L. Carnevali, A. Sgoifo, et al., Heart rate variability in neonatal seizures: investigation and implications for management, Neurophysiol. Clin. 51 (2021) 483–492, https://doi.org/10.1016/j.neucli.2021.10.002.
[24] I. Bersani, F. Piersigilli, D. Gazzolo, et al., Heart rate variability as possible marker of brain damage in neonates with hypoxic ischemic encephalopathy: a systematic review, Eur. J. Nucl. Med. Mol. Imaging 180 (2021) 1335–1345, https://doi.org/10.1007/s00431-020-03882-3.
[25] M. Lucchini, W.P. Fifer, R. Sahni, MG. Signorini, Novel heart rate parameters for the assessment of autonomic nervous system function in premature infants, Physiol. Meas. 37 (2016) 1436–1446, https://doi.org/10.1088/0967-3334/37/9/1436.
[26] R. Statello, L. Carnevali, D. Alinovi, F. Pisani, A. Sgoifo, Heart rate variability in neonatal patients with seizures, Clin. Neurophysiol. 129 (2018) 2534–2540, https://doi.org/10.1016/j.clinph.2018.10.001.
[27] Fontanet LN, Pimpel B, Lawrence N, Boylan G, Fons C, Pressler R. Management of neonatal seizures: insights from EPICARE study survey. *Abstracts of the 16th European Paediatric Neurology Society Congress (EPNS); 2025 Jul 8–12; Germany*. Abstract book, p. 104/963.
[28] C.J. Wusthoff, R.A. Shellhaas, RR. Clancy, Limitations of single-channel EEG on the forehead for neonatal seizure detection, J. Perinatol. 29 (2009) 237–242, https://doi.org/10.1038/jp.2008.195.
[29] T. Skoric, M. Djermanovic, J. Kljajic, S. Spasojevic, J.M. O'Toole, Early prediction of electrographic seizures in neonatal hypoxic-ischemic encephalopathy based on amplitude-integrated EEG and clinical data, in: Proceeding of 32nd European Signal Processing Conference (EUSIPCO), 2024, pp. 1581–1585.
[30] K. Konstantinides, K. Yao, Statistical analysis of effective singular value in matrix rank determination, IEEE Trans Acoust Speech Signal Process 36 (1988) 757–763, https://doi.org/10.1109/29.1585.
[31] S.M. Usman, S. Khalid, R. Akhtar, et al., Using scalp EEG and intracranial EEG signals for predicting epileptic seizures: review of available methodologies, Seizure 71 (2019) 258–269, https://doi.org/10.1016/j.seizure.2019.08.006.
[32] L. Li, W. Chen, X. Shao, Z. Wang, Analysis of amplitude-integrated EEG in the newborn based on approximate entropy, IEEE Trans. Biomed. Eng. 57 (2010) 2459–2466, https://doi.org/10.1109/TBME.2010.2055863.
[33] J.M. Yentes, N. Hunt, K.K. Schmid, J.P. Kaipust, D. McGrath, N. Stergiou, The appropriate use of approximate entropy and sample entropy with short data sets, Ann. Biomed. Eng. 41 (2013) 349–365, https://doi.org/10.1007/s10439-012-0668-3.
[34] M. Błażkiewicz, J. Kędziorek, A. Hadamus, The impact of visual input and support area manipulation on postural control in subjects after osteoporotic vertebral fracture, Entropy 23 (2021) 375, https://doi.org/10.3390/e23030375.
[35] T. Skoric, O. Mohamoud, B. Milovanovic, N. Japundzic-Zigon, D. Bajic, Binarized cross-approximate entropy in crowdsensing environment, Comput. Biol. Med. 80 (2017) 137–147, https://doi.org/10.1016/j.compbiomed.2016.11.019.
[36] B. Olmi, C. Manfredi, L. Frassineti, et al., Heart rate variability analysis for seizure detection in neonatal intensive care units, Bioengineering 9 (2022) 165, https://doi.org/10.3390/bioengineering9040165.
[37] L. Frassineti, A. Lanatà, B. Olmi, et al., Multiscale entropy analysis of heart rate variability in neonatal patients with and without seizures, Bioengineering 8 (2021) 122, https://doi.org/10.3390/bioengineering8090122.
[38] B. Selvaraj, E. Rajasekar, JBB. Rayappan, Machine learning approaches: detecting the disease variants in human-exhaled breath biomarkers, ACS. Omega 9 (2024) 215–226, https://doi.org/10.1021/acsomega.3c03755.
[39] A. Odabaee, W.J. Freeman, P.B. Colditz, et al., Spatial patterning of the neonatal EEG suggests a need for a high number of electrodes, Neuroimage 68 (2013) 229–235, https://doi.org/10.1016/j.neuroimage.2012.11.062.
[40] Kljajic, J., O'Toole, J.M., Hogan, R., Skoric, T., 2025. Honest and reliable evaluation and expert equivalence testing of automated neonatal seizure detection. *arXiv preprint*, arXiv:2508.04899.
[41] A. O'Shea, G. Lightbody, G.B. Boylan, A. Temko, Neonatal seizure detection from raw multi-channel EEG using a fully convolutional architecture, Neural Netw. 123 (2020) 12–25, https://doi.org/10.1016/j.neunet.2019.11.023.
[42] D.Y. Isaev, A. Temko, G. Lightbody, G. Boylan, Attention-based network for weak labels in neonatal seizure detection, in: Machine Learning for Healthcare Conference, 2020, pp. 479–507. September 2020.
[43] McDermott MB, Zhang H, Hansen LH, Angelotti G, Gallifant J. A closer look at AUROC and AUPRC under class imbalance In: Proc 38th Conf Neural Inf Process Syst (NeurIPS); 2024. p. 1–59.
[44] SPEED GitHub repository. https://github.com/cergenx/SPEED.
[45] F. Magarelli, G.B. Boylan, S. Montazeri, F. O'Sullivan, D. Lightbody, M. Ashoori, T. Skoric, J.M. O'Toole, Machine-learning competition to grade EEG background patterns in newborns with hypoxic-ischaemic encephalopathy, *arXiv preprint* (2025) arXiv:2509.09695.
[46] A. Gramacki, J. Gramacki, A deep learning framework for epileptic seizure detection based on neonatal EEG signals, Sci. Rep. 12 (2022) 13010, https://doi.org/10.1038/s41598-022-15830-2.
[47] R. Elakkiya, H. Chandra, N.E. Pears, et al., Lung image quality assessment and diagnosis using generative autoencoders in unsupervised ensemble learning, Biomed. Signal. Process. Control 102 (2025) 107268, https://doi.org/10.1016/j.bspc.2024.107268.
[48] R. Elakkiya, S. Vairavasundaram, K. Kotecha, N. Herencsar, SignCraft: transformative learning based sign gesture generation for consumer electronics, IEEE Trans. Consumer Electr. 72 (2026) 3601–3613, https://doi.org/10.1109/TCE.2026.3658738.
[49] L. Kuhlmann, P. Karoly, D.R. Freestone, et al., Epilepsyecosystem.Org: crowd-sourcing reproducible seizure prediction with long-term human intracranial EEG, Brain 141 (2018) 2619–2630, https://doi.org/10.1093/brain/awy210.
[50] V. Srinivasan, C. Eswaran, N. Sriraam, Approximate entropy-based epileptic EEG detection using artificial neural networks, IEEe Trans. Inf. Technol. Biomed. 11 (2007) 288–295, https://doi.org/10.1109/TITB.2006.884369.
[51] Mark E. O'Sullivan, Gordon Lightbody, Sean R. Mathieson, William P. Marnane, Geraldine B. Boylan, John M. O'Toole, Development of an EEG artefact detection algorithm and its application in grading neonatal hypoxic-ischemic encephalopathy, Expert. Syst. Appl. 213 (2023) 118917.