applied sciences

MDPI

Article

Validation of an Automated Seizure Detection Procedure for Multi-Channel Neonatal EEG

Cris Micheli [1,*], Antonia Thelen [2], Maarten De Vos [3], Anneleen Dereymaeker [3], Jens Haueisen [4,5] and Patrique Fiedler [4,*]

1 NeuroTechX Chapter, 12161 Berlin, Germany
2 Department of Psychology, Brandenburg Medical School, 16816 Neuruppin, Germany
3 Department of Development & Regeneration, Katholieke Universiteit Leuven, 3000 Leuven, Belgium; anneleen.dereymaeker@uzleuven.be (A.D.)
4 Institute of Biomedical Engineering and Informatics, Technische Universität Ilmenau, 98693 Ilmenau, Germany
5 Biomagnetic Center, Department of Neurology, University Hospital Jena, 07747 Jena, Germany
* Correspondence: cris.micheli@neurotechx.com (C.M.); patrique.fiedler@tu-ilmenau.de (P.F.); Tel.: +49-367-769-2865 (P.F.)

Abstract

This study validates an automated seizure detection algorithm for multi-channel neonatal EEG, adapting a previously published method to a dataset with fewer electrodes. The Python-based implementation, SDApy, was applied to EEG recordings from 23 neonates to classify seizure and non-seizure epochs using a support vector machine trained on an independent dataset. The algorithm employs time- and frequency-domain features and maintains high generalization across different recording setups, achieving robust performance despite using only nine electrodes instead of nineteen. Evaluation metrics, including F1 scores and precision—recall curves, confirmed strong agreement between algorithm predictions and expert annotations for most patients. SDApy's open-source implementation enhances accessibility compared with earlier MatLab versions, offering a transparent and cost-effective approach to clinical EEG analysis. The pipeline can operate with labels from a single expert, supports data pre-labeling for deep learning, and integrates well into neonatal intensive care unit monitoring workflows. Overall, SDApy demonstrates reliable adaptation to reduced-channel EEG and shows potential for real-time seizure detection, personalized threshold optimization, and integration into multimodal neurophysiological monitoring systems.

Keywords: electroencephalography; neonatal seizures; seizure detection; machine learning; signal processing

Check for updates

Academic Editors: Agnieszka Duraj and Piotr S. Szczepaniak

Received: 24 October 2025
Revised: 15 December 2025
Accepted: 16 December 2025
Published: 20 December 2025

Copyright: © 2025 by the authors. Licensee MDPI, Basel, Switzerland. This article is an open access article distributed under the terms and conditions of the Creative Commons Attribution (CC BY) license.

1. Introduction

Continuous electroencephalographic (EEG) monitoring in neonatal intensive care units (NICUs) is essential for diagnosing neurological pathologies and associated risks due to severe conditions [1–3]. EEG provides crucial information to clinicians, aiding in monitoring brain development and identifying neurological disorders [2,4,5]. Early detection and intervention are vital due to the rapid structural and functional changes in the neonatal brain, helping to avoid or minimize long-term adverse effects [6–8] and supporting positive neurodevelopmental outcomes [9,10]. Electrographic seizures, both clinical and subclinical, are among the most common neurological emergencies in the

NICU, with an incidence rate ranging from 1.5 to 5.5 per 1000 live births [11]. These seizures and their underlying causes can significantly impact brain development and have a high mortality rate ranging from 20 to 45% [8,12]. Seizures are typically identified through long-term EEG recordings reviewed by medical experts, and neonatal seizures are defined as electrographic events with sudden, paroxysmal, and abnormal alterations of activity with a minimal length of 10 s [13,14]. Despite being the gold standard, this manual review process has several limitations:

1. Time-consuming and labor-intensive for medical experts, reducing availability for on-site care;
2. Subjective interpretation and classification [15,16];
3. Clinical inter-rater agreement difficulties [17];
4. Risk of delayed diagnosis, impacting therapeutic outcomes [11];
5. Post hoc diagnosis can be affected by misleading signal artifacts, especially without synchronous video acquisition [16,18].

Seizure detection algorithms on noninvasive EEG often use various signal processing and machine learning techniques, including threshold-based detection [19,20], frequency-based methods [21,22], nonlinear dynamics [23], random forests [24,25], neural networks [26–28], k-nearest neighbors or KNN [29], hidden Markov models [30], and support vector machines [31,32]. Temko and colleagues [33] proposed a personalized neonatal seizure detection system combining patient-adaptive generative and patient-independent discriminative classifiers, demonstrating improved seizure detection rates of up to 70% in a clinical dataset of over 816 h of EEG recordings from 18 full-term newborns. More recently, advanced deep learning approaches have shown promising results in neonatal seizure detection [27,34,35]. For example, O'Shea and colleagues [28] developed a fully convolutional neural network for detecting seizures from raw multi-channel neonatal EEG, achieving high performance with an area under the receiver operating characteristic curve of 98.5%.

Although impressive in classifier performance, these methods need tedious manual work to label the seizure events in the pre-training phase of the algorithm. Additionally, neural networks, though effective at capturing long-term dependencies, are at risk of being overfit, complex, and computationally expensive, requiring large datasets, whereas threshold- and frequency-based methods are simple to implement but, on the other hand, may miss non-rhythmic or subtle seizures and are sensitive to noise. KNNs are easy to implement but may be sensitive to irrelevant features and inefficient for large datasets. Wavelet methods, while effective for transient seizures, are computationally demanding and complex. Markov models are good for modeling state transitions but may miss long-term interactions. Random forests handle high-dimensional spaces well and are less prone to overfitting, but they can become complex and hard to interpret. Support vector machines are effective in high-dimensional spaces with a clear margin of separation but require careful kernel selection (we selected a Gaussian kernel; see the Section 2). These advantages and disadvantages are consistent with those summarized in a recent general review of machine-learning algorithms [36,37].

To address the constraints hindering the adoption of such algorithms in daily clinical routine, Tapani and colleagues proposed an algorithm using time and frequency domain EEG correlates to train a multi-parametric support vector machine for classifying neonatal EEG recordings into seizure or non-seizure epochs [32]. This method overcomes many disadvantages of other automatic seizure detection algorithms previously used in the literature, offering a more robust and reliable approach. Specifically, this offline algorithm selects a specific montage from the classic 19-electrode 10–20 layout, using 18 bipolar adjacent channels in geodesic frontal-occipital directions. This approach aims to increase local SNR

by reducing distant noise and allows for a systematic reduction in channels to understand the impact on reliability and clinical practice. Additionally, because of the bipolar derivation noise-reduction properties, the technique is robust to slightly different bipolar montages [16,38,39], ensuring its adaptability and effectiveness across various setups.

To address the need for seizure detection in the NICU, we adapted Tapani and colleagues' offline MatLab 2017a algorithm (The MathWorks Inc., https://www.mathworks.com) for automatic EEG processing and implemented it in Python 3.11. We tested the algorithm on new data collections using a Python-based pipeline (Figures 1 and 2), thereby demonstrating successful external validation of the classifier without adjusting any hyperparameters. This paper highlights that the originally trained model can be effectively extrapolated and applied across various practical contexts.

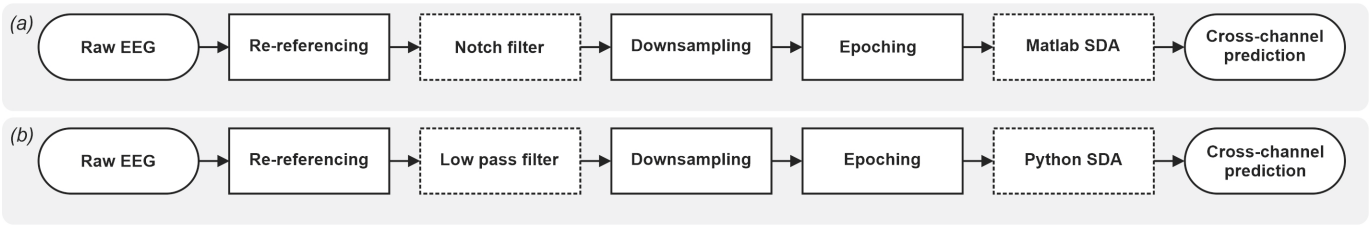


Figure 1. Schematic comparison of MatLab and Python seizure detection pipelines. Panel (**a**) illustrates the original MatLab implementation described by Tapani et al. [32], which applies a 50 Hz notch filter but no additional low-pass filtering or post-processing across channels. Panel (**b**) illustrates the current Python (SDApy) implementation, which omits the 50 Hz notch filter, includes a 32 Hz low-pass filter, and applies additional post-processing by aggregating the SVM decision function across channels and over time before thresholding (see also Figure 2). For both pipelines: re-referencing applies a bipolar montage (see Figure 3) to the raw signals, Downsampling resamples data from 250 to 64 samples/s. Epoching refers to the process of data aggregation before feature calculation (28 s overlap and 4 s epochs or time segments). This highlights three key differences between the pipelines: (1) notch filtering (MatLab only), (2) low-pass filtering (Python only), and (3) multi-channel/post-processing aggregation (Python only).

In brief, the goals of this study are to demonstrate that only minor extensions are required for the algorithm's efficacy to generalize to a different neonatal EEG dataset recorded at another site and with a different recording setup, which features a lower channel count and different 10–20 bipolar derivations. Additionally, the study aims to suggest that this technique can be integrated into AI and deep learning pipelines for automatic neonatal seizure detection, serving as a useful online pre-labeling tool for NICU monitoring systems and thereby reducing the time required for manual expert labeling. Finally, by translating the pipeline from MatLab to Python, the study makes this tool freely available and more accessible, encouraging broader adoption within the clinical and research communities.

In the Section 3, we will present a comprehensive multi-parameter performance analysis of the original trained Support Vector Machine (SVM) model first established by Temko and colleagues [40] on new data to show its seizure detection properties, and in the Section 4, we will comment on the extensibility and generalizability of this method and its clinical use practicalities.

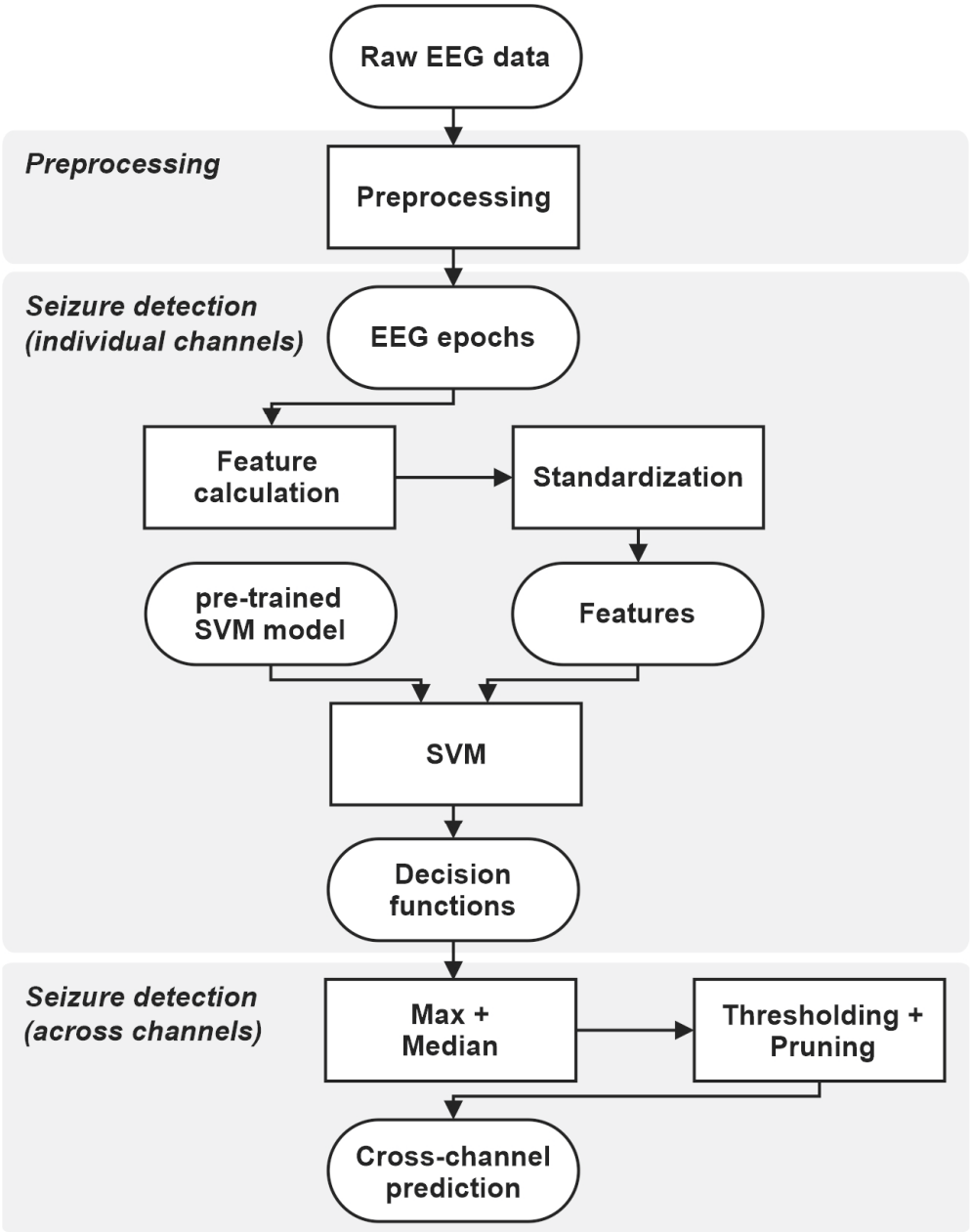


Figure 2. Flowchart of the seizure detection algorithm implementation (SDApy). Preprocessing as in Figure 1. EEG Epochs represent 4 s distanced time windows, each 32 s long, which are superimposed with an overlap of 87.5% (28 out of 32 s). Standardization refers to the process of subtracting the mean and dividing by the std of Helsinki's data (see Section 2.3). The Max of all channels' decision functions is performed across channels, then the median over three epochs of the resulting aggregate function is calculated for each epoch. Thresholding is applied to the resulting time function, and isolated <10 s episodes above threshold are excluded (Pruning).

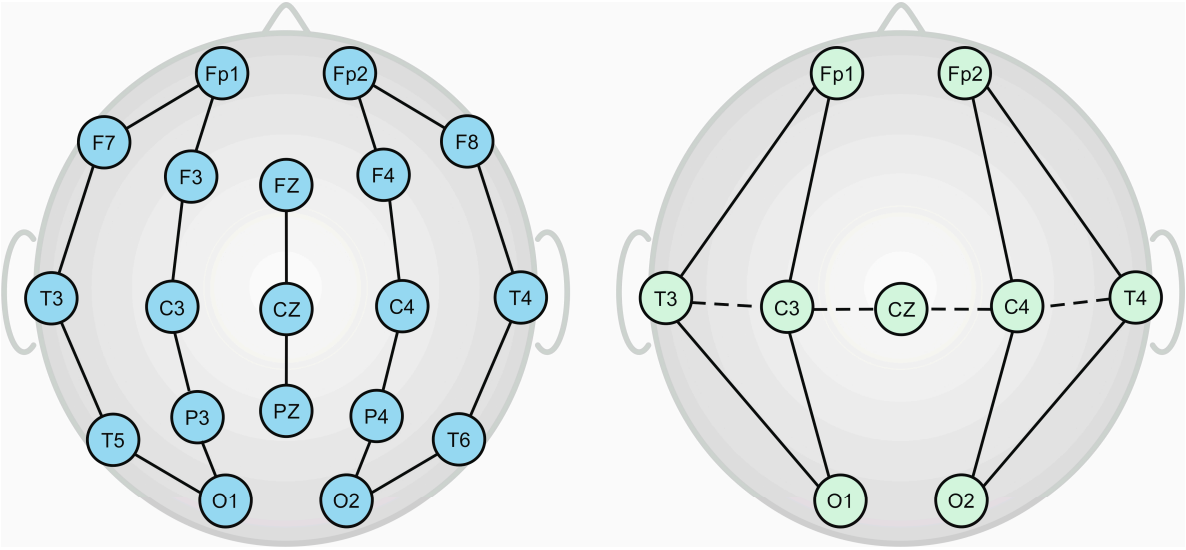


Figure 3. Bipolar EEG montages used in training and validation datasets. **Left**: bipolar montage used in the pre-trained model. **Right**: bipolar montage used in this study. The latter montage shows a reduced channel count. The dark edges represent the bipolar channel pairs. The dashed edges in the second layout represent 4 additional bipolar derivations adopted in this work. These derivations are used for neonates, as this area covers the watershed regions, the most vulnerable regions for brain damage [2].

2. Methods

2.1. Data Recordings

For the automatic seizure classification, we utilize the model trained on the publicly available dataset of Stevenson and colleagues [39] to validate the seizure detection algorithm. The trained model's clinical data was obtained at HUS Medical Imaging Center, Helsinki University Hospital, Finland, and consists of multi-channel EEG recordings from 79 neonatal patients with a total duration of approximately 112 h (see Table 1). The datasets are available at https://doi.org/10.5281/zenodo.1280683 [39]. The dataset comprises annotations by three independent experts on a per-channel basis. We present the description of these data for comparison (see Table 1) and to illustrate the type of data utilized in training the model for the seizure-prediction procedure. However, we employed only the Helsinki study's [32] pre-trained model and not the raw datasets.

Table 1. Characteristics and comparison of the datasets used for model training and validation of the automatic seizure detection procedure.

	Model Training Dataset	Validation Dataset
Clinical sites	1 (Helsinki)	1 (Leuven)
No. of recordings	79	23
No. bipolar channels (electrodes)	18 (19)	12 (9)
No. of patients (with seizures)	79 (45)	23 (23)
No of off-label recordings (*)	1	23
Overall duration (hours)	112	44
No. of annotated seizures	460 (+)	168
Median dataset duration (hours)	1	2
No. of expert reviewers	3	2

(*) Tapani and colleagues used a leave-one-out method to train the models, which resulted in all 79 datasets being used alternatively in the training. The validation dataset makes use of Tapani's SVM model; hence, none of the new datasets were used for the training. (+) In the case of Tapani and colleagues, this is an average across expert reviewers.

An independent, unseen second dataset was used for the validation. A total of 23 neonates were enrolled from the NICU of University Hospital, Leuven. They presented with seizures in the neonatal period, and the cohort of continuous EEG monitoring was recorded during the period January 2013–July 2015 as part of standard clinical care. 65% of these patients had acute symptomatic seizures (e.g., Hypoxic–ischemic encephalopathy or HIE, stroke, hemorrhage, brain infection). However, some neonates with suspected metabolic and genetic causes were also included.

The datasets were recorded at 250 Hz using Brain RT EEG systems (Micromed NATUS, BVBA, Rumst, Belgium). Nine electrodes were applied according to the restricted 10–20 system (Fp1, Fp2, T3, T4, C4, C3, Cz, O2, O1). These derivations are commonly used for neonates [2,41], as this area encompasses the watershed regions, which are the most vulnerable to brain damage and seizures (see Figure 3). All recordings were filtered in the annotation software between 1 and 20 Hz before visual expert review. EEGs were interpreted by a clinical neurophysiologist (EEGs were scored visually at 15 mm/s) as part of clinical care, and anti-epileptic drug (AED) treatment was initiated and followed according to the local hospital protocol. The study was conducted with approval from the Ethical Review Board in Leuven. All recordings were provided fully anonymized, following approval by the responsible Ethical Review Boards of the clinical site and written informed consent from the patient's parents.

The seizure classification algorithm was validated on the clinical recordings from the University Hospital of Leuven (multi-channel EEG recordings from 23 neonatal patients, with a total duration of approximately 44 h). The Leuven data (rightmost column, Table 1) was reviewed by two medical experts, and seizure events were annotated per channel. Labeled events shorter than 10 s were successively discarded for compatibility with the abovementioned definition of neonatal seizure [13,15]. During annotating, the expert had access to additional physiological recordings (ECG, EMG, respiration) and video data, which was consulted for annotations but not part of the seizure detection pipeline. A comparative overview of the key characteristics of the trained model's recordings and the validation datasets is provided in Table 1.

2.2. Algorithms

We provide an overview of the two compared Seizure Detection Algorithms (SDA) in Figure 1, which depicts in functional blocks the original MatLab algorithm (Figure 1a), by Tapani and colleagues [32], found in the following GitHub repository: https://github.com/ktapani/Neonatal_Seizure_Detection (accessed on 14 December 2025), https://doi.org/10.5281/zenodo.1280683 and the newly adapted Python algorithm (Figure 1b), applied offline to the dataset employed in the present work. In the discussion, we give recommendations for an online pipeline implementation of our Python pipeline. Both implementations use the same SVM method but differ in the preprocessing and postprocessing (see Figures 1 and 2) since the Python implementation is more efficient.

Throughout the rest of the Section 2, we describe in detail the Python implementation of the SDA algorithm (Figure 1b). We selected Python for the implementation of the seizure detection algorithm because of its accessibility, free licensing scheme, and the vast presence of machine learning libraries. We called the new implementation SDApy (Seizure Detection Algorithm for Python). In the next section, we describe the details of the SDApy pipeline.

The attentive reader might notice that this preprocessing part of the pipeline (first four modules of Figure 1) is kept lean and minimalistic. This was done by design to allow for translation of the full algorithm into real-time automatic seizure detection applications for the NICU (see Section 4). The translational part of this algorithm is not analyzed in detail in this paper; however, we emphasize that the main requirement of such real-world

algorithms is speed of execution, and therefore, the processing has to be optimized for time efficiency on medical hardware.

2.3. Seizure Detection Algorithm for Python (SDApy)

A comprehensive overview of the SDApy algorithm is provided in Figure 2. The details of the MatLab–Python algorithmic differences are depicted in Figures 1 and 2, as well as in Table 2, and described in the following paragraphs.

Table 2. Key differences between the original MatLab implementation and the current Python (SDApy) implementation.

Aspect	MatLab Implementation (Tapani et al. [32])	Python Implementation (SDApy)
Implementation language	MatLab	Python
Notch filter at 50 Hz	Applied	Not applied
Low-pass filter	Not applied	32 Hz low-pass filter applied
Feature extraction	Spiking and time–frequency features	Same
Classifier	SVM with Gaussian kernel	Same
Decision function	Per channel	Same
Post-processing across channels	Not applied	Maximum of decision function across channels
Temporal aggregation	Not applied	Median of decision function over 3 consecutive epochs and pruning < 10 s episodes before thresholding
Output	Channel-wise seizure/non-seizure labels	Aggregated seizure/non-seizure labels

Initially, individual analysis epochs of 32 s in length and 28 s overlap (87.5% overlap) were extracted for all channels from the overall recording, resulting in a timeline of 1 epoch every 4 s. After re-referencing to a bipolar montage, the EEG data were low-pass filtered (5th order Butterworth low-pass filter with 32 Hz cut-off, see Figures 1b and 2, preprocessing) and resampled to 64 samples per second, using SciPy's 'resample' method. This method employs a trigonometric frequency resampling method, which samples the Fourier circle at equally distant points. These filtering parameters are different from the original MatLab pipeline (Figure 1a) and were chosen to prevent aliasing during the resampling operation. Additionally, a 32 Hz low-pass filter sufficiently reduced the 50 Hz interference. Therefore, the Python pipeline does not apply an additional notch filter afterward (as in the MatLab pipeline) to avoid signal distortions.

All analysis features (see Table 3) were calculated according to the original MatLab pipeline [32] and include a high-pass filtering step at 1 Hz. The 21 features describe the spikiness (features 1 to 10) and the time-frequency characteristics (features 11 to 21) of each EEG epoch. Specifically, the features are shown in Table 3.

Table 3. List of all features calculated per channel and per epoch.

Features Type	Description
Spike and SNLEO-based features (1–10) (SNLEO: smoothed nonlinear energy operator)	(1) Skewness and (2) variability of the SNLEO output, (3) number of spikes per epoch, (4) median spike width and (5) inter-spike interval, (6) mean and (7) variability of spike correlations, (8) mean and (9) standard deviation of the SNLEO, and (10) mean spike peak intensity relative to background
Time–frequency correlation features (11, 21)	(11) Median correlation between scaled time-slices of the time-frequency distribution (TFD) and (21) sum of the maximum TFD of each slice divided by the total TFD
Amplitude envelope and broadband power (12, 13)	(12) Mean amplitude envelope and (13) natural logarithm of spectral power in the 0.5–30 Hz band
Harmonic power features (14–16)	(14) Power in the first three harmonics relative to total PSD, (15) power in the fundamental relative to total PSD, and (16) logarithm of power in the fundamental
Relative band power features (17–20)	(17) Relative delta (0.5–4 Hz), (18) theta (4–8 Hz), (19) alpha (8–12 Hz), and (20) beta (12–30 Hz) band powers
Normalized Max TFD (21) (TFD: time-frequency distribution)	Sum of the maximum TFD of each slice divided by the sum of the TFD

The resulting features are a tensor with 3 dimensions: bipolar channels (12), features (21), and epochs (1800 on average; see Table 1 and the Supplementary Material). The standardization step ensures that the newly calculated features are divided by the STD of the training dataset's features after demeaning the average of Tapani's features.

The SVM classifier uses a pretrained model trained on the Helsinki datasets (see Table 1). The original features were calculated for each epoch and channel, with 18 bipolar channels taken from 19 electrodes distributed in a 10–20 layout (see Figure 3). In comparison, the SDApy pipeline uses datasets of 12 bipolar channels, with a 10–20 subset configuration of only 9 electrodes (Figure 3). Note that all 9 electrodes were also used in the original publication; however, the bipolar channels of this study are chosen from a subset of the original 10–20 layout; therefore, inter-electrode distances are increased (see Figure 3), except for the newly introduced derivations: T3-C3, C3-Cz, Cz-C4, and C4-T4.

The output of the SVM block (Figure 2) is a decision function calculated per epoch and channel, which represents, after thresholding, the class attribution (a seizure epoch with label '1' if the threshold is crossed, or a non-seizure epoch with label '0'). Since the classifier operates on the individual channel level, no theoretical restrictions regarding minimum and maximum channel numbers exist. The decision function of the SVM is defined according to Equation (1):

$$Decision_{ep,ch} = \sum_{j=1}^{n} \alpha_j y_j G\left(\vec{f}_{ep,ch}, \vec{x}_j\right) + b \quad (1)$$

Herein, $Decision_{ep,ch}$ indicates the decision function for epoch ep and channel ch; α_j denotes the Lagrangian dual coefficient for support vector j (used to make a constrained maximization problem unconstrained and solvable by means of Lagrangian multipliers, called dual variables), b is the model bias coefficient, and G is the SVM's Gaussian kernel function, $\vec{x}_j$ the support vector j, $\vec{f}_{ep,ch}$ the features vector for epoch ep and channel ch, y_j the support vector label (either -1 or 1), and n is the number of support vectors. The kernel is

defined according to Equation (2), with $\left\| \vec{f_{ep,ch}} - \vec{x_j} \right\|^2$ being the Euclidean distance between the support vector $\vec{x_j}$ and the features vector $\vec{f_{ep,ch}}$, and ks being the kernel scaling factor:

$$G\left(\vec{f_{ep,ch}}, \vec{x_j}\right) = \exp\left(-\frac{\left\| \vec{f_{ep,ch}} - \vec{x_j} \right\|^2}{ks^2}\right) \quad (2)$$

The selection of the Gaussian kernel for SVM training in Tapani and colleagues was driven by its strong capability to handle non-linear, complex, and noisy epileptic EEG data [42–44], and we applied the same logic here. Finally, a decision threshold t_d was applied to the calculated likelihood to derive the binary classification. This pre-set threshold was shown to maximize the inter-observer agreement of the SDA on the previous verification dataset used by Tapani and colleagues and is recommended for the online version of the algorithm (a modified version of Figure 1b, see Section 4), whereby threshold tuning is not possible. We made use of the same threshold for the SDApy pipeline to show that it also generalizes well in our offline case (see the Section 3). The model parameters x_j and b, and hyperparameters α_j, t_d, ks, and y_j are derived from the pre-trained model inherited from Tapani and colleagues and available on Zenodo (https://doi.org/10.5281/zenodo.1280683), whereas the features are calculated from the new dataset.

The final steps of SDApy apply additional processing on the decision function by deriving the maximum across all channels and calculating the median over 3 epochs, before thresholding. The decision function is compiled by aggregating channel-wise SVM outputs, using a combination of maximum across channels and temporal median, to account for the spatial variability and propagation of seizure activity. This approach increases sensitivity by leveraging the distributed nature of seizure foci and underlying brain connectivity and avoids a simple winner-takes-all strategy by employing weighted voting across all channels.

Finally, after thresholding and hence determining the class of seizure (1) and non-seizure (0) per epoch and per channel, we converted epochs to seconds (upsampling the decision function to 1 sample/s, i.e., repeating the samples of each epoch 4 times) for compatibility with the annotation timeline (defined in seconds). Thresholding was performed on the upsampled decision function using the threshold from the pre-trained model, which is a hyperparameter determined during training. Events shorter than 10 s were removed.

2.4. Model Performance Assessment

For the validation of the SDApy pipeline, we compared the output of the classification results of SDApy to the ground truth, which was the respective expert evaluation. We employed temporal event-based statistics [20,31], which is a comparison between expert assessment and classification output (Figure 4), to generate the confusion matrix and its derivatives (see Supplementary Material). Specifically, in terms of temporal statistics, we compare the annotations of the clinical expert against the predictions of the SDApy for each second of a given recording, grouping false positives, false negatives, true positives, and true negatives. The ultimate goal of an SDA is to provide seizure predictions that are indistinguishable from the current "gold standard", i.e., the human expert. In Figure 4, we represent the criterion used to generate the confusion matrix for each patient.

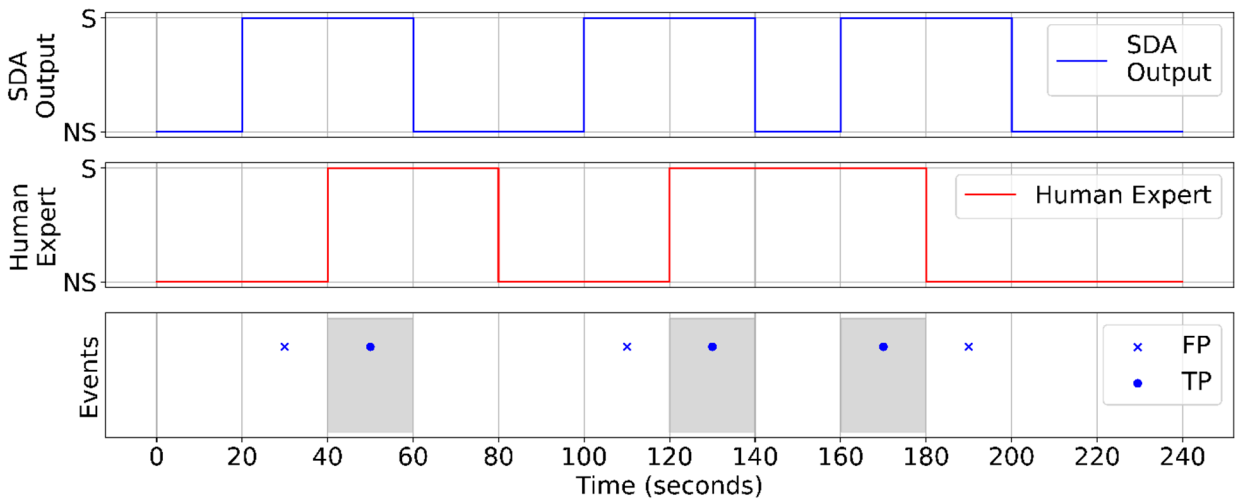


Figure 4. Temporal mapping of expert annotations and SDApy decision function for confusion matrix generation. On the *y*-axis: S—Seizure, NS—No Seizure, FP—false positive, TP—true positive) for every single patient. The confusion matrix (lower track) is determined by comparing the thresholded maximum decision function across channels (blue line, top track) and the logic OR across channels for the expert annotations (red line, middle track). Abbreviations: SDA: Seizure Detection Algorithm, FP: False Positives, TP: True Positives.

The F1 score was selected as a quality parameter because it performs well when the ratio between positive and negative labels (0 in this case) is asymmetric [45,46]. It is also recommended that the cost of false positives and false negatives be equally high, since it accounts for both precision and recall [47,48]. The F1 score metric depends on the decision function's threshold. To calculate all threshold-dependent instances of the F1 value, 1000 thresholds are generated, ranging linearly from the minimum to the maximum value of the overall channels' time course (the function's maximum across all channels).

Additionally, we calculated the precision–recall curve, i.e., a graphical representation used to evaluate the performance of a binary classification model, especially when the classes are imbalanced. It shows the trade-off between precision and recall for different threshold settings (see Section 3), corresponding to the same thresholds calculated for the F1 metric.

3. Results

Figure 5 presents SDApy model performance across two exemplary patients: patient 1 representing good performance and patient 14 illustrating poor performance. For patient 1 (top half), the decision function (blue line) remains mostly stable, only occasionally spiking above the decision threshold (red line at 0.641) around the 200 and 5000 s marks. Expert annotations (black lines) show clear, distinct event occurrences during the same periods. False positive and false negative rates remain low throughout, with the cyan track representing baseline true positives and negatives as a flat line, suggesting strong agreement between model predictions and ground truth.

In contrast, patient 14 (lower half) shows much lower performance. The decision function fluctuates heavily throughout, frequently crossing the decision threshold. Expert annotations appear more scattered and less distinct. The cyan track shows many instances of both false positives and false negatives deviating from the flat line, indicating the model is not accurately detecting continuous seizure events.

Figure 6 shows the performance of the SDApy classifier across all patients using two evaluation metrics: F1 scores (top panel) and precision—recall curves (bottom panel). As discussed at the end of the Section 2, these metrics were chosen due to the imbalance

between seizure and non-seizure events. Presenting both allows us to demonstrate the consistency of the results across two complementary performance measures.

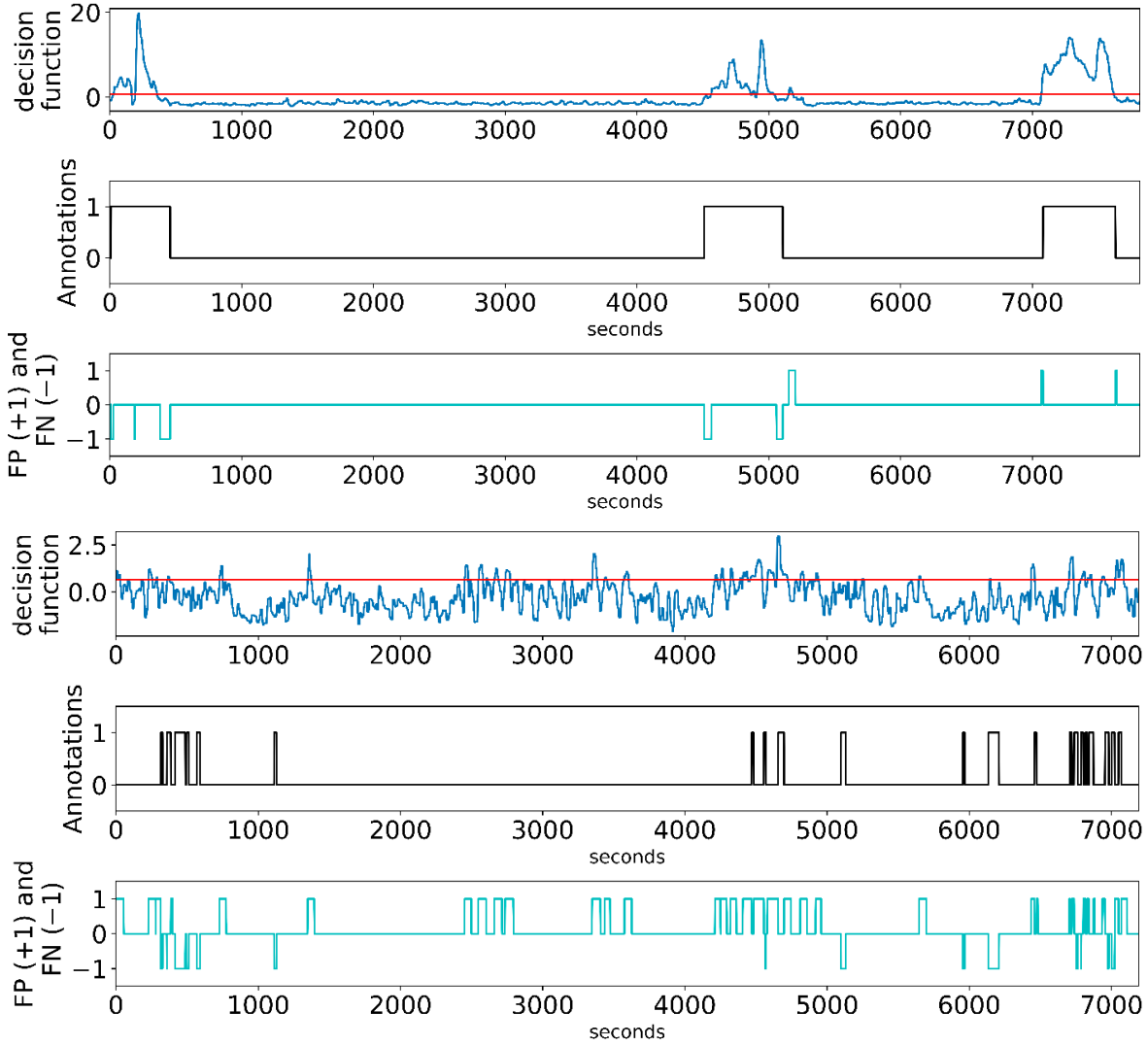


Figure 5. Example SDApy decision functions for a high-performing and a low-performing patient. Above: patient 1 (good performance, F1 metric) and below: patient 14 (low performance). Each panel shows the decision function in blue, the expert annotations in black, and the false positives/false negatives track in cyan. The blue track represents the maximum across all channels of all decision functions, the black track represents the logic OR across all channels of all annotated channels, and the cyan track visualizes the confusion matrix in time with a fixed empirical threshold of 0.641 (this threshold represents a hyperparameter of the pretrained model used for all patients).

In the panel above, the F1 score (a metric balancing precision and recall) is plotted against varying threshold values. Each patient shows different optimal thresholds (look at the peak of the F1 curves), indicating performance variability depending on individual data. The red dashed line marks threshold 0.641, a generalized threshold from the pre-trained model, though the diverse F1 curves show this may not be optimal for all patients. The lower panel shows precision—recall curves for each patient. Some curves show high initial precision with a steep decline as recall increases (low-performing models). Patients 20 and 8 show better overall precision—recall performance, while patients 21, 19, and 14 exhibit low performance with rapidly degrading precision and recall.

Figure 7 displays bar charts with patients sorted by performance metrics (percentage decreasing from left to right): the top chart shows the maximum F1 score per patient,

while the bottom shows the average precision (AP) score, representing the area under the precision—recall curve. Bars are color-coded consistently with Figure 6.

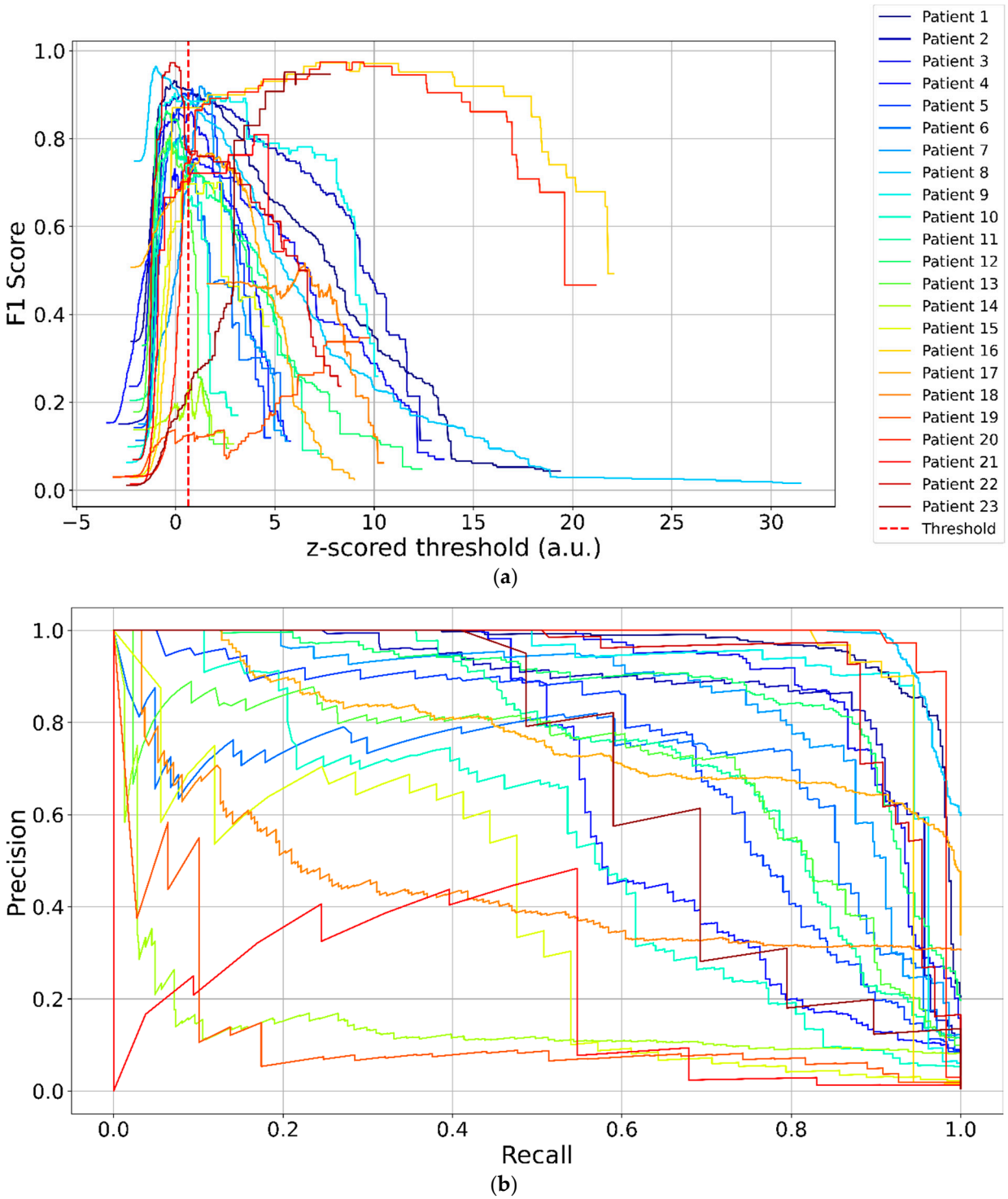


Figure 6. Above: (**a**) Threshold-dependent F1 scores for SDApy across all 23 patients (different colors, see legend). Each curve shows F1 score as a function of 1000 thresholds spanning the global decision-function range, illustrating inter-patient variability in optimal thresholds. The vertical dashed line indicates the fixed threshold 0.641 inherited from the pre-trained model. **Below**: (**b**) Precision–recall curves for SDApy across all 23 patients. Each curve summarizes the trade-off between precision and recall over the same 1000 thresholds used in the upper panel, highlighting patients with consistently high precision and recall versus those with rapidly degrading performance.

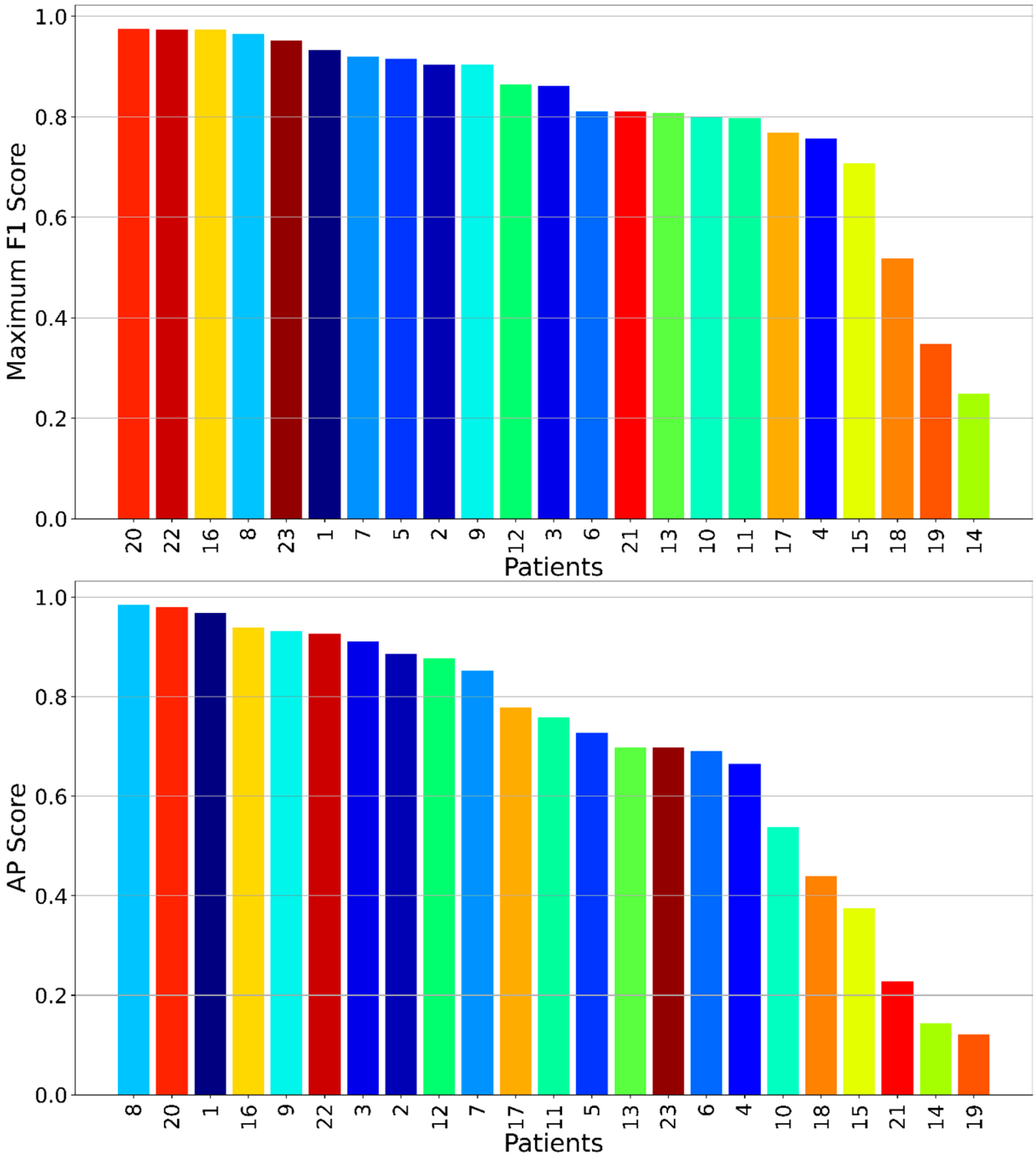


Figure 7. Patient-wise rankings based on F1 and average precision scores. Above: We evaluate and rank the quality of the patients' seizure predictions using the F1 score's peak or maximum value (see Figure 6, upper panel—threshold-dependent F1 score). The leftmost bin is the highest quality, the rightmost the lowest quality. Below: Patients ranked by average precision score. The AP score is the area under the precision—recall curves of Figure 6 (lower panel). The colors represent the same patients as in Figure 6.

In the top panel, patients ranked by maximum F1 score show patients 20, 8, 16, and 9 with excellent performance (F1 > 0.9), indicating highly effective seizure identification with minimal false positives or negatives. Patients 14 and 19 show significantly lower F1 scores (<0.3), indicating the classifier struggles. The bottom panel orders patients by AP score, with similar results: patients 8, 20, 16, and 9 perform well with high AP scores, indicating

maintained precision and recall across threshold ranges. Low-performing patients 14 and 19 again show much lower AP scores.

Figure 8 shows patients ranked by the default fixed threshold's F1 (0.641 in Figure 6, above, from [32]), aiming to visually compare patients in terms of model performance and assess whether a fixed threshold can be used for automatic seizure labeling independently of the performance metric. Comparison with Figures 6 and 7 (upper panel) shows a discrepancy between the ideal F1 value (peak of F1 curves, not settable a priori) and the fixed threshold (settable a priori and suggested as default by Tapani and colleagues).

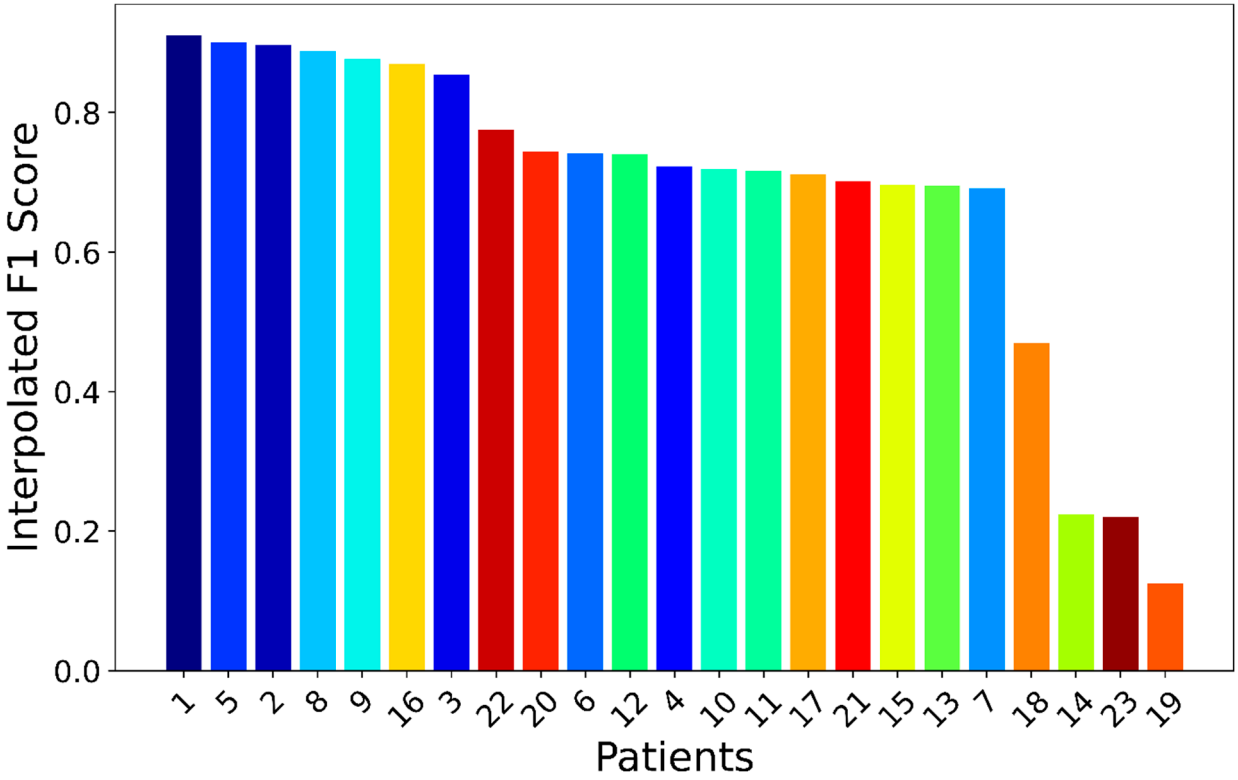


Figure 8. Patient-wise F1 scores at the fixed default threshold. We show the F1 scores for a pre-fixed default threshold value (0.641). Comparing the present rankings with Figure 7, upper panel, will give an impression of classifier results (this figure) with a priori threshold versus the ground truth.

The figure resembles Figure 7 (above) but plots F1 values corresponding to the a priori fixed threshold (intersection of patient curves with red dashed line in Figure 6, upper panel) instead of F1 maxima per patient. Notably, only two patients (16 and 20) show F1 peak values significantly different from the adopted fixed threshold value.

4. Discussion

This study demonstrates the successful adaptation and validation of an automated seizure detection algorithm for neonatal EEG, originally developed by Tapani and colleagues, to a new dataset with fewer channels and different recording conditions. The Python-based implementation (SDApy) showed robust performance across most patients, with some variability in effectiveness.

The per-patient time courses confirm that SDApy can closely track expert annotations in many neonates while still failing in a minority of more challenging cases. Rather than relying on individual examples, the group variability in performance is more clearly quantified by the patient-wise F1 scores and precision—recall curves and rankings, which reveal a group of patients with consistently high detection quality and a smaller group

with low scores, suggesting in both metrics that patient-specific EEG noise characteristics and seizure morphology strongly influence how well the algorithm performs.

F1 scores and precision—recall curves show that SDApy performs well on imbalanced seizure versus non-seizure data, but also that a single global decision threshold is not optimal for all patients. Patient-wise maxima of F1 and average precision indicate that some neonates reach excellent detection quality only at very different thresholds, which supports the need for patient-specific calibration or adaptive thresholds to improve clinical utility. From a clinical perspective, these results suggest that SDApy is suitable as an assistive tool for automated pre-labeling and seizure surveillance in a large fraction of NICU patients, while still flagging cases where algorithm confidence is low. Figure 8, which reports patient-wise F1 scores at the fixed default threshold, further shows that this default operating point already yields acceptable performance for most high-quality recordings, making it a reasonable starting threshold in practice. Future implementations could then refine this starting point via online estimation of patient-specific performance (e.g., from short calibration segments) to adjust the threshold dynamically or trigger human-in-the-loop review for difficult recordings. In fact, SDApy is designed to be efficient and can run on off-the-shelf computer hardware. The algorithm processes data in 4 s epochs with 28 s overlap, allowing near real-time analysis with a minimum 4 s delay.

Despite using a reduced channel count compared to the original study, the algorithm's core functionality generalized well for similar but different recording montages, which is crucial for practical clinical applications [2]. The fact that these results were obtained on a reduced 9-electrode montage further supports the feasibility of deploying SDApy in NICU settings where full 10–20 coverage is not available.

One limitation is the lack of specialized artifact detection, which could potentially improve overall performance; to partially compensate for this, we introduced a more pronounced low-pass filter because in European EEG recordings many technical and muscle artifacts, including 50 Hz power-line noise, predominantly affect frequencies above approximately 40 Hz, and we deliberately avoided an explicit 50 Hz notch filter to reduce the risk of introducing distortions in the signal during real-time processing, while at the same time keeping the preprocessing simple and compatible with fast online use. De Vos and colleagues [16] demonstrated that automated artifact removal as preprocessing can refine neonatal seizure detection. However, for low-performance patients, artifact removal alone may not enhance classification performance. Recent evidence from Leal and colleagues [49] suggests background noise around ictal events can influence classification, distinguishing it from the standard EEG baseline in healthy individuals, and we discuss future artifact removal developments in Section 4.2.

Beyond the SVM-based approach used here, several other techniques have been proposed for neonatal seizure detection, including neural networks and advanced deep learning architectures. Recent deep learning–based detectors, such as fully convolutional networks and hybrid CNN models [27,34], have achieved very high performance—for example, O'Shea and colleagues [28] reported an AUC of 98.5% using a fully convolutional network on raw multi-channel neonatal EEG, and scaled CNN architectures have reached expert-level detection accuracy [35]. However, these methods typically require large, labeled datasets, higher computational resources, and more complex maintenance than classical machine-learning approaches. In contrast, SDApy is positioned as a lightweight, open-source implementation of a validated SVM-based algorithm that is compatible with reduced-channel neonatal EEG and real-time processing constraints, making it suitable as a practical building block in clinical NICU environments and as a pre-labeling tool that can support future deep learning pipelines. From an implementation perspective, SDApy can theoretically run continuously on bedside EEG, provide online seizure surveillance, flag

suspicious segments for expert review, and support human-in-the-loop adjustment of decision thresholds when needed. A systematic, head-to-head comparison with contemporary deep learning and other state-of-the-art methods on the same datasets, under harmonized preprocessing and evaluation protocols, is an important next step and is planned as future work once an appropriate benchmarking framework is established.

We propose that SDApy can complement these methodologies, enhancing the labeling and classification of neonatal seizures. Its compatibility with commercial EEG/aEEG systems, thanks to reduced electrode requirements, makes it practical for widespread clinical adoption.

While Temko and colleagues [39] established SVM use for seizure detection, more recent work [50,51] continues to employ SVM-based pipelines for epileptic seizure detection and highlights the breadth and competitiveness of SVM methods for imbalanced classification problems. Our SVM choice aligns with proven robustness in neonatal EEG analysis, where dataset sizes often preclude deep learning effectiveness.

Though Gaussian kernel SVMs are less interpretable than linear variants, post hoc methods like SHAP [52] enable clinical feature insights. Future work could integrate deep learning for latent pattern discovery with larger datasets while leveraging hybrid interpretability frameworks.

Potential use cases for SDApy include online warning systems for expert review and automated data labeling for deep learning algorithms. Online settings would use fixed thresholds, while offline versions can employ 1000 thresholds to infer optimal performance. This has the disadvantage of requiring multiple algorithm runs at different threshold steps to discern data quality. For automatic labeling where annotations are unknown, it might be beneficial to start with the fixed threshold and then use more complex strategies accounting for individual variability. This could involve recording variations in classified events across different runs to identify cases where thresholds might be significantly off. We should also consider alternative feature standardization techniques accounting for inter-patient variability and commonalities, as most seizures manifest in central/temporal regions [2,53]. Dynamic baseline approaches might achieve robust seizure detection through better patient-specific adaptation by continuously updating baseline buffers.

Another limitation is relying on two expert raters versus the three used in Stevenson and colleagues [38]. This could introduce bias in ground truth data, but it has the advantages of lower labeling time and cost. Future studies should incorporate multiple expert ratings or alternative automatic labeling techniques. The feature selection process, based on Tapani and colleagues [32], could be refined further. Exploring additional features or employing feature selection techniques might improve performance, especially for struggling patients. Incorporating additional information sources could enhance algorithm performance. Lee and colleagues [3] and Falsaperla and colleagues [12] showed value in integrating amplitude-integrated EEG (aEEG) and other physiological signals. Future SDApy iterations could explore such multimodal approaches for improved accuracy and robustness. While the pre-trained model developed at a different location demonstrated good generalization, we lacked access to the original noise vs. signal characteristics, which could impact performance. The model could be fine-tuned upon request to Tapani and colleagues, potentially leading to better performance on specific datasets or recording setups.

4.1. How to Use SDApy in Hospital Settings

SDApy can be integrated into routine NICU workflows as a real-time decision-support tool with only minor adaptations to the current offline pipeline (see Table 4). The core feature extraction and SVM classification blocks can be kept unchanged, while the main modifications concern continuous data streaming, online thresholding, and user-facing

visualization. In practice, SDApy would run continuously on bedside EEG, generate seizure likelihood traces in real time, and raise alarms or visual flags when the decision function crosses a configurable threshold for at least 10 s, allowing clinicians to quickly review suspicious segments instead of scanning the entire recording manually. Human-in-the-loop threshold adaptation, for example, based on a short patient-specific calibration phase, can further tune performance for each neonate while preserving a simple and robust default configuration.

Table 4. A concise summary of suggested design choices for real-time hospital use.

Component	Keep as in Current SDApy	Modify for Real-Time Hospital Use
Preprocessing	Bipolar montage, 32 Hz low-pass filter, 1 Hz high-pass	Implement as low-latency streaming filters; avoid extra complex steps
Feature extraction and SVM	Same 21 features and a pre-trained SVM model	Run on >32 s length circular buffer with 4 s step in an online fashion
Thresholding and event logic	Same default decision threshold and $\geq$10 s minimum event	Add online minimum-duration filtering and allow optional patient-specific threshold tuning
Output representation	Decision function and binary labels per epoch	Display real-time probability/decision trace plus binary alarm channel at the bedside
Clinical workflow integration	Offline analysis and retrospective evaluation	Continuous NICU surveillance, automated pre-labeling, and flagged segments for expert review
Movement artifact detection	Not available for offline data	Automatic online detection of patient movements
Signal degradation detection	Not available for offline data	Automatic online detection of signal degradation (e.g., monitoring for noise bursts, baseline shifts, or sustained spatial or spectral distortion)
Start and end of session impedance calculation	Not available for offline data	Record and log impedances at the beginning and end of each recording session

4.2. Future Directions for Artifact and Noise Reduction in Online Neonatal EEG Analysis

While artifact rejection is not within the immediate scope of this technical development, noise reduction remains a central focus and is adequately achieved through filtering techniques. Sufficient filtering already mitigates a large portion of the noise affecting neonatal EEG recordings. Nonetheless, both artifact and noise reduction may contribute to lowering variability and improving the reliability and reproducibility of online seizure detection results.

In contrast to offline processing, where numerous advanced artifact rejection methods have been developed [54], only a limited set of algorithms are compatible with real-time, online use. Future developments may therefore focus on adapting offline approaches such as artifact subspace reconstruction (ASR, [55]) and generalized eigen-decomposition (GED, [56]) for real-time deployment. Additionally, online-oriented noise reduction could target unavoidable physiological artifacts and incorporate frequency- and spatial-domain filtering strategies [57] that maintain compatibility with real-time system constraints.

5. Conclusions

In conclusion, SDApy demonstrates the successful adaptation of a neonatal seizure detection algorithm to a new dataset with fewer channels, showing robust performance across most patients. The algorithm's ability to generalize well to a different recording setup and its compatibility with existing EEG systems make it a valuable tool for neonatal

intensive care units. The study reveals significant variability in model performance across patients, highlighting the need for patient-specific considerations in clinical applications. While a fixed threshold approach shows promise, there is potential for more sophisticated thresholding strategies to improve performance further. Future work should focus on addressing identified limitations, such as incorporating artifact detection, investigating applicability to other sensor types [58], and exploring multimodal approaches to enhance the algorithm's capabilities.

Supplementary Materials: The following supporting information can be downloaded at: https://www.mdpi.com/article/10.3390/app16010052/s1, Table S1: Classification performance metrics for all validation subjects with fixed threshold.

Author Contributions: C.M.: Conceptualization, Methodology, Software (SDApy), Formal analysis, Writing—Original Draft, A.T.: Data curation (Validation dataset acquisition), Writing—Original Draft, Writing—Review and Editing, M.D.V.: Methodology, Supervision, Writing—Review and Editing, Resources, A.D.: Data curation (Clinical data acquisition), Investigation (Expert EEG annotation), Writing—Review and Editing, J.H.: Supervision, Resources, Writing—Review and Editing, P.F.: Project administration, Supervision, Writing—Review and Editing, Visualization, Corresponding author. All authors have read and agreed to the published version of the manuscript.

Funding: This research was supported by the European EUREKA program (grant ID 01QE2316C), COST action AI-4-NICU (grant ID CA20124), BMFTR (grant ID 01DS23002), and the European Union within the Horizon 2020 call H2020-MSCA-RISE-2019 (grant ID 101007521).

Institutional Review Board Statement: This study was approved by the Ethics Committee of the Review Board of the UZ Leuven hospital (study S54590 approved on 14/02/2014) and was conducted in accordance with the Declaration of Helsinki. All EEG recordings were fully anonymized prior to analysis.

Informed Consent Statement: Written informed consent was obtained from the parents or legal guardians of all participants.

Data Availability Statement: The Helsinki training dataset is publicly available at https://doi.org/10.5281/zenodo.1280683. The Leuven validation dataset is not publicly available due to patient privacy restrictions but can be made available from the corresponding authors upon reasonable request and with appropriate ethical approvals. Requests for access to the Leuven NICU validation dataset should be directed to the corresponding authors via their institutional email address, who will coordinate the necessary steps with the University Hospital Leuven NICU and the relevant ethics bodies. The Python implementation of the seizure detection algorithm (SDApy) developed in this study is available from the corresponding author upon reasonable request. The original MATLAB code is accessible at https://github.com/ktapani/Neonatal_Seizure_Detection (accessed on 14 December 2025).

Acknowledgments: We are grateful to the three anonymous reviewers for their thoughtful feedback and valuable suggestions that strengthened this work. We also acknowledge support for the publication costs by the Open Access Publication Fund of the Technische Universität Ilmenau.

Conflicts of Interest: The authors declare no conflict of interest.

References

1. Dilena, R.; Raviglione, F.; Cantalupo, G.; Cordelli, D.M.; De Liso, P.; Di Capua, M.; Falsaperla, R.; Ferrari, F.; Fumagalli, M.; Lori, S.; et al. Consensus Protocol for EEG and Amplitude-Integrated EEG Assessment and Monitoring in Neonates. *Clin. Neurophysiol.* **2021**, *132*, 886–903. [CrossRef]
2. Dereymaeker, A.; Pillay, K.; Vervisch, J.; De Vos, M.; Van Huffel, S.; Jansen, K.; Naulaers, G. Review of Sleep-EEG in Preterm and Term Neonates. *Early Hum. Dev.* **2017**, *113*, 87–103. [CrossRef]

3. Lee, I.-C.; Hong, S.-Y.; Weng, Y.-H.; Chen, Y.-T. Amplitude Integrated Electroencephalography and Continuous Electroencephalography Monitoring Is Crucial in High-Risk Infants and Their Findings Correlate with Neurodevelopmental Outcomes. *Front. Pediatr.* **2021**, *9*, 691764. [CrossRef] [PubMed]
4. Khazaei, M.; Raeisi, K.; Croce, P.; Tamburro, G.; Tokariev, A.; Vanhatalo, S.; Zappasodi, F.; Comani, S. Characterization of the Functional Dynamics in the Neonatal Brain during REM and NREM Sleep States by Means of Microstate Analysis. *Brain Topogr.* **2021**, *34*, 555–567. [CrossRef]
5. Giampietri, M.; Bartalena, L.; Guzzetta, A.; Boldrini, A.; Ghirri, P. New Techniques in the Study of the Brain Development in Newborn. *Front. Hum. Neurosci.* **2015**, *8*, 1069. [CrossRef]
6. Pisani, F.; Spagnoli, C. Neonatal Seizures: A Review of Outcomes and Outcome Predictors. *Neuropediatrics* **2015**, *47*, 012–019. [CrossRef]
7. Vasudevan, C.; Levene, M. Epidemiology and Aetiology of Neonatal Seizures. *Semin. Fetal Neonatal Med.* **2013**, *18*, 185–191. [CrossRef]
8. Sands, T.T.; McDonough, T.L. Recent Advances in Neonatal Seizures. *Curr. Neurol. Neurosci. Rep.* **2016**, *16*, 92. [CrossRef]
9. Berchicci, M.; Comani, S. Editorial: Bridging the Gap before and after Birth: Methods and Technologies to Explore the Functional Neural Development in Humans. *Front. Hum. Neurosci.* **2015**, *9*, 571. [CrossRef] [PubMed]
10. Martins, R.; Coelho, J.; dos Santos, T.P.; Moreno, T.; Quintas, S.; Levy Gomes, A. Predictores Pronósticos de Desarrollo Neurológico En Recién Nacidos a Término Con Crisis Neonatales. *Rev. Neurol.* **2019**, *69*, 370. [CrossRef] [PubMed]
11. Ziobro, J.; Shellhaas, R.A. Neonatal Seizures: Diagnosis, Etiologies, and Management. *Semin. Neurol.* **2020**, *40*, 246–256. [CrossRef]
12. Falsaperla, R.; Scalia, B.; Giugno, A.; Pavone, P.; Motta, M.; Caccamo, M.; Ruggieri, M. Treating the Symptom or Treating the Disease in Neonatal Seizures: A Systematic Review of the Literature. *Ital. J. Pediatr.* **2021**, *47*, 85. [CrossRef]
13. Kim, E.-H.; Shin, J.; Lee, B.K. Neonatal Seizures: Diagnostic Updates Based on New Definition and Classification. *Clin. Exp. Pediatr.* **2022**, *65*, 387–397. [CrossRef] [PubMed]
14. Ziobro, J.; Pilon, B.; Wusthoff, C.J.; Benedetti, G.M.; Massey, S.L.; Yozawitz, E.; Numis, A.L.; Pressler, R.; Shellhaas, R.A. Neonatal Seizures: New Evidence, Classification, and Guidelines. *Epilepsy Curr.* **2024**. [CrossRef]
15. Pressler, R.M.; Cilio, M.R.; Mizrahi, E.M.; Moshé, S.L.; Nunes, M.L.; Plouin, P.; Vanhatalo, S.; Yozawitz, E.; de Vries, L.S.; Puthenveettil Vinayan, K.; et al. The ILAE Classification of Seizures and the Epilepsies: Modification for Seizures in the Neonate. Position Paper by the ILAE Task Force on Neonatal Seizures. *Epilepsia* **2021**, *62*, 615–628. [CrossRef] [PubMed]
16. De Vos, M.; Deburchgraeve, W.; Cherian, P.J.; Matic, V.; Swarte, R.M.; Govaert, P.; Visser, G.H.; Van Huffel, S. Automated Artifact Removal as Preprocessing Refines Neonatal Seizure Detection. *Clin. Neurophysiol.* **2011**, *122*, 2345–2354. [CrossRef]
17. Dereymaeker, A.; Ansari, A.H.; Jansen, K.; Cherian, P.J.; Vervisch, J.; Govaert, P.; De Wispelaere, L.; Dielman, C.; Matic, V.; Dorado, A.C.; et al. Interrater Agreement in Visual Scoring of Neonatal Seizures Based on Majority Voting on a Web-Based System: The Neoguard EEG Database. *Clin. Neurophysiol.* **2017**, *128*, 1737–1745. [CrossRef]
18. Tatum, W.O.; Mani, J.; Jin, K.; Halford, J.J.; Gloss, D.; Fahoum, F.; Maillard, L.; Mothersill, I.; Beniczky, S. Minimum Standards for Inpatient Long-term Video-electroencephalographic Monitoring: A Clinical Practice Guideline of the International League Against Epilepsy and International Federation of Clinical Neurophysiology. *Epilepsia* **2022**, *63*, 290–315. [CrossRef]
19. Kelly, K.M.; Shiau, D.S.; Kern, R.T.; Chien, J.H.; Yang, M.C.K.; Yandora, K.A.; Valeriano, J.P.; Halford, J.J.; Sackellares, J.C. Assessment of a Scalp EEG-Based Automated Seizure Detection System. *Clin. Neurophysiol.* **2010**, *121*, 1832–1843. [CrossRef] [PubMed]
20. Navakatikyan, M.A.; Colditz, P.B.; Burke, C.J.; Inder, T.E.; Richmond, J.; Williams, C.E. Seizure Detection Algorithm for Neonates Based on Wave-Sequence Analysis. *Clin. Neurophysiol.* **2006**, *117*, 1190–1203. [CrossRef]
21. Faust, O.; Acharya, U.R.; Adeli, H.; Adeli, A. Wavelet-Based EEG Processing for Computer-Aided Seizure Detection and Epilepsy Diagnosis. *Seizure* **2015**, *26*, 56–64. [CrossRef] [PubMed]
22. Indiradevi, K.P.; Elias, E.; Sathidevi, P.S.; Dinesh Nayak, S.; Radhakrishnan, K. A Multi-Level Wavelet Approach for Automatic Detection of Epileptic Spikes in the Electroencephalogram. *Comput. Biol. Med.* **2008**, *38*, 805–816. [CrossRef] [PubMed]
23. Akbarian, B.; Erfanian, A. Automatic Seizure Detection Based on Nonlinear Dynamical Analysis of EEG Signals and Mutual Information. *Basic Clin. Neurosci. J.* **2018**, 227–240. [CrossRef] [PubMed]
24. Stewart, C.; Fung, W.K.; Fough, N.; Prabhu, R. Automated Tonic-Clonic Seizure Detection Using Random Forests and Spectral Analysis on Electroencephalography Data. In Proceedings of the 6th International Conference on System-Integrated Intelligence (SysInt 2022), Genova, Italy, 7–9 September 2022; pp. 679–688.
25. Hussain, L. Detecting Epileptic Seizure with Different Feature Extracting Strategies Using Robust Machine Learning Classification Techniques by Applying Advance Parameter Optimization Approach. *Cogn. Neurodyn.* **2018**, *12*, 271–294. [CrossRef]
26. Pathak, A.; Ramesh, A.; Mitra, A.; Majumdar, K. Automatic Seizure Detection by Modified Line Length and Mahalanobis Distance Function. *Biomed. Signal Process. Control* **2018**, *44*, 279–287. [CrossRef]
27. Ansari, A.H.; Cherian, P.J.; Caicedo, A.; Naulaers, G.; De Vos, M.; Van Huffel, S. Neonatal Seizure Detection Using Deep Convolutional Neural Networks. *Int. J. Neural Syst.* **2019**, *29*, 1850011. [CrossRef]

28. O'Shea, A.; Lightbody, G.; Boylan, G.; Temko, A. Neonatal Seizure Detection from Raw Multi-Channel EEG Using a Fully Convolutional Architecture. *Neural Netw.* **2020**, *123*, 12–25. [CrossRef]
29. Ibrahim, S.W.; Djemal, R.; Alsuwailem, A.; Gannouni, S. Electroencephalography (EEG)-Based Epileptic Seizure Prediction Using Entropy and K-Nearest Neighbor (KNN). *Commun. Sci. Technol.* **2017**, *2*, 6–10. [CrossRef]
30. Dash, D.P.; Kolekar, M.H.; Jha, K. Multi-Channel EEG Based Automatic Epileptic Seizure Detection Using Iterative Filtering Decomposition and Hidden Markov Model. *Comput. Biol. Med.* **2020**, *116*, 103571. [CrossRef]
31. Mathieson, S.R.; Stevenson, N.J.; Low, E.; Marnane, W.P.; Rennie, J.M.; Temko, A.; Lightbody, G.; Boylan, G.B. Validation of an Automated Seizure Detection Algorithm for Term Neonates. *Clin. Neurophysiol.* **2016**, *127*, 156–168. [CrossRef]
32. Tapani, K.T.; Vanhatalo, S.; Stevenson, N.J. Time-Varying EEG Correlations Improve Automated Neonatal Seizure Detection. *Int. J. Neural Syst.* **2019**, *29*, 1850030. [CrossRef]
33. Temko, A.; Sarkar, A.K.; Boylan, G.B.; Mathieson, S.; Marnane, W.P.; Lightbody, G. Toward a Personalized Real-Time Diagnosis in Neonatal Seizure Detection. *IEEE J. Transl. Eng. Health Med.* **2017**, *5*, 1–14. [CrossRef]
34. Gramacki, A.; Gramacki, J. A Deep Learning Framework for Epileptic Seizure Detection Based on Neonatal EEG Signals. *Sci. Rep.* **2022**, *12*, 13010. [CrossRef]
35. Hogan, R.; Mathieson, S.R.; Luca, A.; Ventura, S.; Griffin, S.; Boylan, G.B.; O'Toole, J.M. Scaling Convolutional Neural Networks Achieves Expert Level Seizure Detection in Neonatal EEG. *NPJ Digit. Med.* **2025**, *8*, 17. [CrossRef] [PubMed]
36. Rahaman, M.J. A Comprehensive Review to Understand the Definitions, Advantages, Disadvantages and Applications of Machine Learning Algorithms. *Int. J. Comput. Appl.* **2024**, *186*, 43–47. [CrossRef]
37. Orban, M.; Zhang, X.; Lu, Z.; Zhang, Y.; Li, H. An Approach for Accurate Pattern Recognition of Four Hand Gestures Based on sEMG Signals. In Proceedings of the 2019 2nd International Conference on Control and Robot Technology, Phuket, Thailand, 25–27 October 2019; pp. 145–150. [CrossRef]
38. Koolen, N.; Dereymaeker, A.; RÃ¤sÃ¤nen, O.; Jansen, K.; Vervisch, J.; Matic, V.; De Vos, M.; Van Huffel, S.; Naulaers, G.; Vanhatalo, S. Interhemispheric Synchrony in the Neonatal EEG Revisited: Activation Synchrony Index as a Promising Classifier. *Front. Hum. Neurosci.* **2014**, *8*, 1030. [CrossRef]
39. Stevenson, N.J.; Tapani, K.; Lauronen, L.; Vanhatalo, S. A Dataset of Neonatal EEG Recordings with Seizure Annotations. *Sci. Data* **2019**, *6*, 190039. [CrossRef] [PubMed]
40. Temko, A.; Thomas, E.; Marnane, W.; Lightbody, G.; Boylan, G. EEG-Based Neonatal Seizure Detection with Support Vector Machines. *Clin. Neurophysiol.* **2011**, *122*, 464–473. [CrossRef]
41. Lamblin, M.D.; André, M.; Auzoux, M.; Bednarek, N.; Bour, F.; Charollais, A.; Cheliout-Heraut, F.; D'Allest, A.M.; De Bellecize, J.; Delanoe, C.; et al. Indications de l'électroencéphalogramme En Période Néonatale. *Arch. Pédiatrie* **2004**, *11*, 829–833. [CrossRef]
42. Lima, C.A.M.; Coelho, A.L.V.; Eisencraft, M. Tackling EEG Signal Classification with Least Squares Support Vector Machines: A Sensitivity Analysis Study. *Comput. Biol. Med.* **2010**, *40*, 705–714. [CrossRef]
43. Sha'abani, M.N.A.H.; Fuad, N.; Jamal, N.; Ismail, M.F. KNN and SVM Classification for EEG: A Review. In Proceedings of the 5th International Conference on Electrical, Control & Computer Engineering, Kuantan, Malaysia, 29th July 2019; pp. 555–565.
44. Jaiswal, A.K.; Banka, H. Epileptic Seizure Detection in EEG Signal with GModPCA and Support Vector Machine. *Biomed. Mater. Eng.* **2017**, *28*, 141–157. [CrossRef]
45. Haibo, H.; Garcia, E.A. Learning from Imbalanced Data. *IEEE Trans. Knowl. Data Eng.* **2009**, *21*, 1263–1284. [CrossRef]
46. Lipton, Z.C.; Elkan, C.; Naryanaswamy, B. Optimal Thresholding of Classifiers to Maximize F1 Measure. In Proceedings of the European Conference, ECML PKDD 2014, Nancy, France, 15–19 September 2014; pp. 225–239.
47. Sokolova, M.; Lapalme, G. A Systematic Analysis of Performance Measures for Classification Tasks. *Inf. Process Manag.* **2009**, *45*, 427–437. [CrossRef]
48. Hand, D.J. Measuring Classifier Performance: A Coherent Alternative to the Area under the ROC Curve. *Mach. Learn.* **2009**, *77*, 103–123. [CrossRef]
49. Leal, A.; Curty, J.; Lopes, F.; Pinto, M.F.; Oliveira, A.; Sales, F.; Bianchi, A.M.; Ruano, M.G.; Dourado, A.; Henriques, J.; et al. Unsupervised EEG Preictal Interval Identification in Patients with Drug-Resistant Epilepsy. *Sci. Rep.* **2023**, *13*, 784. [CrossRef]
50. Atlam, H.F.; Aderibigbe, G.E.; Nadeem, M.S. Effective Epileptic Seizure Detection with Hybrid Feature Selection and SMOTE-Based Data Balancing Using SVM Classifier. *Appl. Sci.* **2025**, *15*, 4690. [CrossRef]
51. Rezvani, S.; Pourpanah, F.; Lim, C.P.; Wu, Q.M.J. Methods for Class-Imbalanced Learning with Support Vector Machines: A Review and an Empirical Evaluation. *Soft Comput.* **2024**, *28*, 11873–11894. [CrossRef]
52. Feng, J.; Liang, J.; Qiang, Z.; Hao, Y.; Li, X.; Li, L.; Chen, Q.; Liu, G.; Wei, H. A Hybrid Stacked Ensemble and Kernel SHAP-Based Model for Intelligent Cardiotocography Classification and Interpretability. *BMC Med. Inf. Decis. Mak.* **2023**, *23*, 273. [CrossRef]
53. Bogaarts, J.G.; Hilkman, D.M.W.; Gommer, E.D.; van Kranen-Mastenbroek, V.H.J.M.; Reulen, J.P.H. Improved Epileptic Seizure Detection Combining Dynamic Feature Normalization with EEG Novelty Detection. *Med. Biol. Eng. Comput.* **2016**, *54*, 1883–1892. [CrossRef]

54. Schmoigl-Tonis, M.; Schranz, C.; Müller-Putz, G.R. Methods for motion artifact reduction in online brain-computer interface experiments: A systematic review. *Front. Hum. Neurosci.* **2023**, *17*, 1251690. [CrossRef]
55. Kim, H.; Chang, C.Y.; Kothe, C.; Iversen, J.R.; Miyakoshi, M. Juggler's ASR: Unpacking the principles of artifact subspace reconstruction for revision toward extreme MoBI. *J. Neurosci. Methods* **2025**, *420*, 110465. [CrossRef] [PubMed]
56. Cohen, M.X. A tutorial on generalized eigendecomposition for denoising, contrast enhancement, and dimension reduction in multichannel electrophysiology. *NeuroImage* **2022**, *247*, 118809. [CrossRef] [PubMed]
57. Graichen, U.; Klee, S.; Fiedler, P.; Hofmann, L.; Haueisen, J. Head-Specific Spatial Spectra of Electroencephalography Explained: A Sphara and BEM Investigation. *Biosensors* **2025**, *15*, 585. [CrossRef]
58. Asayesh, A.; Warsito, I.F.; Haueisen, J.; Fiedler, P.; Vanhatalo, S. Neonatal EEG Recording with a Dry Electrode Cap: A Feasibility study. *Sensors* **2025**, *25*, 966. [CrossRef] [PubMed]

Disclaimer/Publisher's Note: The statements, opinions and data contained in all publications are solely those of the individual author(s) and contributor(s) and not of MDPI and/or the editor(s). MDPI and/or the editor(s) disclaim responsibility for any injury to people or property resulting from any ideas, methods, instructions or products referred to in the content.