UDC 004.89

# Feature Selection for Electrical Brain Activity Classification in Newborns in Case of Painful Events

*Bondarev V. R.*[1], *Ivanko K. O.*[1], *Popov A. O.*[1], *Karplyuk Ye. S.*[1], *Korogod N. S.*[2]

[1]National Technical University of Ukraine "Igor Sikorsky Kyiv Polytechnic Institute", Kyiv, Ukraine
[2]HESAV School of Health Sciences - Vaud, HES-SO University of Applied Sciences and Arts Western Switzerland

E-mail: *vbondarev-ee23@lll.kpi.ua*, *ivanko-ee@lll.kpi.ua*, *popov-ee@lll.kpi.ua*, *yk-ee@lll.kpi.ua*, *natalya.korogod@hesav.ch*

The understanding of pain mechanisms in infants is critically important because newborns lack verbal communication abilities to report their pain experiences. This study focuses on analyzing electrical brain activity features in time and time-frequency domains using electroencephalographic (EEG) signals collected during clinically required noxious stimuli in newborns. Different feature extraction techniques are explored by applying a combined feature selection approach with forward feature selection and statistical measures involved. Six machine learning algorithms, namely Logistic Regression, Linear Discriminant, K-Nearest Neighbors, Support Vector Machines, Random Forest, and Gaussian Naive Bayes, were used and compared with the purpose of painful events recognition in newborns. Two binary classification tasks were considered: the first task was to distinguish between EEG signals before painful stimulus and after it (painless and painful state of the patient) and the second task was to distinguish between EEG signals on the background of the painful event (heel lance for blood sampling in neonates) and signals without painful events (audio stimulation).

In the task of EEG signals classification into pre- and post-painful stimulus segments, the support vector machines classifier showed the best accuracy estimate of 93.5% with the pre-painful EEG segments classification accuracy of 100% and post-painful segment classification accuracy of 86.9%. In the task of distinguishing between EEG signals containing painful events as heel lance and signals without painful events, the linear discriminant algorithm showed the best accuracy estimate of 84% with 76.9% correctly determined EEG segments containing painful events and 91.6% correctly determined EEG segments without painful events. Results demonstrate the potential of using features that focus on spectral power in alpha, beta, and gamma frequency bands and machine learning techniques for advancing pain detection in neonates.

*Keywords:* electroencephalography; pain markers; newborns; machine learning; feature selection; classification; classification accuracy; biosignal analysis

DOI: 10.64915/RADAP.2025.102.40-50

## Introduction

Modern medicine faces significant challenges in both detecting pain in newborns and measuring its severity [1]. Medical practitioners use indirect signs to determine pain levels in nonverbal patients who cannot express their discomfort [1]. The development of effective pain-management strategies and prevention of long-term pain effects requires a better understanding of the neuro-physiological mechanisms that control neonatal pain responses. Several validated scoring systems exist in medical practice [1], but there is no universally accepted standard. The current assessment methods combine objective physiological indicators, including heart rate and blood pressure and cortisol levels, with clinical observations of facial expressions and crying patterns and movement behaviors [1].

Research conducted on newborn electroencephalographic (EEG) signals during painful medical procedures [2–4] shows that machine learning technology demonstrates potential for detecting pain occurrences of newborns, creating automated pain assessment techniques as well as analyzing pain related markers based on the large amount of features. The automated assessment method reduces the need for subjective interpretation when evaluating newborn health status during painful medical procedures. The creation of reliable algorithms for infant pain detection presents substantial opportunities to enhance pain management practices [2–4]. The development of these advancements could help reduce the potential negative developmental effects that result from early pain exposure.

It was shown [5] that neural networks are highly effective in EEG signal analysis for pain-related pattern detection since they are able to model complex nonlinear patterns in high-dimensional data. The analysis of EEG signals for pain detection benefits from convolutional neural networks (CNNs) because these networks automatically extract meaningful features and identify pain-related brain activity patterns in unprocessed EEG data. Models using band power features, such as Shallow ConvNet, outperform other models for classification of EEG segments regarding presence or absence of painful events. But since neural networks provide direct EEG signal analysis, the current methods do not enable researchers to explain or fully understand how the brain's reaction to pain is manifested in the EEG signal, which limits the interpretability of the results. In order to better understand and interpret the results of artificial intelligence models by doctors and identify pain-related markers in newborns, it is necessary to study various features in time and time-frequency domains while performing feature selection and developing machine learning models.

On the other hand, the task of investigating electrical brain activity in newborns in case of painful events remains relatively new, so the most informative EEG leads and rhythms have not yet been clearly established. The research findings in the study by Norman, E. et al. [2] showed that pain-related manifestations appeared mainly in the beta rhythm but the study provided by Marianne van der Vaart [3] discovered that beta rhythm features provided less information than lower-frequency rhythms. Also, Talebi S. et al [4] used the Student's t-test and pseudo-sequential forward feature selection methods and spectral analysis to determine that theta activity in the Cz EEG lead and delta activity in C4 were the most important features for pain vs no-pain classification. No useful features were found in the alpha and gamma bands.

Despite the fact that machine learning techniques are already used for pain recognition in newborns, there is still a need for deep feature analysis. This work considers the search of pain-related markers in EEG signals of newborns based on the power features in different frequency bands, providing feature selection and machine learning techniques to identify and study pain-related markers of newborns.

# 1 Materials and methods

The task of automatic recognition of painful and non-painful states in infants using machine learning methods is relevant since newborns are nonverbal patients who cannot assess their pain and inform the doctor about its degree. Machine learning techniques as well as neural network models, which leverage power features in different frequency bands, can be an appropriate tool to identify painful and non-painful states with high precision. Based on the neural network models evaluation, band power analysis may be more suitable for detecting pain-related markers in neonates compared to an event-related approach but the understanding of the exact features and pain markers is still a complex task. The pre-training approach also requires more in-depth analysis to find a correlation between features and prove the approach for the neural network models training.

This work considers the search for pain-related markers in newborns by EEG signals analysis based on the feature detection in different frequency bands and machine learning models training. Two datasets with responses to painful stimuli were used in this study. The first dataset *Cortical, behavioral, and physiological responses following a single, clinically required noxious stimulus in neonatal subjects* by Jones, L. et al [6] contained EEG data of the newborns collected during blood collection from the heel. The second dataset *Distinct patterns of brain activity mediate perceptual and motor and autonomic responses to noxious stimuli* by Tiemann, L. et al [7] consisted of EEG data registered in adults and contained responses to painful laser stimuli applied to the dorsum of the left hand. The dataset was used to evaluate the potential of using adults data for the machine learning pre-training in the tasks where the amount of data is crucial, like neural network studying, and if the approach can be used in the future. The expectation was to determine whether there are similarities in the pain markers in the electrical activity of the brains of adults and infants.

EEG signals from both datasets were pre-processed and segmented as 2 seconds pre- and 2 seconds post-stimulus and were labeled as two classes: pre- and post-stimulus (painless and painful state of the patient). EEG signals of newborns were also used to distinguish between the signals containing painful events (heel lance) and those without painful events (audio stimulation).

## 1.1 EEG dataset of newborns with reactions to a painful procedure

The neonates dataset used in this study is a dataset of *EEG, behavioral, and physiological responses to a painful procedure in human neonates with relevant medical history* by Jones et al recruited from the postnatal, special care, or intensive care wards at the Elizabeth Garrett Anderson Obstetric Wing, University College London Hospital [6]. 112 neonates participated in a research (29–47 weeks of gestational age at study). Cortical activity was registered as a 20-channel EEG recording using a modified international 10/10 electrode placement system, with high-density central-parietal and posterior temporal coverage. Each signal is associated with a research event such as a heel lance as a painful stimulus or an audio control as a non-painful stimulus. Signals were digitized with a sampling rate of 2 kHz and a resolution of 24-bit.

Database documentation was also provided by Jones et al [6], including the quality of the neonates' EEG signals recorded using the modified 10/10 electrode placement system and the stimulus information, such as the Premature Infant Pain Profile (PIPP) score. EEG channels with poor EEG quality, defined by Jones et al [6] were not involved in the analysis. Signals were grouped after the pre-processing based on the available PIPP scores. Records with a PIPP score from 0 to 5 were marked as a "low pain level" group (46 patients), records with scores from 6 to 12 – as a "mid pain level" group (27 patients), and records with scores higher than 12 – as a "high pain level" group (5 patients). Signals without available PIPP scores (33 patients) were also used in training according to the experiment design and considered binary classification tasks: pre- and post-painful event signals classification and distinguishing between EEG signals with and without painful events. For the EEG signals without painful events, 60 patients are available in the "low pain level" group, 8 patients in the "mid pain level" group and 32 patients without available PIPP score.

With the two classification tasks set in this study, the neonates dataset was used to form two data groups. EEG signals from newborns were segmented into two fragments: 2 seconds pre- and 2 seconds post-stimulus and were labeled as two classes: pre- and post-stimulus. A total of 220 EEG segments of the neonates' EEG signals (110 for each class) were used for machine learning. For the second classification task, we used EEG signals with painful stimulus (heel lance) and without (audio stimulation). 210 EEG signals of the neonates data were available (110 EEG signals class containing painful events and 100 EEG signals without painful events).

A 4th-order Butterworth band-pass filter with cutoff frequencies of 0.5 Hz and 70 Hz and a band-stop notch filter with a quality factor of 20 were applied to filter signals, including 50 Hz power supply frequency. Soft automatic and tunable artifact removal algorithm [8] was used to remove EEG artifacts. Window-wise wavelet packet decomposition and reconstruction with a soft thresholding window size of 1 sec (2000 samples) was used to employ statistical measures such as standard and median absolute deviations and wavelet coefficient distribution to establish thresholds that were computed and applied to filter the wavelet coefficients for each window for separating artifacts from physiological components.

## 1.2 Adult EEG dataset with responses to a painful procedure

To find the features reflecting the EEG responses to painful procedures, EEG signals collected from adults with responses to noxious stimuli were investigated. EEG records of 51 healthy adult participants from the dataset *Distinct patterns of brain activity mediate perceptual and motor and autonomic responses to noxious stimuli* [7] were used for this purpose. The electrode montage included all electrodes according to the International 10–20 system of scalp electrodes location. EEG signals were referenced to the FCz electrode, sampled at 1000 Hz, high-pass filtered at 0.015 Hz, and low-pass filtered at 250 Hz. 60 brief painful laser stimuli were applied to the left hand of the volunteers. Different intensities of the laser stimuli were used (light, medium and strong): 20 painful laser stimuli per record. A total of 9180 events were available for analysis across all experiments.

## 1.3 Feature engineering for painful EEG responses detection

EEG signals of newborns and adults underwent analysis through time and frequency domain methods as well as wavelet decomposition for time-scale analysis. Multiple statistical and morphological parameters were calculated from the time domain EEG data. The time domain analysis included measures of the mean value and mode. Standard deviation, together with variance, served to evaluate the signal variability. The standard deviation shows how much data points spread out from their mean value, indicating the spread of EEG signal amplitude values, while variance measures the total spread of values in relation to their expected value.

The analysis included parameters that measured both the signal's magnitude and its distribution shape. The root-mean-square value evaluates the signal amplitude and signal energy in the time domain. Kurtosis [9] characterizes how heavily the tails of a distribution differ from those of a normal distribution, while skewness measures the asymmetry of the probability distribution around the mean. The Higuchi [10] coefficient evaluates the complexity and irregularity of EEG signals.

The following formula was used to calculate the coefficient of kurtosis:

$$k = \frac{\frac{1}{n}\sum_{i=1}^{n}(x_i - \overline{X})^4}{\left(\frac{1}{n}\sum_{i=1}^{n}(x_i - \overline{X})^2\right)^2}\ , \tag{1}$$

where $n$ – signal length in samples, $\overline{X}$ – average signal value, $x_i$ – value of the $i$-th sample of the signal.

The asymmetry index was calculated as follows:

$$S = \frac{\frac{1}{n}\sum_{i=1}^{n}(x_i - \overline{X})^3}{\left(\sqrt{\frac{1}{n}\sum_{i=1}^{n}(x_i - \overline{X})^2}\right)^3} \tag{2}$$

Hjorth [11, 12] parameters, that are a set of features (activity, mobility, and complexity) that quantify the shape and frequency characteristics of EEG, were also used for analyzed EEG signals. The activity measurement determines signal variance which directly corresponds to signal amplitude. The strength of signal intensity or neural activation corresponds to high activity while flat segments indicate low activity. This is represented by the following equation:

$$Activity = Variance(x(t)) = \frac{1}{N}\sum_{t=1}^{N}\left(x(t)-\overline{X}\right)^2, \quad (3)$$

where $\overline{X}$ represents the mean of the signal $x(t)$.

The mobility parameter represents the mean frequency of the signal. It estimates how quickly the signal changes over time. The signal contains fast transitions when mobility values are high, but it shows slow waveforms when mobility values are low.

$$Mobility = \sqrt{\frac{Variance\left(\frac{dx(t)}{dt}\right)}{Variance(x(t))}} \quad (4)$$

Complexity measures the change in mobility within the signal. The measure indicates how much the waveform deviates from a basic harmonic form. The complexity value increases when the signal contains complex temporal structures, which include abrupt changes and multiple frequency elements.

$$Complexity = \frac{Mobility\left(\frac{dx(t)}{dt}\right)}{Mobility(x(t))} \quad (5)$$

From the perspective of entropy-based analysis, several measures were calculated to assess the complexity and regularity of the EEG signals. First, singular-value decomposition (SVD) entropy [13] was computed. The metric shows the minimum number of eigenvectors needed to explain the dataset structure which indicates its total complexity level. The system complexity was measured by permutation entropy through the analysis of consecutive value order and the generation of probability distributions of their patterns. Finally, approximate entropy and sample entropy were calculated to evaluate EEG time series regularity through the assessment of pattern similarity probability across the series duration.

The same parameters were calculated for the EEG signals filtered in different rhythms of electrical activity of the brain using a 4th-order Butterworth band-pass filter. So, having 17 EEG channels for each patient and 16 parameters calculated, 1632 features were calculated in the time domain.

Brief analysis of the neonates EEG power distribution between classes (pre-painful EEG segment and post-painful EEG segment, as well as EEG signal containing painful events or without painful events) showed the potential of using features from the frequency and time-frequency domains to identify pain markers (Fig. 1).

To obtain features in the frequency domain, a Short-Time Fourier Transform (STFT) was applied to the signal, after which the resulting spectrum was divided into parts corresponding to different rhythms of electrical activity of the brain: delta (0.5 - 3.0 Hz), theta (4.0 - 6.0 Hz), alpha (8.0 - 14.0 Hz), beta (14.0 - 40.0 Hz), gamma (40.0 - 70.0 Hz). A Blackman window with a STFT length of 1000 samples was used to provide time-frequency analysis. The feature selection scheme is shown in Fig. 2.

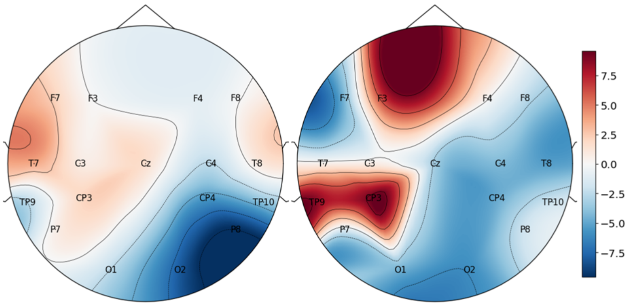


Fig. 1. Total energy spatial distribution of the "high" group neonate's EEG for the two states: pre-painful and post-painful EEG segments

For the spectrum, as well as for each of the rhythms, a full energy was calculated. Also, signal characteristics that were listed in the time domain analysis were calculated for the time-frequency domain. So, having 13 features for each of the 5 frequency rhythms and each of the 17 channels, 1105 features were collected in frequency and time-frequency domains. The total energy was calculated as:

$$P_r = \sum_{i=0}^{n} F_{ri}^2 \, , \quad (6)$$

where $F_{ri}$ is the value of the $i$-th sample of the amplitude spectrum corresponding to the rhythm $r$.

After downsampling of EEG signals to 512 Hz, the discrete wavelet transform (DWT) with decomposition to 7 levels using the Daubechies wavelet function of the 2nd order was applied to analyze wavelet components corresponding to the EEG rhythms. Aligning wavelet components to the EEG frequency rhythms we obtained the following components: D1 (256 - 512 Hz), D2 (128 - 256 Hz), D3 (64 - 128 Hz), D4 (32 - 64 Hz that corresponds to the gamma rhythm), D5 (16 - 32 Hz – beta rhythm), D6 (8 - 16 Hz – alpha rhythm), D7 (4 - 8 Hz – theta rhythm), A7 (0 - 4 Hz – delta rhythm). D1, D2, D3 detail components were ignored in further analysis.

For each of the obtained wavelet decomposition levels, all parameters were calculated, which were listed in the time domain analysis. So, having 12 features for each of the 5 wavelet decomposition components and each of the 17 channels, 1020 features were derived from the wavelet analysis.

In total, 3757 features were calculated for each analyzed EEG signal and used in further analysis and machine learning for two considered binary classification tasks (pre- and post-painful EEG classification and classification of EEG signals containing painful events and without painful events). The same features were also determined from the adult EEG signals.

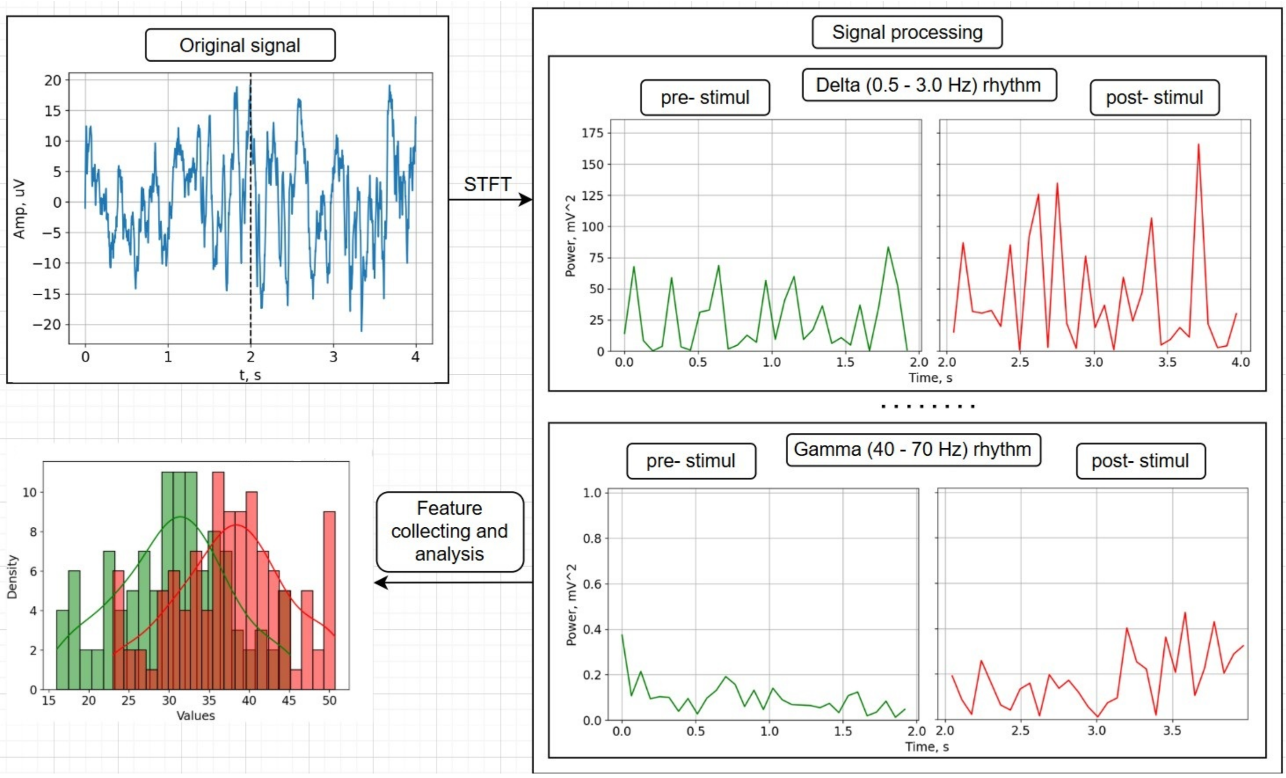


Fig. 2. Neonates EEG signal processing and analysis for feature collection and distribution analysis in a time-frequency domain

The cumulative sum control chart (CUSUM) [14] metric was used for signal change detection. CUSUM tests for changes in the mean by computing cumulative deviations from a reference value (the mean), using the following equation:

$$S_t = \sum_{i=1}^{t} \left(x_i - \overline{X}\right), \tag{7}$$

where $\overline{X}$ is the mean of the signal. Then, the CUSUM statistic is $\max|S_t|$ corresponds to fluctuating around 0 if the mean is constant. A shift in the mean causes a slope in $S_t$ making $\max|S_t|$ large. Another metric was $(\max S_t - \min S_t)$ used to get the direction of accumulated deviation.

## 1.4 Feature selection

One of the main goals of the study was to determine the most important features describing pain markers that manifest themselves in the signal of brain electrical activity of newborns and are the most suitable for machine learning to build models for the painful events recognition in newborns.

The features were cleaned of the outliers using a standardized score. A standardized score (z-score) [15] is a measure of the relative dispersion of an observed or measured value, which shows how many standard deviations the dispersion of the relative mean value corresponds to. Standardized value estimation is calculated according to the formula:

$$z = \frac{x - \overline{X}}{S_x}, \tag{8}$$

where $\overline{X}$ – mean value, $S_x$ – standard deviation calculated for a set of data $x_i$. Since the distribution of z-scores is approximated by a standard normal distribution, there is a one-to-one correspondence between percentiles and z values. This allows for an unambiguous translation of the rank scale or scores into z-score values and back (thus, the value z = 3 corresponds to the 0.13 percentile, z = 2 – the 2.3rd percentile, z = 1 – the 15.9th percentile, etc.). Outliers were considered to have absolute z-score value more than 2, values that were smaller or larger than the given range were assigned a minimum or maximum value respectively. Outliers removal was only applied to provide a statistical assessment of the features and the forward feature selection algorithm, while training and validation of the models were performed with the original feature sets.

Statistical assessment of the features was used to identify the features that can be used as pain markers. This allowed to assess features independently from the chosen machine learning algorithms and to analyze the full set of features without the need to specify the target number of features, as it is needed for the forward feature selection. On the other hand, forward feature selection was used to identify the starting number of features to analyze and to validate feature values for the specific machine learning algorithms as described further.

To evaluate the discriminative relevance of features between the two classes, Student's t-test [16] was used as a statistical test for determining whether the difference between the responses of the two groups is statistically significant or not. In testing the null hypothesis that the sample mean (the mean value of the tested feature) is equal to a specified value 0, the following equation is used:

$$t = \frac{\sqrt{n}\,(\overline{X} - \mu_0)}{s}, \tag{9}$$

where $\overline{X}$ is the sample mean, $s$ – the sample standard deviation and $n$ is the sample size. The degrees of freedom used in this test are $n - 1$. The independent two-sample t-test was computed for each feature, treating the values from the two classes as separate distributions. The t-statistic measures the difference between class means relative to within-class variability to show the strength of feature separation between classes. Features with higher absolute t-values exhibit stronger discriminative power. Unlike p-values, which provide a significance level but are influenced by sample size, the t-statistic directly reflects the effect size in relation to variance, making it a suitable criterion for ranking features [17].

To identify the starting amount of features to analyze and provide further analysis, an iterative Forward Feature Selection [18] method was used. This method is a wrapper method, so it depends on the machine learning model for which the feature selection is performed. Direct sequential feature selection consists of gradually increasing the number of features: in the first iteration, the model is trained on each of the independent variables separately and for further training, the feature that provides the highest classification accuracy is selected. In subsequent iterations, the algorithm checks how the addition of a new feature to the existing subset affects. The feature whose addition led to the greatest increase in accuracy is added to the feature subset. The absence of an increase in classification accuracy when adding a new feature was chosen as the stopping criterion of the iterative algorithm.

The obtained features were also analyzed using Student's t-test and the top 20 features by the t-score were chosen for further machine learning. The example of the statistical feature analysis is shown in Fig. 3. Also, if a specific feature was selected across all machine learning methods using the Forward Feature Selection method, it was also included in the resulting set of features. Since not all available channels were marked as valid, a threshold value of feature list length was set prior to applying Student's t-test. Features with the 2.3rd percentile (features that are available for at least 97.7% of patients) were used to apply Student's t-test.

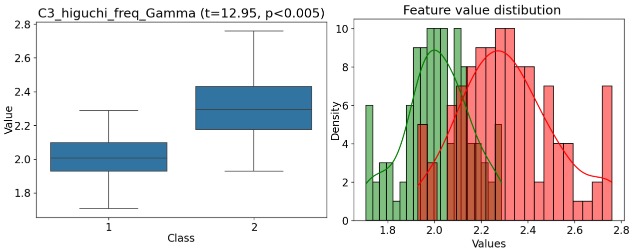


Fig. 3. Distribution of the Higuchi coefficient in Gamma rhythm for C3 channel for the pre-painful (class 1) and the post-painful (class 2) EEG segments of newborns based on the Student's t-test assessment

## 1.5 Machine learning application for painful events detection in newborns

Selected features were used to provide machine learning training for two classification tasks: pre- and post-painful EEG segment classification and classifying EEG signals containing painful events and without painful events. Six machine learning algorithms were chosen for the considered tasks.

Logistic Regression (LR) [19] is a supervised classification algorithm used to predict the probability of an input belonging to a specific class. The algorithm transforms input features into predicted class probabilities through the logistic (sigmoid) function. The model determines feature coefficients to achieve the highest possible likelihood of correct training example classification.

The supervised Linear Discriminant (LD) [20] method transforms input features into a reduced-dimensional space to achieve optimal class discrimination. The method requires data to follow normal distribution patterns with equal covariance measurements between classes. The algorithm identifies linear feature combinations that optimize class separation through the maximization of between-class variance relative to within-class variance.

The K-Nearest Neighbors (KNN) algorithm [21] functions as a non-parametric instance-based learning method that determines new sample classifications through the majority class found among its $K$ nearest neighbors in feature space. The most typical distance metrics for defining neighborhood closeness include Euclidean and Manhattan. The algorithm lacks distribution assumptions about data which provides flexibility yet remains vulnerable to feature scaling and $K$ value selection. The algorithm performs best with small datasets.

The supervised learning model C-Support Vector Classification (C-SVC) [22] with a sigmoid kernel performs binary classification by transforming input data into a higher-dimensional space through sigmoid function mapping. The decision boundary is determined by support vectors that lie closest to the separating surface. The sigmoid kernel function creates non-linear decision boundaries. The regularization parameter $C$ determines how much to prioritize margin size against classification error.

The Random Forest (RF) [23] ensemble learning method trains multiple decision trees and selects the class that receives the most votes from all trees. The training process of each tree uses both random data samples and random feature subsets to prevent overfitting and enhance generalization capabilities. The robustness of the Random Forest method against noise and its ability to handle high-dimensional data makes it an appropriate choice for complex classification tasks, including EEG signal analysis.

The Gaussian Naive Bayes (GNB) [24] classifier uses Bayes' theorem to classify data while assuming that features become independent when the class label is known. The model represents continuous feature distributions through normal distributions that apply to each class. The model achieves good results on high-dimensional data despite its basic nature and strict independence requirements. The algorithm operates efficiently while delivering optimal results when the features match normal distribution patterns.

Both datasets (pre- and post- painful EEG segments and EEG signals containing painful event or without painful events) were split into training (180 EEG segments and 171 EEG signals accordingly) and test (44 EEG segments and 43 EEG signals accordingly) data parts with a split factor of 0.2 which corresponds to 80% of the training data and 20% of the testing data. The train set of 180 EEG segments and 171 EEG signals were additionally split with a split factor of 0.2 for the training of the models to provide validation accuracy estimates. Data were split taking into account the distribution between classes to save data distribution. Also, for the pre- and post-painful EEG segments classification data was split, taking into account patient affiliation: pre- and post-painful EEG segments of one patient were used either as training or validation data. For the task of EEG signals containing painful event or without painful events classification additional work was carried out to analyze only "mid pain level" (PIPP scores from 6 to 12) and "high pain level" (PIPP scores higher than 12) groups of signals containing painful events and "low pain level" (PIPP score from 0 to 5) group of signals without painful events to eliminate data collection errors. Classes were balanced with EEG signals containing painful events from the "low pain level" group (PIPP score from 0 to 5) with higher scores. A total number of 125 signals was available in this case.

A few techniques were chosen to improve model training performance. It was decided to pre-tune the models using hierarchical search. Grid search [25] is a method that iterates through all possible combinations of settings for each parameter. The following models were used with variations of the following parameters: KNN (selecting the number of neighbors used) and RF (selecting the number of estimators). Also, collected features were normalized by subtracting the mean and scaling the variance to 1 with storing the mean and variance values to be able to use the trained classifiers on new data sets:

$$z = \frac{x - \overline{X}}{s}\ , \tag{10}$$

where $\overline{X}$ is the mean of the training samples and $s$ is the standard deviation.

Model training was performed with the use of common techniques [19–24] declared for each machine learning algorithm validating performance metrics, including accuracy, during training and cross-validation. The best model was then used to provide total accuracy estimates and evaluation across classes.

Total accuracy estimates were performed on the previously unused testing data and was calculated based on a total amount of true predictions across:

$$Accuracy = \frac{(T_1 + T_2)}{(T_1 + T_2 + F_1 + F_2)}\ , \tag{11}$$

where $T_1$ is the number of true predictions in first class, $T_2$ is the number of true predictions in second class, $F_1$ is the number of false predictions in first class, $F_2$ is the number of false predictions in second class.

Accuracy for each class separately (Recall) was calculated based on the number of true predictions to the total number of analyzed signals in class:

$$Accuracy_1 = \frac{T_1}{T_1 + F_1} \tag{12}$$

Permutation importance [26] was used with the trained models to analyze feature importance from the concrete model perspective and additionally analyze feature importance compared to the statistical analysis. The permutation importance is defined to be the difference between the baseline metric and the metric from permutating the feature column. The idea behind the method is that randomly re-ordering a single column should cause less accurate predictions, since the resulting data no longer corresponds to real data. Results obtained for newborns were also analyzed in comparison with the results for adults.

## 2 Results

Using the Forward Feature Selection method, 20 features from the total number of 3757 features calculated for each patient were selected for further statistical analysis. Also, 6 machine learning algorithms (LR, LD, KNN, C-SVC, RF, GNB) were used for training in the scope of the two classification tasks: pre- and post-painful EEG segments and classification of the EEG signals containing painful events and without the presence of painful events.

For the task of the pre- and post-painful EEG segments classification, it can be seen that C3, C4, Cz, T7 and T8 EEG channels are the most valuable for the pain event marker identification according to the Student's t-test statistical distribution. The Higuchi coefficient, which is used to calculate the fractal dimension of an EEG signal and SVD entropy, which is an indicator of the number of eigenvectors needed for an adequate explanation of the data set in the gamma frequency range are proposed as valuable pain-related markers. Also, root mean square (RMS), standard deviation and the mean signal value in a delta and theta rhythms showed differences between the two investigated classes according to the Student's t-test (Fig. 4).

Selected features were used to train machine learning models. Gaussian Naive Bayes and the C-SVC showed the highest total accuracy of 93.5%. Results of the machine learning models training for the task of the pre- and post-painful EEG segments classification are presented in Table 1.

Table 1 Results of the machine learning models training for the task of pre- and post-painful neonates EEG segments classification

| | LR | LD | KNN | C-SVC | RF | GNB |
|---|---|---|---|---|---|---|
| 1st class, % (pre-painful EEG segments) | 100 | 100 | 91.3 | 100 | 95.6 | 100 |
| 2nd class, % (post-painful EEG segments) | 74 | 82.6 | 78.2 | 86.9 | 86.9 | 86.9 |
| Total accuracy, % | 87 | 91.3 | 84.8 | 93.5 | 91.3 | 93.5 |

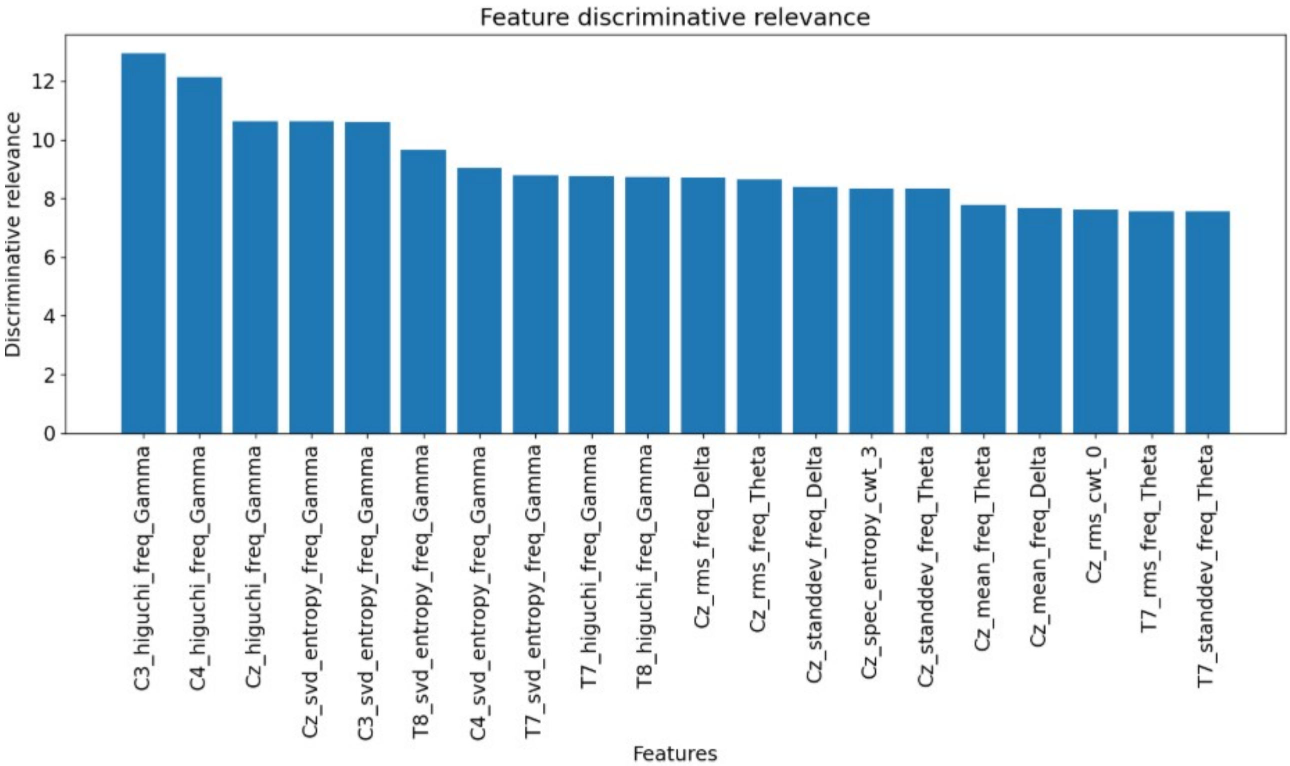


Fig. 4. Pain markers selected using discriminative relevance by the Student's t-test for the task of the pre- and post-painful neonates EEG segments classification

The same approach was used to identify features for the task of classification of EEG signals containing painful events and those without painful events (Fig. 5). Features that can monitor change detection were added to the analysis (CUSUM). C4 and Cz can also be noted as valuable channels for the pain events detection. RMS, standard deviation and the mean parameters in a delta, theta and gamma rhythms, as well as CUSUM parameters, can become valuable pain-related markers.

The same machine learning models were trained with the selected features. Taking into account unsatisfactory accuracy when using the full data set, a subset of "mid pain level" (PIPP scores from 6 to 12) and "high pain level" (PIPP scores higher than 12) groups of signals containing painful events and "low pain level" (PIPP score from 0 to 5) group of signals without painful events were analyzed. This helped to eliminate data collection errors. Results of the machine learning models training are grouped in Table 2. LD showed the best total accuracy of 84%.

Table 2 Results of the machine learning models training for the classification task of neonates EEG signals with and without painful events

| | LR | LD | KNN | C-SVC | RF | GNB |
|---|---|---|---|---|---|---|
| 1st class, % (signals without painful event) | 75 | 91.6 | 83.3 | 75 | 58.3 | 75 |
| 2nd class, % (signals with painful event) | 76.9 | 76.9 | 61.5 | 84.6 | 53.8 | 76.9 |
| Total accuracy, % | 76 | 84 | 72 | 80 | 56 | 76 |

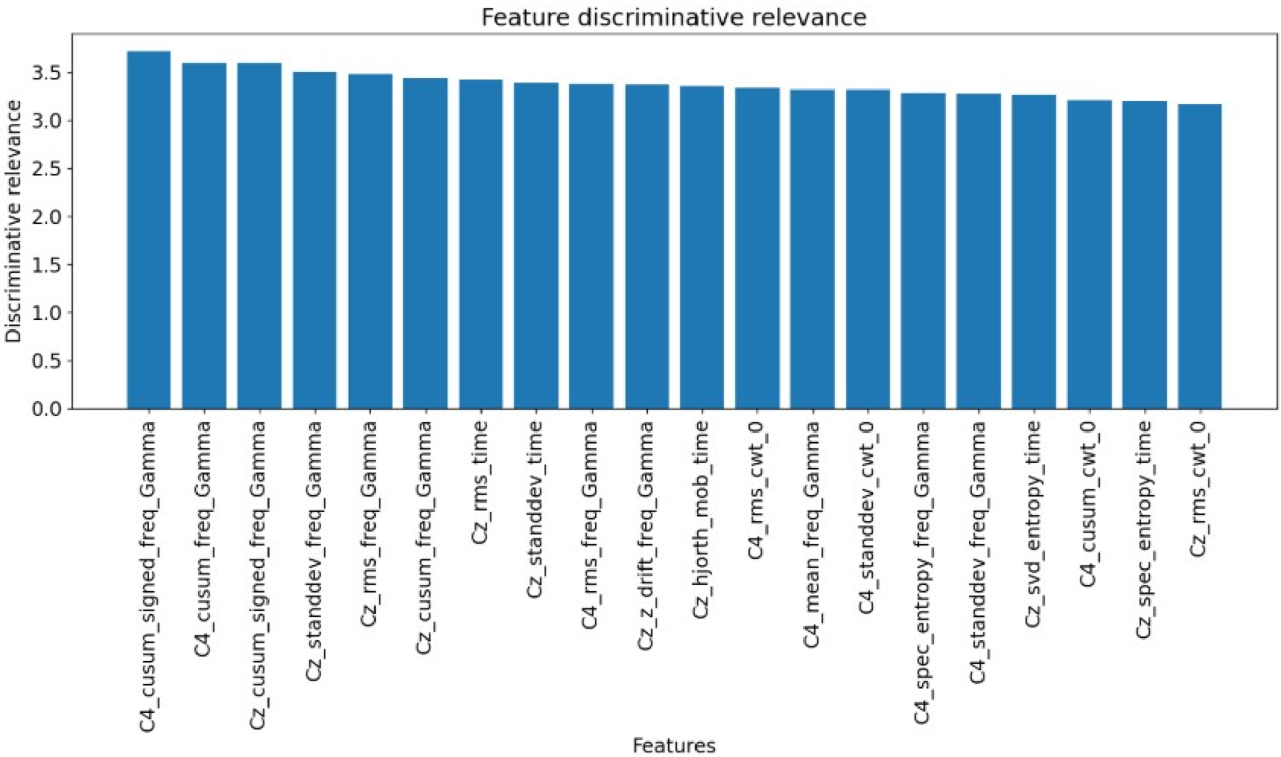


Fig. 5. Pain markers selected using discriminative relevance by the Student's t-test for the task of neonates EEG signals classification of the containing painful events and without painful events.

It can be noted that C3, C4, Cz and T7 channels are the most valuable for the pain event marker detection and RMS, standard deviation and the mean parameters in delta, theta and gamma rhythms as well as Higuchi coefficient and SVD entropy in gamma rhythm can be marked as pain markers of newborns. If the task is to detect painful events without a specific time point, features that can monitor EEG signal change should be included in the models.

Features collected using the Forward Feature Selection method were also analyzed and feature importance for each machine learning algorithm was also estimated. A similar distribution of features can be seen, but divergence with the statistical measures should also be noted. Feature importance also differs across different machine learning techniques.

The same analysis was performed for the adult EEG signals, which showed a slight correlation between the most significant features obtained for newborns. The Cz and T7 channels have also proven to be valuable channels for collecting pain-related markers in adults. RMS and standard deviation have shown themselves as valuable features for classifying painful events in adult EEG signals. This shows the potential of using adults data for pre-training of the machine learning models to improve the classification results, especially in a task where limited amount of data is crucial like neural networks studying, but also highlights the need for detailed experiment design data selection in such cases.

## 3 Discussion

3757 features were calculated using different techniques in time, frequency and time-frequency domains. Statistical measures such as Student's t-test were chosen as a feature selection approach. The Forward Feature Selection approach showed correlation between features selected for different machine learning algorithms and the statistical analysis, but unique features for each machine learning algorithm with a low t-criteria can also be noted. Also, the number and value of features selected using the Forward Feature Selection approach depends on the selected machine learning method, which complicates the definition of feature selection. On the other hand, Forward Feature Selection is a required step in statistical measures and can help to identify non-obvious features that are selected across multiple machine learning methods, as well as defining the initial number of the features for analysis. So, statistical measures can be a more valuable approach for selecting features in the task of identifying pain-related markers of newborns, but may also be improved using Forward Feature Selection.

In a task of classification of EEG signals containing painful event and those without painful event it was noted the high level of errors in case of using the full set of data without separating a subset of "mid pain level" (PIPP scores from 6 to 12) and "high pain level" (PIPP scores higher than 12) groups of signals containing painful events and "low pain level" (PIPP score from 0 to 5) group of signals without painful events which may be related to the data collection methods and may be proved by analyzing stimulation information for the signals not containing painful events where PIPP scores were distributed across all pain levels. This shows the need to collect more data from newborns for further research.

Features selected using statistical measures and trained machine learning models showed that features that focus on band power features in alpha, beta, and gamma frequency bands show good potential as a pain marker for the pain detection of newborns. The same analysis of the adult EEG signals showed a slight correlation between features, where Cz and T7 can be used to collect pain-related markers, and RMS and standard deviation are valuable features.

## Conclusions

This study analyzed the total amount of 3757 features derived from time and time-frequency domains with a focus on band power features in alpha, beta, and gamma frequency bands for the two classification task: pre- and post-stimulus EEG segments classification (painless and painful state of the patient) and EEG signals classification containing painful event (heel lance) and those without painful event (audio control). A combined forward feature selection and statistical measures approach was used to collect 20 features that can be named as pain-related markers of newborns. Statistical measures of features is a valuable approach for selecting features in the task of identifying pain-related markers of newborns.

6 machine learning algorithms were trained with the selected features in the two aforementioned classification tasks. For the task of pre- and post-stimulus EEG segments classification, the C-SVC algorithm showed the best accuracy estimate of 93.5% with a pre-painful event segment classification accuracy of 100% and a post-painful event segment classification accuracy of 86.9%. For the task of EEG segments classification containing painful event and without painful event, after separating a subset of "mid pain level" (PIPP scores from 6 to 12) and "high pain level" (PIPP scores higher than 12) groups of signals containing painful events and "low pain level" (PIPP score from 0 to 5) group of signals without painful events, LD algorithm showed the best accuracy estimate of 84% with 91.6% of correctly identified EEG segments containing painful event and 76.9% of correctly identified EEG segments without painful event. The study shows that features that focus on band power features in alpha, beta, and gamma frequency bands show good perspective as a pain marker for the pain detection of newborns.

In conclusion, this research demonstrates the potential of using features that focus on band power features in alpha, beta, and gamma frequency bands and machine learning techniques for advancing pain detection in neonates. However, the challenges associated with data availability highlight the need for continued efforts to develop larger, high-quality neonatal EEG datasets. These efforts are crucial for improving pain diagnosis and management strategies in this vulnerable patient population.

## References


[1] Witt, N., Coynor, S., Edwards, C., Bradshaw, H. (2016). A Guide to Pain Assessment and Management in the Neonate. *Curr Emerg Hosp Med Rep*, Vol. 4, pp. 1-10. PMID: 27073748; PMCID: PMC4819510. doi: 10.1007/s40138-016-0089-y.

[2] Norman, E., Rosén, I., Vanhatalo, S. et al. (2008). Electroencephalographic Response to Procedural Pain in Healthy Term Newborn Infants. *Pediatr Res*, Vol. 64, pp. 429–434. doi: 10.1203/PDR.0b013e3181825487.

[3] Van der Vaart, M., Hartley, C., Baxter, L., Mellado, G. S. et al. (2022). Premature infants display discriminable behavioral, physiological, and brain responses to noxious and non-noxious stimuli. *Cereb Cortex*, Vol. 32, Iss. 17, pp. 3799-3815, PMID: 34958675; PMCID: PMC9433423. doi: 10.1093/cercor/bhab449. Erratum in: Cereb Cortex, Vol. 32, Iss. 9, 2056. doi: 10.1093/cercor/bhac138.

[4] Talebi, S., Frounchi, J., Tazehkand, B. M. (2022). A Novel Channel Selection Approach for Human Neonate's Pain EEG Data Analysis. *Research Square*. doi: 10.21203/rs.3.rs-2390234/v1.

[5] Thomas, K. R., Sudhakaran, P., Emerson, S., Rahul, R, Kurian, S. M. (2021). Automatic Neonatal Pain Detection for Pediatrics using CNN. *International journal of engineering research & technology (IJERT) ICCIDT*, Vol. 09, Iss. 07. doi: 10.17577/IJERTCONV9IS07005.

[6] Jones, L., Laudiano-Dray, M., Whitehead, K. et al. (2018). EEG, behavioural and physiological recordings following a painful procedure in human neonates. *Sci Data*, Vol. 5, 180248. doi: 10.1038/sdata.2018.248.

[7] Tiemann, L., Hohn, V. D., Ta Dinh, S. et al. (2018). Distinct patterns of brain activity mediate perceptual and motor and autonomic responses to noxious stimuli. *Nat Commun*, Vol. 9, 4487. doi: 10.1038/s41467-018-06875-x.

[8] Bajaj, N., Carrion, J. R., Bellotti, F., Berta, R., De Gloria, A. (2020). Automatic and tunable algorithm for EEG artifact removal using wavelet decomposition with applications in predictive modeling during auditory tasks. *Biomedical Signal Processing and Control*, Vol. 55, 101624. doi: 10.1016/j.bspc.2019.101624.

[9] Xiang, J., Maue, E., Fan, Y., Qi, L., Mangano, F. T., et al. (2020). Kurtosis and skewness of high-frequency brain signals are altered in paediatric epilepsy. *Brain Commun*, Vol. 2, Iss. 1, fcaa036. doi: 10.1093/braincomms/fcaa036.

[10] Gladun, K. V. (2021). Higuchi Fractal Dimension as a Method for Assessing Response to Sound Stimuli in Patients with Diffuse Axonal Brain Injury. *Sovrem Tekhnologii Med*, Vol. 12, Iss. 4, pp. 63-70. doi: 10.17691/stm2020.12.4.08.

[11] Yogarajan, G., Alsubaie, N., Rajasekaran, G., Revathi, T., Alqahtani, M. S., et al. (2023). EEG-based epileptic seizure detection using binary dragonfly algorithm and deep neural network. *Sci Rep*, Vol. 13, 17710. doi: 10.1038/s41598-023-44318-w.

[12] Hjorth, B. (1970). EEG analysis based on time domain properties. *Electroencephalog Clin Neurophysiol*, Vol. 29, Iss. 3, pp. 306-310. doi: 10.1016/0013-4694(70)90143-4.

[13] Lal, U., Chikkankod, A., Longo, L. (2023). Leveraging SVD Entropy and Explainable Machine Learning for Alzheimer's and Frontotemporal Dementia Detection using EEG. *TechRxiv*. doi: 10.36227/techrxiv.23992554.v2.

[14] Grigg, O. A., Farewell, V. T., Spiegelhalter, D. J. (2003). Use of risk-adjusted CUSUM and RSPRT charts for monitoring in medical contexts. *Stat Methods Med Res*, Vol. 12, Iss. 2, pp. 147-70. doi: 10.1177/096228020301200205.

[15] Andrade, C. (2021). Z Scores, Standard Scores, and Composite Test Scores Explained. *Indian J Psychol Med*, Vol. 43, Iss. 6, pp. 555-557. doi: 10.1177/02537176211046525.

[16] Mishra, P., Singh, U., Pandey, C. M., Mishra, P., Pandey, G. (2019). Application of student's $t$-test, analysis of variance, and covariance. *Ann Card Anaesth*, Vol. 22, Iss. 4, pp. 407-411. doi: 10.4103/aca.ACA_94_19. PMID: 31621677; PMCID: PMC6813708.

[17] Abdulmohsin, H., Abdul, H., Hossen, A. (2021). A New Hybrid Feature Selection Method Using T-test and Fitness Function. *Computers, Materials & Continua*, Vol. 68, Iss. 3, pp. 3997-4016. doi: 10.32604/cmc.2021.014840.

[18] Chen, W., Cai, Y., Li, A. et al. (2023). EEG feature selection method based on maximum information coefficient and quantum particle swarm. *Sci Rep*, Vol. 13, 14515. doi: 10.1038/s41598-023-41682-5.

[19] Pan, C., Shi, C., Mu, H., Li, J., Gao, X. (2020). EEG-Based Emotion Recognition Using Logistic Regression with Gaussian Kernel and Laplacian Prior and Investigation of Critical Frequency Bands. *Applied Sciences*, Vol. 10, Iss. 5, 1619. doi: 10.3390/app10051619.

[20] Fu, R., Tian, Y., Bao, T., Meng, Z., Shi, P. (2019). Improvement Motor Imagery EEG Classification Based on Regularized Linear Discriminant Analysis. *J Med Syst*, Vol. 43, Iss. 6, 169, doi: 10.1007/s10916-019-1270-0. PMID: 31062175.

[21] Ginting, A. S., Simanjuntak, R. M., Lumbantoruan, N. and Sitanggang D. (2024). EEG Signal Classification using K-Nearest Neighbor Method to Measure Impulsivity Level. *Jurnal Sisfokom*, Vol. 13, Iss. 2, pp. 261-266. doi: 10.32736/sisfokom.v13i2.2154.

[22] Himalyan, S., Gupta, V. (2022). Support Vector Machine-Based Epileptic Seizure Detection Using EEG Signals. *Engineering Proceedings*, Vol. 18, Iss. 1, 73. doi: 10.3390/ecsa-11-20506.

[23] Edla, D., Mangalorekar, K., Dhavalikar, G., Dodia, S. (2018). Classification of EEG data for human mental state analysis using Random Forest Classifier. *Procedia Computer Science*, Vol. 132, pp. 1523-1532. doi: 10.1016/j.procs.2018.05.116.

[24] Machado, J., Balbinot, A. (2014). Executed Movement Using EEG Signals through a Naive Bayes Classifier. *Micromachines*, Vol. 5, Iss. 4, pp. 1082-1105. doi: 10.3390/mi5041082.

[25] Bigoni, C., Cadic-Melchior, A., Morishita, T., Hummel, F. C. (2023). Optimization of phase prediction for brain-state dependent stimulation: a grid-search approach. *J Neural Eng*, Vol. 20, Iss. 1. doi: 10.1088/1741-2552/acb1d8. PMID: 36626830.

[26] Kaneko, H. (2022). Cross-validated permutation feature importance considering correlation between features. *Analytical Science Advances*, Vol. 3, Iss. 9-10, pp. 278–287. doi: 10.1002/ansa.202200018.


## Вибір ознак для класифікації електричної активності мозку у новонароджених у разі больових подій

*Бондарев В. Р., Іванько К. О., Попов А. О, Карплюк Є. С., Корогод Н. С.*

Розуміння механізмів болю та виявлення больових подій у новонароджених є критично важливим, оскільки новонароджені не мають здатності до вербального спілкування, щоб повідомляти про свої больові відчуття. Це дослідження зосереджено на аналізі особливостей електричної активності мозку новонароджених в часовій та часово-частотній областях з використанням електроенцефалографічних (ЕЕГ) сигналів, зареєстрованих під час больових подій, а саме забору крові шляхом проколу шкіри у новонароджених із використанням п'яткового ланцету. З метою автоматизованого виявлення больових подій у новонароджених було побудовано масив ознак із використанням широкого набору методів розрахунку ознак в часовій та часо-частотній областях і застосовано комбінований підхід до їх відбору, що включає метод прямого відбору ознак та статистичні оцінки. Для вирішення задачі виявлення больових подій за даними аналізу електричної активності мозку новонароджених було використано та порівняно шість алгоритмів машинного навчання, а саме: логістичну регресію, модель на основі лінійного дискримінантного аналізу, метод K-найближчих сусідів, метод опорних векторів, випадковий ліс дерев рішень та наївний байєсів класифікатор. Для задачі класифікації сегментів ЕЕГ як таких, що були зафіксовані до та після больового стимулу, модель на основі методу опорних векторів показала точність 93,5% правильно класифікованих ЕЕГ сегментів. Для задачі класифікації ЕЕГ-сигналів, що відповідають наявності такої больової події як прокол шкіри, та ЕЕГ сигналів у стані спокою, модель на основі лінійного дискримінантного аналізу показала найкращу оцінку точності у 84%, при цьому точність класифікації ЕЕГ сигналів, що відповідають наявності больової події становить 76,9%, а точність розпізнавання ЕЕГ у стані спокою склала 91,6%. Результати демонструють потенціал використання ознак, що зосереджені на спектральній потужності в альфа-, бета- та гамма-діапазонах частот, та методів машинного навчання для покращення виявлення болю у новонароджених.

*Ключові слова:* електроенцефалографія; больові маркери; новонароджені; машинне навчання; вибір ознак; класифікація; точність класифікації; аналіз біосигналів