Feature Selection for Automatic Burst Detection in Neonatal Electroencephalogram

Sourya Bhattacharyya, Arunava Biswas, Jayanta Mukherjee, *Senior Member, IEEE*, Arun Kumar Majumdar, *Senior Member, IEEE*, Bandana Majumdar, Suchandra Mukherjee, and Arun Kumar Singh

***Abstract*—Monitoring neonatal electroencephalogram (EEG) signal is useful in identifying neonatal convulsions which might be clinically invisible. Presence of burst suppression pattern in neonate EEG is a clear indication of epilepsy. Visual identification of burst patterns from recorded continuous raw EEG data is time consuming. On the other hand, automatic burst detection techniques mentioned in the standard literature mostly rely on comparison with respect to predefined static voltage or energy thresholds, thus becoming too specific. Burst detection using ratio information of quantitative feature values between burst segment and neighborhood background EEG segment is proposed in this paper. Features like ratio of mean nonlinear energy, power spectral density, variance and absolute voltage, when applied as an input to a support vector machine (SVM) classifier, provides high degree of separability between burst and normal (nonburst) EEG segments. Exhaustive simulation using various literature specified features and proposed feature combinations shows that the proposed feature set provides best classification accuracy compared to other reported burst detection methods. The results documented in this paper can be used as a reference of optimum quantitative EEG feature sets for distinguishing between burst and normal (nonburst) EEG segments.**

***Index Terms*—Electroencephalogram (EEG), EEG burst suppression, mean nonlinear energy (MNLE), neonatal intensive care unit (NICU), power spectral density (PSD), support vector machine (SVM).**

I. Introduction

ELECTROENCEPHALOGRAM (EEG) monitors cerebral electrical activities through electrodes placed on scalp and provides a sensitive real time graphical representation of brain function [1]. Especially for neonates, neurophysiological disorders and seizures are mostly diagnosed by visual inspection of EEG signals. The reason behind this is that, unlike the seizure cases in adults or matured children, neonates commonly do not exhibit clinical sign and symptoms [2] for seizures. So, proper recording and analysis of EEG signals is extremely essential for monitoring brain function in neonates. The frequency of neonatal EEG lies within the range 0.1–100 Hz [1]. However, the low frequency range of 0.1–0.5 Hz is dominated by sweating artifacts [13].

Fig. 1. Multichannel normal or nonburst EEG patterns. Display sensitivity is 5 μV/mm.

EEG waves are classified according to their frequency as—β (>13 Hz), α (8–13 Hz), θ (5–7 Hz), and δ (0–4 Hz). Usually, types of activities seen in neonatal EEG are in the frequency ranges of θ and δ. α frequency components are rather infrequent. However, in case of seizures, neonatal EEG can exhibit much more occurrences of α frequency components. For neonatal EEG, β activities, especially in high β frequency range (>17 Hz) occur rarely [1], [3]. In newborns, EEG patterns that are free from any burst or artifact or any other stereotyped seizure patterns are called normal or background patterns [1]. In the current study, we refer such patterns as nonburst EEG. The frequency range of such patterns are mostly confined within 0.5–30 Hz [1], [4]. Fig. 1 shows an example of multichannel normal or background EEG segment.

A burst pattern is basically a high energy activity, usually lasting 1–10 s and characterized by high voltage (75–250 μV) θ and δ activity with intermixed spikes and waves. On the other hand, a suppression pattern is characterized by long periods (usually >10 s) of low amplitude (<5 μV) [4]. Repeated occurrences of burst-suppression patterns produce burst suppression cycle or event, which indicates seizure. For neonatal EEG, burst-suppression is the most frequently observed seizure pattern. However, existence of a burst pattern is not always associated with suppression. As an example, burst patterns, with and without suppression, are shown in Figs. 2 and 3, respectively.

Manuscript received June 13, 2011; revised October 13, 2011; accepted November 28, 2011. Date of publication January 17, 2012; date of current version February 01, 2012. This work was supported by the Ministry of Communications and Information Technology, Government of India under project Grant 1(4)/2009-ME&TMD. This paper was recommended by Guest Editor J. Sanchez.

S. Bhattacharyya, J. Mukherjee, A. Majumdar, and B. Majumdar are with the Department of Computer Science and Engineering, Indian Institute of Technology, Kharagpur WB 721302, India (e-mail: sourya.bhatta@gmail.com; jay@cse.iitkgp.ernet.in; akmj@cse.iitkgp.ernet.in; bandana.majumdar@gmail.com).

A. Biswas is with the School of Medical Science and Technology, Indian Institute of Technology, Kharagpur WB 721302, India (e-mail: mail2arunavabiswas@gmail.com).

S. Mukherjee and A. Singh are with the Department of Neonatology, Institute of Post Graduate Medical Education and Research and Seth Sukhlal Karnani Memorial Hospital, Kolkata WB 700020, India (e-mail: drsmukherjee70@gmail.com; drarunsingh61@yahoo.co.in).

Digital Object Identifier 10.1109/JETCAS.2011.2180834

2156-3357/$26.00 © 2011 IEEE

Fig. 2. Multichannel burst patterns (within 4.4–7.6 s time interval) associated with clear suppression patterns. Deep vertical lines represent 1-s time interval and dotted vertical lines represent 0.2-s time interval. Display sensitivity is 5 μV/mm.

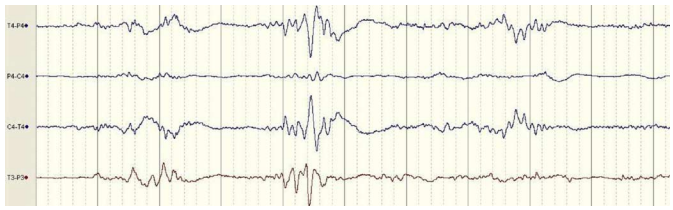

Fig. 3. Multichannel burst patterns (within 3.8–5.4 s time interval) associated with background EEG (but not suppression) patterns. Deep vertical lines represent 1 s time interval and dotted vertical lines represent 0.2-s time interval. Display sensitivity is 5 μV/mm.

Visual detection of burst patterns is very much time consuming and also subjective to the respective viewer [5]–[7]. Thus, considering huge amount of recorded data, replacement of visual inspection by automated detection of burst patterns is necessary. In the following sections, past works regarding automatic burst detection are summarized first. Next, EEG dataset acquisition and labeling process are described. Features used in the current study are explained and they are compared with the features mentioned in existing literature. Several feature selection methods are discussed, followed by the description of classification methods using different feature combinations and comparison of their relative performances. Obtained accuracy of best feature set is compared with existing burst detection methodologies. Finally the application of this method over a long continuous recorded data set is briefly outlined.

II. Related Works

Burst detection algorithms mentioned in the existing literature, along with their relative advantages and disadvantages, are summarized in Table I. The methods fall in either of the following two categories:

- supervised learning with feature based classification;
- static amplitude or energy threshold based comparison.

A. Supervised Learning Based Burst Detection

This approach uses feature based classification between burst and normal segments. Various time domain, frequency domain, and statistical features are computed and given as the inputs to a support vector machine or a neural network classifier [23] for burst detection.

1) In the method discussed in [11], values of features like Spectral Edge Frequency (SEF95) and 10 Hz power are computed using labeled burst and nonburst segments. Those feature values together with the label are used for supervised learning using a neural network (NN) classifier. However, the algorithm does not model the variation of background EEG.
2) Approaches discussed in [5] and [6] compute various frequency domain and statistical features like entropy, variance, kurtosis, skewness, and cepstral coefficients using labeled burst and nonburst segments. The classifier used is Fisher's linear discriminant (FLD) classifier [23]. Here, background information is modeled by taking the median of all the channel data. Sliding window based data segment analysis and feature extraction procedure is used to predict the burst segment.

B. Static Amplitude or Energy Based Burst Detection

In this approach, amplitude or energy values are computed for the current EEG segment. Based on comparison of the computed values with respect to defined static thresholds, existence of a burst pattern is decided.

1) Burst detection using nonlinear energy operator (NLEO) [8], [9] is based on comparison with respect to a predefined static threshold. Here, EEG signal is first divided in two frequency bands—1) EEG band (0.1–8 Hz) obtained by applying first-order high pass butterworth filter and sixth-order low pass elliptic filter [19], and 2) artifact band (47–49 Hz) obtained by applying eighth-order band pass elliptic filter. If x_i is the value of band filtered EEG at ith sample, then NLEO output for that sample is

$$\mathrm{NLEO}(x_i) = x_i x_{i-3} - x_{i-1} x_{i-2}. \tag{1}$$

NLEO outputs for ith sample of EEG band and Artifact band signals are referred as NEB(i) and NAB(i), respectively. If the difference between these two quantities, $\mathrm{DIFF}(i) = \mathrm{NEB}(i) - \mathrm{NAB}(i)$ is persistently greater than certain predefined fixed burst threshold value for a minimum specified duration, the algorithm notifies occurrence of a burst pattern. Similarly, if this difference stays below another fixed suppression threshold for a certain period of time, it indicates the occurrence of a suppression pattern.
2) Method based on integrated instantaneous amplitude and static threshold based comparison [7] calculates the root mean square of all the different channel data to generate the integrated instantaneous amplitude. Then this data is smoothed with respect to a moving window of length 0.5 s and 0 overlap. The amplitude of resulting signal is compared to defined burst and suppression thresholds to get the burst and suppression intervals.

Fig. 4 shows an example of false burst detection, using the NLEO based algorithm [8], [9]. The data is recorded from a six-day-old neonate. The display shown has sensitivity 5 μV/mm and timebase 15 mm/s. Bursts (marked in yellow rectangles) are detected based on static amplitude based thresholding. However from the figure, it is evident that many of the detected burst portions by NLEO algorithm are not clearly distinguishable from background EEG patterns (specially for the channels T4-P4 and C4-T4, which are first and third channels, respectively, from the top of the display). Thus, the number of false positive cases of the burst detection is rather high. Increasing the static threshold

TABLE I
OVERVIEW OF VARIOUS BURST DETECTION ALGORITHMS

Method	Advantage	Disadvantage
Supervised learning with Neural Network Classifier [9]	Demonstration of a supervised learning model with frequency domain features for burst detection.	Not adaptive to varying background EEG, since the visual difference of burst patterns with respect to background EEG is not modeled. Sensitive to different recording channels of varying amplitudes, thus increasing false positive or false negative cases.
Burst detection using non linear energy operator (NLEO) and static threshold based comparison [10], [11]	Burst detection based on comparison to background EEG is modeled by NLEO operator. Furthermore, burst onset time calculation is analyzed deeply in [10] by taking into account the sample-wise results for burst detection.	Recorded EEG signal amplitudes or energy values vary for different neonates. A fixed threshold value based decision thus is not suitable for generalized burst detection over neonates. In addition, ranges of EEG amplitudes or energy values vary on different recording channels. For example, involvement of occipital channels (O1 or O2) results in generation of higher EEG amplitude than EEGs from frontal channel (FP1 or FP2). Thus a fixed threshold cannot properly detect burst in all channels.
Feature based background EEG modeling and burst detection [6], [7]	Background model prediction and burst detection based on sliding window is useful for application in large continuous data. Similar approach is also used in the proposed scheme.	Ratio of feature values between burst and background segments are not experimented. So a trained classifier with input of absolute feature values (rather than using ratio of feature values) can perform poorly for different neonates or different recording channels having varying degree of amplitude or energy information.
Integrated instantaneous amplitude and static threshold based comparison [8]	Burst detection using all the channel data is done. Easy to implement and computationally efficient. It has the provision of dynamically setting the amplitude threshold based on visual interpretation by user.	Once the amplitude threshold is set, it applies simultaneously to the data for all different channels. Thus, this method does not support individual channel data based threshold adaptation, producing variability of results for data across different neonates and also for different recording channels.
Feature ratio comparison with static thresholding [12]	Adaptation with respect to varying background EEG helps to detect bursts in EEGs of varying degree of amplitudes or recording channels.	Static threshold based comparison of the feature ratio values sometimes fail to detect the presence of bursts (false negative case).

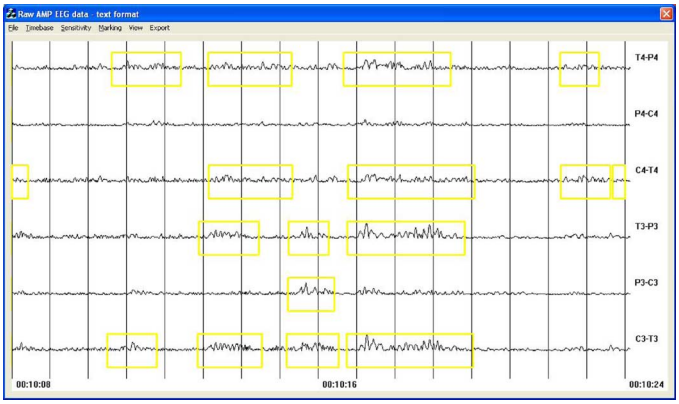

Fig. 4. False burst detection (marked in yellow rectangles) for NLEO based algorithm for a six-day-old neonate, for channels T4-P4 and C4-T4. The display shown has sensitivity 5 μV/mm and time base 15 mm/s. Marked burst portions for the channels are not much distinguishable from background EEG patterns.

value leads to reduction of false positive cases, but simultaneously the chance of false negative case increases.

The major drawback of the mentioned static threshold based burst detection techniques is that, there is no mechanism of individual channel data based threshold adaptation. Thus, they produce variability of results for data across different neonates and also for different recording channels. To overcome this problem, in our earlier work [10], ratio of features, such as mean nonlinear energy (MNLE), power spectral density (PSD), variance and sum of absolute amplitude, are computed in current EEG segment and neighborhood background segment. These ratio information are compared with static predefined thresholds for detection of burst. It provides the adaptation with respect to dynamic background. However, feature ratio values are compared with static thresholds, thus sometimes producing false negative cases.

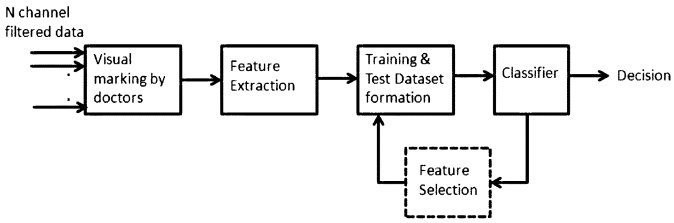

Fig. 5. Schematic of proposed supervised learning based burst detection method.

C. Proposed Model for Burst Detection

A general schematic of supervised learning based burst detection method, using single or multichannel raw EEG data is depicted in Fig. 5 (adapted from the methods in [13], [6], and [12].

Multiple experts independently marked different burst and normal segments within recorded continuous raw EEG by visual detection. Marked data segments for each channel were extracted from the input data. Feature values computed using the data segments along with corresponding labels (burst or normal) constitute the training and test datasets for supervised learning. Various feature combinations were experimented to get the best classification accuracy between burst and normal EEG segments. The performance metric selected in this study is an weighted function of sample sensitivity and specificity values. This metric and analysis procedure are described in Section IV.

In current analysis, the ratio based features mentioned in [10], are combined with some other burst detection features mentioned in standard literature [5], [6], [11], [12]. However, using a huge set of features for classification can lead to poor performance due to feature irrelevance, redundancy and curse of dimensionality. Thus, after feature extraction step, selection of best feature subset [13] in terms of maximum classification

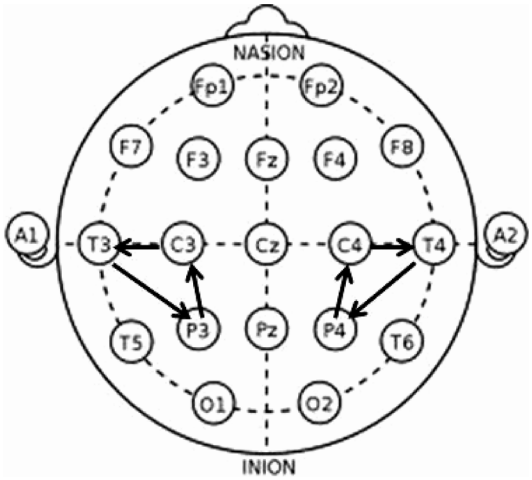

Fig. 6. Bipolar montage used in the EEG recording. The joining arrows indicate two channels participating in bipolar montage configuration.

accuracy is carried out. In this study, three different feature selection methods are experimented—1) exhaustive feature subset generation, 2) backward feature reduction, and 3) principal component analysis [22].

Current study models the transition of feature values from burst to background EEG or vice versa, for discrimination between burst and background EEG patterns. Ratio of features like power spectral density (PSD), mean nonlinear energy (MNLE), variance and sum of absolute voltage between one EEG segment and corresponding background EEG are computed. These values are not compared with any predefined static thresholds—rather they are given as inputs to a support vector machine (SVM) classifier for further generalization and adaptability of the algorithm.

III. Dataset Description

A. Data Acquisition

The current study is performed over EEG data acquired from 12 term newborns. That is, all these newborns are born at a gestational age (the time period between conception and birth during which the fetus grows and develops inside the mother's womb) between 37 and 42 completed weeks. All recordings were made in the Neonatal Intensive Care Unit at Department of Neonatology, SSKM Hospital, Kolkata, India, using NicOne 5.3 EEG system [26]. During data recording, bipolar montage with six electrodes were used, according to international 10–20 standard [1]. Bipolar montage was selected because this montage setting is less susceptible to noise than referential montage [13]. In neonates, seizures can be identified mostly in central (C), temporal (T), and parietal (P) brain regions of neonates [4]. So the electrode montage (shown in Fig. 6) used for recording consisted of six bipolar channels: 1) T4-P4, 2) P4-C4, 3) C4-T4, 4) T3-P3, 5) P3-C3, and 6) C3-T3. Before recording, the electrode sites were cleaned and rubbed until the skin impedance reduced to below 5 kΩ. That was due to the fact that, low, stable, and equal impedance of electrodes ensures rejection of common-mode signals such as 50-Hz mains interference.

Out of the 12 data files recorded from 12 term newborns, three files have duration less than 15 min; four data files have duration between 15 min and 1 h; five data files have duration within 1–5 h. Generally for adult EEG, recording of maximum 1 h is sufficient. But for neonatal EEG, prolonged recordings are used depending upon critical condition of the infant [1]. All the acquired EEG signals have presence of clear burst patterns, sometimes associated with clear suppression patterns as well, thus confirming existence of seizure activity.

B. Data Preprocessing

Out of original recording of 12 data files, seven files had sampling rate of 125 Hz; remaining files had sampling rate of 2000 Hz. EEG processing and viewing in NicOne 5.3 EEG system [26] is done without any down sampling. The variation in sampling rate during data acquisition is caused by recording settings. But it does not affect the artifact detection or any other EEG processing methods. So in our approach, no down sampling operation is performed. As conventional EEG signal analysis methods rely on information extracted from the frequency components above 0.5 Hz and sweat artifacts compromise recordings at the lowest frequencies ($<$0.5 Hz) [13], each EEG data segment was band pass filtered between 0.5–35 Hz. Filter passband step was set as 0.1 Hz. For filtering, the stop band attenuation and the passband ripple were set as 80 dB and 0.1 dB, respectively. An FIR filter of Kaiser window [20] was used for the filtering. To meet the passband step and the ripple specifications, the filter order was set as 101. The design and implementation of the filter along with the filter order specification were performed using the signal processing toolbox of MATLAB 7.8.0.

C. Visual Marking by Experts

After preprocessing of the continuous EEG recordings, each of the data files were visually marked by three doctors independently. For event marking, same display settings were used for all the doctors—filtering in the frequency range 0.5–35 Hz, sensitivity 5 μV/mm and time base 15 mm/s. The doctors marked the burst and normal EEG segments; in addition, some segments were marked as unknown if they were not properly concluded as belonging to either burst, artifact, normal, or any other typical seizure patterns.

From the observations obtained by independent marking of three reviewers, the labeled data set is generated by considering their agreement and disagreement over the class of a segment. As suggested in [8], we have avoided using kappa statistics and instead relied on complete agreement. If a particular segment is labeled unanimously of same type (burst, normal, or unknown) then that segment together with the label is used for further analysis. On the other hand, if at least two different labels are given for a particular data segment then final label of that segment is put as unknown. While marking, minimum duration of burst is taken as 1 s [10]. Value of maximum burst duration has no such predefined constraint, but a few reported work [1], [4] put the general maximum limit as 10 s. This constraint for maximum burst duration is used in the dataset.

Since there are variations among the reviewers in their judgment of onset and termination of burst segment, the final burst

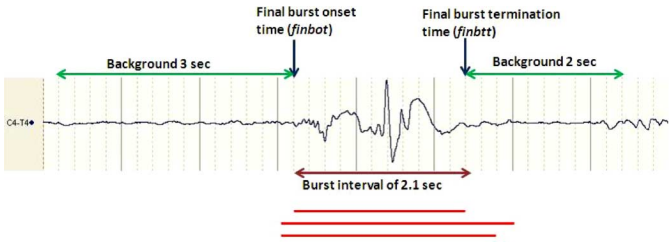

Fig. 7. Example of a marked burst segment of 2.2 s length, preceded by 3 s and followed by 2 s of background EEG information. Deep vertical lines correspond to 1-s interval; dotted vertical lines mean 0.2-s interval.

onset time (*finbot*) taken is taken as the maximum of the individual onset timestamps, while the final burst termination time (*finbtt*) is taken as minimum of the individual timestamps. The interval between these two final timestamp values is treated as the burst interval. The patterns surrounding the marked burst period constitute the background EEG. A schematic of marked burst interval and background EEG boundary is shown in Fig. 7. The red lines correspond to bursts marked by three reviewers, from which the final burst interval is obtained.

D. Extraction of Labeled Short Data Segments

After finalizing the labels of segments, along with the timestamps, short segments of single channel labeled EEG data are extracted from the original data set. Each short segment consists of marked burst or normal segment, preceded by 3 s and followed by 2 s of background EEG information, as shown in Fig. 7. Extraction of background EEG information is required for comparison between quantitative feature values both in burst and background EEG segments. For the final dataset, the marked burst or normal segments which have preceding 3 s and following 2 s of background EEG information (without any other overlapping marked segments) are considered only. If there exists any other labeled segment, overlapping within the boundary of current segment, then none of those segments is considered for final dataset. Overall, a total of 1000 normal EEG segments and 1800 burst segments are generated. In addition, there are total 1000 segments having label of unknown, thus not considered for the final dataset. EEG data set consisting of normal and burst segments is equally divided in training and test data sets.

IV. Methodology

A. Feature Description

A complete list of features used in current study is summarized in Table II. As shown in Fig. 7, any particular feature value for burst or normal segments is computed using the middle portion of extracted data segment. Feature values corresponding to background EEG are computed by considering both the starting 3 s and the final 2 s; thus total $(3+2)=5$ s of signal is treated as background EEG segment. The feature values together with the label constitute the training and test dataset.

In Table II, features 2, 4, 7, and 11 are the proposed ratio information of the mean nonlinear energy (MNLE), mean absolute amplitude, power spectral density (PSD) and variance between burst or normal segments and background EEG segments [10] respectively. These feature values have the maximum classification accuracy, which is discussed in Section V. Features 3, 5, 6, and 10 deal with absolute values of respective features computed in burst or normal EEG segments. Rest of the features are reported in previous works [5], [6], [11], [12].

TABLE II
Features for Classifying Between Burst and Normal Segments of EEG Data (Features Involving Burst/Normal to Background Segment Feature Value Ratio Are 2, 4, 7, 11)

Feature Number	Feature Name
1	Higuchi fractal dimension (HFD) in burst / normal segment
2	Ratio between MNLE within burst / normal segment EEG and background segment EEG
3	Mean of absolute voltage value in burst / normal segment
4	Ratio between mean value of absolute voltage within burst / normal segment EEG and background segment EEG
5	MNLE within burst / normal segment
6	Mean of PSD within burst / normal segment
7	Ratio between mean value of PSD within burst / normal segment EEG and background segment EEG
8	3 Hz power in burst / normal segment
9	10 Hz power in burst / normal segment
10	Variance in burst / normal segment
11	Ratio between variance within burst / normal segment EEG and background segment EEG
12	Kurtosis in burst / normal segment
13	95% Spectral Edge Frequency (SEF95) in burst / normal segment
14	Shannon entropy in burst / normal segment
15	Skewness in burst / normal segment

Higuchi fractal dimension (HFD) (feature 1) is a measure of signal complexity [15]. Specifically, it is a statistical quantity that gives an indication of how completely a fractal appears to fill space, as one zooms down to finer and finer scales. HFD is computed following the algorithm described in [14].

If an EEG segment has N samples and x_i is ith sample of this EEG segment, and μ is the sample mean of the segment, then mean nonlinear energy (MNLE), variance (VAR), and mean of absolute value of signal amplitudes (ABSAMP) of this EEG segment [10], [12], [5] are defined below

$$\text{MNLE} = \frac{\sum \left(x_i^2 - x_{i-1}x_{i+1}\right)}{N} \tag{2}$$

$$\text{VAR} = \frac{\sum (x_i - \mu)^2}{N-1} \tag{3}$$

$$\text{ABSAMP} = \frac{\sum \text{abs}(x_i)}{N}. \tag{4}$$

The feature 6, power spectral density (PSD) shows the distribution of signal power with respect to frequency [16]. Power in 3 and 10 Hz frequency [11] are measured by first computing the PSD of current segment using a fast fourier transform (FFT) [16]. For input data having sampling rate of 125 and 2000 Hz, FFT window length applied is 128 and 2048, respectively. A Hamming window is employed during FFT computation. Power at 3 and 10 Hz frequencies are derived using power spectrum at those frequencies surrounded by 1 Hz bandwidth. So power spectrum within 2.5–3.5 Hz and 9.5–10.5 Hz are used to compute 3 and 10 Hz power, respectively.

The feature 13, spectral edge frequency (SEF95) [11] is the frequency under which 95% of the signal power resides. Computation of SEF95 is done using PSD. Kurtosis and skewness are

measures of "peakedness" and asymmetry, respectively, for the probability distribution of a real-valued random variable [17]. Shannon entropy is a measure of uncertainty of a random variable [6], [12].

B. Classifier

To test the classification performance with the above mentioned feature set, a binary classification scheme using support vector machine (SVM) [21] is used. The classifier output is binary—0 signifies normal EEG segment and 1 indicates burst EEG segments. SVM maps input data to a higher dimensional feature space through some selected nonlinear mapping method or kernel. A linear decision surface is then constructed in this high dimensional feature space [18]. Kernel function for a certain problem is dependent on the specific data. Here, radial-basis function (RBF) kernel function is selected as the nonlinear mapping method. If inputs in original parameter space and support vectors in the nonlinear feature space are denoted by x and x' respectively, then the RBF kernel [22] is defined by following equation:

$$\mathrm{RBFK}(x, x') = \exp\left[-\frac{\|x - x'\|^2}{\sigma^2}\right]. \tag{5}$$

Here, parameter σ, supplied by the user, acts as the scaling factor or radius of RBF kernel. Small value of σ produces smooth decision boundaries, preventing overfitting of the model to the data samples used in training procedure. But the decision boundary can have less than optimum complexity, yielding underfitting of the model. On the other hand, high value of σ generates complex decision boundaries which sometimes can be too specific to the training examples, resulting overfitting of the model. In our study, we used value of σ from the set $S_\sigma = [0.1, 0.2, 0.3, 0.4, 0.5, 0.6, 0.7, 0.9, 1, 2, 3, 4, 5]$, covering range from small to high values. All the feature extraction and classifier implementation procedures are done using MATLAB version 7.8.0.

C. Performance Metric

For performance evaluation of SVM classification, sensitivity and specificity values (as defined below) are used as performance metrics

$$\text{Sensitivity} = \frac{\mathrm{TP}}{\mathrm{TP} + \mathrm{FN}} \tag{6}$$

$$\text{Specificity} = \frac{\mathrm{TN}}{\mathrm{TN} + \mathrm{FP}}. \tag{7}$$

Here, true positive (TP) means that a burst segment is correctly detected. True negative (TN) means that a normal (non-burst) segment is correctly detected. False positive (FP) means that a normal segment is incorrectly detected as burst segment. Finally, false negative (FN) means that a burst segment is incorrectly detected as normal segment.

For each feature, variation of output sensitivity with respect to output specificity for different input σ is plotted in Fig. 8. Here, only the features showing significant values of sensitivity and specificity are used for comparison.

From Fig. 8, it is evident that classifier using feature 2 (ratio of MNLE between burst/normal segment to background segment) has the best classification performance. Classifiers generated by using features 11 (ratio of variance between burst/normal segment to background segment), 5 (MNLE within burst/normal segment), 7 (ratio of PSD between burst/normal segment to background segment) also have quite good classification performances. Features 9 (10 Hz power) and 4 (ratio of mean absolute amplitude between burst/normal segment to background segment) also perform well to some extent. Thus, our proposed features, even when executed alone, perform very well in discriminating burst from normal segments.

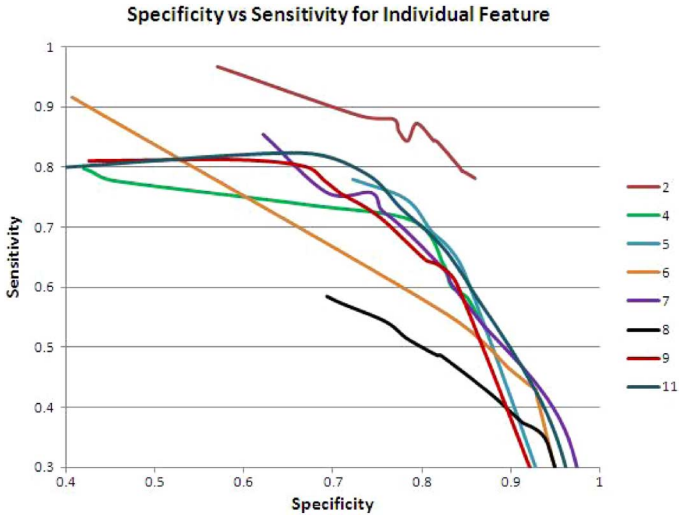

Fig. 8. Comparison of specificity versus sensitivity for different quantitative EEG features.

Furthermore, it is observed that for different σ, specificity and sensitivity values can considerably vary, even when a fixed feature is used for classification. Thus the performance metric should be a function related to both sensitivity and specificity values. Two such different metrics are mentioned below.

1) Weighted sensitivity plus specificity (WSS): It is a linear combination of sensitivity and specificity. This metric would enable the users to vary relative importance to sensitivity and specificity during comparison

$$\mathrm{WSS} = \alpha * \text{sensitivity} + (1 - \alpha) * \text{specificity} \tag{8}$$

where $0 \leq \alpha \leq 1$. For equal weights given to both sensitivity and specificity, α is set as 0.5.

2) F score: It is the harmonic mean of precision and sensitivity values. Precision is the accuracy to predict the true positive (TP) while minimizing the prediction of false positive (FP). Higher F score confirms correct detection, both in positive (burst EEG) and negative (normal EEG) cases

$$\text{Precision} = \frac{\mathrm{TP}}{\mathrm{TP} + \mathrm{FP}} \tag{9}$$

$$\text{F score} = \frac{2 \times \text{Precision} \times \text{Sensitivity}}{\text{Precision} + \text{Sensitivity}}. \tag{10}$$

For feature based classification, due to variation of sensitivity and specificity over different σ, corresponding WSS and F score also vary. Figs. 9 and 10 depict such variations. Only the features showing significant values of the evaluation metrics are used in

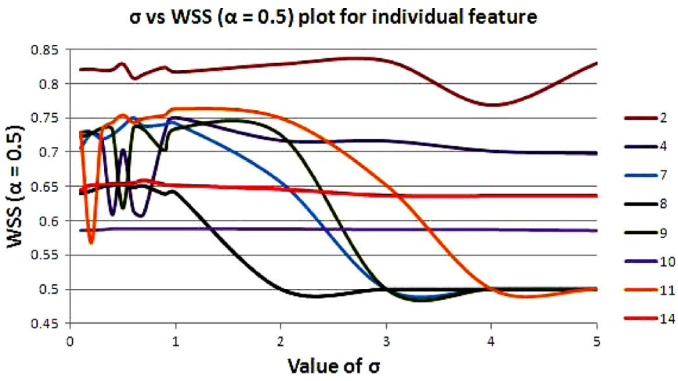

Fig. 9. Comparison of WSS versus σ for different quantitative EEG features.

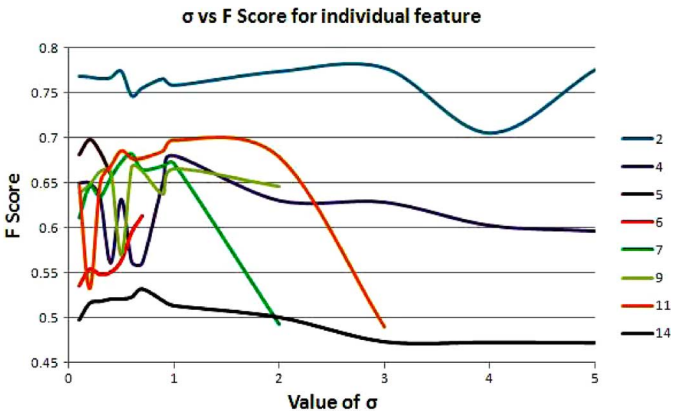

Fig. 10. Comparison of F score versus σ for different quantitative EEG features.

the plots. From the figures, features having small variation of WSS and F score with respect to σ may be considered as the prime discriminator between burst and normal segments. On the other hand, classification scheme based on features that produce large variation of WSS or F score with changing σ may not be robust.

From Figs. 9 and 10, it is evident that the classifier using the feature 2 produces best values of WSS and F score, with consistent high scores for different σ. The classifier using feature 4 does also produces quite consistent high values of WSS and F score for different σ. These observations justify the utility of using the ratio of feature values in burst detection. Performances of all other feature based classifiers are sensitive to changes in σ, indicating that changes in test data can produce variations in their execution performances.

Of course, rather than using a single feature, combination of several features may produce better output metric values. Generation of different feature combinations is discussed in the next section. To compare the performance of different feature combinations, following output metrics are used.

(M_1) Average F score per feature combination: It is the mean of generated F score values for different σ during execution of a particular feature combination. Higher value of this parameter signifies consistent performance of that particular feature combination even in changing σ, thus being more reliable to unknown test dataset.

(M_2) Average WSS value per feature combination: Mean of generated WSS values (with $\alpha = 0.5$) for different σ with a particular feature combination. Higher value of this parameter indicates higher reliability to unknown test data.

(M_3) Best F score per feature combination: It is the maximum F score that can be obtained with a particular feature combination over all input σ parameters.

(M_4) Best WSS value per feature combination: Maximum WSS value (with $\alpha = 0.5$) that can be obtained for a particular feature combination over all input σ parameters.

D. Feature Selection

After defining the output metrics M_1 to M_4 for comparing the performances of different feature combinations, the next goal is to find the feature subsets which produce best values for these metrics. From the various feature selection methods mentioned in the literature [13], [22], [24], [25], following approaches are used in the current study.

1) Exhaustive Feature Subset Generation: A brute force approach is to carry out exhaustive generation of all the feature subsets. For example, with a feature set of cardinality 15, total number of unique feature subsets that can be formed is

$$\sum_{r=1}^{15} \binom{15}{r} = (2^{15} - 1).$$

Here, index r is the cardinality of a feature subset, with value of r varying from 1 to 15.

The objective of current study is to find the feature set which provides best output metric value. If two feature sets simultaneously produce best output metric value, then the feature set with lower cardinality is selected. In this quest, for each r (r varies from 1 to 15), feature sets of cardinality r producing best output metric value are identified. Since there are four different output metric parameters M_1 to M_4, so for each r, at most four different best performing feature sets are identified. Suppose f_{ir} denotes the best performing feature set for a particular r and output metric M_i, where value of r varies from 1 to 15, and i varies from 1 to 4. The optimum discriminator for a particular output metric M_i $(1 \leq i \leq 4)$ is the feature set F_i having maximum output value among the feature sets $[f_{i1}, f_{i2}, \ldots, f_{i15}]$. Finally we obtain at most four distinct optimum discriminators F_1 to F_4 corresponding to four different output metrics M_1 to M_4.

2) Backward Feature Subset Selection: Backward feature subset selection [22] is a greedy method for feature subset selection—it discards the features one by one at each step of execution, provided that removing this feature does not reduce the classification accuracy. The steps of this algorithm is described in Algorithm 1.

TABLE III
PERFORMANCE BY FEATURE SUBSETS GENERATED BY EXHAUSTIVE COMBINATION, FOR DIFFERENT OUTPUT METRICS

No of features (r)	Best average F score (M_1) per feature combination (covering all σ cases)	{Feature set having best average F score (F_1)}	Best average WSS with α = 0.5 (M_2) per feature combination (covering all σ cases)	{Feature set having best average WSS (F_2)}	Max F score (M_3) per feature combination	{Feature set having max F score (F_3)}, corresponding σ value	Max WSS with α = 0.5 (M_4) per feature combination	{Feature set having max WSS (F_4)}, corresponding σ value
1	0.76	{2}	0.82	{2}	0.78	{2}, 3	0.83	{2}, 3
2	0.8	{2, 4}	0.85	{2, 4}	0.81	{2, 4}, 0.6	0.85	{2, 4}, 2
3	0.8	{2, 4, 6}	0.85	{2, 4, 6}	0.81	{2, 4, 10}, 0.9	0.86	{2, 4, 10}, 0.9
4	0.8	{2, 4, 6, 9}	0.85	{2, 4, 6, 9}	0.82	{2 to 4, 6}, 0.6	0.86	{2 to 4, 6}, 0.6
5	0.8	{2, 4, 6, 7, 9}	0.85	{2, 4 to 6, 9}	0.82	{2, 4, 5, 10, 14}, 1	0.86	{2, 4, 5, 10, 14}, 1
6	0.8	{2, 4 to 6, 8, 9}	0.85	{2, 4, 5, 7, 9, 11}	0.82	{2 to 6, 10}, 0.4	0.86	{2, 4, 5, 9, 10, 14}, 0.6
7	0.8	{2, 4 to 9}	0.85	{2, 4 to 7, 9, 11}	0.82	{2, 4 to 6, 9 to 11}, 0.4	0.87	{2, 4 to 6, 9 to 11}, 0.4
8	0.8	{2, 4 to 9, 11}	0.85	{2, 4 to 9, 11}	0.83	{2, 4 to 6, 8, 10, 12, 14}, 0.9	0.86	{2, 4 to 6, 8, 10, 12, 14}, 0.9
9	0.8	{2 to 7, 9 to 11}	0.85	{2 to 7, 9 to 11}	0.83	{2, 4 to 8, 10, 12, 14}, 0.1	0.86	{2, 4 to 8, 10, 12, 14}, 0.1
10	0.79	{2 to 11}	0.84	{2 to 11}	0.83	{2, 4 to 8, 10 to 12, 14}, 0.9	0.86	{2, 4 to 8, 10 to 12, 14}, 0.9
11	0.76	{2 to 12}	0.82	{2 to 12}	0.82	{2 to 10, 12, 14}, 0.9	0.86	{2 to 10, 12, 14}, 0.9
12	0.72	{2 to 12, 14}	0.8	{2 to 12, 14}	0.82	{2 to 12, 14}, 0.9	0.86	{2 to 12, 14}, 0.9
13	0.67	{1 to 12, 14}	0.78	{1 to 12, 14}	0.82	{1 to 10, 12 to 14}, 2	0.86	{1 to 10, 12 to 14}, 2
14	0.62	{1 to 14}	0.75	{1 to 14}	0.82	{1 to 14}, 2	0.85	{1 to 14}, 2
15	0.57	{1 to 15}	0.73	{1 to 15}	0.8	{1 to 15}, 2	0.84	{1 to 15}, 2

Algorithm 1 Pseudo code of backward feature subset selection method

1: step 1—Calculate the value of M for current feature set.
2: **for** All the features within the current feature set **do**
3: Remove one feature at a time.
4: Using remaining features, calculate value of M for different σ.
5: **end for**
6: step 2—Discard the feature whose removal produced the maximum value of M for a particular σ. Note down the value of M along with particular σ.
7: Current feature set will consist of the remaining features. Go to step 1 if current feature set has cardinality greater than 1.
8: Finally select the feature set and the σ for which the best value of M is produced. In case of a tie, feature set with lower cardinality is selected.

3) Principal Component Analysis: For a feature set of dimensionality d, the features themselves may be correlated, carrying redundant or irrelevant information. Principal component analysis (PCA) is an unsupervised feature extraction method [22] which generates a new set of k features called principal components $(\mathrm{k} \leq \mathrm{d})$. Each component is a combination of original d features. The components are mutually orthogonal (uncorrelated) and they are sorted in the order of decreasing variance. Using only a few components, almost whole variability of the input data can be covered, leading to optimization of the feature set cardinality. Generated principal components can be fed as inputs to the SVM for classification. Here, principal components are generated out of total 15 features.

The above mentioned feature selection methods may produce different feature subsets of reduced cardinality. Each of these feature subsets are expected to provide maximum classification accuracy (measured as per a selected metric, e.g., $M_1, \ldots, M_4$). Similarly for PCA, the number of principal components which are used for final classification, varies according to the classification metrics $(M_1, \ldots, M_4)$.

V. EXPERIMENTAL RESULTS

Table III lists the best feature subsets of different cardinality values (1 to 15). Best feature subset implies that they produce maximum classification accuracies (as compared to all other feature subsets of corresponding cardinality) for different evaluation metrics M_i $(1 \leq i \leq 4)$. All the different feature subsets are generated by executing the exhaustive feature subset generation method.

From Table III, it is observed that the feature set $\{2, 4\}$ produces best values for metric M_1 (average F score) and metric M_2 (average WSS with $\alpha = 0.5$) as 0.8 and 0.85, respectively. Addition of extra feature does not significantly increase these scores. The best value of metric M_3 (maximum F score) is 0.83, generated by the feature set $\{2, 4, 5, 6, 8, 10, 12, 14\}$ with σ 0.9. The Feature set $\{2, 4\}$, with σ 0.6, or the feature set $\{2, 3, 4, 6\}$ with σ 0.6 produces output F score of 0.81 and 0.82, respectively. Similarly, the best value of metric M_4 (maximum WSS with $\alpha = 0.5$) is 0.87, generated by the feature set

TABLE IV
PERFORMANCE BY BACKWARD FEATURE REDUCTION METHOD, FOR OUTPUT METRICS M_3 AND M_4

Max F score (M_3)	{Feature set having max F score (F_3)}	σ value	Precision	Sensitivity	Max WSS (M_4)	{Feature set having max WSS (F_4)}	σ value	Sensitivity	Specificity
0.82	{2, 4, 6, 12, 14}	0.7	0.85	0.79	0.86	{2, 4, 10}	0.9	0.83	0.88

TABLE V
PERFORMANCE OBTAINED BY PCA

Number of Principal Components (PC)	Average F score (M_1) (covering all σ cases)	Average WSS (M_2) (covering all σ cases)	Max F score (M_3)	Corresponding σ value	Max WSS (M_4)	Corresponding σ value
1	0.7	0.77	0.72	0.9	0.78	0.9
2	0.53	0.67	0.71	3	0.77	3
3	0.38	0.6	0.48	0.4	0.64	0.4

TABLE VI
COMPARISON OF CLASSIFICATION ACCURACY OBTAINED BY OUR APPROACH FOR BURST PATTERN DETECTION COMPARED TO REFERENCE METHODS

Author	Method Description	Sensitivity	Specificity	Precision	WSS	F score
Leistritz et. al. [9]	SEF95 and 10 Hz power based classification	0.6	0.84	0.67	0.72	0.64
Sarkela et. al. and Palmu. et. al. [10], [11]	NLEO based static thresholding	0.29	0.92	0.87	0.61	0.43
Bhattacharyya et. al. [12]	Feature value ratio between current and background segment, with static thresholding	0.67	0.92	0.94	0.8	0.79
Proposed Method	Feature value ratio between current and background segment, with SVM classification	0.81	0.89	0.8	0.85	0.81

$\{2, 4, 5, 6, 9, 10, 11\}$ with σ 0.4. The feature set $\{2, 4\}$, with σ 2, or the feature set $\{2, 4, 10\}$ with σ 0.9 produces output WSS of 0.85 and 0.86, respectively. These observations indicate that the feature set $\{2, 4\}$ produces best or close to the best results in all the cases considered. Thus, the proposed approach of using ratio of feature values produces optimal results for each of the performance metrics. Table IV summarizes the best values for output metrics M_3 (best F score) and M_4 (best WSS with $\alpha = 0.5$) for backward feature subset selection method. It also lists corresponding feature set, used value of σ, sensitivity, specificity and precision values. Both of the feature subsets, corresponding to two different evaluation metrics, contain the features 2 and 4. It indicates the utility of proposed feature ratio values.

Table V summarizes three sets of maximum values for output metrics M_i $(1 \leq i \leq 4)$ which are obtained by applying SVM based classification using one, two and three principal components respectively. From Table V, it is clear that though a single principal component produces best values for all output metrics, the overall performance is less compared to the classifier output using a single feature 2 (the first row of Table III). This further strengthens the utility of the proposed features.

We observe that for different evaluation metrics, various feature subsets of reduced cardinality perform much better compared to using full feature set or PCA technique. Finally, performance comparison based on the proposed methodology with existing burst detection methods is summarized in Table VI. All these methods are evaluated with same test dataset. Since the feature set $\{2, 4\}$ produces best or close to the best results for all the different performance metrics considered, so the results obtained by applying this feature set with $\sigma = 0.6$ is tabulated for comparison. It is clear that the proposed method outperforms all other methods in terms of final WSS and F score values. All the other methods exhibit comparatively lower sensitivity values. Especially, static thresholding based approaches in [8], [9] have very low sensitivity, indicating that the burst detection is too specific. The proposed method has high values for both sensitivity (81%) and specificity (89%), indicating its utility to assist the doctors or technicians for better summarization of EEG burst segments.

VI. APPLICATION TO LONG TIME CONTINUOUS EEG DATA

So far, the current study concentrated on the supervised learning approach using prior labeled short data segments, which are used to identify feature subsets having maximum classification accuracy. To deal with long term continuous (unlabeled) recorded EEG, the proposed burst detection method needs to be extended to perform burst detection during real time recording. In such an environment, the received data from multiple recording channels can be first segmented by applying a sliding window of duration 1 s and displacement of 0.1 s (thus having 90% overlap). An EEG segment under the sliding window interval then can be used for feature extraction.

Feature values computed using the most recent nonburst or normal EEG intervals (under the sliding window) are estimated and mean of these feature values can be used to model the background EEG pattern. A dedicated circular queue for each feature is employed to store the computed feature values. Queue structure removes the least recent element, thus keeping the most recent information. Circular mechanism for insertion of elements prevents the queue from going out of bound. For instance, if the feature values of previous 2.5-s nonburst intervals (with respect to current EEG segment under sliding window interval) are stored for background estimation, then a total of 25 values per feature needs to be stored in the corresponding queue

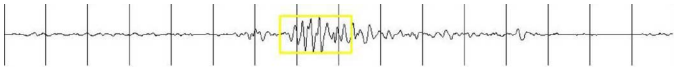

Fig. 11. Snapshot of a single channel segment recorded from channel T4-P4. The segment is of duration 16 s and the automatic detected burst interval (shown in the marked rectangle) lies between 6.5–8.5 s.

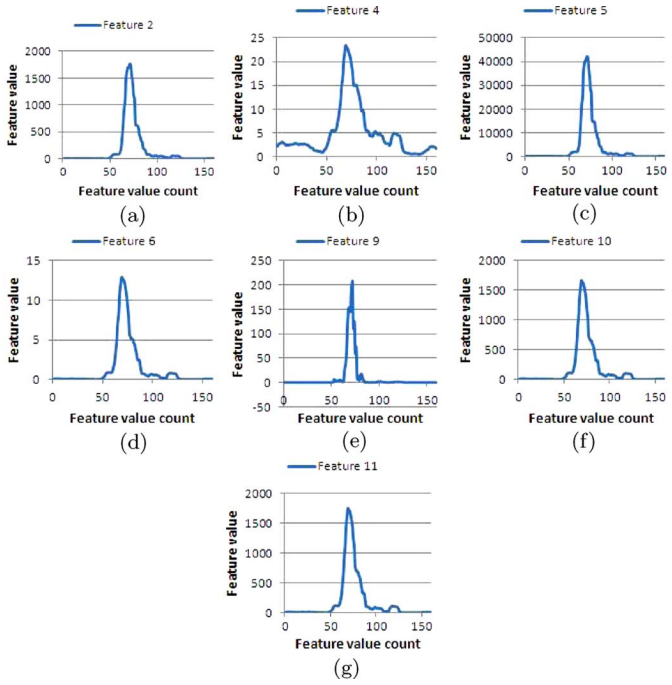

Fig. 12. Variation of individual feature values in the detected burst segment as compared to the background EEG. (a) Feature 2. (b) Feature 4. (c) Feature 5. (d) Feature 6. (e) Feature 9. (f) Feature 10. (g) Feature 11.

(since the sliding window displacement is 0.1 s, so there are total 25 sliding windowed segments over 2.5-s interval). Mean of these 25 feature values stored in the queue is used to estimate the feature value representing latest background EEG pattern. As the sliding window moves, the feature values corresponding to latest encountered nonburst (normal) EEG interval can be used by the classifier while dealing with next segment in the sequence.

Thus, the latest estimate of the background EEG feature values are used to compute the absolute and ratio based feature values to be used as input of the designed classifier. The classifier is already trained with a previous mentioned dataset of labeled short data segments. With the new sets of feature inputs (computed using the current EEG segment under the sliding window), the classifier may recognize the occurrence of a new burst interval—in which case, the start time of current sliding window interval is treated as the burst start time. On the other hand, if the classifier output indicates end of an ongoing burst, then the end time of current window interval is marked as the burst end time.

For example, a single channel burst segment recorded from channel T4-P4 is shown in Fig. 11. This single channel segment of duration 16 s, contains burst interval between 6.5–8.5 s. This burst interval is automatically detected by the proposed method. The feature set producing the best value for metric M_4 (features 2, 4, 5, 6, 9, 10, 11) is used for burst detection. Fig. 12 shows the variation of individual feature values in the detected burst interval as compared to surrounding nonburst interval.

With the help of Fig. 12, we can see that the proposed features exhibit significant fluctuations in case of a burst as compared to the surrounding background EEG. However, the efficiency of this feature based classification method needs to be validated with more continuous multichannel long duration EEG data.

VII. Conclusion and Future Works

In this paper, an automatic burst detection method based on the ratio of feature values, computed in current and background EEG segment is described. Combination of features like ratio of mean nonlinear energy and ratio of mean absolute voltage performs very well. A study on various feature combinations is also presented. As the proposed approach is free from any static thresholding based decision, it can be tested with EEG data from matured children or adults, to make the approach more generally applicable. Of course, it requires more testing with new set of continuous long time recorded data with appropriate feedback from the experts in order to finally judge the overall accuracy of the proposed feature based classification system.

Acknowledgment

The authors would like to thank their colleagues and the doctors of Department of Neonatology, SSKM Hospital, for helping them in data collection, marking, and analysis.

References

[1] J. M. Rennie, C. F. Hagmann, and N. J. Robertson, *Neonatal Cerebral Investigation*, 1st ed. Cambridge, U.K.: Cambridge Univ. Press, 2008.

[2] J. Connell, L. S. Oozeer, L. S. De Vries, L. M. S. Dubowitz, and V. Dubowitz, "Continuous EEG monitoring of neonatal seizures: Diagnostic and prognostic considerations," *Arch. Disease Childhood*, vol. 64, pp. 452–458, 1989.

[3] S. Sanei and J. A. Chambers, *EEG Signal Processing*. Hoboken, NJ: Wiley, 2007.

[4] J. Volpe, *Neurology of the Newborn*, 5th ed. New York: Elsevier Health Sciences, 2008.

[5] J. L. Löfhede, N. Löfgren, M. Thordstein, A. Flisberg, I. Kjellmer, and K. Lindecrantz, "Classification of burst and suppression in the neonatal electroencephalogram," *J. Neural Eng.*, vol. 5, no. 4, pp. 402–410, 2008.

[6] J. L. Löfhede, M. Thordstein, N. Löfgren, A. Flisberg, M. Rosa-Zurera, I. Kjellmer, and K. Lindecrantz, "Automatic classification of background EEG activity in healthy and sick neonates," *J. Neural Eng.*, vol. 7, no. 1, pp. 1–12, 2010.

[7] Y. Wang and R. Agarwal, "Automatic detection of burst suppression," in *IEEE EMBS Int. Conf.*, 2007, pp. 553–556.

[8] K. Palmu, S. Wikström, E. Hippeläinen, G. Boylan, L. Hellström-Westas, and S. Vanhatalo, "Detection of 'EEG bursts' in the early preterm EEG: Visual vs. automated detection," *J. Clin. Neurophysiol.*, vol. 121, no. 7, pp. 1015–1022, 2010.

[9] M. Särkelä, S. Mustola, T. Seppänen, M. Koskinen, P. Lepola, K. Suominen, T. Juvonen, H. Tolvanen-Laakso, and V. Jäntti, "Automatic analysis and monitoring of burst suppression in anesthesia," *J. Clin. Monit. Comput.*, vol. 17, no. 2, pp. 125–34, 2002.

[10] S. Bhattacharyya, J. Mukhopadhyay, A. Majumdar, B. Majumdar, A. Singh, and C. Saha, "Automated burst detection in neonatal EEG," in *Int. Conf. Bio-Inspired Syst. Sig. Process.*, 2011, pp. 15–21.

[11] L. Leistritz, H. Jäger, C. Schelenz, H. Witte, P. Putsche, M. Specht, and K. Reinhart, "New approaches for the detection and analysis of electroencephalographic burst-suppression patterns in patients under sedation," *J. Clin Monit.*, vol. 15, no. 6, pp. 357–367, 1999.

[12] B. R. Greene, S. Faul, W. P. Marnane, G. Lightbody, I. Korotchikova, and G. B. Boylan, "A comparison of quantitative EEG features for neonatal seizure detection," *J. Clin. Neurophysiol.*, vol. 119, no. 6, pp. 1248–1261, 2008.

[13] A. Aarabi, F. Wallois, and R. Grebe, "Automated neonatal seizure detection: A multistage classification system through feature selection based on relevance and redundancy analysis," *J. Clin. Neurophysiol.*, vol. 117, no. 2, pp. 328–340, 2006.
[14] F. S. Bao, D. Y.-C. Lie, and Y. Zhang, "A new approach to automated epileptic diagnosis using EEG and probabilistic neural network," in *IEEE Int. Conf. Tools Artif. Intell.*, 2008, pp. 482–86.
[15] T. Higuchi, "Approach to an irregular time series on the basis of the fractal theory," *Physica D*, vol. 31, no. 2, pp. 277–283, 1988.
[16] P. D. Welch, "The use of fast Fourier transform for the estimation of power spectra: A method based on time averaging over short, modified periodograms," *IEEE Trans. Audio Electroacoust.*, vol. 15, no. 2, pp. 70–73, 1967.
[17] D. N. Joanes and C. A. Gill, "Comparing measures of sample skewness and kurtosis," *J. R. Stat. Soc. (Series D): Statistician*, vol. 47, no. 1, pp. 183–189, 1998.
[18] I. Güler and E. D. Übeyli, "Multiclass support vector machines for EEG-signals classification," *IEEE Trans. Inf. Technol. Biomed.*, vol. 11, no. 2, pp. 117–26, Mar. 2007.
[19] A. V. Oppenheim and R. W. Schafer, *Discrete-Time Signal Processing*. Upper Saddle River, NJ: Prentice-Hall, 1989.
[20] J. F. Kaiser, "Nonrecursive digital filter design using the I0- sinh window function," in *IEEE Symp. Circuits Syst.*, Apr. 1974, pp. 20–23.
[21] C. Cortes and V. Vapnik, "Support vector networks," *Mach. Learn.*, vol. 20, no. 3, pp. 273–297, 1995.
[22] E. Alpaydin, *Introduction to Machine Learning*. Cambridge, NJ: MIT Press, 2008.
[23] C. M. Bishop, *Pattern Recognition and Machine Learning (Information Science and Statistics)*. New York: Springer, 2007.
[24] Z. Zhao, F. Morstatter, S. Sharma, S. Alelyani, A. Anand, and H. Liu, Advancing feature selection research Arizona State Univ., 2010 [Online]. Available: http://featureselection.asu.edu/featureselection_techreport.pdf
[25] P. Somol, J. Novovicova, and P. Pudil, "Efficient feature subset selection and subset size optimization," *Pattern Recognit.*, pp. 75–97, 2010.
[26] VIASYS Healthcare Official [Online]. Available: http://www.viasyshealthcare.com/

Sourya Bhattacharyya received the B.E. degree from Jadavpur University, Kolkata, India, in 2006. Since July 2009, he is pursuing the M.S. degree in the Department of Computer Science and Engineering, IIT Kharagpur, India.

His research area is biomedical signal processing.

Arunava Biswas received the M.B.B.S. degree from Calcutta Medical College, India, in 2009. He is currently pursuing the Ph.D. degree from School of Medical Science and Technology, IIT Kharagpur, India.

His research topic is neonatal health care.

Jayanta Mukherjee (M'90–SM'04) received the Ph.D. degree in electronics and electrical communication engineering from the IIT Kharagpur, in 1990.

He is currently the Head of the Department of Computer Science and Engineering and the School of Information and Technology, IIT Kharagpur. He was a Humboldt Research Fellow at the Technical University of Munich, Germany and also associated with the University of California, Santa Barbara, University of Southern California, and the National University of Singapore. His research interests are in image processing, pattern recognition, computer graphics, multimedia systems and medical informatics.

Arun Kumar Majumdar (M'74–SM'84) received the M.Tech. and Ph.D. degrees in applied physics from the University of Calcutta, India, in 1968 and 1973, respectively, and the Ph.D. degree in electrical engineering from the University of Florida, Gainesville, in 1976.

He is currently Deputy Director of IIT Kharagpur and Professor in the Computer Science and Engineering Department. He was previously associated with ISI Kolkata, Jawaharlal Nehru University, New Delhi, University of Guelph, Canada, and George Mason University, Fairfax, VA. His research interests include data and knowledge based systems, multimedia systems, medical information systems, and design automation.

Bandana Majumdar received the Ph.D. degree from University of Calcutta, India, in 1978.

She has been associated with IIT Delhi, IIT Kharagpur, ISI Kolkata, India and University of Guelph, ON, Canada, in various capacities and is currently a consultant with the Telemedicine research group in the Department of Computer Science and Engineering, IIT Kharagpur. She has published over 50 research papers. Her research interests include VLSI design and medical informatics.

Suchandra Mukherjee received the M.B.B.S. degree from R.G. Kar Medical College, Kolkata, India, in 1993, and the M.D. degree with a specialization in pediatrics from M.K.C.G. Medical College, Berhampur, India, in 2000.

Currently she is an Associate Professor and in-charge of Neuro-developmental Follow up Clinic for High-Risk Neonates of the Department of Neonatology, IPGMER & SSKM Hospital, Kolkata, India.

Dr. Mukherjee is a life member of Indian Academy of Pediatrics and National Neonatology Forum.

Arun Kumar Singh received the M.B.B.S. degree in 1983 from NRS Medical College, Kolkata, India; DCH degree in 1987 from Calcutta Medical College, Kolkata, India; the M.D. degree with a specialization in pediatrics from AIIMS, New Delhi, India, in 1999.

Currently, he is an Associate Professor and Head of the Department of Neonatology at IPGMER & S.S.K.M. Hospital, Kolkata, India. He was previously associated with the Department of Pediatrics AIIMS. He has been involved in a number of collaborative research projects with various government agencies and institutes. He has published many research papers in national and international journals and conferences. He has also co-authored two books.