

A Multi-Channel Fusion Based Newborn Seizure Detection

Malarvili Balakrishnan¹, Paul Colditz², Boualeum Boashash^{2,3}

¹Faculty of Biosciences and Medical Engineering, Universiti Teknologi Malaysia, Skudai, Malaysia

²Royal Brisbane & Women's Hospital Campus, Centre for Clinical Research, University of Queensland, Brisbane, Australia

³College of Engineering-Qatar University, Doha, Qatar

Email: malarvili@biomedical.utm.my

Received 15 April 2014; revised 30 May 2014; accepted 12 June 2014

Copyright © 2014 by authors and Scientific Research Publishing Inc.

This work is licensed under the Creative Commons Attribution International License (CC BY).

<http://creativecommons.org/licenses/by/4.0/>



Open Access

Abstract

We propose and compare two multi-channel fusion schemes to utilize the information extracted from simultaneously recorded multiple newborn electroencephalogram (EEG) channels for seizure detection. The first approach is known as the multi-channel feature fusion. It involves concatenating EEG feature vectors independently obtained from the different EEG channels to form a single feature vector. The second approach, called the multi-channel decision/classifier fusion, is achieved by combining the independent decisions of the different EEG channels to form an overall decision as to the existence of a newborn EEG seizure. The first approach suffers from the large dimensionality problem. In order to overcome this problem, three different dimensionality reduction techniques based on the sum, Fisher's linear discriminant and symmetrical uncertainty (SU) were considered. It was found that feature fusion based on SU technique outperformed the other two techniques. It was also shown that feature fusion, which was developed on the basis that there was inter-dependence between recorded EEG channels, was superior to the independent decision fusion.

Keywords

EEG, Newborn Seizure Detection, Multi-Channel, Feature Fusion, Decision/Classifier Fusion

1. Introduction

Seizures are among the most common and important signs of acute newborn encephalopathy (degenerative disease of the brain). Newborn seizures have been associated with increased rates of long-term chronic neurological morbidity and newborn mortality, with neurological sequelae in as many as 50% to 70% [1]. Newborns with

seizures are reported to be 55 - 70 times more likely to have severe cerebral palsy and 18 times more likely to have epilepsy than those without seizures [1]. Therefore, early detection of seizure in the newborn is very important to prevent long term neurological damage.

Newborn seizures can be identified through a number of clinical and EEG manifestations. The clinical manifestations involve some stereotypical physical behaviors such as sustained eye opening with ocular fixation, repetitive blinking or fluttering of the eyelids, drooling, sucking and other slight facial manifestations [2]. These characteristics are also part of the repertoire of normal behavior in newborns, which makes the identification of the newborn seizure based on clinical signs a difficult task. Therefore, EEG is used as the primary tool by neurophysiologists in monitoring and diagnosing seizures in newborns.

Analysis of EEG by human experts has many limitations [3]. As a consequence, growing attention has been directed toward the development of computerized methods for automatic newborn seizure detection based on the EEG. There are a number of techniques proposed by different authors for detecting newborn EEG seizure. These techniques use information extracted from the time [4]-[6], frequency [7] and time-frequency (TF)/time-scale (TS) [8]-[13] domains. With all these developments, an efficient automatic newborn seizure detector remains elusive.

Faul *et al.* [14] recently performed a comparison of a number of automated newborn seizure detection methods that include those proposed by Liu *et al.* [4], Gotman *et al.* [7] and Celka *et al.* [5] using a common database. This data consisted of 34 one-minute EEG seizure epochs and 43 one-minute non-seizure EEG epochs selected from recordings obtained from 13 newborns. It was found that all 3 methods failed to reliably identify newborn seizures. It must be pointed out here that Gotman's algorithm is the most rigorously tested method and has been made available commercially.

A relative weakness in most of the existing EEG-based newborn seizure detectors is that the algorithms were limited to using single channel EEG. Such algorithms do not utilize information that may present in simultaneously recorded multi-channel EEG. Although Liu's algorithm is based on multiple channels, it only uses a simple decision fusion rule. It declares the presence of a seizure if at least two channels simultaneously report detection. A newborn seizure detector should be able to use the available information provided by the multi-channel EEG. To overcome this limitation, we propose, in this paper, two multi-channel fusion approaches to account for the EEG information spread over the different channels. The first approach is called the multi-channel feature fusion while the one is called the multi-channel decision/classifier fusion.

Feature fusion, also known in the literature as early integration or feature-level combining [15], is a common framework for the fusion of different feature sets as shown in **Figure 1**. Feature fusion combines features derived from the different single-channel EEG into a more global feature set to be used for classification of multiple channel EEG, thus, requires only one classifier.

Decision fusion, referred to in the literature as late integration or classifier combining [15], is illustrated in **Figure 2**. In this approach, the single-channel based decisions are combined to produce an overall decision as to the existence of the seizure in the multichannel EEG.

This paper is organized as follows. Section 2 addresses each step involved in development of both the proposed multi-channel fusion approaches. These steps include data acquisition, preprocessing of the EEG to remove unwanted noise, extracting time, frequency and TF/TS domain features from the multi-channel EEG, selecting non-redundant and discriminative EEG features from the larger extracted set and finally fusing the selected EEG features at two different levels as mentioned above. In section 3, the performances of both proposed fusion approaches are evaluated and discussed. Performance comparison between the proposed newborn multi-channel EEG seizure detection with two of the most widely cited newborn EEG seizure detection algorithms is also included in the section. Finally, a conclusion of our work is presented. Method

Overview of both the proposed newborn multi-channel EEG seizure detection algorithm is summarized in **Figure 3**. Both proposed EEG multi-channel fusion approaches consist of a sequence of processing steps, namely; preprocessing of the EEG signal, extraction of the EEG features and selection of optimal feature subset from the larger set. This process is done for each channel separately. This is followed by a fusion scheme where the extracted information from the different channels is combined at either feature level or decision level. Each of these steps is detailed in following subsections.

1.1. Data Acquisition and Preprocessing of EEG

The newborn EEG data used in this study was recorded at the Royal Brisbane and Women's Hospital in Brisbane, Australia. Each newborn recording consists of 20 channels of EEG recorded using Medelec Profile System

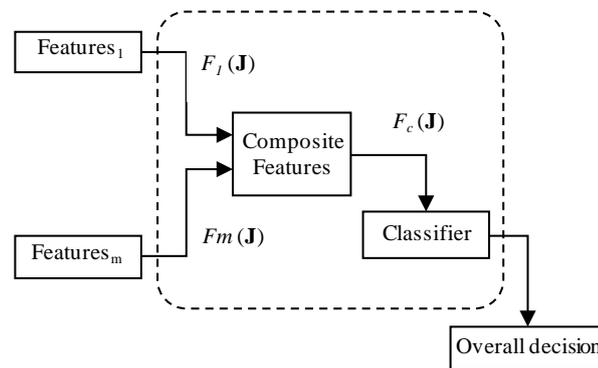


Figure 1. A common framework for feature fusion using feature sets from m different sources (in this case $m = 2$).

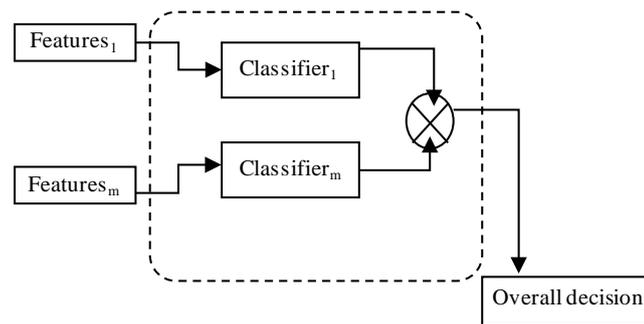


Figure 2. A common framework for classifier/decision fusion using feature sets from m different sources (in this case $m = 2$).

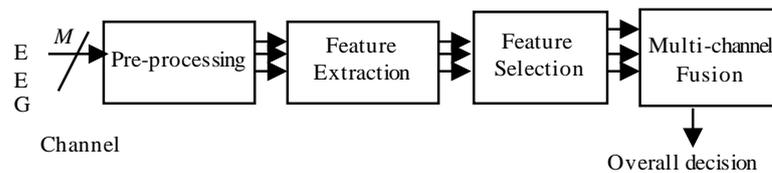


Figure 3. Overview of proposed multi-channel newborn EEG seizure detection.

(Oxford Instruments, UK). The EEG channels were obtained from 14 electrodes, placed according to the international 10 - 20 standard, using longitudinal bipolar montage. The EEG seizure and non-seizure segments were identified and annotated by a pediatric neurologist. The raw EEG data was filtered with a band pass filter whose cutoff frequencies is 0.5 Hz and 70 Hz and then sampled at rate of 256 Hz. Since the majority of spectral energy (*i.e.* > 95% of spectral energy) in the newborn EEG is concentrated in the delta and theta frequency ranges (0.4 - 8 Hz) [16], the EEG from each channel was filtered with low pass filter with a cutoff frequency of 8 Hz and re-sample at a rate of 20 Hz. In this study, we analyzed 112 seizure and 69 non-seizure epochs of 12 seconds each from 8 newborns.

1.2. Feature Extraction of EEG

Since the aim of the paper is to show the added value of multi-channel EEG based seizure detections over the single channel EEG based seizure detection, we have used a wide range of features that were derived in previous studies on EEG. In this approach, a set of d -dimensional feature vectors are extracted from the M different EEG channels and fed to supervised statistical classifiers to discriminate between EEG seizure and background (non-EEG seizure). To do so, a number of parameters have been identified as potential features. These include parameters extracted from time, frequency, and time-frequency and time-scale representations. The description of these features is as follows:

1.2.1. Time Domain Features

In the time domain, a number of features have been identified and extracted from the 12-second EEG epochs. These include the mean, standard deviation, skewness, kurtosis, coefficient of variation, RMS and zero-crossings [17]. The Hjorth parameters, which describe the signal characteristics in terms of activity, mobility, and complexity, were also computed [18]. The total nonlinear score, proposed by Liu *et al.* [4] as a measure of periodicity, was also used as a feature.

1.2.2. Frequency Domain Features

The frequency domain representation is achieved using Fast Fourier Transform with Hamming window of length 256 points using Equation (1) where $x(n)$ is the EEG signal and N is the length of 12 second EEG epoch. The frequency domain features adopted here are similar to the one proposed by Gotman *et al.* in [7]. These include the peak and maximum frequency, the bandwidth, and the spectral power of the dominant spectral peak of the 12-second EEG epochs.

$$X(k) = \sum_{n=0}^{N-1} x(n) e^{-\frac{j2\pi nk}{N}} \quad (1)$$

1.2.3. Time-Scale Features

The wavelet transform of the signal $x(t)$ is defined in terms of the mother wavelet $\psi(t)$ as Equation (2) [13];

$$W_x(a, b) = \frac{1}{\sqrt{a}} \int_{-\infty}^{\infty} x(t) \psi\left(\frac{t-b}{a}\right) dt \quad (2)$$

where a and b are respectively the scaling and shifting parameters. Discrete wavelet transform has been used as a basis for EEG seizure detection [13] for EEG seizure detection in newborn. The Daubechies 4-tab wavelet has been used for decomposing the EEG segments into 9 scales. We have extracted the optimal features identified in [13]. These include variance of $cd3$, $cd6$, $cd7$, $cd4$, $cd2$, $ca9$, $d5$ and mean of $d9$, $d6$, $d5$; where cdi and di refer to the detail coefficient and detail component of i^{th} scale respectively.

1.2.4. Time-Frequency Features

The time-frequency distribution (TFD) of a signal, $s(t)$, represents its energy density in the joint time-frequency domain. The general expression for the quadratic shift-invariant TFD is given by Equation (3) [19]:

$$\rho_z(t, f) = \iiint e^{j2\pi(vt-vu-f\tau)} g(v, \tau) z(t+\tau/2) z^*(t-\tau/2) dv d\tau du \quad (3)$$

where $z(t)$ is the analytic associate of the real signal $s(t)$ [19]. The integrations are taken from $-\infty$ to ∞ unless otherwise stated. The function $g(v, \tau)$, defined in the Doppler-lag domain, (v, τ) , is known as the Doppler-lag kernel which determines the characteristics of the TFD. There is a large number of TFDs that can be used to represent a signal in the TF domain by varying the $g(v, \tau)$ function. The choice of a suitable one depends on both the characteristic of the signal under analysis and on the application. The TF representation in this paper is achieved using the Modified B distribution (MBD) [19]. The MBD has been chosen to represent the EEG in the TF domain as it was found in previous studies to realize the best compromise in terms of cross-term reduction and TF resolution in [19]. The function $g(v, \tau)$ for MBD is given by Equation (4)

$$g(v, \tau) = \frac{|\Gamma(\beta + j\pi v)|^2}{\Gamma^2(\beta)} \quad (4)$$

where $\Gamma(\cdot)$ is the gamma function and β is a positive number that controls the trade-off between components' resolution and cross-terms suppression [19]. The optimal value of β for this study is found to be around 0.02.

Figure 4 and **Figure 5** shows the TF representation of EEG background and EEG seizure respectively. The TF features comprise the total TF energy, the largest and smallest singular values obtained from singular value decomposition of the MBD [10] and the number of TF components with a predefined duration. In our context, a true TF component is a component whose duration is larger than 9.5 seconds [19]. This last feature is extracted using the multicomponent IF estimation procedure proposed in [19]. In this study, it was observed that the EEG seizure epochs have at least one noticeable component whereas the EEG non-seizure epochs have no significant (noticeable) TF components (refer to **Figure 4** and **Figure 5**). This finding is consistent with the results in [8].

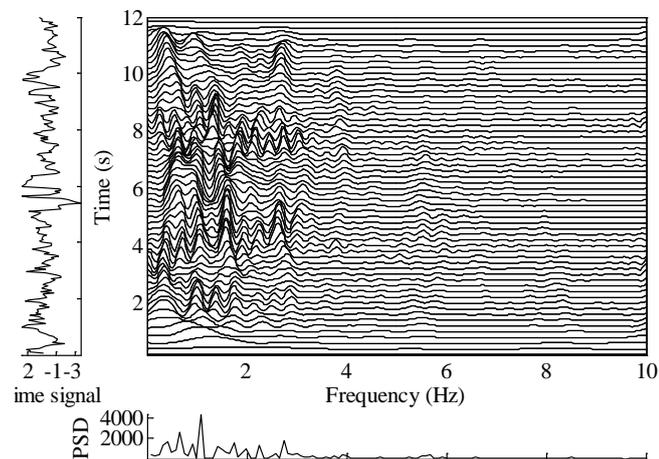


Figure 4. TF representation (using MBD) of EEG non-seizure.

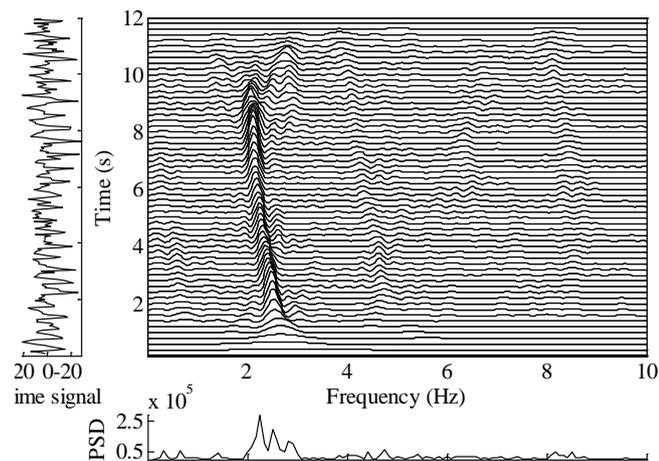


Figure 5. TF representation (using MBD) of EEG seizure.

All the extracted features were normalized to have zero mean and unit standard deviation for ease of comparison. They were further processed in order to identify the features with high discriminatory power. This last point is addressed in the next section.

1.3. Feature Selection

Feature selection is a process of selecting an optimal subset from the original set of extracted features that is both relevant and non-redundant [20]. The extracted features may contain a large number of redundant and/or irrelevant feature subsets that may significantly affect the classification performance [20]. Existing feature selection approaches generally belong to the following two categories: wrappers and filters [20]-[22].

Wrappers are classifier-dependent as they include the classifier as part of their performance evaluation while filters are classifier-independent. For a pre-selected classifier, wrappers tend to give superior performance as they select features better suited to that classifier but at a high computational cost. Filters, on the other hand, are computationally efficient but, since they are classifier-independent, they tend to select sub-optimal features. To exploit the merits of the two methods, we have used the filter-wrapper-based feature selection method developed earlier in [22] to select an optimal subset of EEG features from the larger set extracted in section 2.2. The feature selection process is considered successful if the dimensionality of the data is reduced while the accuracy of the classification is either improved or remains unaffected.

The filter-wrapper-based feature selection method in [22] is a two-phase process used to reduce the computation load associated with the conventional wrappers by reducing the dimension of the search space and, there-

fore, speeding up the convergence process. The first phase, which acts as pre-processing phase, involves a filter [21] that uses discriminant and redundancy analysis to select a feature subset with high discriminatory power between the two newborn EEG classes. As the result of this, a set of relevant features, \mathbf{f} , with minimum redundancy and maximum class discriminability is obtained. In the second phase, the feature subsets, \mathbf{f} , is presented as an input to a wrapper. The wrapper optimizes a measure of performance, such as the probability of classification error, to obtain the optimal feature set. The results are presented in section 3.

1.4. Multi-Channel EEG Fusion Configuration

Two fusion methods for combining selected features from 20 channels of EEG data are outlined below.

1.4.1. Multi-Channel Feature Fusion

The proposed multi-channel EEG feature fusion configuration for newborn seizure detection is shown in Figure 6 [23]. In this fusion scheme, the feature fusion is achieved by concatenating the feature vectors, extracted from each epoch of the M channels into a single composite feature matrix, \mathbf{J}_c , such that Equation (5)

$$\mathbf{J}_c = [\mathbf{J}_1 | \mathbf{J}_2, \dots, | \mathbf{J}_M] \tag{5}$$

where \mathbf{J}_i is the d -dimensional feature vector associated with the i^{th} EEG channels and containing the d selected features using the feature selection method discussed earlier. The resulting M -channel composite feature matrix, \mathbf{J}_c , with a dimension of $d \times M$ is given by Equation (6)

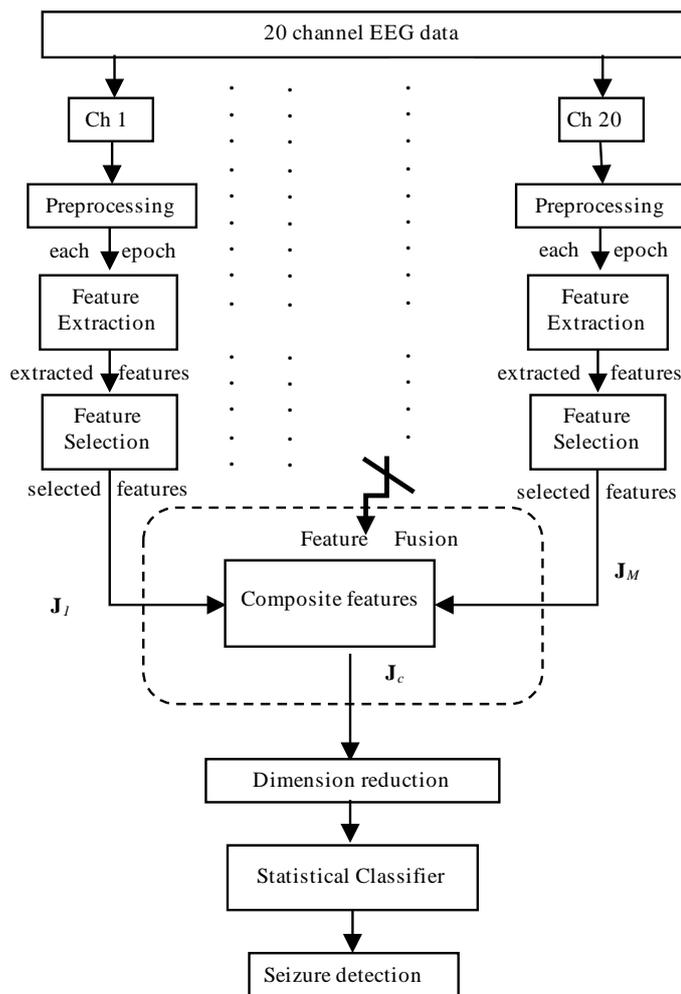


Figure 6. Multi-channel feature fusion for newborn seizure detection.

$$\mathbf{J}_c = \begin{bmatrix} J_{11} & J_{12} & \cdots & J_{1M} \\ J_{21} & J_{22} & \cdots & J_{2M} \\ \vdots & \vdots & & \vdots \\ J_{d1} & J_{d2} & \cdots & J_{dM} \end{bmatrix} \quad (6)$$

where J_{ij} is the j^{th} components of the vector \mathbf{J}_i .

The advantage of concatenation fusion is that the EEG features from all channels are contained in the composite feature matrix. Practical applications of combining multiple feature sets have shown promising results. One such application is the audio-visual feature fusion for speech recognition [24]. The integration of audio and visual features for speech recognition has been reported to enhance the recognition system. In another research, feature-level combining of voice and facial features for person identification have been documented to increase the recognition accuracy compared to audio only recognition [25]. Thus, we have adopted the feature vector concatenation for feature fusion in our work.

The drawback of feature concatenation is that in cases involving a large number of channels, as in our case, the dimension of the composite feature matrix becomes large. As mentioned before, using excessive number of features in classification process tends to reduce the efficiency of the classification process. Therefore, a further dimension reduction technique¹ is employed in order to reduce the dimension of the composite feature matrix, \mathbf{J}_c .

There a number of dimensionality reduction algorithm available in the literature for decreasing the dimension of the combined feature matrix in cases which involves multi-channel data. A review of feature dimension reduction methods and some general results are given in [26]-[28]. In this work, we have considered 3 different techniques which we refer to as sum-based, Fisher's linear discriminant-based and symmetrical uncertainty (SU)-based dimensionality reduction techniques. The first approach is usually chosen for its computational simplicity, thus, has been used widely in feature fusion [29].

In the sum-based approach, the elements in each row of the \mathbf{J}_c are summed such that Equation (7)

$$\Psi_i = \sum_{j=1}^M J_{ij} \quad (7)$$

where $i \in \{1, \dots, d\}$. That is, each element (feature) in the feature vector, \mathbf{J}_i , is summed across all the channels to form a reduced composite feature vector which has the same dimension as the individual feature vectors, \mathbf{J}_i . This is allowed as the features in each row of \mathbf{J}_c are of the same type. This idea was proven effective when applied to evoked potentials for classification of different brain activity [30]. The composite feature vector of reduced dimension, \mathbf{J}^S , can be written as Equation (8)

$$\mathbf{J}^S = \begin{bmatrix} J_1^S \\ \vdots \\ J_d^S \end{bmatrix}. \quad (8)$$

Another way of reducing the dimension is through a linear combination of the elements in each rows of the matrix \mathbf{J}_c using the Fisher's linear discriminant function [27]. Each row of the \mathbf{J}_c is considered as an independent (individual) vector to which the Fisher's linear discriminant function was applied. As a consequence, the dimension of each row in the \mathbf{J}_c is reduced from M to 1. Therefore, the new composite feature vector, \mathbf{J}^F , is represented as Equation (9)

$$\mathbf{J}^F = \begin{bmatrix} J_1^F \\ \vdots \\ J_d^F \end{bmatrix} \quad (9)$$

where J_i^F represents the value obtained from linear combinations of the elements in each rows of the \mathbf{J}_c .

In SU-based dimension reduction technique, the correlation is used as a basis to systematically decrease the dimension of \mathbf{J}_c by averaging the most correlated elements in each rows of the \mathbf{J}_c . In real world applications,

¹This dimension reduction technique is used to decrease the dimension of the combined feature vector obtained from multi-channel EEG while the feature selection method in section 3.3 is used to discard the irrelevant and redundant features from larger set extracted.

features are not always linearly correlated. Thus, we use the symmetrical uncertainty (SU) instead of the linear correlation. The SU is a nonlinear correlation measure based on the information-theoretical concept of entropy and its values range between $[0, 1]$ [28]. As in the two other methods, each row of the \mathbf{J}_c is considered as an independent (individual) vector and the correlation operation was employed to each row. A threshold SU value, $th = 0.75$, is used to identify the correlated elements in each row. These elements are then summed. As a result, the dimension of each row in the \mathbf{J}_c is decreased to one iteratively by summing the most correlated elements. The composite feature vector of reduced dimension, \mathbf{J}^{SU} , is then given by Equation (10)

$$\mathbf{J}^{SU} = \begin{bmatrix} J_1^{SU} \\ \vdots \\ J_d^{SU} \end{bmatrix} \tag{10}$$

The dimension of the reduced composite feature vector, \mathbf{J}^{SU} , is the same as the dimension of the feature vectors, \mathbf{J}_i , from each channel.

The reduced composite feature vector is then fed to the statistical classifiers investigated in the process of newborn seizure detection. The performance of the proposed multi-channel newborn EEG seizure detection using the three different dimension reduction methods is presented in Section 3.

1.4.2. Multi-Channel Decision Fusion

In decision-based fusion configuration, M separate statistical classifiers were utilized; each one used a d -dimensional feature vector extracted from the corresponding channel at for EEG epoch. The decisions from the different classifiers were then combined using the different combination rules described in [31] to achieve an overall classification of the EEG epochs. The proposed newborn multi-channel EEG decision fusion scheme is illustrated in Figure 7 [23]. The investigated combining rules were [31]: the mean, the max, the min, the product, the sum and the majority rules.

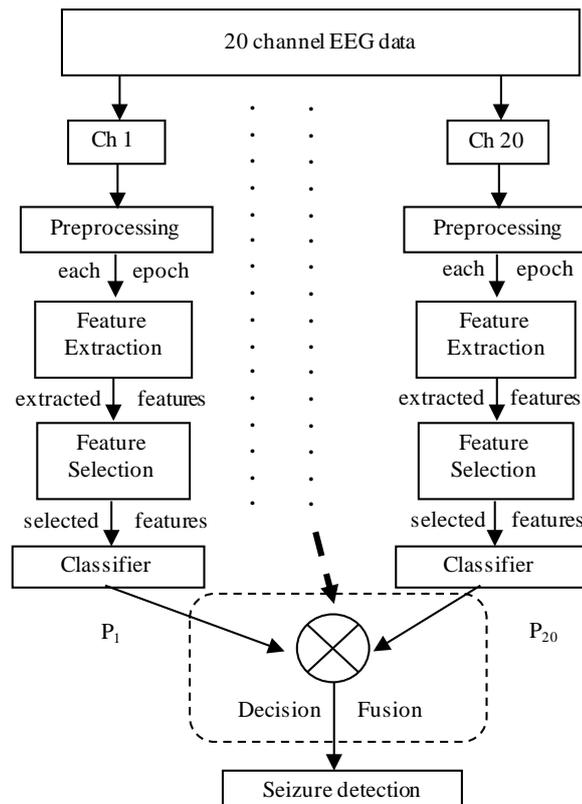


Figure 7. Multi-channel decision fusion for newborn seizure detection.

1.4.3. Supervised Statistical Classification

In this work, supervised statistical classifiers were used to classify the multi-channel EEG epochs into seizure class and background class. In this approach, a set of d -dimensional feature vectors are used as input to a statistical supervised classifier. As in practice no single classifier is optimal in all circumstances, different statistical classifiers are tested in order to find the most suitable one for the application at hand. Therefore, the following three classifiers were investigated in this study: linear classifier, quadratic classifier and k -nearest neighbor (k -NN) where $k = 1, 3$ and 5 .

The statistical classifiers has been chosen to investigate this work because among the various frameworks in which pattern recognition has been traditionally formulated, the statistical approach has been most intensively studied and used in practice [32]. Statistical classifiers fall into two categories; parametric and non-parametric [32]. The linear and quadratic classifiers are of the parametric type. In this type, the classification rules are based on models of the probability density function of the data. Both linear and quadratic classifiers are based on the assumption that classes have multivariate Gaussian distributions. The k -NN is a nonparametric classification procedure and hence no assumption of the form of the underlying probability density functions is required.

The statistical classifiers are evaluated using a leave-one-out cross-validation [33]. This method is also known as n -fold cross-validation, where n stands for the number of subsets or folds. The process is performed by splitting the data set D into n mutually exclusive subsets D_1, D_2, \dots, D_n . The classifier is trained and tested n times; each time $k = 1, \dots, n$, it is trained on $D \setminus D_k$ and tested on D_k . As the leave-one-out variant is used when the size of the data is small, it is adopted here [33]. All the classification results are presented in section 3.

1.4.4. Classification Performance Measure

In this paper, the classification performance is expressed in terms of sensitivity and specificity. The specificity is defined as the ratio of non-seizure epochs correctly recognized by the classifier (the non-seizure detection rate) while the sensitivity is the ratio of seizure events correctly recognized by the classifier (the seizure detection rate) [34]. The sensitivity, Sn , and specificity, Sp , are mathematically defined as Equation (11):

$$Sn = \frac{TP}{TP + FN} \quad Sp = \frac{TN}{TN + FP} \quad (11)$$

where TP , TN , FN , and FP respectively represent the number of true positive, true negative, false negative and false positive. The error, or misclassification, rate is computed using the following expression Equation (12) [35]:

$$Er = 1 - \frac{(Sn + Sp)}{2} \quad (12)$$

2. Performance Evaluation

In this section, we present and discuss the performance of proposed methods above.

2.1. Performance of Feature Selection Method

Since a large number of features were extracted from each of the 12-second EEG epochs and in order to achieve efficient automatic newborn EEG seizure classification, a two step filter-wrapper-based feature selection method using discriminant and redundancy analysis was used to reduce the dimensionality of the feature vectors. The results of this selection for the different statistical classifiers considered are summarized in **Table 1**.

Table 1 shows the error rates for the different classifiers using both the full feature set and the reduced feature subsets. As mentioned above, the filter-wrapper-based feature selection method is a two step process which includes of a filter [21] that uses discriminant and redundancy analysis to select a feature subset with high discriminatory power. The output of this filter stage is 5 features, which are; numbers of TF components, (*TF features*), peak frequency and bandwidth of the peak frequency (*frequency-domain feature*) and total score from Liu's technique (*time-domain feature*).

The error rates associated with the reduced feature subsets are noticeably smaller than those related to the full feature set. This shows that the classification accuracy is increased by discarding non-discriminating and redundant features. For example, the linear classifier has an error rate of 23.9% when using the selected features compared to the 38.9% when using the full feature set. The 3-NN classifier achieved an error rate of 37.4% down from 52.8%.

Table 1. Comparison of classifiers accuracy when using the original feature set and the optimal feature subset obtained from the filter and filter-wrapper based feature selection for the different statistical classifiers.

Classifier	Full Set	Filter (5 features)	Filter-wrapper	
	Error rate	Error rate	No. Features	Error rate
Linear	0.389	0.239	5	0.239
Quadratic	0.429	0.370	4	0.357
1-Nearest Neighbor	0.451	0.333	5	0.333
3-Nearest Neighbor	0.528	0.390	4	0.374
5-Nearest Neighbor	0.474	0.302	5	0.302

We can also see that, for the filter-wrapper case, the dimension of the optimal subset (the one achieving the lowest error rate) differs for different classifiers. The classifier which gives the lowest error rate among all the classifiers tested is considered as the optimal classifier for the application at hand. In this case, the linear classifier achieved its lowest error rate of 23.9% when the number of dominant feature is 5. As a consequence of these results, the selected 5 features and the linear classifier were used in both the multi-channel EEG fusion approaches.

2.2. Performance of Proposed Fusion Approach

We proposed two methods for fusing information extracted from multiple newborn EEG channels, namely, feature fusion and decision fusion. As mentioned before, the multi-channel feature fusion is achieved by concatenating EEG feature vectors from multiple channels. The dimension of the concatenated feature matrix was reduced using three different dimension reduction techniques as explained in section 2.4. The multi-channel decision fusion, on the other hand, was accomplished by combining the independent decision of each EEG channel. We first compared the performance of the newborn multi-channel EEG feature fusion for seizure detection using the three different dimension reduction methods. The results are summarized in **Table 2**. We then compared the performances of the two multi-channel fusion techniques in order to determine the overall optimal newborn multi-channel-based EEG classifier for seizure detection.

From **Table 2**, it is evident that newborn multi-channel EEG feature fusion using the *SU*-based dimension reduction technique outperforms the other two techniques with 80.9% sensitivity and 86.5% specificity. Newborn multi-channel EEG feature fusion using the Fisher's linear discriminant-based dimension reduction method gave the lowest sensitivity and specificity of 65.3% and 69.8% respectively. Unlike the *SU*-based dimension reduction technique, the other two methods are based on the assumption that there is an equal manifestation of seizure across all EEG channels. That is, the stereotypical EEG seizure morphology was assumed to be present on all the EEG channels simultaneously. This assumption is obviously not true, especially in the case of the newborn where the generalized seizure is unlikely, and therefore negatively impacts on the performance of the corresponding classifiers. Another problem with linear discriminant-based dimension reduction is its dependence on the invalid assumption of linear correlation between features selected from different channels.

As for the multi-channel decision fusion, the independent decision from each of the classifiers were fused using the combination rules described in [30] for the overall classification of EEG. As mentioned before, the combining rules tested were the mean, the max, the min, the product, the sum and the majority rules. Among these, the sum rule was found to produce the best classification performance with a sensitivity of 72.3% and a specificity of 70.5% respectively. In the sum combination rule, the output probability of each classifier is summed for the overall decision. As noted by Kittler *et al.* in [31], the superior performance of the sum-rule could be due to the fact that it is not significantly affected by the probability of errors.

The results shown in **Table 2** indicate that the multichannel newborn EEG seizure detection based on feature fusion outperforms the decision fusion-based method. In classifier fusion strategy, the independent decision of each classifier was combined to derive a final decision. Thus, it was developed based on assumption that all EEG channels are statistically independent from one and another; and the simultaneously recorded nature of EEG channels was unexploited. As such, the performance of the multi-channel decision fusion approach is limited. As we mentioned before, the feature fusion strategy was developed based on the assumption that there is

Table 2. Performances of the fusion approach for automatic newborn multi-channel EEG seizure detection.

Performance Measure	Feature Fusion			Decision Fusion-sum rule
	Dimension Reduction Technique			
	Sum	Fisher's	SU	
Sn (%)	67.5	65.3	80.9	72.3
Sp (%)	70.2	69.8	86.5	70.5

inter-dependence between some of the recorded EEG channels during seizure and the simultaneously recorded nature of EEG channels was exploited. Therefore, our results indicate that the best performance for combining information extracted from multi-channel EEG can be achieved by treating the seizure manifested EEG channels as interrelated to one another.

2.3. Comparison with Existing Newborn EEG Seizure Detection Techniques

A similar study on combining information extracted from newborn multi-channel EEG was conducted by Greene *et al.* in [36]. Two methods were proposed for combining the EEG information from multiple channels. The methods were referred to as early integration (EI) and late integration (LI). Both methods were considered in a patient-specific and patient-independent framework for newborn seizure detection. It was found that the EI scheme outperformed the LI scheme. This result is consistent with our finding. However, their feature level fusion achieved 65.36% sensitivity and 78.27% specificity in the patient-specific case and 69.3% sensitivity and 63.57% specificity in the patient-independent case. As such, there is significant difference between our results and the results found in [36]. From our investigation, we found two factors are responsible for these differences.

Firstly, our method uses a number of features extracted from time domain, frequency domain and TF domain. Thus, the proposed method made comprehensive use of the available stationary and non-stationary information while the algorithm in [36] only uses stationary features. This means the non-stationary property of the EEG was overlooked. Secondly, our proposed feature fusion used a dimension reduction technique to reduce the dimension of the concatenated feature matrix while the method in [36] used a very large feature vector for classification process which is a potential source for poor performance as mentioned previously.

To further evaluate the performance of our proposed feature fusion based multichannel newborn EEG classifier, we compared it with two widely cited newborn EEG seizure detection algorithms. The two algorithms considered are those proposed by Liu *et al.* [4] and Gotman *et al.* [6]. These two methods use stationary features extracted from time (autocorrelation function) and frequency (spectrum) domains respectively in an attempt to find periodicity in the EEG which is thought to characterise newborn EEG seizure. In these algorithms, the thresholds were modified slightly from their original values in order to obtain improved detection rates.

Table 3 shows the results of the different detection methods. With a sensitivity of 60.6% and a specificity of 61.5%, Liu's algorithm had the lowest performance. Gotman's algorithm performed better with 72.1% sensitivity and 69.2% specificity. The results clearly demonstrate the superiority of our proposed newborn multi-channel EEG seizure detection algorithm. The good results can be attributed to a number of factors. Firstly, our method uses a number of discriminative and non-redundant EEG features to improve classification accuracy. Moreover, by selecting TF features, our method accounted for nonstationarity, which is an important property of newborn EEG. The non-stationarity assumption of the newborn EEG is a major factor which restricts the performance of the stationary techniques such as those of Liu and Gotman. Another critical factor affecting the performance of Gotman's technique is that it is based on a single channel EEG and as such does not fully utilize information spatially spread through different channels. Although Liu's algorithm is based on multi-channel, it only involves a simple decision fusion rule.

3. Conclusion

Two classification approaches based on multi-channel feature fusion and multi-channel decision fusion have been introduced in order to exploit EEG information from simultaneously recorded multiple EEG channels to detect newborn seizure. The multi-channel feature fusion is achieved by concatenating EEG feature vectors extracted from the different EEG channels. Since this process tends to increase the dimensionality of the combined feature vector (or matrix), a number of dimension reduction techniques have been investigated for this purpose.

Table 3. Newborn EEG seizure detection algorithm performance comparison.

Algorithm	Performance Measure	
	Sn (%)	Sp (%)
Proposed EEG multi-channel feature fusion	80.9	86.5
Liu et al. [4]	60.6	61.5
Gotman et al. [6]	72.1	69.2

Among these techniques, the one based on symmetrical uncertainty was found to have the best performance. The second proposed classification technique was based multi-channel decision fusion. It was accomplished by combining the independent decisions of the single channel based classifiers. Our results show that the feature fusion based EEG classification performed better than the decision fusion one. The feature fusion classification technique was also shown to outperform some of the widely cited newborn EEG seizure detections techniques.

References

- [1] Aylward, G.P. (1989) Outcome of the High-Risk Infant: Fact versus Fiction. In: Gottlieb, M.I. and Williams, J.E., Eds., *Developmental-Behavioral Disorders: Selected Topics*, Chapter 2, Volume 2, Plenum Publishing Corporation, New York, 18-25.
- [2] Volpe, J.J. (1989) Neonatal Seizures: Current Concepts and Revised Classification. *Pediatrics*, **84**, 422-428.
- [3] Rangayyan, R.M. (2001) *Biomedical Signal Analysis: A Case-Study Approach*. John Wiley & Sons Inc., Hoboken.
- [4] Liu, A., Hahn, J.S., Heldt, G.P. and Coen, R.W. (1992) Detection of Neonatal Seizures through Computerized EEG Analysis. *Electroencephalography and Clinical Neurophysiology*, **82**, 363-369. [http://dx.doi.org/10.1016/0013-4694\(92\)90179-1](http://dx.doi.org/10.1016/0013-4694(92)90179-1)
- [5] Celka, P. and Colditz, P. (2002) A Computer-Aided Detection of EEG Seizures in Infants: A Singular-Spectrum Approach and Performance Comparison. *IEEE Transactions on Biomedical Engineering*, **49**, 455-462. <http://dx.doi.org/10.1109/10.995684>
- [6] Navakatikyan, M.A., Colditz, P.B., Burke, C.J., Inder, T.E., Richmond, J. and Williams, C.E. (2006) Seizure Detection Algorithm for Neonates Based on Wave-Sequence Analysis. *Clinical Neurophysiology*, **117**, 1190-1203. <http://dx.doi.org/10.1016/j.clinph.2006.02.016>
- [7] Gotman, J., Flanagan, D., Zhang, J. and Rosenblatt, B. (1997) Automatic Seizure Detection in the Newborn: Methods and Initial Evaluation. *Electroencephalography and Clinical Neurophysiology*, **103**, 356-362. [http://dx.doi.org/10.1016/s0013-4694\(97\)00003-9](http://dx.doi.org/10.1016/s0013-4694(97)00003-9)
- [8] Boashash, B. and Mesbah, M. (2001) A Time-Frequency Approach for Newborn Seizure Detection. *IEEE Engineering in Medicine and Biology Magazine*, **20**, 54-64. <http://dx.doi.org/10.1109/51.956820>
- [9] Boashash, B., Mesbah, M. and Colditz, P. (2001) Newborn EEG Seizure Pattern Characterization Using Time-Frequency Analysis. *Proc. of IEEE Int. Conf. on Acoustics, Speech and Signal Processing*, **2**, 1041-1044.
- [10] Hassanpour, H., Mesbah, M. and Boashash, B. (2004) Time-Frequency Feature Extraction of Newborn EEG Seizure Using SVD-Based Techniques. *EURASIP Journal on Applied Signal Processing* **16**, 2544-2554. <http://dx.doi.org/10.1155/s1110865704406167>
- [11] Hassanpour, H. and Mesbah, M. (2003) Neonatal EEG Seizure Detection Using Spike Signatures in the Time-Frequency Domain. *International Symposium on Signal Processing and Its Application*, **2**, 41-44. <http://dx.doi.org/10.1109/isspa.2003.1224810>
- [12] Hassanpour, H., Mesbah, M. and Boashash, B. (2004) Time-Frequency Based Newborn EEG Seizure Detection Using Low and High Frequency Signatures. *Physiological Measurement*, **25**, 935-944. <http://dx.doi.org/10.1088/0967-3334/25/4/012>
- [13] Zarjam, P., Mesbah, M. and Boashash, B. (2003) Detection of Newborns EEG Seizure Using Optimal Features Based on Discrete Wavelet Transform. *IEEE International Conference on Acoustics, Speech, and Signal Processing*, **2**, 265-268.
- [14] Faul, S., Boylan, G., Connolly, S., Liam, M. and Gordon, L. (2005) An Evaluation of Automated Neonatal Seizure Detection Methods. *Clinical Neurophysiology*, **116**, 1533-1541. <http://dx.doi.org/10.1016/j.clinph.2005.03.006>
- [15] Wald, L. (1999) Some Terms of References in Data Fusion. *IEEE Transactions on Geoscience and Remote Sensing*, **37**, 1190-1193. <http://dx.doi.org/10.1109/36.763269>
- [16] Scher, M.S., Sun, M., Steppe, D.A., Guthrie, R.D. and Scwabassi, R.J. (1994) Comparison of EEG Spectral and Corre-

- lation Measures Between Healthy Term and Preterm Infants. *Pediatric Neurology*, **10**, 104-108. [http://dx.doi.org/10.1016/0887-8994\(94\)90041-8](http://dx.doi.org/10.1016/0887-8994(94)90041-8)
- [17] Hjorth, B. (1973) The Physical Significance of Time Domain Descriptors in EEG Analysis. *Electroencephalography and Clinical Neurophysiology*, **34**, 321-325. [http://dx.doi.org/10.1016/0013-4694\(73\)90260-5](http://dx.doi.org/10.1016/0013-4694(73)90260-5)
- [18] D'Alessandro, M., Esteller, R., Vachtsevanos, G., Hinson, A., Echauz, J. and Litt, B. (2003) Epileptic Seizure Prediction Using Hybrid Feature Selection over Multiple Intracranial EEG Electrode Contacts: A Report of Four Patients. *IEEE Transactions on Biomedical Engineering*, **50**, 603-615. <http://dx.doi.org/10.1109/tbme.2003.810706>
- [19] Rankine, L., Mesbah, M. and Boashash, B. (2007) IF Estimation for Multicomponent Signals Using Image Processing Techniques in the Time-Frequency Domain. *Signal Processing*, **87**, 1234-1250. <http://dx.doi.org/10.1016/j.sigpro.2006.10.013>
- [20] Dash, M. and Liu, H. (1997) Feature Selection for Classification. *Intelligent Data Analysis: An International Journal*, **1**, 131-156. [http://dx.doi.org/10.1016/s1088-467x\(97\)00008-5](http://dx.doi.org/10.1016/s1088-467x(97)00008-5)
- [21] Malarvili, M.B., Mesbah, M. and Boashash, B. (2007) HRV Feature Selection Based on Discriminant and Redundancy Analysis for Neonatal Seizure Detection. *International Conference on Information, Communications and Signal Processing, ICICS*, Singapore, 10-13 December 2007, 1-5. <http://dx.doi.org/10.1109/icics.2007.4449765>
- [22] Malarvili, M.B., Mesbah, M. and Boashash, B. (2007) HRV Feature Selection for Neonatal Seizure Detection: A Wrapper Approach. *IEEE International Conference on Signal Processing and Communications*, Dubai, 24-27 November 2007, 864-867. <http://dx.doi.org/10.1109/icspc.2007.4728456>
- [23] Malarvili, M.B., Mesbah, M. and Boashash, B. (2008) Newborn Seizure Detection Based on Fusion of Multi-Channel EEG. *Proceedings of the 5th International Workshop on Signal Processing and its Application (WOSPA 08)*, Sharjah, 18-20 March 2008, 99-102.
- [24] Potamianos, G., Neti, C., Cravier, G., Garg, A. and Senior, A.W. (2003) Recent Advances in the Automatic Recognition of Audiovisual Speech. *Proceedings of the IEEE*, **91**, 1306-1326. <http://dx.doi.org/10.1109/jproc.2003.817150>
- [25] Chibelushi, C.C., Mason, J.S.D. and Deravi, F. (1997) Feature-Level: Data Fusion for Bimodal Person Recognition. *International Conference on Image Processing and Its Applications*, **1**, 399-403.
- [26] Fodor, I.K. (2002) A Survey of Dimension Reduction Techniques. Technical Report, UCRL-ID-148494, Lawrence Livermore National Laboratory, Livermore. <http://dx.doi.org/10.2172/15002155>
- [27] Fisher, R.A. (1936) The Use of Multiple Measurements in Taxonomic Problems. *Annals of Eugenics*, **7**, 179-188. <http://dx.doi.org/10.1111/j.1469-1809.1936.tb02137.x>
- [28] Chappelow, J., Madabhushi, A., Rosen, M., Tomaszewski, J. and Feldman, M. (2007) A Combined Feature Ensemble Based Mutual Information Scheme for Robust Inter-Modal, Inter-Protocol Image Registration. *4th IEEE International Symposium on Biomedical Imaging: From Nano to Macro*, Arlington, 12-15 April 2007, 644-647. <http://dx.doi.org/10.1109/isbi.2007.356934>
- [29] Yilmaz, A. (2007) Sensor Fusion in Computer Vision. Urban Remote Sensing Joint Event, 1-5.
- [30] Gupta, L., Chung, B., Srinath, M.D., Molfese, D.L. and Kook, H. (2005) Multichannel Fusion Models for the Parametric Classification of Differential Brain Activity. *IEEE Transactions on Biomedical Engineering*, **52**, 1869-1881. <http://dx.doi.org/10.1109/tbme.2005.856272>
- [31] Kittler, J., Hatef, M., Duin, R. and Matas, J. (1998) On Combining Classifiers. *IEEE Transactions Pattern Analysis and Machine Intelligence*, **20**, 226-239. <http://dx.doi.org/10.1109/34.667881>
- [32] Fukunaga, K. (1990) Introduction to Statistical Pattern Recognition. Academic Press, New York. <http://dx.doi.org/10.1016/b978-0-08-047865-4.50007-7>
- [33] Devijver, P.A. and Kittler, I. (1982) Pattern Recognition: A Statistical Approach. Prentice-Hall, Englewood Clis.
- [34] Armitage, P. and Colton, T. (1998) The Encyclopedia of Biostatistics. John Wiley & Sons, New York.
- [35] Press, W.H. (1988) Numerical Recipes in C: The Art of Scientific Computing. Cambridge University Press, Cambridge.
- [36] Greene, B.R., Reilly, R.B., Boylan, G., de Chazal, P. and Connolly, S. (2006) Multi-Channel EEG Based Neonatal Seizure Detection. *IEEE Engineering in Medicine and Biology Conference*, New York, 30 August-3 September 2006, 4679-884. <http://dx.doi.org/10.1109/iembs.2006.260461>

Scientific Research Publishing (SCIRP) is one of the largest Open Access journal publishers. It is currently publishing more than 200 open access, online, peer-reviewed journals covering a wide range of academic disciplines. SCIRP serves the worldwide academic communities and contributes to the progress and application of science with its publication.

Other selected journals from SCIRP are listed as below. Submit your manuscript to us via either submit@scirp.org or [Online Submission Portal](#).

