33rd Annual International Conference of the IEEE EMBS
Boston, Massachusetts USA, August 30 - September 3, 2011

Wepilet, Optimal Orthogonal Wavelets for Epileptic Seizure Prediction with one Single Surface Channel

Mojtaba Bandarabadi, Cesar A. Teixeira, Francisco Sales, Antonio Dourado, *Member IEEE*

Abstract— **Wepilet is a series of novel orthogonal wavelets optimized for Electroencephalogram (EEG) signals, specialized for epileptic seizure prediction. The main idea is to design a mother wavelet that when applied to EEG signal to create the feature space, should enable a better classification of the brain state. Wepilet is developed by an iterative optimization process, employing Genetic Algorithm (GA). Frequency sub-band features are first extracted using wepilet under design for the EEG signal captured by one single surface channel. These features are then fed to Support Vector Machines (SVMs) that classify the cerebral state in preictal and inter-ictal classes. The results of the classification are then used to compute the Probability of Error Rate (PER), which in turn is the GA objective function to be minimized. Results in a group of four patients, indicate the efficiency of optimized mother wavelet compared to the well-known Daubechies wavelet in EEG processing.**

I. Introduction

Around 20 Million people worldwide suffer from refractory epilepsy. They cannot be treated neither with medication nor with surgery, and must live with the seizures that can happen anytime, anywhere, “like a bolt from the sky” [1]. During the last twenty years many efforts have been made to develop the possibility to predict seizures, aiming to improve the living conditions of such patients. The problem has not yet found a solution. On one side, one must achieve high prediction capabilities, and on the other side low false alarm rates are needed, in order to allow the building of a transportable device clinically usable.

Epileptic seizure prediction was faced traditionally by applying thresholds to a given measures (feature) extracted from the EEG [2] or by nonlinear analysis [3]. More recently, classification methods based on high-dimensional feature spaces were used to detect the preictal state [4][5][6].

One of the methods of extracting features from EEG is to consider several frequency sub-bands, since these signals contain fast as well as gradual events, and they are highly non-stationary. Thus, wavelets with their time and frequency localization characteristics are considered as naturally powerful tools for signal processing in general, and to extract features from EEG in particular. However, the choice of good mother wavelet is essential to achieve good results. Several studies have been done using wavelets, mainly using the Daubechies family.

Optimal wavelet design has drawn a lot of attention in the past years [7][8][9]. The basic idea of optimal wavelet design is to look for some prototype filters that optimize some criteria. In addressing optimal wavelet problem, two issues arise: the first one is the choice of the criteria for optimality and the second one is the optimization algorithm to be used. Especially after the criterion is fixed, the optimization algorithm becomes more important. After selecting a decomposition filter bank corresponding to a scaling function $\varphi(t)$ and its related mother wavelet $\psi(t)$, one has to evaluate its suitability to be applied in the particular signal processing problem of interest. Thus a criterion function is necessary, so as to measure the performances of various filter banks.

The objective of this work is to design specific wavelets that should result in an improved classification of the preictal and inter-ictal states in patients suffering from refractory epilepsy. Preictal is the state just before one seizure that one wants to predict. Inter-ictal is normal brain state between two consecutive seizures.

A new approach is developed in the present work based on the theory of orthogonal filter banks to design the wavelets, on the Support Vector Machines (SVM) as the classifier, and on Genetic Algorithms as the optimization algorithm.

The remainder of the paper is organized as follows. Section II introduces basic concepts and describes steps and tools for designing the optimal wavelet (Wepilet). The evaluation of the proposed approach is presented in section III. Concluding remarks are presented in section IV.

II. Methodology

The methods and materials for designing the Wepilet and the details of the proposed algorithm are described in the following.

A. Wavelet Transform

Unlike the Fourier transform, in which basis functions are sinusoidal and redundant, the wavelet transforms are based on short-duration waves, of different frequencies and limited lengths. This characteristic makes them a favorable choice, providing us with frequency as well as temporal information for a given signal. In wavelet analysis, signal is decomposed into a sum of scaled and translated variations of the mother wavelet. A mother wavelet is simply a wavy function carefully constructed so as to have certain mathematical

Acknowledgments: this work was funded by the European Project EPILEPSIAE (EU FP7 Grant 211713).

M. Bandarabadi, C. A. Teixeira and A. Dourado are with the Centre for Informatics and Systems (CISUC), University of Coimbra, Portugal (phone:+351-239-790000; e-mail: {mojtaba, cteixei, dourado}@dei.uc.pt).

F. Sales is with the Hospitais da Universidade de Coimbra (HUC), Coimbra, Portugal, 3000-075. franciscosales@HUC.Min-Saude.pt

properties. Wavelet transform is in fact a measure of similarity between a signal and basis functions (mother wavelets). Here similarity means the analogy between frequency contents. In other words, wavelet coefficients indicate the closeness of signal to the wavelet in the desired scale. Thus, if the desired signal has a major component in the frequency corresponding to the scale under analysis, the scaled wavelet will be similar to that signal. Therefore the coefficient of the wavelet transform calculated for this scale would have a rather large value.

B. Parameterization of Orthogonal Wavelets

Compactly supported orthogonal wavelets have found common applications in various signal processing problems. Daubechies orthogonal wavelets are a very important class, which have been widely used by research community. The following parametric equations introduced in [10], are used for building length-8 compactly supported scaling functions of orthogonal mother wavelets:

$$\begin{aligned}
a_1 &= 0.125 + (\cos\alpha + 2 * \cos\beta * \cos\gamma)/4\sqrt{2} \\
a_2 &= 0.125 + (\sin\alpha\ + 2 * \sin\beta * \cos\theta)/4\sqrt{2} \\
a_3 &= 0.125 - (\cos\alpha - 2 * \cos\beta * \sin\gamma)/4\sqrt{2} \\
a_4 &= 0.125 - (\sin\alpha\ - 2 * \sin\beta * \sin\theta)/4\sqrt{2} \\
a_5 &= 0.125 + (\cos\alpha - 2 * \cos\beta * \cos\gamma)/4\sqrt{2} \\
a_6 &= 0.125 + (\sin\alpha\ - 2 * \sin\beta * \cos\theta)/4\sqrt{2} \\
a_7 &= 0.125 - (\cos\alpha + 2 * \cos\beta * \sin\gamma)/4\sqrt{2} \\
a_8 &= 0.125 - (\sin\alpha\ + 2 * \sin\beta * \sin\theta)/4\sqrt{2}
\end{aligned} \tag{1}$$

where -π≤ [α (Alpha), β (Beta), γ (Gamma), θ (Theta)] ≤ π must satisfy $f_x = 0$ as stated in (2):

$$\begin{aligned}
f_x = {} & 2\cos\theta * \sin\beta - 2\cos\theta * \sin\alpha * \sin\beta + \\
& 2\cos\beta * \cos\gamma - \sin\gamma - 4\cos^2\beta * \cos\gamma * \sin\gamma - \\
& 2\cos\alpha * \cos\beta * \cos\gamma + \sin\gamma - 2\sin\beta * \sin\theta - \\
& 2\sin\alpha * \sin\beta * \sin\theta - 4\cos\theta * \sin\theta * \sin^2\beta
\end{aligned} \tag{2}$$

To satisfy equation (2) one of the four parameters α, β, γ, θ has to be found depending on the other three parameters. Thus one has three degrees of freedom to select a wavelet.

To solve these equations two major steps are needed: (a) Three out of four parameters (α, β, γ) are selected freely and (b) Theta (θ) is then selected so that the equation (2) is fulfilled. To solve this equation (2) for Theta, Brent's method [11] is employed, which is a combination of three root-finding algorithms, namely the bisection method, the secant method and inverse quadratic interpolation method. Although the choice of the three parameters can be done freely, yet only some optimized sets of these parameters exist for specific applications and should be found. To carry out this, an optimization problem has to be formulated and solved. Considering the very large search space of this optimization problem, not every search method can be effectively employed. In such cases, particular search methods are usually employed such as GA, Simulated Annealing (SA), Particle Swarm Optimization (PSO), etc. In the next GA is presented to obtain the optimized parameters.

C. Genetic Algorithm

Genetic Algorithm (GA) is one of the commonly used approaches in optimization problems, introduced by John Holland [12]. Crossover and fitness factor are the two characteristics of the GA algorithm having major impact the resulting outcome. In contrast, mutation plays a less important role. Crossover determines the way through which pairs of parents are mixed to generate new off-springs. In this work the scattered crossover [13] is employed. On the other hand, mutation is required to make small changes in the genes' population to provide diversity and luckily to enable GA to escape local minima. Adaptive Feasible mutation function embedded in the Matlab toolbox is used in our study, which generates random best feasible directions, to produce mutated children. The population size is dependent on the number of variables to be optimized as well as the range of the changes of these variables. In our study, population size of 400 chromosomes is selected to allow a sufficiently large coverage of the search space.

The fitness function drives the population toward better solutions [14]. So, the definition of a good fitness function that rewards the right population of individuals is the most important stage in optimizing a wavelet using genetic algorithms. As soon as a wavelet is designed, the performance of the wavelet should be evaluated. Thus, a quality criterion is required, for which we've employed the Probability of Error Rate (PER) of the classifier. This quality criterion is fed back to the GA to change the three variables and to optimize them to develop new seizure prediction specific wavelets.

D. Probability of Error Rate

The prediction accuracy is closely related to the accuracy of inter-ictal/preictal classification, which is in turn a good quality criterion to be used as GA fitness function. One has to increase the classification performance to achieve better predictions. Thus our goal is to enhance the accuracy of classification using the features extracted through the wavelet transform with optimized mother wavelets. On the other hand, the same classification approaches could be used in the design stage.

The fitness function chosen to be minimized (and thus improve accuracy) by the GA is the Probability of Error Rate (PER) [15], which is calculated by summing the percentages of all samples belonging to the undesired classes, and is given by (3), where *Prob*(*k*) is the probability that sample *k* belong to the undesired class.

$$PER = \sum_{k=1}^{n} Prob(k)\,, n = Number\ of\ Samples \tag{3}$$

The criterion PER can achieve better results compared to pure error rate, as it provides more details about the separability of features in the feature space. The Support Vector Machines (SVM) is selected among different of classifiers because they are considered to have the best separation capability.

E. *Support Vector Machines*

Support vector machines (SVMs) are a set of commonly used supervised learning methods employed for classification problems [16][17]. SVM classifiers in their simplest form use linear boundaries to classify two sets of data. To classify datasets with nonlinear boundaries, SVM employ a kernel function to transform the nonlinear boundary into a linear one. The popular Gaussian Radial Basis Function (RBF) kernel (4) is used,

$$K(x,y) = exp\,(\frac{-|x-y|^2}{2\sigma^2}) \qquad (4)$$

where σ is the scale parameter, *x, y* are feature vectors in the input space. The Gaussian kernel has two hyper parameters to control classification performance: the cost C and the scale parameterσ.

Parameter C controls the tradeoff between maximization of the margin width and the minimization of the number of misclassified samples in the training set [16]. Also, the σ parameter in (4) controls the width of the Gaussian surface of the RBF kernel. These two parameters are coded in the chromosomes of the GA population, and are optimized inside the fitness function through an ordinary search method.

F. *Proposed Algorithm*

Initially, Alpha, Beta and Gamma parameters are generated by the GA, and then GA passes these parameters into the fitness function. The fourth parameter Theta is also calculated by solving the equation (2) $f_x = 0$, and employing Brent's method. These four parameters are then fed into the parameterized wavelet equations to obtain the scaling function of the mother wavelet. Fig.1 illustrates the block diagram of the employed fitness function.

The designed wavelets are applied to the raw EEG segments of both inter-ictal and preictal classes to decompose every segment into its corresponding sub-bands. Then energy of each sub-band is calculated, i.e., the square of the Wepilet coefficients, originating the feature space. Classification is carried out using LibSVM toolbox [18], SVMs are trained in a part of the data and tested in a different one. Finally, the PER of the classification, i.e., the output of the fitness function is calculated in the testing data. The calculation of PER becomes possible by activating related probabilistic outputs' option in the LibSVM toolbox. We also optimized the parameters of the SVM classifier inside the fitness function to find the best results that could be achieved with the designed wavelet.

The procedures above are then repeated by GA, until the algorithm converges to an optimized solution.

III. Experimental Results

Data from four epileptic patients from the EPILEPSIAE database [19] with long-term continuous multichannel EEG recordings were used to evaluate the proposed method (Table I). All patients had focal seizures. For each patient

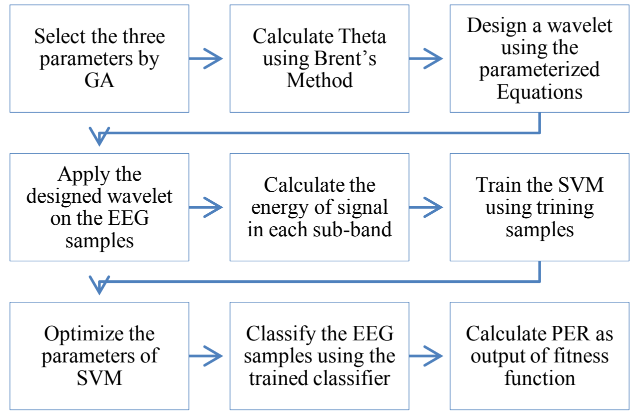

Fig. 1. Block Diagram of Fitness Function.

TABLE I
Information of Patients' Database

Pat.	No. Seizures	Sampling Rate	Rec.Time (hours)
1	6	1024 Hz	66.7 h
2	4	1024 Hz	108.1 h
3	5	512 Hz	35 h
4	9	256 Hz	143.8 h

one electrode located over or close to the seizure focus was selected, with its signal relative to a reference electrode. The selected patients had a total of 24 epileptic seizures. For each patient half of the seizures were chosen for wavelet design. The other half was selected to test the optimized wavelet.

Preictal periods are the time interval from 10 minutes to 1 minute before every seizure. For inter-ictal, the data outside the interval starting at 90 minutes before every seizure and ending at 30 minutes after that same seizure was selected. This assures that there is no preictal neither post-ictal activity in the data selected as inter-ictal. Sampling rates of these EEG recording were 1024Hz for two first patients, 512Hz for the third one and 256Hz for the last one. Before starting decomposition, all signals were first down-sampled into 256Hz in order to reduce the computational costs and hypothesizing that there is no important information in the frequencies above 128Hz. During every 6-level decomposition iteration, signals are decomposed to obtain seven sub-bands of 0-2, 2-4, 4-8, 8-16, 16-32, 32-64, 64-128 Hz. These bands have been chosen by a trial and error procedure, showing to be the best division from the classification capability perspective.

The optimal wavelets were then computed for each patient (Table II) and applied to the respective one channel EEG signal. The wavelet coefficients were then given to a SVM classifier. The classification results achieved are presented in table III, and compared by those obtained using Daubechies-4 wavelet.

The Area Under Curve (AUC) [20] criterion is also shown in table III. This criterion is considered by some authors to be superior to the error rate criterion, to measure the quality of a classification algorithm [20]. The higher the AUC, the better the classifier; AUC = 1 is the best possibility.

TABLE II
OPTIMIZED WAVELETS FOR EACH PATIENT

Param.	Db4	Wp_1	Wp_2	Wp_3	Wp_4
Alpha	+2.2401	+0.4151	+0.7111	-1.3640	-0.5355
Beta	+0.7535	+0.9046	+0.9853	+2.6769	+1.1791
Gama	+0.9614	+0.0046	+1.9100	+1.5285	+0.6085
Theta	-0.0254	-2.0668	-1.9472	-2.1915	-1.4478
a_1	+0.1629	+0.5053	+0.1939	+0.1479	+0.3878
a_2	+0.5055	+0.0640	+0.1321	-0.1402	+0.0748
a_3	+0.4461	-0.0358	+0.1753	-0.2271	+0.0501
a_4	-0.0198	-0.1907	-0.26441	+0.1691	-0.1091
a_5	-0.1323	+0.0689	+0.3239	+0.1747	+0.1663
a_6	+0.0218	+0.3286	+0.3487	+0.0441	-0.0053
a_7	+0.0233	-0.0377	-0.1932	+0.4045	-0.1042
a_8	-0.0075	+0.2982	+0.2836	+0.4269	+0.5395

TABLE III
RESULTS OF CLASSIFICATION

	Preictal*	Inter-ictal*	AUC**
Db4	41.07 %	82.88 %	0.5351
Wp_1	45.03 %	89.23 %	0.5754
Db4	52.59 %	75.93 %	0.6792
Wp_2	57.04 %	77.92 %	0.7204
Db4	21.97 %	78.96 %	0.5214
Wp_3	34.32 %	75.34 %	0.5627
Db4	62.95 %	69.72 %	0.7226
Wp_4	69.31 %	75.08 %	0.7752

* Classification Accuracy Percentage for Preictal/Inter-ictal Samples
** Area Under the ROC Curve

IV. CONCLUSION

A new approach was introduced for designing the optimal orthogonal wavelets for seizure prediction problem, using only one EEG surface electrode (plus a reference electrode).

An average improvement of the classification accuracy of 6.8 % was obtained for preictal test samples with respect to the more traditional Daubechies wavelets. Results show the ability of the proposed technique to classify preictal periods, and thus present a promising tool to predict seizures. However further work is needed to improve the performance. One way may be to use more than one EEG channel. Our research aims at good seizure predictors with a low number of channels (less than 6) in order to allow the development of transportable devices for incoming seizure warning.

In addition to simple energy feature employed here, obtained by the square of the band coefficients extracted by the wavelet under design, other features such as percentage of energy of the signal in each sub-band, statistical moments of each sub-band, and so on, can as well be considered to design optimized wavelets which may achieve better classification results and shall be our future work.

REFERENCES

[1] A. Dourado, R. Martins, J. ao Duarte, and B. Direito, "Towards Personalized Neural Networks for Epileptic Seizure Prediction," *Proceedings of the 18th international conference on Artificial Neural Networks, Part II*, Berlin, Heidelberg: Springer-Verlag, 2008, p. 479–487.

[2] B. Schelter, M. Winterhalder, T. Maiwald, A. Brandt, A. Schad, A. Schulze-Bonhage, and J. Timmer, "Testing statistical significance of multivariate time series analysis techniques for epileptic seizure prediction," *Chaos (Woodbury, N.Y.)*, vol. 16, Mar. 2006, p. 013108.

[3] M. Le Van Quyen, J. Martinerie, M. Baulac, and F. Varela, "Anticipating epileptic seizures in real time by a non-linear analysis of similarity between EEG recordings," *Neuroreport*, vol. 10, Jul. 1999, pp. 2149-2155.

[4] B. Direito, A. Dourado, M. Vieira, and F. Sales, "Combining Energy and Wavelet Transform for Epileptic Seizure Prediction in an Advanced Computational System," *BioMedical Engineering and Informatics, International Conference on*, Los Alamitos, CA, USA: IEEE Computer Society, 2008, pp. 380-385.

[5] P.W. Mirowski, Y. Lecun, D. Madhavan, and R. Kuzniecky, "Comparing SVM and Convolutional Networks for Epileptic Seizure Prediction from Intracranial EEG."

[6] L. Chisci, A. Mavino, G. Perferi, M. Sciandrone, C. Anile, G. Colicchio, and F. Fuggetta, "Real-time epileptic seizure prediction using AR models and support vector machines," *IEEE Transactions on Bio-Medical Engineering*, vol. 57, May. 2010, pp. 1124-1132.

[7] R.S.S. Kumari, S. Bharathi, and V. Sadasivam, "Design of Optimal Discrete Wavelet for ECG Signal Using Orthogonal Filter Bank," *Computational Intelligence and Multimedia Applications, International Conference on*, Los Alamitos, CA, USA: IEEE Computer Society, 2007, pp. 525-529.

[8] J. M.H. Karel, R. L.M. Peeters and R. L. Westra and K. M.S. Moermans and S. A.P. Haddad, , and , W. A. Serdijn, "Optimal discrete wavelet design for cardiac signal processing," 2005, pp. 2769-2772.

[9] E. Jones, P. Runkle, N. Dasgupta, L. Couchman, and L. Carin, "Genetic Algorithm Wavelet Design for Signal Classification," *IEEE Transactions on Pattern Analysis and Machine Intelligence*, vol. 23, 2001, pp. 890-895.

[10] M.-J. Lai and D.W. Roach, "Parameterization of univariate orthogonal wavelets with short support," *Approximation Theory X: Wavelets, Splines, and Applications*, Vanderbilt Univ. Press, 2002, pp. 369-384.

[11] R.P. Brent, *Algorithms for Minimisation without Derivatives*, Prentice Hall, 1972.

[12] J.H. Holland, *Adaptation in Natural and Artificial Systems: An Introductory Analysis with Applications to Biology, Control, and Artificial Intelligence*, The MIT Press, 1992.

[13] The Mathworks, "Matlab's Genetic Algorithm and Direct Search ToolboxUser's Guide.," 2005.

[14] L.R. Knight and R.L. Wainwright, "HYPERGEN - A Distributed Genetic Algorithm on a Hypercube," *IEEE*, Williamsburg: IEEE Press, 1992, pp. 232-235.

[15] J.C. Platt, "Probabilistic Outputs for Support Vector Machines and Comparisons to Regularized Likelihood Methods," *ADVANCES IN LARGE MARGIN CLASSIFIERS*, 1999, p. 61--74.

[16] B. Schlkopf and A.J. Smola, *Learning with Kernels: Support Vector Machines, Regularization, Optimization, and Beyond*, The MIT Press, 2001.

[17] B.E. Boser, I.M. Guyon, and V.N. Vapnik, "A training algorithm for optimal margin classifiers," *Proceedings of the fifth annual workshop on Computational learning theory*, New York, NY, USA: ACM, 1992, p. 144–152.

[18] C.-chung Chang and C.-J. Lin, "LIBSVM: a Library for Support Vector Machines," 2001.

[19] M. Ihle, H. Feldwitch-Drentrup, C.A. Teixeira, A. Witon, B. Schelter, J. Timmer, and A. Schulze-Bonhage, "EPILEPSIAE - A common database for research on seizure prediction," *Computer Methods and Programs in Biomedicine*, 2010.

[20] C. Cortes and M. Mohri, "AUC optimization vs. error rate minimization," *IN ADVANCES IN NEURAL INFORMATION PROCESSING SYSTEMS*, vol. 16, 2004.