

7-10-2009

A Research Platform for Artificial Neural Networks with Applications in Pediatric Epilepsy

Melvin Ayala

Florida International University, melvinayala@yahoo.com

DOI: 10.25148/etd.FI09080600

Follow this and additional works at: <https://digitalcommons.fiu.edu/etd>

Recommended Citation

Ayala, Melvin, "A Research Platform for Artificial Neural Networks with Applications in Pediatric Epilepsy" (2009). *FIU Electronic Theses and Dissertations*. 90.

<https://digitalcommons.fiu.edu/etd/90>

This work is brought to you for free and open access by the University Graduate School at FIU Digital Commons. It has been accepted for inclusion in FIU Electronic Theses and Dissertations by an authorized administrator of FIU Digital Commons. For more information, please contact dcc@fiu.edu.

FLORIDA INTERNATIONAL UNIVERSITY

Miami, Florida

A RESEARCH PLATFORM FOR ARTIFICIAL NEURAL NETWORKS WITH
APPLICATIONS IN PEDIATRIC EPILEPSY

A dissertation submitted in partial fulfillment of the

requirements for the degree of

DOCTOR OF PHILOSOPHY

in

ELECTRICAL ENGINEERING

by

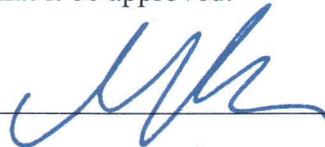
Melvin Ayala

2009

To: Dean Amir Mirmiran
College of Engineering and Computing

This dissertation, written by Melvin Ayala, and entitled A Research Platform for Artificial Neural Networks with Applications in Pediatric Epilepsy, having been approved in respect to style and intellectual content, is referred to you for judgment.

We have read this dissertation and recommend that it be approved.



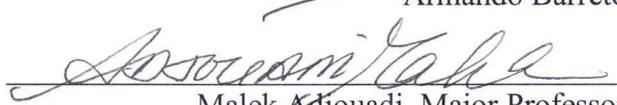
Naphtali Rishe



Jean Andrian



Armando Barreto



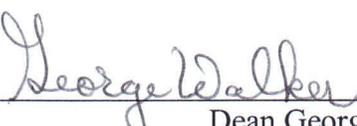
Malek Adjouadi, Major Professor

Date of Defense: July 10, 2009

The dissertation of Melvin Ayala is approved.



Dean Amir Mirmiran
College of Engineering and Computing



Dean George Walker
University Graduate School

Florida International University, 2009

DEDICATION

To my family and friends

ACKNOWLEDGMENTS

My deepest appreciation goes to my advisor, Dr. Malek Adjouadi, for the admirable support and guidance provided for the completion of this work, as well as for his incomparable dedication to mentoring.

My appreciation extends also to my committee members: Dr. Armando Barreto, Dr. Naphtali Rische, and Dr. Jean Andrian, for the time they spent advising and encouraging me through this research.

My gratitude is also to the Center for Advanced Technology and Education (CATE), for giving me access to the equipment and software that allowed me to perform this research. I thank all the CATE lab members, especially Mercedes Cabrerizo, Magno Guillen and Xiaozhen You for their assistance on different tasks related to my research.

I am thankful for the clinical support and the data provided through the Ware Foundation and the joint Neuro-Engineering Program with Miami Children's Hospital.

Last, but definitely not least, I appreciate the support provided by the National Science Foundation under grants HRD-0317692, CNS-0426125, CNS-0520811, CNS-0540592, HRD-0833093, CNS-0426125, CNS-0520811, CNS-0540592, and IIS-0308155.

ABSTRACT OF THE DISSERTATION

A RESEARCH PLATFORM FOR ARTIFICIAL NEURAL NETWORKS WITH APPLICATIONS IN PEDIATRIC EPILEPSY

by

Melvin Ayala

Florida International University, 2009

Miami, Florida

Professor Malek Adjouadi, Major Professor

This dissertation established a state-of-the-art programming tool for designing and training artificial neural networks (ANNs) and showed its applicability to brain research. The developed tool, called NeuralStudio, allows users without programming skills to conduct studies based on ANNs in a powerful and very user friendly interface.

A series of unique features has been implemented in NeuralStudio, such as ROC analysis, cross-validation, network averaging, topology optimization, and optimization of the activation function's slopes. It also included a Support Vector Machines module for comparison purposes. Once the tool was fully developed, it was applied to two studies in brain research. In the first study, the goal was to create and train an ANN to detect epileptic seizures from subdural EEG. This analysis involved extracting features from the spectral power in the gamma frequencies. In the second application, a unique method was devised to link EEG recordings to epileptic and non-epileptic subjects. The contribution of this method consisted of developing a descriptor matrix that can be used to represent any EEG file regarding its duration and the number of electrodes.

The first study showed that the inter-electrode mean of the spectral power in the gamma frequencies and its duration above a specific threshold performs better than the other frequencies in seizure detection, exhibiting an accuracy of 95.90%, a sensitivity of 92.59%, and a specificity of 96.84%. The second study yielded that Hjorth's parameter activity is sufficient to accurately relate EEG to epileptic and non-epileptic subjects. After testing, accuracy, sensitivity and specificity of the classifier were all above 0.9667. Statistical tests measured the superiority of activity at over 99.99 % certainty.

It was demonstrated that 1) the spectral power in the gamma frequencies is highly effective in locating seizures from EEG and 2) activity can be used to link EEG recordings to epileptic and non-epileptic subjects. These two studies required high computational load and could be addressed thanks to NeuralStudio. From a medical perspective, both methods proved the merits of NeuralStudio in brain research applications. For its outstanding features, NeuralStudio has been recently awarded a patent (US patent No. 7502763).

TABLE OF CONTENTS

CHAPTER	PAGE
1. INTRODUCTION.....	1
2. PHYSIOLOGY AND DETECTION OF EPILEPTIFORM ACTIVITY	8
2.1. Processing Epileptiform Activity in EEG	8
2.1.1. Basics	8
2.1.2. Epilepsy: A Synopsis	13
2.1.3. Epileptiform Activity	14
2.1.4. Interictal Spikes.....	15
2.1.5. Seizures	17
2.1.6. Feature Extraction in Time Domain.....	18
2.1.7. Statistical Features.....	19
2.1.8. Hjorth’s Descriptors	20
2.1.9. Non-Linear Features.....	21
2.1.10. Other Features	24
2.1.11. Feature Extraction in Frequency Domain	26
2.1.12. Power Spectrum and Relevant Frequency Bands	27
2.1.13. Spectral Features	29
2.2. Overview of Methods for the Automated Detection of EFA and Seizures.....	32
2.2.1. Basic Operational Steps of Existing Methods.....	32
2.2.2. Performance Evaluation of Detection Algorithms.....	34
3. DESIGN AND IMPLEMENTATION OF NEURALSTUDIO: AN ARTIFICIAL NEURAL NETWORKS DESIGN AND EVALUATION TOOL	40
3.1. Introduction	40
3.2. Significance of the NeuralStudio Design	42
3.3. General Description and Functionality.....	45
3.4. Design of the Graphical User Interface.....	49
3.4.1. Data Entry and Preprocessing.....	49
3.4.2. Creating and Training Neural Networks.....	54
3.4.2.1. Feedforward Networks Design and Training	54
3.4.2.2. Self-Organizing Maps.....	60
3.4.2.3. Clustering Networks.....	62
3.4.2.4. Pattern Association Networks	64
3.4.2.5. Fixed-Weight Networks	65
3.4.3. Activation Function Module	66
3.4.4. Support Vector Machines Module	68
3.4.5. Code-Insight Module.....	69
3.4.6. Performance Metric Module	71
3.4.7. Help Module.....	72

4.	AN APPLICATION OF NEURALSTUDIO TO THE DETECTION OF EPILEPTIC SEIZURES FROM SUBDURAL EEG	75
4.1.	Objectives	75
4.2.	Experimental Setup.....	78
4.3.	Data Preprocessing	79
4.4.	Extracting the Power Spectrum in the Gamma Band	80
4.5.	Aggregating Features.....	82
4.6.	Establishing the Inputs of the ANN.....	83
4.7.	ANN Configuration	88
4.8.	Training Procedure	89
4.9.	Training Results.....	93
4.9.1.	Findings Regarding the Number of Hidden Neurons.....	93
4.9.2.	Findings Regarding Data Arrangement.....	96
4.10.	Testing Results	98
4.11.	Comparative Analysis under Different Frequency Bands.....	101
4.12.	Conclusion of the Method	102
5.	AN APPLICATION OF NEURALSTUDIO TO THE CLASSIFICATION OF EPILEPTIC AND NON-EPILEPTIC SUBJECTS FROM SCALP EEG.....	105
5.1.	Objectives of the Method	105
5.2.	Data Preprocessing	106
5.3.	Feature Extraction.....	107
5.4.	Dimensionality Analysis.....	112
5.5.	Training Setup	116
5.5.1.	Data Organization	116
5.5.2.	ANN Configuration and Training Procedure	118
5.6.	Testing Results	121
5.7.	Statistical Analysis of the Results	124
5.8.	Concluding Remarks	128
6.	CONCLUSIONS	134
	BIBLIOGRAPHY.....	138
	VITA.....	157

LIST OF TABLES

TABLE		PAGE
Table 2. 1:	Entries of a confusion matrix.....	34
Table 4.1:	Data set organization for the study, showing the amount of interictal and ictal files used from each patient to train and test the classifier.	80
Table 4.2:	Setup of two arrangements with indication of the files pertaining to the training, cross-validation and testing (N_{neg} = number of interictal files, N_{pos} = number of ictal files, N_{total} = total number of files).	92
Table 4.3:	Five best performing topologies and their performance in arrangements A and B, sorted in descending order of the average pattern square error on the validation set (L: linear activation, S: logsig activation, G: Gaussian activation). Best values are in bold.	96
Table 4.4:	Five best performing topologies on both arrangements when sorted in descending order of the average pattern square error on training and validation set.	97
Table 4.5:	Errors obtained after training and testing the corresponding topologies on the three subsets of arrangements A and B: pure training, cross-validation, and testing. Best values are in bold.	99
Table 4.6:	Testing errors obtained after training a 2-5-1 topology on the data set extracted for each frequency band. Best values are in bold.	101
Table 5.1:	Subjects' information and data used	106
Table 5.2:	Segment descriptor matrix for a particular feature	111
Table 5.3:	Common header structure of training, cross-validation and testing tables, including the nine statistical parameters proposed and the target column.....	113
Table 5.4:	Data arrangement for training purposes.....	117
Table 5.5:	Procedure implemented to generate the 10 solutions per network after random starts. Accuracy, sensitivity and specificity are represented by X1, X2, and X3, respectively.....	122
Table 5.6:	Best solutions found in all trials. Within each group, the best solution is in bold.	123

Table 5.7:	t-test results that prove that the minimum metric found for the activity descriptor is significantly higher than the remaining metrics. For a 0.05 significance, the critical t-value is 1.795885, but the computed t-value was much higher (8.495415) representing a one-tail p-value of only $1.836425 \cdot 10^{-6}$. $P(x \geq t \text{ Critical})$ is the probability that an observation x is higher than t Critical if the null hypothesis is true, whereas the null hypothesis is that the minimum metrics for activity and the remaining descriptors do not differ in average. The test assumes unequal variances.....	125
Table 5.8:	ANOVA test results that prove that the variance of the minimum metrics between the descriptors is not different by chance (df: degrees of freedom, SS: sum of squares, MS: mean of squares or variance).....	126
Table 5.9:	All descriptors, sorted in descending order of their minimum performance for EEG classification.....	127

LIST OF FIGURES

FIGURE	PAGE
Figure 2.1: An EEG of a patient showing epileptiform activity (Interictal spikes) in some channel pairs (courtesy of Miami Children’s Hospital).....	9
Figure 2.2: Location of the electrodes in the 10-20 system	10
Figure 2.3: Location of the electrodes in the modified combinatorial nomenclature MCN	10
Figure 2.4: An EEG cap with complete set (Image taken with the permission of a student volunteer from the CATE lab at FIU).	11
Figure 2.5: An electrode grid being implanted on the cortex of the brain of a patient (Courtesy of the Brain Institute at Miami Children’s Hospital).	12
Figure 2.6: Interictal spikes occurring synchronously in four distinct electrodes or channels	16
Figure 2.7: Characterizing features of an interictal spike	16
Figure 2.8: An EEG containing the transition from interictal to ictal state, showing electrode pairs that initiate the seizure (courtesy of Miami Children’s Hospital).....	17
Figure 2.9: Two ROC curves plotted from the confusion matrixes of two classifiers, showing how erratic their performance can be.....	38
Figure 3.1: NS’s main window, displaying a main menu and tool bar that gives access to all network designing tools as well as to the patterns editor, the activation functions, the performance evaluation module, and an option and help menu, among others.	46
Figure 3.2: Pattern window with a particular data set entered and plotted. A toolbar facilitates data editing and also grants access to different supporting tools to preprocess the data. In this case, there is only one input and one output, i.e. two variables.	50
Figure 3.3: Sensitivity tools as given in NS provide an easy way to dispose unwanted data: (a) Contradictory patterns search option, giving a warning of all data that is contradictory. (b) Column-wise analysis, where all variables are checked for similarity.	51
Figure 3.4: 2D map generation in NS allows creating data points by using the mouse pointer.	52

Figure 3.5: A set of options are provided to configure the k-fold cross-validation algorithm.....	53
Figure 3.6: The bootstrapping technique is configured with only two options.....	54
Figure 3.7: Feedforward networks module, displaying the network editor and the input/output panel	56
Figure 3.8: Snapshot of the supervised training module during a training session to approximate a cosine function.	57
Figure 3.9: View of the topology optimization panel for backpropagation networks ...	59
Figure 3.10: Representation of the winner-takes-all algorithm implemented in NS. The node closest to the input pattern is selected as the winner neuron and its weights are partially adapted according to the input pattern weights. Graphically, the weight update makes the winner neuron move towards the input pattern.	61
Figure 3.11: Configuration panel of SOM module, where basic map features and input source are selected	62
Figure 3.12: NS's module for self-organizing maps during training	62
Figure 3.13: A snapshot of NS's module for clustering networks during a training session. The user can customize the training by setting the amount of clusters and deciding which clustering method to use.....	63
Figure 3.14: Module for pattern association while performing character codifications ..	64
Figure 3.15: Module of the fixed weight networks while performing an optimization ...	65
Figure 3.16: Activation functions module displaying options to assign parameters to the logistic sigmoid function.....	67
Figure 3.17: A snapshot of the SVM module during training.....	69
Figure 3.18: The code-insight module provides pseudo-code based in PASCAL for programming-interested users.....	70
Figure 3.19: A snapshot of the performance metric module, showing the first page with a demo that allows users to understand the basic concepts of ROC....	72
Figure 3.20: NS's help contents file, opened to the full extent	73
Figure 3.21: NS's help file, showing a specific topic previously selected from the contents file.....	73

Figure 4.1:	An illustrative example of gamma power for all electrodes versus time (shown in terms of samples) for 4 different seizures from different patients. The vertical red line is the seizure onset as identified by medical experts. Note that the scales are different for the different subjects, thus making absolute thresholds impossible to use.	82
Figure 4.2:	Top figure shows the behavior of the gamma power for each of the 48 electrodes used for subject 1, seizure 5, and the bottom figure displays its respective inter-electrode mean signal. Seizure onset is around 100 seconds.	84
Figure 4.3:	Illustration of the variation of the inter-electrode mean signal S_{μ} within the same patient: (a) seizure 1 of patient 1, (b) seizure 2 of patient 2, (c) zoomed in view of seizure 2 of patient 2.	85
Figure 4.4:	Positive (ictal files) and negative (interictal files) points used for training (a) and testing (b). The x-axis represents the duration in seconds (normalized with respect to the maximum), whereas the y-axis represents the maximum value of the inter-electrode gamma power (also normalized).	87
Figure 4.5:	The 2-5-1 ANN topology chosen in NS to detect seizures based on the power spectrum in the gamma band.	89
Figure 4.6:	NS's training module while performing a topology optimization in the 2-X-1 group with linear activation in the first two layers and logsig in the last layer. The number of hidden neurons is being varied from 1 to 10.	91
Figure 4.7:	NS while performing topology variation training in a loop, showing the plot of the average accuracy after ten trials in each set for every specific amount of hidden neurons.	95
Figure 4.8:	Behavior of the grand average of the mean pattern square error (in training and validation set) vs. the number of hidden neurons in L-L-L (left) and L-S-S (right) topologies.	95
Figure 5.1:	(a) First step in feature extraction from EEG recordings, illustrating the extraction of EEG activity from 1-second time windows for a set of electrodes. Window activity $A(E_k)$ is computed for each electrode E_k for each time window. (b) Second step in feature extraction, using inter-electrode average as statistical parameter to analyze EEG activity in time independent of the number of electrodes. (c) Last step of the proposed feature extraction procedure, using grand average over time of the inter-electrode average as overall parameter to analyze EEG activity of a specific EEG recording file.	107

Figure 5.2: Compilation of all steps needed to extract a global descriptor for a particular feature for an entire EEG file. Param1 is an inter-electrode statistical parameter, and Param2 is a statistical parameter of the time series of Param1.....	109
Figure 5.3: Proposed approach to feature extraction from EEG data segments, reduced to a 3x3 matrix obtained from the common statistical parameters avg, std and snr.....	110
Figure 5.4: A generalization of the global descriptor with three dimensions (feature cube): one vertical for the inter-electrode aggregation, one horizontal for the time aggregation, and another for the statistics across subjects. Keep in mind that additional parameters can be used.....	112
Figure 5.5: Plots of the gamma descriptors for all 105 files together, showing combinations of dimension pairs (dimension1-dimension2) with different clustering abilities. Files from non-epileptic and epileptic subjects are represented by blue dots and red crosses, respectively.....	114
Figure 5.6: Zoom in of plot 1-2 of Figure 5.5, showing a better cluster organization once outliers have been excluded.....	115
Figure 5.7: Ratio of file usage during classifier training.....	116
Figure 5.8: Automation of table separation in NeuralStudio. The user can load the whole data into a table and from there make a selection, row by row, of where to send each row. By entering 1, 2, or 3 in the last column provided, the rows can be sent automatically to the training, cross-validation and testing tables, respectively.....	117
Figure 5.9: Illustration of the cross-validation criterion for training stop. The point where the network starts memorizing and losing its generalization ability is signaled by $N_{critical}$. The testing error can have any behavior but is not part of the stop procedure.....	120
Figure 5.10: Screenshot of NeuralStudio's training window, showing a panel that allows the user loading multiple ANN files at once and train each of them several times. In this process, the best networks are automatically stored in the interface and also saved to disk.....	120
Figure 5.11: Variation of the nine descriptors. The box plot suggests the size of the F-statistic and the p-value. Large differences in the center lines of the boxes correspond to large values of F and correspondingly small values of p.....	126

- Figure 5.12: A representation of the spectral power of a non-epileptic person's EEG in the delta, theta, alpha, beta I, beta II, and gamma frequencies is given in (a) through (e), respectively. No trend can be observed as the frequency band changes. 130
- Figure 5.13: A plot of the spectral power of an epileptic patient's scalp EEG in the delta, theta, alpha, beta I, beta II, and gamma frequencies is given in (a) through (e), respectively. A narrowing of the plots is observable as the frequency band changes. 131
- Figure 5.14: A plot of the spectral power of an epileptic patient's iEEG in the delta, theta, alpha, beta I, beta II, and gamma frequencies is given in (a) through (e), respectively. A narrowing of the plots is observable as the frequency band changes. 132

SYMBOLS AND ABBREVIATIONS

2D	two-dimensional
9D	nine-dimensional
AF	Activation Function
AI	Artificial Intelligence
ANN	Artificial Neural Networks
ANOVA	Analysis of Variance
AVG	Average
avg	Average
CCR	Correct Classification Rate
CT	Computed Tomography
df	Degrees of freedom
ECG	Electrocardiogram
ECoG	Electrocorticography
EFA	Epileptiform Activity
EEG	Electroencephalograph
FL	Fuzz Logic
FN	False Negatives
FNr	False Negative rate
FP	False Positives
FPr	False Positive rate
GA	Genetic Algorithms
iEEG	Intracranial EEG

logsig	Logistic sigmoid function
Max	Maximum
MaxMin	MaximumMinimum procedure in which the solution is the maximum of all minimum values.
MCH	Miami Children's Hospital
MRI	Magnetic Resonance Imaging
MS	Mean of squares (variance)
Min	Minimum
fMRI	Functional Magnetic Resonance Imaging
MCR	Misclassification Rate
NS	NeuralStudio
N_{total}	Total number of files
N_{neg}	Number of interictal files
N_{pos}	Number of ictal files
Purelin	Pure linear activation function
Purelinnorm	Pure linear activation function with normalized output
ROC	Receiver Operating Characteristics
PCA	Principal Component analysis
PET	Positron Emission Tomography
SMR	Sensory-Motor-Rhythm (former description of low beta (13-15 Hz) waves)
SOM	Self-Organizing Maps
SPECT	Single Photon Emission Computed Tomography
SS	Sum of squares

SNR	Signal to Noise Ratio
snr	Signal to Noise Ratio
STD	Standard Deviation
std	Standard Deviation
tansig	Tangent sigmoid function
trivalent	Activation function with three outputs (-1, 0, and 1)
TN	True Negatives
TNf	True Negative fraction
TNr	True Negative rate
TP	True Positives
TPf	True Positive fraction
TPr	True positive rate
t-Test	Student's test
n1-n2-n3	ANN topology, where n1, n2, and n3 are the amount of neurons in input, hidden and output layer, respectively
L	Linear activation
S	Sigmoid activation
G	Gaussian activation

CHAPTER 1

1. INTRODUCTION

This dissertation proposes a research platform for the application of artificial neural networks (ANNs) in pediatric epilepsy. This research platform is designed with a generalized topology that will allow researchers to apply ANNs in different research disciplines. Given the strong collaboration between FIU and MCH, the research methodology and the practical implications of this dissertation focus on brain research towards a better understanding of seizures and their causality.

The first part of this dissertation dwells in the design approach taken in establishing the artificial neural network platform we call NeuralStudio (NS). The novelty of this platform has led to a U.S. patent that has been granted in March 2009 [Ayala and Adjouadi 2009].

In order to assess the research and practical merits of this NeuralStudio, new algorithms have been developed within this research platform to address the problems associated with the complex problems of detection and ultimately prediction of epileptic seizures. In this research endeavor, the key aspects in establishing ANN architectures with the optimal weights for all the necessary decision functions are provided in details. What is unique in this approach is that it allows researchers to make use of this NeuralStudio to deliberate on what constitutes optimal decision planes and optimal decision functions.

The term Artificial Intelligence (AI) refers to a group of computerized methods that emerged in the 1950s and use non-conventional techniques to solve practical problems in

Engineering that were otherwise impossible to solve with conventional mathematics [Fogel et al. 1966] [Rich and Knight 1991] [Winston 1992]. Advances in AI are a direct consequence of the invention of computers, and of a constant improvement in their performance as well as their massive use in education and research. Areas which experienced the highest progress and have to date made the most impact are genetic algorithms, fuzzy logic, and artificial neural networks.

In the 1950s, the first ideas of what would be later called the theory of genetic algorithms (GAs) [Barricelli 1954] [Barricelli 1957] [Fraser 1957] were published. Soon after that, the theory experienced a fast development and became very attractive due to its applicability to solve practical real life problems [Fraser 1970] [Rechenberg 1973]. The theory of GAs was considerably improved by the works of John Holland in the 1970s [Holland 1975] and it has become an important tool to perform system optimization [Goldberg 1988] [Goldberg 1989] [Fogel 2006].

Genetic algorithms are optimization algorithms that use heuristic search methods that manipulate the variables resembling concepts found in Genetics such as mutation, cross-over, and selection [Fogel 2000] [Langdon and Poli 2002] [Grosan and Abraham 2007] [Goldberg 2007]. Genetic algorithms are suitable for constrained optimization problems and are widely used in the automated design of industrial equipment, automotive components for crashworthiness, sophisticated trading systems in the financial sector, water distribution systems, distributed computer network topologies and electronic circuit, among others. They have been also applied to gene expression profiling analysis,

marketing analysis, container loading optimization, code-breaking and the search for hardware bugs.

In 1965, the foundations of the Fuzzy Logic (FL) theory were established in a paper by Lotfi Zadeh [Zadeh 1965]. In the beginnings, the theory was received with skepticism, but with time it attracted the attention of the scientific community as it showed its undisputable usability in solving many practical problems, especially real-time control problems. Zadeh showed that many complex real-world problems lack precise information and therefore do not need to be modeled by traditional system analysis techniques. Instead, FL could be used to obtain model results that would be as accurate as desired and at the same time, of great practical use. In the subsequent years, the FL theory was enriched and strongly associated with the solution to many control problems [Zadeh 1965] [Zadeh 1968] [Zadeh 1989] [Zadeh 1976] [Yen and Langary 1994]. FL has extended its applications to robotics, consumer products (vacuum cleaners, rice cookers, refrigerators), the automotive industry (power train and transmission control, engine control), Industrial process control (embedded control applications, manufacturing) and also aerospace. Other areas of application have been information processing, financial banking and trading, decision making, and pattern recognition [Klir and Yuan 1995] [Dimitrov and Korotkich 2002].

In the meantime, the creation of the perceptron by Rosenblatt in 1958 [Rosenblatt 1958] constituted the starting point of the development of the theory of ANNs [Minsky and Papert 1969] [Widrow and Lehr 1990], and thereafter, different types of networks were created, such as the ADALINE [Specht 1990] [Piche 1995] [Chan et al. 2003] [He et al.

2005] and MADALINE nets [Winter and Widrow 1988] [Widrow and Lehr 1990], as well as the self-organizing maps, also called Kohonen features maps by the name of its inventor, the Finnish professor Teuvo Kohonen [Ferrán et al. 1994] [Santini 1996] [Kohonen 2001]. Further advances led to the design of other types of nets, such as clustering nets [Pal et al. 1993], pattern association nets [Nooralahiyan et al. 1994] and fixed weight nets [Feldkamp et al. 1996], among others. Each of them was developed in an effort to model different types of real-world problems. What is so attractive about the perceptron and ANNs in general is that they resemble the way biological neurons interact to process information.

Many practical problems can be indistinctly solved either with FL, ANNs, and even GAs as in the case of control systems [Patrikar and Provence 1993] [Shibuchi et al. 1994] [Chen et al. 1995] [Chung and Chiang 1997] [Hagan et al. 2002] [Ayala et al. 2002a] [Ayala et al. 2002b] [Ayala and Adjouadi 2003] [Pal et al. 2003] [Pal et al. 2003]. A classic example is the traveling salesman problem which can be solved with GAs, FL and ANNs [Braun 1991] [Favata and Walker 1991] [Xu and Vukovich 1993] [Julstrom 1995] [Budinich 1996] [Li and Kwan 2003]. An example of how to set up a FL system using GAs is given in [Shibuchi et al. 1994]. Similarly, many problems of pattern recognition and classification can be solved with FL and ANNs [Kandel 1982] [Kwan and Cai 1994] [Lo et al 1995] [Ripley 1996] [Rowley et al 1998] [Baraldi and Blonda 1999] [Osowski and Linh 2001] [Bishop 2005] [Ajiboye and Weir 2005].

Beyond the countless industrial applications of AI algorithms, it is in the biomedical field where especially ANNs have clearly outperformed the other areas. Their applications have increased in proportion to the development of screening technologies such as

Magnetic Resonance Imaging (MRI), Computer Tomography (CT), Positron Emission Tomography (PET), Single Photon Emission Computed Tomography (SPECT) as well as with the extended usage of flow cytometry, to name a few. Besides segmentation of MRI images [Hall et al. 1992] [Reddick et al. 1997], such applications are mainly targeted to the recognition of patterns that are indicative of diseases, such as:

- Leukemia [Kothari et al. 1996] [Zong 2005] [Zong 2006]
- Cancer [Khan et al. 2001]
- Breast cancer [Wilding et al. 1994] [Chen et al. 2000]
- Brain tumors [Reddick et al. 1998],
- Pulmonary cancer and related diseases [Gurney and Swensen 1995] [Scott et al. 2000] [Eng 2002]
- Alzheimer [Horn et al. 1993] [Kippenhan et al. 1994] [Warkentin et al. 2004] [Tandona et al. 2006]
- Hepatitis [Ozyilmaz and Yildirim 2003]
- Heart diseases [Das et al. 2009]

Especially in epilepsy research, a focus of this dissertation, many AI applications deal with the automation of the detection of epileptiform activity (EFA) from Electroencephalograph (EEG) recordings. Epilepsy has become a critical area of research and has received great attention from medical and scientific institutions in the last two decades. This attention is due to the overwhelming number of persons suffering from that disease, which reportedly affects over 3 million people in the United States [Epilepsy Foundation of America 2008]. It is estimated that in about 80 percent of those diagnosed

with epilepsy, seizures can be controlled with some medicines. In the remaining cases, surgical intervention is mostly the only alternative, and the precise detection of seizure focus becomes the most critical part of the pre-surgical evaluation due to its sensibility to surgery outcome. Seizure focus localization is then performed by tracking EFA, which is done with different modalities in order to assure high accuracy. Despite being a relatively old technique, EEG is still the key tool in epilepsy treatment and research. Computer aided algorithms are used to automate focus localization process and make it free from subjectivity [Yaylali 1991] [Gabor 1992] [Tarassenko et al. 1998] [Adjouadi and Ayala 2003] [Subasi 2006].

This dissertation is devoted to the crucial role that ANN-based EFA detection methods play in epilepsy research. For that reason, the aforementioned NS programming tool, which was originally designed and developed to target a general scientific audience, will be presented. The tool was later improved over the years to be suitable for use in detecting EFA from EEG recordings. This dissertation will present two methods for detecting seizures and classifying EEG recordings, which have been designed and modeled with the NS tool.

Chapter 2 of this dissertation provides the research foundation of ANN-based algorithms for detecting EFA, with emphasis in the extraction of features from EEG in time and frequency domain. After a review of existing detection methods, the bases of the method used for performance evaluation are described.

Chapter 3 presents an overview of the design and functionality of the NS tool and will give special attention to the ability of designing and training ANNs to detect EFA.

In Chapter 4, the experimental setup that was used as a basis for all methods is presented. Additionally, a seizure detection method is discussed and presented, emphasizing its modeling with the NS tool. A performance evaluation analysis of the detector is made.

Chapter 5 proposes a new methodology to analyze feature trends in EEG files and compares non-epileptic with epileptic patients as an attempt to delineate the differences between them. This methodology is evaluated using Receiver Operating Characteristics (ROC) analysis within the NS tool.

Concluding remarks are provided in Chapter 6.

CHAPTER 2

2. PHYSIOLOGY AND DETECTION OF EPILEPTIFORM ACTIVITY

2.1. Processing Epileptiform Activity in EEG

2.1.1. Basics

EEG plays a decisive role in epilepsy research and treatment. EEG was first applied by Hans Berger in 1929 in Germany in an attempt to demonstrate that the measured activity changes depending on the brain status (epilepsy, anesthesia, sleep). In its basic representation, an EEG is a plot of the measurements of electric signals from the scalp as illustrated in Figure 2.1. The measurements, usually in microvolt, are taken from electrodes that are placed on the scalp and fixed to it with some type of material of high conductivity. One electrode always serves as ground or reference and the others are placed in specific locations on the scalp. Most EEG recordings include an electrode to record electrocardiogram (ECG) activity. This is done with the intent of using that signal to cancel its interference in the EEG [Tong et al. 2001].

EEG was initially plotted directly on paper-rolls (analog), but nowadays, EEG recordings are digitally recorded and displayed using computers, usually in a referential montage, and stored for subsequent processing and analysis.

Through the years, the number of electrodes has been standardized, with the 10-20 system (Figure 2.1) [Homan et al. 1987] and the Modified Combinatorial Nomenclature

(MCN) montage (Figure 2.2) [Gevins and Remonds 1987] [Adjouadi et al. 2004b] [Adjouadi et al. 2004c] [Cambier et al. 2001] being the most used ones.

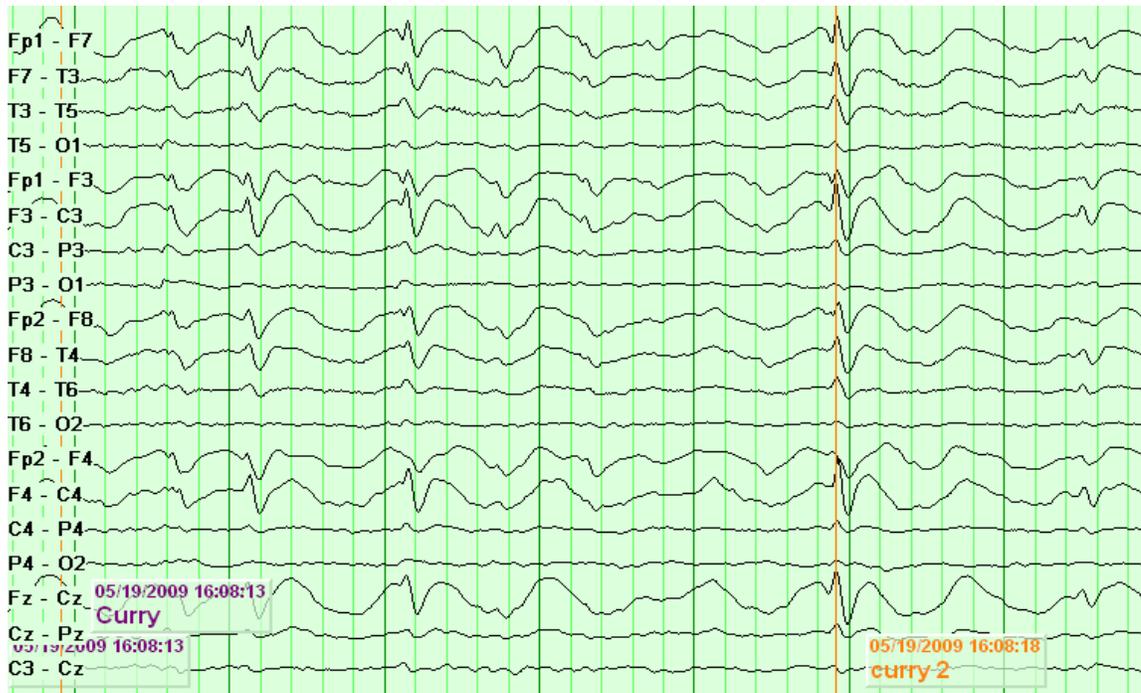


Figure 2.1: An EEG of a patient showing epileptiform activity (Interictal spikes) in some channel pairs (courtesy of Miami Children's Hospital)

The 10-20 system is however more practical for its simplicity of use and for the bigger spacing between electrodes which reduces the possibility of inter electrode interference. Since EEG records voltage differences between each electrode and a ground electrode, the plots thus represent a voltage difference between two electrodes. By simple transformations, the raw plot of one live electrode vs. ground (referential montage) can be always transformed into another representation to show the voltage difference of two specific live electrodes (bipolar montage). This allows personalizing the analysis without the need of repeating the recordings.

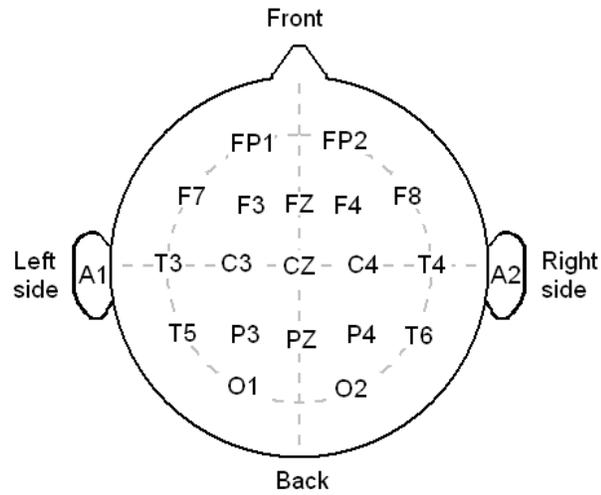


Figure 2.2: Location of the electrodes in the 10-20 system

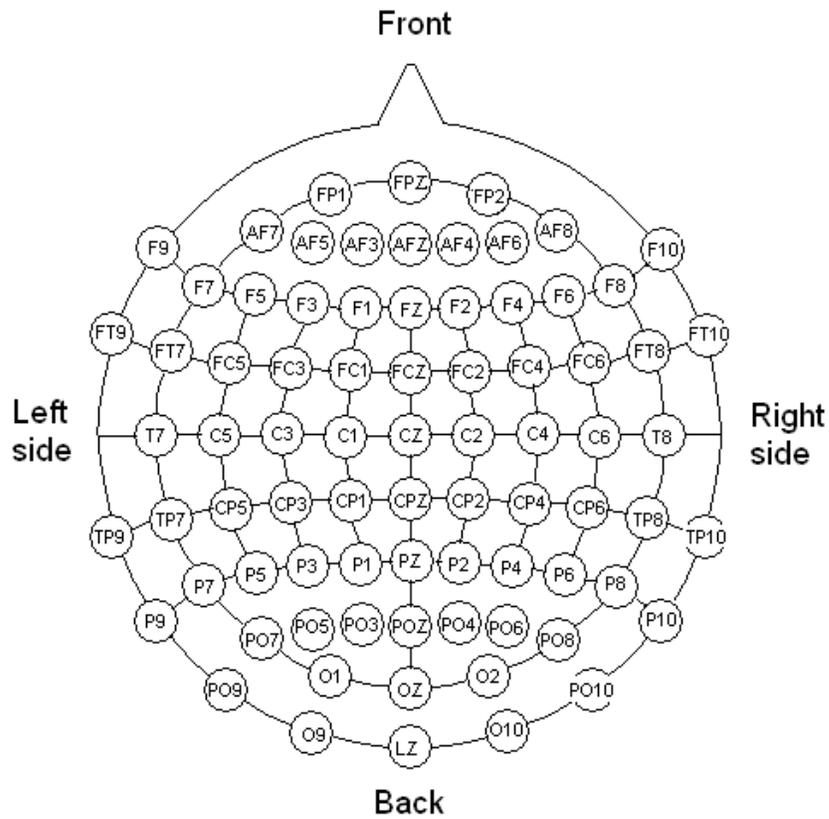


Figure 2.3: Location of the electrodes in the modified combinatorial nomenclature MCN

In the 10-20 system, each channel location has a label and/or number to identify its sub-cranial lobe and hemispherical location. For example, “FP” as the Front-Polar or prefrontal lobe, “F” as Frontal lobe, “T” as Temporal lobe, “C” as Central lobe, “P” as Parietal lobe, and “O” as Occipital lobe. There are two important reference landmarks known as nasion, which is in the front of the head (point between the forehead and nose) and inion which is located at the back of the skull. To simplify and standardize electrode placement, caps are used which have a maximum number of openings to place as much electrodes as needed (Figure 2.4).



Figure 2.4: An EEG cap with complete set (Image taken with the permission of a student volunteer from the CATE lab at FIU).

As opposed to MRI, CT, PET and SPECT, EEG is a relative inexpensive technology to estimate locations of tumors and epileptic foci in the brain. By reading EEG, experts can find specific patterns which are indicative of seizure activity. If these patterns consistently repeat in specific electrodes (or channels), then the approximate location of the seizure focus is expected to be in a region enclosing these electrodes. Therefore, EEG provides a simple way of roughly locating the seizure focus.

With time, procedures to study epilepsy have become more sophisticated and, when necessitated, invasive. For example, part of the pre-surgical evaluation of persons with

refractory seizures is the implantation of electrode arrays [Tran et al. 1997] which are placed in the cortex of the brain called Electrocorticography (ECoG). The placement of these arrays coincides with the location where the seizure focus was suspected by using scalp EEG. ECoG is considered in clinical practice the golden standard for locating epileptogenic zones due to its high spatial resolution and lower degrees of noise than the scalp EEG, whose recordings are attenuated due to high scalp inductivity [Ferree et al. 2001]. Figure 2.5 shows an electrode grid as it is implanted in the brain of a patient.

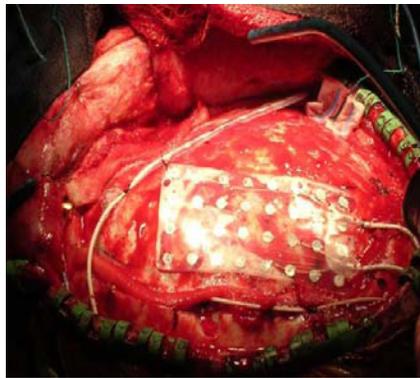


Figure 2.5: An electrode grid being implanted on the cortex of the brain of a patient (Courtesy of the Brain Institute at Miami Children's Hospital).

EEG is also used clinically to diagnosis coma and encephalopathy conditions as well as to monitor anesthesia, to name a few. Regardless of these merits of EEG, other technologies such as MRI are often used in parallel before surgical treatment of epileptic patients to improve surgical success.

Most EEG is nowadays digitally sampled between 256 and 512 Hz, although higher frequencies are used sometimes for research purposes. The recordings are then low-pass filtered to remove low-frequency artifacts (usually between 0.5 and 1 Hz) and high-pass

filtered to remove high-frequency artifacts (usually between 35 and 70 Hz). However, most EEG machines have configurable noise cancellation options already built-in.

As opposed to scalp EEG, ECoG is typically recorded at higher sampling rates since higher frequencies are better revealed in subdural signals.

2.1.2. Epilepsy: A Synopsis

Epilepsy is a chronic neurological disease manifested by abnormal electric discharges in the brain leading to seizures [Adams and Victor 1989]. It affects people of all ages, although it is predominantly a pediatric disorder, with the mean age of epilepsy onset defined in the range between 8 and 10 years. Epilepsy is considered as one of the most common neurological disorders affecting 3 million people in the United States alone. According to Centers for Disease Control and Prevention, one out of 100 adults has active epilepsy [CDC 2008]. According to the World Health Organization, it is estimated that 50 million people worldwide have epilepsy, especially children and adolescents, with millions more that go unreported in poor and developing countries.

It is estimated that 30% of the epileptic population has poor response to medication [Cascino 1994] [Engel 1996], and around 10% undergo surgical intervention [Epilepsy Foundation of America 2008].

Treatment options for intractable seizures are limited with some resorting to focal resections of abnormal brain tissue when the epileptogenic region can be accurately defined; a critical task that may require intracranial EEG recordings of seizures to define their onset and region of involvement. More recently, an alternative treatment option has evolved where chronic intracranial implants apply electrical stimuli directly to the brain

surface with the ultimate aim of preventing or aborting seizures. The implants continuously record the EEG activity and apply the stimuli when seizures are detected or are presumed to be imminent.

Although congenital factors, head traumas and vascular diseases are considered risk factors, the etiology of epilepsy is unknown for approximately three fourths of all cases [Hauser and Kurland 1975]. Patients with epilepsy are initially treated with anticonvulsant medication; but in difficult clinical cases, surgery becomes the only alternative for them.

2.1.3. Epileptiform Activity

According to [Chatrian 1994], epileptiform activity (EFA) is a term used in EEG to describe waves that are clearly distinguishable from the background activity and are similar to the waves found in EEG from epileptic subjects. EFA refers to the waves recorded in the interictal activity (the time between seizures) but not during the seizure itself.

According to [IFSCN 1974], EFA can be divided in spikes, sharp waves, spike-and-slow-wave complex, and multiple spike-and-slow-wave complexes. The distinctions are as follows:

- A Sharp wave is a transient distinguishable from EEG background which lasts 70 to 200 milliseconds
- A Spike is a sharp wave with a duration of 20 to 70 milliseconds

- A Spike-and-slow-wave complex is a spike followed by a slow wave, whereas the later has usually higher amplitude
- Multiple spike-and-slow-wave complex is a concatenation of spike-and-slow-wave complexes

In practice, however, it is more important to distinguish them from the background activity than to detect their morphological distinctions.

2.1.4. Interictal Spikes

Interictal spikes are spikes recorded in the time between seizures, while the subject is not having any seizures. Their detection is shown experimentally to be critical in locating the seizure focus. Interictal spikes usually occur in neighboring electrodes (spatio-temporal context) and at the same time, as depicted in Figure 2.6. Key aspects of the morphology of an interictal spike are illustrated in the RPF wave in Figure 2.7, consisting on raising and falling amplitude and duration. In [Gevins and Remond 1987], an excellent overview of the interictal spike morphology is presented. The conditions under which a spike can be considered an interictal spike are summarized in detail in [Adjouadi et al. 2005a].

Medical experts and neuroscientists [Jayakar et al. 1989] have established several criteria as necessary conditions to declare the existence of an interictal spike, and their detection and extraction were successful using algorithms that exploit such characteristics [Adjouadi et al. 2004a] [Adjouadi et al. 2005a].

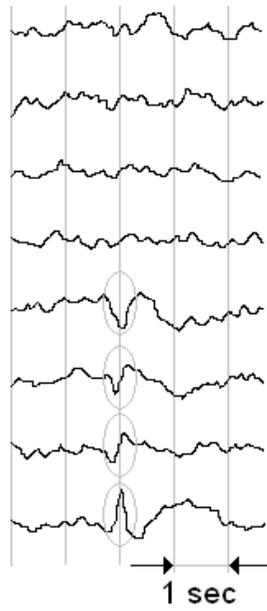


Figure 2.6: Interictal spikes occurring synchronously in four distinct electrodes or channels

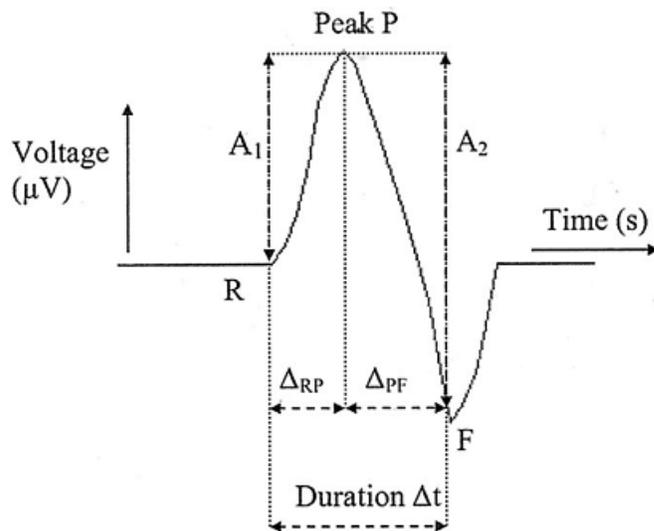


Figure 2.7: Characterizing features of an interictal spike

2.1.5. Seizures

Brain discharges lasting more than a few seconds usually represent ictal activity (called a seizure) rather than EFA. During a seizure, the EEG clearly shows high activity in most channels, often manifested with much higher frequencies and amplitudes as shown in Figure 2.8. The reader can clearly observe the increase in frequency and amplitude during the seizure or ictal state.

Usually, signs of the seizure are first observed in specific channels, called “channels that initiate the seizure”. In Figure 2.8, electrode pairs FP1-F7, F7-T3, T3-T5, F8-T4, and T4-T6 initiate the seizure. It has been recently found [Cabrerizo et al. 2009] that these channels are closer to seizure focus or tumor and have much lower coherence than the remaining channels.

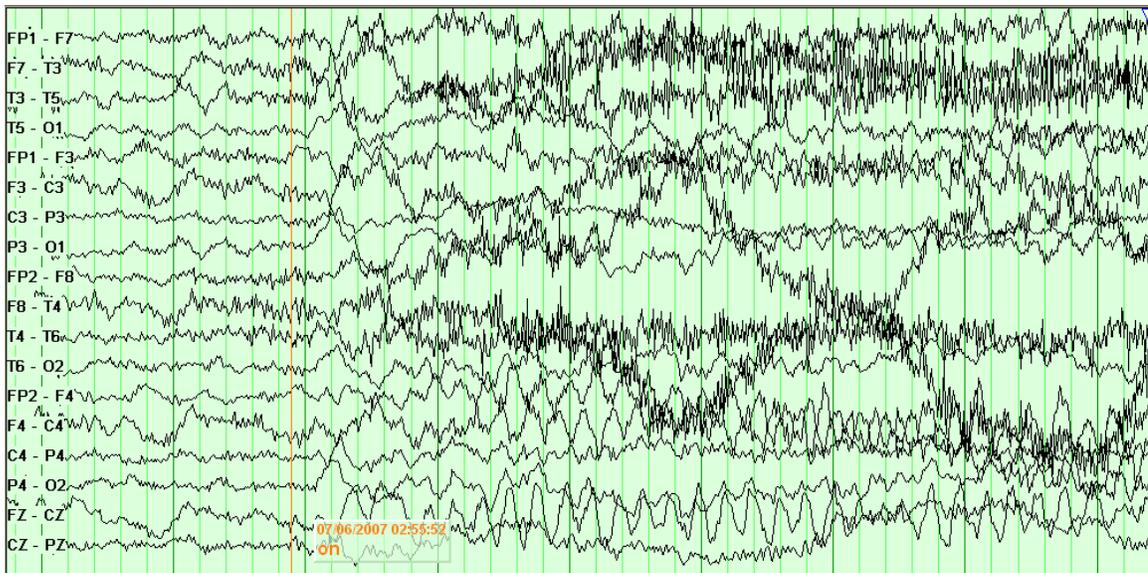


Figure 2.8: An EEG containing the transition from interictal to ictal state, showing electrode pairs that initiate the seizure (courtesy of Miami Children’s Hospital)

As explained earlier, during pre-surgical evaluation, an electrode grid is usually implanted in the brain of patients for a better localization of the seizure focus. Patients are then put in an observation room until it is deemed that sufficient data containing interictal spikes or seizures is collected. On occasions, this process may last for more than a week. Then, the patient is sent to the operation room for surgery. But before that, an EEG expert has to visually scan the EEG to find the interictal spikes and seizures, if any.

Considering that EEG recording sessions can last for days, reviewing such EEG data can be an exhausting process, besides that it is prone to subjective assessment. However, the detection of seizures during monitoring is crucial for the surgery outcome since the grid electrodes that initiate the seizure are consequently used to pinpoint the seizure focus more accurately. For that reason, methods for the automated detection of interictal spikes and seizures can serve as valuable tools for the scrutiny of EEG data in a more objective and computationally efficient manner.

2.1.6. Feature Extraction in Time Domain

EEG sampled at a specific rate would obviously yield the same amount of data points in each and every second. At 512 Hz, the prevalent frequency used in this dissertation, one second would thus be made up of 512 data points. Just as in any real-world problem, large amounts of data would increase the computational load that often lead to detection algorithms that are time consuming. However, in many practical cases, there is no need to evaluate each data point as in the specific case of interictal spike detection. In this case, only inflection points and their distance to each other will be sufficient to detect the

spike. In a more general sense, signals can be represented by a reduced number of so-called features which, depending on the case at hand, can help in identifying a specific pattern.

Since EEG is a time series, one can subdivide the recording in windows (let's say of one second) and then extract some statistical parameters (called features) out of each window. In doing so, each window can be represented in time as a set of features, and the change of these features over time can serve as means to forecast or identify a specific pattern or event of interest. Since the amount of features is often much less than the raw data used to compute them, this helps at reducing the computational load and simplifying the detection procedure.

Time-domain EEG features can be grouped in the following categories:

- Statistical features (average, standard deviation, moments)
- Hjorth's descriptors (activity, mobility, complexity)
- Non-linear features (correlation dimension, Lyapunov exponent)
- Other features obtained from convolution kernels, Principal Component Analysis (PCA), autocorrelation, and entropy

2.1.7. Statistical Features

Results from empirical evaluations indicated that the easiest way to assess the merit of the aforementioned features is to make use of the following statistical parameters:

- Minimum value in the window
- Maximum value in the window

- Spread (difference Maximum-Minimum) in the window
- Average value in the window
- Standard deviation of all values in the window
- Signal-to-noise ratio (average to standard deviation ratio)
- Skewness: Is the 3rd standardized moment and is computed as

$$y = \frac{\sum (x_i - \bar{x})^3}{\sigma^3} \quad (2.1)$$

- Kurtosis: Is the 4th moment and is computed as

$$k = \frac{\sum (x_i - \bar{x})^4}{\sigma^4} \quad (2.2)$$

Most of the related studies use the signal-to-noise ratio because as it combines the ratio of these two critical signals. Often, only a combination of such features could produce desired outcomes that can be validated. For example, in [Adjouadi et al. 2005b], a method is described, which extracts average, standard deviation, kurtosis and skewness from the raw data from blood samples and applies Support Vector Machines [Vapnik 1995] [Cristianini and Taylor 2000] to detect leukemia. The raw data of that study was so huge that it was impossible to be handled without feature extraction.

2.1.8. Hjorth's Descriptors

Neurologists to this day have relied on the use of the so-called Hjorth's descriptors [Hjorth 1970] in assessing EEG activity. These parameters are widely used and have been specifically designed to describe EEG patterns.

Hjorth's parameters are described as follows:

- Activity which defines the variance of the signal in the window
- Mobility which is the 1st derivative of the STD of the signal in the window
- Complexity which is the 2nd derivative of the STD of the signal in the window

Although the use of the Hjorth's parameters is a common practice, researchers have created additional, less conventional features to assess brain activity using nonlinear features as described next.

2.1.9. Non-Linear Features

Non-linear features are relatively new to EEG analysis and are still in a development state, given by the constant variation and increased sophistication of methods that have been lately developed. Some of the most important non-linear parameters are described below:

- **Correlation dimension**: Important studies on epilepsy [Pritchard 1992] [Theiler 1995] [Martinerie et al. 1998] [Lehnertz and Elger 1998] [Stam 2005] have applied chaos theory [Alligood 1997] to analyze EEG signals. A fundamental measure in chaos is the correlation dimension (also used in Fractal theory), which characterizes how chaotic a system is.

The correlation dimension has been extensively used to describe chaos in EEG in different studies, such as in [Frank G et al. 1990] [Iasemidis and Sackellares 1996] and [Guevara 1997]. In [Grassberger and Procaccia 1983], a numerical technique was introduced to compute an estimate of the correlation dimension, however, it is widely

accepted by the scientific community that the numerical techniques for computing the correlation dimension are not easy to implement.

The correlation dimension is computed through the correlation integral $C(\varepsilon)$:

$$C(\varepsilon) = \lim_{N \rightarrow \infty} \frac{N_V}{N^2} \quad (2.3)$$

where N_V is called correlation sum and represents the total number of reconstructed vectors $\underline{V}(i)$ in a data segment whose difference is less than a small distance ε . The reconstructed vectors are obtained from the raw data with some arrangement of equidistant data points x_i such that $\underline{V}(i) = [x_1(i), x_2(i), \dots, x_m(i)]$. When the number of points tends to infinity and N_V to zero, the correlation integral $C(\varepsilon)$ converges to ε^v . Power v is called the correlation dimension.

Numerical techniques to compute the correlation dimension are difficult to implement. However, when comparing EEG segments of constant N , $C(\varepsilon)$ becomes only dependent of N_V . Therefore, the correlation sum may represent an alternative to look for changes in the correlation dimension of signals.

In an effort to overcome computational load and improve the accuracy, [Tito and Adjouadi 2007] used the correlation sum to detect seizures in EEG segments.

- **Lyapunov exponent**: Another significant non-linear parameter is the Lyapunov exponent, which measures how fast signal trajectories separate from each other. The Lyapunov exponent has been used to some extent in EEG analysis [Iasemidis and Sackellares 1991] [Pradhan and Sadasivan 1996] [Osowski et al. 2007] and even in

seizure prediction [Moser et al. 1999] [Isaemidis et al. 1998] [Isaemidis and Sackellares 2001]. However, the use of the Lyapunov exponent for the particular case of seizure prediction remains controversial to this day [Lai et al. 2003] [Lai et al. 2004]. In the challenging area of seizure prediction, our research group in the CATE center at FIU is trying to consolidate key findings reported in the literature and augment the current knowledge through an extensive investigation on several patients that have been monitored under pre-surgical evaluations combining both intracranial EEG (iEEG) and MRI modalities. Our preliminary results suggest the following: (1) that the coherence behavior in frequencies from 30 to 60 Hz is suitable to classify electrodes that initiate the seizure; and at the time of the seizure there is a significant decrement of the coherence values between frequencies from 30 to 60 Hz; and those electrodes that initiate the seizure keep their low coherence values until at least 10 minutes before a seizure onset. (2) The standard deviation of the power of the frequency bands, calculated in epochs of 8 seconds, using files of 30 to 40 minutes of duration before and during a seizure varies across frequencies. There is an evident decrement in the standard deviation of the power from Delta to Gamma prior to an ictal state. (3) Using orthogonal transformations of the iEEG activities, there is a significant difference of the standard deviation for those electrodes that initiate a seizure in contrast to those that do not, and that a few seconds prior to an ictal state, interlocking between all electrodes is experienced.

2.1.10. Other Features

Apart from the aforementioned features, there are several types of parameters that are also increasingly applied to EEG analysis. Some of them are listed below:

- **Convolution kernels**: Convolution kernel are operators that are passed through the data sample by sample, usually producing one single output, as it is the case of the derivatives and filters. In [Adjouadi et al. 2005a], the Walsh transform (which can be regarded as a configurable derivative) is used as a convolution kernel for interictal spike detection from EEG recordings of 31 patients with focal epilepsy, yielding a precision of 92%.
- **PCA Analysis**: Principal Component Analysis is an important de-correlation technique that processes multidimensional data (EEG channels) aiming at finding uncorrelated variables or dimensions in another space so as to reduce data covariance. PCA involves matrix operations known as eigenvalue or singular value decomposition. Closely related to PCA are independent component analysis techniques, which are used to separate mixed variables into independent ones. The use of this technique is extensive and is described in detail in several books, with an excellent explanation in [Jolliffe 2002]. [Cabrerizo et al. 2006a] use PCA in conjunction with power spectrum to validate functional brain mappings associated with auditory/comprehension tasks. [Jung et al. 1997] use PCA on EEG power spectrum to train an ANN to estimate levels of alertness in volunteers performing auditory tasks. [You et al. 2008] use PCA to automatically classify Functional Magnetic Resonance Imaging (fMRI) patterns. Since EEG is recorded simultaneously

at different channels, this technique has major applications since it can find the degree of correlation between channel signals. The technique has been used in separating artifacts from channel recordings [Barlett et al. 1995] [Congedo et al. 2008], to approximately locate focus of temporal brain activity [Gomez-Herrero et al. 2008], and to automate seizure onset detection [De Vos et al. 2007] [Hesse and James 2005], among other applications.

- **Autocorrelation**: Algorithms that apply the autocorrelation function to EEG analysis are moderately reported in the literature. In [Creutzfeldt et al. 1985], it is shown that the autocorrelation produces different outcomes when segmenting normal and abnormal EEG patterns. [Tito and Adjouadi 2007] have also used it to automatically detect seizures.
- **Entropy**: There are several definitions of entropy, and some of them differ in formulation. Entropy in information theory measures the degree of uncertainty in systems. Therefore, it is used to model the chaotic nature of EEG. In EEG studies, the Shannon entropy [Shannon 1948] is applied in most cases. The Shannon entropy of a signal $y(t)$ is defined as:

$$E = -K \sum_{t=1}^N y(t) \log_2(y(t)) \quad (2.4)$$

where K is some constant. [Steuer et al. 2004] apply the Shannon entropy to determine activity changes in long term EEG monitoring. Entropy can also be used in conjunction with the power spectrum (Spectral entropy) [Anderson and Jakobsson 2004] during anesthetic induction.

The Kolmogorov-Sinai entropy is yet another parameter that allows classification of signals by their information contents. In [Micheloyannis et al. 1998], this version of entropy is successfully used in characterizing EEG changes in volunteers performing different mental tasks. [Pezard et al. 1994] have shown that this parameter can detect variations in the EEG of individuals performing different tasks (closed eyes at rest, closed eyes and counting even numbers, staring at a spotlight, passive and active auditive odd-ball tasks).

The so-called approximated entropy is a derivation of the Kolmogorov-Sinai entropy and it has been applied to investigate how the level of anaesthesia in patients is reflected in their EEG activity [Bruhn et al. 2000]. [Abásolo et al. 2005] use it to compare EEG background of subjects with Alzheimer's disease and normal individuals in the same age range.

2.1.11. Feature Extraction in Frequency Domain

Another important domain from which features are usually extracted is the frequency domain. Since EEG is a digitized time series, it can be represented as a superimposition of signals of different frequencies. Therefore, patterns that are not visible in the time series may be revealed in the frequency domain.

Frequency related features are extracted from the frequency spectrum, as obtained using the Fourier transform. In the case of a time series such as EEG, it is always better to divide the time in windows and then extract features from the spectrum of each particular window, rather than computing the spectrum of the whole time series.

A windowed Fourier transform done in time is called short-time Fourier transform and is one way to increase time resolution in frequency analysis of time series. The spectrum of each window is computed by the Fourier transform and generates as many coefficients as there are data points in the window, whereas for one-second windows, only half of the coefficients may be used due to the Nyquist criterion [Shannon 1949]. It is common practice to represent the spectrum in frequency bins rather than Fourier coefficients, where frequency bins and Fourier coefficients are related by:

$$\frac{N}{F_s} = \frac{k}{f} \quad (2.5)$$

where N is the number of Fourier coefficients (equal to the samples in the window), F_s is the sampling rate, and f is the frequency bin that corresponds to the k^{th} Fourier coefficient. However, due to the Nyquist criterion, the maximum frequency analyzed may never exceed half of the sampling rate ($F_s/2$).

2.1.12. Power Spectrum and Relevant Frequency Bands

The most common parameters extracted from EEG spectrums are: the power spectrum, the mean and peak frequency, spectral entropy, coherence, and spectral phase. They are listed below:

- Power spectrum: A measure of how much of the signal varies in a specific frequency range is given by the power spectrum. The power spectrum is computed the conventional way: by squaring all Fourier coefficients that fall in a specific frequency

range (f_1, f_2) , and adding them. Since the Fourier coefficients are complex numbers, squaring is substituted by the norm as follows:

$$P(f_1, f_2) = \sum_{f=f_1}^{f_2} |S(f)|^2 \quad (2.6)$$

where f_1 and f_2 are the lower and upper bounds of the desired frequency range and $S(f)$ is the power spectrum at frequency bin f .

In EEG analysis, the following established frequency ranges have been thoroughly investigated in this dissertation:

- **Delta (< 4 Hz):** These waves are normal in adults during sleep, although are the dominant rhythm in infants up to 1 year of age.
- **Theta (4 - 8 Hz):** Theta waves are strong during alertness and meditation and reflect the transition from sleep to wakefulness. They are normal in children up to 13 years old.
- **Alpha (8 - 13 Hz):** These waves are observed by closing the eyes and by relaxation.
- **Beta (13 - 36 H z):** Beta waves reflect active concentration and are dominant in adults who are alert and have their eyes open. This band has a relative large band and it has been subdivided in several sub-bands: low beta or SMR (13-15 Hz), midrange beta (15-18 Hz) and high beta (>18 Hz). A subdivision into beta I (13-20 Hz) and beta II (20-36 Hz) is also found in the literature.
- **Gamma (3 6 - 44 Hz):** Are present in almost all areas of the brain and it is associated with perception and consciousness. It is also sometimes defined as the

frequencies above 36 Hz. In [Niedermeyer 2004] it is defined between 36 and 70 Hz.

The use of power spectrum in EEG studies is almost a golden rule. A study found in [Adjouadi et al. 2004b] shows how the power spectrum obtained from scalp EEG of patients performing auditory/comprehension tests during pre-surgical evaluation is useful in mapping critical brain regions that control speech and language functions. In [Tito et al. 2007], a comparative study is performed on the behavior in time of the power spectrum in the aforementioned frequency bands prior to seizure onset. In this study, other linear and non-linear parameters are also compared as well. Algorithms using power spectrum to automatically detect seizures are described in [Murro 1991] [Hopfengartner et al. 2007] and [Friedman et al. 2008], to cite a few.

2.1.13. Spectral Features

Other useful features that have been investigated include:

- **Mean Spectral Frequency:** This feature is the average frequency of the spectrum and it can be computed as:

$$\bar{F} = \frac{\sum_{k=1}^N k |F(k)|^2}{\sum_{k=1}^N |F(k)|^2} \quad (2.7)$$

The mean frequency divides the spectrum in two parts of approximately equal areas. The summation in Eq. (2.7) starts intentionally at $k=1$ to avoid adding the DC-value at $k=0$, where N is the total number of Fourier coefficients or half of the sampling rate.

For example, in a study by [Jeong 2004] it has been found that EEG abnormalities of patients with Alzheimer's are characterized by slow mean spectral frequency.

- **Peak Spectral Frequency**: This feature is the frequency with the most impact in the spectrum and corresponds to the highest Fourier coefficient. As an example, a study on the variability of the peak alpha frequency involving epileptic and non-epileptic subjects is found in [Larsson and Kostov 2005].
- **Spectral entropy**: Entropy as formulated earlier in Eq. (2.4) is also applied to the power spectrum as a way of measuring how chaotic a signal is. In [Anderson and Jakobsson 2004], an interesting algorithm computes the entropy from power spectra to classify the level of anesthesia in patients.
- **Coherence**: The use of this parameter is still in its early states, and relevant findings using this rather important measure are not well documented in the scientific literature. Coherence measures how similar the spectra of two signals by evaluating the level or extent of synchronization between them. The coherence between two electrode signals x and y at a specific frequency f is computed as follows:

$$C_{xy}(f) = \frac{|F_{xy}(f)|^2}{F_{xx}(f) \cdot F_{yy}(f)} \quad (2.8)$$

where $F_{xy}(f)$ is the cross-spectrum of time-domain signals x and y at frequency f , and $F_{xx}(f)$ and $F_{yy}(f)$ are the auto-spectra of time domain signals x and y , respectively, at frequency f . When computing coherence for a specific frequency band, numerator and denominator of Eq. (2.8) are substituted by summation signs over all frequencies in the specific range. An excellent study that shows an evident change in the coherence

of electrode initiating the seizure can be found in [Cabrerizo et al. 2009]. A method that applies coherence to estimate the extent of seizures from EEG of patients performing specific tasks is proposed in [Towle et al. 2007]. An application of coherence to noise cancellation from EEG is found in [Wahlberg and Lantz 2002].

- **Spectral Phase**: Spectral phase is computed as the arctangent (*arctan*) between the imaginary and the real part of the Fourier coefficients. It can be used to detect dynamical changes in EEG. In a study involving 8 children with medical refractory seizures, [Ayala et al. 2009] report that the phase diagram of the signal of electrodes that initiate the seizure clearly differs from those of the remaining electrodes

Wavelets are yet another way of extracting features from EEG. They allow a representation of signals in both time and frequency. When an event in a signal is represented in the time-frequency plane, it will mark a region instead of a point. As opposed to conventional methods of frequency analysis (such as the short time Fourier transform), Wavelets allow for a low to fine resolution in the time and frequency scales. Similarly to the Fourier transform, wavelets generate coefficients which are used directly as features or for feature extraction. For example, in [Kiyimik and Subasi 2004], the wavelets coefficients are taken as inputs to an ANN to detect alertness from EEG. In [Li 2006], a method is proposed to automatically detect seizures from EEG recordings using wavelet spectral entropy.

There are many other methods that can be used in assessing EEG functionality which are less frequently found in the scientific literature, and which will not be described as they

are beyond the context of this dissertation. In retrospect, in all the methods that are based on the spectrum, frequencies are always analyzed up to $F_s/2$.

In relation to the window size, it should be also noted that most research studies divide the recordings into one-second windows, although some neurologists and neuroscientists tend to use up to 8-second windows in an attempt to assess EEG dynamics with added context.

2.2. Overview of Methods for the Automated Detection of EFA and Seizures

2.2.1. Basic Operational Steps of Existing Methods

The simplest detection of EFA and seizures is performed on EEG by searching for patterns that are similar to valid EFA or seizure patterns. All detection methods have in common the need to “learn” the conditions under which a pattern may be considered an EFA or a seizure. One set of rules use imitative means of representation (i.e, mimetic) on individual channels and the other set takes into account the spatio-temporal aspects of multichannel EFA. After the rule base has been extracted with some method, the algorithm is tested on unknown data to evaluate its performance.

The steps undertaken in the context of this dissertation can be generalized in the following way:

- 1) Data collection and pre-processing (including data partitioning for pattern extraction and later testing)
- 2) Feature extraction

- 3) Rule extraction (training)
- 4) Event detection/classification (testing)

For the detection of patterns, rules are implemented in the time domain. There are many ways of detecting such patterns: either by rule extraction or by using more sophisticated methods such as artificial neural networks (ANN).

The two most reliable characteristics for the detection of interictal spikes are their fast rise and decay and the peak sharpness. These characteristics have been assessed by simple rule-based methods in [Birkemeier et al. 1978] [Gevins and Remond 1987] [Glover et al. 1989] [Gotman 1985]. More accurate methods include the spatio-temporal aspect of the interictal spikes, which have been assessed in several studies, such as in [Dingle et al. 1993] [Jayakar et al. 1991]. In [Gotman and Wang 1992], a detection algorithm is implemented based on the subject state (sleep, wakefulness, etc).

The patterns are usually extracted mainly with some specific technique. For example, an implementation of the Walsh transform to spike detection is provided by [Adjouadi et al. 2004a]. A detection method based on Wavelet theory is given in [Calvagno et al. 2000] and [Popescu 1998]. In [Barreto et al. 1993], a study is described for the detection of interictal spikes using electrocorticographs and Lagrange derivatives.

Methods implementing ANNs have been proposed by [Subasia and Ercelebi 2005] [Ayala et al. 2004] [Hellmann 1999] [Ko and Chung 2000] [Kurth et al. 2000] [Tarassenko et al. 1998] [Gabor and Seyal 1992] and [Acir et al. 2005], to cite a few. [Gulera et al. 2005] proposes an interesting EEG signal classification and seizure prediction algorithm that uses Lyapunov exponents and ANNs.

Also in [Ubeyli and Guler 2004], ANNs are feed with Lyapunov exponents in order to assess EEG changes in epileptic patients. Expert systems have also been applied in [Davey et al. 1989] and [Dingle et al., 1993]. It is also common to combine several algorithms in detecting dynamical changes in EEG activity as exemplified in [Jerger et al. 2001].

When comparing detection or classification algorithms in terms of performance, some sort of criterion is needed in order to select the best algorithm. Testing such algorithms produces different types of errors, which makes any comparison difficult. Next section will describe how the algorithms are evaluated in terms of performance.

2.2.2. Performance Evaluation of Detection Algorithms

Evaluation of the Performance of detection algorithms is usually performed with Receiver Operating Characteristics (ROC) Analysis [Kohavi et al. 1998] [Tilbury et al. 2000]. Such analysis begins by establishing a confusion matrix which contains information about the actual classification of the data being tested and the outcome of the classification system. Table 2.1 shows the main entries of the confusion matrix for a two-class classifier.

Table 2.1: Entries of a confusion matrix

		Detected as	
		Negative	Positive
Actual	Negative	TN	FP
	Positive	FN	TP

The four table entries are defined as follows: TP (true positives) is the number of correct classifications that an instance is positive; FN (false negatives) is the number of incorrect classifications that an instance is negative; FP (false positives) is the number of incorrect classifications that an instance is positive; and TN (true negatives) is the number of correct classifications that an instance is negative.

Positive and negative refers to the outcome given by the classifier, whereas true and false refers to the correctness of this outcome (i.e. right or wrong with respect to the actual state of the patient). The sum of the first and second row is the total number of positive and negative instances being under test, respectively, whereas instances are just all data values to be classified, regardless of their class. Similarly, the sum of the first and second column is the total number of positive and negative detections by the system, respectively. The grand total is the total number of instances being classified.

Rows and columns summarize to the following:

- N_{neg} is the total number of negative instances: $N_{neg} = TN + FP$ (2.9)

- N_{pos} is the total number of positive instances: $N_{pos} = TP + FN$ (2.10)

- C_{neg} is the total number of negative classifications: $C_{neg} = TN + FN$ (2.11)

- C_{pos} is the total number of positive classifications: $C_{pos} = TP + FP$ (2.12)

- N_{tot} is the total number of instances being detected:

$$N_{tot} = N_{neg} + N_{pos} = C_{neg} + C_{pos} = TP + TN + FP + FN \quad (2.13)$$

The following quantities can be extracted from the confusion matrixes:

- Correct Classification Rate (CCR, also called accuracy): Is the proportion of all correct classifications to the total number of instances:

$$CCR = \frac{TP + TN}{N_{tot}} = \frac{TP + TN}{TP + TN + FP + FN} \quad (2.14)$$

- Misclassification Rate (MCR): Is the proportion of all incorrect classifications to the total number of instances:

$$MCR = \frac{FP + FN}{N_{tot}} = \frac{FP + FN}{TP + TN + FP + FN} = 1 - CCR \quad (2.15)$$

- True Positive Fraction (also known as Sensitivity, hit rate or recall): Is the proportion of TP to the total number of positive instances:

$$TP_f = \frac{TP}{N_{pos}} = \frac{TP}{TP + FN} \quad (2.16)$$

- True Negative Fraction (also known as Specificity): Is the proportion of TN to the total number of negative instances:

$$TN_f = \frac{TN}{N_{neg}} = \frac{TN}{TN + FP} \quad (2.17)$$

- True Positive Rate (also known as Precision): Is the proportion of TP to the number of positive classifications:

$$TP_r = \frac{TP}{C_{pos}} = \frac{TP}{TP + FP} \quad (2.18)$$

- True Negative Rate: Is the proportion of TN to the number of negative classifications:

$$TN_r = \frac{TN}{C_{neg}} = \frac{TN}{TN + FN} \quad (2.19)$$

- False Negative Rate: Is the proportion of FN to the number of positive instances:

$$FN_r = \frac{FN}{N_{pos}} = \frac{FN}{TP + FN} = 1 - TP_f \quad (2.20)$$

- False Positive Rate (also called False Alarm Rate): Is the proportion of FP to the number of positive instances:

$$FP_r = \frac{FP}{N_{neg}} = \frac{FP}{TN + FP} \quad (2.21)$$

The reader should note that with FN_r and FP_r , low values are better than high values.

When two classifiers are being compared and the results of only one test are available, accuracy plays the most important role, unless one wants to favor TP or TN, which is common in the medical field. In these cases, sensitivity or specificity should be used instead.

Varying a classifier threshold can have contradictory effects. Increasing the TP rate can also increase the FP rate, which is an undesired collateral effect. This trade-off is best given by the so-called ROC curves [Marcum 1960], which are parametric curves that are constructed based on the values of the TP and FP rate. In Figure 2.9, two ROC curves from two different classifiers are plotted. Observe that the classifiers excel each other in different regions of the plot.

Depending on the problem at hand, any particular ROC measure can be chosen in favor of another. For example, if one prefers to maximize TP and minimize FN, then TP_f is the best choice. Sometimes, the compromise is made by using the area under the ROC curve, and the classifier with the highest area under the ROC curve is said to be the best. Depending on the problem at hand, this dissertation will select the appropriate measure for algorithm evaluation.

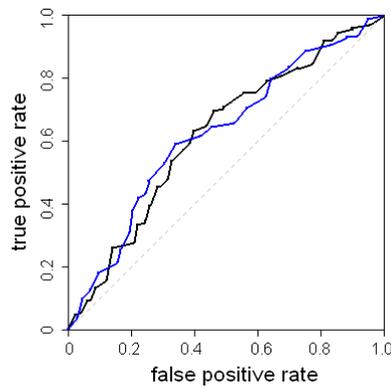


Figure 2.9: Two ROC curves plotted from the confusion matrixes of two classifiers, showing how erratic their performance can be.

As mentioned earlier, the methods for feature extraction can be combined to produce more robust classifiers. Once features are extracted, the algorithms used to create the classifier can differ from each other. For example, once the power spectrum of an EEG is extracted, it can be used to feed a Fuzzy system, an expert system, an ANN, or even a combination of them in order to train a classifier to detect EFA or seizures.

For the reasons given in Chapter 1, this dissertation chooses ANNs to create the classifiers. Given the difficulty in designing optimal ANNs, a computational tool named

NeuralStudio has been created in the context of this dissertation to support researchers in the design and training of ANNs. Therefore, the methods that are described in the next chapters are implemented and tested using NeuralStudio. As such, the next chapter will provide a detailed overview of this newly created tool.

CHAPTER 3

3. DESIGN AND IMPLEMENTATION OF NEURALSTUDIO: AN ARTIFICIAL NEURAL NETWORKS DESIGN AND EVALUATION TOOL

3.1. Introduction

Recent advances in artificial intelligence have benefited from the problem solving capabilities of ANNS in such areas as time series prediction, pattern recognition, data classification, and decision theory [Widrow and Lehr 1990] [Specht 1990] [Patrikar and Provence 1993] [Joyce 1999]. Artificial neural networks have also been relied upon in applications such as forecast studies, object identification, and process control, to name a few [Hagan et al. 2002]. Indeed, the wide range of applications enabled by artificial neural networks has not been limited to a few specialized-context problems, but rather expanded into different industrial applications and in the development of various commercial products.

Artificial neural network theory can only be applied to practical problems with the use of computers. Various programming tools haven been developed and are available in the market. However, despite the wide use of ANNs, the tools available for designing artificial neural networks are in general limited either in functionality, user friendliness, or both. Efforts to provide user friendly tools may be complicated by aspects of artificial neural network theory, including the complex mathematics involved. Nonetheless, past programming tools often complicate the process further by requiring programming in proprietary script or other languages. The user must first master a programming language or script before designing and training ANNs. Such requirements and other non-user

friendly details may then obscure aspects and features of ANNs to the user in training, as well as frustrate implementation and use for more experienced users.

The data processing requirements of ANNs have likely been another source of complications for designing ANNs. More specifically, a considerable amount of data often needs to be processed to train an ANN. Efforts that would benefit from observation and analysis of the processing steps directed toward one network, or preferably many networks, may be impeded by difficulties arising from the creation, handling and processing of the training data. In fact, the inability to teach students with examples has limited the usefulness of current ANN software design solutions. Thus, the entry, handling and other processing of the data sets have acted as a barrier against effective teaching of ANN theory.

One widely used ANN programming tool is provided as a toolbox within the MATLAB software package available from The MathWorks, Inc. (Natick, MA, www.mathworks.com). Unfortunately, knowledge of MATLAB's script language is generally required in order to access the full suite of programming options and features of the toolbox. Making matters worse, the tool user is forced to enter the script language instructions via a command line. Thus, the design and other programming of ANNs are at times inconvenient and slow, even when the scripting language may be familiar to the user.

The MATLAB toolbox also provides a network/data manager to support the implementation of certain programming tasks outside of the command line. Unfortunately, the network/data manager does not present or support all of the

functionality available via the toolbox, thereby forcing the user to utilize the command line at times. As a result, the network/data manager is primarily useful as a preliminary interface for users designing networks and data sets of relatively low complexity.

More generally, the ANN programming tools commercially available for use in research, industry or educational settings often fail to provide comprehensive coverage of the ANN field in the sense that, for instance, not all network types are supported or, for those types that are supported; the designs are limited due to the absence of design options, training features, etc. For example, ALNfit Pro, a software tool available from Dendronic Decisions Ltd. (Edmonton, Alberta, www.dendronic.com), generally supports one network type, an adaptive logic network (ALN), that utilizes a single type of multilayer perceptron, or feedforward network, for application only to Boolean function-based computations. Moreover, the graphical user interface generated by ALNfit Pro provides a limited number of options for configuration and training. The programming interface provided by Attrasoft, Inc. (Savannah, GA, attrasoft.com), and its Attrasoft Boltzmann Machines (ABM) software, is similarly limited, insofar as the software supports only two network types, the Hopfield Model and the Boltzmann Model.

3.2. Significance of the NeuralStudio Design

From educational and scientific points of view, currently available ANN tools present issues that prevent them from being widely used. There are also problems that are related to their accessibility.

A summary of the issues is as follows:

- There are no tools for unlicensed use in academia,
- There are not enough academically oriented tools albeit licensed that are of value to the student,
- There is lack of information about the features and the use of the tools (users guide being difficult to read, and lacking concise descriptions of desirable features),
- Some tools are not designed to run under the Windows operating system (often opting for UNIX, LINUX or other platforms),
- Some tools are only suited for specific areas of study (e.g., economics, business), and
- No open source tools have been found to the best knowledge of this dissertation author.

On the other hand, an assessment of desired design and operational characteristics for an educational tool could be summarized as follows:

- Convenient and user-friendly interface: The windows should allow users to interact with the system primarily through mouse clicks, making keyboard inputs only occasionally.
- Removed requirement for programming: This feature is necessary to remove the constraints imposed by powerful tools such as MATLAB, which provides a neural networks toolbox, but still requires knowledge on a pseudo-language based on C.

- Open source: Easy to understand source code of key calculation processes must be made available from within the application.
- Demonstration modules: Examples on how to use the programming tool must be given such as to alleviate the difficulties encountered in the mathematical details or in terms of implementation.
- Freeware: Enhanced accessibility for educational institutions must be provided.
- Compatibility with Windows operating system: The software designed needs to run under the Windows operating system, as it is widely used in academia.
- Stopping options: Options to stop the calculations at specified points of the algorithms need to be included (to allow viewing intermediate results to test for convergence issues or for additional debugging opportunities).
- Stand-alone application: The stand-alone application ensures that the tool box is not dependent of another application.
- Configurable display: Configurable display capabilities (graphics, charts, plots, and options to turn the software into a high-speed calculation tool) should be made available.

Seeking to realize these aforementioned design and operational characteristics as described above in an integrated teaching and research tool for ANNs was precisely the motivation for developing NeuralStudio (NS). This development constitutes one the main contribution of this dissertation. This integrated programming tool is proposed to be used by researchers and academia due to simplicity of use and the characteristics that make it an excellent platform to learn and apply ANNs to real-world problems. In this dissertation, the applications of this tool will focus on epilepsy research, as it will be

shown in chapters 4 and 5. The focus on this research topic is intentional, given the strong and ongoing collaboration with Miami Children's Hospital (MCH) through the joint FIU-MCH Neuro-engineering program. In subsequent sections, the tool will be described, giving special emphasis to its enhanced features.

3.3. General Description and Functionality

NS was primarily developed for use in postgraduate teaching, but with time, further improvements were made especially in speeding up calculations providing different types of graphical outputs, and analyzing and pre-processing data. In retrospect, the programming platform that was chosen for developing the application was Borland Delphi [Pacheco et al. 2001] which uses object PASCAL [Rachele 2000], a code easy to understand. The Delphi language was deemed appropriate for this work because its graphical user interface provides a large library of efficient components for easily creating charts and other graphical outputs with short and simple programming code. The compiled Delphi code is also fast and computationally efficient under the proposed design configuration.

NS includes network configuration modules to provide display interface to support configuration of different types of ANNs, such as feedforward nets, self-organizing maps, clustering nets, pattern association nets, and fixed weight nets. Additionally, a pattern data module is included to provide a display interface to support establishment and modification of pattern data sets for training and testing the ANNs.

Support modules for the simulation of ANNs are provided, consisting on an activation function editor and a performance evaluation module. Help and wizard modules are also provided, as well as a module to simulate Support Vector Machines. The latter has been included with the purpose of comparing its efficiency with the ANNs designed and implemented.

NS's main window is illustrated in Figure 3.1 and consists of an editor for a multilayer network, information panels, and a table for network input and a corresponding results table. NS users are able to freely design the network, and to customize the neurons as well as their interconnections. Interconnections can be established by drag and drop operations between neurons. Neurons are represented with circles and interconnections are represented by lines drawn between pairs of neurons.

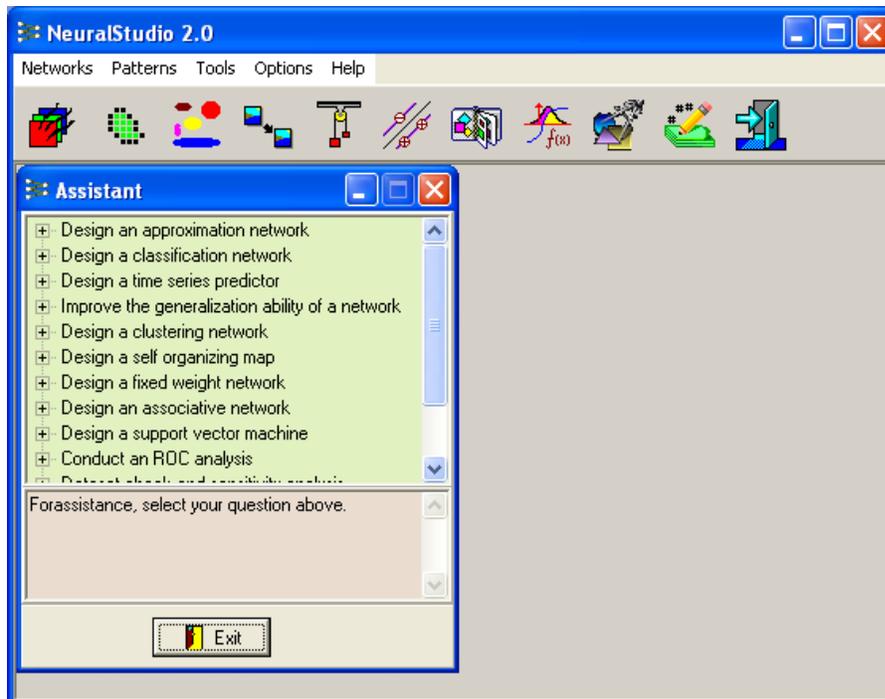


Figure 3.1: NS's main window, displaying a main menu and tool bar that gives access to all network designing tools as well as to the patterns editor, the activation functions, the performance evaluation module, and an option and help menu, among others.

The natural flow of a working session in NS is straightforward: The NS user can open the module of the particular network to design its topology and configure neurons and weights. Once done, the user can enter the input data that will be used for training. Because there are several options to facilitate data entry, there is a specific module called “patterns module” for that particular purpose, which is accessible from the network design module.

After all data has been entered, the user can exit the module and start the training. For backpropagation networks, training is performed in a module different from the design module. For the rest of the networks, both design and training are in the same module.

Once training is concluded, the user can go back to the pattern module to test the trained network with unknown data, which is data that was eventually entered as testing data in that module.

Another module of singular importance is the performance evaluation module. This module can be used to evaluate a network that has been tested several times on different data, particularly by means of ROC analysis [Kohavi et al. 1998] [Tilbury et al. 2000].

The main window of NS contains a menu for accessing the rest of the NS modules, having the structure described below:

- **Networks**: This option provides access to the different types of networks, which are the following:
 - Feedforward nets
 - Self-organizing maps

- Clustering nets
 - Pattern association nets
 - Fixed weight nets
 - Support vector machines (SVM), included for comparative purposes
- **Patterns**: This option opens the pattern module, where the user can edit the training and testing patterns.
- **Tools**: This command provides access to the following support modules:
- Activation functions: This command opens the activation functions module, where the activation functions can be chosen and customized.
 - Performance metric: This module provides tools to perform an ROC analysis on the networks.
 - Random number generator: This window allows defining the settings of the random number generator. The settings can be changed at any time during the working session with NS.
 - Average networks: This option allows loading a group of networks with similar topologies to compute their average weights and biases.
- **Options**: Available options through this command are:
- Number format: It allows defining the precision of the number outputs.
 - Code insight: It gives access to a module where the code of the most important algorithms used in NS is shown.

- **Help:** This command gives access to different help files:
 - Show Assistant: It opens a small assistant window that guides the user in the use of NS.
 - Contents: It loads and shows the NS help file.
 - Demo: It opens the demo module.
 - About: It shows NS copyright information.

Shortcut icons for the most frequently used commands are placed in a tool bar below the main menu.

3.4. Design of the Graphical User Interface

3.4.1. Data Entry and Preprocessing

Input data entry is done via a pattern module, which not only provides a data editor, but also includes a series of supporting tools to visualize and manipulate the data as well as to check it for inconsistencies, redundancies and other characteristics revealed in the so-called sensitivity analysis.

In this module, the data is entered in two different tables, the training set table and the testing set table, and it can be plotted. Figure 3.2 shows a snapshot of the module, where all data points are plotted on the right panel so as to provide a better insight into the complexity and overlap of the data. Common editing buttons in the toolbar are available for inserting, deleting and sorting data.

Sensitivity analysis as included in NS allows the data to be primarily audited against inconsistencies. A tab for this purpose (Figure 3.3) directs the user to a panel where a cross-analysis of the data set is performed, specifically row-wise as well as column-wise. The first option allows identifying duplicative or conflicting patterns in the data.

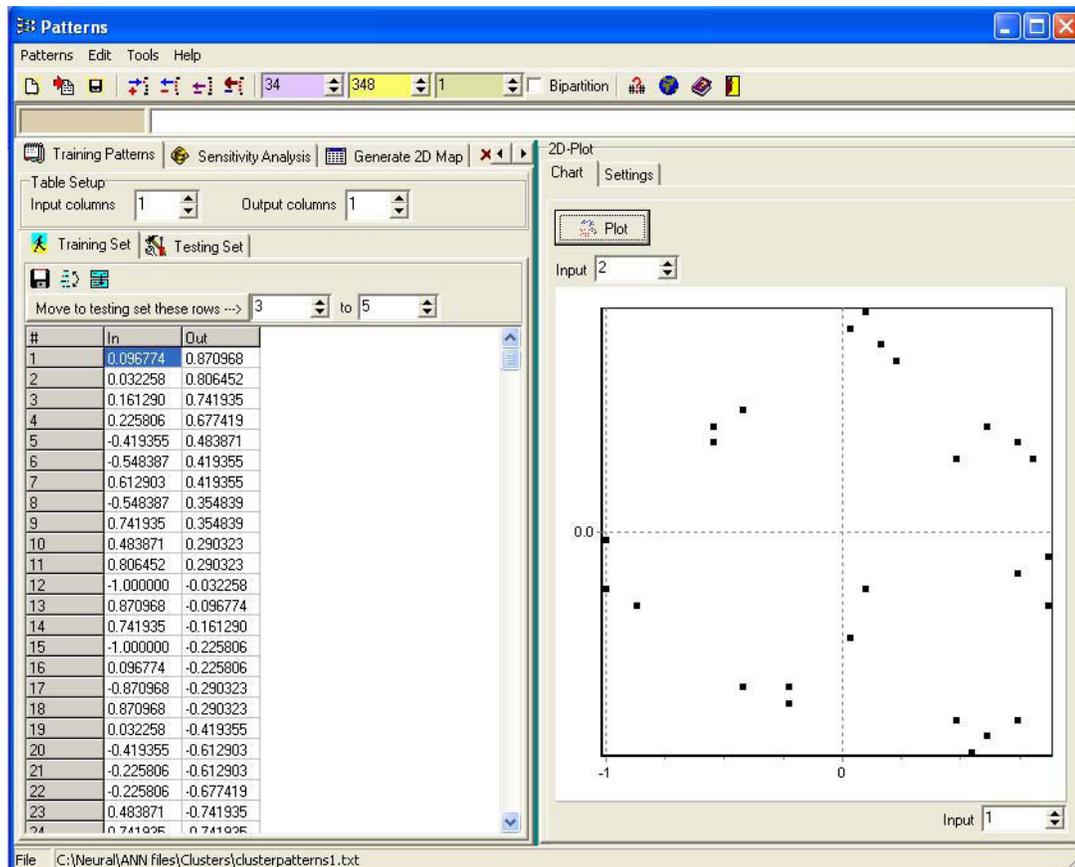


Figure 3.2: Pattern window with a particular data set entered and plotted. A toolbar facilitates data editing and also grants access to different supporting tools to preprocess the data. In this case, there is only one input and one output, i.e. two variables.

Although similar patterns are harmless and can be (but do not need to be) removed from the data set, contradictory patterns contain similar inputs but different outputs. Such patterns considerably affect the performance of the training and need to be detected and

removed from the data set prior to training. NS provides a fast and simple way of getting rid of such unwanted data.

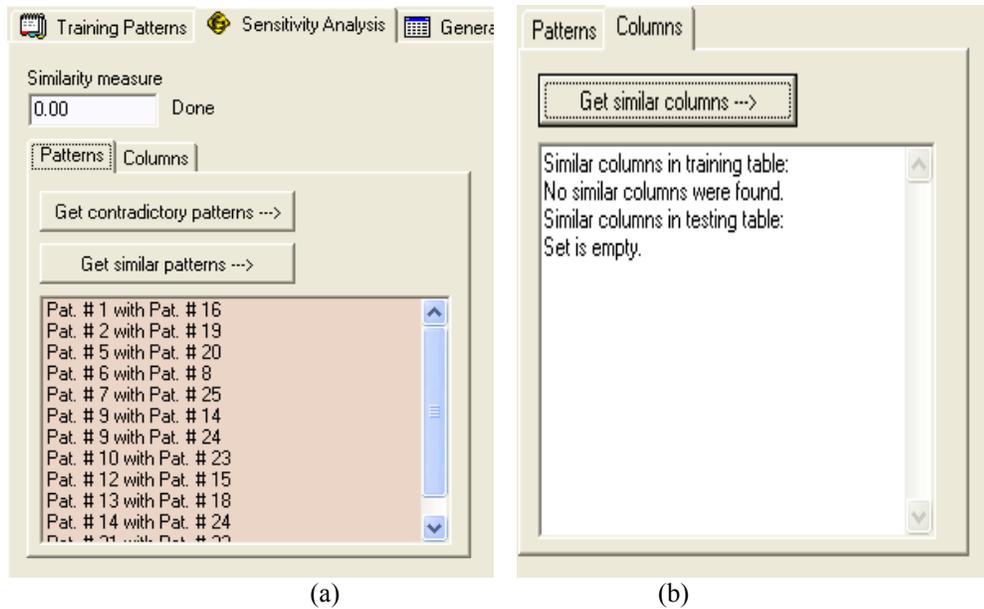


Figure 3.3: Sensitivity tools as given in NS provide an easy way to dispose unwanted data: (a) Contradictory patterns search option, giving a warning of all data that is contradictory. (b) Column-wise analysis, where all variables are checked for similarity.

As opposed to pattern checking, variable checking can be also performed in the second tab. Clicking on the “Get similar columns” button alerts the user that one particular variable is similar to another one and is therefore unnecessary. This option is crucial for cleaning up the dataset and reducing dimensionality.

The pattern module also provides a panel to generate 2D data by mouse pointer movements. A snapshot of this panel is depicted in Figure 3.4 and the option is of particular interest for teaching self-organizing maps.

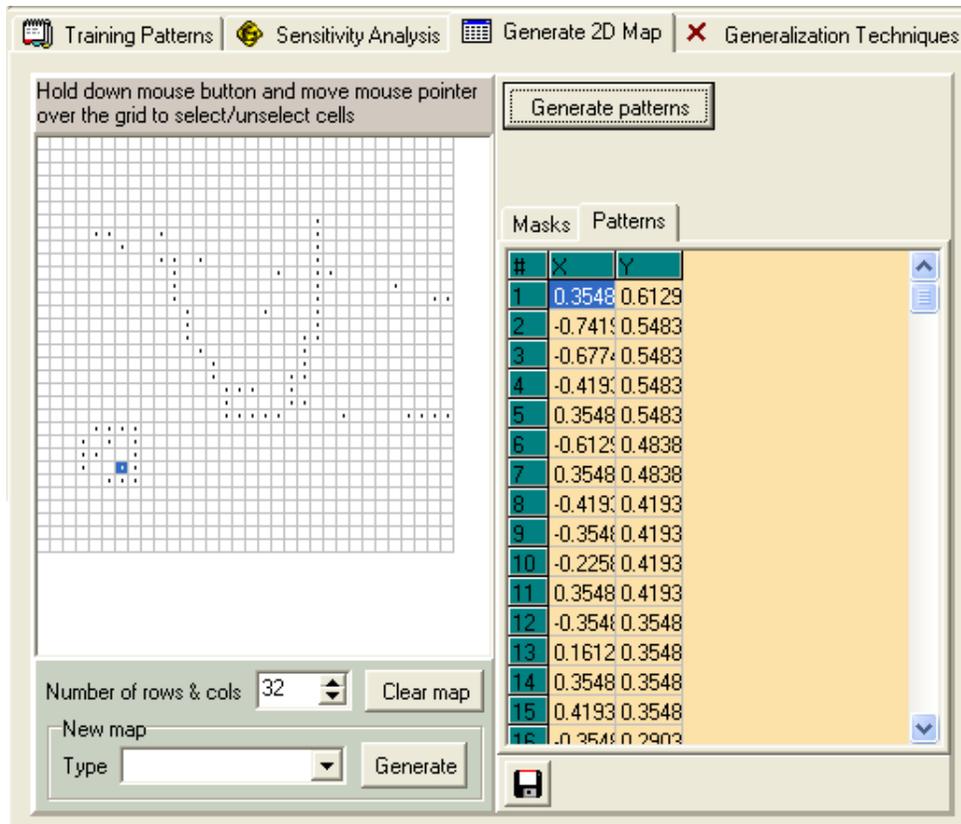


Figure 3.4: 2D map generation in NS allows creating data points by using the mouse pointer.

The data generated here can be easily exported to the patterns module to be used for training self-organizing maps.

The last option included in this module is reserved for configuring the network training, which actually occurs in the training module. It is called “Generalization Techniques” (see Figure 3.5) and allows configuring the training procedure such that it is stopped at the moment where the network starts to memorize the data and loses its ability to generalize when unknown data is presented.

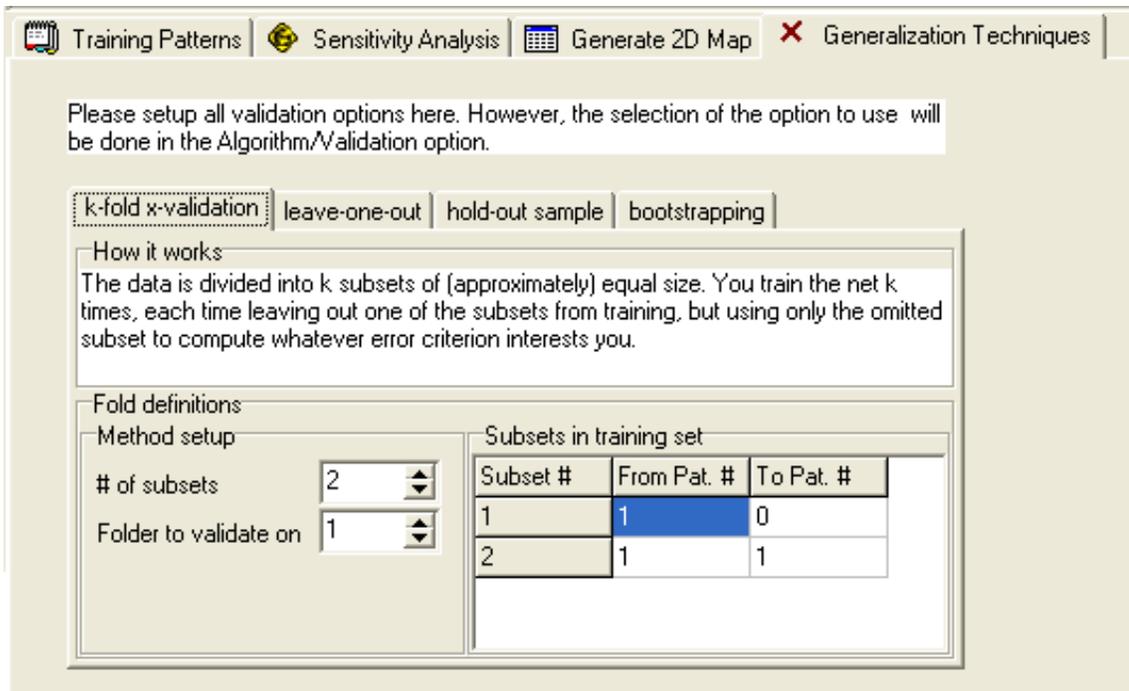


Figure 3.5: A set of options are provided to configure the k-fold cross-validation algorithm.

NS uses four different ways to handle the memorization-generalization trade off:

1. **k-fold cross-validation:** The data is divided into k subsets of (approximately) equal size. The network is trained k times, each time leaving out one of the subsets from training, but using only the omitted subset to compute whatever error criterion is of interest (refer back to Figure 3.5).
2. **leave-one-out:** The network is trained as many times as there are samples in the set, each time leaving out one sample from training, but using only the omitted sample to compute whatever error criterion is of interest.
3. **hold-out-sample:** Only a single subset (the validation set) is used to estimate the generalization error instead of using k different subsets; i.e., there is no "crossing".

4. **bootstrapping**: Instead of repeatedly analyzing subsets of the data, only subsamples of the data are repeatedly analyzed. Each subsample is a random sample (see Figure 3.6).

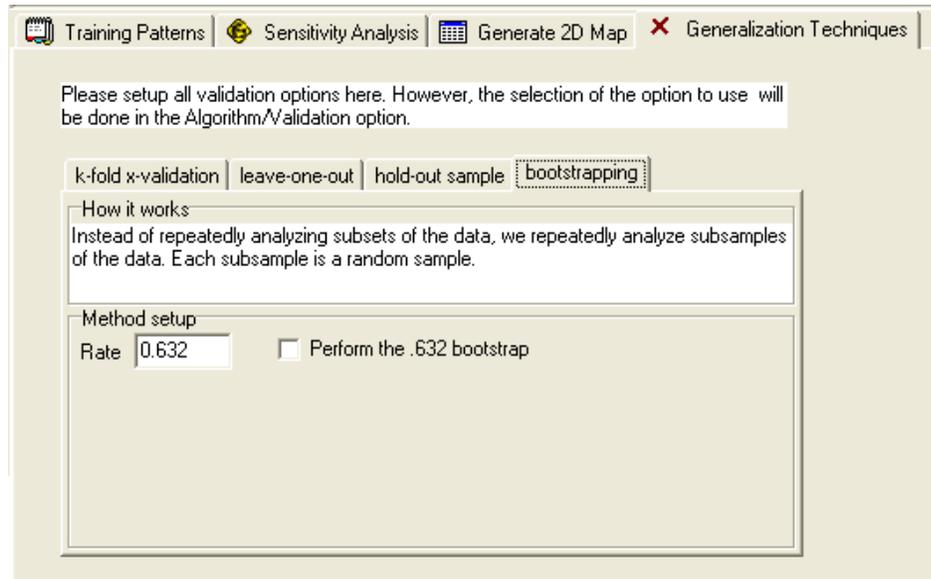


Figure 3.6: The bootstrapping technique is configured with only two options.

From these 4 options, only the first and the last are configurable, the other two options have fixed configurations and therefore can not be changed by the user. By clicking on the 2nd and 3rd tab, information will be provided on how the algorithm will be performed to comply with that particular type of training. No user interaction is performed here.

3.4.2. Creating and Training Neural Networks

3.4.2.1. Feedforward Networks Design and Training

Feedforward networks as considered in NS are multilayer perceptrons [Minsky and Papert 1969] [Widrow and Lehr 1990] with connections only between neurons of

adjacent layers. Feedback connections are not allowed from any neuron back to neurons of prior layers or to layers others than the next layer. Similarly, no connections between neurons from the same layer are allowed. This type of network is easy to train by means of the much-cited backpropagation algorithm, which is based on successive optimization by backpropagation of errors. The foundations of this unique mathematical approach were laid by Paul Werbos in his groundbreaking 1974 Harvard doctoral thesis, whose full text is published in [Werbos 1994]. The algorithm is one of the most used by scientists, engineers and researchers involved in neural networks.

NS's feedforward networks module is illustrated in Figure 3.7 and is designed to contain an editor for a multilayer network. It also consists of information panels, editing and processing tools as well as a table for network input and a corresponding results table. NS users are able to freely design the network, and to customize the neurons as well as their interconnections. Interconnections can be established by drag-and-drop operations between neurons. Neurons are represented with circles and interconnections are represented by lines drawn between pairs of neurons.

Details of the network components (neurons and their interconnections) as well as global configuration features can be viewed and edited in the network inspector (NI). Selections can be done by clicking on components in the drawing area as well as by selecting objects from a list in the NI. Once a specific neuron or weight is selected, the NI shows the features of the selected object. The object can also be selected using the selection box located at the top of the NI. Clicking on the drawing surface somewhere outside all neurons and weights takes the user back to the network configuration panel. Supervised

training occurs when the network outputs are compared with the targets and the resulting output errors are used as reference for updating weights and biases.

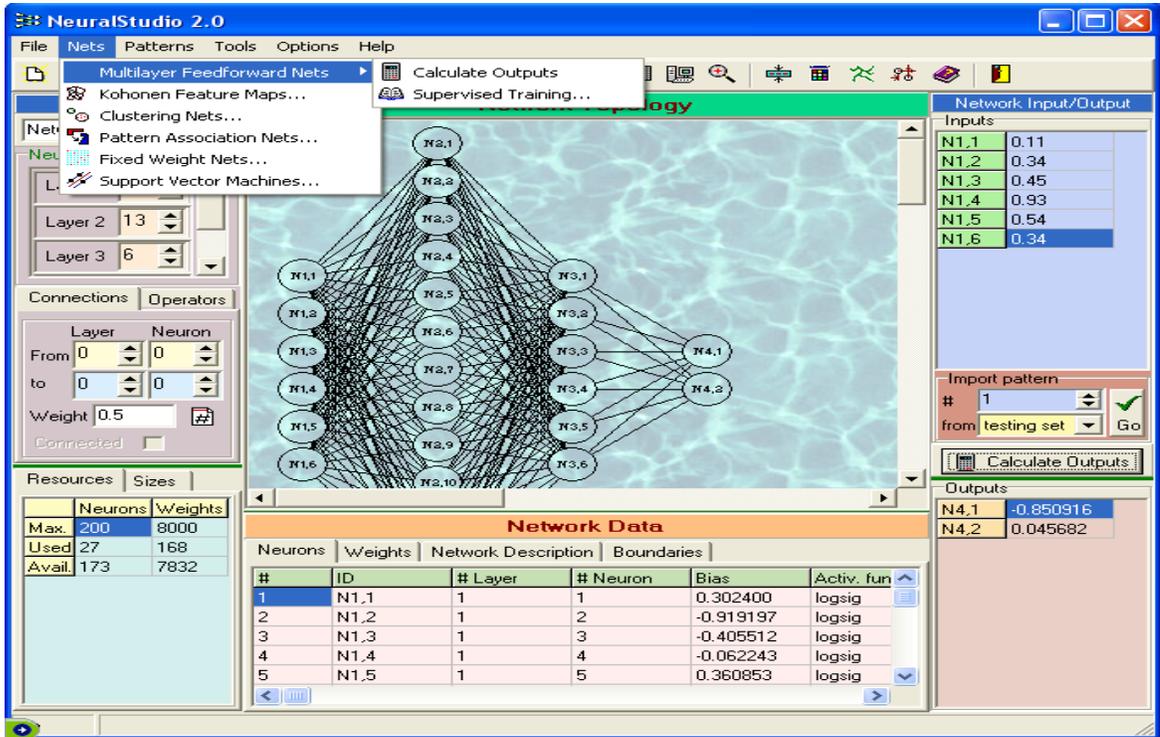


Figure 3.7: Feedforward networks module, displaying the network editor and the input/output panel

NS applies the back-propagation method to train the feed-forward networks designed in the main window. Training is performed in a training module as shown in Figure 3.8, however, the information about how to perform training is taken from the training module as well as from the network design module. The data used for training is taken from the pattern module.

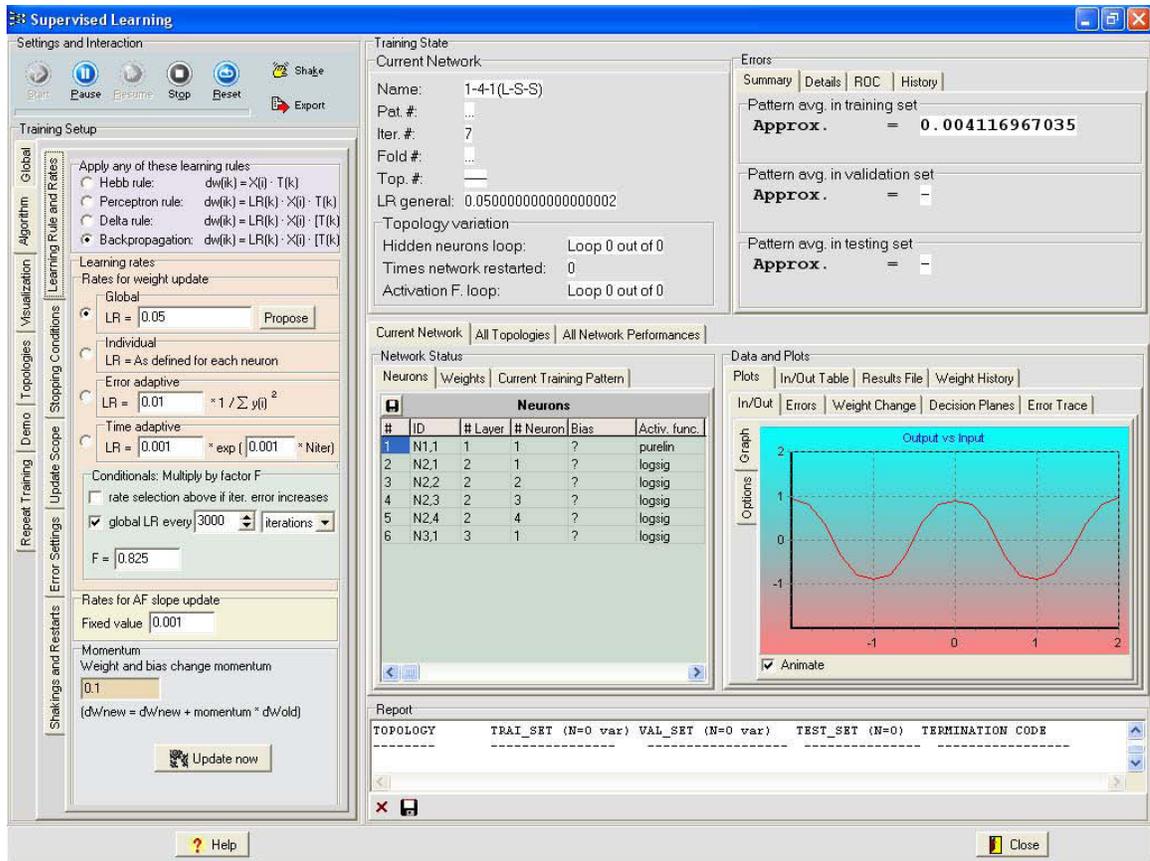


Figure 3.8: Snapshot of the supervised training module during a training session to approximate a cosine function.

Graphical outputs can be customized in the training module depending on the nature of the problem addressed. For example, the problem of approximating a multidimensional input/output relation can be visualized by plotting a pair of input/output neurons. In classification problems, animated charts enhance the results by showing classification regions with multiple boundaries. Errors of different types can be graphically traced along the iterations.

One important feature of NS is its ability to approximate different one-dimensional mathematical functions and provide animated training. These functions can be

parameterized and used for network training. During this process, animated graphical outputs can be obtained, helping NS users in understanding the training process and showing how the network's ability to generalize is influenced by the number of iterations.

The training module also allows customizing several output charts. Additionally, recall that the training can be temporarily stopped at any time, thus allowing NS users to look into intermediate results, such as current status of neurons and weights.

The training module for the feedforward networks is based on the topology of the network as established in the network editor window. Training is performed here to optimize the weights and biases so as to better accommodate the network outputs to the targets. While the number of input and output neurons is always dictated by the specific problem to be modeled by the network and can not be changed, the number of hidden layers and the number of neurons contained in each layer can always be arbitrarily chosen by the NS users. Less hidden neurons require less training time but make convergence slower. Because the total number of neurons determines the duration of the training process, NS incorporates a topology optimization option as shown in Figure 3.9, by which the number of neurons in each hidden layer can be optimized.

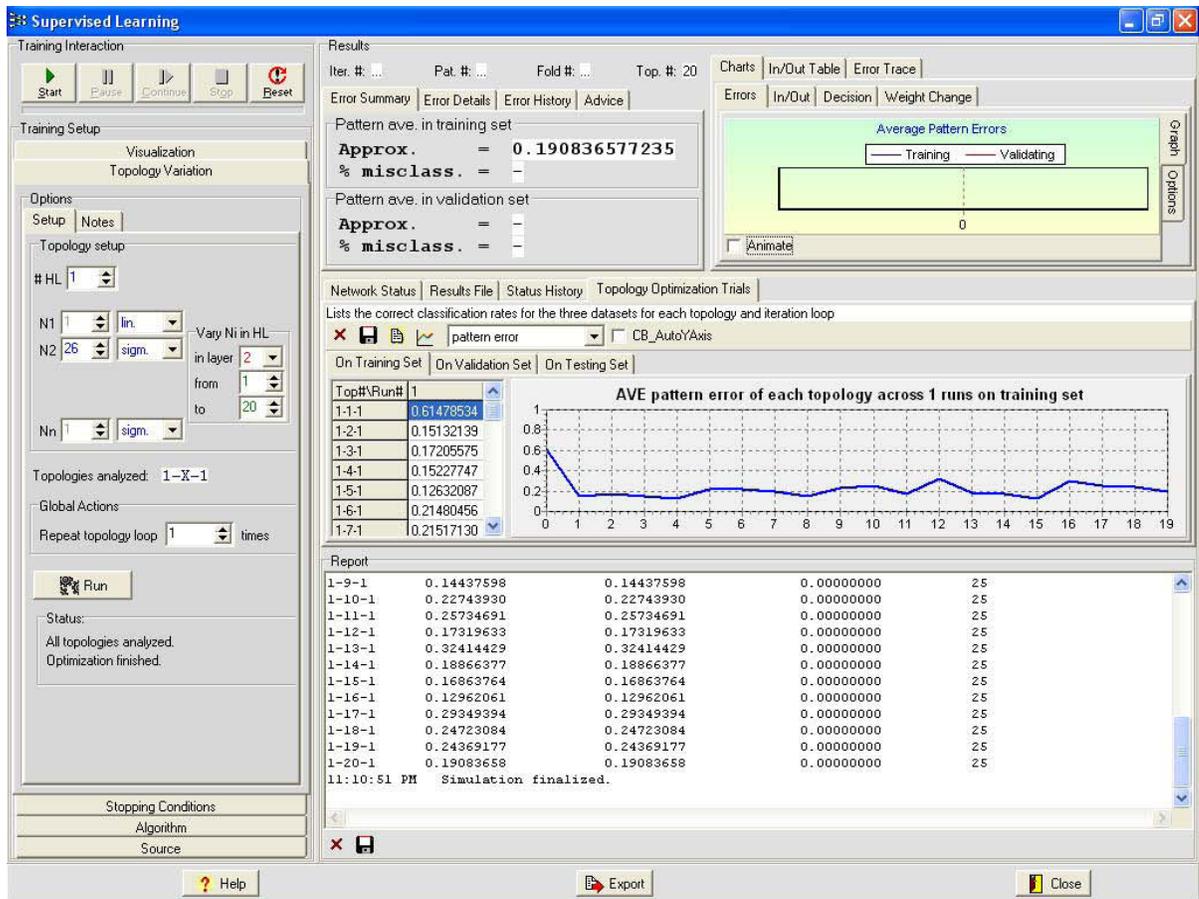


Figure 3.9: View of the topology optimization panel for backpropagation networks

The best topology is one whose average pattern error is the smallest. Before training, the user can select in the stopping conditions window whether the pattern error is computed in the training, in the validation, or in the testing set. The topology optimization is heuristic, i.e. NS generates all topologies in a specific range and computes for each one a specific performance measure that is always compared to the previous one to determine if the new topology is better or worse than the best so far.

3.4.2.2. Self-Organizing Maps

As opposed to backpropagation networks, which are used to model input-output relationships from a black box, other types of neural networks can work with only input data, without the need of having an output set which indicates some input-output relationship. In this case, there are ANNs that can be used to extract statistical information out of the input data, such as average, standard deviation, clusters, etc. Because there are no targets associated to the input patterns such that penalty/rewards to the network weights can be applied, these networks are trained without supervision.

During unsupervised learning, the network is not trained towards specified outputs. Instead, the network seeks to find patterns or regularity in the input data. The mapping implies clustering of the data. NS offers two separate modules for this type of training: the self-organizing map and the clustering networks. They are applied when no output data to correlate with is available.

The self-organizing map (SOM) [Kohonen 2001] (also called Kohonen feature map) is a map resulting from plotting the 2D weights of the neurons in a special type of network. No supervision is performed during training.

The map itself consists of a 2D grid arrangement of neurons, where each node is considered a neuron with its two weights, one in each dimension. The nodes act as cluster centroids, i.e. if an arrangement of 2x2 neurons is selected, there will be 4 clusters in the network. The training procedure is a winner-takes-all algorithm configured as follows: after selecting the desired number of nodes, each input pattern is sequentially presented to the grid. Each time a pattern is presented, the node that is closest in Euclidean distance to

that pattern is selected as the winner neuron and its weights are slightly moved towards the input pattern at a ratio dictated by the learning rate (see Figure 3.10). To stabilize the weight update and assure network convergence, the learning rate is set to gradually decrease with the number of iterations. SOM have proven to be of great use for its many practical applications, especially in parameter identification in industrial processes hard to model, where statistical features are extracted.

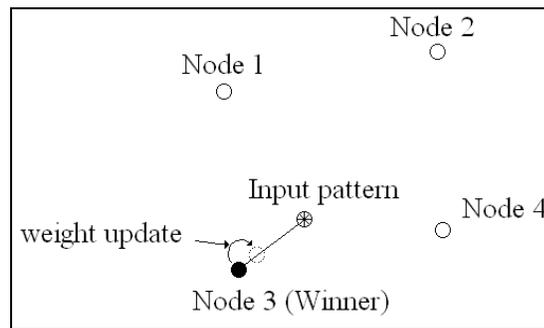


Figure 3.10: Representation of the winner-takes-all algorithm implemented in NS. The node closest to the input pattern is selected as the winner neuron and its weights are partially adapted according to the input pattern weights. Graphically, the weight update makes the winner neuron move towards the input pattern.

The main window of the SOM module is as shown in Figure 3.11. NS supports up to 50x50 maps and provides an animation during network training, such that the user can stop the iterations at any moment. The training set can be taken from the pattern module as well as from a training set generator available in the module. Once the SOM is configured, the training session can be started. Typically, the number of nodes is selected according to the intended search objective. The amount of nodes can be set from the beginning to the maximum (50x50) and then gradually decreased according to how the nodes visually arrange to create a specific amount of clusters and optimally accommodate them. The basic training procedure used in this window is similar to the one used in clustering networks, as described in the next session. Figure 3.12 shows a partially

concluded training session where the input data is circular in 2D. In this module, a set of parameters is used to configure the training and the visualization, such as drawing neuron interconnections.

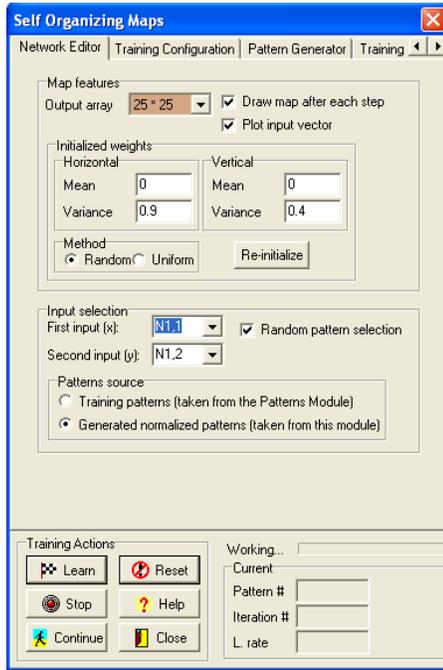


Figure 3.11: Configuration panel of SOM module, where basic map features and input source are selected

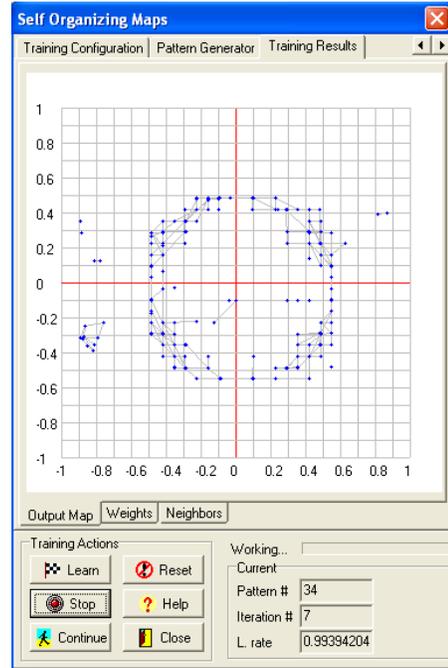


Figure 3.12: NS's module for self-organizing maps during training

3.4.2.3. Clustering Networks

This approach clusters the data together into groups, attempting to put similar ones together. This can reveal patterns hidden in the data. Predictions can then be made by comparing recent data with the different identified clusters. This type of training requires only the input and determines by itself which classes exist in the input and which input belongs to which class. The technique used here is to calculate a multidimensional feature map where each neuron represents one cluster. As with the self-organizing maps, the

winner-takes-all method is used. The algorithm finds the clusters and their centroids. In Figure 3.13, a dataset example is grouped into 7 different clusters.

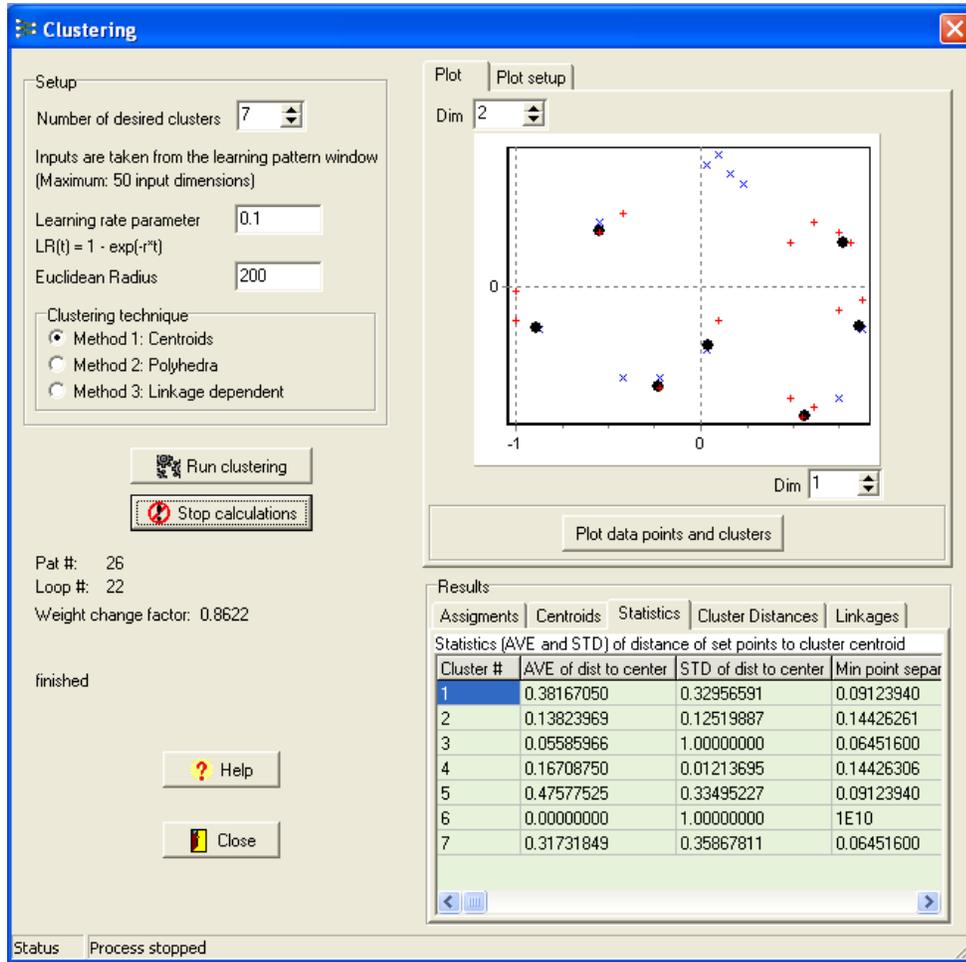


Figure 3.13: A snapshot of NS's module for clustering networks during a training session. The user can customize the training by setting the amount of clusters and deciding which clustering method to use.

In a second program stage, the algorithm is recalled and all the patterns are presented to the clusters and assigned to the best matching cluster using the Euclidean distance as grouping criterion.

3.4.2.4. Pattern Association Networks

NS introduces also a module dedicated to memory association, which allows auto-association as well as hetero-association of patterns. A view of this module is given in Figure 3.14. The module imposes no limit on network size, and lets the user choose from a list the weight adjustment rule as well as the type of activation functions.

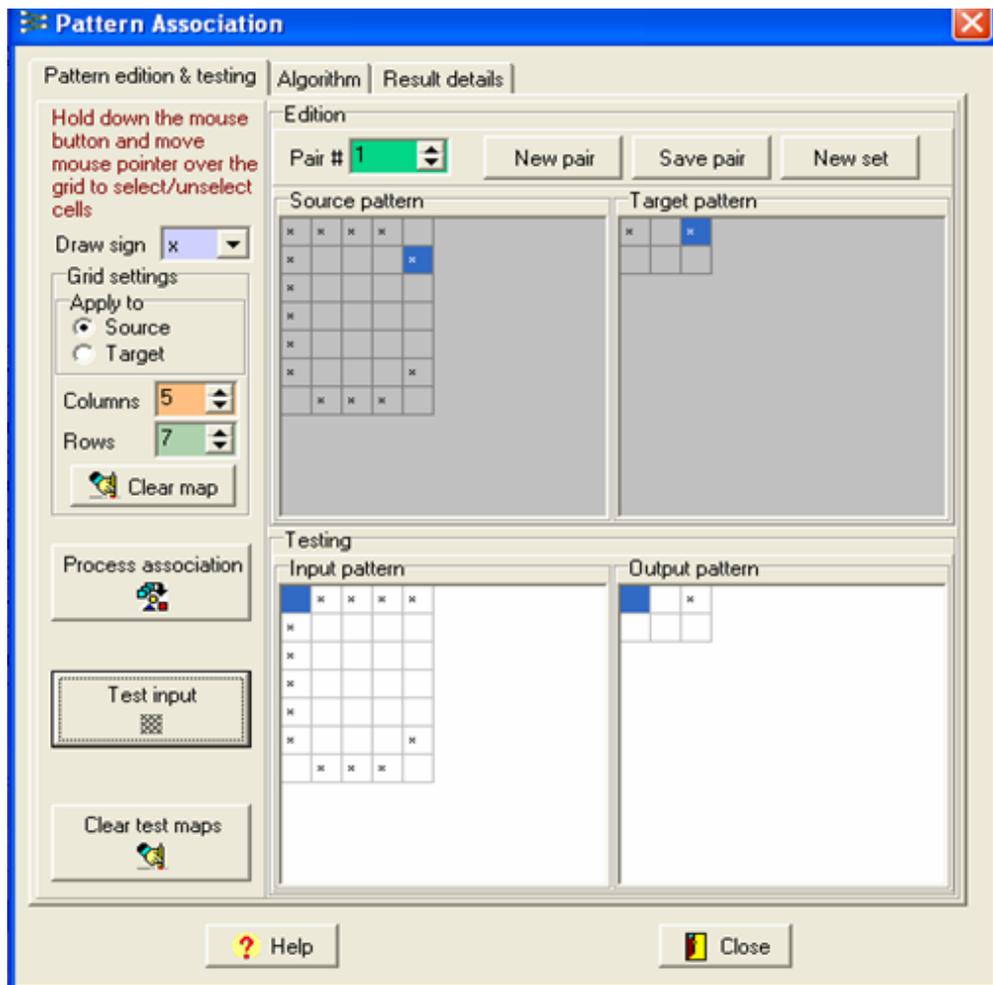


Figure 3.14: Module for pattern association while performing character codifications

3.4.2.5. Fixed-Weight Networks

An additional module for simulating fixed weight networks such as Boltzmann and Cauchy machines is provided in NS.

The options provided in this module as illustrated in Figure 3.15 are as follows:

- Number of neurons in each dimension
- Weight setting between neurons
- Cooling schedule (initial temperature and temperature decrement)

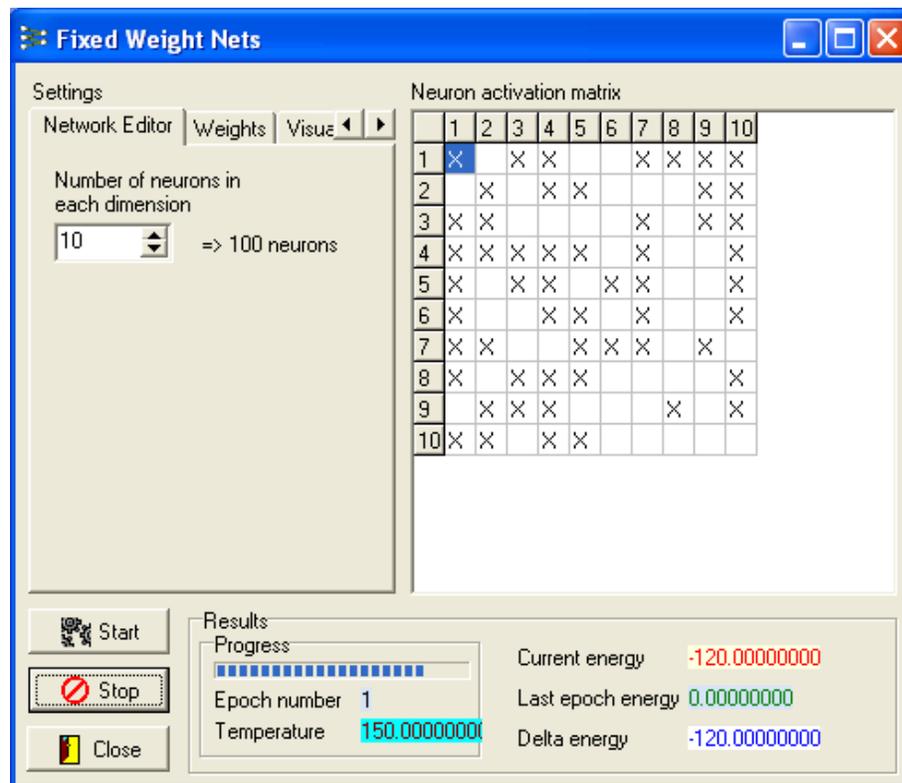


Figure 3.15: Module of the fixed weight networks while performing an optimization

By selecting appropriate weights between the neurons, many practical problems can be simulated and solved, especially transport problems with constraints as detailed in [Aarts and Korst 1987] [Aarts and Korst 1989a] [Aarts and Korst 1989b] [Teoh et al. 2008].

3.4.3. Activation Function Module

The selection of type and parameters of the activation function (AF) for the neurons of the feedforward networks is done in the activation function module. This module, as shown in Figure 3.16, provides options to assign parameters to the following AFs:

- *logistic sigmoid (logsig)*, parameterized as follows:

$$F_{\log sig}(x) = \frac{(y_{\max} - y_{\min})}{1 + e^{-a(x-b)}} + y_{\min} \quad (3.1)$$

- *Gaussian*, parameterized as follows:

$$F_{Gaussian}(x) = m e^{-\frac{1}{2}\left(\frac{x-b}{a}\right)^2} \quad (3.2)$$

- *tangent sigmoid or hyperbolic tangent (tansig)*, parameterized as follows:

$$F_{tansig}(x) = a \tanh(b x) = a \frac{e^{bx} - e^{-bx}}{e^{bx} + e^{-bx}} \quad (3.3)$$

- *linear (purelin)*, with output = input
- *binary*, with outputs = [0,1]
- *bipolar*, with outputs = [-1,1]
- *linear normalized (purelinnorm)*, with any real outputs in range from -1 to 1
- *trivalent*, with integer outputs = [-1,0,1]

In Eq. (3.1) through (3.3), the m , a , and b parameters are real numbers (usually in the range from 0.1 to 5) that are set such that the output range of the functions includes the training targets.

Of all these functions, the *logsig*, *Gaussian* and *tansig* functions are mostly applied in backpropagation networks, since their slope is not constant and therefore provide a very effective way of updating weights depending on the extent of the gradient of the error. In addition, the *logsig* is popular since its slope y' can be expressed in terms of its output as $y' = y(1-y)$, which makes it very practical for the backpropagation algorithm. The input and output neurons of feedforward networks can have linear AFs as long as there exists a hidden layer with *logsig*, *Gaussian* and *tansig* activation functions. For classification networks, it is recommended to use *Gaussian* or *logsig* AFs at the output layer.

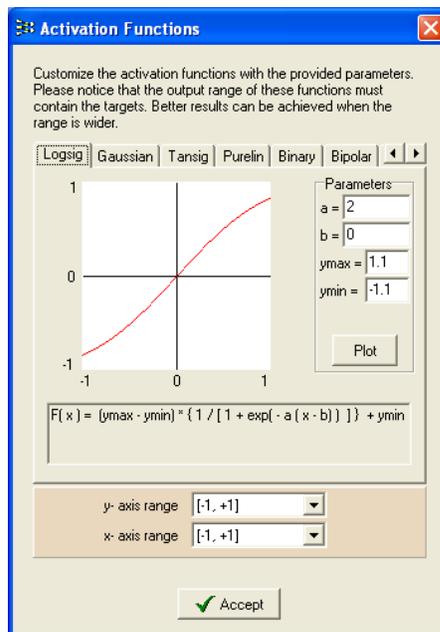


Figure 3.16: Activation functions module displaying options to assign parameters to the logistic sigmoid function

The *purelinnorm* function is linear in the range from -1 to +1 and then becomes either -1 or +1, depending on whether it is on the left side (-1) or the right side (+1) of the y -axis. This range can be made wider or inverted if multiplied by a constant. The function is useful to avoid instability during network training.

The *trivalent* function has three possible values: -1, 0, and +1, where the values can be multiplied by a constant if desired. Once all AFs are configured in this module, the user only needs to enter the name of the AF for each neuron in the feedforward networks module without the need to specify their parameters. This way, the configuration of the neurons in the network design module is simplified.

3.4.4. Support Vector Machines Module

As earlier explained, NS provides Support Vector Machines module as shown in Figure 3.17 for comparative purposes during network training and testing.

This module allows selecting four different kernels for the training and provides different visualization options. The results table shows the support vectors as well as the Lagrange multipliers of each data point.

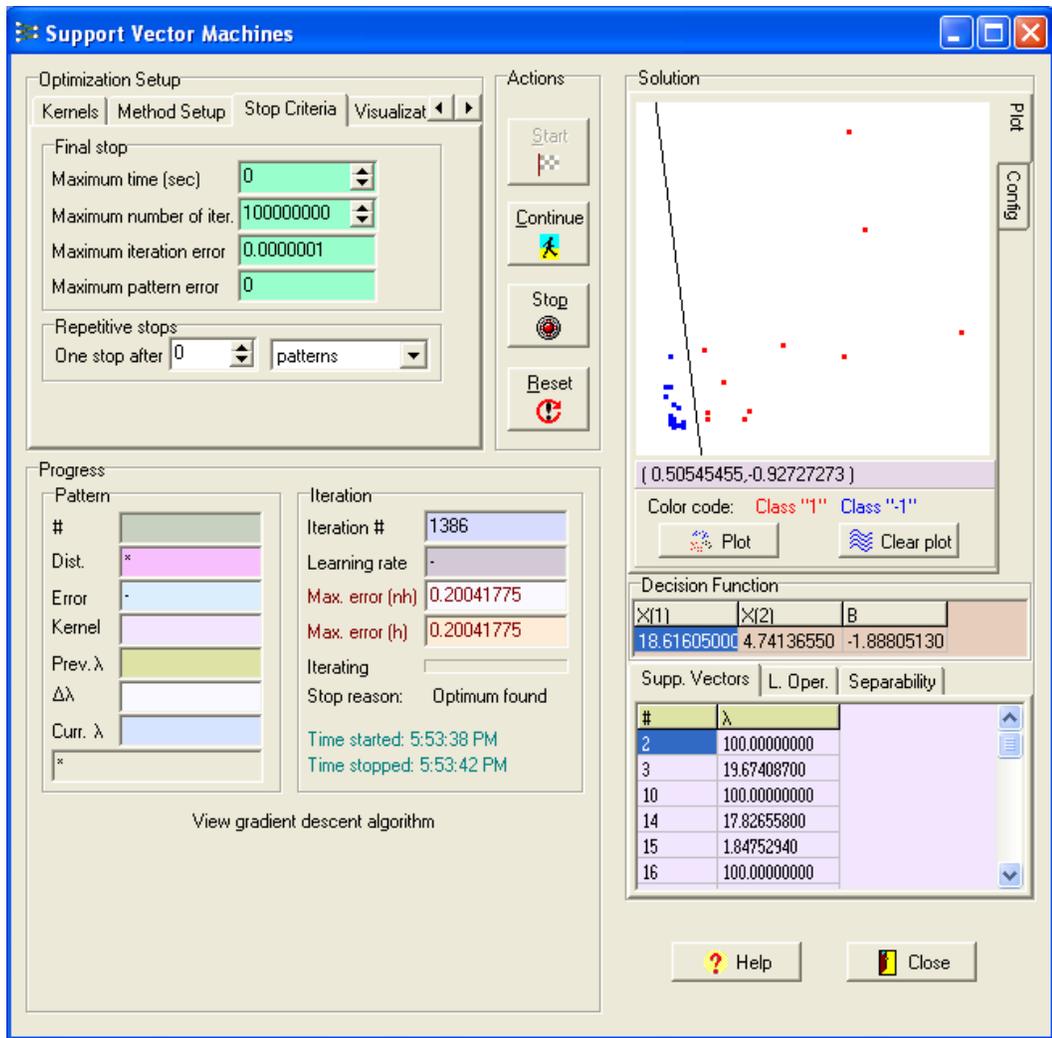
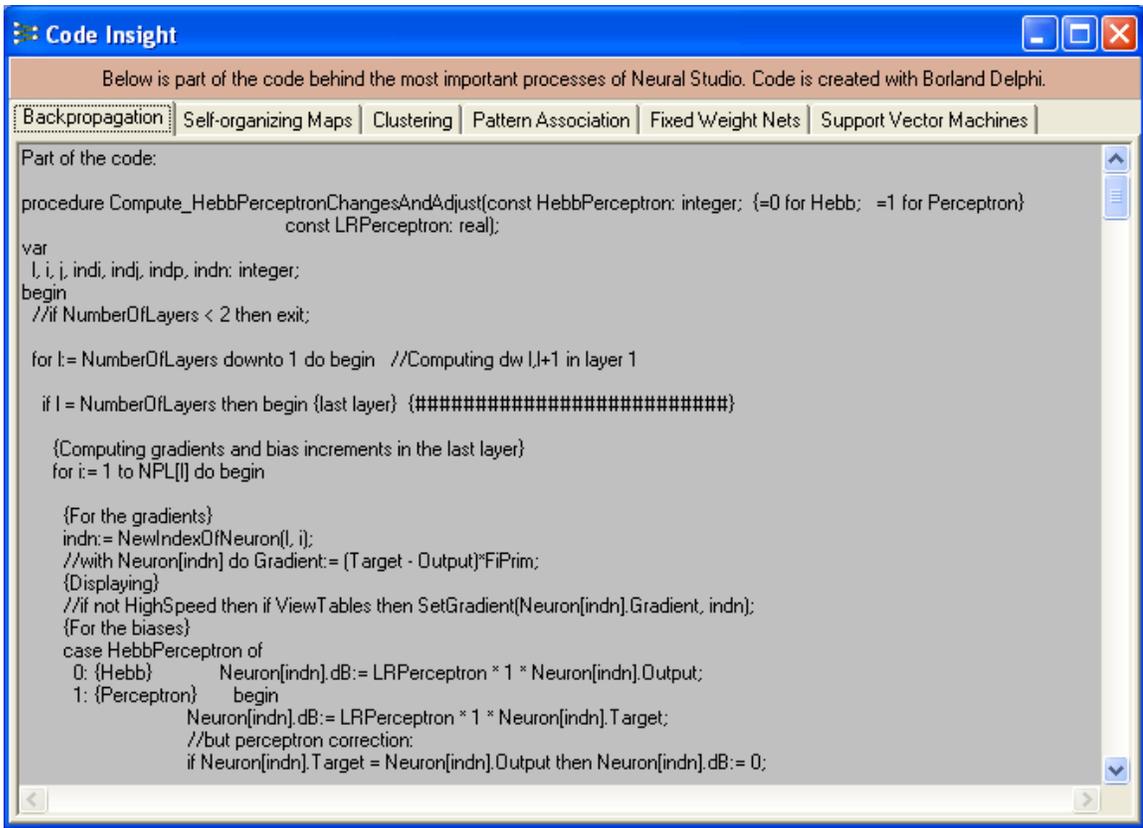


Figure 3.17: A snapshot of the SVM module during training

3.4.5. Code-Insight Module

A desired feature for learning purposes is to give NS users access to the code, especially the utilized formulas and programming steps. A code-insight module as shown in Figure 3.18 has been created specifically to make available the code related to the most important NS tasks. This possibility is useful for advanced NS users as well as for users

who need details about a specific algorithm so as to improve their knowledge and programming skills.



```
Code Insight
Below is part of the code behind the most important processes of Neural Studio. Code is created with Borland Delphi.
Backpropagation | Self-organizing Maps | Clustering | Pattern Association | Fixed Weight Nets | Support Vector Machines
Part of the code:
procedure Compute_HebbPerceptronChangesAndAdjust(const HebbPerceptron: integer; (=0 for Hebb; =1 for Perceptron)
const LRPerceptron: real);
var
  l, i, j, indl, indj, indp, indn: integer;
begin
  //if NumberOfLayers < 2 then exit;

  for l:= NumberOfLayers downto 1 do begin //Computing dw l,i+1 in layer 1

    if l = NumberOfLayers then begin {last layer} {#####}

      {Computing gradients and bias increments in the last layer}
      for i:= 1 to NPL[l] do begin

        {For the gradients}
        indn:= NewIndexOfNeuron(l, i);
        //with Neuron[indn] do Gradient:= (Target - Output)*FiPrim;
        {Displaying}
        //if not HighSpeed then if ViewTables then SetGradient(Neuron[indn].Gradient, indn);
        {For the biases}
        case HebbPerceptron of
          0: {Hebb}      Neuron[indn].dB:= LRPerceptron * 1 * Neuron[indn].Output;
          1: {Perceptron} begin
              Neuron[indn].dB:= LRPerceptron * 1 * Neuron[indn].Target;
              //but perceptron correction:
              if Neuron[indn].Target = Neuron[indn].Output then Neuron[indn].dB:= 0;
```

Figure 3.18: The code-insight module provides pseudo-code based in PASCAL for programming-interested users.

The code-insight module can be accessed with the “*Options/Code-Insight*” menu option. It is divided into the following pages: - *Backpropagation*, - *Self-organizing maps*, - *Clustering*, - *Pattern association*, - *Fixed Weight Nets*, and - *Support Vector Machines*.

Each page shows the most important code corresponding to the specified NS task. The code shown is the same code used for developing NS and has been written in Delphi, an easy to read object PASCAL code. Hence, the code is easy to understand.

3.4.6. Performance Metric Module

The Evaluation of the network performance is done in the Performance Metric Module given in Figure 3.19, which provides a platform that runs an ROC analysis with the testing results of the designed feedforward networks. The module consists of a window with three pages to show a demo and compute two-class or multi-class confusion matrix.

The first page serves as an introduction to ROC and provides a figure with sliders to change statistical parameters of two histograms that are displayed. When NS users interact with these sliders, they can see the effect in the ROC results. A detailed explanation of the terms TP, TN, FP and FN is also included.

A 2x2 table is provided in the second page for the user to enter the classification results of the current study and performs all ROC calculations automatically. It explains all terms used to measure performance in detail, which have been previously addressed in Chapter 2 of this dissertation.

Finally, the third page provides a $n \times n$ table that can be filled up with results obtained when there are more than 2 classes involved, where n is the number of classes. This page computes other performance metrics that are of interest in this type of cases.

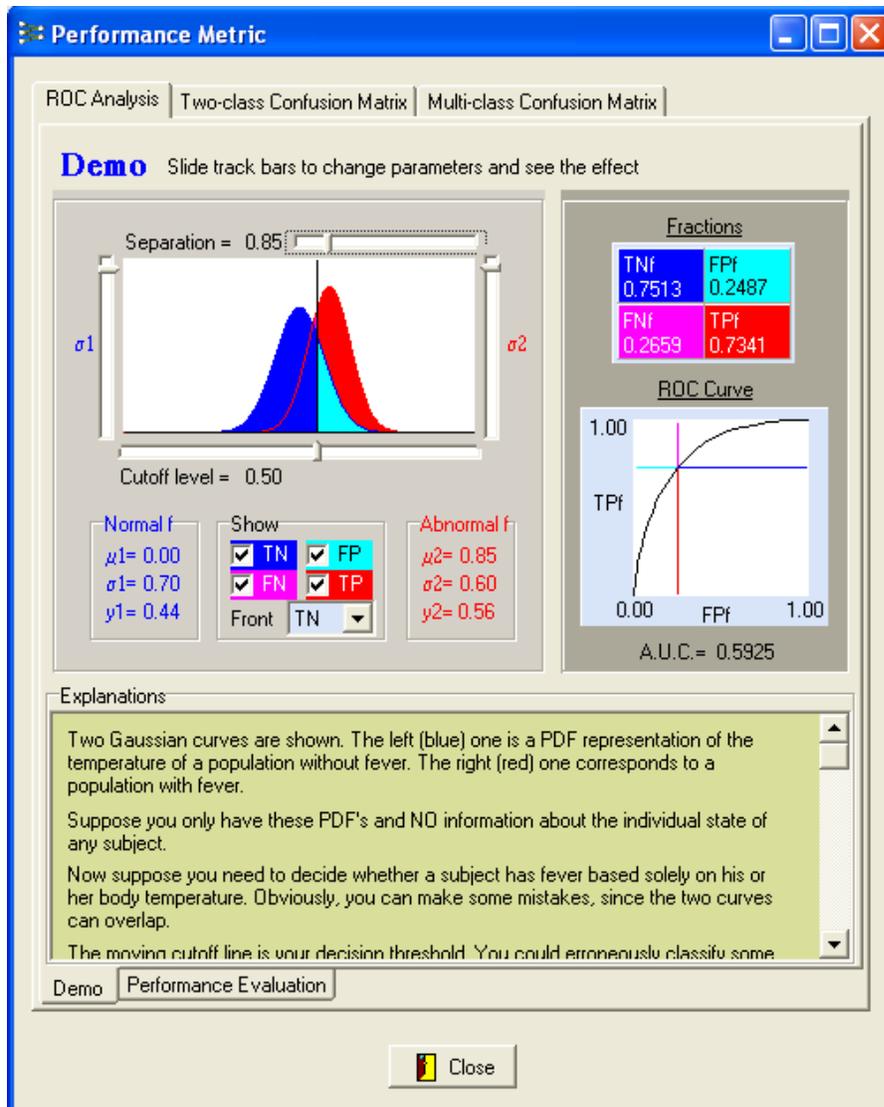


Figure 3.19: A snapshot of the performance metric module, showing the first page with a demo that allows users to understand the basic concepts of ROC.

3.4.7. Help Module

NS's main help module can be accessed from the main window and consists of a contents and a help file as shown in Figure 3.20 and Figure 3.21, respectively. The help project was developed using Microsoft's Help Workshop and the rtf file used for the compilation consists of 11 pages. The help file is accessed via the "Help/Contents" menu option,

which opens the contents file (Figure 3.20). Clicking on each of the entries opens a specific page in the help file, as shown in Figure 3.21.

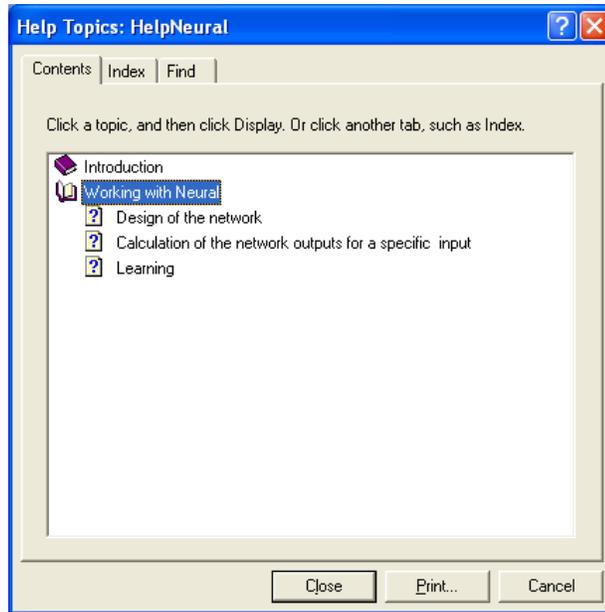


Figure 3.20: NS's help contents file, opened to the full extent

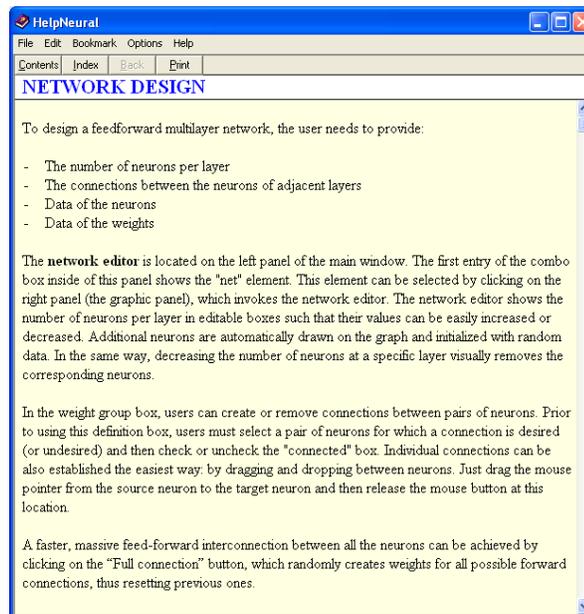


Figure 3.21: NS's help file, showing a specific topic previously selected from the contents file.

All NS modules provide a link to their corresponding help page in the help file.

The modules included in NS were provided as a result of recommendations of students who have used NS in graduate courses at FIU as well as researchers who have applied it to flow-cytometry data for leukemia blood sample classification, to EFA and seizure detection, and to security risk modeling of information systems. The modules have been designed such that each of them has the minimum set of options needed to perform the tasks and share information with other modules. Once NS was fully designed and tested for functionality, it was used to implement several methods for epilepsy analysis.

Due to the ongoing and strong collaboration between FIU and Miami Children's Hospital, the next two chapters will focus on the merits of the application of the NS tool to two examples in brain research using EEG data.

CHAPTER 4

4. AN APPLICATION OF NEURALSTUDIO TO THE DETECTION OF EPILEPTIC SEIZURES FROM SUBDURAL EEG

4.1. Objectives

The intent of this research endeavor is initially focused on developing an automated algorithm for the detection of seizures offline, based on intracranial EEG (iEEG) data that would satisfy the two primary aims: (1) a high sensitivity (i.e., minimum number of false negatives), and (2) a high specificity (i.e., minimum number of false positives). A stopping condition for delineating seizure from non-seizure files assumes therefore the highest accuracy attainable to ensure that that both primary aims are concurrently met.

Seizures occur intermittently and unpredictably. Hence, whether the planned treatment option is focal brain resection or chronic stimulation implants, massive amounts of EEG or iEEG data needs to be analyzed offline to detect seizures; a challenge that can only be met through reliable and computationally efficient seizure detection paradigm.

Over the past two decades, several automated seizure detection paradigms focusing mainly on EEG recorded from the scalp surface have been reported with different degrees of success and inherent challenges [Adjouadi et al. 2004a] [Bragin et al. 2005] [Calvagno et al. 2000] [Chander et al. 2006] [Gotman 1982] [Gotman 1999] [Smart et al. 2007]. These studies include the use of multi-channel trends, application of neural networks, use of orthogonal transforms, genetic programming, and all dwell in either time or frequency domains.

The scalp EEG however is known to have a much lower signal to noise ratio compared to EEG recorded directly from the cortical surface of the brain. In the area of epilepsy, where one of the most important goals is to detect and ultimately predict seizures, the use of many measures has been practiced by various research groups for many years with varying degrees of success. In the context of this study, many of the methods currently available in the specialized literature have been tested yielding different results. The experience gained through the different implementations reveal that the issue of contention is not in the implementation of such measures, but in determining which ones are more suitable. In this initial assessment, it was determined that the correlation sum is the temporal measure that performed best, while the gamma power is found to be the most revealing frequency range for seizure detection purposes.

EEG and iEEG signals have played an important role in the modeling of the brain's cortical dynamics and have been analyzed over the past two decades with much effort towards a better understanding of the functional characteristics of the brain, including the complex and yet to be resolved problem of seizure prediction [D'Alessandro et al. 2003] [Good et al. 2007] [Iasemidis et al. 1998] [Iasemidis and Sackellares 2001] [Lai et al. 2003]. Researchers have thus considered different approaches using a diversity of linear and nonlinear parameters in order to automate processes of seizure detection, eliciting a better understanding of the chaotic dynamics in biological systems [Bezerianos et al. 2003] [Frank et al. 1990] [Guevara 1997] [Iasemidis et al. 1994], and where promising results have been substantiated [Abend et al. 2008] [Adjouadi et al. 2005a] [Gabor 1998] [Martinerie et al. 1998] [Shoeb et al. 2004].

In using iEEG data in this study, it was considered that scalp EEG has a much lower signal to noise ratio compared to EEG recorded directly from the cortical surface of the brain. Furthermore, because of the considerable attenuation effect of the skull, the scalp EEG generally reveals very little fast activity exceeding the beta range [>30 Hz], limiting as a result the application scope of seizure detection algorithms that rely on scalp EEG recordings.

A seizure detection algorithm should then be sufficiently sensitive and specific, and any detection paradigm will thus need to capture the main features characterizing the ictal transformation that differentiate it from the interictal state in between seizures. The latter state shows relatively random distribution of frequencies over a broad range from 0-30 Hz. These frequencies may also be seen during a seizure but show considerable inter-subject and intra-subject variability. By contrast, a sustained increase in the very high frequency activity exceeding 30 Hz, defined here as the gamma band, is seen only at ictal onset and during early evolution of the seizure. These empirical facts served as the foundation of this research endeavor.

In this study, the role of the gamma frequency band is explored in order to develop a reliable offline seizure detection algorithm for EEG recorded intracranially. For this purpose, ANN architecture in NS is established and a training procedure is implemented to confront the complex nature of iEEG data. The proposed method is based on aggregating the power in the 36-44 Hz frequency range and analyzing its behavior in time, looking for patterns indicative of seizure evolution. The performance of the algorithm, which was evaluated by means of the ROC terminology, relied on two primary

aims: (1) establishing a decision space most suitable for iEEG data classification, and (2) implementing an ANN that is trained to generate the weights for the highest classification accuracy possible.

The proposed method looks at all seizure and non-seizure files together with the purpose of creating an inter-patient classifier that would be applied irrespective of the particular patient under test.

Also it should be mentioned that since all data was collected only from epileptic and not from both epileptic and non-epileptic subjects, the seizure detector only should be applied to detect seizures on EEG recordings of epileptic patients.

4.2. Experimental Setup

The data used in each one of the methods described in this study was obtained sequentially from a significant sample of 14 patients who underwent two-stage epilepsy surgery with subdural recordings. The age of the subjects varied from 3 to 17 years. The number and configuration of the subdural electrodes differed between subjects, and was determined by clinical judgment at the time of implantation. Grid, strip, and depth electrodes were used, with a total number of contacts varying between 20 and 88. The amount of data available for analysis was influenced by recording duration, and by the degree to which the interictal EEG was “pruned” prior to storage in the permanent medical record. The intracranial EEG (iEEE) data was recorded at Miami’s Children Hospital (MCH) using XLTEK Neuroworks Ver.3.0.5, equipment manufactured by Excel

Tech Ltd. Ontario, Canada. The data was collected at 500 Hz sampling frequency and filtered to remove the DC component.

4.3. Data Preprocessing

The file length for all patients was approximately 10 minutes. Out of the total 157 files considered, 35 (21 interictal and 14 ictal) iEEG data files or 22% were selected randomly and used initially in a training phase to ascertain the reliability of the gamma power in the seizure detection process. The remaining 122 iEEG data files or 78% were then used in the testing phase to assess the merits in selecting gamma power as means to detect a seizure. The 35 files selected for training are from patients 1 to 7, and patients 8 to 14 were those used subsequently in the testing phase in order to validate the classifier's ability to perform well on an inter-patient level.

The ictal and interictal files used for each patient are summarized in Table 4.1. Files containing a seizure were considered to be "ictal" files. Otherwise, the files were considered to be "interictal". It is worthy to note that even though a file was classified as "ictal", the files usually lasted longer than an individual seizure. Therefore, it was possible for "ictal" files to include some interictal data, which nonetheless the classifier needed to handle correctly. Each file was categorized by whether or not it contained a seizure, and were randomly assigned to avoid unwanted biases to either the small training set (22% of the data) or the large testing set (the remaining 78% of the data).

Table 4.1: Data set organization for the study, showing the amount of interictal and ictal files used from each patient to train and test the classifier.

(N_{neg} = number of interictal files, N_{pos} = number of ictal files)

Patient	Training Files		Testing Files		Total Nneg	Total Npos	Total
	Nneg	Npos	Nneg	Npos			
1	0	3	0	2	0	5	5
2	0	1	0	1	0	2	2
3	0	2	0	1	0	3	3
4	0	1	0	1	0	2	2
5	6	3	30	3	36	6	42
6	5	2	10	1	15	3	18
7	10	2	5	1	15	3	18
8	0	0	12	2	12	2	14
9	0	0	9	5	9	5	14
10	0	0	15	2	15	2	17
11	0	0	4	2	4	2	6
12	0	0	5	3	5	3	8
13	0	0	1	1	1	1	2
14	0	0	4	2	4	2	6
Total	21	14	95	27	116	41	157

4.4. Extracting the Power Spectrum in the Gamma Band

The gamma frequency power of each electrode was calculated from the EEG data using consecutive one-second windows (500 samples).

Due to the high volume of information contained in the pre-filtered iEEG data files, two key preprocessing steps were performed in order to (1) reduce the data to be analyzed, minimizing as a consequence the computational requirements, and (2) seek a transformation of the raw iEEG data in order to enhance the accuracy, specificity and sensitivity of the seizure detection algorithm. In step (2) the transformation chosen is that of gamma frequency component after a thorough evaluation of several other standard parameters in the time domain such as mobility, complexity and activity [Hjorth 1970a], as well as in the frequency domain by evaluating all other frequency ranges. The iEEG

data files were further analyzed with one second (1-sec) timed windows and the power of the gamma frequency component was extracted for all these 1-sec windows and for each electrode.

The power spectrum of the gamma frequency band was computed as given in Eq. (4.1):

$$P_g = \int_{b_{start}}^{b_{end}} |F(w)|^2 dw \quad (4.1)$$

where b_{start} and b_{end} are its starting and ending frequencies in the 36 to 44 Hz band, with $F(w)$ defining the complex fast Fourier coefficient at frequency w .

Figure 4.1 shows the gamma power near the time of seizure onset using iEEG data from four different patients. It shows the intricate and yet informative nature of the gamma power signal. At the time of seizure onset (vertical red line), there is an abrupt change in magnitude almost in synchrony for all electrodes. The vertical red line represents the seizure onset previously labeled at the observation room by the EEG expert. It should be noted that the onset as marked by the EEG expert and the results of the synchronized increment in magnitude of the gamma frequency do not coincide exactly in time, which only heightens the relevance and need for an automated seizure detection process. Further clinical evaluations reveal that the synchronized increment in power spectrum in the gamma band does actually coincide with the actual clinical onset of the seizure. This is viewed as another interesting finding of this study. Note in this figure that the scales are different for the different subjects, making absolute thresholds impossible to use. This is another important issue that is addressed and resolved in this study.

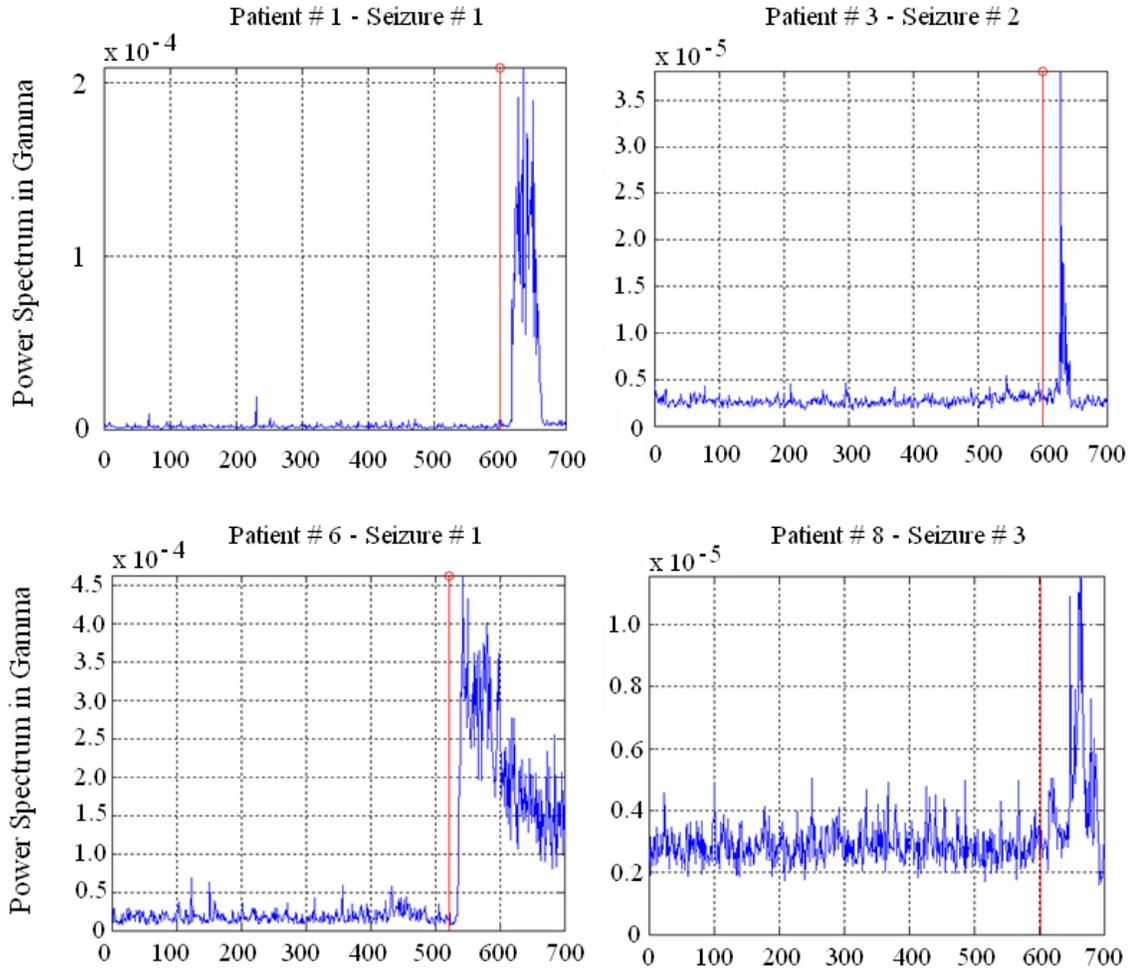


Figure 4.1: An illustrative example of gamma power for all electrodes versus time (shown in terms of samples) for 4 different seizures from different patients. The vertical red line is the seizure onset as identified by medical experts. Note that the scales are different for the different subjects, thus making absolute thresholds impossible to use.

4.5. Aggregating Features

In order to handle the variable number of electrodes used from patient to patient, the power spectrum in the gamma band across all electrodes was averaged. This averaging process, which is referred to as the inter-electrode mean, was used as input to the classifier. The use of the inter-electrode mean is a result of the experimental studies [Albano et al. 2000] [Arnhold et al. 1999] [Cabrerizo et al. 2006b] [Larter and Speelman

1999] that reveal that electrodes tend to interlock in behavior at the onset moment of a seizure. Therefore, this average process of all electrodes did not distort the results, and yet allowed for uniformity in the implementation process across patients independent of the varied number of electrodes used for each.

With this fact, it is emphasized that the concept of averaging for a representative signal does not sidetrack from the main intent of detecting a seizure with the highest accuracy, specificity, and sensitivity possible. At the same time, such a step minimizes to a great extent the computational burden [Cabrerizo et al. 2007] that would have been required in dealing with all of the iEEG data as input to the classifier, and simplifies greatly the seizure detection process as only one representative signal is fed into the ANN-classifier. Figure 4.2 illustrates this assertion by comparing the contributions of individual electrodes to the behavior of gamma frequency for every single electrode used for a particular patient, and the results obtained using averaging or the so-called inter-electrode mean signal S_{μ} for an arbitrary section of iEEG.

4.6. Establishing the Inputs of the ANN

A threshold had to be established before the inter-electrode mean of the power spectrum in the gamma band can be used to detect a seizure. If the threshold was crossed at any point during a file, then the entire file is classified as a possible “seizure” file. If the threshold was never crossed, then it was classified as a “non-seizure” file.

The seizure detection method tests a threshold based on the inter-electrode mean signal S_{μ} , which has a natural variability even in a single patient, and even in the absence

of a seizure. Using the same threshold in many patients makes the variability even greater, but increases the clinical usefulness of the test. This is essentially a dilemma that is faced due to unreliable and changing thresholds and varying standard deviations that can be experienced even within a single patient.

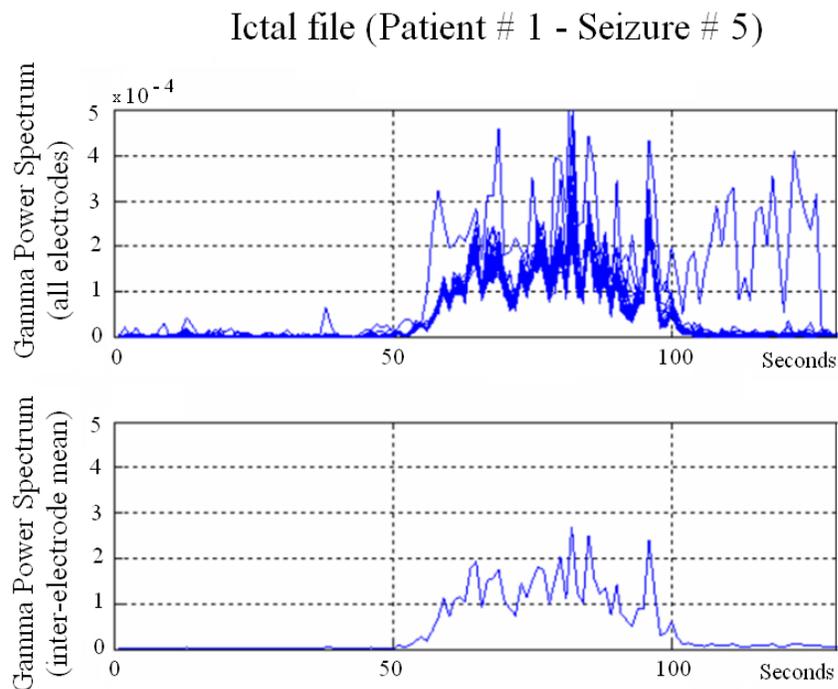


Figure 4.2: Top figure shows the behavior of the gamma power for each of the 48 electrodes used for subject 1, seizure 5, and the bottom figure displays its respective inter-electrode mean signal. Seizure onset is around 100 seconds.

With these observable facts, the problem becomes difficult to contain not only in terms of these noted variations, but also in ascertaining in a meaningful way the performance evaluation of the classifier. The example given in Figure 4.3 illustrates perfectly this dilemma. Note how different is the variation of the magnitude of the inter-electrode

mean signal between patient 1 and patient 2. However if one is to rescale the y axis for patient 2 it will reveal that an ictal change similar to that of patient 1 is indeed present, which only amplifies that a singular threshold computed on the basis of patient 1 would have missed the ictal change in patient 2 observed after rescaling.

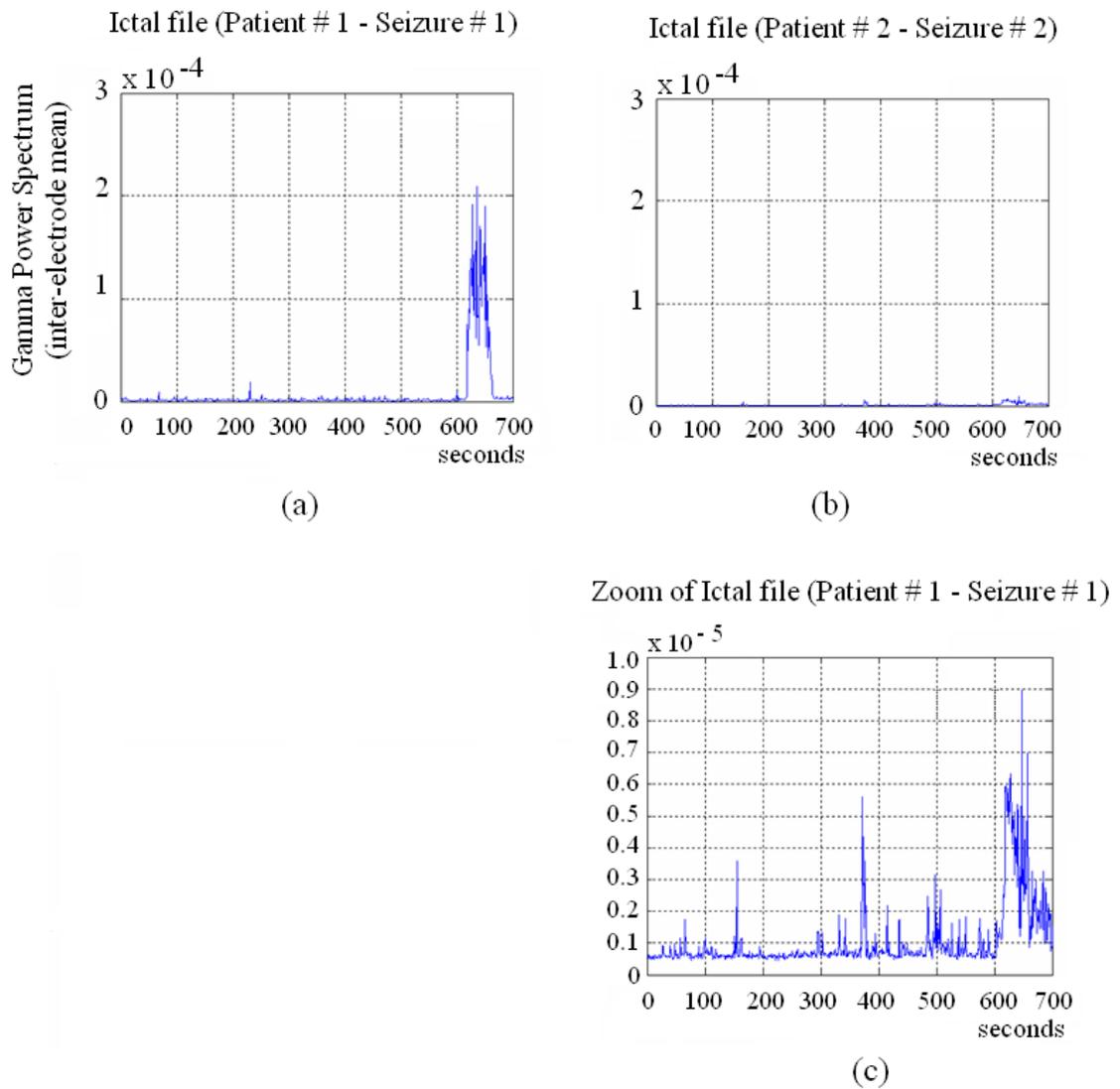


Figure 4.3: Illustration of the variation of the inter-electrode mean signal S_{μ} within the same patient: (a) seizure 1 of patient 1, (b) seizure 2 of patient 2, (c) zoomed in view of seizure 2 of patient 2.

As a consequence of this observation, a generalized statistical threshold was established for the power spectrum of the gamma band that can work across all patients independently of the signal's magnitude. This threshold is thus defined by the average of inter-electrode mean plus one standard deviation of the inter-electrode mean signal as defined in Eq. (4.2)

$$T = \overline{S_{\mu}} + \sigma \quad (4.2)$$

The suitability of the average spectral power between electrodes for seizure detection has been proved in an early study by [Tito et al. 2007].

In terms of the proposed classifier, any point which exceeded the threshold was considered a point belonging to a potential seizure, with all subsequent consecutive points exceeding this threshold defining the duration of such a seizure. Such a determination constitutes a first and most critical requirement for the proposed algorithm.

The next stage of the algorithm was to determine additional means to validate that such potential seizure can indeed be declared an actual seizure. At this stage of the investigation, two measurements were taken into consideration. The first measurement is the duration in time in which the signal S_{μ} was consistently above the aforementioned threshold, and the second measurement is found to be the maximum value of S_{μ} in that interval. These two measurements were deemed sufficient for the detection algorithm to work properly.

With these two measurements in place, a table was constructed to train the seizure ANN-based detector. The table contains as many records as data files were used in the training,

whereas each record contains three values: the two aforementioned measurements and also a target (+1 for seizure file and -1 for non-seizure file). Recall that non-seizure data files that did not meet the first requirement (not having a single point that passed the set threshold) were not used in this table to begin with, since they were already identified as true negatives.

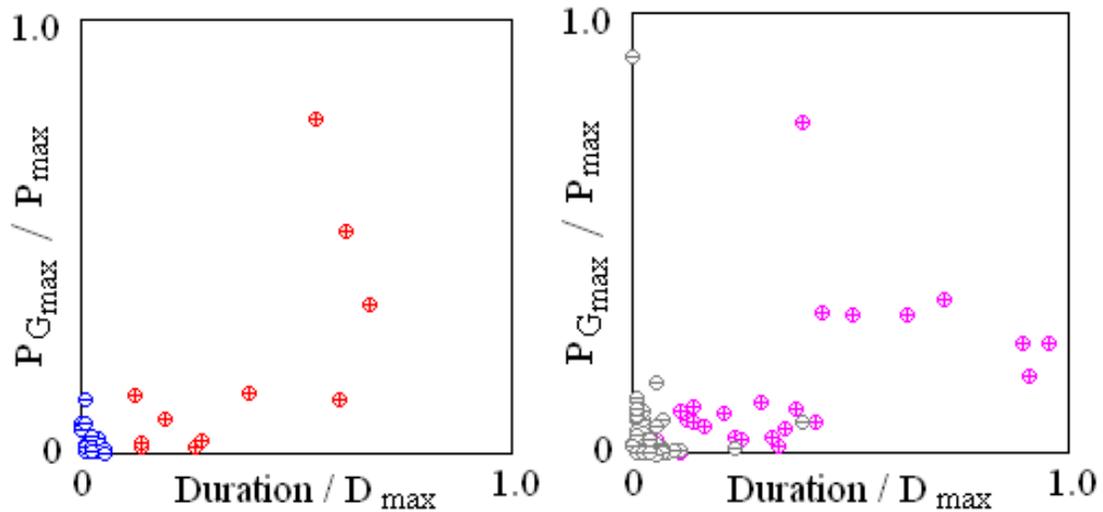


Figure 4.4: Positive (ictal files) and negative (interictal files) points used for training (a) and testing (b). The x-axis represents the duration in seconds (normalized with respect to the maximum), whereas the y-axis represents the maximum value of the inter-electrode gamma power (also normalized). Color code: red and purple: ictal files for training and testing, respectively; blue and gray: interictal files for training and testing, respectively.

Figure 4.4 helps in visualizing the geometrical placement of the two pattern classes in the input space, for both training and testing set. In this 2D space, the x -axis represents the duration which was divided by a normalization factor of 1000 in order to accelerate the convergence of the ANN, and thus facilitate the determination of the optimal weights of the network; while the y -axis represents the maximum of S_{μ} in the interval above the threshold.

When this 2D space is chosen appropriately, which was the most challenging part of this research problem, patterns of the same class will tend to cluster together and the classification algorithm is logically expected to yield optimal results.

4.7. ANN Configuration

To address this problem and begin its implementation steps, a 3-layer ANN with two inputs, 5 hidden neurons and 1 output was designed with NS as illustrated in Figure 4.5. The 5 hidden neurons were initially chosen following an empirical rule that proposes for the number of hidden neurons twice the number of inputs plus one, whereas the exact number of neurons will be later investigated by a heuristic approach. To simplify the study, only three types of activation functions were considered: linear (L), logsig (S), and Gaussian (G), as they provide the three basic shapes of activation functions. These functions were used for the hidden and output layer, while the input layer was set to have always linear activation with zero bias.

To comply with the bipolarity of targets, the parameters of the activation functions were set such that their output falls in a range from -1 to +1.

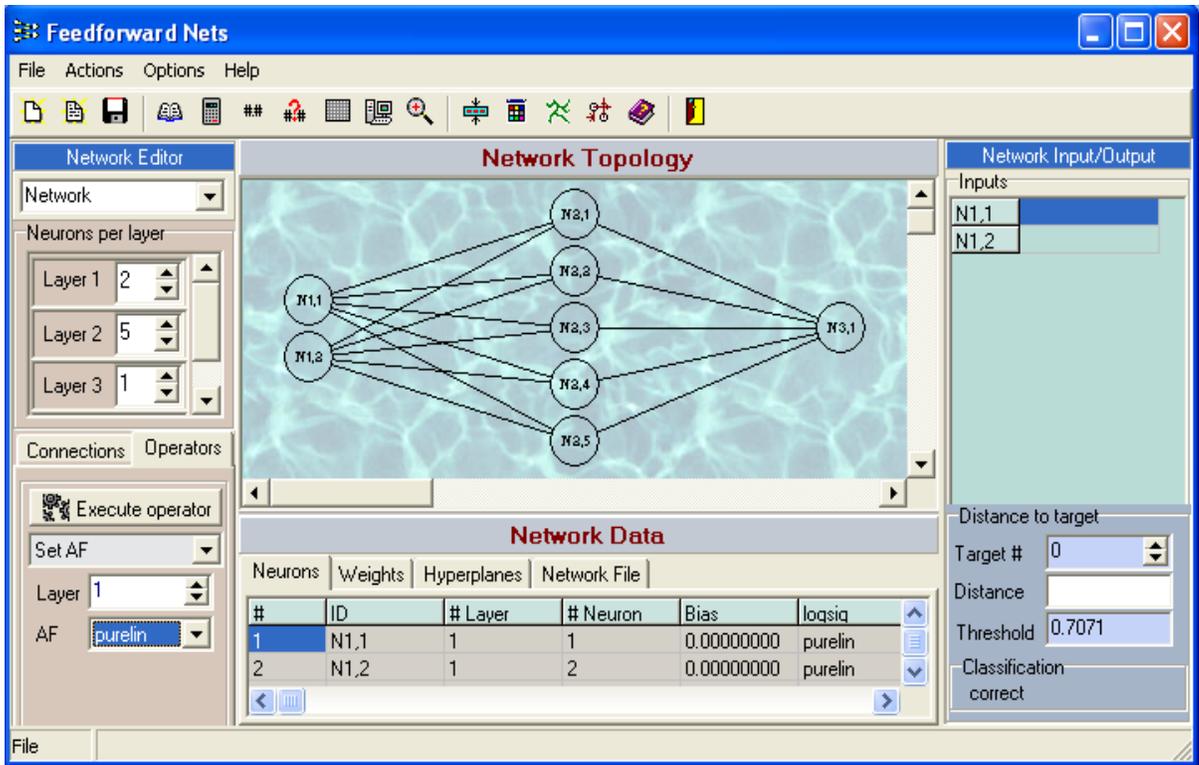


Figure 4.5: The 2-5-1 ANN topology chosen in NS to detect seizures based on the power spectrum in the gamma band.

4.8. Training Procedure

Training was performed with the backpropagation algorithm and finalized with early stopping (a cross-validation strategy) as a regularization procedure to avoid network memorization. The procedure was set as follows: every 3 iteration loops, the average square error on the validation set is computed and compared to the previous 5 values computed thus far. If the last error is higher, the iteration is stopped, because this could represent an increasing error trend.

NS offers different options to define misclassification. Since the targets in this study are either -1 or +1, misclassification was defined to happen when the sign of the output differs from the sign of the targets.

For comparison purposes, different topologies with two (2-1) and three layers (2-X-1) were sequentially trained, using in each layer either the linear (L), logsig (S), or the Gaussian (G) activation function, with exception of the input layer for which linear activation functions were used at all times. The hidden neurons were varied in a range from 1 to 10. When creating all combinations, a total of 3 possible 2-1 topologies (two inputs and 1 output), were trained, namely: L-L, L-S, and L-G. In the 2-X-1 topologies, a total of $9 \times 10 = 90$ possible combinations were created and trained, where 9 is the number of allowed combinations of activation functions across the layers (L-L-L, L-L-S, ... , L-G-G) and 10 is the number of topologies created for each activation combination (2-1-1, 2-2-1, ..., 2-10-1).

During an exploratory training phase on a group of networks, the shape parameters of the logsig and Gaussian activation functions (in both cases, the “ a ” parameter) were varied from 0.1 to 5 in steps of 0.1, yielding no significant variation in the results, i.e., the weight update algorithm always yielded the same results regardless of the parameter value. Therefore, the shape parameter was set to $a=5$ for the logsig (positioned in the numerator of the power) and $a=0.5$ for the Gaussian (positioned in the denominator of the power).

Figure 4.6 shows a snapshot of NS performing a topology search loop, where each topology is trained until the cross-validation condition is satisfied.

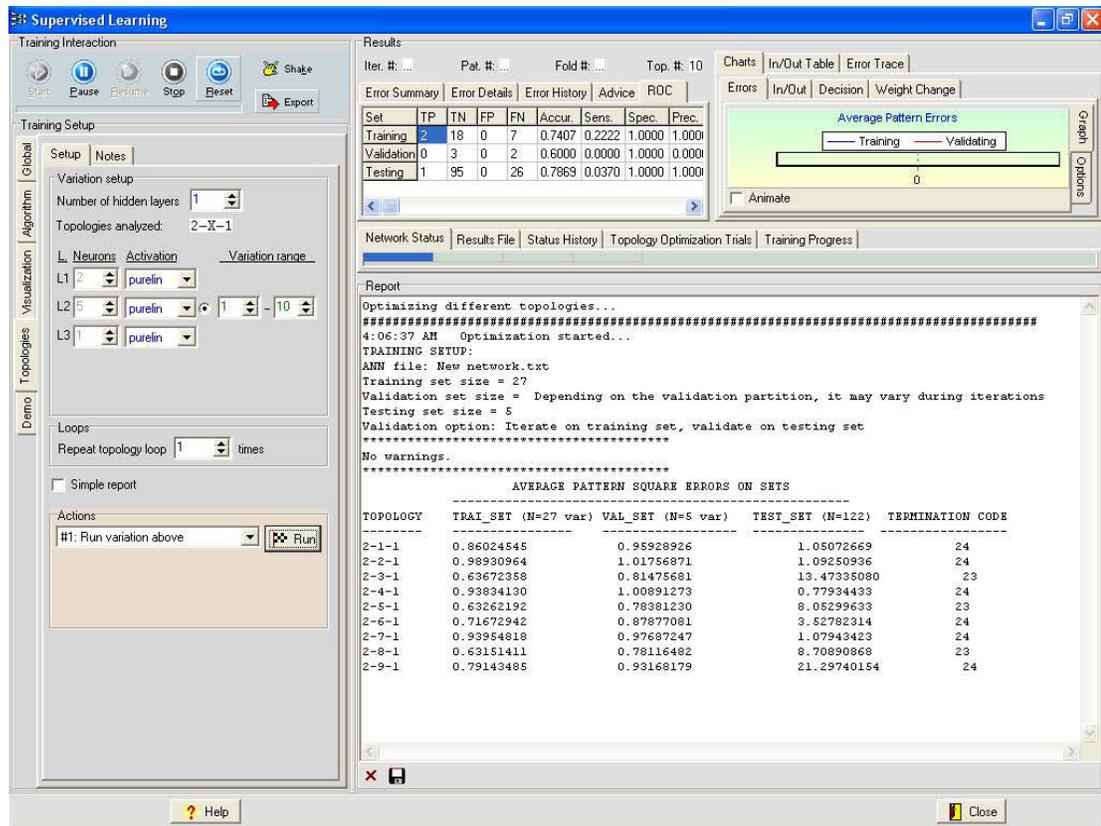


Figure 4.6: NS's training module while performing a topology optimization in the 2-X-1 group with linear activation in the first two layers and logsig in the last layer. The number of hidden neurons is being varied from 1 to 10.

The data set was arranged in two ways as defined in Table 4.2. Data processing details of arrangement A were given earlier in section 4.3., where 50 % of the patients (1-7) were used for training and another 50 % (8-14) for testing. The training set was sub-divided into pure training and cross-validation subset. Additionally to this data setup, arrangement B was prepared in which data from all subjects was used for training and testing.

The reason of these two arrangements was to confront two very different scenarios: One in which testing occurs on patients not involved in training, and another in which all

patients are involved in training and testing as well. The key question with the first arrangement was to elucidate whether the dynamics of iEEG during the transition to seizure onset are somewhat similar regardless of the patient. If a detector trained in this way performs well on files from other patients, then one can assume that the power spectrum in the gamma band is an excellent instrument to detect seizures offline since it is not restricted to the dynamics of the EEG of any particular patient.

Another hypothesis is that if the detectors trained with the two arrangements don't differ substantially from each other in terms of performance, then it can be concluded that the dynamics of the power spectrum in the gamma band in the transition to seizure is patient-independent. Table 4.2 shows how the data was distributed for the two arrangements. In both arrangements, the intent was to use about 20 % of the training files for cross-validation and not for weight update, whereas that proportion was difficult to obtain due to the insufficient amount of ictal and interictal files per patient. In general, training on the pure training subset was performed 3 times before validation in the cross-validation subset.

Table 4.2: Setup of two arrangements with indication of the files pertaining to the training, cross-validation and testing (N_{neg} = number of interictal files, N_{pos} = number of ictal files, N_{total} = total number of files).

	Training	Validation (20% of training set)	Testing
Arrangement A	Patients 1-7 (subset) $N_{neg}=17 = 21 - 4$ (20%) $N_{pos}=11 = 14 - 3$ (20%) $N_{total} = 28 = 35 - 7$	Patients 1-7 (subset) $N_{neg}= 4$ $N_{pos}= 3$ $N_{total} = 7$	Patients 1 - 14 (subset) $N_{neg}=116-21 = 95$ $N_{pos}=41 - 14 = 27$ $N_{total} = 157 - 35 = 122$
Arrangement B	Patients 1 - 14 (subset) $N_{neg}= 59$ $N_{pos}= 19$ $N_{total} = 78$	Patients 1 - 14 (subset) $N_{neg}=12$ $N_{pos}=6$ $N_{total} = 18$	Patients 1 - 14 (subset) $N_{neg}=45$ $N_{pos}=15$ $N_{total} = 60$

After training all 93 possible topologies, they were compared to each other in terms of performance on the validation set. For simplification, selection of the best topology was done with the average pattern square error on the validation set rather than with ROC terminology. Also the limited amount of data did not allow generating ROC curves to compare the performances of all topologies. The pattern square error is the square of the difference between the network output and the target for a specific pattern, and the average is taken after each iteration loop across all patterns in the table.

In the course of the training, several preliminary findings concerning network topology and data set organization were made. They will be presented before discussing the classification results.

4.9. Training Results

4.9.1. Findings Regarding the Number of Hidden Neurons

In [Berry and Linoff 1997] [Blum 1992] [Boger and Guterman 1997] [Swingler 1996], different rules for selecting the number of hidden units are proposed. Intuitively, increasing the number of identical hidden neurons represents adding more degrees of freedom to the backpropagation algorithm in a multilayer perceptron trained with early stopping. To investigate the effect of adding hidden neurons, all topologies between 1 and 40 hidden neurons with all possible combinations of activation functions were initialized with random weights and trained with early stopping. This type of analysis was easily done because NS allows for automatically training and modifying a network

topology several times. A snapshot of the training module performing this action is illustrated in Figure 4.7.

To make this discussion independent of how misclassification is defined, approximation errors rather than misclassification errors will be shown. In total, 10 trials were performed for each specific topology. The grand average of the average pattern square errors across all trials for the L-L-L and L-S-S topology groups and for both training and validation set are shown in Figure 4.8, where in the training set it is hard to observe any clear trend, because of the recurring peaks. This proved that adding more hidden neurons (as reflected by the x-axis) does not considerably help reduce the error on the training set. This phenomenon could be possibly associated to the use of numerical methods which always require a starting solution that can trap networks in local minima. Also, note that the recurring peaks in the validation errors tend to increase in value, which reduces generalization ability. If the starting weights were kept the same, the error would certainly reduce in trend. However, there is no way to know in advance which initial set of weights will yield the lowest error, and as a conclusion, it is possible for a topology with less hidden units to achieve lower errors than one with more hidden units, if the starting point is different.

This proves that the conclusions made by [Barron 1993] [DeVore et al. 1989] [Lawrence et al. 1997] [Sarle 1995] [Tetko et al. 1995] and [Weigend 1994] regarding the convenience of adding more hidden units to a multilayer perceptron, trained with backpropagation and early stopping in order to find a better local minimum and reduce

both training and generalization error, do not always apply because of drawbacks related to random selection of starting solution.

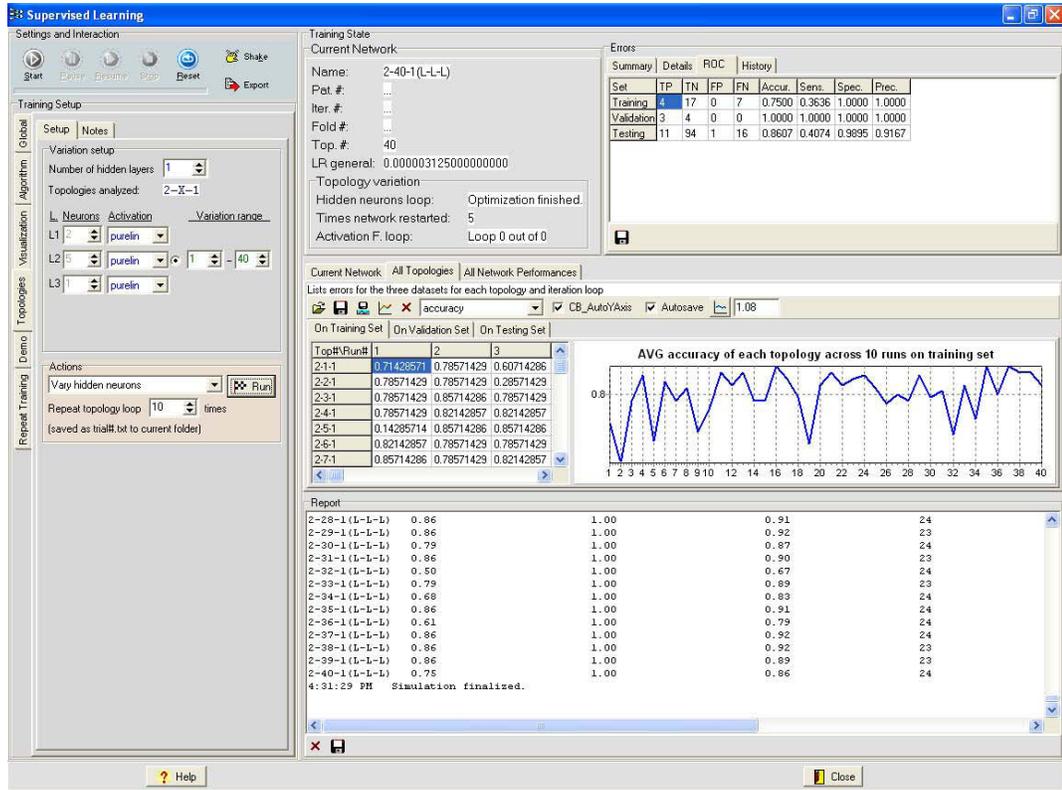


Figure 4.7: NS while performing topology variation training in a loop, showing the plot of the average accuracy after ten trials in each set for every specific amount of hidden neurons.

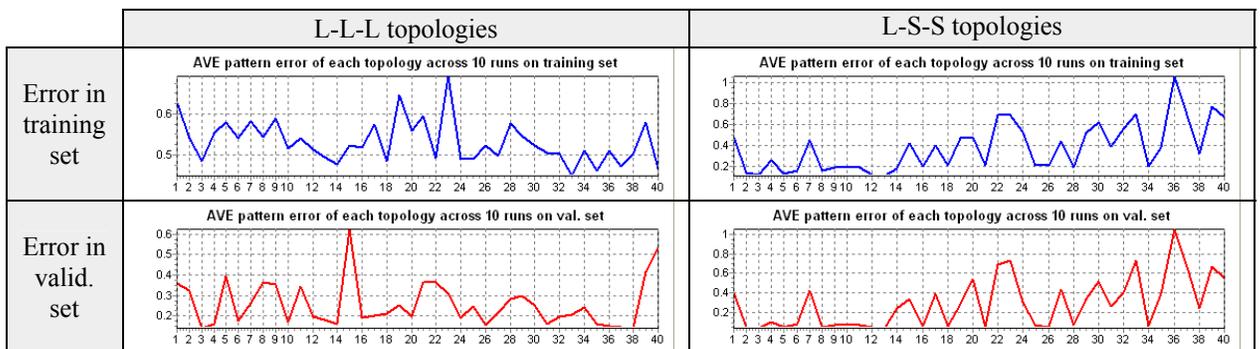


Figure 4.8: Behavior of the grand average of the mean pattern square error (in training and validation set) vs. the number of hidden neurons in L-L-L (left) and L-S-S (right) topologies. Color code: blue and red: training and validation errors, respectively.

Since the optimum number of hidden neurons was impossible to obtain empirically, it was opted to include between 1 and 10 hidden neurons. Considering the facts stated above, different trials were performed in order to find the best ANN in each arrangement.

4.9.2. Findings Regarding Data Arrangement

After training the networks in the two arrangements, the optimum topologies were sorted in descending order of the average validation set error. Table 4.3 shows the five best performing topologies, placing the 2-5-1(L-S-S) topology on top in arrangement A and the 2-10-1(L-S-S) topology in arrangement B. As it can be noted from this table, the error difference with respect to the next four best networks within the arrangements is insignificant for a classification network. However, the differences between the two arrangements are noticeable.

When sorted by the validation error, arrangement B proved to be more convenient, because the average pattern square errors in the validation set were in average about 21 % smaller than in the A arrangement. However, the average pattern square errors in the training set for arrangement A were about 19 % smaller than in arrangement B.

Table 4.3: Five best performing topologies and their performance in arrangements A and B, sorted in descending order of the average pattern square error on the validation set (L: linear activation, S: logsig activation, G: Gaussian activation). Best values are in bold.

	Best 5 topologies	Error on training set	Error on validation set
Arrangement A	2-5-1(L-S-S)	0.08266383	0.01977263
	2-6-1(L-S-S)	0.06727370	0.01983868
	2-8-1(L-S-S)	0.07115250	0.02405285
	2-8-1(L-L-S)	0.09079637	0.02775257
	2-5-1(L-L-S)	0.09733627	0.02964879
	Avg. error	0.08184453	0.02421310

	Best 5 topologies	Error on training set	Error on validation set
Arrangement B	2-10-1(L-S-S)	0.09466055	0.01484767
	2-9-1(L-L-S)	0.09908852	0.01838150
	2-9-1(L-S-S)	0.10115675	0.02043153
	2-3-1(L-S-S)	0.10267141	0.02076315
	2-5-1(L-L-S)	0.10312260	0.02119864
	Avg. error	0.10013997	0.01912450
Ratio		Avg(A) / Avg(B) = 0.8173	Avg(B) / Avg(A) = 0.7898

The conclusion that validation is better for arrangement B proves that omitting patients 8-14 from the training set affects the performance of the detectors when tested on the same patients, which reflects that there are significant variations in the EEG dynamics from the validation files used. Summarizing, sorting all networks by their performance on the training and testing set yielded the ranking given in Table 4.4., best being on top.

Table 4.4: Five best performing topologies on both arrangements when sorted in descending order of the average pattern square error on training and validation set.

	Sorted by performance on training set	Sorted by performance on validation set
Arrangement A	2-6-1(L-S-S)	2-5-1(L-S-S)
	2-8-1(L-S-S)	2-6-1(L-S-S)
	2-5-1(L-S-S)	2-8-1(L-S-S)
	2-8-1(L-L-S)	2-8-1(L-L-S)
	2-4-1(L-S-S)	2-5-1(L-L-S)
Arrangement B	2-10-1(L-S-S)	2-10-1(L-S-S)
	2-9-1(L-L-S)	2-9-1(L-L-S)
	2-9-1(L-S-S)	2-9-1(L-S-S)
	2-3-1(L-S-S)	2-3-1(L-S-S)
	2-8-1(L-S-S)	2-5-1(L-L-S)

Because the validation set is data not used for the weight update, performance on this set was given priority when deciding which topology to select. From Table 4.4, the 2-10-1(L-S-S) topology performs better on arrangement B for the validation set, whereas for arrangement A, the 2-5-1 L-S-S topology performs better on the same set. Referring back

to Figure 4.8, the reader should observe that the plots shown are indeed proof that adding more hidden neurons does not lower errors, because all starting solutions were purposely randomized in order to avoid local minima. Therefore, it can be concluded that there is a possibility that the 2-10-1 topology could have performed better in another 5 trials. But again, due to the uncertainty inherent to the numerical problem at hand, the course of the investigation had to be based on one choice, and this choice was the 2-5-1 topology.

After selecting the two ANNs, they were trained with early stopping on the corresponding arrangements, and then tested on the corresponding testing data for a complete analysis. Next section presents the ROC analysis performed on the testing results.

4.10. Testing Results

After training the 2-5-1 and 2-10-1 L-S-S networks on arrangements A and B, respectively, they were tested on the testing data as well as re-tested on the training and validation data of the corresponding arrangements. Results of the test in ROC terminology are given in Table 4.5, which for simplification purposes, only shows accuracy, sensitivity, and specificity. In this table, keep in mind that there are two networks (one for each arrangement) rather than only one network, thus the three value columns of this table were obtained with one network and the remaining columns with another network with the same topology, but whose weights are different.

Table 4.5: Errors obtained after training and testing the corresponding topologies on the three subsets of arrangements A and B: pure training, cross-validation, and testing. Best values are in bold.

	Performance of 2-5-1 (L-S-S) ANN on the Arrangement A Sets			Performance of 2-10-1 (L-S-S) ANN on the Arrangement B Sets		
	Training	Validation	Testing	Training	Validation	Testing
TP	8	3	22	18	3	15
TN	17	4	93	52	13	48
FP	0	0	2	1	0	2
FN	3	0	5	2	1	2
Accuracy	89.29 %	100.00 %	94.26 %	95.89 %	94.12 %	94.03 %
Sensitivity	72.73 %	100.00 %	81.48 %	90.00 %	75.00 %	88.24 %
Specificity	100.00 %	100.00 %	97.89 %	98.11 %	100.00 %	96.00 %

Table 4.5 shows how the optimum networks perform in each set. A closer observation reveals inconsistency in the behavior of the indicators. For example, while accuracy and sensitivity are lower in the training set of arrangement A as compared to arrangement B, specificity is better. If one turns the attention to the most important set, i.e. the testing set, the following can be noticed:

- accuracy is better in arrangement A
- sensitivity is worst in arrangement A
- specificity is better in arrangement A

This inconsistency made it impossible to change the balance in favor of one arrangement or the other, not even when arrangement A is supposed to be less biased since it included unseen patients as testing data, besides having a testing set which is more than 4 times bigger than the training set. This proved that there is no substantial difference in the gamma power spectrum of the EEG of the seizure between patients and therefore, no specific proposition can be done as to which arrangement to select. It is therefore concluded that the spectrum in the gamma band is an excellent instrument to detect seizures offline at inter-patient level and is not restricted to EEG dynamics of any

particular patient; in other words, it is patient-independent. Further analysis was intentionally done using arrangement A since a test on data from unseen patients is the most adverse condition for a classifier.

Despite the numerical analysis that was performed, neither the number of hidden neurons nor the data arrangement yielded particular results. One should also expect that a particular data arrangement can change network performance significantly. For example, one arrangement with overlap in the training set and perfect separation in the testing set will yield a network that performs much better in the testing than in the training set. For all these reasons, it was opted to disregard the topology/arrangement analysis during the study of the remaining frequency bands, and use only the 2-5-1 L-S-S topology within arrangement A.

The reader will notice that the analysis is done qualitatively rather than using any particular indicator. A complete analysis could only be done if the classifiers were tested on different data sets and under different circumstances, which would allow generating ROC curves for each one and compute the area under the curves as an overall performance criterion.

Considering that this study is applied to clinical data that involves seizures from different patients, the performance of this ANN-based seizure detector is within acceptable standards. The selection of the adequate arrangement could be only achieved with the features and computation speed of the NS tool.

4.11. Comparative Analysis under Different Frequency Bands

In order to demonstrate the superiority of the spectral power in the gamma range, other frequency bands were also analyzed and compared in terms of their classification performance. Six 2-5-1 ANNs were constructed, trained from random starting solutions and tested in their corresponding data sets. The gamma-based classifier was also trained again. From Table 4.6 it can be observed that the gamma band yielded the most accurate seizure detector.

Table 4.6: Testing errors obtained after training a 2-5-1 topology on the data set extracted for each frequency band. Best values are in bold.

	Alpha	Beta I	Beta II	Delta	Gamma	Theta
TP	17	19	23	16	25	14
TN	93	90	90	94	92	94
FP	2	5	5	1	3	1
FN	9	6	3	11	2	11
Accuracy	90.91 %	90.83 %	93.39 %	90.16 %	95.90 %	90.00 %
Sensitivity	65.38 %	76.00 %	88.46 %	59.26 %	92.59 %	56.00 %
Specificity	97.89 %	94.74 %	94.74 %	98.95 %	96.84 %	98.95 %

In comparing the results obtained for the gamma frequency band, it was noted that equivalent methods developed in recent years have a tendency to use sensitivity as performance criterion. [Navakatikyan et al. 2006] apply a wave-feature extraction algorithm to scalp EEG from 55 neonates and obtains sensitivity ranging 83 % to 95 %. [Hopfengärtner et al. 2009] attain 85.2 % to 90.8 % analyzing two frequency bands of intracranial EEG of 15 patients. Another method by [Chan et al. 2008] uses SVM on features extracted from five frequency bands, resulting in sensitivities between 80% and 98%. [Deburchgraeve et al. 2008] attain 88 % sensitivity when applying correlation and frequency activity to scalp EEG of 26 neonatal patients. Also an algorithm by [Schad et al. 2008] reports sensitivities up to 63 % and 72 % at scalp and intracranial level,

respectively. When compared to these methods, the 92.59 % sensitivity displayed in Table 4.6 for the gamma frequency band of the method proposed in this dissertation is more than satisfactory.

At this juncture it is important to note that even though different ROC terms are listed in Table 4.6, one should consider accuracy as one of the most important descriptor since it measures the degree in which correct classifications were done without favoring either TP or TN classifications. Favoring TP assures high sensitivity, whereas by favoring TN, one would opt for higher specificity. Note that even though delta yielded best specificity, it performs poorly in terms of sensitivity. It can be concluded that in general the gamma frequency range performed better across the board in terms of accuracy, sensitivity and specificity over the remaining frequency bands.

4.12. Conclusion of the Method

This study demonstrated the feasibility of detecting seizures on intracranial EEG using an ANN based on the gamma frequency band for classification purposes. It was shown that the power measurement in the gamma range from 36 to 44 Hz contains the information needed to discriminate seizure files from non-seizure files with an accuracy of 95.90%, a sensitivity of 92.59%, and a specificity of 96.84%.

These results were obtained with a 2-5-1 network topology with linear activation functions in the input layer and sigmoid activation functions in the 2nd and 3rd layer. The two most discriminating features that constituted the 2-D decisional space were determined to be : (1) The time duration (the number of consecutive points) where the

value of each given point in the inter-electrode mean signal S_{μ} exceeded the set statistical threshold $T = \overline{S_{\mu}} + \sigma$, and (2) The maximum value of S_{μ} in that specific interval.

Of particular value is the generalized nature of the algorithm, and its feasibility in the absence of patient-specific training data. This feature is demonstrated most clearly by the test characteristics in patients who were not part of the training data. In general it is worth mentioning that although only 29% (35/122) of the files were used randomly for training the detector, high measures in sensitivity, specificity, and accuracy were still achieved in the remaining files which were subsequently used for testing.

By investigating as to what constitute the optimum number of hidden units and which arrangement of data will be more useful, the following interesting observations can be made:

- (1) No particular number of hidden neurons could be found to minimize the testing error. The reason probably lies in the nature of the training algorithm, which is a numeric method that starts with random solutions.
- (2) Excluding patients from the training set did not have significant impact in the test. This fact clearly proved that the gamma-based classifier is effective even when tested on unseen patients.

The computational requirements for creating the ANNs during the training phase and the ensuing results during the testing phase reveal additional findings that are quite interesting:

- (1) the data clusters of seizure-files seem more spread out than those data clusters of files without seizures, which clearly proves that seizures which are atypical events obviously vary greatly among subjects.
- (2) the choice of the ANN is critical for delineating such wide-ranging behaviors, as long as the appropriate decisional space is cautiously established.

The study has so far included 14 patients who underwent two-stage epilepsy surgery with subdural recordings, and whose iEEG data were obtained sequentially. More insight will be gained into the findings of this study as more patients in the future will consent to be included, since the more data is collected, the more influential will be the clustering characteristics of epileptogenic data.

Another interesting question that can be posed, along the same reasoning adopted thus far, is whether it can be determined if an EEG recording can be associated to epileptic subjects or to non-epileptic persons who have never experienced an epileptic seizure. In the next chapter, a method is proposed to address this problem through a newly designed algorithm and applying the same NS tool. The intent of such a new approach is to position this research towards the ultimate goal of seizure prediction.

CHAPTER 5

5. AN APPLICATION OF NEURALSTUDIO TO THE CLASSIFICATION OF EPILEPTIC AND NON-EPILEPTIC SUBJECTS FROM SCALP EEG

5.1. Objectives of the Method

The main objective of this method was to associate EEG file segments to non-epileptic or epileptic subjects by means of an ANN. The purpose was to identify potential epileptic subjects based on aggregated features before running any algorithm to detect interictal spikes.

With the empirical experience thus gained, the research method of this chapter will be focused on the investigation of following EEG features:

- activity
- mobility
- complexity
- spectral power in each of the frequency bands (delta, theta, alpha, beta I, beta II, and gamma)

The ANN classifier will analyze EEG file segments in text format and classify them as “non-epileptic” or “epileptic” to provide an automatic evaluation to medical doctors. For this purpose, a method will be proposed to extract features from EEG segments in such a way that they are independent of the number of electrodes and the duration of the recordings.

5.2.Data Preprocessing

For this study, scalp EEG recordings from three non-epileptic subjects and three epileptic patients were used. For an efficient comparison of files, the segments from the epileptic patients were extracted from sections without seizures. Recordings were performed at Miami Children’s Hospital using XLTEK Neuroworks Ver. 3.0.5 equipment and sampled at 512 Hz and 500 Hz for non-epileptic subjects and epileptic patients, indistinctly. Electrode setup followed the 10-20 system.

In total, 10 to 30 segments were extracted from each recording. To prevent biasing the analysis, all segments were free of artifacts and mostly without observable epileptiform activity. As mentioned above, the file segments from epileptic patients contained only interictal data, i.e. they did not have seizures. All files were 4 to 45 seconds long. An overview of the subjects’ information is given in Table 5.1.

Table 5.1: Subjects’ information and data used

Subject	Status	Age	Gender	Diagnosis	Number of files	Sampling rate (Hz)
1	Non-epileptic	15	Female	-	10	512
2	Non-epileptic	10	Male	-	15	512
3	Non-epileptic	47	Male	-	20	512
4	epileptic	7	Male	Focal area of epileptogenesis within the left frontal region involving the frontal pole and posterior frontal lobe near the midline region primarily.	20	512
5	epileptic	4	Female	Intractable seizures over the right frontocentro-temporal epileptogenicity	30	500
6	epileptic	15	Female	Medically intractable seizures with epileptogenicity involving both frontal regions	10	500

5.3.Feature Extraction

As stated earlier, EEG recordings can be analyzed and compared to each other by extracting features in time and frequency domains. One problem associated with feature extraction is that EEGs from different subjects and recorded at different times can have different numbers of electrodes. Furthermore, these electrodes also can be arranged in different locations and could be placed either on the scalp or at the subdural level. This fact makes pair-to-pair electrode recording comparison impossible.

The standard way of feature extraction in time domain when different electrodes are present is illustrated in Figure 5.1(a). For each time window, the entire window recording is replaced by a particular feature of interest. In the example, Hjorth's parameter "activity" is extracted out of 1 second windows for each electrode.

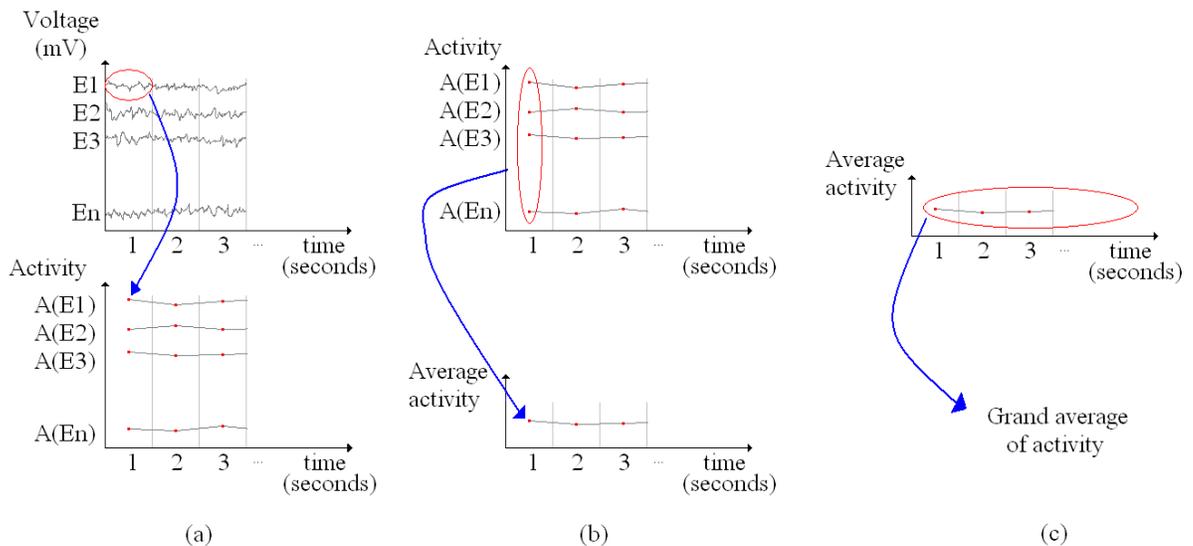


Figure 5.1: (a) First step in feature extraction from EEG recordings, illustrating the extraction of EEG activity from 1-second time windows for a set of electrodes. Window activity $A(E_k)$ is computed for each electrode E_k for each time window. (b) Second step in feature extraction, using inter-electrode average as statistical parameter to analyze EEG activity in time independent of the number of electrodes. (c) Last step of the proposed feature extraction procedure, using grand average over time of the inter-electrode average as overall parameter to analyze EEG activity of a specific EEG recording file.

As can be seen from Figure 5.1(a), extracting electrode features has the disadvantage of generating a set of values which depends on the number of electrodes. To cope with this issue, statistical parameters such as average, standard deviation, and signal to noise ratio can be computed across all electrode features for each time window. This is depicted in Figure 5.1(b), using the grand average across all electrodes as a representative feature.

Since the purpose of this method is the analysis of EEG recordings regardless of how much time they take, further statistical parameter extraction can be performed along the time axis to provide a single feature descriptor for the whole EEG file. Figure 5.1(c) illustrates this last step using the grand average of the inter-electrode average activity.

In general, all steps of the global feature extraction procedure for any particular feature are illustrated in Figure 5.2, where it can be observed that the entire EEG segments can be described by a single global value for the specified feature.

In Figure 5.2, Param1 represents a statistical parameter such as average or standard deviation, which is applied to all electrodes within a time window. Likewise, Param2 is also a statistical parameter such as average or standard deviation which is applied to the time series of the aforementioned parameter Param1.

The statistical parameters chosen at this point to extract global features from EEG file segments are the average (AVG), standard deviation (STD) and signal to noise ratio (SNR) as shown in Figure 5.3.

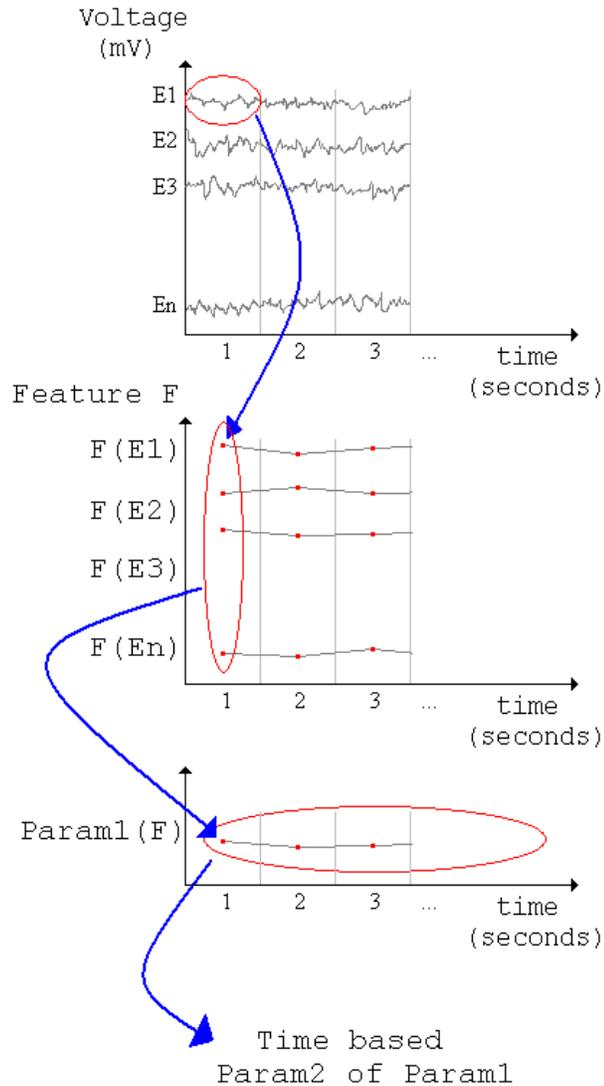


Figure 5.2: Compilation of all steps needed to extract a global descriptor for a particular feature for an entire EEG file. Param1 is an inter-electrode statistical parameter, and Param2 is a statistical parameter of the time series of Param1.

These parameters will be extracted across electrodes as well as over time. To differentiate, uppercase is used for the inter-electrode parameters, and lower case is used for time series parameters. Since there are three parameters across electrodes and another three for each resulting time series, there are in total nine parameters, as shown in Table

5.2, starting with the avg(AVG), i.e., the average over time (avg) of the inter-electrode average (AVG), and ending with snr(SNR).

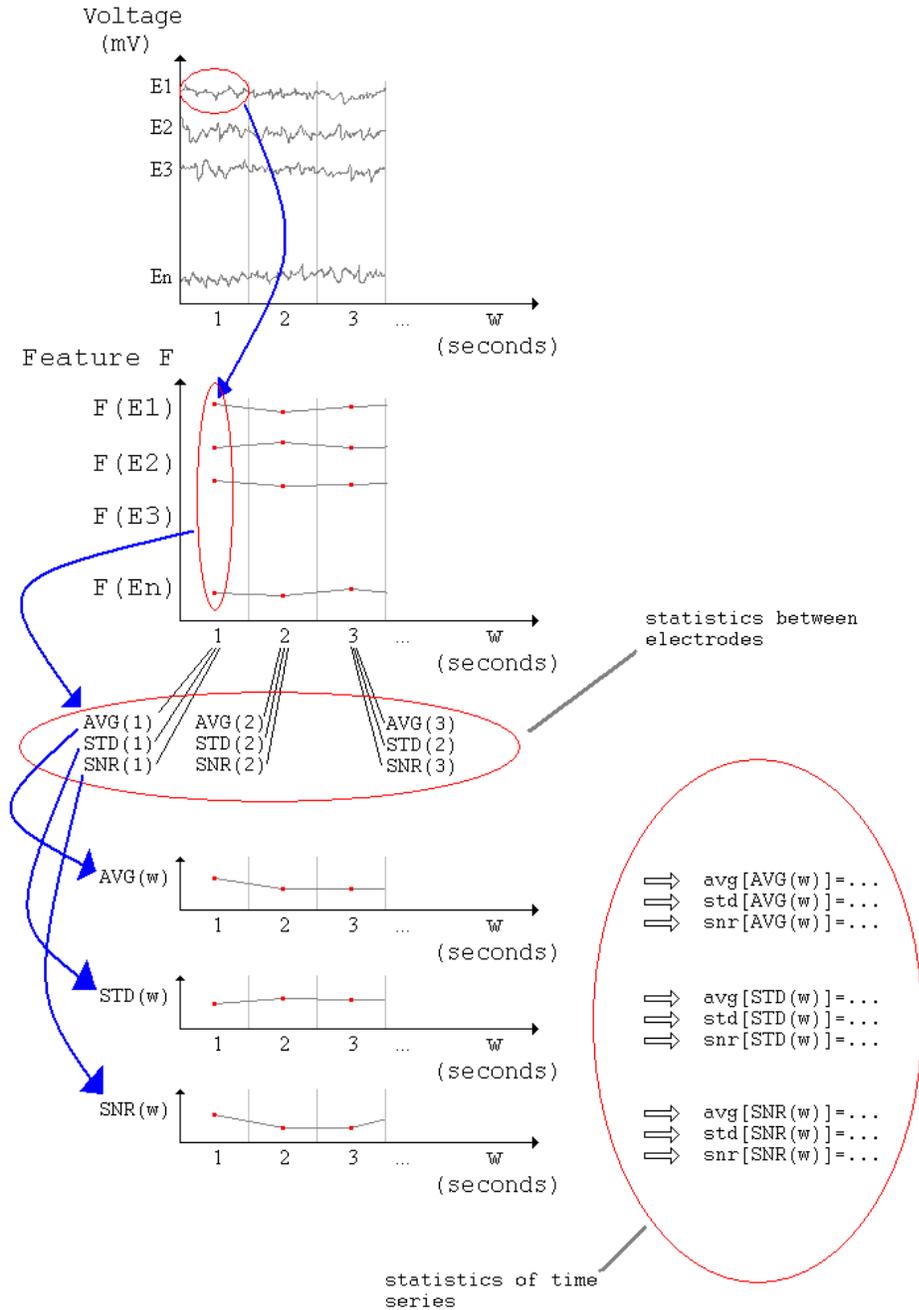


Figure 5.3: Proposed approach to feature extraction from EEG data segments, reduced to a 3x3 matrix obtained from the common statistical parameters avg, std and snr.

Table 5.2: Segment descriptor matrix for a particular feature

	avg	std	snr
AVG	avg(AVG)	std(AVG)	snr(AVG)
STD	avg(STD)	std(STD)	snr(STD)
SNR	avg(SNR)	std(SNR)	snr(SNR)

With this approach, whole EEG segments will be assigned a feature descriptor matrix for each particular feature. For example, activity, mobility and complexity will each have a corresponding segment descriptor matrix. The same applies to the spectral power in the different frequency bands. In total, for the 9 features that will be focused on, there are 9 matrixes, or $9 \times 9 = 81$ variables that will be used to classify EEG segments.

In general, besides avg, std and snr, any other type of statistical parameter could have been used (such as skewness and kurtosis), because it was not known beforehand which one would be best for discriminating EEG segments. However, for simplification purposes, the analysis was limited to these three parameters. The two dimensions of the global feature descriptor come from the fact that both electrode and timing information are represented by statistical parameters. Likewise, if one were to compute a descriptor for an entire group of subjects with a specific health condition, a third dimension would be required, as is shown in Figure 5.4. Generalizing, if the number of statistical parameters n is kept the same in each of the d dimensions, then the number of elements of the descriptor would amount to n^d .

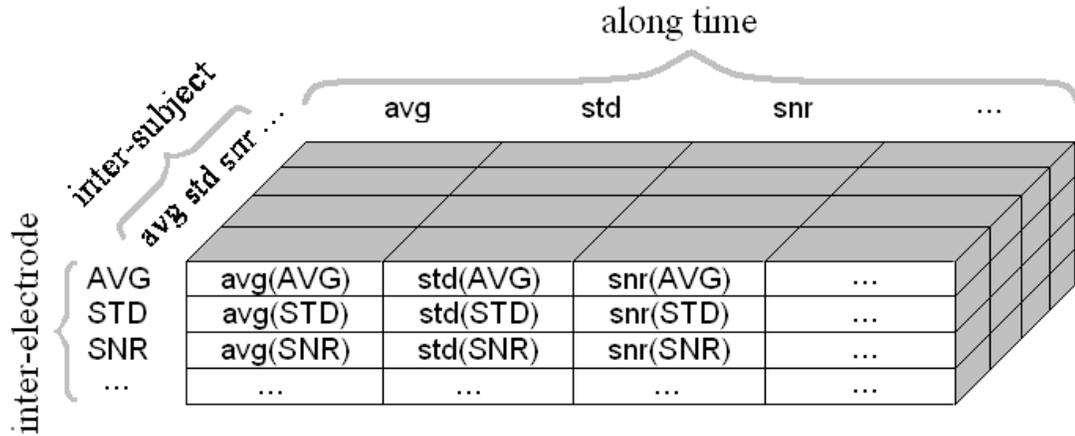


Figure 5.4: A generalization of the global descriptor with three dimensions (feature cube): one vertical for the inter-electrode aggregation, one horizontal for the time aggregation, and another for the statistics across subjects. Keep in mind that additional parameters can be used.

Following the procedure previously explained, the descriptor matrices for all 105 files and each feature were obtained using a general window size of 1 sec in each file. equipped with these descriptors, a classifier for each feature could be created. The next section will explain in detail how the inputs of the classifiers were determined.

5.4. Dimensionality Analysis

Before creating the classifiers, the descriptor matrices of all files and features were put in tables with a structure as shown in Table 5.3 containing 9 columns. The first column corresponded to avg(AVG), the second to std(AVG) and so on, up to the ninth, which has the value for snr(SNR). The last column was reserved for the target, codified as -1 for files of non-epileptic subjects and 1 for files of epileptic patients. This setup allows the global feature of each EEG file to be represented by a 9-dimensional point. In total, 9 tables were obtained, one for each feature (activity, mobility, complexity, and spectral

power in delta, theta, alpha, beta I, beta II and gamma). Each table contained 105 rows (45 for non-epileptic subjects and 60 for epileptic patients).

Table 5.3: Common header structure of training, cross-validation and testing tables, including the nine statistical parameters proposed and the target column

File #	avg(AVG)	Std(AVG)	snr(AVG)	avg(STD)	std(STD)	snr(STD)	...
1
2	=>
3
...

...	avg(SNR)	std(SNR)	snr(SNR)	Target
...
=>
...
...

The tables obtained would require the use of 9-dimensional classifiers. A first attempt to deal with this data was to question whether all 9 dimensions were necessary so as to reduce data dimensionality. At this juncture, all tables were displayed in 2D plots by using all 36 possible pairs of dimensions for the 9 dimensions that were considered. Figure 5.5 shows all plots for the spectral power in gamma. The reason why the gamma feature is shown is because of the suitability of this frequency range for seizure detection purposes, as was proved in the previous Chapter.

In Figure 5.5, files from non-epileptic subjects are represented by blue dots and files from epileptic patients are represented by red crosses. From the first view, plots involving dimensions 3, 6, 7, 8, 9 in Figure 5.5 show in general a wider spread of clusters and better separation. Most of these clusters are somewhat organized and separated from each other.

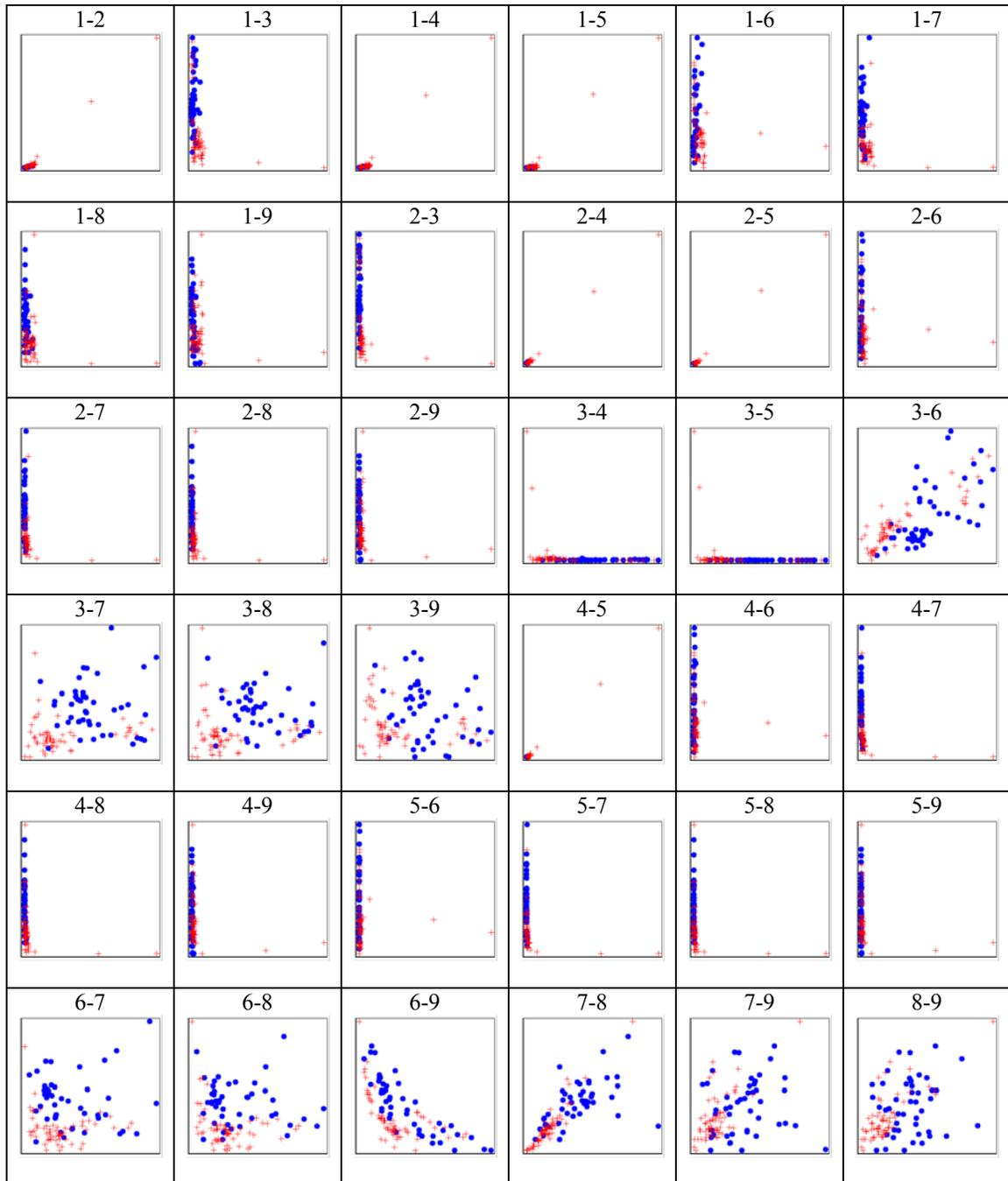


Figure 5.5: Plots of the gamma descriptors for all 105 files together, showing combinations of dimension pairs (dimension1-dimension2) with different clustering abilities. Files from non-epileptic and epileptic subjects are represented by blue dots and red crosses, respectively.

However, because of the presence of outliers, it was suspected that the remaining plots could still show clear clusters if they were properly magnified. Figure 5.6 shows the

original 1-2 plot and the zoom in that excludes the two outliers in the center and the upper right.

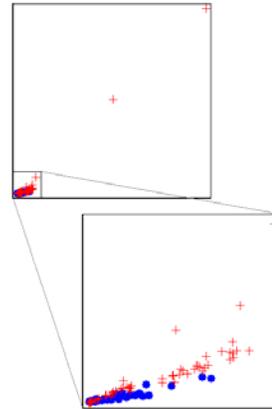


Figure 5.6: Zoom in of plot 1-2 of Figure 5.5, showing a better cluster organization once outliers have been excluded.

Figure 5.6 clearly shows better cluster organization after outliers were excluded. Plots of the remaining features are not provided here for space consideration, but closer examination revealed the same issue. This fact proved that the plots cannot be used to find suitability of axes. Therefore, the study had to include all nine dimensions and leave it up to NeuralStudio to find the best classifier.

Another reason why all nine dimensions were kept is because the accuracy obtained by the 9D classifier cannot be lower than the highest accuracy obtained by any of these 36 2D classifiers. For example, the 8-9 plane in the gamma descriptor plot of Figure 5.5, despite having high overlap, does not affect the overall accuracy of the 9D classifier, which in turn can be higher than the maximum accuracy of any of the 36 2D classifiers. Moreover, if one removes a dimension, despite having too much overlap, the accuracy of the 9D classifier could be reduced.

Once the dimensionality of the problem was established, training was performed. Details are provided in the next section.

5.5. Training Setup

5.5.1. Data Organization

After determining that all nine dimensions would have to be used, all files were selected for training, cross-validation or testing. For this study, about half of the files were reserved for training and the other half for testing, whereas when possible, about 20 % of the training files were kept for cross-validation, as shown in Figure 5.7. In this setup, distribution was done so that each subject had approximately the same balance of files for training, cross-validation, and testing, as can be observed in Table 5.4.

The procedure of data selection for training, cross-validation and testing is automated in NS to allow fast preparation of files. A snapshot of the pattern window is illustrated in Figure 5.8. By selecting the “Tools/Management/Split Table” menu option, a panel is visualized on the right, containing instructions so as to step-by-step separate all data available into training, cross-validation, and testing tables.

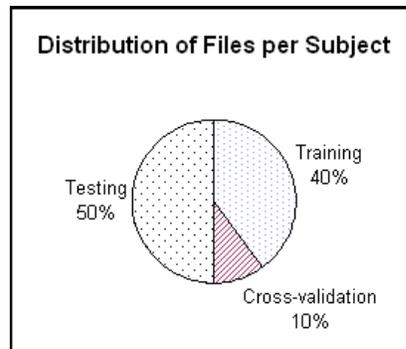


Figure 5.7: Ratio of file usage during classifier training. This ratio was used for all subjects.

Table 5.4: Data arrangement for training purposes

Subject	File usage ratios	Training Files	Cross-validation files	Testing Files	Total
Subject 1 (non-epileptic)	4/1/5	First 4	Next 1	Remaining 5	10
Subject 2 (non-epileptic)	7/1/7	First 7	Next 1	Remaining 7	15
Subject 3 (non-epileptic)	8/2/10	First 8	Next 2	Remaining 10	20
Subject 4 (epileptic)	8/2/10	First 8	Next 2	Remaining 10	20
Subject 5 (epileptic)	12/3/15	First 12	Next 3	Remaining 15	30
Subject 6 (epileptic)	4/1/5	First 4	Next 1	Remaining 5	10
Total		43	10	52	105

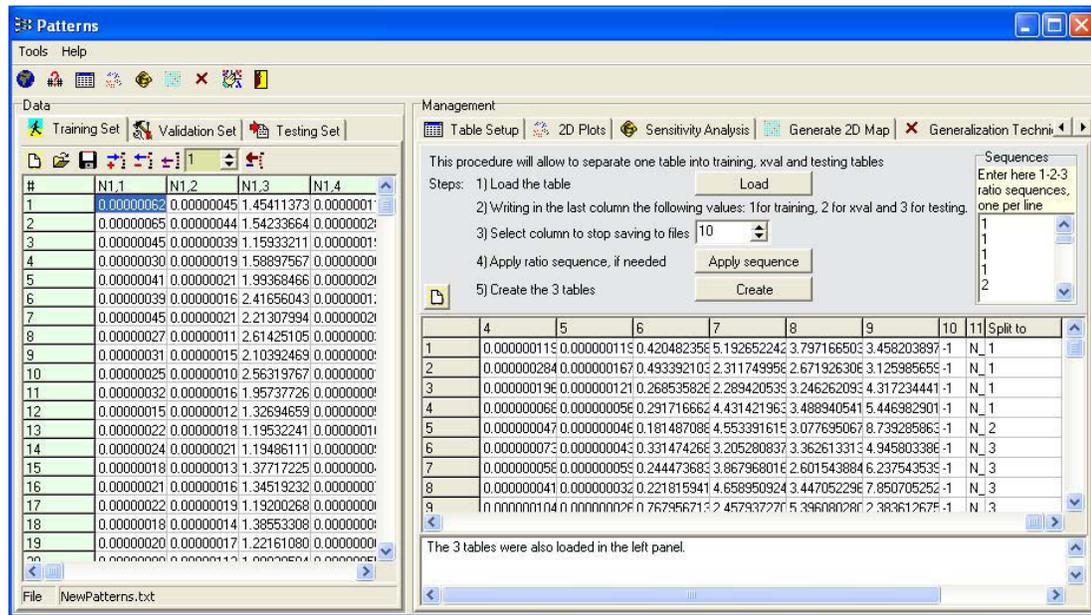


Figure 5.8: Automation of table separation in NeuralStudio. The user can load the whole data into a table and from there make a selection, row by row, of where to send each row. By entering 1, 2, or 3 in the last column provided, the rows can be sent automatically to the training, cross-validation and testing tables, respectively.

By using this feature of NS, the tedious task of opening four text files at once (the whole table and the three tables for training) and performing manual distribution of rows across the 4 files was fully automated in order to avoid human errors.

5.5.2. ANN Configuration and Training Procedure

For the problem at hand, a 9-19-1 topology was selected. The reason why 19 and no other number of hidden units was selected was because of the experiments that have been reported in section 4.1, which indicated that an optimization of the number of hidden units was unpractical. A well-known rule of thumb was used, which advises to use twice the number of input units, plus one more.

The input units were assigned linear activation functions, and the hidden and output units were assigned logsig functions, with the same parameters that were used in section 4.1.

The cross-validation strategy implemented in this case was similar to the one applied in section 4.1. Training was stopped once an increment in the cross-validation set for 5 consecutive times was detected. Under normal conditions, a training error can be 50% higher or lower than the previous one. By allowing the error no more than 5 consecutive times to increase, one is assuming that the error will keep increasing with a probability of $1 - (0.5)^5 = 1 - 0.03125 = 0.96875$, which is a high-confidence value. The learning rate was set in all ANNs to 0.0001.

Because the minimum error obtained by an ANN depends on the starting condition, there is no way to know in advance which will be the best set of weights that will lead the network to a global minimum. The only way is to perform several trials, each starting with random weights, and to store the set of weights that yielded the best solution. This study opted to find the best solution for each ANN after a number of trials.

Each ANN was trained 10 times in different trials as follows:

- 4 times (each 10 repetitions), with 1 minute maximum for each repetition, generating 4 best solutions
- 2 times (each 10 repetitions), with 2 minutes maximum each repetition, generating 2 best solutions
- 1 time (each 10 repetitions), with half a minute maximum each repetition, generating 1 best solution
- 1 time (each 10 repetitions), with 15 seconds maximum each repetition, generating 1 best solution

The intent of this approach with so many repeated trainings is to rule out the possibility that any differences in the results are just by chance (analogous to the t-Test).

Since cross-validation was performed, all training repetitions were stopped either by the cross-validation stop criterion or by reaching the time limit that was set. The procedure was applied consistently across all 9 ANNs. In general, each ANN was allocated a time span of $40 \cdot 1 + 20 \cdot 2 + 10 \cdot (1/2) + 10 \cdot (1/4) = 87.5$ minutes for training, thus totaling 13 hours and 7.5 minutes for training all 9 ANNs.

The reason why 15 seconds of maximum duration was also considered was to avoid the testing error to become too large. As it is known, cross-validation stops only when the validation error starts to increase, as shown in Figure 5.9. But before reaching that point, the error on the testing set can either increase or decrease, although most of the times this error is expected to follow the trend of the cross-validation error, at least for the first few iterations. In this experiment, this tendency was observed in most trainings iterations, but

not in all. Because the training algorithm has purposely no feedback from the testing error, limiting the search time is a good way to avoid high testing errors.

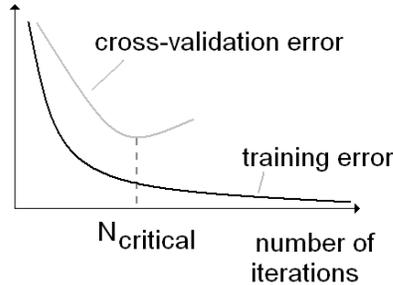


Figure 5.9: Illustration of the cross-validation criterion for training stop. The point where the network starts memorizing and loosing its generalization ability is signaled by $N_{critical}$. The testing error can have any behavior but is not part of the stop procedure.

Figure 5.10 shows the panel in the training window where all ANN files can be loaded at once and trained repeated times.

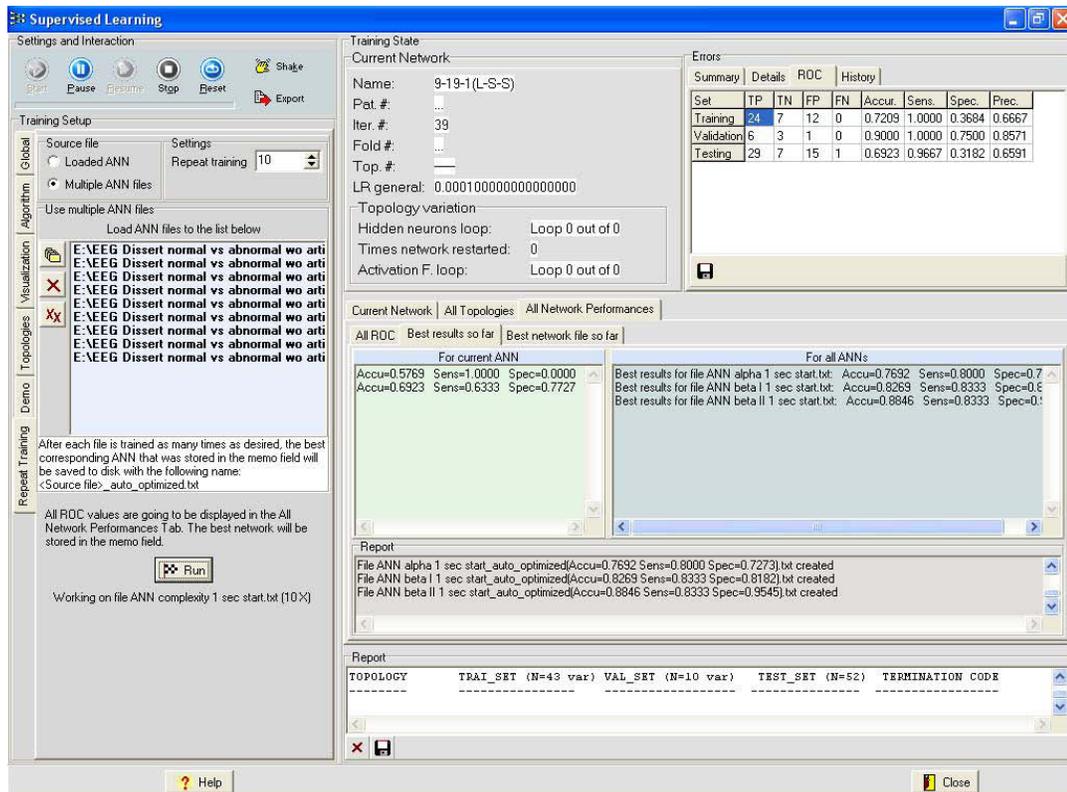


Figure 5.10: Screenshot of NeuralStudio’s training window, showing a panel that allows the user loading multiple ANN files at once and train each of them several times. In this process, the best networks are automatically stored in the interface and also saved to disk.

During training, each time the performance was better than before, the solution was stored in the interface as the best one so far. After training the networks as many times as defined, the best performing ANN was stored and properly named for later use.

5.6. Testing Results

The discussion about the suitability of the descriptors will focus exclusively on the results obtained in the testing set as this was the set that was not used for weight update.

Before proceeding with the analysis, it is important to recall that these classifiers could not be tested on other data than the testing data that was available for the study. For that reason, ROC curves can not be constructed to find the best classifier.

The difficulty associated with the selection of the best descriptor is because there are in total three performance metrics, namely accuracy, sensitivity and specificity. Although accuracy is a widely accepted metric, medical doctors tend to favor sensitivity and specificity. In this section, the intention was to include all terms at once. At this juncture, an approach different to the one implemented in section 4.1 was followed. The solution is a MaxMin procedure in which the best candidate descriptor was considered the one whose lowest performance metric was the highest among all descriptors, and which, at the same time, has the highest possible performance metric among others of similar or higher minimum performance. This procedure is represented in Table 5.5.

Table 5.5: Procedure implemented to generate the 10 solutions per network after random starts. Accuracy, sensitivity and specificity are represented by X1, X2, and X3, respectively.

(1)	(2)	(3)	(4)	(5)
Repetition number	X1	X2	X3	Minimum of Columns (2) to (4)
1
2
3
4
5
6
7
8
9
10

Maximum of column (5)
= best solution of trial

With this criterion at hand, 8 trials were performed for each network, as explained in the previous section. In each trial the network was trained 10 times and the best solution was stored for further analysis. The best solutions for each trial are presented in Table 5.6.

To be consistent with the criterion applied to find each one of the solutions presented in Table 5.6, the same criterion was used to select the best descriptor from the table. It can be easily found that the highest minimum corresponds to activity with a minimum of 0.9667 and a maximum of 1.000 followed by delta with a minimum of 0.9000 and a maximum of 0.9091. These values are highlighted. However, further investigation was needed at this point in order to know how much certainty there was to affirm that the activity is much better than the other descriptors.

Table 5.6: Best solutions found in all trials. Within each group, the best solution is in bold.

	Accuracy	Sensitivity	Specificity	Minimum	Maximum
Activity	0.8846	0.9333	0.8182	0.8182	0.9333
	0.9615	0.9667	0.9545	0.9545	0.9667
	0.9423	0.9333	0.9545	0.9333	0.9545
	0.9615	0.9667	0.9545	0.9545	0.9667
	0.9615	0.9667	0.9545	0.9545	0.9667
	0.9808	0.9667	1.0000	0.9667	1.0000
	0.9423	0.9667	0.9091	0.9091	0.9667
	0.9423	0.9667	0.9091	0.9091	0.9667
Mobility	0.7885	0.8000	0.7727	0.7727	0.8000
	0.8846	0.8000	1.0000	0.8000	1.0000
	0.8846	0.8333	0.9545	0.8333	0.9545
	0.8462	0.8667	0.8182	0.8182	0.8667
	0.8462	0.9000	0.7727	0.7727	0.9000
	0.8462	0.8333	0.8636	0.8333	0.8636
	0.8269	0.8667	0.7727	0.7727	0.8667
	0.7115	0.7333	0.6818	0.6818	0.7333
Complexity	0.7885	0.8667	0.6818	0.6818	0.8667
	0.7885	0.8333	0.7273	0.7273	0.8333
	0.7885	0.8000	0.7727	0.7727	0.8000
	0.7115	0.7000	0.7273	0.7000	0.7273
	0.7500	0.7667	0.7273	0.7273	0.7667
	0.7308	0.8333	0.5909	0.5909	0.8333
	0.7308	0.7667	0.6818	0.6818	0.7667
	0.6923	0.6333	0.7727	0.6333	0.7727
Delta	0.9231	0.9667	0.8636	0.8636	0.9667
	0.8462	0.9000	0.7727	0.7727	0.9000
	0.8462	0.8667	0.8182	0.8182	0.8667
	0.9038	0.9333	0.8636	0.8636	0.9333
	0.8654	0.9000	0.8182	0.8182	0.9000
	0.8846	0.9000	0.8636	0.8636	0.9000
	0.9038	0.9000	0.9091	0.9000	0.9091
	0.9231	1.0000	0.8182	0.8182	1.0000
Theta	0.7885	0.7333	0.8636	0.7333	0.8636
	0.7308	0.6667	0.8182	0.6667	0.8182
	0.6154	0.6000	0.6364	0.6000	0.6364
	0.7308	0.6667	0.8182	0.6667	0.8182
	0.6923	0.6333	0.7727	0.6333	0.7727
	0.7308	0.6333	0.8636	0.6333	0.8636
	0.6731	0.6667	0.6818	0.6667	0.6818
	0.7115	0.7000	0.7273	0.7000	0.7273
Alpha	0.7115	0.7000	0.7273	0.7000	0.7273
	0.7692	0.8000	0.7273	0.7273	0.8000
	0.7885	0.8333	0.7273	0.7273	0.8333
	0.8077	0.8667	0.7273	0.7273	0.8667

	Accuracy	Sensitivity	Specificity	Minimum	Maximum
	0.7885	0.8667	0.6818	0.6818	0.8667
	0.7692	0.8000	0.7273	0.7273	0.8000
	0.7885	0.8667	0.6818	0.6818	0.8667
	0.7692	0.8000	0.7273	0.7273	0.8000
Beta I	0.8846	0.9000	0.8636	0.8636	0.9000
	0.7885	0.7667	0.8182	0.7667	0.8182
	0.8269	0.8000	0.8636	0.8000	0.8636
	0.8269	0.8000	0.8636	0.8000	0.8636
	0.8269	0.8667	0.7727	0.7727	0.8667
	0.8077	0.8000	0.8182	0.8000	0.8182
	0.8077	0.7667	0.8636	0.7667	0.8636
	0.8269	0.8333	0.8182	0.8182	0.8333
Beta II	0.9038	0.8333	1.0000	0.8333	1.0000
	0.8269	0.7667	0.9091	0.7667	0.9091
	0.8846	0.8333	0.9545	0.8333	0.9545
	0.8846	0.8333	0.9545	0.8333	0.9545
	0.9231	0.8667	1.0000	0.8667	1.0000
	0.8846	0.8000	1.0000	0.8000	1.0000
	0.8846	0.8000	1.0000	0.8000	1.0000
	0.8846	0.8333	0.9545	0.8333	0.9545
Gamma	0.7500	0.7667	0.7273	0.7273	0.7667
	0.7885	0.8000	0.7727	0.7727	0.8000
	0.8077	0.8333	0.7727	0.7727	0.8333
	0.8077	0.7667	0.8636	0.7667	0.8636
	0.8462	0.8667	0.8182	0.8182	0.8667
	0.8077	0.8000	0.8182	0.8000	0.8182
	0.8462	0.8667	0.8182	0.8182	0.8667
	0.8269	0.8667	0.7727	0.7727	0.8667

5.7. Statistical Analysis of the Results

To overcome the aforementioned limitation regarding the inability of conducting ROC analysis, a large number of trainings stages was performed for each descriptor. Repeating the training with the same data made sense since training was always started from random solutions so as to have better changes not to always fall in local minima. By generating a large number of solutions, statistical tests can be applied to validate the

hypothesis that the activity descriptor is better than the others for the classification problem at hand.

At first glance, it can be concluded that the activity is by far the most appropriate descriptor for the classification problem discussed here, with less than 5% error in the metrics. However, to rule out the possibility that the results are just by chance, a t-Test was applied on the minimum metrics, which revealed that it is above 99.99 % (1 - 1.836425×10^{-6}) certainty to say that the activity is indeed better than the remaining descriptors. Table 5.7 displays all results of this test. In other words, the hypothesis that the difference that was found in the performance between activity and the other descriptors can be ascribed to chance alone can now be rejected.

Table 5.7: t-test results that prove that the minimum metric found for the activity descriptor is significantly higher than the remaining metrics. For a 0.05 significance, the critical t-value is 1.795885, but the computed t-value was much higher (8.495415) representing a one-tail p-value of only 1.836425×10^{-6} . $P(x \geq t \text{ Critical})$ is the probability that an observation x is higher than t Critical if the null hypothesis is true, whereas the null hypothesis is that the minimum metrics for activity and the remaining descriptors do not differ in average. The test assumes unequal variances.

	<i>Activity</i>	<i>Remaining descriptors</i>
Mean	0.924988	0.761266
Variance	0.002333	0.005106
Observations	8	64
Pooled Variance	0.004829	
Hypothesized Mean Difference	0	
Degrees of Freedom	70	
t Statistics	8.495415	
$P(x \geq t \text{ Critical})$ one-tail	1.836425×10^{-6}	
t Critical one-tail	1.795885	

To further examine if the difference in the minimum metrics obtained between the descriptors is not just by chance, ANOVA (analysis of variance) test was applied, with results displayed in Table 5.8. In this case, the test yielded that the descriptors are different with a certainty degree of almost 100 % (1 - 1.42×10^{-19}) as depicted in Table

5.8. These results of the ANOVA test are not surprising considering the clear variation in the mean of the nine descriptors as can be observed from Figure 5.11.

Table 5.8: ANOVA test results that prove that the variance of the minimum metrics between the descriptors is not different by chance (df: degrees of freedom, SS: sum of squares, MS: mean of squares or variance)

SUMMARY

<i>G</i>	<i>roups</i>	<i>Count</i>	<i>Sum</i>	<i>Average</i>	<i>Variance</i>
	Activity	8	7.3999	0.924988	0.002333
	Mobility	8	6.2847	0.785588	0.002442
	Complexity	8	5.5151	0.689388	0.00327
	Delta	8	6.7181	0.839763	0.001604
	Theta	8	5.3	0.6625	0.001726
	Alpha	8	5.7001	0.712513	0.000448
	Beta I	8	6.3879	0.798488	0.001046
	Beta II	8	6.5666	0.820825	0.000932
	Gamma	8	6.2485	0.781063	0.000918

<i>Source of Variation</i>	<i>SS</i>	<i>df</i>	<i>MS</i>	<i>F</i>	<i>P-value</i>	<i>F crit</i>
Between Groups	0.42562	8	0.053203	32.53271	1.42E-19	2.089185
Within Groups	0.103027	63	0.001635			
Total	0.528648	71				

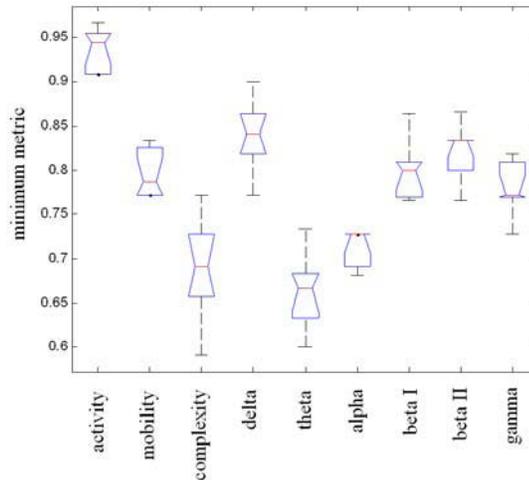


Figure 5.11: Variation of the nine descriptors. The box plot suggests the size of the F-statistic and the p-value. Large differences in the center lines of the boxes correspond to large values of F and correspondingly small values of p.

Table 5.9 shows all descriptors sorted in descending order of their performance.

Table 5.9: All descriptors, sorted in descending order of their minimum performance for EEG classification.

	Accuracy	Sensitivity	Specificity	Minimum	Maximum
Activity	0.9808	0.9667	1.0000	0.9667	1.0000
Delta	0.9038	0.9000	0.9091	0.9000	0.9091
Beta II	0.9231	0.8667	1.0000	0.8667	1.0000
Beta I	0.8846	0.9000	0.8636	0.8636	0.9000
Mobility	0.8846	0.8333	0.9545	0.8333	0.9545
Gamma	0.8462	0.8667	0.8182	0.8182	0.8667
Complexity	0.7885	0.8000	0.7727	0.7727	0.8000
Theta	0.7885	0.7333	0.8636	0.7333	0.8636
Alpha	0.7885	0.8333	0.7273	0.7273	0.8333

As clearly shown in Table 5.9, the difference in performance between activity and delta is remarkable. Then, beta II and beta I have almost the same minimum metrics, followed by mobility. Gamma, as opposed to the results presented in section 4.1, performs poorly in classifying the EEG, and is followed by complexity and theta. The worst descriptor turned out to be alpha.

Note that considering the fact that the input for the statistical tests are best solutions that were obtained after a total of 80 trials for each descriptor, such values follow an extreme value distribution and have therefore smaller STD. In this study, where the minima were obtained out of 3 values, it was experimentally found that the STD of the distribution of the minima was about 75 % of the STD of the distribution of the original values. For this reason, the assertion that activity is most suitable for EEG classification is much stronger, and it can be concluded that the certainty of this hypothesis is actually higher than the already high certainty (around 100 %) obtained in these statistical tests.

5.8. Concluding Remarks

A combination of solutions has made this study possible. First, from the viewpoint of computation, the use of NeuralStudio proved crucial in automating the optimization which otherwise would have taken weeks due to the presence of multiple network files and solutions. NeuralStudio's training module provided the right options that facilitated network optimization and made computations easy and fast.

Second, the approach proposed here to link EEG recordings to non-epileptic or epileptic subjects are, to the knowledge of the author, unique. Prior to developing the algorithm, a thorough search in the literature was done, but no equivalent EEG classification methods could be found. Therefore, the method was developed entirely in a unique way.

The uniqueness of the method is that it proved that a new descriptor can be extracted from a scalp EEG segment in order to associate it to either a non-epileptic or a epileptic subject. The study describes how to obtain the descriptor as a 3x3 matrix whose great advantage is that it is irrespective to the number and location of the electrodes, as well as the duration of the recording. These matrixes constitute the simplest way of comparing EEG recordings. While one can assume that the grand AVG of the activity across all electrodes and time windows can be compared between EEG files, this method expanded this reasoning to STD and SNR along both time and electrodes. By doing this, a 3x3 matrix is constructed, which can detect differences in EEG patterns better than just averaging them.

After applying these descriptors to the time and frequency domain features, the method proved that Hjorth's parameter activity is by far the best feature to link EEG to non-

epileptic or epileptic subjects. When training this classifier 80 times with cross-validation, accuracy, sensitivity and specificity on the testing set were all above 0.9667. The delta classifier ranked second, with minimum performance metric of 0.9000. The worst was found to be the alpha classifier, with a minimum performance of 0.7273.

The superiority of the activity descriptor was then reiterated by means of the t-Test, which showed that activity is with 99.99 % certainly better than the remaining features when used to relate EEG recordings to non-epileptic or epileptic subjects.

This method represents a first solution to the classification problem and can be enhanced in several ways. For example, it is important to investigate whether other features which have been used lately in EEG analysis, such as coherence and entropy, could also be suitable. Also, window sizes other than 1 sec should be investigated to see if they affect the results. It is also important to add more subjects to the analysis in the future, rather than using more segments from each subject. Investigations have been started on this matter. Based on this line of thought, some of the empirical results that were obtained lead us to believe that some course of action can be taken to begin addressing the complex problem of seizure prediction. Figure 5.12, Figure 5.13, and Figure 5.14 illustrate the dynamics of different frequency bands in interictal EEG and iEEG data comparing EEG of a subject with no history of seizures (Figure 5.12) to both EEG and iEEG of a epileptic patient.

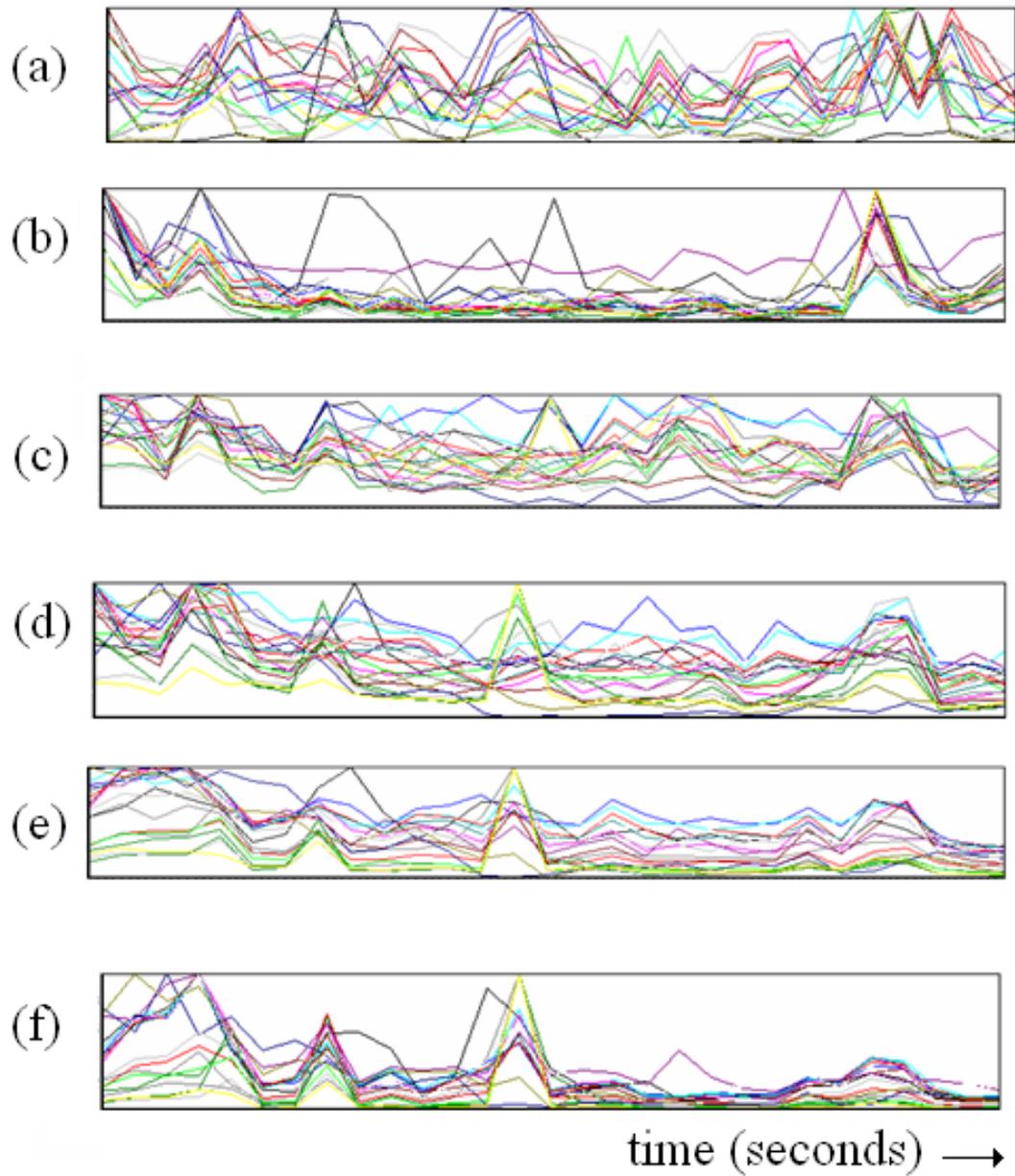


Figure 5.12: A representation of the spectral power of a non-epileptic person's EEG in the delta, theta, alpha, beta I, beta II, and gamma frequencies is given in (a) through (e), respectively. No trend can be observed as the frequency band changes.

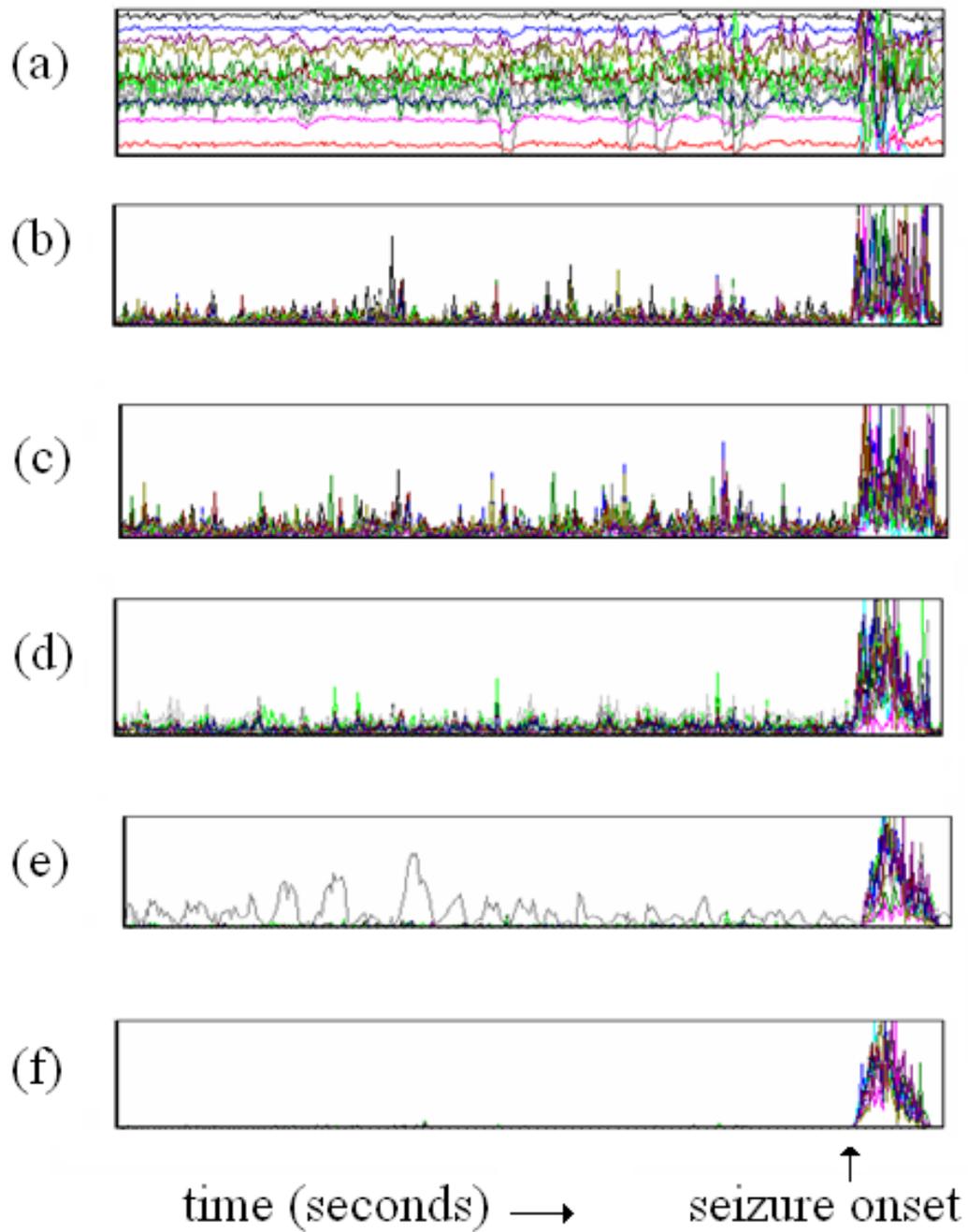


Figure 5.13: A plot of the spectral power of an epileptic patient's scalp EEG in the delta, theta, alpha, beta I, beta II, and gamma frequencies is given in (a) through (e), respectively. A narrowing of the plots is observable as the frequency band changes.

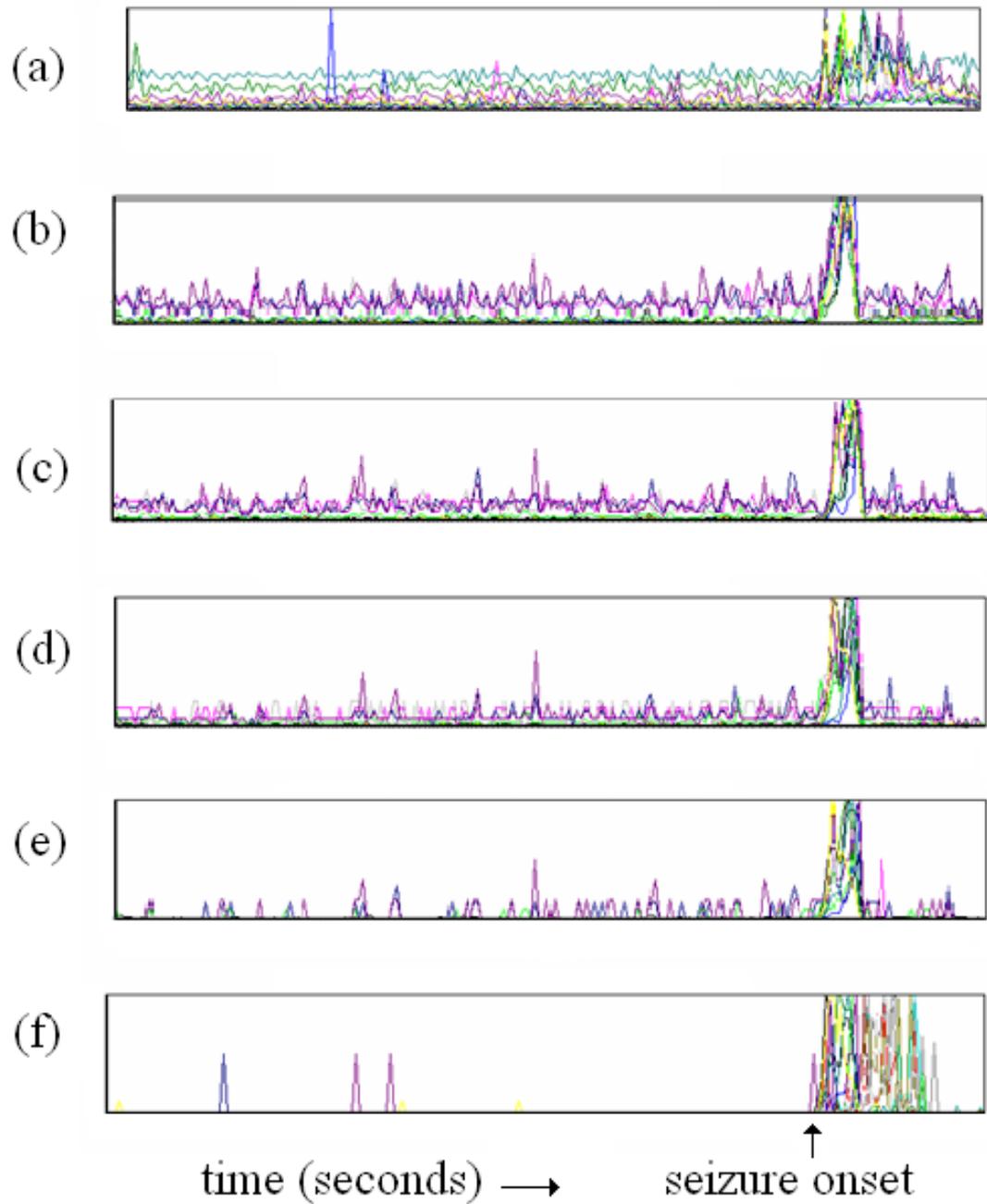


Figure 5.14: A plot of the spectral power of an epileptic patient's iEEG in the delta, theta, alpha, beta I, beta II, and gamma frequencies is given in (a) through (e), respectively. A narrowing of the plots is observable as the frequency band changes.

Based on these empirical evaluations, the following observations can be made on the challenging problem of seizure prediction:

- (1) There is an evident decrement in the standard deviation of the spectral power from delta to gamma prior to seizure;
- (2) There is a significant difference of the standard deviation for those electrodes that initiate a seizure in contrast to those that do not, and
- (3) a few seconds prior to seizure, interlocking between all electrodes is experienced.

It is hoped that combining these findings with the approach presented in this dissertation will be beneficial for future development of seizure prediction algorithms.

CHAPTER 6

6. CONCLUSIONS

The main intention of this dissertation was to present the NS tool and show its applicability to two specific neuroscience problems which were also fully addressed from a medical application perspective and related mathematical framework.

The NS tool developed is capable of simulating and training ANNs in a user friendly interface, offering modules to select the best alternatives out of a large number of solutions which were otherwise impossible to create by other available software tools.

One of the most attractive features of NS is that it can be operated without any programming skills and without knowing the mathematical details inherent to the design of ANNs. As shown in Chapter 3, NS users can generate ANNs easily just by knowing how many inputs and outputs are needed, and they can use a pattern module to load and edit the training data. The training module is fully configurable and can provide visual feedback on the training progress while operating at high speed. For its capabilities, NS has been granted a patent from the US Patent Office (Patent # 7502763, issued on March 10, 2009).

Chapter 1 of this dissertation gives an overview of recent applications of Artificial Intelligence to neuroscience problems. A foundation in algorithms for detecting epileptiform activity was provided in Chapter 2, which describes the most common features used in EEG analysis. This chapter was crucial for understanding the basics of the methods later developed in this dissertation.

A description of the NS tool was presented in Chapter 3, emphasizing the modules related to design, training and data management.

In pursuing practical applications in neuroscience, Chapter 4 described the design and implementation of a method aimed at detecting seizures in subdural EEG files from epileptic patients. The data was provided by MCH and involved 14 patients who underwent two-stage epilepsy surgery with subdural recordings. The approach implemented consisted of extracting the spectral power in different frequency bands and using both average and time information as features to identify seizures in subdural EEG files. More specifically, the 2 dimensions of the ANN were defined as 1) the inter-electrode mean of the spectral power and 2) the time duration in which the inter-electrode mean of the spectral power is above its AVG + one STD in the entire file.

It was found that the classifier performed better when using the gamma frequency band (36 - 44 Hz), yielding 95.90 % accuracy, 92.59 % sensitivity, and 96.84 % specificity. It was proved that blind selection of files for training, validation and testing did not considerably change the final results. Even testing the classifiers in files from patients who were never used for training did not affect the results. This confirms that the patterns found in the 2D plane by means of the spectral power in the gamma frequencies are patient independent.

Increasing the number of hidden neurons did not improve the results. This observation was made after a large number of training iterations, all stopped with cross-validation. Possible explanations for this can be as follows: 1) starting points for the iterations are always different and 2) local minima are always possible. Since the backpropagation

algorithm used in training the networks is always started from random solutions to avoid local minima, it is perfectly possible to achieve higher testing and even training errors if more hidden units are added. To reach this conclusion, up to ten repetitions were done and averaged in each topology. More repetitions could have been performed for each topology, however, with increasing data size and number of neurons, the time needed for the optimization increases so that the test becomes impractical.

This algorithm might be capable of making a substantial contribution to the diagnostic gain of long-term EEG monitoring during presurgical assessment.

In Chapter 5, an innovative solution has been conceived that allows linking scalp EEG recordings to either persons who have never experienced a seizure or to epileptic patients. For this method, scalp EEG was recorded at MCH from 3 volunteers and 3 epileptic subjects. The method was based on establishing a descriptor matrix for any EEG feature which proved to be unique in terms of simplicity and robustness. The algorithm for computing the matrix is straightforward and involves statistical operations on features across electrodes and along time. The matrix used was limited to the AVG, STD and SNR. The attractiveness of this matrix is that it can be used to represent and compare EEG files with different duration and number of electrodes, which could also be placed in different locations.

This study investigated time features (Hjorth's parameters) as well as frequency features (all frequency bands). The matrixes were computed in all EEG files and were then used as inputs of an ANN that was trained to distinguish EEG recordings from normal and epileptic subjects. The analysis was performed for one different feature at a time.

Training all ANNs yielded interesting results: the activity-based ANN performed by far better than the other features, with accuracy, sensitivity and specificity all above 96.67 %. Statistical tests confirmed the superiority of activity with a certainty degree higher than 99.99 %.

The results obtained in these two studies provide clues as to which EEG characteristics can be addressed for seizure detection and EEG classification analysis.

The aforementioned results were all achieved with the ability of NeuralStudio (NS) to automatically generate and optimize a large number of topologies in a short time. It has been shown that the NS tool is efficient in solving practical neuroscience problems such as the ones that have been described here, namely: 1) detecting seizures from intracranial EEG and 2) associating EEG recordings to non-epileptic and epileptic subjects.

Besides the features that make NS capable of performing state-of-the-art studies with ANNs, the two methods described here are unique and represent, to the knowledge of the author, a contribution to the study of epilepsy as it can be used by doctors for patient evaluation and medical assessments, and for researchers in the field to pursue these findings further towards the ultimate goal of seizure prediction.

BIBLIOGRAPHY

Aarts EHL, Korst JHM, Boltzmann Machines and their Applications, in Parallel Architectures and Languages in Europe, Lecture Notes in Computer Science 258, Volume 1, 1987, De Bakker JW, Nijman AJ, Treleven PC, Eds, Springer-Verlag, Berlin, pp. 34-50.

Aarts EHL, Korst JHM, Boltzmann Machines for Travelling Salesman Problems, European Journal of Operational Research 1989, Volume 39, pp. 79-95.

Aarts EHL, Korst JHM, Simulated Annealing and Boltzmann Machines: A Stochastic Approach to Combinatorial Optimization and Neural Computing, Wiley, New-York, NY, 1989.

Abásolo D, Hornero R, Espino P, Poza J, Sánchez, CI, De la Rosa, R, Analysis of Regularity in the EEG Background Activity of Alzheimer's Disease Patients with Approximate Entropy, August 2005, Volume 116, Issue 8, pp. 1826-1834.

Abend NS, Dlugos D, S. Herman, Neonatal Seizure Detection using Multichannel Display of Envelope Trend, Epilepsia, 2008, Volume 49, Issue 2, pp. 349-352.

Acir N, Oztura I, Kuntalp M, Baklan B, Güzeliş C, Automatic Detection of Epileptiform Events in EEG by a Three-stage Procedure based on Artificial Neural Networks, IEEE Transactions on Biomedical Engineering, Jan. 2005, Volume 52, No. 1, pp. 30-40.

Adams RD, Victor M., Principles of Neurology. 4th ed. New York: McGraw-Hill, 1989.

Adjouadi M, Ayala M, Making Waves Useful, IEEE Potentials, Feb/March 2003, Volume 22, No. 1, pp. 6-11.

Adjouadi M, Sanchez D, Cabrerizo M, Ayala M, Jayakar P, Yaylali I, Barreto A, Interictal Spike Detection using the Walsh Transform, IEEE Transactions on Biomedical Engineering, May 2004, Volume 51, No. 5, pp. 868-873.

Adjouadi M, Cabrerizo M, Yaylali I, Jayakar P, Interpreting EEG Functional Brain Activity, IEEE Potentials, Feb.-March 2004, Volume 23, Issue 1, ISSN: 0278-6648, pp. 8-13.

Adjouadi M, Cabrerizo M, Ayala M, Nunez K, , An Approach to Functional Brain Mapping using an Inverse Solution Based on the Principal Component Transform, Inverse Problems, Design and Optimization Symposium, Rio de Janeiro, Brazil, March 17-19, 2004, Available at: <http://www.lmt.coppe.ufrj.br/ipdo/>.

Adjouadi M, Cabrerizo M, Ayala M, Sanchez D, Yaylali I, Jayakar P, Barreto A, Detection of Interictal Spikes and Artifactual Data Through Orthogonal Transformations,

Journal of Clinical Neurophysiology, January/February 2005, Volume 22, Issue 1, pp. 53-64.

Adjouadi M, Zong N, Ayala M, Multidimensional Pattern Recognition and Classification of White Blood Cells using Support Vector Machines, Particle and Particle Systems Characterization, Wiley-VCH, September 2005, Volume 22, Issue 2, pp. 107-118.

Ajiboye AB, Weir RF, A Heuristic Fuzzy Logic Approach to EMG Pattern Recognition for Multifunctional Prosthesis Control, IEEE Transactions on Neural Systems and Rehabilitation Engineering, Sept. 2005, Volume 13, Issue: 3, ISSN: 1534-4320, pp. 280-291.

Albano AM, Cellucci CJ, Harner RN, Rapp PE, Optimization of Embedding Parameters for Prediction of Seizure Onset with Mutual Information, in: MEES, A. I. (Ed.): Nonlinear Dynamics and Statistics: Proceedings of an Isaac Newton Institute Birkhauser, Boston, 2000, pp. 435-451.

Alligood KT, Chaos: An Introduction to Dynamical Systems. Springer-Verlag, New York, 1997, LLC. ISBN 0-387-94677-2.

Anderson RE, Jakobsson JG, Entropy of EEG during Anaesthetic Induction: A Comparative Study with Propofol or Nitrous Oxide as Sole Agent, British Journal of Anaesthesia, 2004, Volume 92, No. 2, pp. 167-170.

Arnhold J, Lehnertz K, Grassberger P, Elger CE, A Robust Method for Detecting Interdependences: Application to Intracranially Recorded EEG, Physica D, 1999, 134, pp. 419-430.

Ayala M, Botura G, Maldonado O, AI Automates Substation Control, IEEE Computer Applications in Power, Jan. 2002, Volume 15, No. 1, pp. 41-46.

Ayala M, Botura G, Maldonado O, An AI Tool for Supervising Substations, IEEE Potentials, Dec. 2001 / Jan. 2002, Volume 20, No. 5, pp. 13-18.

Ayala M, Adjouadi M, An Optimized Artificial Neural Network Approach for Epileptiform Activity Recognition, Proceedings of the IASTED International Conference Modeling and Simulation MS 2003, Palm Springs, California, February 24-26, 2003.

Ayala M, Adjouadi M, Artificial Neural Network Design and Evaluation Tool, United States Patent No. 7,502,763, issued March 10, 2009 (available at <http://patft.uspto.gov/>).

Ayala M, Adjouadi M, Yaylali I, Jayakar P, An Optimization Approach to Recognition of Epileptogenic Data using Neural Networks with Simplified Input Layers, Biomedical Sciences Instrumentation, Volume 40, 2004, pp. 181-186.

- Ayala M, Adjouadi M, Cabrerizo M, Tito M, Lizárraga G, Detecting Electrodes Leading to Seizure from Subdural EEG using Spectral Phase Analysis towards Seizure Prediction, Computing Alliance of Hispanic-Serving Institutions (CAHSI), 3rd Annual Meeting Proceedings, Jan. 15-18, 2009, Google Inc, San Francisco, CA, USA
- Baraldi A, Blonda P, A Survey of Fuzzy Clustering Algorithms for Pattern Recognition - Part II, IEEE Transactions on Systems, Man, and Cybernetics, Part B, Dec 1999, Volume 29, Issue 6, ISSN: 1083-4419, pp. 786-801.
- Bartlett M, Makeig S, Bell AJ, Jung T-P, Sejnowski TJ, Independent Component Analysis of EEG Data, Society for Neuroscience Abstracts, 21, 1995, pp. 437.
- Barreto A, Principe JC, Reid SA, STL: A Spatio-temporal Characterization of Focal Interictal Events, Brain Topography 1993, Volume 5, pp. 215-228.
- Barricelli NA, Esempi Numerici di Processi di Evoluzione, Métodos, 1954, pp. 45-68.
- Barricelli NA, Symbiogenetic Evolution Processes Realized by Artificial Methods, Methodos, 1957, pp. 143-182.
- Barron AR, Universal Approximation Bounds for Superpositions of a Sigmoid Function, IEEE Transactions on Information Theory, 39, 1993, pp. 930-945.
- Berry MJA, Linoff G, Data Mining Techniques, NY: John Wiley & Sons, 1997.
- Bezerianos A, Tong S, Thakor N, Time-dependent Entropy Estimation of EEG Rhythm Changes following Brain Ischemia, Annals of Biomedical Engineering, Volume 31, 2003, pp. 221-232.
- Birkemeier WP, Fontaine AB, Celesia GG, Ma KM, Pattern Recognition Techniques for the Detection of Epileptic Transients in EEG, IEEE Transactions on Biomedical Engineering, Volume 25, No. 3, May 1978, pp. 213-217.
- Bishop CM, Neural Networks for Pattern Recognition, Oxford University Press, 2005, ISBN 0198538642, 9780198538646.
- Blum A, Neural Networks in C++, NY: Wiley, 1992.
- Boger Z, Guterman H, Knowledge Extraction from Artificial Neural Network Models, IEEE Systems, Man, and Cybernetics Conference, Orlando, FL, 1997.
- Bragin A, Wilson CL, Fields T, Fried I, Engel JJr, Analysis of Seizure Onset on the Basis of Wideband EEG Recordings, Epilepsia, No. 46, 2005, pp. 59-63.

Braun H, On Solving Traveling Salesman Problems by Genetic Algorithms, Lecture Notes in Computer Science, Springer Berlin / Heidelberg, Volume 496/1991, pp. 129-133, Chapter from book: Parallel Problem Solving from Nature, ISBN 978-3-540-54148-6.

Bruhn J, Röpcke H, Hoefft A, Approximate Entropy as an Electroencephalographic Measure of Anesthetic Drug Effect during Desflurane Anesthesia, Anesthesiology, Volume 92, Issue 3, March 2000, pp. 715-726.

Budinich M, A Self-Organizing Neural Network for the Traveling Salesman Problem That Is Competitive with Simulated Annealing, Neural Computation, MIT Press, ISSN 0899-7667, February 15, 1996, Volume 8, No. 2, pp. 416-424.

Cabrerizo M, Adjouadi M, Ayala M, Núñez K, An Inverse Solution to Functional Brain Mapping of the Auditory Process using an Eigensystem Study, Inverse Problems in Science and Engineering, Volume 14, No. 4, June 2006, pp. 437-452.

Cabrerizo M, Adjouadi M, Ayala M, Tito M, Pattern Extraction in Interictal EEG Recordings towards Detection of Electrodes Leading to Seizures, Biomedical Sciences Instrumentation, 42 (ISA Volume 464), 2006, pp. 243-248.

Cabrerizo M, Tito M, Ayala M, Adjouadi M, Barreto A, Jayakar P, An analysis of Subdural EEG Parameters for Epileptic Seizure Evaluation, CD Proceedings of the 16th International Conference on Computing CIC-2007, Mexico City, Mexico 2007.

Cabrerizo M, Adjouadi M, Ayala M, Tito M, Lizárraga G, Coherence Analysis of Subdural EEG towards the Delineation of Electrodes that Initiates Epileptic Seizures, Computing Alliance of Hispanic-Serving Institutions (CAHSI), 3rd Annual Meeting Proceedings, Jan. 15-18, 2009, Google Inc, San Francisco, CA, USA.

Calvagno G, Ermani M, Rinaldo R, Sartoretto F, A Multiresolution Approach to Spike Detection in EEG, Proceedings of the EEE International Conference on Acoustics Speech and Signal Processing, 2000, Volume 6, pp. 3582-3585.

Cambier DM, Cascino GD, So EL, Marsh WR, Video-EEG monitoring in patients with hippocampal atrophy, Acta Neurologica Scandinavica, April 2001, Volume 103, Issue 4, pp. 231-237.

Cascino GD, Epilepsy: Contemporary Perspectives on Evaluation and Treatment, Mayo Clinic Proceedings 1994, Volume 69, pp. 1199-1211.

Chander R, Urrestarrazu E, Gotman J, Automatic Detection of High Frequency Oscillations in Human Intracerebral EEGs, Epilepsia, 47, No. 4, 2006, pp. 37-37.

Centers for Disease Control and Prevention, Epilepsy Surveillance Among Adults — 19 States, Behavioral Risk Factor Surveillance System, 2005, Morbidity and Mortality Weekly Report, Surveillance Summaries, Aug 8, 2008, Volume 57, No. SS-6.

Chan CC, Shi CZ, Jin W, Wang DN, Improving the Wavelength Detection Accuracy of FBG Sensors using an ADALINE Network, *Photonics Technology Letters, IEEE*, Aug. 2003, Volume 15, Issue 8, ISSN: 1041-1135, pp. 1126-1128.

Chan AM, Sun FT, Boto EH, Wingeier BM, Automated Seizure Onset Detection for Accurate Onset Time Determination in Intracranial EEG, *Clinical Neurophysiology*, Volume 119, Issue 12, Dec. 2008, pp. 2687-2696.

Chatrian GE, Bergamini L, Dondey M, Klass DW, Lennox-Buchthal M, Petersén I, A Glossary of Terms most commonly used by Clinical Electroencephalographers, *Electroencephalography and Clinical Neurophysiology*, 1994, Volume 37, pp. 538-548.

Chen Y, Narita M, Yamada T, Nuclear Reactor Diagnostic System using Genetic Algorithm (GA)-trained neural networks, *Electrical Engineering in Japan*, Volume 115, Issue 5, August 1995, pp. 88-99.

Chen DR, Chang RF, Huang YL, Chou YH, Tiu CM, Tsai PP, Texture Analysis of Breast Tumors on Sonograms, *Semin. Ultrasound CT MR*. Aug 2000, 21, Volume 4, pp. 308-316.

Chung HY, Chiang CK, A Self-learning and Tuning Fuzzy Logic Controller based on Genetic Algorithms and Reinforcements, *International Journal of Intelligent Systems*, Volume 12, Issue 9, September 1997, pp. 673-694.

Congedo M, Gouy-Pailler C, Jutten C, On the Blind Source Separation of Human Electroencephalogram by Approximate joint Diagonalization of Second Order Statistics, *Clinical Neurophysiology*, Dec. 2008, Volume 119, Issue 12, pp. 2677-2686.

Creutzfeldt OD, Bodenstein G, Barlow JS, Computerized EEG Pattern Classification by Adaptive Segmentation and Probability Density Function Classification, Clinical evaluation, *Electroencephalography and Clinical Neurophysiology*, May 1985, Volume 60, No. 5, pp. 373-393.

Cristianini N, Shawe-Taylor J, *An Introduction to Support Vector Machines and other Kernel-based Learning Methods*, New York: Cambridge University Press; 2000.

D'Alessandro M, Esteller R, Vachtsevanos G, Hinson A, Echaz J, Litt B, Epileptic Seizure Prediction using Hybrid Feature Selection over Multiple Intracranial EEG Electrode Contacts: A report of four Patients, *IEEE Transactions on Biomedical Engineering*, 2003, Volume 50, No. 5, pp. 603-615.

Das R, Turkoglu I, Sengur A, Diagnosis of Valvular Heart Disease through Neural Networks Ensembles, *Computer Methods and Programs in Biomedicine*, Volume 93, Issue 2, February 2009, ISSN:0169-2607, pp. 185-191.

Davey BK, Fright WR, Carroll GJ, Jones RD, Expert System Approach to Detection of Epileptiform Activity in EEG. *Med Biol Eng Comput.* 27, 1989, pp. 365-70.

De Vos M, Vergult A, De Lathauwer L, De Clercq W, Van Huffel S, Dupont P, Palmini A, Van Paesschen W, Canonical Decomposition of Ictal Scalp EEG Reliably Detects the Seizure Onset Zone, *NEUROIMAGE*, Sep 1, 2007, Volume 37, Issue 3, pp. 844-854.

Deburchgraeve W, Cherian PJ, De Vos M, Swarte RM, Blok JH, Visser GH, Govaert P, Van Huffel S, Automated Neonatal Seizure Detection Mimicking a Human Observer Reading EEG, *Clinical Neurophysiology*, Volume 119, Issue 11, Nov. 2008, pp. 2447-2454.

DeVore, RA, Howard R, Micchelli CA, Optimal Nonlinear Approximation, *Manuscripta Mathematica*, 63, 1989, pp. 469-478.

Dimitrov V, Korotkich V, *Fuzzy Logic: A Framework for the New Millennium*, Springer, 2002, ISBN 3790814253, 9783790814255.

Dingle AA, Jones RD, Carroll GJ, Fright RW, A Multiscale System to Detect Epileptiform Activity in the EEG, *IEEE Transactions on Biomedical Engineering*, Volume 40, 1993, pp. 1260-1268.

Eng J, Predicting the Presence of Acute Pulmonary Embolism: A Comparative Analysis of the Artificial Neural Network, Logistic Regression, and Threshold Models, *Am. J. Roentgenol.*, October 1, 2002, Volume 179, Issue 4, pp. 869-874.

Engel J. Jr., Surgery for Seizures. *NEJM* 1996, pp. 647-652.

Epilepsy Foundation of America, August 2008, Available at: <http://www.epilepsyfoundation.org/about/statistics.cfm>.

Favata F, Walker R, A study of the Application of Kohonen-type Neural Networks to the Traveling Salesman Problem, *Biological Cybernetics*, Springer Berlin/Heidelberg, Volume 64, No. 6, April, 1991, pp. 463-468.

Feldkamp LA, Puskorius GV, Moore PC, Adaptation from Fixed Weight Dynamic Networks, *IEEE International Conference on Neural Networks*, 1996, Jun 1996, Volume 1, ISBN: 0-7803-3210-5, pp. 155-160.

Ferrán EA, Pflugfelder B, Ferrara P, Self-organized Neural Maps of Human Protein Sequences, *Protein Science*, Volume 3, Issue 3, March 1994, pp. 507-521.

Ferree TC, Luu P, Russell GS, Tucker DM, Scalp Electrode Impedance, Infection Risk, and EEG Data Quality, *Clinical Neurophysiology*, Volume 112, No. 3, Elsevier, March 2001, pp. 536-544.

Fogel DB, *Evolutionary Computation: Towards a New Philosophy of Machine Intelligence*. New York: IEEE Press 2000, pp. 140.

Fogel DB, *Evolutionary Computation: Toward a New Philosophy of Machine Intelligence*, IEEE Press, Piscataway, NJ. Third Edition, Wiley-IEEE Press 2006.

Fogel LJ, Owens AJ, Walsh MJ, *Artificial Intelligence through Simulated Evolution*, Wiley, New York, 1996.

Frank GW, Lookman T, Nerenberg MAH, Essex C, Lemieux J, Blume W, Chaotic Time Series Analyses of Epileptic Seizures, *Physica D*, 1990, Volume 46, Issue 3, pp. 427-438.

Fraser A, Simulation of Genetic Systems by Automatic Digital Computers. I. Introduction, *Aust. J. Biol. Sci.*, 1957, Volume 10, pp. 484-491.

Fraser A, Burnell D, *Computer Models in Genetics*. New York: McGraw-Hill, 1970.

Friedman DE, Schevon C, Emerson RG, Hirsch LJ, Cyclic Electrographic Seizures in Critically ill Patients, *Epilepsia*, Feb 2008, Volume 49, Issue 2, pp. 281-287.

Gabor AJ, Seizure Detection using a Self-organizing Neural Network: Validation and Comparison with other Detection Strategies, *Electroencephalography and Clinical Neurophysiology*, Volume 107, Number 1, 1998, pp. 27-32.

Gabor AJ, Seyal M, Automated Interictal EEG Spike Detection using Artificial Neural Networks, *Electroencephalography and Clinical Neurophysiology*, Nov. 1992, Volume 83, pp. 271-280.

Gevins AS, Remond A, *Methods of Analysis of Brain Electrical and Magnetic Signals*, in *Handbook of Electroencephalography and Clinical Neurophysiology*, Revised Series, Volume 1, Elsevier Science Ltd; Amsterdam, N.Y., Oxford, 1987.

Guevara MR, Chaos in Electrophysiology, in: *Concepts and Techniques in Bioelectric Measurements: Is the Medium Carrying the Message?*, edited by LeBlanc, A-R, Billette, J., Editions de l'Ecole Polytechnique de Montreal, Montreal 1997, pp. 67-87.

Glover JR, Raghavan N, Ktonas PY, Frost JD, Context-based Automated Detection of Epileptogenic Sharp Transients in the EEG: Elimination of False Positives, *IEEE Transactions on Biomedical Engineering*, 1989, Volume 36, pp. 519-527.

Goldberg DE, Genetic Algorithms, Addison Wesley, 1988.

Goldberg DE, Genetic Algorithms in Search, Optimization, and Machine Learning, Addison-Wesley Publishing Company ISBN 0-201-15767-5, 1989.

Goldberg DE., Genetic Algorithms in Search, Optimization, and Machine Learning, Addison Wesley, 2007, ISBN: 0201157675.

Gomez-Herrero G, Atienza M, Egiazarian K, Cantero J.L, Measuring Directional Coupling between EEG Sources, NEUROIMAGE, Nov. 15, 2008, Volume 43, Issue 3, pp. 497-508.

Good LB, Sabesan S, Marsh ST, Tsakalis K, Iasemidis LD, Treiman DM, Automated Seizure Prediction and Deep Brain Stimulation Control in Epileptic Rats, Epilepsia 2007, Volume 48, Issue 6, pp. 278-278.

Gotman J, Automatic Recognition of Epileptic Seizures in the EEG, Electroencephalography and Clinical Neurophysiology, 54, 1982, pp. 530-540.

Gotman J, Practical Use of Computer-assisted EEG Interpretation in Epilepsy, Journal of Clinical Neurophysiology, Volume 2, 1985, pp. 251-65.

Gotman J, Automatic Detection of Seizures and Spikes, Journal of Clinical Neurophysiology, Volume 16, 1999, pp. 130-140.

Gotman J, Wang LY, State Dependent Spike Detection: Validation, Electroencephalography and Clinical Neurophysiology, 1992, Volume 83, pp. 12-18.

Grassberger P, Procaccia I, Measuring the Strangeness of Strange, Attractors, Physica D: Nonlinear Phenomena, 1983, Volume 9, Issue 1-2, pp. 189-208.

Grosan C, Abraham A, Hybrid Evolutionary Algorithms, SpringerLink (Online service, Hisao Ishibuchi, Springer, 2007, ISBN 3540732969, 9783540732969.

Gulera NF, Ubeylib ED, Guler I, Recurrent Neural Networks Employing Lyapunov Exponents for EEG Signals Classification, Expert Systems with Applications, 2005, Volume 29, pp. 506-514.

Gurney JW, Swensen SJ, Solitary Pulmonary Nodules: Determining the Likelihood of Malignancy with Neural Network Analysis, Radiology, Volume 196, 1995, pp. 823-829.

Hagan MT, Demuth HB, De Jesús O, An Introduction to the Use of Neural Networks in Control Systems, International Journal of Robust and Nonlinear Control, Volume 12, Issue 11, September 2002, pp. 959-985.

Hall LO, Bensaid AM, Clarke LP, Velthuizen RP, Silbiger MS, Bezdek JC, A Comparison of Neural Network and Fuzzy Clustering Techniques in Segmenting Magnetic Resonance Images of the Brain, Sep 1992, ISSN: 1045-9227, Volume 3, Issue 5, pp. 672-682.

Hauser WA, Kurland LT, The Epidemiology of Epilepsy in Rochester, Minnesota, 1935 through 1967, *Epilepsia* 1975, 16(1), pp. 1-66.

He WG, Li SF, Hu GW, Blur Identification using an Adaptive ADALINE Network, Proceedings of 2005 International Conference on Machine Learning and Cybernetics, 2005, Aug. 2005, ISBN: 0-7803-9091-1, Volume 9, pp. 5314-5317.

Hellmann G, Multifold Features Determine Linear Equation for Automatic Spike Detection applying Neural Networks in Interictal EcoG, *Clinical Neurophysiology*, 1999, May 11, pp. 887-94.

Hesse CW, James CJ, Tracking and Detection of Epileptiform Activity in Multichannel Ictal EEG using Signal Subspace Correlation of Seizure Source Scalp Topographies, *Medical & Biological Engineering & Computing*, Nov 2005, Volume 43, Issue 6, pp. 764-770.

Hjorth B, EEG Analysis based on Time Domain Properties, *Electroencephalography and Clinical Neurophysiology*, Volume 29, 1970, pp. 306-310.

Hjorth B, The Physical Significance of Time Domain Descriptors in EEG analysis, *Electroencephalograms and Clinical Neurophysiology*, Volume 34, March 1973, pp. 321-325.

Holland JH, *Adaptation in Natural and Artificial Systems*, University of Michigan Press, Ann Arbor, 1975; MIT Press, Cambridge, second edition, 1992.

Homan RW, Herman J, Purdy P, Cerebral Location of International 10-20 System Electrode Placement, *Electroencephalography and Clinical Neurophysiology*, ISSN 0013-4694, 1987, Volume 66, No. 4, pp. 376-382.

Hopfengartner R, Kerling F, Bauer V, Stefan H, An Efficient, Robust and Fast Method for the Offline Detection of Epileptic Seizures in Long-term Scalp EEG Recordings, *Clinical Neurophysiology*, Nov 2007, Volume 118, Issue 11, pp. 2332-2343.

Hopfengärtner R, Kasper B, Kerling F, Weigel D, Buchfelder M, Stefan H, An Automated Seizure Detection Algorithm for Long-Term ECoG Recordings: An Efficient and Fast Screening Tool for Presurgical Evaluation, *Epilepsia*, Volume 50, Suppl. 4, 2009, pp. 52-52.

Horn D, Ruppin E, Usher M, Herrmann M, Neural Network Modeling of Memory Deterioration in Alzheimer's Disease, *Neural Computation*, Volume 5, Issue 5, September 1993, ISSN: 0899-7667, pp. 736-749.

Larter R, Speelman B, A Coupled Ordinary Differential Equation Lattice Model for the Simulation of Epileptic Seizures, *Chaos*, 9, 1999, pp. 795-804.

Iasemidis LD, Sackellares JC, The evolution with Time of the Spatial Distribution of the Largest Lyapunov Exponent on the Human Epileptic Cortex. In Duke DW, Pritchard WS (Eds), *Measuring Chaos in the Human Brain*, World Scientific., pp. 49-82.

Iasemidis LD, Sackellares JC, Chaos Theory and Epilepsy, *The Neuroscientist*, 1996, Volume 2, pp. 118-126.

Iasemidis LD, Sackellares JC, US Patent 6304775 - Seizure Warning and Prediction, Issued on October 16, 2001.

Iasemidis LD, Sackellares JC, Gilmore RL, Roper SN, Automated Seizure Prediction Paradigm, *Epilepsia*, Volume 39, 1998, pp. 56.

Iasemidis LD, Barreto A, Gilmore RL, Uthman BM, Roper SA, Sackellares JC, Spatiotemporal Evolution of Dynamical Measures Precedes Onset of Mesial Temporal Lobe Seizures, *Epilepsia*, Volume 35, 1994, pp. 133-134.

International Federation of Societies for Clinical Neurophysiology, A Glossary of Terms Most Commonly used by Clinical Electroencephalographers, *Electroencephalography and Clinical Neurophysiology*, Nov 1974, Volume 37(5), pp. 538-48.

Jayakar P, Patrick JP, Shwedyk E, Seshia SS, Automated Rule Based Graded Analysis of Ambulatory Cassette EEGs, *Electroencephalography and Clinical Neurophysiology*, Volume 72(2), Feb. 1989, pp. 165-75.

Jayakar P, Resnick TJ, Duchowny MS, Alvarez LA, Localization of Epileptogenic Foci using a Simple Reference-subtraction Montage to Detect Small Inter-channel Time Differences, *Journal of Clinical Neurophysiology*, Volume 8, 1991, pp. 212-215.

Jerger KK, Netoff TI, Francis JT, Sauer T, Pecora L, Weinstein SL, Schiff SJ, Early Seizure Detection, *Journal of Clinical Neurophysiology*, May 2001, Volume 18, Issue 3, pp. 259-268.

Jeong JS, EEG Dynamics in Patients with Alzheimer's Disease, *Clinical Neurophysiology*, Jul 2004, Volume 115, Issue 7, pp. 1490-1505.

Jolliffe IT, *Principal Component Analysis*, 2nd Edition, Springer Verlag New York, Inc., ISBN 0-387-95442-2, 2002

Joyce J, *The Foundations of Causal Decision Theory*, Cambridge Studies in Probability, Induction, and Decision Theory, Cambridge University Press 1999, ISBN 0-521-64164-0.

Julstrom BA, Very Greedy Crossover in a Genetic Algorithm for the Traveling Salesman Problem, Proceedings of the 1995 ACM symposium on Applied Computing, p.324-328, February 26-28, 1995, Nashville, Tennessee, USA.

Jung TP, Makeig S, Stensmo M, Sejnowski TJ, Estimating Alertness from the EEG Power Spectrum, *IEEE Transactions on Biomedical Engineering*, ISSN: 0018-9294, Jan. 1997, Volume 44, Issue 1, pp. 60-69.

Kandel A, *Fuzzy Techniques in Pattern Recognition*, New York, Wiley, 1982, ISBN: 0471091367: 9780471091363.

Khan J, Wei JS, Ringnér M, Saal LH, Ladanyi M, Westermann F, Berthold F, Schwab M, Antonescu CR, Peterson C, Meltzer PS, Classification and Diagnostic Prediction of Cancers using Gene Expression Profiling and Artificial Neural Networks, *Nature Medicine* 7, 2001, pp. 673-679.

Kippenhan JS, Barker WW, Nagel J, Grady C, Duara R, Neural-Network Classification of Normal and Alzheimer's Disease Subjects using High-Resolution and Low-Resolution PET Cameras, *The Journal of Nuclear Medicine* Volume 35, No. 1, 1994, pp. 7-15.

Kiyimik MK, Akin M, Subasi A, Automatic Recognition of Alertness Level by using Wavelet Transform and Artificial Neural Network, *Journal of Neuroscience Methods*, Volume 139(2), 2004, pp. 231-240.

Klir GJ, Yuan B, *Fuzzy Sets and Fuzzy Logic: Theory and Applications*, Prentice Hall PTR, 1995, ISBN 0131011715, 9780131011717.

Ko CW, Chung HW, Automatic Spike Detection via an Artificial Neural Network using Raw EEG Data, *Clinical Neurophysiology*, 2000, 111, pp. 477-81.

Kohavi R, Provost F, Glossary of Terms, *Machine Learning*, 30, 1998, pp. 271-274.

Kohonen T, *Self-organizing Maps*, 3rd Edition, Springer 2001, ISBN 3540679219, 9783540679219

Kothari R, Cualing H, Balachander T, Neural Network Analysis of Flow Cytometry Immunophenotype Data, Context-based Automated Detection of Epileptogenic Sharp Transients in the EEG: Elimination of False Positives, Aug. 1996, Volume 43, No. 8, pp. 803-810.

Kurth C, Gilliam F, Steinhoff B, EEG Spike Detection with a Kohonen Feature Map. *J Ann Biomed Eng* 28, 2000, pp. 1362-1369.

Kwan HK, Cai Y, A Fuzzy Neural Network and its Application to Pattern Recognition, *IEEE Transactions on Fuzzy Systems*, Aug 1994, Volume 2, Issue 3, ISSN: 1063-6706, pp. 185-193.

Lai YC, Harrison MAF, Frei MG, Osorio I, Inability of Lyapunov Exponents to Predict Epileptic Seizures, *Physical Review Letters*, Volume 91, Issue 6, Aug. 8, 2003, pp. 068102-1 to 068102-4.

Lai YC, Harrison MAF, Frei MG, Osorio I, Controlled Test for Predictive Power of Lyapunov Exponents: Their inability to Predict Epileptic Seizures, *CHAOS*, Sep 2004, Volume 14, Issue 3, pp. 630-642.

Langdon WB, Poli R, *Foundations of Genetic Programming*, Springer-Verlag, Berlin Heidelberg New York, ISBN 3-540-42451-2, 2002.

Larsson PG, Kostov H, Lower Frequency Variability in the Alpha Activity in EEG Among Patients with Epilepsy, *Clinical Neurophysiology*, Nov 2005, Volume 116, Issue 11, pp. 2701-2706.

Lawrence S, Giles CL, Tsoi AC, Lessons in Neural Network Training: Overfitting May be Harder than Expected, *Proceedings of the Fourteenth National Conference on Artificial Intelligence, AAAI-97*, AAAI Press, Menlo Park, California, 1997, pp. 540-545.

Lehnertz K, Elger CE, Can Epileptic Seizures be Predicted? Evidence from nonlinear time series analysis of brain electrical activity, *Phys. Rev. Lett.*, 80, 1998, pp. 5019-5022.

Li XL, Wavelet Spectral Entropy for Indication of Epileptic Seizure in Extracranial EEG, *Neural Information Processing, Pt 3, Proceedings, Book Series: Lecture Notes in Computer Science*, 2006, Volume 4234, pp. 66-73.

Li J, Kwan RSK, A Fuzzy Genetic Algorithm for Driver Scheduling, *European Journal of Operational Research*, Volume 147, Issue 2, June 1st, 2003, pp. 334-344.

Lo SCB, Chan HP, Lin JS, Li H, Freedman MT, Mun SK, Artificial Convolution Neural Network for Medical Image Pattern Recognition, *Neural Networks*, Volume 8, Issues 7-8, 1995, pp. 1201-1214.

Marcum J, *Statistical Theory of Target Detection by Pulsed Radar*, *IEEE Trans. Info. Thry.*, Apr. 1960.

Martinerie J, Adam C, Van Quyen L, Baulac M, Clemenceau M, Renault S, Varela FJB, Epileptic Seizures can be Anticipated by Non-linear Analysis, *Nature Med.*, 4, 1998, pp. 1173-1176.

Micheloyannis S, Flitzanis N, Papanikolaou E, Bourkas M, Terzakis D, Arvanitis S, Stam CJ, Usefulness of non-linear EEG analysis, *Acta Neurologica Scandinavica*, Volume 97(1), January 1998, pp 13-19.

Minsky ML, Papert SA, *Perceptrons* (Cambridge, MA: MIT Press), 1969.

Moser HR, Weber B, Wieser HG, Meier PF, Electroencephalograms in Epilepsy: Analysis and Seizure Prediction within the Framework of Lyapunov Theory, *Physica D-Nonlinear Phenomena*, Volume 130, 1999, Issue 3-4, pp. 291-305.

Murro AM, Computerized Seizure Detection of Complex Partial Seizures, *Electroencephalography and Clinical Neurophysiology*, Volume 79, pp. 330-333, 1991.

Navakatikyan MA, Colditz PB, Burke CJ, Inder TE, Jane Richmond J, Williams CE, Seizure Detection Algorithm for Neonates based on Wave-sequence Analysis, *Clinical Neurophysiology*, Volume 117, Issue 6, pp. 1190-1203, June 2006.

Niedermeyer E, Lopes da Silva F, *Electroencephalography: Basic Principles, Clinical Applications and Related Fields*. Lippincott Williams & Wilkins; Fifth Edition (November 1, 2004), ISBN-10: 0781751268, ISBN-13: 978-0781751261.

Nooralahiyan AY, Hoyle BS, Bailey NJ, Neural Network for Pattern Association in Electrical Capacitance Tomography, *IEE Proceedings on Circuits, Devices and Systems*, Dec 1994, Volume 141, Issue 6, pp. 517-521, ISSN: 1350-2409.

Osowski S, Linh TH, ECG Beat Recognition using Fuzzy Hybrid Neural Network, *IEEE Transactions on Biomedical Engineering*, Nov 2001, Volume 48, Issue 11, ISSN: 0018-9294, pp. 1265-1271.

Osowski S, Swiderski B, Cichocki A, Rysz A, Epileptic Seizure Characterization by Lyapunov Exponent of EEG signal, *Compel - The International Journal for Computation and Mathematics in Electrical and Electronic Engineering*, Volume 26, 2007, Issue 5, pp. 1276-1287.

Ozyilmaz L, Yildirim T, Artificial Neural Networks for Diagnosis of Hepatitis Disease, *Proceedings of the International Joint Conference on Neural Networks (IJCNN)*, Portland, OR, USA, July 20-24 2003, Volume 1, pp. 586-589.

Teixeira S, Pacheco X, Borland Delphi 6 developer's Guide, Sams Publishing 2002, ISBN 0-672-32115-7

Pal NR, Bezdek JC, Tsao ECK, Generalized Clustering Networks and Kohonen's Self-organizing Scheme, IEEE Transactions on Neural Networks, Jul 1993, Volume 4, Issue 4, ISSN: 1045-9227, pp. 549-557.

Pal T, Pal NR, Pal M, Learning Fuzzy Rules for Controllers with Genetic Algorithms, International Journal of Intelligent Systems, Volume 18, Issue 5, May 2003, pp. 569-592.

Patrikar A, Provence J, Control of Dynamic Systems using Fuzzy Logic and Neural Networks, International Journal of Intelligent Systems, Volume 8, Issue 6, 1993, pp. 727-748.

Pezard L, Martinerie J, Breton F, Bourzeix JC, Renault B, Non-linear Forecasting Measurements of Multichannel EEG dynamics, Electroencephalography and Clinical Neurophysiology, 1994, Nov; 91(5), pp. 383-391.

Piche SW, The Selection of Weight Accuracies for MADALINES, IEEE Transactions on Neural Networks, Volume 6, Issue 2, March 1995, pp. 432-445.

Popescu S, Trained Wavelets used to Detect Epileptic Spikes, Proceedings of the IEEE-SP Int. Symp., Time-Frequency and Time-Scale Analysis, 1998, pp. 285-288.

Pradhan N, Sadasivan PK, The Nature of Dominant Lyapunov Exponent and Attractor Dimension Curves of EEG in Sleep, Volume 26, Issue 5, September 1996, pp. 419-428.

Pritchard WS, Measuring Chaos in the Brain: A Tutorial Review of Nonlinear Dynamical EEG Analysis, Intern. J. Neuroscience, 1992, Volume 67, pp. 31-80.

Rachele W, Learn Object Pascal with Delphi, Plano, Tex: Wordware Pub., 2001, ISBN 1-55622-719-1.

Rechenberg I, Evolutionsstrategie, Stuttgart, Friedrich Frommann Verlag, 1973.

Reddick WE, Glass JO, Cook EN, Elkin TD, Deaton RJ, Automated Segmentation and Classification of Multispectral Magnetic Resonance Images of Brain using Artificial Neural Networks, IEEE Transactions on Medical Imaging, Dec. 1997, Volume 16, Issue 6, pp. 911-918.

Reddick WE, Mulhern RK, Elkin TD, Glass JO, Merchant TE, Langston JW, A Hybrid Neural Network Analysis of Subtle Brain Volume Differences in Children Surviving Brain Tumors, Magnetic Resonance Imaging, Elsevier Science, New York, NY, ISSN 0730-725X, 1998, Volume 16, No 4, pp. 413-421.

Rich E, Knight K, Artificial Intelligence, McGraw-Hill, New York, 1991, pp. 105-130.

Ripley BD, Pattern Recognition and Neural Networks, 8th Edition, Cambridge University Press, 1996, ISBN 0521460867, 9780521460866.

Rosenblatt F, The Perceptron: A Probabilistic Model for Information Storage and Organization in the Brain, Cornell Aeronautical Laboratory, Psychological Review, 1958, Volume 65, No. 6, pp. 386-408.

Rowley HA, Baluja S, Kanade T, Neural Network-based Face Detection, IEEE Transactions on Pattern Analysis and Machine Intelligence, Jan 1998, Volume 20, Issue 1, ISSN: 0162-8828.

Santini S, The Self-organizing Field (Kohonen Maps), IEEE Transactions on Neural Networks, Nov. 1996, Volume 7, Issue 6, ISSN: 1045-9227, pp. 1415-1423.

Sarle WS, Stopped Training and Other Remedies for Overfitting, Proceedings of the 27th Symposium on the Interface of Computing Science and Statistics, 1995, pp. 352-360, Available at: <ftp://ftp.sas.com/pub/neural/inter95.ps.Z>

Scott JA, Palmer EL, Fischman AJ, How Well Can Radiologists using Neural Network Software Diagnose Pulmonary Embolism?, AJR 2000, Volume 175, pp. 399-405.

Schad A, Schindler K, Schelter B, Maiwald T, Brandt A, Timmer J, Schulze-Bonhage A, Application of a Multivariate Seizure Detection and Prediction Method to Non-invasive and Intracranial Long-term EEG Recordings, Clinical Neurophysiology, Volume 119, Issue 1, Jan 2008, pp. 197-211.

Shannon CE, A Mathematical Theory of Communication, Bell Syst. Tech. J., 27, 1948, 379-423, pp. 623-656.

Shannon CE, Communication in the Presence of Noise, Proc. Institute of Radio Engineers, Volume 37, No.1, Jan. 1949, pp. 10-21. Reprint as classic paper in: Proc. IEEE, Volume 86, No. 2, (Feb 1998).

Ishibuchi H, Nozaki K, Yamamoto N, Tanaka H, Selection of Fuzzy If-Then Rules by a Genetic Method, Electronics and Communications in Japan (Part III: Fundamental Electronic Science), Volume 77, Issue 2, 1994, pp. 94-104.

Shoeb A, Edwards H, Connolly J, Bourgeois B, Treves T, Gutttag J, Patient-Specific Seizure Onset Detection, Proceedings of the 26th Annual International Conference of the IEEE EMBS, San Francisco, CA, USA, September 1-5, 2004, Volume 1, ISBN: 0-7803-8439-3, pp. 419-422.

Smart O, Firpi H, Vachtsevanos G, Genetic Programming of Conventional Features to Detect Seizure Precursors, Engineering Applications of Artificial Intelligence, Volume 20 (8), 2007, pp. 1070-1085.

Specht DF, Probabilistic Neural Networks and the Polynomial ADALINE as Complementary Techniques for Classification, IEEE Transactions on Neural Networks, Mar 1990, Volume 1, Issue 1, ISSN: 1045-9227, pp. 111-121.

Stam C, Nonlinear dynamical analysis of EEG and MEG: Review of an Emerging Field, *Clinical Neurophysiology*, Volume 116, Issue 10, October 2005, pp. 2266-2301.

Steuer R, Ebeling W, Bengner T, Dehnicke C, Hättig H, Meencke HJ, Entropy And Complexity Analysis of Intracranially Recorded EEG, *International Journal of Bifurcation and Chaos*, 2004, Volume 14, Issue 2, February 2004, pp. 815-823.

Subasi A, Automatic Detection of Epileptic Seizure using Dynamic Fuzzy Neural Networks, *Expert Systems with Applications*, Volume 31, Issue 2, Aug. 2006, pp. 320-328.

Subasia A, Ercelebi E, Classification of EEG Signals using Neural Network and Logistic Regression, *Computer Methods and Programs in Biomedicine*, Volume 78, 2005, pp. 87-99.

Swingler K, *Applying Neural Networks: A Practical Guide*, London: Academic Press, 1996.

Tandona R, Adaka S, Kayeb JA, Neural Networks for Longitudinal Studies in Alzheimer's Disease, Volume 36, Issue 3, 2006, pp. 245-255.

Tarassenko L, Khan YU, Holt MRG, Identification of Interictal Spikes in the EEG using Neural Network Analysis, *IEEE Proceedings: Science, Measurement and Technology*, 1998, Volume 145, Issue 6, pp. 270-278.

Teoh EJ, Tan KC, Tang HJ, Xiang C, Goh CK, An Asynchronous Recurrent Linear Threshold Network Approach to Solving the Traveling Salesman Problem, *Neurocomputing*, Volume 71, Issue 7-9, MAR 2008, pp. 1359-1372.

Tetko IV, Livingstone DJ, Luik AI, Neural Network Studies 1, Comparison of Overfitting and Overtraining, *J. Chem. Info. Comp. Sci.*, 35, 1995, pp. 826-833.

Theiler J, On the Evidence for Low-dimensional Chaos in an Epileptic Electroencephalogram, *Physics Letters A*, Volume 196, Issues 5-6, 2, January 1995, pp. 335-341.

Tito M, Adjouadi M, A Unified Approach to Seizure Detection on Subdural Recordings, Proceedings and Poster presentation at the 2nd Annual CAHSI (Computing Alliance for Hispanic Serving Institutions) Workshop, Miami, Florida, December 15-18, 2007, pp. 75-77.

Tilbury J, Eetvelt P, Garibaldi J, Curnow J, Ifeachor E, Receiver Operating Characteristic Analysis for Intelligent Medical Systems - A New Approach for Finding Confidence Intervals, *IEEE Transactions on Biomedical Engineering*, 2000, Volume 47, No.7, pp. 952-963.

Tito M, Ayala M, Yaylali I, Cabrerizo M, Barreto A, Rische N, Adjouadi M, Can EEG Processing Reveal Seizure Prediction Patterns?, *Proc. of the IASTED International Conference on Graphics and Visualization in Engineering (GVE 2007)*, January 3-5, 2007, Clearwater, Florida, USA, ISBN: 978-0-88986-625-6, pp. 47-52.

Tong S, Bezerianos A, Paul J, Zhu Y, Thakor N, Removal of ECG Interference from the EEG Recordings in Small Animals using Independent Component Analysis, *Journal of Neuroscience Methods*, Volume 108, Issue 1, 15 July 2001, pp. 11-17.

Towle VL, Hunter JD, Edgar JC, Chkhenkeli SA, Castelle MC, Frim DM, Kohrman M, Hecox KE, Frequency Domain Analysis of Human Subdural Recordings, *Journal of Clinical Neurophysiology*, Apr 2007, Volume 24, Issue 2, pp. 205-213.

Tran TA, Spencer SS, Javidan M, Pacia S, Marks D, Spencer DD, Significance of Spikes Recorded on Intraoperative Electroencephalography in Patients with Brain Tumor and Epilepsy, *Epilepsia*, Volume 38, Issue 10, Lippincott-Raven Publishers, pp. 1132-1139.

Ubeyli ED, Guler I, Detection of Electrocardiographic Changes in Partial Epileptic Patients using Lyapunov Exponents with Multilayer Perceptron Neural Networks, *Engineering Applications of Artificial Intelligence*, Volume 17(6), 2004, pp. 567-576.

Vapnik VN, *The Nature of Statistical Learning Theory*, New York : Springer, 1995.

Wahlberg P, Lantz G, Approximate Time-variable Coherence Analysis of Multichannel Signals, *Multidimensional Systems and Signal Processing*, Jul 2002, Volume 13, Issue 3, pp. 237-264.

Siegbert Warkentin S, Ohlsson M, Wollmer P, Edenbrandt L, Minthon L, Regional Cerebral Blood Flow in Alzheimer's Disease: Classification and Analysis of Heterogeneity, *Dement Geriatr Cogn Disord*, 2004, Volume 17, pp. 207-214.

Weigend A, On Overfitting and the Effective Number of Hidden Units, *Proceedings of the 1993 Connectionist Models Summer School*, 1994, pp. 335-342.

Werbos PJ, *The Roots of Backpropagation: From Ordered Derivatives to Neural Networks and Political Forecasting*, Series on Adaptive and Learning Systems for Signal Processing, Communications and Control, Wiley-Interscience, John Wiley & Sons, Inc., New York, NY 10158-0012, Feb 1994, ISBN 0-471-59897-6.

Widrow B, Lehr MA, 30 years of Adaptive Neural Networks: Perceptron, MADALINE, and Backpropagation, Proceedings of the IEEE, Sep 1990, Volume 78, Issue 9, ISSN: 0018-9219, pp. 1415-1442.

Wilding P, Morgan MA, Grygotis AE, Shoffner MA, Rosato EF, Application of Backpropagation Neural Networks to Diagnosis of Breast and Ovarian Cancer, Cancer Lett., 1994, Mar 15; 77(2-3), pp. 145-53.

Winston PH, Artificial Intelligence, 3rd Edition, Addison-Wesley, Menlo Park, 1992, pp. 16-22.

Winter R, Widrow B, MADALINE RULE II: a Training Algorithm for Neural Networks, IEEE International Conference on Neural Networks, 1988, Jul 1988, Volume 1, pp. 401-408.

Xu HY, Vukovich G, A Fuzzy Genetic Algorithm with Effective Search and Optimization, Proceedings of 1993 International Joint Conference on Neural Networks, 1993, IJCNN '93-Nagoya, ISBN: 0-7803-1421-2, Volume 3, pp. 2967-2970.

Yaylali I, Neural Network Based EEG Spike Detection System, University of Miami, Miami, FL, 1991.

Yen J, Langari R, Fuzzy Logic – Intelligence, Control, and Information, Prentice-Hall Inc., 1994, ISBN 0-13-525817-0.

You X, Adjouadi M, Guillen MR, Bernal B, Ayala M, Barreto A, Rische N, Gaillard WD, Automatic Decision Making Process for Classification of fMRI Patterns Integrating the Principal Component Analysis and the Laterality Index in Pediatric Epilepsy, Proc. 30th Annual Meeting of the Society for Medical Decision Making SMDM, Philadelphia, Pennsylvania, p. 154, Oct.19-22, 2008.

Zadeh LA, Fuzzy Sets, Information and Control 8, 338-353, 1965.

Zadeh LA, Probability Measures of Fuzzy Events, Jour. Math. Analysis and Appl. 23, 1968, pp. 421-427.

Zadeh LA, A Fuzzy-algorithmic Approach to the Definition of Complex or Imprecise Concepts, Int. Jour. Man-Machine Studies 8, 1976, 249-291.

Zadeh LA, Knowledge Representation in Fuzzy Logic, IEEE Trans. on Knowledge and Data Engineering 1, 1989, 89-100.

Zong N, Adjouadi M, Ayala M, Artificial Neural Networks Approaches for Multidimensional Classification of Acute Lymphoblastic Leukemia Gene Expression

Samples, WSEAS Transactions on Information Science and Applications, Volume 2 (8), August 2005, pp. 1071-1078.

Zong M, Adjouadi M, Ayala M, Optimizing the Classification of Acute Lymphoblastic Leukemia and Acute Myeloid Leukemia Samples using Artificial Neural Networks, Biomedical Sciences Instrumentation, Volume 42, pp. 261-266. Also presented at the Rocky Mountain Bioengineering Symposium, Terre Haute, Indiana, April 7-9, 2006.

VITA

MELVIN AYALA

EDUCATION

- 1980 - 1984 B.S. in Economic Engineering, University of Applied Sciences (former Zittau Engineering Institute), Zittau, Germany
- 1984-1987 Ph.D. in Economic Engineering, University of Applied Sciences (former Zittau Engineering Institute), Zittau, Germany
- 1987-1994 Professor at the Polytechnic Institute of Havana Electric Research and Test Center, Havana, Cuba
Duties: Taught undergrad and graduate courses in the fields of computer sciences, programming languages, and reliability of electrical systems.
- 1994 - 2000 Researcher at the Polytechnic Institute of Havana Electric Research and Test Center, Havana, Cuba
Duties: Consultant to the Cuban Electric Company, Development of programming tools for the CEC, taught graduate courses in programming languages and for electrical distribution systems.
- 2000 - 2001 Invited professor and researcher
State Univ. of São Paulo (UNESP), Guaratingueta Campus, Brazil
Duties: Taught undergraduate courses in programming languages and Fuzzy Logic. Research project also done under a grant from the Research Foundation of the State of São Paulo, Brazil.
- 2002 - present Research Supervisor
Center for Advanced Technology and Education, Florida International University, Miami, FL.

PUBLICATIONS AND PRESENTATIONS

(Ten out of 45 journal publications and 46 conference publications)

1. Cabrerizo M, Adjouadi M, Ayala M, Jayakar P, Subdural Interictal EEG Analysis for Extracting Discriminating Features Towards Electrode Classification Using Artificial Neural Networks, Brain Mapping Research Progress (Book Chapter 5) by Girard IC, Andre JS (Eds.), pp. 87-108, Nova Science Publishers - May 2009, ISBN: 1604567848 : 9781604567847.
2. Tito M, Cabrerizo M, Ayala M, Barreto A, Miller M, Jayakar P, Adjouadi M, Classification of electroencephalographic seizure recordings into ictal and interictal files using correlation sum, Computers in Biology and Medicine, Volume 39, 2009, Issue 7, pp. 604-614 .

3. Sesin A, Adjouadi M, Cabrerizo M, Ayala M, Barreto A, An Eye Gaze Tracking System Using Customized User Profiles to Help Persons with Motor Challenges Access Computers, Lecture Notes in Computer Science, K. Miesenberger et al., Eds., Springer-Verlag 2008 (in press).
4. Ayala M, Adjouadi M, Cabrerizo M, Barreto A, A Windows-based Interface for Teaching Image Processing, Computer Applications in Engineering Education (2007, in press).
5. Cabrerizo M, Adjouadi M, Ayala M, An Application of Eigensystem and Frequency Analysis in Brain Functional Mapping, Progress in Brain Mapping Research (Book Chapter 5), pp. 177-204, Nova Science Publishers, Inc. 2006, ISBN I-59454-580-4.
6. Simon A, Adjouadi M, Ayala M, A .NET Solution for Distributed Computing Applications, IEEE Potentials, Vol. 25 (2), pp. 24-28, March/April 2006.
7. Rossman M, Adjouadi M, Ayala M, Yaylali I, An Interactive Interface for Seizure Focus Localization Using SPECT Image Analysis, Computers in Biology and Medicine. Volume 36, Issue 1, pp. 70-88 January 2006.
8. Cabrerizo M, Adjouadi M, Ayala M, Nunez K, Jayakar P, Yaylali I, Integrated Study of Topographical Functional Based on an Auditory-Comprehension Paradigm Using an Eigensystem Study and Spectrum Analysis, Brain Topography, Vol. 17 (3), pp. 151-163, March 2005.
9. Adjouadi M, Cabrerizo M, Ayala M, Mirkovic N, Seizing lesions in 3-D, IEEE Potentials, Vol. 24, Issue 5, pp. 11-17, December 2005.
10. Adjouadi M, Cabrerizo M, Ayala M, Sanchez D, Jayakar P, Yaylali I, Barreto A, Detection of Interictal Spikes and Artifactual Data through Orthogonal Transformations, Journal of Clinical Neurophysiology, Vol. 22(1), pp. 53-64, January/February 2005.

PATENTS

1. NEURALSTUDIO: Artificial Neural Network Design and Evaluation Tool (with Dr. Malek Adjouadi), United States Patent No. 7,502,763, issued March 10, 2009 (available at <http://patft.uspto.gov/>).
2. ESEC: Software for Technical Studies in Secondary Distribution Circuits (with Dr. Raul Carvajal), Cuban Patent No. 05107-5107, issued 2000. Used officially in about 83 electric company branches in Cuba to analyze, design and test improvements on secondary distribution circuits.