A Neural Network Seismic Detector

Guilherme Madureira*, António E. Ruano**

*** Institute of Meteorology, Geophysical Center of S. Teotónio,
7630-585 Portugal
e-mail: guilherme.madureira@meteo.pt
** Centre for Intelligent Systems, University of Algarve,
8005-139 Portugal
e-mail: aruano@ualg.pt**

Abstract: This experimental study focuses on a detection system at the seismic station level that should have a similar role to the detection algorithms based on the ratio STA/LTA. We tested two types of neural network: Multi-Layer Perceptrons and Support Vector Machines, trained in supervised mode. The universe of data consisted of 2903 patterns extracted from records of the PVAQ station, of the seismography network of the Institute of Meteorology of Portugal. The spectral characteristics of the records and its variation in time were reflected in the input patterns, consisting in a set of values of power spectral density in selected frequencies, extracted from a spectrogram calculated over a segment of record of pre-determined duration. The universe of data was divided, with about 60% for the training and the remainder reserved for testing and validation. To ensure that all patterns in the universe of data were within the range of variation of the training set, we used an algorithm to separate the universe of data by hyper-convex polyhedrons, determining in this manner a set of patterns that have a mandatory part of the training set. Additionally, an active learning strategy was conducted, by iteratively incorporating poorly classified cases in the training set. The best results, in terms of sensitivity and selectivity in the whole data ranged between 98% and 100%. These results compare very favorably with the ones obtained by the existing detection system, 50%.

Keywords**:** *Seismic detector, neural networks, support vector machines, spectrogram.*

1. Introduction

There is a growing interest in seismology for increasing the speed and the reliability of the automatic processing of seismic data acquired by the monitoring system. The application of Artificial Neural Networks (ANN) in this field, and more specifically, the automatic detection of seismic events, has been tested for some years and is a promising path of current research.

We propose a seismic detection system, to be implemented at the seismic station, using ANN. This system should be able to distinguish segments of seismic records containing signal caused by local and regional events, from all other situations. The aim is to build a classifier that assigns one of two class periods of the seismic record of pre-determined fixed duration, Class 1, local and regional natural earthquakes, and class 2, all the other possibilities.

In the last two decades several researchers worked in the field of automatic seismic detection with neural networks. Some of those studies are presented below.

(Masotti et al., 2006) applied a Support Vector Machine (SVM) to classify volcanic tremor data at Etna volcano, Italy. Trained in a supervised way, the classifier should recognize patterns belonging to four classes; pre-eruptive, lava fountains, eruptive, and post-eruptive. 425 spectrogram based feature vectors were used for training. The system correctly classified 94.7 ± 2.4% of the data in validation.

In (Abu-Elsoud et al., 2004) an automatic system is proposed to discriminate between local earthquakes and local explosions in the Suez Gulf area, Egypt. The system is ANN-based and is composed of two modules; a feature extractor that quantifies the seismogram signatures using a Linear Prediction Code and a classifier to discriminate the seismic events. The data used is a set of 320 seismic events recorded by the Egyptian National Seismic Network; 142 records are explosions and 178 are local earthquakes. Validation results achieved 93.7% of correct classifications.

To detect distant seismic events automatically, (Tiira, 1999), proposes a Multi-Layer Perceptron (MLP) trained with the Error-Back-Propagation algorithm. The entries in this network are instantaneous values of STA/LTA (see Section 2.2) calculated with 4 different windows of STA, in 7 frequency bands. 193 distant seismic events were used in the training process. Comparing with the Murdock-Hutt detector (Murdock and Hutt, 1983), this system detected 25% more events, and produced 50% less false alarms.

(Dai and MacBeth, 1997) proposed a Back-Propagation Neural Network (BPNN) to identify P (Primary) and S (Secondary) arrivals (Udías, 2000) from three-component recordings of local earthquake data. The BPNN was trained by selecting trace segments of P and S waves and noise bursts, converted into an attribute space based on the Degree of Polarization (DOP). 1363 seismic records were used for training and validation. Compared with a manual analysis, the trained system can correctly identify between 76.6% and 82.3% of the P arrivals, and between 60.5% and 62.6% of the S arrivals.

The detection of seismic events was the objective of the study presented in (Wang and Teng, 1995). Two ANN were trained in supervised mode with different types of inputs: In one case the ratio STA/LTA was used, the other used spectrogram as input feature. Experiments have shown that these systems performed better than those algorithms based on a threshold of the STA/LTA ratio.

In this work we used data collected from the seismic station PVAQ[1], located in Vaqueiros, Algarve, in southern Portugal.

[1] In general, Portuguese seismic stations begin with a “P”, that stands for Portugal, followed by an abbreviation of the location name, in this case “VAQ” stands for Vaqueiros.

The structure of the paper is as follows. In section 2, the procedures used for data collection and feature extraction are described. The training methods used in the experiment are also indicated in this section. In section 3 the experiments are described and the results analyzed. Conclusions and future work are expressed in section 4.

2. Data and Training Methods

2.1. Input Data

Non-stationary signals occur naturally in many real-world applications: Examples include speech, music, biomedical signals, radar, sonar and seismic waves. Time-frequency representations such as the spectrograms are important tools for processing such time-varying signals. In this work, the spectrogram is used as the first stage of earthquake detection.

The Power Spectrum Density (PSD) is estimated using periodogram averaging (Welch, 1967). Only positive frequencies are taken into account (the so-called one-sided PSD). PSD values are slightly smoothed by taking the average of PSD values in a constant relative bandwidth of 1/10 of a decade. The procedure to achieve that smoothness was as follows: Let $P(f)$ be the PSD values in some set of discrete frequencies F. Starting with the lowest frequency of F, (f_{min}), we created a sequence of frequencies separated by 1/10 of a decade,

$$f_k = f_{\min} 10^{\frac{k-1}{10}}, \quad k = 1, 2, \ldots \tag{1}$$

We then split F into disjoint subsets D_k,

$$D_k = \{f\} : f_k \le f \le f_{k+1}, f \in F, k = 1, 2, \ldots \tag{2}$$

each set D_k is associated with a frequency f_k as defined above. The smoothed PSD, $P_s(f_k)$, is given by,

$$P_S(f_k) = \frac{1}{\# D_k} \sum_{f \in D_k} P(f) \tag{3}$$

We have divided segments of 120 seconds into 5 non-overlapping intervals. For each one of them we computed the PSD. This is done with standard Matlab functions. We then picked the power at 6 frequencies 1, 2, 4, 8, 10 and 15 Hz. This means that 30 different features will be used for the classifier. This was a constraint that we imposed, in order to limit the classifiers complexity. Fig. 1 illustrates a seismic-record and its spectrogram, highlighting the frequencies selected.

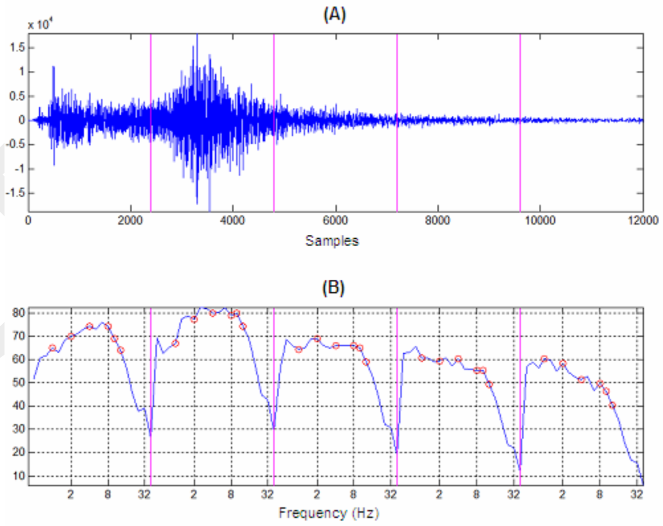

Figure 1. (A) 120 sec of seismic record (B) Spectrogram

In most experiments, a Butterworth digital high-pass filter was applied to the signal previous to PSD computation. The cut-off frequency was 0.5 Hz and the order of the filter was 5. This procedure intended to remove low frequency content from the spectrum, since for local and regional seismic events those frequencies are out of the main bandwidth of interest.

2.2. Target Data

Seismic data, previously classified was collected from the PVAQ station of the seismic monitoring system of the Institute of Meteorology of Portugal (IM). Seismic data was classified by seismologists of the National Data Center (NDC) at IM. The seismic detector used at a station level is a standard STA/LTA ratio based detector (Stewart, 1977). Fig. 2 outlines the operation of such a detector.

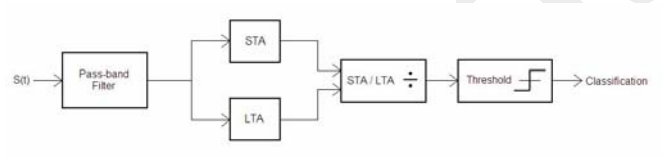

Figure 2. Block diagram of a typical STA/LTA detector

The input data is band-pass filtered to maximise sensitivity within a specific frequency band of interest, and to reject noise outside this band. Averages of the modulus of signal amplitude are computed over two user-defined time periods, a short time average (STA) and a long time average (LTA), and the ratio of the two, (STA/LTA), at each sample point is computed. If this ratio exceeds a user-defined threshold, then a trigger is declared, and the system remains in a triggered state until the ratio falls below the defined threshold.

These detectors based on the ratio STA/LTA at the seismic station show in general very modest performance, i.e., large numbers of non detected seismic events and several false alarms. However, a seismic network can drastically improve the overall performance considering clusters of stations. The likelihood of noise events occurring in a given time interval at various stations is very small, thus reducing the likelihood of making false alarms. In addition, an event that is not detected by a particular station is likely to be detected by other stations of the group, thereby increasing dramatically the ability of detection. However, the automatic system at the NDC is always supervised by seismologists.

2.3. Collected Data

From the year of 2007, 2903 examples were collected, 502 representing the positive class (classified as earthquake by the seismologists at NDC, and where seismic phases were identified in the PVAQ records), and the other 2401 classified as non-seism. In the former case, the station detection system miss-classified 50% of the events. In the latter class, 50% of the examples were randomly selected representing events that triggered the detection system, but that were not classified as seismic by the NDC, while the rest of the examples were selected randomly, neither coinciding with events detected by the system nor classified as earthquakes by the NDC. This way, the station automatic detected system achieved values of 50% of Sensitivity and Specificity (measures introduced latter) in the data collected.

2.4. Training Methods

In this work, MLPs were used as classifiers. We shall briefly describe here the training method employed. For more information, the reader is referred to, for instance (Ruano et al., 2005).

First of all we assign to each positive example the value of +1, and to each negative example, the value of -1. Input data is scaled and the classifier nonlinear parameters are initialized with a stochastic procedure which does not exacerbate the condition number of the Jacobean matrix of the model.

Parameter estimation is achieved by applying the Levenberg-Marquardt algorithm (Ruano et al., 1992) for the minimization of a criterion that exploits the separability of the classifier parameters, as linear parameters are used in the output layer (Ruano et al., 1991). This process is applied to the training data, and terminates whether a local minimum is found, or the performance in another set, denoted here as a test set, deteriorates. This is the well-known method of early stopping (Haykin, 1999).

As indirectly, the test set is used in the determination of the classifiers, their performance is assessed in a third data set, denoted here as the validation set.

3. Results

3.1. First Experiment

The first experiment was conducted by assigning, randomly, 60% of the data to the training set, 20% to the test set, and 20% to the validation set. It was only ensured that a similar percentage of positive cases was assigned to each data set. The training set consisted of 1744 examples, with 307 positive cases; the test set had 582 examples, with 99 positive cases; the validation set consisted of 577 events, with 96 positive cases.

In this first experiment, 20 different topologies of MLPs were tried, each one with 20 different parameters initializations. Moreover, the use (or not) of the filter described above was tested, resulting in 800 different classifiers.

The results are presented in terms of the Sensitivity, or Recall (R) criterion, defined as:

$$R = \frac{TP}{TP + FN}, \tag{4}$$

and in terms of Specificity (S) criterion, defined as…

$$S = \frac{TN}{FP + TN}, \tag{5}$$

where *TP*, *TN*, *FP* and *FN* denote the number of True Positives, True Negatives, False Positives and False Negatives, respectively. Moreover, these criteria are applied to the training, test and validation sets, separately, and to all the data. As the classification is casted as a multi-objective problem, we do not have a single optimum; instead a set of Non-Dominated (ND) solutions is obtained, where the elements have the property that no one is better (larger in this case) in all objectives than the other solutions belonging to the set. The following tables show the ND solutions found, for the three data sets, individually considered.

A line in italic indicates that the same ND classifier is present in the training and in the validation sets, while an underlined line indicates that a common ND classifier is obtained in the test and in the validation sets. Please note that as 20 different initializations were conducted for the same topology, equal entries in the topology column is not an indication that the same classifier is used. A mark in the column labelled as *F* indicates if filtering of the input data has been applied. The columns labelled as *R(All)* and *S(All)* show the Recall and the Specificity values computed for the whole data (the union of the training, test and validation data sets). The topology column shows the number of neurons in the first and the second hidden layers.

Table 1. Training set

Topology	F	R	S	R(All)	S(All)
[7 2]		91.86	99.23	92.43	99.25
[5 7]		96.74	96.66	96.81	97.17
[4 16]		96.09	97.56	94.82	97.88
[5 11]		93.49	98.96	93.43	98.96
<u>[7 2]</u>		<u>94.79</u>	<u>98.12</u>	<u>95.62</u>	<u>98.25</u>
[6 5]		94.46	98.75	94.62	98.88
[6 3]	*	92.18	99.10	93.23	99.13
[5 7]	*	95.44	98.05	95.82	98.50
[7 2]	*	96.42	97.49	96.61	98.00
[4 15]	*	92.51	99.03	93.23	99.04

Table 2. Test set

Topology	F	R	S	R(All)	S(All)
[7 2]		96.97	99.38	94.22	98.71
[5 8]		98.99	98.96	94.82	97.33
[4 19]		95.96	99.79	93.82	98.04
[6 2]		*94.95*	*100.00*	*94.82*	*98.79*

Table 3. Validation set

Topology	F	R	S	R(All)	S(All)
[6 2]		*97.92*	*99.17*	*94.82*	*98.79*
[7 2]		98.96	98.13	95.62	98.25
[6 3]	*	95.83	99.38	94.82	98.96
[7 2]	*	90.63	99.79	90.44	98.67

If we perform the same analysis for the three data sets together (i.e., considering as criteria the Selectivity and the Specificity for the training, the test and the validation sets, and subsequently determining the ND solutions), we obtain the union of the ND solutions for the three data sets considered separately, plus a significant number of additional Pareto solutions. In the total, 51 ND solutions were obtained. If we select the classifier by the total number of misclassifications (both positive and negative) in the whole data, 3 models achieve the smallest number, 51, in the full 2903 examples. One of the three solutions is shown in the 3rd line of Table 3, and the other two belong to additional ND solutions.

The results can also be presented as a ROC (Receiver Operating Characteristics) curve (Swets, 1988). The next three figures present these results, where, in every case, the ND solutions obtained considering the corresponding data set are shown as a red circle, and the ND solutions, considered the 6 criteria, are shown as blue diamonds.

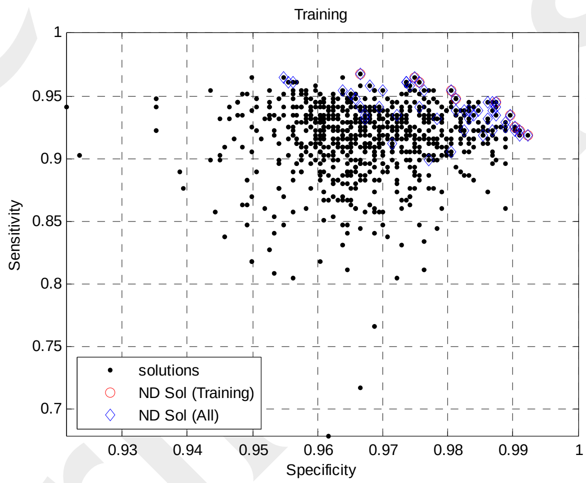

Figure 3. ROC for the training set

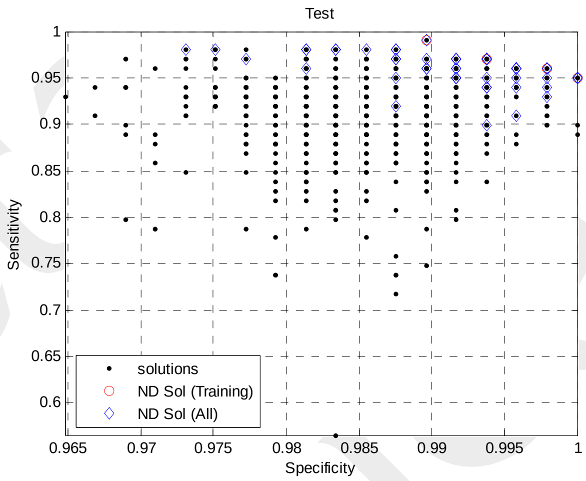

Figure 4. ROC for the test set

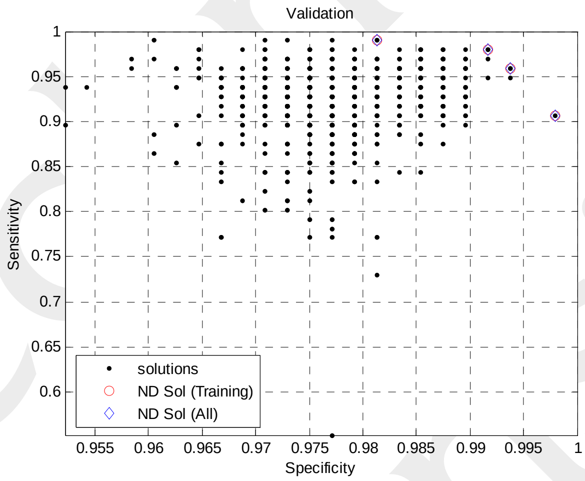

Figure 5. ROC for the validation set

We were therefore able to obtain classifiers with Recall and Specificity values above 95% (compared with the 50% values obtained by the existing detection system), and with a total number of misclassifications in the order of 50, compared with 1450, achieved by the existing system.

These results are also able to highlight that the use or not of the filter did not produce any significant difference. Filtered data will be used from now on.

3.2. Support Vector Machines

Another set of experiences regarding different partitioning of data between the training, test and validation sets was conducted. First of all, an approximate convex hull of the input data has been obtained, and the examples that lie in the hull were integrated in the training set. In order to maintain an approximate distribution of 60%, 20% and 20% of the data to the three sets, examples of the original training set were moved to the other two sets. With this data partitioning, a Support Vector Machine (SVM) classifier, with a Gaussian kernel, was experimented. The implementation described in (Frieß et al., 1998) was used.

In this case the examples in the test and validation set were used as a single validation set. With a spread value of 0.237, the following results were obtained:

Table 4. SVM performance

SVs	R	S	R(All)	S(All)
583	100.00	100.00	99.62	99.35

Subsquently, a form of active learning (Cohn et al., 1994) was applied. The examples badly classified were incorporated in the training set, and randomly removed the same number of examples to the validation set, provided they were not in the approximate convex hull previously determined. This procedure was repeated three times. The results are presented in Table 5.

Table 5. SVM performance with active learning

SVs	R	S	R(All)	S(All)
609	100.00	100.00	99.72	99.66
626	100.00	100.00	99.76	99.72
640	100.00	100.00	100.00	99.93

This represents an almost perfect performance (only 2 misclassifications in the whole data). The major problem is the large complexity of the classifier, consisting of 640 support vectors. We therefore tried, with this new partitioning of data, to improve the performance of the MLP classifiers.

3.3. Further experiments with MLPs

We used 20 different topologies, each one with 10 different initializations. The non-dominated solutions obtained are shown below.

Table 6. Training set

Topology	R	S	R(All)	S(All)
[6 4]	**97.48**	**99.30**	**96.61**	**99.50**
[6 2]	99.37	99.09	96.81	98.82
[5 11]	96.21	99.44	92.23	99.29
[5 12]	95.27	99.79	92.43	98.83

Table 7. Test set

Topology	R	S	R(All)	S(All)
[4 20]	98.90	99.38	97.21	98.46
[5 8]	100.00	98.97	98.41	97.83
[6 4]	**94.51**	**99.79**	**96.61**	**99.50**
[4 19]	96.70	95.59	86.65	96.50

Table 8. Validation set

Topology	R	S	R(All)	S(All)
[6 2]	100.00	99.80	97.81	98.83

A line in bold indicates that the same ND classifier is obtained, considering the training set and the test set. The number of ND solution achieved, considering the three data sets, is 32. The best solution, in terms of the total number of miss-classifications, has a topology of [5 9], and it is not present in tables 6-8.

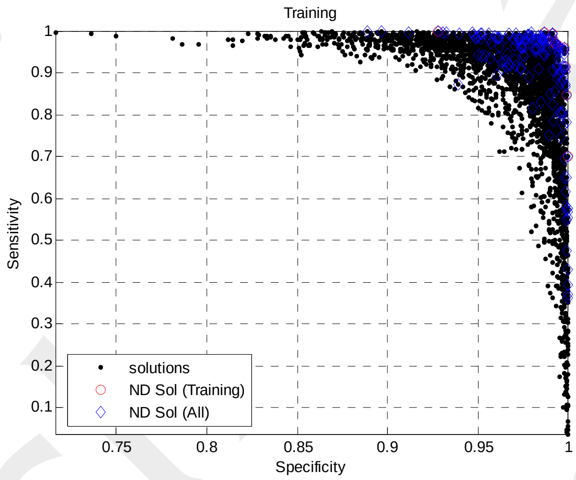

Figure 6. ROC for the training set

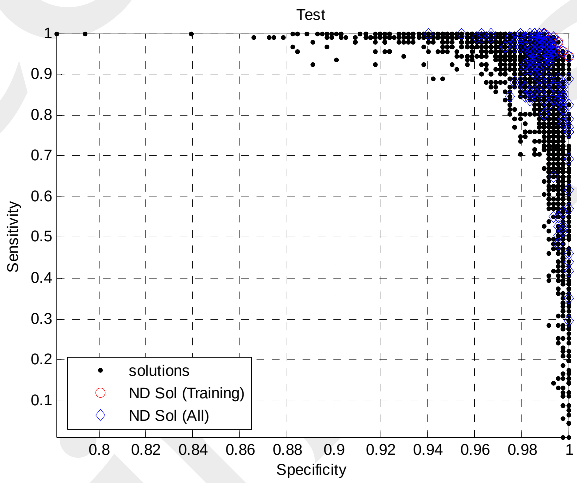

Figure 7. ROC for the test set

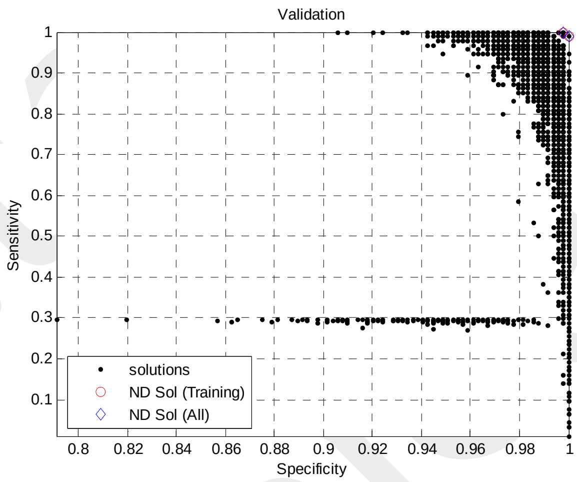

Figure 8. ROC for the validation set

We were able, with just a different data partitioning, to reduce the number of miss-classifications from 51 (please see Section 3.1) down to 29. This classifier presents a complexity, in terms of the number of parameters, of 219, compared with the solution obtained in Section 3.2, with 640 support vectors.

Further experiments were conducted, varying the decision threshold of the classifiers. No improvements, however, were obtained. The following figures show the ROC curves, for the training, test and validation sets.

Conclusions

With the same data that produced 50% Sensitivity and Selectivity values in an existing detection system, based on the LTA/STA ratio, we were able to obtain, in a first step, values greater than 95% for the two criteria. Using an active learning technique, we were able to improve the performance of our MLP classifiers to 98%. An SVM classifier was able to achieve almost perfect classification, albeit at the expense of a large complexity.

Although the results are encouraging, the work described in this paper must be considered as preliminary. At present the performance of the neural classifiers is being assessed in the whole 2007 record of the station employed. The analysis of the results will enable to construct a better off-line classifier. Then, our attention will be focused in on-line learning methods, so that the classifier learns with its on-line performance. Additional features can also be considered and a search for the best to use, together with the classifier topology, can be conducted by meta-heuristics. Finally, the use of data from different stations will be considered.

References

[1] Abu-Elsoud, M. A., Abou-Chadi, F. E. Z., Amin, A. E. M., Mahana, M.: *Classification of Seismic Events in Suez Gulf area, Egypt Using Artificial Neural Network*, ICEEE'04: 2004 International Conference on Electrical, Electronic and Computer Engineering, Proceedings, (2004), pp. 337-340.

[2] Cohn, D., Atlas, L., R, L.: *Improving Generalization with Active Learning*, Machine Learning, 15, (1994), pp. 201-221.

[3] Dai, H. C., MacBeth, C.: *The Application of Back-Propagation Neural Network to Automatic Picking Seismic Arrivals from Single-Component Recordings*, Journal of Geophysical Research-Solid Earth, 102, (1997), pp. 15105-15113.

[4] Frieß, T., Cristianini, N., Campbel, C.: *The Kernel Adatron Algorithm: A Fast and Simple Learning Procedure for Support Vector Machines*, 15th Intl. Conf. Machine Learning. Morgan Kaufmann Publishers, (1998).

[5] Haykin, S.: *Neural Networks: A Comprehensive Foundation*, Prentice Hall, (1999).

[6] Masotti, M., Falsaperla, S., Langer, H., Spampinato, S., Campanini, R.: *Application of Support Vector Machine to the Classification of Volcanic Tremor at Etna*, Italy. Geophysical Research Letters, 33. (2006).

[7] Murdock, J., Hutt, C.: *A New Event Detector Designed for the Seismic Research Observatories*, USGS Open-File-Report, (1983).

[8] Ruano, A. E., Ferreira, P. M., Fonseca, C. M.: *An Overview of Nonlinear Identification and Control with Neural Networks*, In Ruano, A. E. (Ed.) Intelligent Control using Intelligent Computational Techniques. IEE Control Series, (2005).

[9] Ruano, A. E. B., Fleming, P. J., Jones, D. I.: *A Connectionist Approach to PID Autotuning*, IEE Proceedings-D Control Theory and Applications, 139, (1992), pp. 279-285.

[10] Ruano, A. E. B., Jones, D. I., Fleming, P. J.: *A New Formulation of the Learning-Problem for a Neural Network Controller*, Proceedings of the 30th Ieee Conference on Decision and Control, Vols 1-3. (1991).

[11] Stewart, S. W.: *Real Time Detection and Location of Local Seismic Events in Central California*, Bulletin of Seismological Society of America 67, (1977), pp. 433-452.

[12] Swets, J.: *Measuring the Accuracy of Diagnostic Systems*, Science, 240, (1988), pp. 1285-1293.

[13] Tiira, T.: *Detecting Teleseismic Events using Artificial Neural Networks*, Computers & Geosciences, 25, (1999), pp. 929-938.

[14] Udías, A.: *Principles of Seismology*, Cambridge University Press. (2000).

[15] Wang, J., Teng, T. L.: *Artificial Neural-Network-Based Seismic Detector*, Bulletin of the Seismological Society of America, 85, (1995), pp. 308-319.

[16] Welch, P. D.: *Use of Fast Fourier Transform for Estimation of Power Spectra: A Method Based on Time Averaging Over Short Modified Periodograms*, IEEE Transactions on Audio and Electroacoustics, AU15, 70. (1967).