An Hybrid Approach Based on Neural Networks and Regression Tree Models for Fast Dynamic Security Assessment

Helena Vasconcelos (1,2)
hvasconcelos@inescporto.pt

J. A. Peças Lopes (1,2)
jplopes@inescporto.pt

INESC Porto – Instituto de Engenharia de Sistemas e Computadores do Porto
FEUP – Faculty of Engineering of the University of Porto
Porto, Portugal

Abstract: This paper presents a new hybrid automatic learning approach, which combines artificial neural networks (ANN) and regression trees (RT), to perform on-line dynamic security assessment of power systems. In the proposed method, the RT is firstly used to split the vast amount of knowledge data that describes a security problem into several less spread and disjoint problems. Then, an ANN is trained for each of these new smaller problems, resulting in a tree structure with an ANN predicting function associated to each leaf. Moreover, the capability of the RT to perform feature subset selection before ANN training is also tested. With this new method, the advantages of the two techniques are exploited in order to obtained a more accurate model without compromising prediction time. The quality of the approach is illustrated through its application to a major security problem of the power system of Madeira Island (Portugal).

Keywords: dynamic security assessment, isolated power systems, neural networks, regression tree models, renewable power generation.

I. INTRODUCTION

In the last years, automatic learning techniques have been exploited to perform security assessment functions in advanced control systems specially designed to achieve secure and economical operation of isolated power systems with increased renewable penetration. Such work has been developed within the framework of the MORECARE European R&D project [1]-[2], which resulted in the installation of a MORECARE prototype system in Madeira and Crete Island (in Portugal and Greece).

One of the key features of this control applications is related with the capability to assure a robust operation of the power system regarding to expected disturbances, which is achieved by performing on-line and accurate prediction of the system dynamic security for present and future operating scenarios, and also by providing fast preventive control measures when insecurity is detected. In the ongoing project, these security functions where successfully implemented, namely by exploiting Decision Trees [3], backpropagation Artificial Neural Networks [4][5] (ANN) and Kernel Regression Trees [6] (KRT). This last approach, integrates the classical Regression Trees (RT) [7] with kernel regression models to make prediction in the tree leafs.

From the authors previous research work, it was possible to identify that the most performing security assessment structures were ANN. The KRT, although showing a comparable accuracy relatively to ANN, were more demanding in computational requests, namely in terms of [2]:

- memory, since the learning set needs to be stored together with the security rules;
- computational times, because each time a security prediction is performed a regression procedure needs to be performed, which may be a very burden task.

Moreover, as presented in [8], ANN also offer simple and effective mechanisms of computing the derivatives of the system security indices (the ANN outputs) with respect to the power system operating conditions (the ANN inputs), which allows the application of gradient based methods for preventive control purposes.

Although these advantages, it is known from bibliography [9] that backpropagation ANN may lose predicting performance in the boundary of knowledge discontinuities. These discontinuities are assumed to be removed by the training of a regression tree, and this fact is exploited in this research.

Based on this assumptions, an innovative security structure type was developed, named here RT+ANN, which integrates artificial neural networks (ANN) in the regression tree (RT) leafs, in order perform on-line dynamic security assessment of electrical power systems.

As already demonstrated with the KRT approach in [6], by considering a RT+ANN structure, the RT accuracy is with no doubt improved by including a function that highly approximate non-linear functions in the tree leafs. The accuracy improvement regarding to a single ANN is not so obvious, depending of the properties of the security surface to evaluate, and therefore performance evaluation tests must be performed.

In this paper, results obtained from the application of the proposed approach to a major security problem of the Madeira power system are presented. From this case study, as initially expected, an increased accuracy was experienced by the application of RT+ANN security structures, when compared to isolated application of ANN or RT.

II. MAIN STEPS TO GENERATE A RT+ANN STRUCTURE

The following three main steps were followed in order to obtain an accurate RT+ANN structure (i.e. a tree structure with ANN predicting functions in the leafs):

- STEP 1: Design of a regression tree (RT) for all the security knowledge, i.e. a binary tree with the mean value as the model to make prediction at the tree leafs.
- STEP 2: ANN training for all the security knowledge, i.e. an ANN for the root of the tree structure;
- STEP 3: Training an ANN for each of the tree leafs.

Previous to ANN training, the interpretable "if-then-else" rules of the RT structure may also be explored in order to perform feature subset selection. All these steps are performed off-line, being the final product of the procedure – the RT+ANN security structure – to be used within the on-line security assessment framework of power systems control centers. These steps are described next.

A. STEP 1: Design of a RT for the main security problem

1) Growing a Very Large RT

The design of the RT is made by applying a recursive partitioning algorithm, which successively divides the learning knowledge data (Learning Set - LS) into mutually exclusive subsets, aiming to minimize knowledge dispersion. Each tree mode is divided by the application of a splitting test of the following from:

$$\{feature_k(OP) > u_k\}?　\tag{1}$$

where:

$feature_k(OP)$: value of feature k in the operating point;

u_k : optimal threshold value for the chosen feature.

By applying this test to all the set of operating points (OP) in node t , two successor nodes are created, t_L and t_R , which correspond to the two possible instances of the test. This must be performed according to an "optimal" splitting test, which corresponds to the one that provides a maximum amount of information. This is equivalent to divide the LS into disjoint regions, in such a way that in each region the security index is as constant as possible. By considering the mean value as the predicting function in the tree leafs, the goal becomes into reducing the knowledge variance. Therefore, the amount of information provided by each splitting test s in node t is measured by:

$$\Delta var(s,t) = var(t) - \frac{N(t_L)}{N(t)} var(t_L) - \frac{N(t_R)}{N(t)} var(t_R)　\tag{2}$$

where:

$N(t)$: number of operating points stored in node t ;

$var(t)$: variance of the security index in node t .

A detailed description of the Regression Tree method may be found in [7] and [10].

2) Overfitting Control by Tree Pruning

In order to avoid overfitting problems, after growing a very large tree T_{max} , which is supposed to overfit the LS, a pruning algorithm is applied in order to look for the right sized tree.

Even for a moderate sized tree, there is an extremely large number of possible pruned trees and an even larger number of distinct ways of pruning up it to the root node. Regarding this, a selective pruning process is applied, that generates a reasonable number of pruned trees of T_{max} , with decreasing size, such that each subtree is the "best" pruned tree in its size range. The existing approaches differ in the measure used to define the next node to prune, i.e. the weakest non-terminal node to be replaced by a leaf. Namely, in this research the following criteria were used, one each time, to define the weakness of the tree nodes:

MEC – Minimal error-complexity, by CART [7]:

$$min\{MSE_{LS}(t) - MSE_{LS}(T_t)\} + max\{\#\tilde{T}_t - 1\}　\tag{3}$$

MEL – Minimal Error Loss, by Bohanec and Bratko [11]:

$$min\{MSE_{LS}(t) - MSE_{LS}(T_t)\}　\tag{4}$$

newMEC – A new proposed variant of MEC:

$$min\{MSE_{LS}(T - T_t) - MSE_{LS}(T)\} + max\{\#\tilde{T}_t - 1\}　\tag{5}$$

newMEL – A new proposed variant of MEL:

$$min\{MSE_{LS}(T - T_t) - MSE_{LS}(T)\}　\tag{6}$$

LSS – Lowest Statistical Support, by Luis Torgo [10]:

$$min\{N(t)\}　\tag{7}$$

MCV – Maximal Coefficient of Variation, by Luis Torgo [10]:

$$max\{CV(t)\}　\tag{8}$$

where:

$MSE(t)$: Mean Squared Error of the security index in node t , given by

$$MSE(t) = \frac{1}{N(t)} \sum_{OPi \in t} (y_i - \hat{y}_i)^2　\tag{9}$$

$CV(t)$: Coefficient of Variation of the security index in node t , given by

$$CV(t) = SE[MSE(t)] / MSE(t)　\tag{10}$$

$SE(MSE(t))$: standard error estimation of $MSE(t)$.

T : tree structure;

T_t : subtree of T , which results from having node t as root;

$T - T_t$: subtree of T , which results from pruning node t ;

$\tilde{T}$: set of all the leafs in T ;

$\#\tilde{T}$: number of leafs in T ;

y_i : Real value of the security index for OP_i;

$\hat{y}_i$: Predicting value of the security index for OP_i, being, in the applied RT approach, equal to the mean value of y in the set of OP stored in t .

At the end of each pruning process, it results a sequence of nested trees with decreasing size, being each tree accuracy evaluated by the MSE obtained from applying the structure to an independent testing set (TS) – i.e. the $MSE_{TS}(T)$.

In order to choosing the "best" pruning process, and namely the "best" pruned tree among the generated ones, an exhaustive process should be performed, by training a RT+ANN structure for each tree and evaluate TS predicting error, which we believe would be a very demanding and time consuming task. In a first approach, only the smaller trees were considered in order to derive a RT+ANN structure. The "best" candidates trees were selected from the pruning process that provides the most accurate RT structure (i.e. the lowest $MSE_{TS}(T)$) for the range of less complex trees.

3) Exploiting the RT Structure for Feature Selection

Before performing ANN training, a feature subset selection may be performed by exploiting the RT structure. Based on the information gain provided by each feature for growing the most accurate RT, the following two new techniques were tested for measuring feature relevance:

FSS1: The information gain of each feature is calculated by summing the obtained Δvar, only for the splitting tests where the feature was applied to perform tree division. This approach was inspired in the procedure presented in [12], where decision trees were suggested to provide an attribute ranking regarding its contribution to the total tree information.

FSS2: The information gain of each feature is calculated by summing the maximum obtained Δvar with this feature for each of the tree divisions. This second approach was also considered based on the knowledge that the tree structure of a RT (or Decision Tree) is unstable [7] (i.e. small changes in the LS may lead to much different tree structures, however achieving almost the same accuracy). If a feature does not appear in any of the tree splits does not mean that it is not relevant for the problem. In fact, a relevant attribute can be constantly masked by another and thus never be chosen for splitting the nodes.

These information gains are then normalized between 0 and 1, and the features ordering by decreasing values of information gain. From this ranking, only the most relevant features will be selected for training the ANN structures.

B. STEP 2: ANN Training for the main security problem

1) Artificial Neural Networks

The applied ANN approach was a multi-layer feedforward networks with a tan-sigmoid transfer function. ANN parameters, i.e. the network weights and biases, were found through the Adaptive Backpropagation algorithm [5][8][9] by performing batch training. This training algorithm is based on the traditional Backpropagation [4] where, instead of a fixed and unique learning rate, a different adaptative learning rate it uses for each weight and bias, which provides a much faster learning process.

Besides the learning set, which was used for computing error gradients and updating the network weights and biases, a testing set was considered to perform overfitting control. According to the applied technique, besides considering a maximum number of epochs, when the testing error increases consecutively for a specified number of iterations, the training is stopped.

In order to remove offset and measurement scale problems, before starting ANN training, the learning and testing patterns were normalized to have zero mean and a standard deviation of one. In this primary stage, the ANN parameters are randomly initialized between –1 and 1.

2) ANN Training Procedure

From tests performed in the considered power system security problem, the accuracy of the trained ANN showed to be very sensitive to the initial parameters values. Therefore, a special care regarding this issue was considered when defining the RT+ANN training procedure. Namely, the following was considered for training the main ANN (i.e. the ANN trained for all the security knowledge):

- m different sets of initial random ANN parameters;
- for each initial set of ANN parameters, n different initial learning rates.

In the end it results a set of $m \times (n+1)$ different ANN, where the one with the lowest $MSE_{TS}(ANN)$ is selected – the *ANNroot1*. The initial and final parameters of this ANN will be used to initialize the ANN training in each tree leaf. Because of this, the initial error surface of the security structure will be closer to the desired one, and therefore, besides improving the training time, this also usually increases ANN final accuracy [9].

In order to improve the main ANN accuracy, after obtaining the first ANN (*ANNroot1*), a new retraining process was considered initialized with the last obtained parameters. If an accuracy improvement is achieved, then this second ANN will be used for comparing the performance results between the ANN and RT+ANN approaches.

A fixed and identical structure was considered for all the trained ANN, where the more suitable one was obtained from performing several ANN trainings with different structures, for the main security problem.

C. STEP 3: ANN training for each tree leaf

The followed procedure was considered for training an ANN for each tree leaf:

- starting from the initial parameters of *ANNroot1*, n training process are performed by considering n different initial learning rates;
- starting from the final parameters of *ANNroot1*, n training process are equally performed;

At the end it results a set of $2 \times (n+1)$ different ANN security structures, where the one with the lowest $MSE_{TS}(ANN)$ is selected. For this procedure, only the subsets of LS and TS associated to the tree leaf were considered. Before ANN training, these sets were normalized only regarding to the learning and testing data stored in the tree leaf.

III. CASE STUDY AND RESULTS

A. Madeira Security Problem

The quality of the developed approach is illustrated here through its application to the case of Madeira island. This power system is an isolated grid with a peak load of 120 MW and a minimum load of 42.8 MW, comprising utility owned and independent thermal units (134 MW), one independent waste to energy unit (6.4 MW), utility hydro units (46 MW), and utility owned and independent wind parks with asynchronous generators (15.3 MW). The single line diagram of the transmission and generation system is presented in Figure 1. In this figure, *Pw* regard to wind parks, and *Pc* to conventional power plants (hydro, waste to energy and thermal). Due to space limitations the system data cannot be included but it can be obtained upon request.
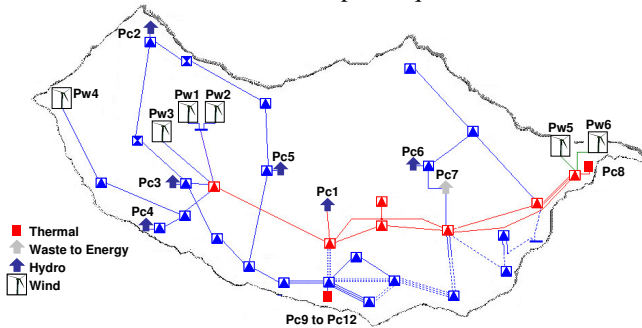

Figure 1 – Single line diagram of Madeira Power System

The RT+ANN approach was applied to this power system, in order to derive security structures to be used for security assessment purposes related with a critical pre-selected disturbance. The following disturbance was considered: short-circuit in the eastern side of the island, causing the disconnection of all wind (Pw1 to Pw6), and Pc7 to Pc8 power plants. This regards to situations where the dynamic security of the system is reduced in case of short-circuits that take place near to power production facilities, leading to these facilities disconnection (due to under-voltage conditions). As wind parks sites are more expose to adverse climatic conditions, short-circuits usually take place near these facilities. This disturbance was selected by the utility as one of the most important to be included as a security restriction within the system operating policies, and therefore to be considered in a security assessment process. In fact, this is a particular severe disturbance that may provoke large frequency drops, leading to load shedding activation, or to system instability. Based on the set-point values of the installed load shedding relays, the system was considered to lose security if the negative frequency deviations (*Δf*) go bellow –2 Hz.

B. Madeira Knowledge Data

The generation of a representative knowledge data set of the system frequency dynamic behavior, for the disturbance under consideration, was a key stage for the success of applying any "learning from example" approach. For generating the Madeira data set, an innovative generation procedure was adopted aiming at building an adequate knowledge base, able to describe implicitly the system dynamic security behavior of power systems with large wind

power production. A description of this procedure may be found in [8].

The final vector of ANN inputs was selected based on engineering judgment, comprises 34 variables with information about:

- *Pload*: total active load;
- *Pwj, Nwj*: active power produced and number of operating units in each set of equal wind generators connected in the same power plant;
- *Pci, SRi*: active power produced and spinning reserve in each set of equal conventional generators connected in the same power plant;
- *SR – Pgloss*: system spinning reserve minus total power loss.

Since the main goals of this research are to perform accurate security assessment and to apply preventive control procedures, the selection of these features was based on the following criteria:

- To be related with the dynamic phenomena under study;
- The number of features should be as low as possible without losing relevant information. (the concept of "equivalent machine" was used to group similar generators operating in parallel in the same plant);
- To use independent (or easy related) and dispatchable variables for further control use.

Each operating point (or pattern) of the data set is therefore characterized by these 34 features and the security index *Δfmin*, i.e. the minimum value reached by the negative frequency deviations that results from the considered disturbance.

A total amount of 7083 patterns were obtained, where 70% of the data was randomly extracted for the training purposes (learning set), and the remaining 30% for performance evaluation purposes (testing set). The number of obtained secure/insecure patterns are summarized in Table I.

Table I - Number of obtained secure/insecure patterns

	Secure	Insecure
LS - Learning Set	3432	1526
TS - Testing Set	1476	649

C. RT for the Main Security Problem

After growing a very large tree and applying each of the earlier described pruning approaches, the smaller trees of the *newMEC* method were selected as the most promising ones in order to extract a RT+ANN structure. In fact, this method provided the lowest $MSE_{TS}(T)$ predicting errors in the subset of the less complex trees. Namely, the following candidate RT structures were selected:

- RT with 3 nodes and $MSE_{TS}(T) = 3.557 \text{ Hz}^2$
- RT with 5 nodes and $MSE_{TS}(T) = 2.387 \text{ Hz}^2$

To give an example of a RT, the tree structure of the obtained RT with 3 nodes is presented in Figure 2.

The most accurate RT structures, having $MSE_{TS}(T) = 0.411$ Hz^2, were provided by the *LSS* and *MCV* methods, however with a very complex tree structure, namely with 758 and 1121 nodes. The *MEC* and *MEL* methods provided the less accurate RT for all the range of tree size. The *newMEL* method, provided the same size-tree/accuracy curve as the one provided by *newMEC*, however with a much higher number of trees, which highlights the advantage, already presented in [7], of pruning a tree trough a compromise between accuracy and complexity. The obtained size-tree/accuracy curves for each of the applied pruning processes are too complex to be presented here.
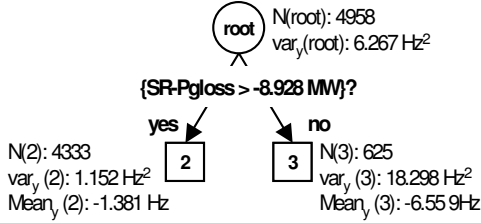
Figure 2 – Obtained RT with 3 nodes, with LS information

D. ANN for the Main Security Problem

For all the trained ANN, a structure defined by 34 inputs (the security features), two hidden layers with 16 and 10 units and one output (the security index) was considered. After applying the training procedure described in Section B, an ANN with the estimated accuracy presented in Table II was obtained for the main security problem.

Table II – TS predicting errors for ANNroot

MSE (Hz^2)	0.037
MAE (Hz)	0.088
Global Classification Error (%)	1.60
False Alarm Error (%)	1.49
Missed Alarm Error (%)	1.85

In this table, besides Mean Squared Error, the following predicting errors were considered:

MAE : Mean Absolute Error, given by

$$MAE_{TS} = \frac{1}{N(TS)} \sum_{OPi \in TS} |y_i - \widehat{y}_i| \tag{11}$$

Global Class. Error, given by

$$\frac{N^\circ \{OP\ of\ the\ TS\ incorrectly\ class.\}}{N^\circ \{OP\ of\ the\ TS\}} \times 100\% \tag{12}$$

False Alarm Error, given by

$$\frac{N^\circ \{"secure"\ OP\ of\ TS\ class.\ as\ "insecure"\}}{N^\circ \{"secure"\ OP\ of\ the\ TS\}} \times 100\% \tag{13}$$

Missed Alarm Error, given by

$$\frac{N^\circ \{"insecure"\ OP\ of\ TS\ class.\ as\ "secure"\}}{N^\circ \{"insecure"\ OP\ of\ the\ TS\}} \times 100\% \tag{14}$$

E. RT+ANN Obtained Results

After applying the training procedure described in Section C for each of the two selected candidate RT structures (with 3 and 5 nodes), two different RT+ANN structures were obtained where an accuracy improvement was observed for all the considered predicting errors. These results are presented in Figure 3 and Figure 4, where a comparison between the accuracy provided by the ANN and RT+ANN approaches is performed. We may still observe that the larger RT provides the lowest predicting errors, which leads to the assumption that RT+ANN accuracy may still benefit from considering some other larger trees.
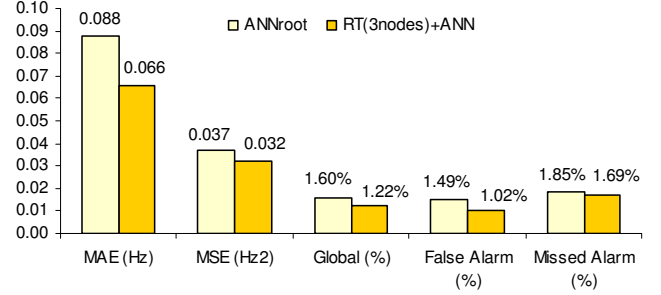

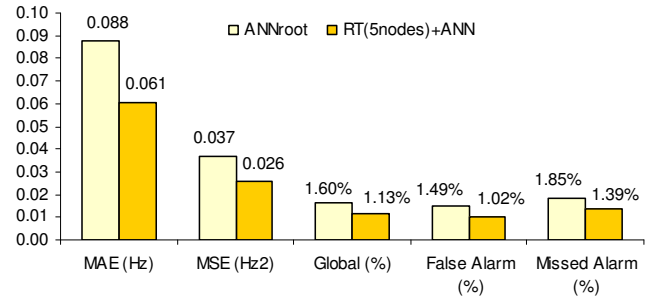Figure 3 – TS predicting errors for ANNroot and RT(3nodes)+ANN

Figure 4 – TS predicting errors for ANNroot and RT(5nodes)+ANN

F. Exploiting the RT Structure for Feature Selection

The automatic learning structures presented in last sections were extracted by considering all the 34 initial features. Namely, Figure 5 presents the set of trained ANN for the main security problem, where the one with the highest testing set accuracy ($MSE_{TS}(ANN)$ of 0.037 Hz^2) was selected.

To analyze the capability of the two earlier described techniques that performed feature selection by exploiting the RT structure, these were applied before starting a new ANN training for the main security problem. The exploited tree structure was the one resulted from the RT that minimizes the testing set predicting error, with 1121 nodes and $MSE_{TS}(T) = 0.411$ Hz^2.

After applying the *FSS1* method, 4 features were eliminated since they don't provide any information gain (i.e. have not been selected in any splitting test). After applying the *FSS2* method, 13 features were eliminated since they provided an information gain lower than 0.3 (normalized value).

After performing the ANN training with these two new set of features (30 and 21), the set of trained ANN presented in Figure 6 and Figure 7 were obtained. These figures present the testing set accuracy of each trained ANN and the minimum, maximum and mean value of the obtained $MSE_{TS}(ANN)$. A

solid line was draw to highlight the minimum error obtained with no feature selection procedure. By comparing these results with the ones presented in Figure 5, we may clearly recognize that the two applied methods improve ANN accuracy, where a higher reduction is observed with *FSS1* method. However, in applications where features reduction is of higher importance, *FSS2* method should be more advantageous, because much more features are eliminated providing almost the same accuracy.
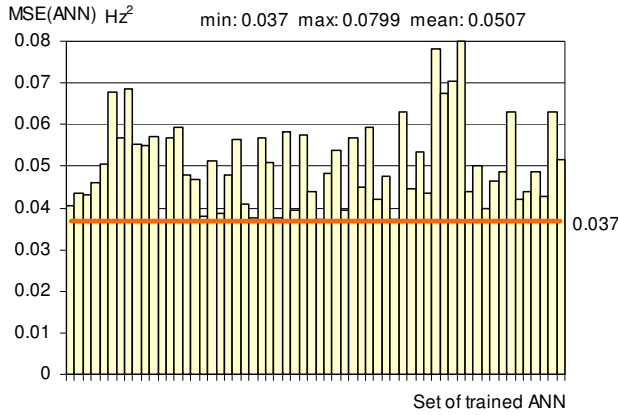

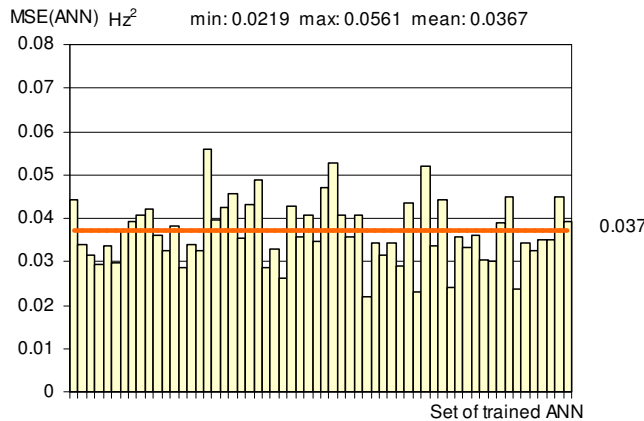Figure 5 – ANNroot training results without feature selection

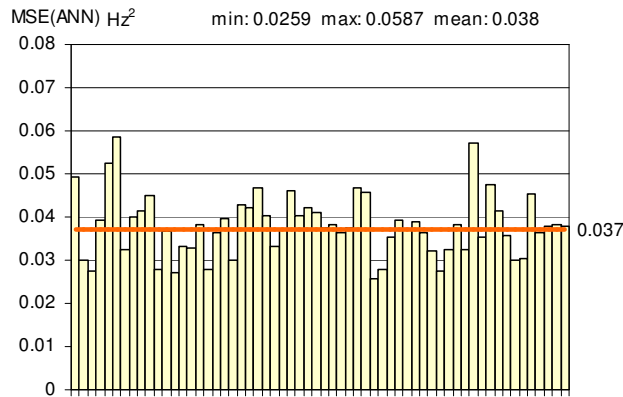Figure 6 – ANNroot training results after FSS1

Figure 7 – ANNroot training results after FSS2

IV. CONCLUSIONS

In this paper, a new promising automatic learning approach, to perform on-line dynamic security assessment of electrical power systems, was presented. With the obtained hybrid security structure, the advantages of artificial neural networks and regression trees are exploited in order to obtained a more accurate model without compromising prediction time. The quality of the approach was illustrated through its application to a major security problem of the power system of Madeira Island.

The contributions of the proposed method is however not restricted to accuracy improvement. In future work, the obtained tree structure with ANN in the tree leafs will be exploit in order to provide fast preventive control. This will be performed by combining the interpretable "if-then-else" tree structure, with the capability of ANN to provide new inputs solutions in order to reach some output threshold, trough the application of gradient based methods.

V. ACKNOWLEDGMENTS

VI. REFERENCES

[1] N. Hatziargyriou et. al., "Preliminary Results From The More Advanced Control Advice Project For Secure Operation Of Isolated Power Systems With Increased Renewable Energy Penetration And Storage – More Care", in Proc. IEEE Power Tech, Porto – Portugal, September 2001.
[2] J. A. Peças Lopes et. al., "Dynamic Security Evaluation Functions in the MoreCare Project", in Proc. MedPower, Athens – Greece, November 2002.
[3] Louis A. Wehenkel, "Automatic Learning Techniques in Power Systems", Kluwer Academic Publishers, ISBN 0-7923-8068-1, 1998.
[4] Simon Haykin, Neural Networks: A Comprehensive Foundation, 2nd ed., Prentice-Hall, 1999.
[5] F. M. Silva, L. B. Almeida, "Acceleration Techniques For The Backpropagation Algorithm", In Neural Networks, L. B. Almeida and C. J. Wellekens (Eds.), Springer-Verlag, 1990.
[6] L. Torgo, "Kernel Regression Trees", Presented at European Conference on Machine Learning (ECML-97), Prague, 1997.
[7] L. Breiman, H. F. Friedman, R. A. Olshen, C. J. Stone, "Classification and Regression Trees", Wadsworth International, 1984.
[8] H. Vasconcelos, J. Fidalgo, J. Peças Lopes, "A General Approach for Security Monitoring and Preventive Control of Networks with Large Wind Power Production", in Proc. 14th PSCC, Sevilla, June 2002.
[9] J. N. Fidalgo, "Operação de Rede Eléctricas Isoladas com Produção Eólica – contributos das redes neuronais", Ph.D. dissertation, Faculdade de Engenharia da Universidade do Porto, Portugal, 1995.
[10] L. Torgo, "Inductive Learning of Tree-Based Regression Models", Ph.D. dissertation, Faculdade de Ciências da Universidade do Porto, Portugal, September 1999.
[11] Bohanec, Bratko, "Trading Accuracy for Simplicity in Decision Trees", Machine Learning, 15-3, 223-250, Kluwer Academic Publishers.
[12] L. Wehenkel et. al., "Automatic Learning Approaches for On-line Transient Stability Preventive Control of the Hydro-Quebec System – Part I. Decision tree approaches", SIPOWER'95, Mexico, December 1995.

VII. BIOGRAPHIES

Maria Helena Vasconcelos was born in Porto, Portugal, on March 19, 1973. She graduated in Electrical Engineer from the Engineering Faculty of Porto University (FEUP) concluded in July 1996, and obtained the MSc. from FEUP in October 1999. Since September 1996 she works as a researcher in the Power System Unit of INESC Porto. Currently, she is a Ph.D. student and Teaching Assistant at FEUP.

João A. Peças Lopes was born in Portugal in May 1958. He graduated in Electrical Engineer from the Engineering Faculty of Porto University (FEUP) in July 1981, and obtained the Ph.D. and Aggregation degrees also from FEUP in October 1988 and November 1996 respectively. Dr. Peças Lopes is presently Associated Professor with Aggregation at FEUP and Assistant Coordinator of the Power Systems unit at INESC Porto.