Support Vector and Multilayer Perceptron Neural Networks Applied to Power Systems Transient Stability Analysis with Input Dimensionality Reduction

L. S. Moulin, *Student Member, IEEE*, A. P. Alves da Silva, *Senior Member, IEEE*,
M. A. El-Sharkawi, *Fellow, IEEE*, and R. J. Marks II, *Fellow, IEEE*

***Abstract*--The Neural Network (NN) approach to the Transient Stability Analysis (TSA) has been presented as a potential tool for on-line applications, but the high dimensionality of the power systems turns it necessary to implement feature extraction techniques to make the application feasible in practice. At the same time, feature extraction can offer sensitivity information to help the identification of input features best suited for control action. This paper presents a new learning-based nonlinear classifier, the Support Vector Machines (SVMs) NNs, showing its suitability for power system TSA. It can be seen as a different approach to cope with the problem of high dimensionality due to its fast training capability, which can be combined with existing feature extraction techniques. SVMs' theoretical motivation is conceptually explained and they are applied to the IEEE 50 Generator system TSA problem. Aspects of model adequacy, training time and classification accuracy are discussed and compared to stability classifications obtained by Multi-Layer Perceptrons (MLPs). Both models are trained with complete and reduced input features sets.**

***Index Terms*--Feature Extraction, Support Vector Machines, Multilayer Perceptrons, Neural Networks, Transient Stability Analysis.**

I. Introduction

THE Transient Stability Analysis (TSA) is a crucial operation procedure to ensure secure performance of a power system experiencing a variety of disturbances and operating condition changes. The power system operates in a secure manner, from the transient stability viewpoint, when the generators maintain synchronism after the system is subjected to severe disturbances.

In the last few decades, TSA methods of practical use have been developed, and the transient stability schemes of current use are mainly based on time-domain simulations [1]. These techniques, however, require the numerical solution of a system of nonlinear equations using time-consuming numerical integrations for each contingency.

With the power systems expansion and the increase in their complexity, the dynamic security analysis has become a very crucial and complex process. The current deregulation trend and the participation of many players in the power market are contributing to the decrease in the security margin [2]. This makes the security evaluation even more important, and demands the investigation of fast and accurate techniques to allow on-line TSA.

The NN approach has been introduced as an alternative solution for the analytical TSA [3], [4], and has been recently studied with potential use for real-world, large-scale power systems [4]-[6]. In such a process, the NN-based TSA would be applied to a group of selected critical contingencies. The nonlinear input/output mapping capability of a NN can be used to produce a security index that classifies the current operating point as secure or insecure [3], [5]-[7].

The NN uses training data sets that are representatives of different loading conditions and generation schedulings, different types of contingencies and different topology configurations.

Although successfully applied to TSA [3], [5], [8], Multi-Layer Perceptrons (MLPs) implementations require extensive training process. In general, this is the major drawback for NN applications in large power systems with hundreds (even thousands) of generators, because such a large grid will require a large number of input variables to train a NN. This can be a prohibitive task. Therefore, a feature extraction/selection method is needed to reduce the dimensionality of the NN's input space. The main objective is to use as little number of inputs as possible to reduce the NN training time, while maintaining a high degree of classification accuracy [9].

A new type of nonlinear learning based classifier has been recently introduced which has very interesting theoretical promises, the Support Vector Machines (SVMs) NNs [10]. They can map complex nonlinear input/output relationships with good accuracy and they seem to be very well suited for the TSA application [11]. SVM classifiers rely on training points located on the boundary of separation between different

This work was supported by the Brazilian Research Council (CNPq).

L. S. Moulin and A. P. Alves da Silva are with the Federal Engineering School at Itajuba/Institute of Electrical Engineering, Av. BPS, 1303 – Itajuba, MG 37500-000 Brazil.

M. A. El-Sharkawi and R. J. Marks II are with the University of Washington/Department of Electrical Engineering, 253 EE/CSE Building – Campus Box 352500 – Seattle, WA 98195-2500 USA.

classes, where the stability evaluation is critical. A good theoretical development of the SVM NN, due to its foundations on the Statistical Learning Theory (SLT) [10] made it possible to devise fast training techniques even with large training sets and high input dimensions [12]-[14]. This feature can be exploited as an approach to address the problem of high input dimension and large training datasets in the TSA problem.

However, the SVMs capabilities cannot be explored without a good understanding of their conceptual details. In this paper, the SVM classifier is explained and analyzed in terms of advantages and disadvantages. The aim is the application to power system TSA, which is developed as follows: three feature extraction techniques are applied to the training/test transient stability data with the objective of dimensionality reduction, as presented in [9]. Stability classifications are obtained by MLP and SVM NNs, comparing the good generalization capacity of both models and exploring SVMs' fast training. Expectations are that input feature dimensionality reduction is of lower concern for SVMs due to their fast training, but accuracy must be checked for complete and reduced features sets. Results for the IEEE 50-Generator system are presented, discussed and compared in terms of modeling characteristics, generalization performance and training time.

The structure of the paper is as follows. Section 2 briefly presents the feature extraction/selection techniques as applied in the TSA. In Section 3, a summarized description of SVM classifiers is sketched with the conceptual ideas and discussions on advantages and disadvantages. In Section 4, the TSA application and results are presented. Finally, the conclusions are drawn in Section 5.

II. Feature Extraction/Selection

The feature extraction problem can be explained by assuming the classification task in 2 disjoint classes with a training set T of ordered pairs (x_i, y_i), $T = \{x_i, y_i\}_{i=1}^{N}$, where x_i is a real-valued n dimensional vector (i.e., $x_i \in R^n$) representing the operating point and $y_i \in \{+1,-1\}$ is a label that represents the security index. The feature extraction goal is to determine a transformation $f = F(A, x)$ from the original space R^n to a subspace R^d (for dimensionality reduction, $d < n$), where A is a matrix of transformation parameters and $f \in R^d$. The original data is represented in a new training set, $T = \{f_i, y_i\}_{i=1}^{N}$. If the feature extraction/selection is successful, a point in R^d can be assigned to one of the 2 classes with a minimum error. Hence, the expected number of misclassifications for a test set should be as low as possible.

Three feature extraction techniques are used in this work, as presented in [9], based in Sequential Search, Genetic Algorithms and Principal Components Analysis to perform dimensionality reduction of the input vector.

III. Support Vector Machines Classification

SVMs are nonlinear models based on theoretical results from the Statistical Learning Theory [10]. This theory formally generalizes the empirical risk minimization principle that is usually applied for NN training when the classifier is determined by minimizing the number of training errors. In NN training, a number of heuristics is traditionally used in order to avoid overfitting and to estimate a NN classifier with adequate complexity for the problem at hand.

An SVM classifier minimizes the generalization error by optimizing the relation between the number of training errors and the so-called Vapnik-Chervonenkis (VC) dimension. This is a new concept of complexity measure that can be used for different types of functions.

A formal theoretical bound exists for the generalization ability of an SVM, which depends on the number of training errors (t), the size of the training set (l), the VC dimension associated to the resulting classifier (h), and a chosen confidence measure for the bound itself (η) [15]:

$$R < \frac{t}{l} + \sqrt{\frac{h(ln(2l/h)+1) - ln(\eta/4)}{l}} \tag{1}$$

The risk R represents the classification error expectation over all the population of input/output pairs, even though the population is only partially known. This Risk is a measure of the actual generalization error and does not require prior knowledge of the probability distribution of the data. Statistical Learning Theory derives inequality (1) to mean that the generalization ability of an SVM is measured by an upper limit of the actual error given by the right hand side of (1), and this upper limit is valid with probability $1-\eta$ $(0 < \eta < 1)$. As h increases, the first summand of the upper bound (1) decreases and the second summand increases, so that there is a balanced compromise between the two terms (complexity and training error).

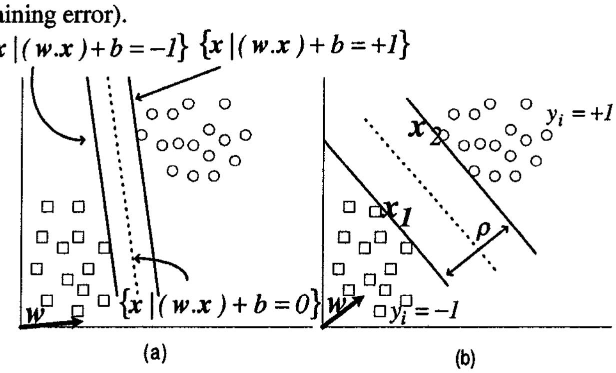

Fig. 1. Maximum margin classifier

The SVMs used for two-class problems are based on linear hyperplanes to separate the data, as shown in Fig. 1. The hyperplane is determined by an orthogonal vector w and a bias b, which identify the points that satisfy $w.x + b = 0$. By finding a hyperplane that maximizes the margin of separation, ρ, it is intuitively expected that the classifier will have a better generalization ability (Fig. 1). The hyperplane with the largest margin on the training set can be completely determined by points that are closest to the hyperplane. Two of such points are x_1 and x_2 in Fig. 1(b), and they are called Support Vectors (SVs) because the hyperplane (i.e., the classifier) depends entirely on them.

Therefore, in their simplest form, SVMs learn linear decision rules

$$f(x) = sign(w^t x + b) \qquad (2)$$

so that (w, b) are determined as to correctly classify the training examples and maximize ρ.

For linearly separable data, as shown in Fig. 1, a linear classifier can be found such that the first summand of bound (1) is zero. It is always possible to scale w and b so that

$$w^t x + b = \pm 1 \qquad (3)$$

for the SVs, with

$$w^t x + b > +1 \text{ and } w^t x + b < -1 \qquad (4)$$

for non-SVs. Using the SVs x_1 and x_2 of Fig. 1 and Equation (3), the margin ρ can be calculated as

$$\rho = \frac{w^t}{\|w\|}(x_2 - x_1) = \frac{2}{\|w\|} \qquad (5)$$

For linearly separable data the VC dimension of SVM classifiers can be assessed by [10]

$$h < min\left\{n, \frac{4D^2}{\rho^2}\right\} + 1 = min\left\{n, D^2\|w\|^2\right\} + 1 \qquad (6)$$

where n is the dimension of the training vectors and D is the minimum radius of a ball which contains the training points. Therefore the risk (1) can be decreased by decreasing the complexity of the SVM, that is, by increasing the margin of separation ρ, which is equivalent to decreasing $\|w\|$.

As practical problems are not likely to be separable by a linear classifier, the linear SVM can be extended to a nonlinear version by mapping the training data to an expanded feature space with a nonlinear transformation:

$$\Phi(x) = (\phi_1(x), \ldots, \phi_m(x)) \in R^m \qquad (7)$$

where $m > n$. Then, the maximum margin classifier of the data on the new space can be determined. With this procedure, the data that are non-separable in the original space may become separable in the expanded feature space. The next step is to estimate the SVM by minimizing

$$V(w) = \tfrac{1}{2} w^t w \qquad (8.1)$$

subject to the condition that all training patterns are correctly classified, that is,

$$y_i\left(w^t \Phi(x_i) + b\right) \geq 1,\ i = 1, \ldots, l \qquad (8.2).$$

However, depending on the type of nonlinear mapping (7), the training points may happen to be not linearly separable, even in the feature space. That means, it will be impossible to find an SVM classifier that fulfills all the conditions (8.2). Therefore, instead of solving (8), a new cost function is used to minimize (1) [10]:

$$V(w, \varepsilon) = \tfrac{1}{2} w^t w + C \sum_{i=1}^{l} \varepsilon_i \qquad (9)$$

where l *slack* variables ε_i are introduced to allow for training errors, that is, training patterns for which $y_i\left(w^t \Phi(x_i) + b\right) \geq 1 - \varepsilon_i$ and $\varepsilon_i > 1$. By minimizing the first summand of (9), the complexity of the SVM is decreased and by minimizing the second summand of (9), the number of training errors is decreased. C is a positive penalty constant that must be chosen to act as a tradeoff between the two terms.

The minimization of the cost function (9) leads to the SVM training as a quadratic optimization problem with unique solution. In practice, the nonlinear mapping (7) is indirectly obtained by the so called Mercer Kernel Functions, which correspond to inner products of data vectors in the feature space, $K(a, b) = \Phi(a)^t \Phi(b)$, $a, b \in R^n$ [14]. In order for this equivalence to be valid, a Kernel function must satisfy some requirements called Mercer Conditions. These conditions have limited the number of Kernel Functions applied in practice so far, and the most commonly used are the Gaussian RBF Kernel

$$K(a, b) = e^{-\frac{\|a-b\|^2}{\sigma^2}} \qquad (10)$$

and the Polynomial Kernel

$$K(a, b) = \left(a^t b + 1\right)^p \qquad (11)$$

where the parameters σ and p in (10) and (11) must be pre-set. Details on the solution of (7) and the final SVM architecture are shown in the Appendix.

In summary, some nonlinear mapping (7) can be indirectly defined by a Kernel Function (i.e., there is no need for specifying (7)), for example (10) or (11). The parameters σ and p affect how sparse and easily separable the data are in feature space, and consequently, affect the complexity of the resulting SVM classifier and the number of training errors. The parameter C also affects the model complexity. Currently, there are no clues on how to set C, how to choose the best Kernel Function (the nonlinear map Φ) and how to set the Kernel parameters. In practice, a range of values has to be tried for C and for the Kernel parameters, and then the performance of the SVM classifier is assessed on each of these values.

IV. Transient Stability Analysis Tests and Results

This section explains how the feature extraction techniques, connected with MLP and SVM NNs' training, are actually applied to obtain power system transient stability evaluations.

The IEEE 50-Generator system has been used [16] to generate training and test examples. Different operating conditions have been created by changing the generation and load patterns of the system. Each case has been validated by a load flow execution. For each operating condition, the same contingency has been simulated in the time domain using the ETMSP software [17] and the corresponding critical clearing time (CCT) has been determined. The complete input features set is composed of the active and reactive powers of each generator and the total active and reactive loads of the system at the moment of the fault, with a total of 102 inputs and 1 output indicating the security class.

The feature extraction techniques presented in [9] have been run on the training set to reduce the input space dimension. Besides the complete set of features, reduced sets have been obtained with $d = 50, 30, 20$ and 10.

Multi-Layer Perceptrons have been trained with the Levenberg-Marquardt backpropagation training algorithm to give security estimations based on binary outputs corresponding to the stale/unstable classes. The classification of the system as stable/unstable is determined based on a given security threshold, which represents the realized clearing time of the contingency. If a given sample output i.e., the simulated CCT, is above the threshold, the input state is considered stable, otherwise it is unstable.

SVMs have also been trained on the examples with binary outputs to indicate the stable and unstable classes. Gaussian RBF (10) and Polynomial (11) Kernels have been used. The parameters for theses Kernel functions have been sought in a heuristic manner. The software SVM^{light} has been used for training[13].

Fig. 2 presents Receiver Operating Characteristic (ROC) curves of the Gaussian RBF SVM performance on the test set, after it has been trained with the complete set of 102 input features. The false dismissal rate on the x-axis is the ratio of test points that have been incorrectly classified as stable. The detection rate on the y-axis is the ratio of stable test points that have been correctly classified. Several SVMs have been trained with increasing values of σ^2 and the corresponding values of the detection and false dismissal rates are shown in Fig. 2. These are two conflicting values that increase together for a fixed value of C and increasing σ^2. Values of $C = 0.1, 1, 10, 100, 1000$ and 10000 have been tried. The solid line in Fig. 2 corresponds to the SVM with $C = 1$. The dashed line corresponds to the SVM with $C = 10, 100, 1000$ and 10000, which show the same results. The curve for the SVM with $C = 0.1$ is not shown in Fig. 2 due to the difference in scales. It presents much higher values of false dismissal rates, going as far as 0.08 and lower values of detection rate than the curves shown in Fig. 2. ROC curves like these can also be drawn when a Polynomial SVM is used, but now the False Dismissal Rate and the Detection Rate change with p, the Polynomial Kernel parameter.

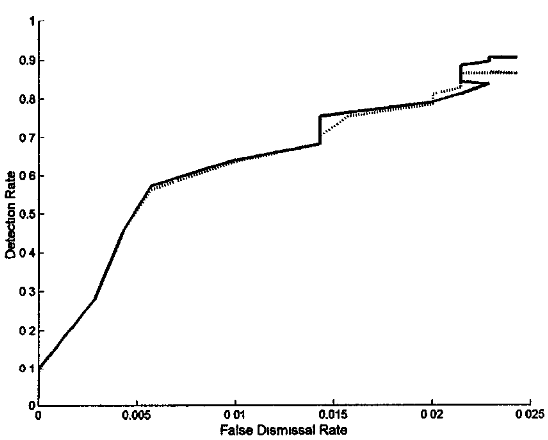

Fig. 2. SVM ROC curve

ROC curves can help to identify a classifier with good performance, which is required to present low False Dismissal Rates and high Detection Rates.

An MLP has also been trained on the complete set of 102 features, but an ROC curve like the one of Fig. 2 cannot be drawn to show the performance on the test set. The number of factors affecting its performance is large and interrelated.

On Table I the performances of the classifiers trained on the complete features sets is indicated with the headings SVM-G102 (RBF Gaussian SVM), SVM-P102 (Polynomial SVM) and MLP-102 (MLP with 5 hidden neurons). Table I shows specifications of the input features, the detection rate, false dismissal rate, error rate (ratio of test points which have been incorrectly classified), training time and number of SVs.

Next, SVM and MLP classifiers have been trained on reduced features sets. For the MLP, a cross-validation training has been performed until the test error started increasing or stopped decreasing.

For SVMs trained with different features sets, ROC curves have been drawn for different values of C. Then, for the values of C that resulted in the best ROC curves, fixed values of False Dismissal Rates have been set: 0.02 and 0.03, which could be reached with specific values of the Kernel parameters. These models are shown in Table II as SVM-G0.02, SVM-G0.03 (RBF Gaussian Kernel), and SVM-P0.02 and SVM-P0.02 (Polynomial Kernel). For the MLP model, MLP-1 is the model with the lowest test error rate and MLP-2 is the model with the smallest training time.

Table II shows the models trained with reduced input features that resulted in the best test performances, in terms of Detection Rate, FalseDismissal Rate and Error Rate. For example, for an RBF Gaussian SVM with parameters set to give 0.03 False Dismissal Rates (SVM-G0.03) the 30 input features selected by the Sequential Search (SS) technique resulted in the best performance; for a Polynomial SVM with parameters set to give 0.03 False Dismissal Rates (SVM-P0.03) the 30 input features selected by the Genetic Algorithm (GA) technique resulted in the best performance.

Tables I and II show that SVM classifiers are a viable model for the TSA application, with performance results that are comparable to the MLP and much faster training times. In these TSA tests, the RBF and Polynomial Kernel showed similar results on both criteria of classifier performance and training time. Table II shows that it is possible to achieve a good reduction in the input features dimensionality, with performance results that are comparable to the complete features set and much lower training times for the MLP model.

It could be noticed that the MLP training time dramatically increased with the number of input features and the number of hidden neurons. The SVM training time depends on the number of input features, on the Kernel parameters values and on the number of SVs of the resulting classifier.

Tables I and II show that the adequacy of feature extraction techniques depends on the classifier, as expected.

V. Conclusions

This paper shows that the SVMs are a new NN model that fits the TSA application. It provides a different strategy to tackle the curse of dimensionality, regarding computational effort, because of very low training times compared to MLPs.

In this application, the reduction on the number of features from *102* to *30* and *50* resulted in classifiers with the best performances. The accompanying reduction in the sparsity of the data has turned the training process into an easier task for SVMs and MLPs. On the other hand, larger training sets could be used for SVMs to improve the performance, while the training time would not be considerably increased.

The SVM model allows a good understanding of its theoretical details, as shown in Section 3, and this can be used to identify the important parameters for the classifier.

The feature extraction techniques shown in this paper are good candidates to be used with artificial intelligence tools in control centers to avoid potentially vulnerable power system states. They can provide not only dimensionality reduction, but also the most important rules to prevent the system from getting closer to unstable situations.

VI. APPENDIX

The computation of the decision boundary of an SVM, $f(x) = sign(w^t\Phi(x)+b)$, for the non-separable case consists in solving the following optimization problem [10]:

$$\begin{aligned} &minimize: V(w,\varepsilon) = \tfrac{1}{2}w^t w + C\sum_{i=1}^{l}\varepsilon_i \\ &subject\ to: y_i\left(w^t\Phi(x_i)+b\right) \geq 1-\varepsilon_i, \quad i=1,...,l \\ &\qquad \varepsilon_i > 0, \quad i=1,..,l \end{aligned} \tag{12}$$

Instead of solving (12) directly, it is much easier to solve the dual problem (13), in terms of the Lagrange multipliers, α_i [10]:

$$\begin{aligned} minimize: W(\alpha) &= -\sum_{i=1}^{l}\alpha_i + \frac{1}{2}\sum_{i=1}^{l}\sum_{j=1}^{l} y_i y_j \alpha_i \alpha_j \Phi(x_i)^t \Phi(x_j) = \\ &= -\sum_{i=1}^{l}\alpha_i + \frac{1}{2}\sum_{i=1}^{l}\sum_{j=1}^{l} y_i y_j \alpha_i \alpha_j K(x_i, x_j) \\ subject\ to: &\sum_{i=1}^{l} y_i \alpha_i = 0 \ and \ 0 \leq \alpha_i \leq C, \quad i=1,...,l \end{aligned} \tag{13}$$

which is a quadratic optimization problem. From the solution, $\alpha_i, \ i=1,...,l$ of (13) the decision rule $f(x)$ can be computed as [10]

$$\begin{aligned} f(x) = w^t\Phi(x)+b &= \sum_{i=1}^{l}\alpha_i y_i \Phi(x_i)^t \Phi(x)+b = \\ &= \sum_{i=1}^{l}\alpha_i y_i K(x_i, x)+b \end{aligned} \tag{14}$$

The training points with $\alpha_i > 0$ are the SVs, and (14) depends entirely on them. The threshold b can be calculated using (3), which is valid for any SV:

$$b = y_{SV} - w^t\Phi(x_{SV}) \tag{15}$$

An SVM can be represented as in Fig. 3, where the number of units $K(x, x_i)$ is determined by the number of SVs.

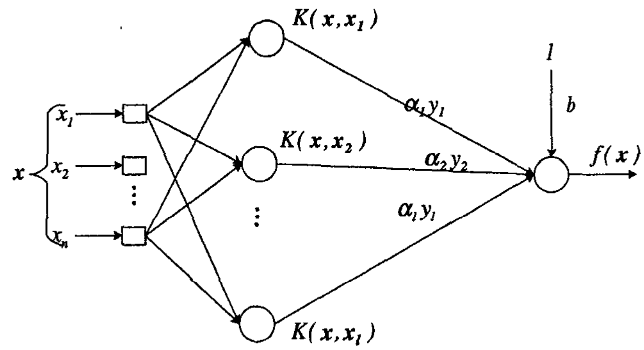

Fig. 3. SVM NN architecture

VII. REFERENCES

[1] P. Kundur, *Power System Stability and Control*, McGraw-Hill, 1994.

[2] Y. Mansour, "Competition and System Stability – The Reward and the Penalty", *Proceedings of the IEEE*, Vol. 882, pp. 228-234, Feb. 2000.

[3] D. J. Sobajic and Y.-H. Pao, "Artificial neural-net based dynamic security assessment for electric power systems", *IEEE Trans. On Power Systems*, Vol. 4, No. 1, pp. 220-228, Feb. 1989.

[4] A. B. R. Kumar, A. Ipakchi, V. Brandwajn, M. El-Sharkawi, and G. Cauley, "Neural networks for dynamic security assessment of large-scale power systems: requirements overview", in *Proceedings of the First International Forum on Applications of Neural Networks to Power Systems*, pp. 65-71, 1991.

[5] Y. Mansour, A. Y. Chang, J. Tamby, E. Vaahedi, and M. A. El-Sharkawi, "Large scale dynamic security screening and ranking using neural networks", *IEEE Trans. On Power Systems*, Vol. 12, No. 2, pp. 954-960, May 1997.

[6] Y. Mansour, E. Vaahedi, and M. A. El-Sharkawi, "Dynamic security contingency screening and ranking using neural networks", *IEEE Transactions on Neural Networks*, Vol. 8, No. 4, pp. 942-950, Jul. 1997.

[7] Y. M. Park, G.-W. Kim, H.-S. Choi, and K. Y. Lee, "A new algorithm for kohonen layer learning with application to power system stability analysis", *IEEE Transactions on Systems, Man, and Cybernetics*, Vol. 27, No. 6, pp. 1030-1033, Dec. 1997.

[8] K. Demaree, T. Athay, K. W. Cheung, Y. Mansour, E. Vaahedi, A. Y. Chang, B. R. Corns, and B. W. Garrett, "An on-line dynamic security analysis system implementation", *IEEE Transactions on Power Systems*, Vol. 9, No. 4, pp. 1716-1722, Nov. 1994.

[9] L. S. Moulin, M. A. El-Sharkawi, R. J. Marks II, and A. P. Alves da Silva, "Automatic feature extraction for neural network based power systems dynamic security evaluation", in *Proc. Of the IEEE ISAP2001 International Conference on Intelligent Systems Applied to Power Systems*, Budapest, Hungary, Paper No. 21.

[10] V. Vapnik, *Statistical Learning Theory*, John Wiley & Sons, New York, 1998.

[11] A. E. Gavoyiannis, D. G. Vogiatzis, and N. D. Hatziargyriou, "Dynamic security classification using support vector machines", in *Proc. of the IEEE ISAP2001 International Conference on Intelligent Systems Applied to Power Systems*, Hungary, Budapest, Paper No. 22.

[12] E. Osuna, R. Freund, and F. Girosi "Training support vector machines: an application to face detection", in *Proceedings of CVPR'97*, Puerto Rico.

[13] T. Joachims, *"Making large-scale SVM learning practical"*, in *Advances in Kernel Methods – Support Vector Learning, Scholkopf, B, Burges, J C C and Smola, A J (editors)*, MIT Press, Cambridge, USA, 1998.

[14] N. Cristianini and J. Shawe-Taylor, *An Introduction to Support Vector Machines and Other Kernel-Based Learning Methods*, Cambridge University Press, 2000.

[15] V. Vapnik, *The Nature of Statistical Learning Theory*, Springer-Verlag, New York, 1995.

[16] V. Vittal, (Chairman) "Transient stability test systems for direct stability methods", *IEEE Transactions on Power Systems*, Vol. 7, No. 1, pp. 37-43, 1992.

[17] ETMSP – *"Extended Transient Midterm Stability Program User's Manual"*, Vol. 1-6, Version 3.1, EPRI TR-102004-V2R1, Final Report, 1994.

VIII. BIOGRAPHIES

Luciano S. Moulin was born in Nanuque, Brazil, in 1972. He received the B.Sc. and M.Sc. degrees in Electrical Engineering (EE) from the Federal Engineering School at Itajuba (EFEI), Brazil, in 1995, 1998, respectively. During 2000, he was a visiting student at the Department of Electrical Engineering, University of Washington. He is currently working towards the D.Sc. degree in EE at EFEI.

Alexandre P. Alves da Silva was born in Rio de Janeiro, Brazil, in 1962. He received the B.Sc, M.Sc. and Ph.D. degrees in Electrical Engineering from the Catholic University of Rio de Janeiro, in 1984 and 1987, and the University of Waterloo, Canada, in 1992, respectively. He worked with the Electric Energy Research Center (CEPEL) from 1987 to 1988. During 1999, he was a Visiting Professor in the Department of Electrical Engineering, University of Washington, USA. Since 1993, he has been with the Federal Engineering School at Itajuba, where he is a Professor in Electrical Engineering. He and his group have developed intelligent forecasting and security assessment systems that are in operation at the control centers of Brazilian electric utilities. Professor Alves da Silva was the Technical Program Committee Chairman of the First Brazilian Conference on Neural Networks, in 1994, and of the International Conference on Intelligent System Applications to Power Systems, in 1999. He is the Chairman of the IEEE Neural Network Council Brazilian Regional Interest Group. He has authored and co-authored about 150 papers on intelligent systems application to power systems.

Mohamed A. El-Sharkawi is a Fellow of the IEEE. He received his B.Sc. in Electrical Engineering in 1971 from Cairo High Institute of Technology, Egypt. His M.Sc. and Ph.D. in British Columbia in 1977 and 1980 respectively. In 1980 he joined the University of Washington as a faculty member where he is currently a Professor of Electrical Engineering. He is the founder of the International Conference on the Application of Neural Networks to Power Systems (ANNPS) which was later merged with the *Expert Systems Conference* and renamed *Intelligent Systems Applications to Power (ISAP)*. He is the co-editor of the IEEE tutorial book on the applications of neural networks to power systems. He has published over 90 papers and book chapters. He hods five patents: three on *Adaptive Var Controller* for distribution systems and two on *Adaptive Sequential Controller* for circuit breakers.

Robert J. Marks, II (Fellow, IEEE) is a Professor and Graduate Program Coordinator in the Department of Electrical Engineering, College of Engineering, University of Washington, Seattle. He is the author of numerous papers and is co-author of the book *Neural Smithing: Supervised Learning in Feedforward Artificial Neural Networks* (MIT Press, 1999). He currently serves as the faculty advisor to the University of Washington's chapter of Campus Crusade for Christ. His hobbies include cartooning, song writing, *Gunsmoke*, and creating things for his website. Dr. Marks is a Fellow of the Optical Society of America. He served as the firs President of the IEEE Neural Networks Council. In 1992 he was given the honorary title of Charter President. He served as the Editor-in-Chief of the IEEE TRANSACTIONS ON NEURAL NETWORKS and as a Topical Editor for Optical Signal Processing and Image Science for the *Journal of the Optical Society on America*. For more information see: cialab.ee.washington.edu/Marks.html.

TABLE I
IEEE 50-GENERATORS SYSTEM TRANSIENT STABILITY CLASSIFICATIONS – COMPLETE FEATURES SET

	SVM-G102	SVM-P102	MLP-102
Input Features	102	102	102
Detection Rate	0.92	0.92	0.91
False Dismissal Rate	0.026	0.026	0.021
Error Rate	0.047	0.047	0.047
Training Time	37 seconds	32 seconds	40min42sec
Nr. Of SVs	433	241	-

TABLE II
IEEE 50-GENERATORS SYSTEM TRANSIENT STABILITY CLASSIFICATIONS – REDUCED FEATURES SET

	SVM-G0.02	SVM-G0.03	SVM-P0.02	SVM-P0.03	MLP-1	MLP-2
Input Features	30-SS	30-SS	30-SS	30-GA	50-SS	10-GA
Detection Rate	0.85	0.95	0.96	0.95	0.95	0.92
False Dismissal Rate	0.02	0.03	0.0215	0.0314	0.023	0.024
Error Rate	0.06	0.043	0.033	0.046	0.036	0.046
Training Time	34 seconds	12 seconds	19 seconds	20 seconds	11min24sec	1min42sec
Nr. Of SVs	1065	298	156	178	-	-