International Journal of Interdisciplinary Innovative Research &Development (IJIIRD)
ISSN: 2456-236X
Vol. 06 Special Issue 01 | 2021

Design And Development of Generalized Supervised Assembled Neural Network

Miss. Sneha Ramchandra Sawale, Prof. Miss. Manjiri Karande

Student, Department of Computer Science and Engg, Padm. Dr. V. B. Kolate College Of Engg, Malkapur- Dist. Buldana State Maharashtra India.

Assistant Professor, Department of Computer Science and Eng, Padm. Dr. V. B. Kolate College Of Engg, Malkapur- Dist. Buldana State Maharashtra India.

Abstract

An Artificial Neural Network is an information processing paradigm that is inspired by the way biological nervous systems, such as the brain, process information. The key element of this paradigm is the novel structure of the information processing system. It is composed of a large number of highly interconnected processing elements (neurons) working in unison to solve specific problems.

In this paper, concepts like, back-propagation approach, information theory, pattern recognition, max-margin principle, stopping criteria are used. This paper proposes three novel training methods, two of them based on the back-propagation approach and a third one based on information theory for multilayer perceptron (MLP) binary classifiers. Both back-propagation methods are based on the maximal-margin (MM) principle. The first one based on the gradient descent with adaptive learning rate algorithm (GDX) and named maximum-margin GDX (MMGDX). The second approach, named minimization of interclass interference (MICI), has an objective function inspired by the Fisher discriminant analysis. Such an algorithm aims to create an MLP hidden output where the patterns have a desirable statistical distribution. The third approach offers a robust training framework able to take the best of each proposed training method. The main idea is to compose a neural model by using neurons extracted from three other neural networks, each one previously trained by MICI, MMGDX, and Levenberg Marquard (LM), respectively. The resulting neural network was named assembled neural network (ASNN).

Keyword- MMGDX, MICI, ASNN Algorithms

1. INTRODUCTION

The field of neural networks can be thought of as being related to artificial intelligence, machine learning, parallel processing, statistics, and other fields. The attraction of neural networks is that they are best suited to solving the problems that are the most difficult to solve by traditional computational methods. Consider an image processing task such as recognizing an everyday object projected against a background of other objects. This is a task that even a small child's brain can solve in a few tenths of a second. But building a conventional serial machine to perform as well is incredibly complex. However, that same child might NOT be capable of calculating 2+2=4, while the serial machine solves it in a few nanoseconds.

A fundamental difference between the image recognition problem and the additional problem is that the former is best solved in a parallel fashion, while simple mathematics is best done serially. Neurobiologists believe that the brain is similar to a massively parallel analog computer, containing about 10^10 simple processors which each require a few milliseconds to respond to input. With neural network technology, we can use parallel processing methods to solve some real-world problems where it is very difficult to define a conventional algorithm.

An Artificial Neural Network is an information processing paradigm that is inspired by the way biological nervous systems, such as the brain, process information. The key element of this paradigm is the novel structure of the information processing system. It is composed of a large number of highly interconnected processing elements (neurons) working in unison to solve specific problems.

The Neural Networking algorithm is modeled after the brain and how it processes the information. The brain is a very efficient tool. Having about 100,000 times slower response times than computer chips, it beats the computer in complex tasks, such as image and sound recognition, motion control, and so on. It is also about 10,000,000,000 times more efficient than the computer chip in terms of energy consumption per operation. The brain is a multi-layer structure with 10^11 neurons, structure, that works as a parallel computer capable of learning from the "feedback" it receives from the world and changing its design by growing new neural links between neurons or altering activities of existing ones. To make a picture a bit more complete, let's also mention, that a typical neuron is connected to 50-100 of the other neurons.

An artificial neuron is a device with many inputs and one output. The neuron has two modes of operation; the training mode and the using mode. In the training mode, the neuron can be trained to fire (or not), for particular input patterns. In the using mode, when a taught input pattern is detected at the input, its associated output becomes the current output. If the input pattern does not belong in the taught list of input patterns, the firing rule is used to determine whether to fire or not.

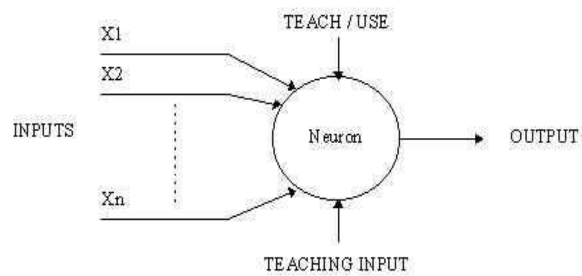

Fig 1.1: A simple neuron

2. CONCEPTS IN NEURAL NETWORK

Back-propagation- is a form of the gradient descent algorithm used with artificial neural networks for minimization and curve-fitting. Back-propagation, an abbreviation for "backward propagation of errors", is a common method of training artificial neural networks. From the desired output, the network learns from many inputs, similar to the way a child learns to identify a dog from examples of dogs. It is a supervised learning method and is a generalization of the delta rule. It requires a dataset of the desired output for many inputs, making up the training set. It is most useful for feed-forward networks. Back-propagation requires that the activation function used by the artificial neurons.

The feed-forward neural network model- If we consider the human brain to be the 'ultimate' neural network, then ideally we would like to build a device that imitates the brain's functions. However, because of limits in our technology, we must settle for a much simpler design. The obvious approach is to design a small electronic device that has a transfer function similar to a biological neuron. This type of electronic model is still rather complex to implement, and we may have difficulty 'teaching' the network to do anything useful. Further constraints are needed to make the design more manageable. First, we change the connectivity between the neurons so that they are in distinct layers, such that each neuron in one layer is connected to every neuron in the next layer. Further, we define that signals flow only in one direction across the network, and we simplify the neuron and synapse design to behave as analog comparators being driven by the other neurons through simple resistors. We now have a feed-forward neural network model that may be practical to build and use.

Referring to figures 2.1 and 2.2, the network functions as follows each neuron receives a signal from the neurons in the previous layer, and each of those signals is multiplied by a separate weight value.

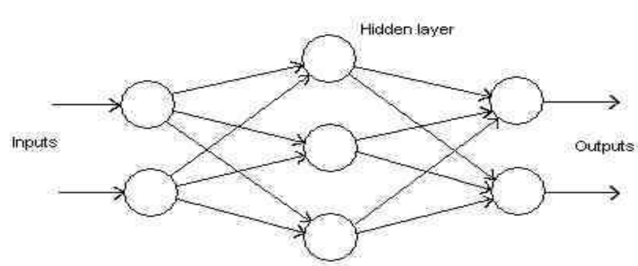

Fig 2.1: A Generalized Network

Stimulation is applied to the inputs of the first layer, and signals propagate through the middle (hidden) layer to the output layer. Each link between neurons has a unique weighting value.

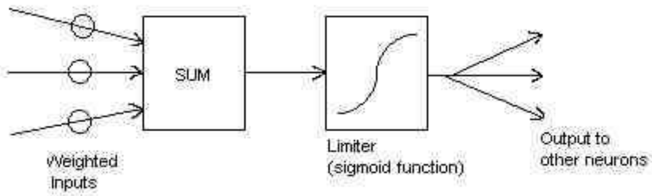

Fig 2.2: The Structure of a Neuron

Information Theory- Information theory is a branch of applied mathematics, electrical engineering, the quantification of information. Information theory was developed by Claude E. Shannon to find fundamental limits on signal processing operations such as compressing data and reliably storing and communicating data. A key measure of information is entropy, which is usually expressed by the average number of bits needed to store or communicate one symbol in a message. Entropy quantifies the uncertainty involved in predicting the value of a random variable. Applications of fundamental topics of information theory include loss data compression. The field is at the intersection of mathematics, statistics, computer science, physics, neurobiology, and electrical engineering. Its impact has been crucial to the success of the Voyager missions to deep space, the invention of the compact disc, the feasibility of mobile phones, the development of the Internet, the study of linguistics and human perception, the understanding of black holes, and numerous other fields. Important sub-fields of information theory are source coding, channel coding, algorithmic complexity theory, algorithmic information theory, information-theoretic security, and measures of information.

Maximum Margin Principle- In geometry, a maximum-margin hyper-plane is a hyper-plane that separates two 'clouds' of points and is at an equal distance from the two. The margin between the hyper-plane and the clouds is maximal. Classifying data is a common task in machine learning. Suppose some given data points each belong to one of two classes, and the goal is to decide which class a new data point will be in. Many hyper-planes might classify the data. One reasonable choice as the best hyper-plane is the one that represents the largest separation, or margin, between the two classes. So we choose the hyper-plane so that the distance from it to the nearest data point on each side is maximized. If such a hyper-plane exists, it is known as the maximum-margin hyper-plane and the linear classifier it defines is known as a maximum margin classifier; or equivalently, the perceptron of optimal stability. In machine learning, support vector machines are supervised learning models with associated learning algorithms that analyze data and recognize patterns, used for classification and regression analysis. The basic SVM takes a set of input data and predicts [10], for each given input, which of two possible classes forms the output, making it a non-probabilistic binary linear classifier. Given a set of training examples, each marked as belonging to one of two categories, an SVM training algorithm builds a model that assigns new examples into one category or another. An SVM model is a representation of the examples as points in space, mapped so that the examples of the separate categories are divided by a clear gap that is as wide as possible. New examples are then mapped into that same space and predicted to belong to a category based on which side of the gap they fall on.

Pattern Recognition- Pattern recognition is the study of how machines can observe the environment, learn to distinguish a pattern of interest from their background, and make sound and reasonable decisions about the categories of the pattern. Despite almost 50 years of research, the design of a general-purpose machine pattern recognizer in most instances is humans, yet we do not understand how humans recognized patterns. Ross emphasized the work of Nobel laureate Herbert Simon whose central finding was that pattern recognition is critical in most human decision-making tasks: “The more relevant pattern at your disposal, the better your decision will be. The pattern recognition may consist of task: supervised classification in which the input pattern is identified as a member of a predefined class.

Stopping Criteria- In the early stopping criterion [11], the available data are divided into three subsets. The first subset is the training data set, which is used for computing the gradient and updating the network weights and biases. The second subset is used as a validation data set and the third subset is used to evaluate the final accuracy of the NN. The error on the validation data set is monitored during the training process. After some number of iterations, the NN begins to overfit the data, and consequently, the error on the validation data set begins to rise. To deal with this problem, when the validation error increases during a specified number of iterations, the algorithm stops the training section and applies the weights and biases at the minimum of the validation error in the NN model. Our proposed approaches also follow the early stop criterion.

3. ALGORITHMS USED

The design and development of the neural network process are divided into three algorithms. The first one is Maximum-Margin gradient descent with an adaptive learning rate algorithm (MMGDX). It stretches out distance between two classes (margin) value to its limit. It uses Area under ROC curve (AUC) information for stopping criterion.

The second one is the Minimization of inter-class interference (MICI), which generates a hidden layer where the pattern has desirable statistical distribution. It also uses AUC for stopping criterion over a validation data set. The last one is a novel algorithm named Assembled neural network (ASNN) which composes hidden neurons from previously trained algorithms. The accuracy of ASNN is found to be better than other state-of-art classifiers.

In this project, a design and development of a Generalized Supervised Assembled Neural Network have to be implemented. Briefly, the work can be summarized into the following phases

- Algorithm 1: Maximum-Margin gradient descent with adaptive learning rate algorithm (MMGDX).
- Algorithm 2: Minimization of inter-class interference (MICI).
- Algorithm 3: Assembled Neural Network (ASNN).
- Algorithm 4: Mutual information calculation.
- Algorithm 5: Neuron selection by Genetic Algorithm (GA).

In this section, the focus is on a non-linear classifier. In machine learning, classification is considered as an instance of supervised learning, i.e. learning where a training set of correctly identified observations is available.

Fig 3.1: Basic framework

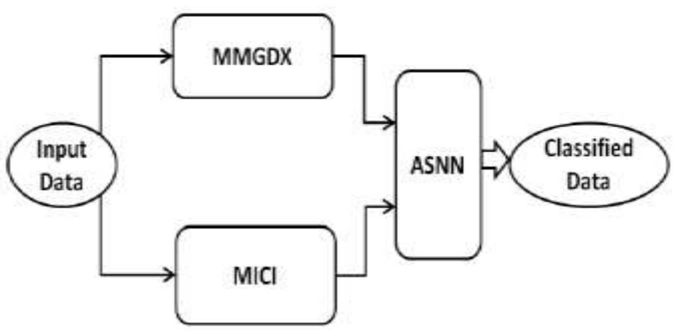

Fig 3.2: Flow of execution

3.1 Maximum-Margin gradient descent with adaptive learning rate (MMGDX)

In gradient descent, an error is equal to target output minus calculated output. However, in MMGDX, the calculation of error is based on support vector norm, target output, and the calculated output. For stopping criteria AUC curve information is used. In the MMGDX algorithm, the hidden layer and output layer are jointly optimized in a single process. The objective function is back-propagated through the output and hidden layers in such a way as to create a hidden output especially oriented toward a larger margin for the output layer.

3.2 Minimization of interclass interference (MICI)

The second training method (MICI) has two phases, the first one is hidden layer training and the second one is output layer training. Hidden layer training is the same as MMGDX, but the output layer is trained based on fisher linear discriminant analysis. It generates a hidden layer where the patterns have a desirable statistical distribution. The MICI creates a hidden space where the Euclidean distance between the prototypes of each class is increased, and the pattern dispersion of each class is decreased.

3.3 Assembled Neural Network (ASNN)

The third and the most important methodology is named assembled neural network (ASNN), which composes a neural model by using neurons extracted from previously trained networks like MICI, MMGDX. As results of MMGDX and MICI shows that accuracy is dependent on the problem. But aim to offer a robust training framework that can take the best of each training method. The choice of neurons aims to maximize the mutual information between target output and hidden output.

4.EXPERIMENTATION AND RESULT

Here we use this algorithm to get more accurate output all these three algorithms are used to design a neural network to test these algorithms we consider some data sets. For evaluating the performance of the algorithms, two kinds of data sets are considered Thyroid data set and the Breast cancer data set. For measuring performance, accuracy is considered a quality factor. The training and testing data sets are from the Neural Information Processing Systems (NIPS), Feature Selection Challenge 2003 [15]. Two types of data sets, Thyroid and Breast cancer data set with different characteristics, in the number of features, the size of data sets, and data distributions. These data sets are used to trained and testing of neural networks.

Data set	Operation	Attribute		Number of samples
		Input	Output	
Thyroid	Training Set	5	1	140
	Testing Set	5	1	75
Breast-Cancer	Training Set	9	1	200
	Testing Set	9	1	77

Table 4.1: Data set Information

Figure 4.1 and 4.2 show data set distribution, and it justifies that the breast cancer data set is more complex and it has more number of examples.

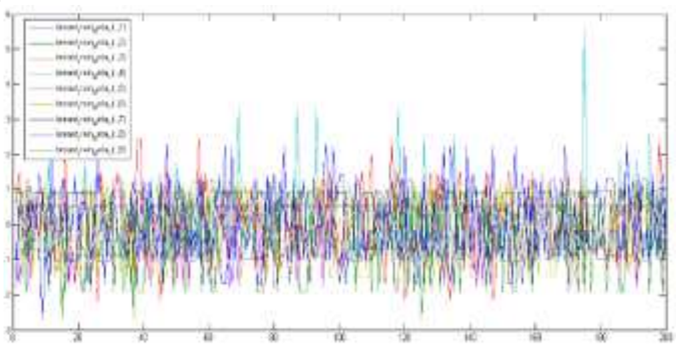

Fig 4.1: Breast cancer data distribution

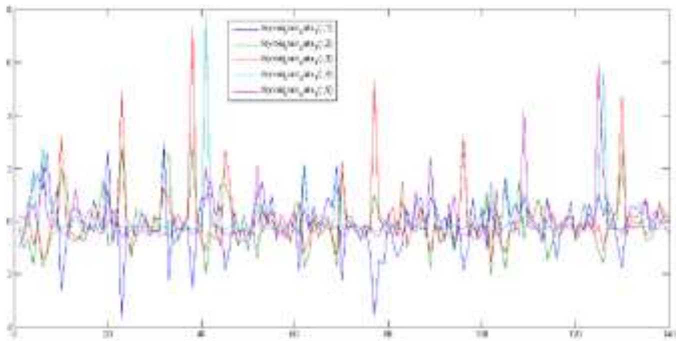

Fig 4.2: Thyroid data distribution

Clustering(K-means)- K-means clustering is a method of cluster analysis that aims to partition *n* observations into *k* clusters in which each observation belongs to the cluster with the nearest mean. It is the simplest unsupervised learning algorithm. K-means algorithm is applied on both data sets for data analysis. K-means result is shown in Table 4.2. It justifies that the k-means give very negligible accuracy.

Cluster (k)	Thyroid data set Accuracy		Breast-Cancer data set Accuracy	
	Training	Testing	Training	Testing
2	17.85	24	18.50	22.07
3	11.42	13.33	13.50	16.80
4	5.71	9.33	14.00	12.98
5	9.28	10.66	6.50	13.38
6	5.43	6.66	3.33	9..9

Table 4.2: Clustering (k-means) results

Accuracy- Experimentation is performed on the algorithms (MMGDX, MICI, and ASNN) with GDX to check the accuracy of both data sets. The GDX is standard Gradient Descent with Adaptive Learning Rate, which is available in the standard MATLAB library. For calculating accuracy on testing data, firstly neural networks are trained using training data and then test data is applied. Figures 4.3 and 4.4 shows the accuracy of testing data on three algorithms with GDX.

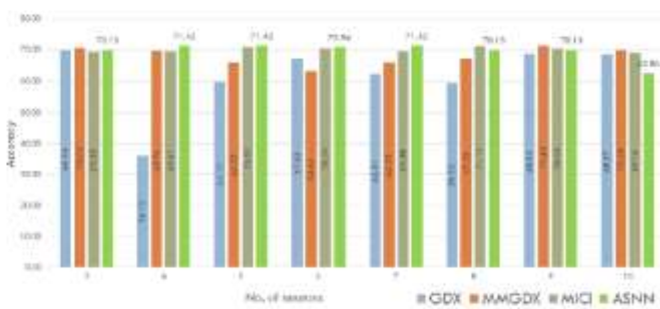

Fig 4.3: Breast Cancer accuracy on Test Data

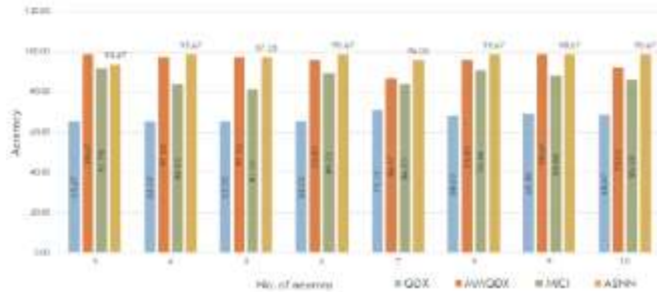

Fig 4.4: Thyroid accuracy on Test Data

Fig 4.5: Breast Cancer accuracy on Train Data

Fig 4.6: Thyroid accuracy on Train Data

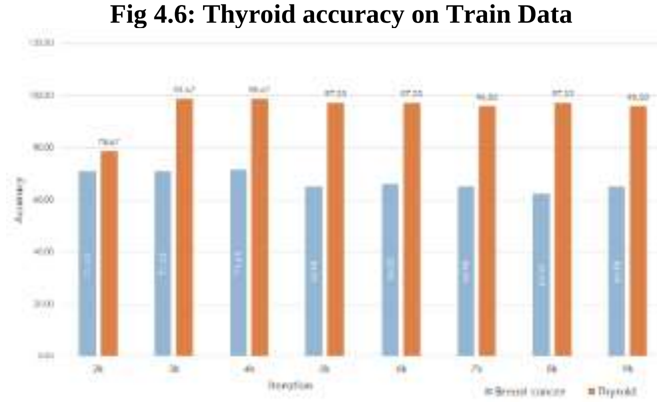

Fig 4.7: Accuracy vs Iteration

For calculating accuracy on training data, firstly proposed neural networks are trained using train data, and then also train data is applied. Figures 4.5 and 4.6 shows the accuracy of training data on three proposed algorithm with GDX. Presented algorithms have better accuracy because it has better error objective function.

In GDX, an error is usually adopted, which permits the undesirable increase of the output matrix $w2$ to achieve the target output y without taking into account the distance d. ASNN gives evidence that an ensemble of a classifier is often better than a single classifier. From figure 4.3 to 4.6, it is justified that all three presented algorithms solve the over-fitting problem because, for GDX, less accuracy is found on test data but accuracy is better on training data. MMGDX, MICI, ASNN provides similar accuracy on both training and testing data.

From figure 4.3 to 4.6, it is justified that accuracy is not increased after the optimal number of neurons in the hidden layer. Finding the optimal number of neurons in the hidden layer is still an open problem.

5. CONCLUSION

The presented MMGDX, MICI, and ASNN algorithms provide better accuracy as compared to another state-off-art classifier. Results obtained after applying it to real-world benchmark data-set show better accuracy. All these algorithms are applied without feature selection. All presented algorithms have a quite similar accuracy on both data sets. ASNN is the first ranked algorithm it has near about 90% accuracy and MMGDX is the second overall ranked algorithm on both data sets, in terms of accuracy having near about 70% accuracy. This gives evidence of problem independent robustness of ASNN. That means an ensemble of the classifier is often better than a single classifier.

In this project, we design and develop a neural network that is based on the back-propagation method. We also used algorithms like maximum margin gradient descent with adaptive learning rate algorithm (MMGDX) and minimization of interclass interference (MICI) based on the back-propagation method to design a neural network. For more accuracy assembled neural network (ASNN) is used. For evaluating the performance of algorithms we used different data sets.

.

REFERENCES

[1] Oswaldo Ludwig and Urbano Nunes, "Novel Maximum-Margin Training Algorithms for Supervised Neural Networks," in IEEE Transactions on Neural Networks, vol. 21, no. 6, pp. 972-984, June 2010.

[2] Paul J. Werbos (1994). The Roots of Backpropagation. From Ordered Derivatives to Neural Networks and Political Forecasting. New York, NY: John Wiley & Sons, Inc.

[3] D. A. White and D. A. Sofge, Eds., Handbook of Intelligent Control Neural, Fuzzy, and Adaptive Approaches. Reinhold, 1992.

[4] P. J. Werbos and X. Pang, “Generalized maze navigation: SRN critics solve what feedforward or Hebbian cannot,” in Proc. Conf. Syst. Man Cybern., 1996, pp. 1764–1769.

[5] A. R. Barron, “Approximation and estimation bounds for artificial neural networks,” Mach. Learn., vol. 14, no. 1, pp. 115–133, 1994.

[6] D. Prokhorov, R. Santiago, and D. Wunsch, “Adaptive critic designs,” IEEE Trans. Neural Netw., vol. 8, no. 5, pp. 997–1007, Sep. 1997.

[7] X. Pang and P. Werbos, “Neural network design for J function approximation in dynamic programming,” Math Model. Sci. Comp. vol. 2, 1996.

[8] Jinghua Wang, JaneYou, QinLi, YongXu "Extract minimum positive and maximum negative features for imbalanced binary classification," Elsevier (2011).

[9] H. Peng, F. Long, and C. Ding, "Feature selection based on mutual information: Criteria of max-dependency, max-relevance, and min-redundancy," IEEE Trans Pattern Anal. Mach. Intell., vol. 27, no. 8, pp. 1226-1238, Aug. 2005.

[10] O. Ludwig, U. Nunes, R. Araujo, L. Schnitman, and H. A. Lepikson, “Applications of information theory, genetic algorithms, and neural models to predict oil flow,” Commun. Nonlinear Sci. Numer. Simul., vol. 14, no. 7, pp. 2870–2885, 2009.

[11] K. Zhang, I. W. Tsang, and J. T. Kwok, “Maximum margin clustering made practical,” IEEE Trans. Neural Netw., vol. 20, no. 4, pp. 583–596, Apr. 2009.

[12] N. K. Treadgold and T. D. Gedeon, “Exploring constructive cascade networks,” IEEE Trans. Neural Netw., vol. 10, no. 6, pp. 1335–1350, Nov. 1999.

[13] G. V. Puskorius and L. A. Feldkamp, “Neuro control of nonlinear dynamical systems with Kalman filter trained recurrent networks,” IEEE Trans. Neural Netw., vol. 5, no. 2, pp. 279–297, Mar. 1994.

[14] L. Feldkamp, D. Prokhorov, C. Eagen, and F. Yuan, “Enhanced multistream Kalman filter training for recurrent networks,” in Nonlinear Modeling: Advanced Black-Box Techniques. Norwell, MA: Kluwer, 1998, pp. 29–53.

[15] F. Rosenblatt, Principles of Neural Dynamic. New York: Spartan, 1962.

[16] T.M. Khoshgoftaar, J. Van Hulse, A. Napolitano, "Supervised neural network modeling: an empirical investigation into learning from imbalanced data with labeling errors," IEEE Transactions on Neural Networks 21 (5) (2010) 813-830.