Available online at www.sciencedirect.com
ScienceDirect
Expert Systems with Applications 36 (2009) 944–949
Expert Systems with Applications
www.elsevier.com/locate/eswa

A comparative study on thyroid disease diagnosis using neural networks

Feyzullah Temurtas *

Sakarya University, Department of Computer Engineering, 54187 Adapazari, Turkey
Sakarya University, Department of Electrical and Electronics Engineering, 54187 Adapazari, Turkey

Abstract

Thyroid hormones produced by the thyroid gland help regulation of the body's metabolism. Abnormalities of thyroid function are usually related to production of too little thyroid hormone (hypothyroidism) or production of too much thyroid hormone (hyperthyroidism). Thyroid disease diagnosis via proper interpretation of the thyroid data is an important classification problem. In this study, a comparative thyroid disease diagnosis were realized by using multilayer, probabilistic, and learning vector quantization neural networks. For this purpose, thyroid disease dataset which is taken from UCI machine learning database was used.

Keywords: Thyroid disease diagnosis; Multilayer neural network; Probabilistic neural network; Learning vector quantization

1. Introduction

Thyroid hormones produced by the thyroid gland helps control the body's metabolism. The thyroid gland produces two active thyroid hormones, levothyroxine (abbreviated T4) and triiodothyronine (abbreviated T3). These hormones are important in the production of proteins, in the regulation of body temperature, and in overall energy production and regulation. The seriousness of thyroid disorders should not be underestimated as thyroid storm (an episode of severe hyperthyroidism) and myxedema coma (the end stage of untreated hypothyroidism) may lead to death in a significant number of cases. (Ozyılmaz & Yıldırım, 2002; Polat, Sahan, & Gunes, 2007; Zhang & Berardi, 1998).

In general, thyroid disease can be divided into two broad groups of disorders: those, which primarily affect the function of the thyroid gland and those, which involve neoplasms, or tumors, of the thyroid. Both types of disorders are relatively common in the general population. Most thyroid problems can be treated successfully. Abnormalities of thyroid function are usually related to production of too little thyroid hormone (hypothyroidism) or production of too much thyroid hormone (hyperthyroidism) (Ozyılmaz & Yıldırım, 2002; Polat et al., 2007).

Hyperthyroidism, or an overactive thyroid, may also be caused by inflammation of the thyroid, various kinds of medications, and lack of control of thyroid hormone production. One of the most common causes is Graves' disease. Graves' disease happens when the body makes proteins that constantly tell the thyroid to make more thyroid hormone (Ozyılmaz & Yıldırım, 2002).

In the correct of diagnosis of thyroid disorders, proper interpretation of the thyroid data besides clinical examination and complementary investigation is an important issue. Thyroid function diagnosis is an important classification problem (Hoshi et al., 2005; Ozyılmaz & Yıldırım, 2002; Polat et al., 2007). Various new methods, such as pattern recognition techniques, fuzzy classifiers, artificial immune recognition system, neural networks, etc., have been used to fit patients into a well defined status (Delen, Walker, & Kadam, 2005; Hoshi et al., 2005; Ozyılmaz & Yıldırım, 2002; Polat et al., 2007).

In this study, a comparative thyroid disease diagnosis were realized by using three different types of neural networks (NNs). The neural network classification models are multilayer neural network (MLNN), probabilistic

* Address: Sakarya University, Department of Computer Engineering, 54187 Adapazari, Turkey. Tel.: +90 532 207 56 33.
E-mail address: temurtas@gmail.com

0957-4174/$ - see front matter © 2007 Elsevier Ltd. All rights reserved.
doi:10.1016/j.eswa.2007.10.010

neural network (PNN) and learning vector quantization neural network (LVQ-NN). In order to perform the study, the UCI machine learning database which is very commonly used among the other classification systems were used (Hoshi et al., 2005; Ozyılmaz & Yıldırım, 2002; Polat et al., 2007; <ftp://ftp.ics.uci.edu/pub/machine-learningdatabases> (last accessed: 15 April 2007)). 3-fold and 10-fold cross-validation technique were performed to compare the accuracy of the neural network models. The results were also compared with the results of the pervious studies reported (Ozyılmaz & Yıldırım, 2002; Polat et al., 2007) focusing on thyroid disease diagnosis and using same database.

2. Method

2.1. Data source

In order to perform the research reported in this manuscript, the thyroid dataset taken from the UCI machine learning respiratory were used (Hoshi et al., 2005; Ozyılmaz & Yıldırım, 2002; Polat et al., 2007; <ftp://ftp.ics.uci.edu/pub/machine-learningdatabases> (last accessed: 15 April 2007)). The reason for using this dataset is that because it is very commonly used among the other classification systems that we have used to compare this study with for thyroid diagnosis problem. The dataset which consists of the thyroid disease measurements contains three classes and 215 samples. The class distribution is

- Class 1: normal (150)
- Class 2: hyper (35)
- Class 3: hypo (30)

All samples have five features. These features are:

- Feature 1: T3-resin uptake test (A percentage).
- Feature 2: Total serum thyroxin as measured by the isotopic displacement method.
- Feature 3: Total serum triiodothyronine as measured by radioimmuno assay.
- Feature 4: Basal thyroid-stimulating hormone (TSH) as measured by radioimmuno assay.
- Feature 5: Maximal absolute difference of TSH value after injection of 200 mg of thyrotropin-releasing hormone as compared to the basal value.

2.2. Previous studies

Classification systems have been used for thyroid disease diagnosis problem as for other clinical diagnosis problems. There have been several studies reported focusing on thyroid disease diagnosis (Ozyılmaz & Yıldırım, 2002; Polat et al., 2007). These studies applied different methods to the given problem and achieved high classification accuracies using the dataset taken from UCI machine learning repository. Ozyılmaz and Yıldırım used MLP with bp ($3\times$FC), MLP with fbp ($3\times$FC), RBF ($3\times$FC) and CSFNN ($3\times$FC) for this dataset and obtained classification accuracies of 86.33%, 89.80%, 79.08% and 91.14%, respectively (Ozyılmaz & Yıldırım, 2002). Polat et al. used AIRS ($10\times$FC) and AIRS with Fuzzy weighted pre-processing ($10\times$FC) for the same dataset and obtained classification accuracies of 81.00% and 85.00%, respectively (Polat et al., 2007).

2.3. Diagnosis of the thyroid disease using multilayer neural network

The multilayer neural networks (MLNNs) have been successfully used in replacing conventional pattern recognition methods for the disease diagnosis systems (Delen et al., 2005; Ozyılmaz & Yıldırım, 2002; Polat et al., 2007).

In the first stage of this study, a multilayer neural network structure was used for the thyroid disease diagnosis. The network structure used for this purpose is shown in Fig. 1. These networks were the multilayer networks (input layer, hidden layers, and output layer). The hidden layer neurons (50 neuron for each hidden layer) and the output layer neurons use nonlinear sigmoid activation functions. In this system, five inputs were features, and three outputs are index of three classes. Equations used in the neural network model are shown in (1)–(3).

Outputs of the first hidden layer neurons are,

$$\vec{X}^{ih1}(n) = 1/(1+\exp(W^{ih1}(n) * \vec{f}(n) + \vec{b}^{ih1}(n))) \quad (1)$$

Outputs of the second hidden layer neurons are,

$$\vec{X}^{ih2}(n) = 1/(1+\exp(W^{ih2}(n) * \vec{X}^{ih1}(n) + \vec{b}^{ih2}(n))) \quad (2)$$

Outputs of the network are,

$$\vec{Y}(n) = 1/(1+\exp(W^{ho}(n) * \vec{X}^{ih2}(n) + \vec{b}^{ho}(n))) \quad (3)$$

where $W^{ih1}(n)$ are the weights from the input to the first hidden layer and $\vec{b}^{ih1}(n)$ are the biases of the first hidden layer, $W^{ih2}(n)$ are the weights from the first hidden layer to the second hidden layer and $\vec{b}^{ih2}(n)$ are the biases of the second hidden layer, $W^{ho}(n)$ the weights from the second hidden layer to the output layer and $\vec{b}^{ho}(n)$ are the biases of the output layer, $\vec{f}(n)$ values the features, $\vec{Y}(n)$ values the outputs for the class index, and n a training pattern index.

The back-propagation (BP) algorithm (Rumelhart, Hinton, & Williams, 1986) is widely recognized as a powerful tool for training of the MLNNs. But, since it applies the steepest descent method to update the weights, it suffers from a slow convergence rate and often yields suboptimal solutions (Brent, 1991; Gori & Tesi, 1992). A variety of related algorithms have been introduced to address that problem. A number of researchers have carried out comparative studies of MLNN training algorithms (Hagan & Menhaj, 1994; Hagan, Demuth, & Beale, 1996; Sagiroglu, Besdok, & Erler, 2000). Levenberg–Marquardt (LM)

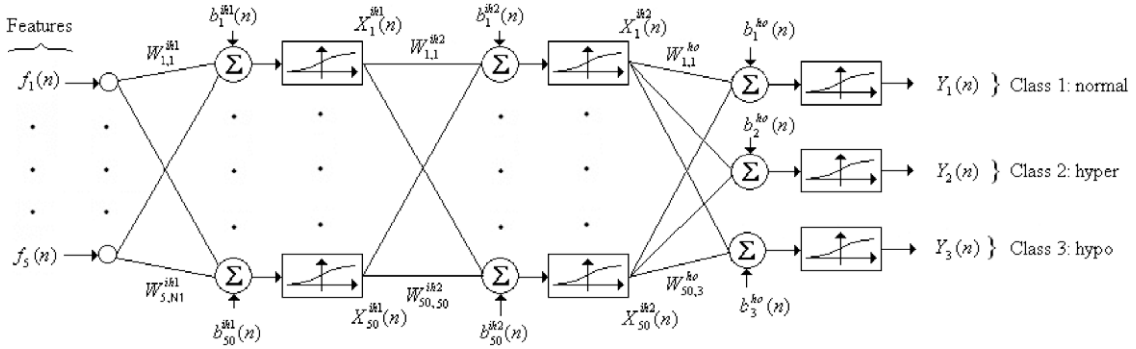

Fig. 1. Implementation of multilayer neural network for the thyroid disease diagnosis.

algorithm (Hagan & Menhaj, 1994) used in this study is one of the fastest type of these algorithms. Detailed computational issues about the application of the training algorithm to MLNN can be found in Gulbag and Temurtas (2006), Matlab Documentation (2004).

2.4. Diagnosis of the thyroid disease using probabilistic neural network

The probabilistic neural network (PNN) developed by Donald Specht (Specht, 1990) is a network formulation of 'probability density estimation'. It is a model based on competitive learning with a 'winner takes all attitude' and the core concept based on multivariate probability. The PNN provides a general solution to pattern classification problems by following an approach developed in statistics, called Bayesian classifiers. The network paradigm also uses Parzen Estimators which were developed to construct the probability density functions required by Bayes theory. The PNN uses a supervised training set to develop distribution functions within a pattern layer. Training of the PNN is much simpler than that of the MLNN. However, the pattern layer can be quite huge if the distinction between categories is varied and at the same time quite similar in special areas (Gulbag, 2006; Yusubov, Gulbag, & Temurtas, 2007).

At the third stage of this study, a probabilistic neural network was used for the the thyroid disease diagnosis. The network structure used for this purpose is shown in Fig. 2.

The PNN structure used in this study has a multilayer structures consisting of a single hidden layer (radial basis layer) of locally tuned units which are fully interconnected to an output layer (competitive layer) of three units, as shown in Fig. 2. In this system, real valued input vector is feature's vector, and three outputs are index of three classes. All hidden units simultaneously receive the five-dimensional real valued input vector. The input vector to the network is passed to the hidden layer nodes via unit connection weights. The hidden layer consists of a set of radial basis functions. Associated with jth hidden unit is a parameter vector, called $\vec{c}_j$ a center. The hidden layer node calculates the Euclidean distance between the center and the network input vector and then passes the result to the radial basis function. All the radial basis functions are the same type (Gaussian). Equations which used in the neural network model are shown in the following

$$X_j = \phi(\|\vec{f} - \vec{c}_j\| * b^{\mathrm{ih}}) \tag{4}$$

$$\phi(x) = \exp(-x^2) \tag{5}$$

$$b^{\mathrm{ih}} = 0.833/s \tag{6}$$

$$S_i = \sum_{j=1}^{h} W_{ji}^{\mathrm{ho}} * X_j \tag{7}$$

$$Y_i = \begin{cases} 1 & \text{if } S_i \text{ is max of } \{S_1, S_2, S_3\} \\ 0 & \text{else} \end{cases} \tag{8}$$

where $i = 1,2,3$, $j = 1,2,\ldots,h$, Y_i is the ith output (classification index), $\vec{f}$ the five-dimensional real valued input vector, W_{ji}^{ho} the weight between the jth hidden node and the ith output node, $\vec{c}_j$ the center vector of the jth hidden node, s the real constant known as spread factor, bth the biasing term of radial basis layer, and $\phi(.)$ the nonlinear radial basis function (Gaussian).

The PNN structures employed in the study utilized the newpnn function implemented in MATLAB. Detailed information about the realisation of the PNN structures can be found in the neural network toolbox part of MATLAB Documentation (Matlab Documentation, 2004).

2.5. Diagnosis of the thyroid disease using learning vector quantization neural network

The learning vector quantization (LVQ) (Kohonen, 1990; Kohonen, 1997) partitions the feature space of the training data into clusters and models prototypes for these clusters. The classification of the LVQ is based on the similarity of the unknown data and these prototypes. An LVQ neural network has a competitive layer and linear output layer. The competitive layer learns to classify input vectors. The linear output layer transforms the competitive layer's classes into target classifications defined by the user. The classes learned by the competitive layer can be referred as subclasses and the classes of the linear output layer can

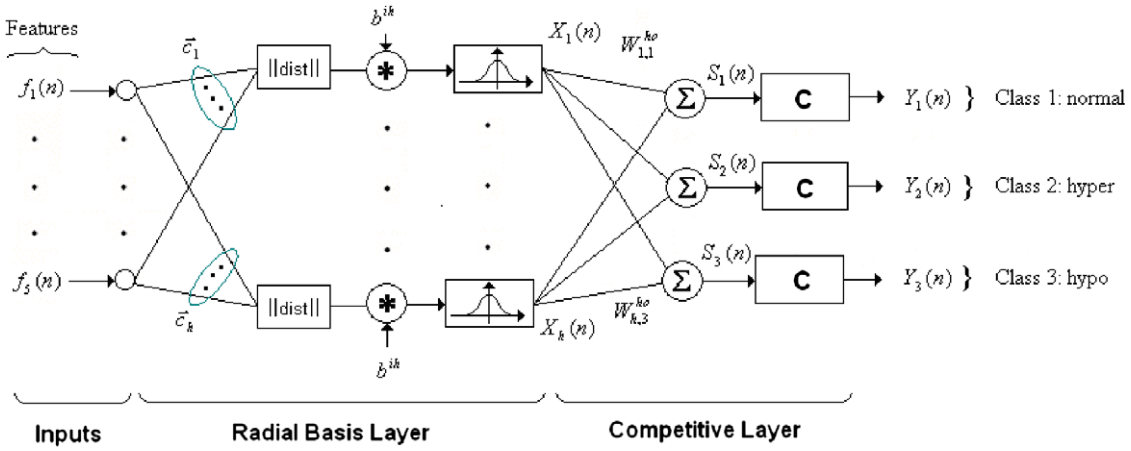

Fig. 2. Implementation of probabilistic neural network for the thyroid disease diagnosis.

be referred as target classes. (Matlab Documentation, 2004).

At the forth stage of this study, a learning vector quantization neural network was used for the the thyroid disease diagnosis. The network structure used for this purpose is shown in Fig. 3.

The LVQ structure used in this study has a multilayer structures consisting of a single hidden layer (competitive layer) and an output layer (linear layer) of three units, as shown in Fig. 3. In this system, real valued input vector is feature's vector, and three outputs are index of three classes. All hidden units simultaneously receive the five-dimensional real valued input vector. The input vector to the network is passed to the hidden layer nodes via unit connection weights. The hidden layer consists of a set of competition functions. Associated with jth hidden unit is a parameter vector, called $\vec{c}_j$ a center. The hidden layer node calculates the Euclidean distance between the center and the network input vector and then passes the result to the competition functions. Equations which used in the neural network model are shown in (9)–(11).

$$S_j = \|\vec{f} - \vec{c}_j\| \tag{9}$$

$$X_j = \begin{cases} 1 & \text{if } S_j \text{ is max of } \{S_1, \ldots, S_h\} \\ 0 & \text{else} \end{cases} \tag{10}$$

$$Y_i = \sum_{j=1}^{h} W^{\text{ho}}_{ji} * X_j \tag{11}$$

where $i = 1,2,3\, j = 1,2,\ldots,h$, Y_i is the ith output (classification index), $\vec{f}$ the five-dimensional real valued input vector, W^{ho}_{ji} the weight between the jth hidden node and the ith output node, $\vec{c}_j$ the center vector of the jth hidden node.

The LVQ structures employed in the study utilized the newlvq function implemented in MATLAB. Detailed information about the realisation of the LVQ structures can be found in the neural network toolbox part of MATLAB Documentation (Matlab Documentation, 2004).

2.6. Measures for performance evaluation

2.6.1. Classification accuracy

In this study, we used the classification accuracies as performance measures (Watkins, 2001):

$$\text{classification accuracy}(N) = \frac{\sum_{i=1}^{|N|} \text{assess}(n_i)}{|N|}, \quad n_i \in N \tag{12}$$

$$\text{assess}(n) = \begin{cases} 1 & \text{if classify}(n) = nc \\ 0 & \text{otherwise} \end{cases} \tag{13}$$

where N is the set of data items to be classified (the test set), $n \in N$, nc is the class of the item n, and classify(n) returns the classification of n by neural networks.

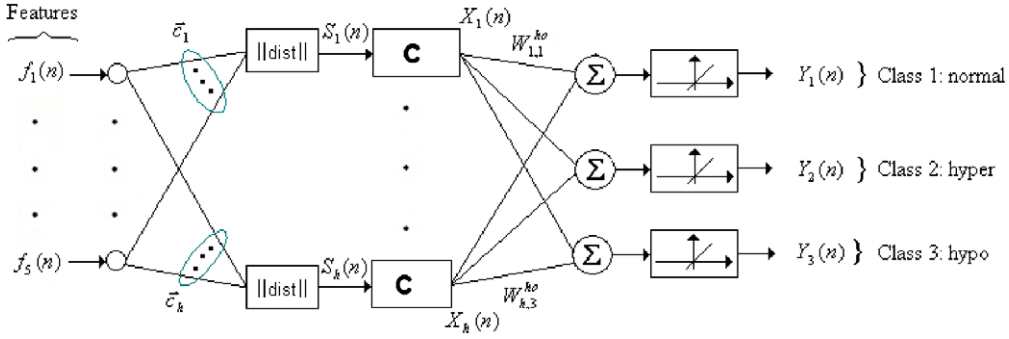

Fig. 3. Implementation of learning vector quantization neural network for the thyroid disease diagnosis.

2.6.2. k-Fold cross-validation

In order to minimize the bias associated with the random sampling of the training and holdout data samples in comparing the predictive accuracy of two or more methods, researchers tend to use *k*-fold cross-validation (Delen et al., 2005; Fayyad, Piatetsky-Shapiro, Smyth, & Uthurusamy, 1996; Ozyılmaz & Yıldırım, 2002; Polat et al., 2007). In *k*-fold cross-validation, whole data are randomly divided to *k* mutually exclusive and approximately equal size subsets. The classification algorithm trained and tested *k* times. In each case, one of the folds is taken as test data and the remaining folds are added to form training data. Thus *k* different test results exist for each training-test configuration (Delen et al., 2005). The average of these results gives the test accuracy of the algorithm. If a neural network learns the training set of a problem, it makes generalisation to that problem. So, this type trained neural network gives similar result for untrained test sets also. But, if a neural network starts to memorize the training set, its generalisation starts to decrease and it's performance may not be improved for untrained test sets (Gulbag & Temurtas, 2006). The *k*-fold cross-validation method shows how good generalisation can be made using neural network structures (Ozyılmaz & Yıldırım, 2002). In this study, 3-fold and 10-fold cross-validation approaches were used to estimate the performance of the used neural networks.

3. Results

The classification accuracies obtained by this and other studies for thyroid disease dataset were presented in Table 1. From this table, it can be seen easily that, 3-fold and 10-fold cross-validation approaches give almost similar classification results for the same methods used in this study. So, the comparison between methods used by this study and other studies can be made easily without any problem.

Table 1
Neural networks classification accuracies for thyroid disease dataset problem with classification accuracies obtained by other studies

Study	Method	Classification accuracy (%)
Ozyılmaz and Yıldırım (2002)	MLP (MLNN) with bp (3 × FC)	86.33
	MLP (MLNN) with fbp (3 × FC)	89.80
	RBF (3 × FC)	79.08
	CSFNN (3 × FC)	91.14
Polat et al. (2007)	AIRS (10 × FC)	81.00
	AIRS with fuzzy weighted pre-processing (10 × FC)	85.00
This study	MLNN with LM (3 × FC)	92.96
	PNN (3 × FC)	94.43
	LVQ (3 × FC)	89.79
	MLNN with LM (10 × FC)	93.19
	PNN (10 × FC)	94.81
	LVQ (10 × FC)	90.05

The classification accuracy of MLNN obtained by this study was better than those obtained by Ozyılmaz and Yıldırım (2002). This can be because of that Levenberg–Marquardt (LM) training algorithm converges better than the training algorithms used by Ozyılmaz and Yıldırım (2002) for thyroid disease classification MLNN structures. The classification accuracies of AIRS and AIRS with Fuzzy weighted pre-processing obtained by Polat et al. (2007) were better than the classification accuracy of RBF obtained by Ozyılmaz and Yıldırım (2002). But, these classification accuracies were worse than classification accuracies of the other type neural networks obtained by Ozyılmaz and Yıldırım (2002) and this study.

From the same table, it can be seen also that the best results for the classification accuracy were obtained from PNN structure used in this study. The second good performance were obtained from MLNN with LM performed in this study. The classification performance of LVQ used in this study were between the classification performances of the CSFNN and MLNN with fbp used by Ozyılmaz and Yıldırım (2002).

4. Conclusions

This paper presents a comparative study on thyroid disease diagnosis by using multilayer, probabilistic and learning vector quantization neural networks. The results were also compared with the results of the pervious studies reported by Ozyılmaz and Yıldırım (2002), and Polat et al. (2007).

As the conclusion, the following results can be summarised;

- It was seen that neural network structures could be successfully used to help diagnosis of thyroid disease.
- The neural network structures used in this study generally show better performances than the methods used by Ozyılmaz and Yıldırım (2002), and Polat et al. (2007).
- And the results shows that probabilistic neural network gives the best classification accuracies for thyroid disease dataset.
- The 3-fold and 10-fold cross-validation method used show how good generalisation can be made using neural network structures. So, the performances of the neural networks for inputs that are not in the training set can be seen.

References

Brent, R. P. (1991). Fast training algorithms for multi-layer neural nets. *IEEE Transactions on Neural Networks, 2*, 346–354.

Delen, D., Walker, G., & Kadam, A. (2005). Predicting breast cancer survivability: A comparison of three data mining methods. *Artificial Intelligence in Medicine Artificial Intelligence in Medicine, 34*(2), 113–127.

Fayyad, U. M., Piatetsky-Shapiro, G., Smyth, P., & Uthurusamy, R. G. R. (1996). *Advances in knowledge discovery and data mining*. Menlo Park, CA: AAAI Press/The MIT Press.

Gori, M., & Tesi, A. (1992). On the problem of local minima in backpropagation. *IEEE Transactions on Pattern Analysis and Machine Intelligence, 14*, 76–85.

Gulbag, A. (2006). PhD Thesis, Sakarya University, Institute of Science and Technology.

Gulbag, A., & Temurtas, F. (2006). A study on quantitative classification of binary gas mixture using neural networks and adaptive neuro fuzzy inference systems. *Sensors and Actuators B, 115*, 252–262.

Hagan, M. T., & Menhaj, M. (1994). Training feed forward networks with the Marquardt algorithm. *IEEE Transactions on Neural Networks, 5*, 989–993.

Hagan, M. T., Demuth, H. B., & Beale, M. H. (1996). *Neural network design*. Boston, MA: PWS Publishing.

Hoshi, K., Kawakami, J., Kumagai, M., Kasahara, S., Nisimura, N., Nakamura, H., et al. (2005). An analysis of thyroid function diagnosis using Bayesian-type and SOM-type neural networks. *Chemical and Pharmaceutical Bulletin, 53*, 1570–1574.

Kohonen, T. (1990). Improved versions of learning vector quantization. In *Proceedings of the IEEE international joint conference on neural networks* (pp. 545–550). New York.

Kohonen, T. (1997). *Self-organizing maps* (2nd ed.). Berlin: Springer-Verlag.

Matlab® Documentation. (2004). Version 7.0, Release 14, The MathWorks, Inc.

Ozyılmaz, L., Yıldırım, T. (2002). Diagnosis of thyroid disease using artificial neural network methods. In *Proceedings of ICONIP'02 9th international conference on neural information processing* (pp. 2033–2036). Singapore: Orchid Country Club.

Polat, K., Sahan, S., & Gunes, S. (2007). A novel hybrid method based on artificial immune recognition system (AIRS) with fuzzy weighted pre-processing for thyroid disease diagnosis. *Expert Systems with Applications, 32*, 1141–1147.

Rumelhart, D. E., Hinton, G. E., & Williams, R. J. (1986). Learning internal representations by error propagation. In D. E. Rumelhart & J. L. McClelland (Eds.). *Parallel distributed processing: Explorations in the microstructure of cognition* (Vol. 1, pp. 318–362). Cambridge, MA: MIT Press.

Sagiroglu, S., Besdok, E., & Erler, M. (2000). Control chart pattern recognition using artificial neural networks. *Turkish Journal of Electronic Engineering, 8*, 137–146.

Specht, D. F. (1990). Probabilistic neural networks. *Neural Networks, 3*, 109–118.

Watkins, A. (2001). *AIRS: A resource limited artificial immune classifier*. Master Thesis, Mississippi State University.

Yusubov, I., Gulbag, A., & Temurtas, F. (2007). A study on mixture classification using neural network. *Electronic Letter Science and Engineering, 3*, 33–38.

Zhang, G., & Berardi, L. V. (1998). An investigation of neural networks in thyroid function diagnosis. *Health Care Management Science*, 29–37.