J. Basic. Appl. Sci. Res., 3(2s)593-598, 2013

ISSN 2090-4304
Journal of Basic and Applied Scientific Research
www.textroad.com

Thyroid Diseases Diagnosis Using Probabilistic Neural Network and Principal Component Analysis

Javad Haddadnia[1*], Omid Rahmani Seryasat[2], Hamidreza Rabiee[3]

[1]Associate Professor, Electrical and Computer Engineering Department, Hakim Sabzevari University & Center for Research of Advanced Medical Technologies, Sabzevar University of Medical Sciences, Sabzevar, Iran
[2,3]Department of Electrical and Computer Engineering, Hakim Sabzevari University, Sabzevar, Iran

Received: June 10 2013
Accepted: July 10 2013

ABSTRACT

In this article PNN classifier is used to classify Thyroid data. Since the features are numerous, statistical PCA algorithm was used for feature extraction. Two classes of thyroid data are used here which are both taken from UCI site. After implementing two classes of thyroid data each kind of them is divided to three classes as high efficient, low efficient, and normal. to classifier and feature extraction algorithm, it was observed that PNN classifier was able to diagnose these three classes correctly.

KEYWORD: Thyroid diseases, PNN, PCA, neural network

1. INTRODUCTION

Thyroid performance has a serious effect on many basic organs of the body. If thyroid disorder is not recognized on time, the patient will suffer from thyroid attack or coma which might lead to his death. Therefore, true diagnosis of thyroid disorder (Thyroid low efficiency and high efficiency) based on laboratory and clinical tests (disease symptom) are quite vital. The use of neural networks as a smart tool for classifying thyroid data is more accurate and flexible than other approaches.

Today, the use of smart methods in control systems, signal processing and sample recognition is a powerful tool in various scientific, technical and engineering, medical, and medical engineering researches has been greatly emphasized. An example of the use of such systems in different medical and medical engineering fields could be the following: research on various diseases, simulation of various body organs, and simulation of body metabolisms. In this regard, the research on thyroid disease and its diagnosis by means of smart neural methods has been emphasized by researchers. In neural network approach, it has been tried to pattern human's brain functions and nervous system. This method is capable of solving complicated issues by relying on learning capacities and parallel processing in natural neural networks [40, 41]. Neural network capabilities and their application in different issues such as signal processing, sample recognition, patterning, identification, prediction, controlling, and optimization have been emphasized in recent decades and these structures are recently used with regard to their learning capability as a common method which is independent of proposed model.[1-2]

PCA approach is used to extract features and to decrease space dimensions. For the data in which the samples are not distributed to each class similarly the features are chosen for which sample distribution is more. Therefore, each class whose samples are more distributed is more easily diagnosed by those features. Consequently, we face more problems in diagnosing classes which have fewer samples or which are less distributed.

Since the omission of each feature will lead to the loss of important information about the patient, the methods should be used which are highly accurate and are able to classify data such as thyroid data without omitting the features. Since there are many features for diagnosing the diseases and since laboratory features are very suitable and reliable for diagnosing the diseases, it is attempted to use algorithms and methods by which features could be used which are based on laboratory tests. Since the centers of data classes are highly correlated and quite nonlinear, it is difficult to classify such data.

In this article PNN classifier is used to classify Thyroid data. Since the features are numerous, statistical PCA algorithm was used for feature extraction. Two classes of thyroid data are used here which are both taken from UCI site. After implementing two classes of thyroid data to classifier and feature extraction algorithm, it was observed that PNN classifier was able to diagnose these three classes correctly.

2. Preprocessing

In the first part of quantity diagram, classification algorithm is shown.

[*]**Corresponding author:** Associate Professor, Electrical and Computer Engineering Department, Hakim Sabzevari University & Center for Research of Advanced Medical Technologies, Sabzevar University of Medical Sciences, Sabzevar, Iran. Email: Haddadnia@sttu.ac.ir

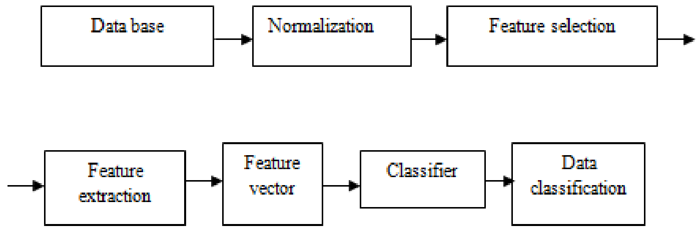

Fig.1: the diagram block of detection of the type of faults.

2.1. Normalization

In many practical situations we face features whose values are within various dynamic ranges. Therefore, high value features might have greater effect on cost function in comparison to low value ones. This problem is solved by normalizing features so that their values are within similar ranges. One normalizing method which is applied in this project is obtained in the following way.

After calculation the means and variance of each feature, the values of each feature are subtracted from the same feature average and then it is divided by the related standard deviation as shown in the following equations:

$$\overline{x_i} = \frac{1}{N}\sum_{i=1}^{N} x_{ik} \qquad k = 1,2,\ldots\ldots,I \tag{1}$$

$$\partial_k^2 = \frac{1}{N-1}\sum_{i=1}^{N}(x_{ik}-\overline{x_i})^2 \tag{2}$$

$$\widehat{x_{ik}} = \frac{x_{ik}-\overline{x_i}}{\partial_1} \tag{3}$$

In these equations *N* is the number of samples and *k* is the number of features. In other words, normalization based on above equations will make the average data be equal to zero and their variance be equal to 1. Data is normalized in two ways:

1. General normalization: In this case the whole date is normalized at the same time.
2. Specific normalization:

In this case, teaching and testing data are separately normalized under certain equation. It should be done that only continuous features are getting normalized.

3. Principal component analysis

Principal component analysis is an unsupervised linear method of feature extraction. If X is vector of rand feature in n-dimensional space with the average of zero and covariance matrix of (Σ) , Matrix Ø is generally obtained by an orthonormal convention (Orthonormal matrix is one in which ǿ =$ø^{-1}$ or ǿø=1). Variable z can be defined as :

$$z = \emptyset\, x \tag{4}$$

In this equation the convention matrix ø is composed of special vectors of matrix Σ.

$$\sum = cov(x) = E\{(x-m)(x-m)'\} \tag{5}$$

$$\emptyset \sum \emptyset = \lambda \tag{6}$$

In above equation, λ is a diagonal matrix whose main diagonal elements are eigenvalues corresponding to eigenvectors which are arranged in an ascending or descending order. Since some especial values are very small, space dimension reduction is justifiable. Therefore, matrix ǿ is defined in such a way that it reduces feature space dimension to (n´ < n) n´ and the new variable z´ is defined as :

$$z' = \left(\emptyset'\right)' x \tag{7}$$

In order to define ǿ we look at special values which are arranged in ascending or descending order respectively. Wherever a big step of special values gets small, eigenvectors are omitted. Therefore, as seen in figure 2, the directions of less data distribution are omitted.

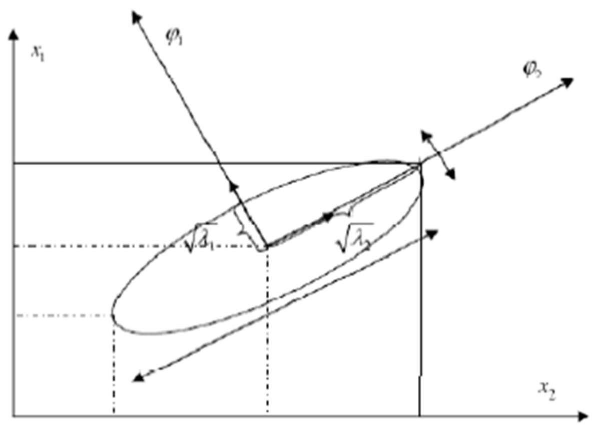

Fig.2: feature space reduction by PCA

In figure 2 x_1 and x_2 are the main directions of input vector, but $ø_1$ and $ø_2$ are distribution directions. As it is seen, distribution direction $ø_1$ is much less than $ø_2$. Therefore, space dimension is reduced by omitting $ø_1$. With regard to the stated points, it is concluded that PCA method considers distribution measure [3].

4. Probabilistic Neural Network

Probabilistic neural network was first introduced by Specht [4]. PNN is in fact a training tool or network which is used in subjects such as pattern recognition, nonlinear planning, estimation of each class data probability, and similarity proportion. Since the function of data density is not identified, nonparametric methods are used for estimating such density functions.

The structure of PNN is like feed-forward neural networks. The difference is that a PNN consists of four layers including input layer, pattern layer, summation layer, and output (decision) layer. Figure 3 shows an example of PNN with binary pattern [5].

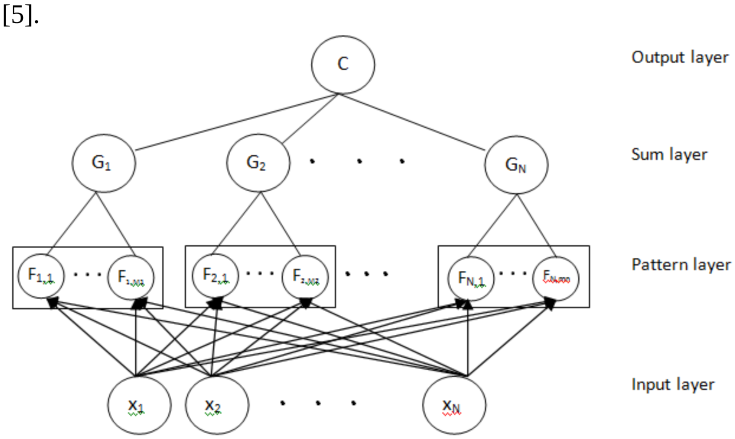

Fig.3: PNN architecture

An input vector $X = (x_1, x_2, \dots . x_n) \subset R^n$ is used for n input neuron and is given to the pattern layer. Pattern layer neurons are divided to k groups each one of which is related to one class. The output of i^{th} neuron of the patter layer in k^{th} group is calculated through the following equation.

$$F_{k,i}(X) = \frac{1}{(2\pi\sigma)^{n/2}} \exp(-\frac{\|X - X_{k,i}\|^2}{2\sigma^2} \tag{8}$$

$R^n \subset X_{k,i}$ is introduced as the core center and σ as smoothing parameter. Summation layer calculates the probabilistic function estimate of conditional class by means of the following equation which is a combination of a priori calculated densities.

$$G_k(X) = \sum_{i=1}^{M_k} w_{ki} F_{k,i}(X), K \in \{1, \dots \dots k\} \tag{9}$$

In equation (9), $M_k$is the number of neurons of class k pattern and w_{ki} is positive factor for which the following equation is computed:

$$\sum_{i=1}^{M_k} w_{ki} = 1 \tag{10}$$

Finally, the vector of x pattern is given to the class which has the highest rate of function of Gk [5].

$$C(X) = \arg \underset{1\le k\le k}{\text{Max}} (G_k) \tag{11}$$

Another structure of probabilistic neural network is shown in figure (4). In this structure PNN could be considered as a radial layer and a comparative layer.

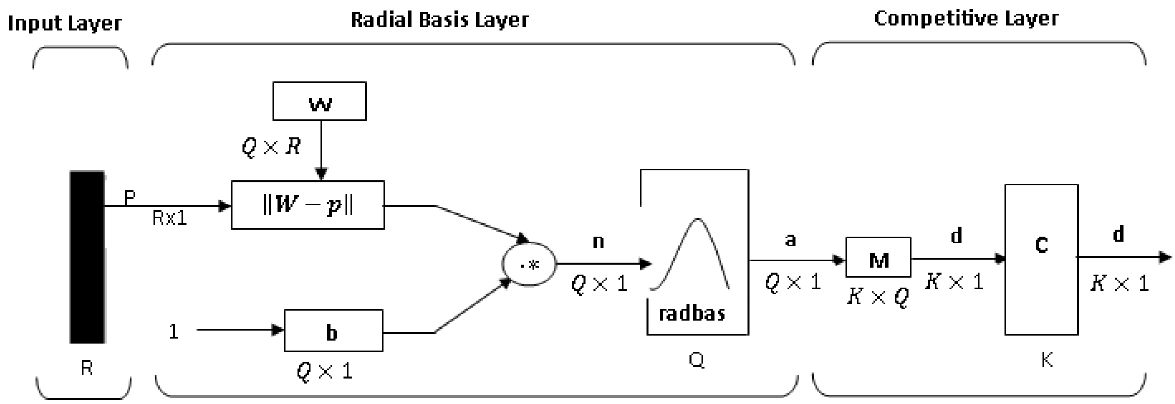

Fig.4: PNN which contains two radial and comparative layers [6]

5. Classifiers criteria

The first criterion which is used for choosing each one of classifier is the rate of classifier separation. However, there are other different criteria for comparing classifiers which are used and defined based on their application. Before defining such criteria, it is necessary to offer some new definitions and then describe the desired criteria in relation to them.

- True positive: person is suffering from thyroid and the classifier has diagnosed the thyroid.
- True negative: Person is healthy and the classifier has diagnosed his health.
- Wrong positive: person is suffering from thyroid, but the classifier has diagnosed as healthy.
- Wrong negative: person is healthy, but the classifier had diagnosed he is suffering from thyroid.

Based on these definitions some criteria are defined for comparing diagnostic systems as the following.

- Sensitivity: a criterion of classifier's capability of true diagnosis of thyroid.

$$SN = \frac{TP}{TP+FN} \times 100 \ \% \tag{12}$$

- Accuracy: a criterion of classifier's capability of true diagnosis of health.

$$SP = \frac{TN}{TN+FP} \times 100\% \tag{13}$$

It should be noted that each one of these two criteria is emphasized for a certain weight which depends on their application. For example, the average rate of diagnosis is defined as follows:

$$\frac{SN+SP}{2} \times 100\% \tag{14}$$

To implement the desired networks, 75% data is used for training and 25% data is used for testing. The network accuracy for the test data is obtained by the following equation:

$$Classification\,accuracy\,(N) = \frac{\sum_{i=1}^{|N|} calculate(n_i)}{|N|} \tag{15}$$

In above equation, N is the total samples which are considered for the test and $n_i \in N$. If n_i is classified correctly then calculate $\{n_i\}$ =1; otherwise, this value will be zero.

6. Simulation Results

The best results obtained from implementing three classes of data in PNN by means of PCA algorithm for choosing the feature are shown in Table 1.

Accuracy diagram of PNN classification training and testing for three classes of data and for PCA structure are also shown in figure 5 and 6 in which vertical axis is PNN classifier and horizontal axis represents PCK.

Table 1: best results of exercising three classes of data in PNN by means of PCA algorithm for selecting features.

Data	The accuracy of Education	The accuracy of test	Percent sensitivity	Percent accuracy
First class	98/77	98/11	100	97/56
Second class	70/72	87/36	33	92/63
Third class	86/98	82/68	87/59	80/81

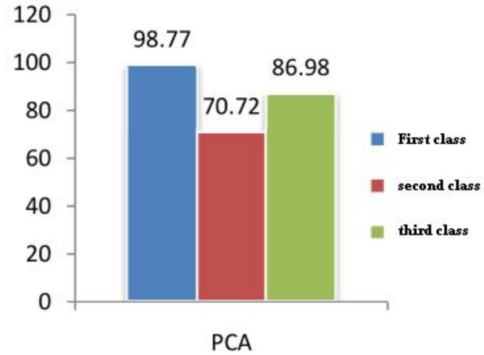

Fig.5: diagram of testing accuracy of classifier PNN (vertical axis) for all three classes of data and PCA structure

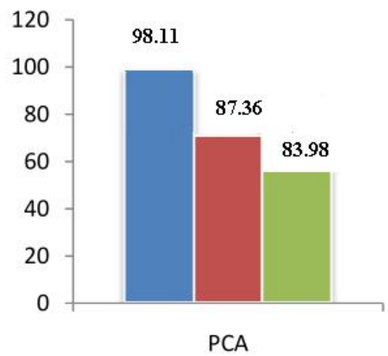

Fig.6: diagram of training accuracy of classifier PNN (vertical axis) for all three classes of data and PCA structure

7. Thyroid datasets

In order to do this research, two classes of data have been used which are shown in VCI site. (University of California, Irvine)[7-9].

The features are as the following:[10]

Table2: The features of data

Main owner	**Danny Coonans**
Contributor	Stephan Aeberhard
Number of samples	215
Number of features	5

In the following table, all constant features are mentioned.

Table3: data features

1	**attraction test(T3RU)**
2	Total level of tiroxin which is measured by Isotopic movement (TT4)
3	Total level of triprotironin measured by radioimnoxi
4	TSH of basic level of Thyroid enhancer measured by radioimnoxi
5	Maximum absolute difference of basic level of TSH and TSH measured by after injecting 200 mg Tirotropin releaser hormone.

Each class listeners are as follows:

Table4: Each class listeners

1	**First class: Normal**	**150 samples**
2	Second class : high efficiency	35 samples
3	Third class : low efficiency	30 samples

8. Conclusion

Since it is very important to diagnose hormonal diseases such as thyroid disease which has a serious effect on body organs, designing a smart network with high diagnostic power could greatly help the doctor recognize the disease on time and prescribe medicine and prevent later problems. In this article PCA approach was used to extract features and PNN was as classifier. The numbers in table shows that class work has been done correctly.

REFRENCES

[1] S. Haykin, "Neural Networks - A Comprehensive Foundation," Macmillan College Publishing Company Inc, New York, 1999.

[2] J. S. R. Jang, C. T. Sun and E. Mizutani, "Neuro-Fuzzy and Soft Computing- A Computational Approach to Learning and Machine Intelligence", Prentice-Hall Inc, ISBN 0-13-261066-3, 1997.

[3] S. Theodoridis, "Pattern Recognition," Second edition Academic Press an imprint of Elsevier Science, ISBN 0 12-685875-6, 2003.

[4] D. F. Specht, "Probabilistic neural networks," Neural Networks, 1(3), pp.109-118, 1990.

[5] V. L. Georgiou, N. G. Pavlidis and K. E. Parsopoulos, "Optimizing the Performance of Probabilistic Neural Networks in a Bioinformatics Task," Department of Mathematics, University of Patras Artificial Intelligence Research center, 2003

[6] http://www.mathworks.com/

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

[8] K. Polat, S. Sahan and S. Gunes, "A novel hybrid method based on artificial immune recognition system (AIRS) with fuzzy weighted preprocessing for thyroid disease diagnosis," Expert Systems with Applications, vol. 32, 2007, pp. 1141-1147.

[9] UCI Repository of Machine Learning Databases, University of California at Irvine, Department of Computer Science. Available: http://www.ics.uci.edu/~mlearn/databases/thyroid-disease/new- thyroid.data (Accessed: 10 May 2010)

[10] ftp://ftp.ics.uci.edu/pub/machine-learning-databases/thyroid-disease