A Comparison of Classification Methods on Diagnosis of Thyroid Diseases

Conference Paper · May 2015

DOI: 10.1109/ISITIA.2015.7219959

3 authors:

I Made Dendi May Sanjaya
Ganesha University of Education

Hanung Adi Nugroho
Universitas Gadjah Mada

Noor Akhmad Setiawan
Universitas Gadjah Mada

A Comparison of Classification Methods on Diagnosis of Thyroid Diseases

I Md. Dendi Maysanjaya[1], Hanung Adi Nugroho[2], Noor Akhmad Setiawan[3]
Department of Electrical Engineering and Information Technology
Faculty of Engineering
Universitas Gadjah Mada
[1]imd.dendi.mti13@mail.ugm.ac.id, [2]adinugroho@ugm.ac.id, [3]noorwewe@ugm.ac.id

Abstract—**Thyroid gland is one of the endocrine glands in the human body which produces thyroid hormone. This gland actively produces two kinds of hormone, namely thyroxine (T4) and triiodothyronine (T3). These hormones aim to produce protein, govern body metabolism, as well as to control body temperature circulation. Either excess or lack of these hormones will disturb those activities. The condition of excessive hormones is called hyperthyroid while the condition of lacking hormones is called hypothyroid. The major factor that influences the volume of the produced T3 and T4 hormones is iodine, because it is the main building-block substance of those hormones. The imbalance condition of this substance prevents thyroid to work properly. To identify the type of thyroid (normal, hypothyroid, hyperthyroid), WEKA (Waikato Environment for Knowledge Analysis) machine learning software is utilized. The thyroid dataset is taken from UCI (University of California – Irvine) machine learning repository as many as 215 instances. The test result shows that among six different methods available in WEKA, MLP (Multilayer Perceptron) method gives result with the highest accuracy, up to 96.74%, while BPA (Back Propagation Algorithm) methods produces result with the lowest accuracy, of 69.77%.**

Keywords—thyroid type; ANN methods comparison, WEKA, UCI machine learning repository

I. Introduction

Thyroid gland is one of the endocrine glands producing thyroid hormone. This gland is located in the neck, exactly under the jaw. The function of thyroid hormone is to help expediting human body metabolism, burn energy, process protein, and control body's sensitivity to other hormonal gland [1][2].

Thyroid gland generally produces two types of active hormones, namely thyroxine (T4) and triiodothyronine (T3) [1][2][3]. These hormones are important in protein production, body temperature circulation, as well as energy production and circulation to every part of the body. Iodine is the main building-block substance of these two thyroid hormones (T3 and T4). The imbalance amount of iodine will prevent the thyroid gland to work properly, which further causes the body's weight fluctuation, makes the heart palpitates, as well as disturbs motoric nervous coordination, body's sensitivity to temperature, and even fertility [1][3][4].

The lack amount of iodine causes hypothyroid while the excessive amount of thyroid hormone released to the body leads to hyperthyroid. Hypothyroid is condition when the amount of produced T3 and T4 hormones are less than the amount needed by the body. Hypothyroid will decelerate body metabolism, increase body's weight drastically, as well as cause joint pain and drowsiness. Hyperthyroid is the inverse of hypothyroid, when the amount of produced hormones exceed the amount needed by the body. This will accelerate body metabolism, decrease body's weight, as well as cause insomnia and unstable emotion [4].

Along with technology development, the problem on thyroid gland can be solved well nowadays. Various methods have been developed to help diagnosing disease symptom on the thyroid gland, such as artificial neural network, fuzzy logic, neuro fuzzy, and so on.

This work aims to compare the level of accuracy among several ANN methods that are used to classify the type of thyroid gland into three classes, namely normal, hyperthyroid, and hypothyroid class, by means of WEKA application.

II. Classification Methods

The Artificial Neural Network (ANN) is one variety of the supervised learning method, which can be used to build an expert system. The ANN is formed by two kind's process, namely training process for the knowledge-based system, and testing process to assess the accuracy of the expert system. The purposed of training process in the ANN is for adjust the strength of connection among the related nodes [1][5][6].

In this research, the test is conducted on 6 types of the ANN methods to classify thyroid gland disease and compare the level of accuracy among those methods. The methods are radial based function (RBF), learning vector quantization (LVQ), multilayer perceptron (MLP), back propagation algorithm (BPA), artificial immune recognition system (AIRS), and perceptron. Here some brief description of each the ANN methods.

A. Radial Based Function (RBF)

RBF has three layers, namely input layer, hidden layer, and output layer. The training stage of RBF starts in the hidden layer by means of supervised learning algorithm, then continues in the output layer using supervised learning algorithm. That is the reason why RBF is among hybrid supervised-unsupervised topology. The weight will be usually adjusted to small value in the network during the training process. Each input unit *Xi* receives input signal and passes it to the next hidden layer known as radial center (*c*). This center is determined based on the input vectors set. RBF will give good result in case the input data is similar to training data, otherwise produces bad result when the input data is not similar to training data [1][6][7].

In the hidden layer, the equation generally used is Gaussian function as written in equation 1 below,

$$\varphi_k\left(\left\|X-C_k\right\|\right)=e^{\left(\frac{1}{2\sigma^2}\left\|X-C_k\right\|\right)} \quad (1)$$

where C_k denotes the center of radial based function of k number of neuron.

B. Learning Vector Quantization (LVQ)

LVQ is hybrid network method which combines supervised learning and unsupervised learning and generally consists of three layers, namely input layer, kohonen layer, and output layer. The input layer and the kohonen layer are fully connected by vector line. The kohonen layer will be divided into several neuron groups based on the class number. To determine the vector distance, the Euclidean distance function is used. The neuron which has the smallest distance in the kohonen layer win and will move toward the input layer, while the other neurons will adjust their weight [6][7][8].

To determine the weight value in the hidden layer, equation 2 is used when the input and the weight have the same class, while equation 3 is utilized when the input and the weight have different class,

$$W_j^{'}=W_j+\alpha\left[x-W_j\right] \quad (2)$$

$$W_j^{'}=W_j-\alpha\left[x-W_j\right] \quad (3)$$

where $W_j^{'}$ denotes the new weight, W_j denotes the old weight, α denotes the learning rate, and x denotes the input.

C. Multilayer Perceptron (MLP)

MLP uses gradient descent to determine the new weight and bias, as well as adjust the network weight quickly. MLP is multilayer neural network, which consists of at least 3 hidden layers and uses back-propagation as learning mechanism. MLP is widely used because of its ability to support parallel implementation, capacity for generalization, fault tolerance, and efficient learning algorithm. On the other hand, serious problem of MLP arises when it is applied to image processing. The problem occurs when the topology structure of input pattern is neglected and the input of MLP is treated as one dimensional vector [9].

D. Back Propagation Algorithm (BPA)

BPA is multi-layered feed-forward neural network, which is built based on the back propagation algorithm. The algorithm receives input vector and passes it into network, makes comparison as needed, produces output based on the input signal, adjust the weight with the derivative of error based on the weight of learning rate [1][7].

E. Artificial Immune Recognition System (AIRS)

AIRS is supervised learning algorithm inspired by human body's immune system. The mechanisms used in this algorithm are resource competition, clonal selection, affinity maturation, and memory cell formation. The feature vector, used for training and testing is called antigen while unit system is called B cells. B cells is represented by artificial recognition balls (ARBs). These ARBs will compete with each other as a resource number. The higher the closeness between ARBs and antigen during the training process, the weight value will be increased. The value of ARBs can be determined by equation 4 below,

$$S_i=\frac{\sum_{j=1}^{|ARB_i|} arb_j\cdot stim}{\left|ARB_i\right|}, arb_j\in ARB_i \quad (4)$$

where i = 1 … nc, S = {S1, S2, …, Snc}, |ARBi| is the number of ARB from i-th class, and arb_j.stim is stimulation level of the j-th ARB from i-th class. This stimulation can be determined by equation 5 below,

$$Stimulation\,(x,y)=\begin{cases} affinity\,(x,y), & if\ class\ of\ x=class\ of\ y \\ 1-affinity\,(x,y), & otherwise \end{cases} \quad (5)$$

where the value of *affinity(x,y)* is given by equation 6,

$$affinity\,(x,y)=1-\sqrt{\sum_{i=1}^{n}\left(x_i-y_i\right)^2} \quad (6)$$

After the antigen of training process has been done, the memory cell used as antigen in the testing process is formed. This algorithm is generally divided into four stages: initialization, memory cell and ARBs generation identification, resources competition and memory cell candidate development, as well as memory cell introduction [10].

F. Perceptron

Perceptron is the simplest artificial neural network used for classifying linearly separable patterns. Perceptron uses hard limit as its activation function or transfer function (f). This function has two kinds of output signal, namely 0 and 1 [7]. The architecture of this perceptron is depicted in Figure 1 below.

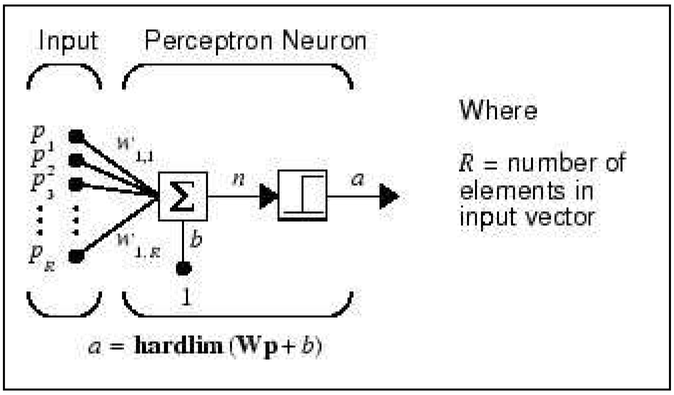

Fig. 1. A Perceptron Neuron Model [11]

The value of *net* and *f(net)* can be determined by equation 7 and 8 respectively.

$$net = \sum_i x_i \cdot w_i + b \tag{7}$$

$$f(net) = \begin{cases} 1, & if\ net > \theta \\ 0, & if\ -\theta \le net \le \theta \\ -1, & if\ net < -\theta \end{cases} \tag{8}$$

where *x* denotes input, *w* denotes weight value, and *b* denotes bias value.

III. WEKA Framework

Waikato Environment for Knowledge Analysis (WEKA) framework is framework developed in University of Waikato [12]. WEKA is open source, specifically designed for pattern recognition and developed by means of JAVA. There are many advantages provided by WEKA such that the user can add various new pattern recognition methods. WEKA support several function of data mining, such as data preprocessing, classification, clustering, association, regression, feature selection, and visualization. The data point is described by a fixed number of attributes, namely numeric, nominal, normally, and other attribute type.

IV. UCI Machine Learning Repository

The University of California – Irvine (UCI) Machine Learning Repository is like a bank of data sets, database, or domain theory for many case study that are used by the machine learning community for the empirical analysis of machine learning algorithms. It is used by anybody as a primary source of machine learning data sets [13].

V. Measure for Performance Evalution

A. Classification Accuracy

In this study, the level of accuracy can be determined by equation (9) as follows [10],

$$Accuracy\ (T) = \frac{\sum_{i=1}^{|T|} assess\ (t_i)}{|T|},\ t_i \in T$$

$$assess\ (t) = \begin{cases} 1, & if\ classify\ (t) = t{\cdot}c \\ 0, & otherwise \end{cases} \tag{9}$$

where *T* denotes dataset that will be classified, $t \in T$, *t.c* denotes the class of item *t*, and *classify(t)* denotes the classification result of *t* resulted by each classification method.

B. k-Fold Cross Validation

To produce better classification result, k-fold cross-validation evaluation method is utilized. This method has several advantages, such as minimizing bias produced during the training process by using random sampling [10]. This method starts by dividing all data randomly into *k* groups with each group is assumed to have the same number of members. The classification algorithm has to execute the training and testing *k* times. In each process, one fold is entered as data test and the other folds serve as training data. Each process will give different results. These values are then averaged and the mean value is used as the accuracy value of the corresponding algorithm.

VI. Result and Discussion

In this work, the thyroid dataset is taken from UCI machine learning repository [14]. The number of sample is 215 (150 first class samples, 35 second class samples, and 30 third class samples) with 5 features (T3-resin uptake test, total serum thyroxin, total serum triiodothyronine, TSH, and maximum absolute difference of TSH), and it consists of three classes (normal, hypothyroid, and hyperthyroid) [2][14].

Table I shows the test result of each classification method as explained in part II (*Classification Method*) using 10-folds cross-validation evaluation method. Table II shows the number of instance classified based on its class as well as accuracy level, precision, recall, and F-Measure value of each ANN methods.

TABLE I. The Comparison Result of Classification Methods

Methods	Normal	Hyper	Hypo	Correctly	Incorrectly
RBF	146	33	26	205	10
LVQ	150	28	23	201	14
MLP	146	33	29	208	7
BPA	150	0	0	150	65
AIRS	147	31	23	201	14
Perceptron	147	26	23	196	19

TABLE II. The Classification Result

Methods	Accuracy	Precision	Recall	F-Measure
RBF	95.35%	0.953	0.953	0.953
LVQ	93.50%	0.94	0.935	0.932
MLP	96.74%	0.968	0.967	0.968
BPA	69.77%	0.487	0.698	0.573
AIRS	93.50%	0.935	0.935	0.933
Perceptron	91.16%	0.932	0.925	0.921

Table I shows the number of instance classified based on its class. For normal class, LVQ and BPA give the best classification results since both of them are able to recognize all instances of normal class (150 instances in total). The classification result with the highest instances for hyperthyroid class is produced by RBF and MLP, as many as 33 of 35 instances, while BPA method is not able to classify the hyperthyroid class at all. All instances in hyperthyroid class are classified as normal class by BPA. The same condition applies in the hypothyroid class, in which all 30 instances are classified as normal class. Otherwise, MLP is able to classify the highest number of instances for hypothyroid class among the six other methods of ANN, as many as 29 of 30 instances.

Table II show that MLP has the highest level of accuracy to classify three types of thyroid disease, namely 96.74%, 1.39% more accurate that the accuracy level of RBF, namely 95.35%. This is caused by the fact that MLP is able to classify instances 3 more than the number of instance classified by RBF. On the other hand, BPA has the lowest level of accuracy among six ANN methods, namely 69.77%. This is caused by the fact that BPA is only able to classify the normal class while the other classes are not able to be classified successfully since all of them are classified as normal instead. Besides, this also can be caused by local minima case, or the presence of extreme value in the dataset. This extreme value affects the classification result produced by BPA.

MLP methods produce the most accurate result (for thyroid case classification) because of its advantages, such as parallel implementation compatibility, generalization of capacity, fault tolerance, and efficient learning algorithm. On the other hand, it has drawbacks when it comes to digital image processing application. The problem occurs when the topological structure of input pattern is neglected and the input of MLP is treated as one-dimensional vector.

RBF gives better classification result when the input data has similarity with the training data. In contrary, when the input data is not similar to training data, it produces bad result. It occurs because the formation of radial center during the training process will be determined based on the input vector set.

Based on the result of Table II, it can be seen that the MLP has the highest precision, recall, and F-Measure value among the six other methods of ANN. Precision values represent the ratio of the number of applicable documents found with the total number of documents found by the system. Recall values refer to the ratio of the number of retrieved applicable documents with the total number of documents in a group of documents which are considered as relevant, and F-measure is the value obtained from the measurement of precision and recall between the classes after clustering with the actual classes contained in the input data. The smaller the value of F-measure is, the worse the quality of a cluster and vice versa.

VII. Conclusion

Based on the results obtained from WEKA framework, it can be concluded that among six ANN methods, MLP and RBF have the highest accuracy (96.74% and 95.35%), while BPA has the lowest accuracy (69.77%). BPA's low accuracy is caused by the problem on the dataset, which has local minima case or the presence of extreme value. On the other hand, MLP have the highest accuracy because of their better fault tolerance.

On the other hand, the classification error is caused by the constraint of each classifier to handle the feature of dataset, such as BPA cannot give the best result if the dataset has an extreme value; RBF produces bad result when the testing data does not have a similar feature to training data; and MLP has a problem when topological structure of input pattern is neglected. In the experiment, BPA has the lowest accuracy value because the dataset contained an extreme value.

References

[1] A. Shukla, R. Tiwari, P. Kaur, and R. R. Janghel, "Diagnosis of Thyroid Disorders using Artificial Neural Networks," in *Advance Computing Conference, 2009. IACC 2009. IEEE International*, 2009, pp. 1016–1020.

[2] L. Ozyilmaz and T. Yildirim, "Diagnosis of thyroid disease using artificial neural network methods," in *Proceedings of the 9th International Conference on Neural Information Processing, 2002. ICONIP '02*, 2002, vol. 4, pp. 2033–2036 vol.4.

[3] F. Saiti, A. A. Naini, M. A. Shoorehdeli, and M. Teshnehlab, "Thyroid Disease Diagnosis Based on Genetic Algorithms Using PNN and SVM," in *3rd International Conference on Bioinformatics and Biomedical Engineering , 2009. ICBBE 2009*, 2009, pp. 1–4.

[4] Asep, "Perbedaan Hipotiroid dan Hipertiroid," *Obat Herbal 21*. [Online]. Available: http://obatherbal21.com/perbedaan-hipertiroid-dan-hipotiroid.html. [Accessed: 06-Dec-2014].

[5] C.-Y. Chang, P. Chung, Y.-C. Hong, and C.-H. Tseng, "A Neural Network for Thyroid Segmentation and Volume Estimation in CT Images," *IEEE Comput. Intell. Mag.*, vol. 6, no. 4, pp. 43–55, 2011.

[6] I. M. D. Maysanjaya, H. A. Nugroho, and N. A. Setiawan, "The classification of fetus gender on ultrasound images using learning vector quantization (LVQ)," in *2014 Makassar International Conference on Electrical Engineering and Informatics (MICEEI)*, 2014, pp. 150–155.

[7] M. Rouhani and K. Mansouri, "Comparison of Several ANN Architectures on the Thyroid Diseases Grades Diagnosis," in *International Association of Computer Science and Information Technology - Spring Conference, 2009. IACSITSC '09*, 2009, pp. 526–528.

[8] M. Swiercz, J. Kochanowicz, J. Weigele, R. Hurst, D. S. Liebeskind, Z. Mariak, E. R. Melhem, and J. Krejza, "Learning Vector Quantization Neural Networks Improve Accuracy of Transcranial Color-coded Duplex Sonography in Detection of Middle Cerebral Artery Spasm," *Neuroinformatics*, vol. 6(4), pp. 279–290, Aug. 2008.

[9] I. S. Isa, N. A. Fauzi, J. M. Sharif, R. Baharudin, and M. H. Abbas, "Comparisons of MLP transfer functions for different classification classes," in *2012 IEEE International Conference on Control System, Computing and Engineering (ICCSCE)*, 2012, pp. 110–114.

[10] K. Polat, S. Şahan, and S. Güneş, "A novel hybrid method based on artificial immune recognition system (AIRS) with fuzzy weighted pre-processing for thyroid disease diagnosis," *Expert Syst. Appl.*, vol. 32, no. 4, pp. 1141–1147, May 2007.

[11] The MathWorks Inc., "Perceptron Neural Networks," *MathWorks*. [Online]. Available: http://www.mathworks.com/help/nnet/ug/perceptron-neural-networks.html. [Accessed: 03-Apr-2015].

[12] M. Hall, E. Frank, G. Holmes, B. Pfahringer, P. Reutemann, and I. H. Witten, "The WEKA Data Mining Software: An Update," *SIGKDD Explor.*, vol. 11, no. 1, 2009.

[13] Center for Machine Learning and Intelligent System, UC-Irvine, "Machine Learning Data Repository [UCI]," *Databib*. [Online]. Available: http://databib.org/repository/654. [Accessed: 04-Apr-2015].

[14] "Thyroid Disease Data Set," *UCI Machine Learning Repository*. [Online]. Available: https://archive.ics.uci.edu/ml/datasets/Thyroid+Disease.