Expert Systems with Applications 38 (2011) 146–150

Contents lists available at ScienceDirect

Expert Systems with Applications

journal homepage: www.elsevier.com/locate/eswa

An expert system based on Generalized Discriminant Analysis and Wavelet Support Vector Machine for diagnosis of thyroid diseases

Esin Dogantekin [a], Akif Dogantekin [b], Derya Avci [c,*]

[a] Firat University, Firat Medicine Center, Department of Microbiology and Clinical Microbiology, 23119 Elazig, Turkey
[b] Firat University, Firat Medicine Center, Department of Internal Diseases, 23119 Elazig, Turkey
[c] Bahcelievler Primary School, Elazig, Turkey

A R T I C L E I N F O

Keywords:
Generalized Discriminant Analysis (GDA)
Wavelet Support Vector Machine (WSVM) classifier
Thyroid gland
Expert diagnosis system
Classification accuracy
Confusion matrix

A B S T R A C T

Nowadays, there are many persons, which suffer from thyroid diseases. Therefore, the correct diagnosis of these diseases are very important topic. In this study, a Generalized Discriminant Analysis and Wavelet Support Vector Machine System (GDA_WSVM) method for diagnosis of thyroid diseases is presented. This proposed system includes three phases. These are feature extraction – feature reduction phase, classification phase, and test of GDA_WSVM for correct diagnosis of thyroid diseases phase, respectively. The correct diagnosis performance of this GDA_WSVM expert system for diagnosis of thyroid diseases is estimated by using classification accuracy and confusion matrix methods, respectively. The classification accuracy of this expert system for diagnosis of thyroid diseases was obtained about 91.86%.

1. Introduction

The thyroid gland is one of the endocrine glands (Özyılmaz & ıldırım, 2002; Thyroid gland, 2007). This forms hormones to regulate physiological functions in body. One of these hormones is thyroid hormone. It regulates the rate at which body carries on its necessary functions.

The thyroid gland is placed in the middle of the lower neck, below the larynx and just above clavicles (Thyroid gland, 2007). The thyroid gland has a shape, which is bow tie. In addition to, it has two lobes. These are right lobe and left lobe.

The thyroid gland has many diseases. The most of these are goiters, thyroid cancer, solitary thyroid nodules, hyperthyroidism, hypothyroidism, thyroiditis, etc. These disorders can be explained as below.

The hypothyroidism is too little thyroid hormone. It is a common problem. Hypothyroidism can even be associated with pregnancy. The diagnosis and treatment for all types of hypothyroidism is usually straightforward. The goiter is a dramatic enlargement of the thyroid gland. Goiters are often removed due to cosmetic reasons. Moreover, these compress other vital structures of the neck including the trachea and the esophagus making breathing and swallowing difficult. The thyroid nodules can take on characteristics of malignancy. Therefore, these require biopsy or surgical excision. In addition to, these contain risks of radiation exposure. The thyroid cancer is a fairly common malignancy. Therefore the diagnosis of this disease is very difficult. The hyperthyroidism is too much thyroid hormone. The radioactive iodine, anti-thyroid drugs, or surgery are common methods used for treating a hyperthyroid patient. The thyroiditis is an inflammatory status for the thyroid gland. This can give with a number of symptoms such as fever and pain, but it can also give as subtle findings of hypo or hyper-thyroidism.

In diagnosis of thyroid diseases literature, many studies have been realized. The some methods used these previous studies can be ordered as below:

- Multi Layer Perception with Back-Propagation (MLP with BP) (Özyılmaz & ıldırım, 2002),
- Radial Basis Function (RBF) (Özyılmaz & ıldırım, 2002),
- Adaptive Conic Section Function Neural Network (CSFNN) (Özyılmaz & ıldırım, 2002),
- Learning Vector Quantizer (LVQ) (Serpen, Jiang, & Allred, 1997),
- Probabilistic Potential Function Neural Network (PPFNN) (Serpen et al., 1997),
- Linear Discriminant Analysis (LDA) (Pasi, 2004),
- C4.5 with default learning parameters (C4.5-1) (Pasi, 2004),
- C4.5 with parameter *c* equal to 5 (C4.5-2) (Pasi, 2004),
- C4.5 with parameter *c* equal to 95 (C4.5-3) (Pasi, 2004),
- DIMLP with two hidden layers and default learning parameters (DIMLP) (Pasi, 2004),
- Expert System for Thyroid Disease Diagnosis with Neuro Fuzzy Classification (ESTDD with NEFCLASS-J) (Keleş & Keleş, 2008).

In this study, a Generalized Discriminant Analysis and Wavelet Support Vector Machine System (GDA_WSVM) method for

* Corresponding author.
E-mail address: derya2344@hotmail.com (D. Avci).

0957-4174/$ - see front matter © 2010 Elsevier Ltd. All rights reserved.
doi:10.1016/j.eswa.2010.06.029

diagnosis of thyroid diseases is introduced. This proposed system includes three phases. These are feature extraction – feature reduction phase, classification phase, and test of GDA_WSVM for correct diagnosis of thyroid diseases phase, respectively. The correct diagnosis performance of this GDA_WSVM expert system for diagnosis of thyroid diseases is estimated by using classification accuracy and confusion matrix methods, respectively. In addition to, the results of this GDA_WSVM expert system with these previous methods used in diagnosis of thyroid disease mentioned above will be compared in Section 4.

The paper is organized as follows: In Section 2, Generalized Discriminant Analysis (GDA) is introduced for the feature reduction on diagnosis of thyroid diseases. In Section 3, The Wavelet Support Vector Machine classifier is explained. In Section 4, the implementing of GDA–WSVM system for thyroid gland is expressed. In Section 4.3, The obtained results are given. In Section 5, discussion and conclusions are presented.

2. Generalized Discriminant Analysis for feature reduction

The purpose of Generalized Discriminant Analysis (GDA) is to maximize the quotient between the inter-classes inertia and the intra-classes inertia in a mapped feature space (Baudat & Anouar, 2000). A set of training patterns from the C classes can be given as below, if there is a S-class problem and letting L_s be the number of samples in class s:

$$L = \sum_{s=1}^{S} L_s x_{si}, \quad s = 1,2,\ldots,S \quad i = 1,2,\ldots,L_s \tag{1}$$

$\phi : R^L \rightarrow T$, which is a nonlinear mapping the set of training samples in the mapped feature space, can be represented as $\{\phi(x_{si}), s = 1,2,\ldots,S; i = 1,2,\ldots,L_s\}$. The S_b and S_w of the training set can be calculated as below.

$$S_w = (1/S)\sum_{s=1}^{S}\frac{1}{L_s}\sum_{i=1}^{L_s}\phi(x_{si})\phi(x_{si})^T \tag{2}$$

$$S_b = (1/S)\sum_{s=1}^{S}(\mu_s - \mu)(\mu_s - \mu)^T \tag{3}$$

GDA calculates the eigenvalues $\lambda \geqslant 0$ and eigenvectors $m \in T \backslash \{0\}$ satisfying

$$\lambda S_W m = S_b m \tag{4}$$

here all solutions m lie in the span of $\phi(x_{11}),\ldots,\phi(x_{si}),\ldots$ and there exist coefficients r_{si} such that

$$m = \sum_{s=1}^{S}\sum_{i=1}^{L_s} r_{si}\phi(x_{si}) \tag{5}$$

the dot product of a sample i from class p and the other sample j from class q in the feature space is performed by using kernel techniques, which are shown as $(A_{zd})_{pq}$, such as radial basis kernel and polynomial kernel as below.

$$(A_{zd})_{pq} = \phi(x_{pz})\cdot\phi(x_{qy}) = k(x_{pz}, x_{qd}) = e^{-|x_{pz}x_{qd}|2/r} \tag{6}$$

If B is a $N \times N$ matrix defined on the class elements by $(B_{pq})_{(p=1,\ldots,s),(q=1,\ldots\ldots,s)}$, here B_{pq} is a matrix composed of dot products between vectors from class p and q in feature space:

$$B_{pq} = (A_{zd})_{(z=1,\ldots,L_p),(d=1,\ldots,L_q)} \tag{7}$$

a $N \times N$ block diagonal matrix can be defined as below:

$$\mathbf{V} = (\mathrm{V}_s)_{s=1,\ldots,S} \tag{8}$$

there V_s is $L_s \times L_s$ a matrix with terms all equal to $1/L_s$.

The solution of Eq. (4) can be accomplished by substituting Eqs. (2), (3) and (5) into (4) and taking inner-product with vector $\phi(x_{zd})$ on both sides:

$$\lambda BBp = BVBp \tag{9}$$

there p represents a column vector with entries p_{si}, $s = 1,\ldots,S$, $i = 1,\ldots,L_s$. The eigenvectors of the matrix $(BB)^{-1}BVB$ are calculated. If these eigenvectors of $(BB)^{-1}BVB$ are found, the solution of **a** in Eq. (9) is completed. Moreover, the matrix B might not be reversible. The eigenvector is found **a** by first diagonalising matrix B (Baudat & Anouar, 2000). Firstly J important eigenvectors are calculated, a projection matrix can be structured as below:

$$G = [a_1 a_2 a_3 \cdots a_J] \tag{10}$$

The projection of x in the J-dimensional GDA space is given by

$$\mathbf{y} = A_{\mathbf{x}} G \tag{11}$$

here

$$\mathrm{A_x} = [A(x, x_{11})\cdots k(x, x_{\mathrm{si}})\cdots k(x, x_{\mathrm{SLs}})] \tag{12}$$

3. Wavelet Support Vector Machine classifier

The Support Vector Machines classifiers performs the classification task for two-class problems by using the separating hyperplane (Yao, Frasconi, & Pontil, 2001). This separating hyperplane is calculated with maximum distance to the closest data points of the training set. These closest data points are called as *Support Vectors*. These data points can be transformed to a High Dimensional Space (HDS) with a nonlinear transformation, if these data points are not linearly separable in the input space. This high dimensional space is named as feature space. These nonlinear transformations are represented with kernel functions $K(.,.)$. It describes an inner product in HDS. The data points in the feature space is divided by the optimal separating hyperplane estimated by using maximum distance to the closest data points of the training set mentioned as above.

$K(.,.)$ kernel function is restricted with conditions, which are named the Mercer condition (Campbell, Singer, Torres-Carrasquillo, & Reynolds, 2004). $K(.,.)$ kernel function can be given as below:

$$K(x, x') = \langle c(x)\cdot c(x')\rangle \tag{13}$$

Here, $c(x)$ is a mapping from low dimension input space to a HDS and x is input data points. Thus, decision function of a SVM is expressed as below:

$$f(x) = \sum_{k=1}^{G} a_k y_k K(x, x_k) + b_k \tag{14}$$

Here, G is the number of data points, the y_k are the target values, $y_k \in \{-1, 1\}$ is the class label of training point x_r. a_k is Lagrangian multipliers. They can be found by solving a quadratic programming problem with linear constraints (Yao et al., 2001; Vapnik, 1998). The value of a_k is nonzero only for support vectors.

The SVMs attempt to find an optimal hyperplane that correctly classifiers input data points by dividing to two classes, which are negative and positive (Frias-Martinez, Sanchez, & Velez, 2006).

The maximizing formulation is expressed as Eq. (15):

$$M = 2/\|w\| \tag{15}$$

Here, M is margin. This is the distance (*dis*) from the separating hyperplane to closest point for both classes of data points (Gunn, 1998; Kindermann, Paass, & Leopold, 2001). w and b_0 are the hyperplane parameters, which are weight and bias vectors.

For a data point, either $\langle w{\cdot}x_k\rangle + b_0 > 0$ or $\langle w{\cdot}x_k\rangle + b_0 < 0$ is valid. Here, y_k outputs are set of labeled input data points. If y_k is 1, the