Computer Methods and Programs in Biomedicine 147 (2017) 37–49

Contents lists available at ScienceDirect

Computer Methods and Programs in Biomedicine

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

Ultrasound-based differentiation of malignant and benign thyroid Nodules: An extreme learning machine approach

Jianfu Xia[a], Huiling Chen[b,*], Qiang Li[b], Minda Zhou[c], Limin Chen[c], Zhennao Cai[b], Yang Fang[a], Hong Zhou[a]

[a] *Department of General Surgery, The Dingli Clinical Institute of Wenzhou Medical University(Wenzhou Central Hospital), Wenzhou, Zhejiang, 325000, China*
[b] *College of Physics and Electronic Information, Wenzhou University, Wenzhou, Zhejiang, 325035, China*
[c] *Department of Ultrasound, The Dingli Clinical Institute of Wenzhou Medical University(Wenzhou Central Hospital), Wenzhou, Zhejiang, 325000, China*

ARTICLE INFO

Article history:
Received 29 September 2016
Revised 23 April 2017
Accepted 20 June 2017

Keywords:
Extreme learning machine
Feature selection
Medical diagnosis
Thyroid cancer
Sonographic features

ABSTRACT

Background and objectives: It is important to be able to accurately distinguish between benign and malignant thyroid nodules in order to make appropriate clinical decisions. The purpose of this study was to improve the effectiveness and efficiency for discriminating the malignant from benign thyroid cancers based on the Ultrasonography (US) features.

Methods: There were 114 benign nodules in 106 patients (82 women and 24 men) and 89 malignant nodules in 81 patients (69 women and 12 men) included in this study. The potential of extreme learning machine (ELM) has been explored for the first time to discriminate malignant and benign thyroid nodules based on the sonographic features in ultrasound images. The influence of two key parameters (the number of hidden neurons and type of activation function) on the performance of ELM was investigated. The relationship between feature subsets obtained by the feature selection method and the classification performance of ELM was also examined. A real-life dataset was used to evaluate the effectiveness of the proposed method in terms of classification accuracy, sensitivity, specificity, and area under the ROC (receiver operating characteristic) curve (AUC).

Results: The results demonstrate that there are significant differences between the malignant and benign thyroid nodules (p-value<0.01), the most discriminative features are echogenicity, calcification, margin, composition and shape. Compared with other methods, the proposed method not only has achieved very promising classification accuracy via 10-fold cross-validation (CV) scheme, but also greatly reduced the computational cost compared to other counterparts. The proposed ELM-based approach achieves 87.72% ACC, 0.8672 AUC, 78.89% sensitivity, and 94.55% specificity.

Conclusions: Based on the empirical analysis, the proposed ELM-based approach for thyroid cancer detection has promising potential in clinical use, and it can be of assistance as an optional tool for the clinicians.

1. Introduction

Thyroid nodules are very common endocrine tumors [1]. Many people in the general population have nodules, but they are asymptomatic. The estimated prevalence rate detected by palpation is only 3%–7% [2,3]; however, when ultrasound is used to detect thyroid nodules the prevalence rate jumps to 19% to 67% [4]. Most thyroid nodules are benign, but between 3% and 7% of cases are malignant [5]. In 2015, it was estimated that in the United States there were 62,450 new diagnoses of thyroid cancer, and about 1950 people died from the disease. (http://www.cancer.gov/cancertopics/types/thyroid).

In order to make a diagnosis, suspected nodules must be biopsied using fine needle aspiration (FNA) [6]. Although, in most cases, FNA biopsy can differentiate malignant nodules from benign nodules, it can result in physical and psychological discomfort because it is invasive. Further, FNA biopsy can produce results that are non-diagnostic or false-negatives [7,8]. Some studies have shown that FNA cytology is indeterminate in 10% to 30% of nodules [9]. Several imaging modalities have been used to identify the nature of ‘thyroid nodules’ clinical settings, including computed tomography (CT), magnetic resonance imaging (MRI), and positron emission tomography (PET) [10–12]. Recently, molecular profiling and gene expression have been explored to differentiate thyroid cancer from

---

* Corresponding author:.
E-mail address: chenhuiling.jlu@gmail.com (H. Chen).

http://dx.doi.org/10.1016/j.cmpb.2017.06.005

benign tumors [13–16]. However, the above-mentioned methods are either very expensive or induce patient anxiety. Ultrasonography (US) provides a safe and fast method for evaluating the thyroid gland. It is highly sensitive in the detection of thyroid nodules, and suspicious features can help differentiate benign from malignant nodules [17,18]. US is currently the most widely accepted imaging method for the preliminary evaluation of thyroid nodules [19–21]. US is the preferred choice for diagnosing thyroid nodules because it is efficient, noninvasive, inexpensive, nonradioactive, and widely available.

Since thyroid cancer diagnoses are mainly made through qualitative inspection by a doctor. This method is not always accurate, and can be rather subjective. There is a strong need to establish a more objective and accurate method or tool to help distinguish malignant from benign nodules in US images. The pattern recognition method could be a suitable tool for automatically detecting thyroid malignancy. Many Computer Aided Diagnostic (CAD) systems based on pattern recognition methods have been developed in recent years for detecting malignancy [21]. These CAD systems utilize two types of features: sonographic features that an endocrinologist observes in the US images and non-clinical features that quantify the visual differences in US images and can be automatically extracted by computer programs [21,22]. In the sonographic features, shape, margin, echogenicity, internal composition, calcification, peripheral halo, and vascularity on the color Doppler are often used to detect malignant nodules [23]. Many studies have been conducted on non-clinical features such as textural features [24–26] and wavelet features [22,24,27–29]. The classifier plays a very important role in the process of screening for benign and malignant thyroid nodules. A good classifier can greatly improve the accuracy and time efficiency of the entire CAD system. Researchers have proposed many pattern recognition methods, including artificial neural network (ANN) [23,30–32], support vector machines (SVM) [26,33], Bayesian classifier [34–38], k-nearest neighbor (KNN) [30,39], logistic regression [28,31], and directionality patterns [40].

Computational time and predictive accuracy are the two most important factors for a thyroid cancer diagnostic system. In this study, we explore a robust pattern recognition method with high generalization capability and a fast learning rate. This method uses an extreme learning machine (ELM) for differentiating between benign and malignant thyroid nodules based on sonographic features. ELM was originally introduced by Huang et al. [41], for single hidden layer feed-forward neural networks (SLFNs). It randomly chooses input weights and hidden biases, and the output weights are analytically determined using a Moore–Penrose (MP) generalized inverse. ELM has been applied in many fields; it performs especially well in medical diagnosis problems including in the diagnosis of erythemato-squamous diseases [42], paraquat-poisoning [43], hepatitis disease [44], and breast cancer [45]. Further, we explore the possibility of adopting feature selection in pre-processing before the ELM model is constructed, in the hope of identifying significant correlating factors to the diagnosis of thyroid malignancy. A 10-fold cross validation (CV) scheme is used to evaluate the effectiveness of the proposed method in terms of classification accuracy, sensitivity, specificity, and the area under the ROC (receiver operating characteristic) curve (AUC) on a real-life dataset from Wenzhou Central Hospital. The ELM-based approach achieves 87.72% ACC, 0.8672 AUC, 78.89% sensitivity, and 94.55% specificity.

The main contributions of this paper are: (1) The potential of ELM is explored through constructing an automatic diagnostic system for effective diagnosis of malignant thyroid nodules; (2) A detailed investigation of the impact of the feature selection on the classification performance of thyroid cancer diagnosis and an interesting discovery are presented; (3) The most relevant features are identified using the feature selection method.

The remainder of this paper is structured as follows: Section 2 offers a detailed description of the proposed method. In Section 3, the experimental design is presented. Section 4 provides the experimental results and discussion. Finally, conclusions and recommendations for future work are summarized in Section 5.

2. Proposed hybrid method for thyroid cancer detection

Fig. 1 shows a flowchart of the proposed ELM-based diagnosis method. In the proposed method, the discriminative features are first identified using the ReliefF feature selection method, and then the obtained feature subsets are evaluated one-by-one on the test set via a cross validation scheme. Finally, the features in the feature subset with the best classification accuracy are the most discriminative features.

2.1. Feature selection via ReliefF method

Feature selection plays an important role in data mining tasks. An efficient feature selection algorithm is an essential part of a machine learning approach. It is a useful tool to eliminate the irrelevant and trivial features that exist in the data. Feature selection helps improve the performance of classifier models, provides more effective models with simpler structures, and offers deeper insights into the underlying mechanism of the problem being investigated [46]. In this study, an incremental feature selection procedure based on the ReliefF method was constructed to find the most discriminative features that could differentiate malignant nodules from benign nodules. During this procedure, features were first ranked according to the score obtained by the ReliefF method. Then, one-by-one, the features were added to the ranked feature list from high to low rank. A total of 8 feature sets were constructed; the set that yielded the best classification performance was considered to be the optimal feature set.

ReliefF [47] is a simple and efficient feature selection method, it is also one of the most well-known instance-based feature weighting algorithms. ReliefF is an extension of the Relief method [48], and is used for dealing with multi-class problems. The core idea of ReliefF is to minimize the intra-class variance and maximize the inter-class variance. ReliefF selects the k nearest samples from the same classes and the different classes respectively, and treats their mean as the weight of each feature. The weight of a feature can be computed according to the following formulation:

$$W(A) = W(A) - \frac{\sum_{j=1}^{k} diff(A,\ R_i,\ H_j)}{m \bullet k} + \frac{\sum_{C \neq class(R_i)} \left[\frac{P(C)}{1-P[class(R_i)]} \times \left(\sum_{j=1}^{k} diff\left[A,\ R_i,\ M_j(C)\right] \right) \right]}{m \bullet k} \tag{1}$$

where A represents a feature, R_i is a sample selected randomly from the set of samples. H_j is one of the k samples closest to R_i found from the samples of same class as R_i; M_j is one of the k samples closest to R_i found from the samples of different classes as R_i. C indicates class and $class(R_i)$ is the class of R_i. Parameter m is the number of samples selected randomly, and the maximum of m is the total number of the set of samples. Parameter k is the number of nearest samples, and the upper limit of k is the number of samples of per class.

In the Eq. (1), the initial value of $W(A)$ is set to zero. $P(C)$ is the proportion of the number of target samples with class C to the total number of samples. *diff* (*Feature, Sample1, Sample2*) is the Euclidean distance of two objects, which is used to measure the dissimilarity of two samples. For discrete features,

$$diff(F,\ S_1,\ S_2) = \begin{cases} 0 & val(F,\ S_1) = val(F,\ S_2) \\ 1 & val(F,\ S_1) \neq val(F,\ S_2) \end{cases} \tag{2}$$