Knowledge-Based Systems 000 (2016) 1–11

Contents lists available at ScienceDirect

Knowledge-Based Systems

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

Thyroid lesion classification in 242 patient population using Gabor transform features from high resolution ultrasound images

U Rajendra Acharya[a,b,c,*], Pradeep Chowriappa[d], Hamido Fujita[e], Shreya Bhat[f], Sumeet Dua[d], Joel E.W. Koh[a], L.W.J. Eugene[a], Pailin Kongmebhol[g], K.H. Ng[h]

[a] Department of Electronics and Computer Engineering, Ngee Ann Polytechnic, Singapore
[b] Department of Biomedical Engineering, School of Science and Technology, SIM University, Singapore
[c] Department of Biomedical Engineering, Faculty of Engineering, University of Malaya, Malaysia
[d] Department of Computer Science, Louisiana Tech University, Ruston, LA 71272, USA
[e] Iwate Prefectural University (IPU), Faculty of Software and Information Science, Iwate 020-0693, Japan
[f] Department of Psychiatry, Research Associate, St. John's Research Institute, Bangalore 560034, India
[g] Department of Radiology, Faculty of Medicine, Chiang Mai University,110 Intawaroros Road, Chiang Mai 50200, Thailand
[h] Department of Biomedical Imaging, Faculty of Medicine, University of Malaya, Kuala Lumpur 50603, Malaysia

ARTICLE INFO

Article history:
Received 9 February 2016
Revised 10 May 2016
Accepted 10 June 2016
Available online xxx

Keywords:
Thyroid nodules
High resolution
SMOTE
Over-sampling strategies
Post hoc test
Relief-F
LSDA

ABSTRACT

Thyroid cancer commences from an atypical growth of thyroid tissue at the edge of the thyroid gland. Initially, it forms a lump in the throat and an over-growth of this tissue leads to the formation of benign or malignant thyroid nodules. Blood test and biopsies are the standard techniques used to diagnose the presence of thyroid nodules. But imaging modalities can improve the diagnosis and are marked as cost-effective, non-invasive and risk-free to identify the stages of thyroid cancer. This study proposes a novel automated system for classification of benign and malignant thyroid nodules. Raw images of thyroid nodules recorded using high resolution ultrasound (HRUS) are subjected to Gabor transform. Various entropy features are extracted from these transformed images and these features are reduced by locality sensitive discriminant analysis (LSDA) and ranked by Relief-F method. Over-sampling strategies with Wilcoxon signed-rank, Friedmans and Iman-Davenport post hoc tests are used to balance the classification data and also to improve the classification performance. Classifiers such as support vector machine (SVM), k-nearest neighbour (kNN), multi-layered perceptron (MLP) and decision tree are used for the characterization of benign and malignant thyroid nodules. We have obtained a classification accuracy of 94.3% with C4.5 decision tree classifier using 242 thyroid HRUS images. Our developed system can be used to screen the thyroid automatically and assist the radiologists.

1. Introduction

Thyroid cancer is the accumulation of malignant cells in the tissues of thyroid gland and constant exposure to radiation, age and gender affects the risk of thyroid cancer. Papillary, anaplastic, follicular and medullary thyroid cancer are four different types of thyroid cancers where anaplastic thyroid cancer is incurable compared to the rest [57].

In the United States, around 62,450 new cases of thyroid cancer are registered every year with papillary thyroid cancer being the most common type and on a yearly basis, around 1950 deaths from thyroid cancer are being reported [72]. In the United Kingdom, around 2700 new cases of thyroid cancer are registered every year, adding up to less than 1% of cancer cases [56]. Thyroid cancer has no age barrier but increases in aggression due to thyroid cancer being common in older population. About 2% of the teen and children population are affected; 2 out of 3 cases diagnosed with thyroid cancer are in the population aged below 55 years. Thyroid nodules are the first signs of thyroid cancer and females are three times more prone to thyroid cancer than males [51]. As much as 75% of the population develops nodules which are benign and less than 1% thyroid nodules become malignant [72]. Table 1 presents the percentage of population affected by thyroid cancer.

The rise in the incidence rate of thyroid cancer has increased the need for cost effective, quantitative and efficient diagnostic systems. In the very beginning, a blood test called thyroid function test is done to measure hormone levels in the blood and confirm the absence of other thyroid disorders. Fine needle aspiration (FNA) biopsy is commonly used in the detection of thyroid

* Corresponding author.
E-mail address: aru@np.edu.sg (U.R. Acharya).

http://dx.doi.org/10.1016/j.knosys.2016.06.010
0950-7051/

Table 1
Percentage of population affected by different types of thyroid cancer [59].

Types of thyroid cancer	Population affected (~ %)
Papillary (Follicular/mixed papillary) thyroid cancer	80
Follicular (Hurthle cell) thyroid cancer	15
Medullary thyroid cancer	3
Anaplastic thyroid cancer	2

nodules. It is a labor intensive technique in large scale screenings and hence, may not be an optimum technique in the thyroid cancer diagnosis [55]. Medical imaging modalities play an important role in this case. Ultrasound imaging is one of the cost effective and non-invasive techniques that can predict the risk of malignancy. Mono et al. [53] used ultrasonographic images of benign and malignant thyroid nodules and evaluated its diagnostic accuracy based on shape, nodule size, echotexture, echogenicity, margin, presence of spongiform appearance, micro and macro-calcification. They observed that an ultrasonographic diagnostic criterion is dependent on tumor size. Presence of spiculated margin, hypoechogenicity, microcalcification and macrocalcification are reported in malignant ultrasonographic images whereas isoechogenicity and spongiform appearance is noticed in benign ultrasonographic images.

Large number of non-palpable thyroid nodules detected using thyroid ultrasound and well-trained personnel are prone to inter-observer variability and poor accuracy in the detection of benign and malignant thyroid nodules. Thus, automated diagnostic and classification systems are developed for accurate diagnosis of benign and malignant thyroid nodules. Acharya et al. [6] reviewed several types of computer-aided diagnostic (CAD) systems that are developed for the automated classification of thyroid nodules. Statistical methods and data mining algorithms are used to extract sonographic features and non-clinical features from the ultrasound images. These features are fed to classifiers to achieve the highest accuracy in classifying benign and malignant thyroid nodules.

Thyroid nodule detection (TND) software for automated detection of malignant thyroid nodule is developed using fuzzy gray level histogram and fuzzy local binary pattern features [49]. Their system is able to diagnose the unknown nodule from normal subjects with an accuracy of more than 95%. Medium-risk and high-risk nodules are classified using textural and shape based features obtained an area under curve of 0.95 [43]. The fusion of fuzzy local binary patterns and fuzzy grey-level histogram features coupled with SVM classifier has yielded a classification accuracy of 97.5% for the classification of malignant nodule from normal subjects [36]. Normal and malignant thyroid nodules are classified using Radon transform based features and k-nearest neighbor classifier using 66 patients [69]. They have reported a classification accuracy of 89.4% using leave one out method.

In this study, a total of 242 thyroid nodule images are selected for feature extraction and pattern classification. Nonlinear algorithms are applied to the pre-processed images followed by feature space reduction by locality sensitive discriminant analysis (LSDA) and feature ranking by Relief-F. The SMOTE oversampling strategy for the data balance and different classifiers such as decision tree, support vector machine (SVM), k-nearest neighbour (kNN), and multi-layered perceptron (MLP) are used for classification of benign and malignant thyroid nodules. Fig. 1 presents the system flow of thyroid nodule classification. Fig. 2 presents typical appearance of benign nodule and Fig. 3 represents malignant nodule.

The proposed system consists of three modules as shown in Fig. 1. As part of phase 1, each image is subjected to Contrast-limited adaptive histogram equalization (CLAHE) for pre-processing and then Gabor transformation. Relevant textural features are obtained using well known feature extraction techniques. In phase 2 of the system, these features are subject to feature space reduction using LSDA to obtain the results in a subset of related features. These reduced set of features are then normalized and ranked. Finally in phase 3, these highly ranked features are used for classification

Due to the inherent data imbalance in the data set, with 211 image representing benign samples (of the 242 samples), phase 3 of the system focuses on the application of relevant oversampling strategies and pairing them with known classification algorithms in the area. Our objective is to provide a systematic validation strategy that helps choose relevant oversampling strategies and an associated classifier that yields results that are significant by taking into consideration the imbalance in data.

This paper is divided into five sections. The Introduction to the paper can be found in Section 1. Section 2 contains details of data acquisition, image pre-processing, feature extraction, feature space reduction, ranking and data normalization. Section 3 explains the classification approach with over-sampling strategies and validation techniques. Results and discussion are discussed briefly in Section 4 followed by the conclusion in Section 6.

2. Material and methods

2.1. Data acquisition

The ultrasound images of 223 patients with benign and malignant thyroid nodules are collected from picture archiving and communication system (PACS) of Chiang Mai University Hospital, Thailand during the period between 1st December 2009 and 30th April 2015. Total of 242 subjects with thyroid nodules were enrolled in the study that included 63 males and 179 females in the age range 12-88 years. Around 22 subjects underwent multiple fine needle aspiration (FNA) biopsies for multiple nodules. LOGIQ_9 and LOGIQ_E9 systems (General Electric, Milwaukee, Wisconsin) with 10-14 MHz linear transducer, Acuson Sequoia 512 (Siemens Ultrasound, MountainView, California) with 5-13 MHz linear transducer, Aplio-XG (Toshiba, Tokyo, Japan) with 10-13 MHz linear transducer, iU22 (Philips Healthcare, Bothell, Washington) with 5-15 MHz linear transducer were used for the recording and mode picture of each thyroid nodule was used for the analysis. In this study, out of 242 thyroid nodules, 211 nodules are benign (89 adenomatoid nodules, 76 colloid nodules, 12 follicular adenomas, and 34 thyroiditis) and 31 nodules are malignant (24 papillary carcinomas, 3 follicular carcinomas, 2 anaplastic carcinomas, 1 lymphoma, 1 poorly differentiated carcinoma). The original image obtained had a resolution of 1024×768. In this work, we have resized the image to 494×610 and stored in BMP format.

2.2. Image pre-processing

The images are subjected to CLAHE [29] and then Gabor transform to obtain the texture information for nonlinear analysis. Nonlinear algorithms are applied to the pre-processed images to extract salient hidden information followed by space reduction, data normalization and feature ranking. Images are classified using machine learning algorithms coupled with validation strategies to achieve efficient classification. The techniques used for the image analysis are briefly explained in this section.

2.2.1. Gabor transform

An image is convolved with Gabor filter to extract its texture information in terms of its frequency and orientation. Gabor kernels are product of Gaussian and Fourier kernels and are used to detect edges, corners and blobs [19,21,37,73]. It is a space-frequency function where Gabor parameters are given by Fourier transform of an