Applied Soft Computing 46 (2016) 652–664

Contents lists available at ScienceDirect

Applied Soft Computing

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

Effective segmentation and classification of thyroid histopathology images

J. Angel Arul Jothi, V. Mary Anita Rajam*

Department of Computer Science and Engineering, College of Engineering Guindy, Anna University, Chennai, India

ARTICLE INFO

Article history:
Received 15 June 2015
Received in revised form
18 December 2015
Accepted 18 February 2016
Available online 9 March 2016

Keywords:
Computer Aided Diagnosis (CAD)
Multilevel thresholding
Segmentation
Feature extraction
Feature reduction
Rule generation
Classification
Particle Swarm Optimization (PSO)
Variable Precision Rough Sets (VPRS)
Histopathology image

ABSTRACT

This paper proposes a Computer Aided Diagnosis (CAD) system that semi-automatically segments and classifies H&E-stained thyroid histopathology images into two classes: Normal Thyroid (NT) or Papillary Thyroid Carcinoma (PTC) based on nuclear texture features. Our system segments the given histopathology image into different binary images using Particle Swarm Optimization (PSO)-based Otsu's multilevel thresholding. From the segmented binary images, a binary image containing the nuclei is chosen manually. Nuclei are extracted from the manually selected binary image by imposing an area constraint and a roundness constraint. The intensity variations of pixels within the nuclei are quantified by extracting texture features. Variable Precision Rough Sets (VPRS)-based β-reduct is used to identify redundant features and generate rules. The rules are then stored in a rule base. A novel closest-matching-rule (CMR) algorithm is proposed to classify a new test sample as PTC or NT using the rules in the rule base. We verified experimentally that the proposed CAD system provides promising results and it is supposed to assist pathologists in their decisions.

1. Introduction

Histopathology is a branch of pathology where biopsy samples of tissues are examined under a microscope by pathologists for the diagnosis of cancer. Specifically, a histopathology study is essential to diagnose different subtypes of a particular type of cancer. The tissue biopsy samples procured are processed, stained and fixed onto glass slides before being examined by a pathologist under a microscope.

Papillary Thyroid Carcinoma (PTC) is the most common histological subtype of thyroid cancer. It accounts for eighty percentage of the total thyroid cancers diagnosed [1–4]. Unlike other histological subtypes of thyroid cancer, an early diagnosis of PTC provides better treatment planning and patient prognosis [4,5].

The diagnosis of PTC is based on a histopathology study that relies on nuclear features [4]. The PTC histopathology images contain orphan annie-eye nuclei. These nuclei are enlarged in size and may contain nuclear grooves and nuclear pseudo-inclusions [1,2,4,5]. Nuclear grooves result from irregularity of nuclear contour [2]. The pseudo-inclusions in the orphan annie-eye nucleus are due to the accumulation of cytoplasm in prominent nuclear grooves. The centres of orphan annie-eye nuclei lack chromatin, while a thin ring of chromatin is present at their periphery. Contrarily, the nuclei in Normal Thyroid (NT) histopathology images are smaller in size and have dense chromatin. The presence of orphan annie-eye nuclei is one of the essential nuclear features for the diagnosis of PTC.

This work is motivated by the following reasons: PTC incidence has increased in recent years [6,7]. Pathologists examine the slides under a microscope to provide diagnosis. This is a highly laborious and time-consuming task for the pathologists that would result in intra-observer and inter-observer variations. Additionally, the accuracy of the diagnosis depends upon the experience of the pathologist. An automated diagnosis system can assist the pathologist in the routine clinical examination of slides.

With recent advancements in digital pathology (enables whole slide images to be acquired, stored and transmitted electronically), there is a need for developing Computer Aided Diagnosis (CAD) systems for histopathology images. CAD of medical images involves computers to process and interpret images using algorithms from image processing and pattern recognition. The CAD systems can act as a second opinion and help the doctors in diagnosis.

* Corresponding author. Tel.: +91 9840700071.
E-mail addresses: jothi@cs.annauniv.edu (J. Angel Arul Jothi), anitav@annauniv.edu (V. Mary Anita Rajam).

http://dx.doi.org/10.1016/j.asoc.2016.02.030
1568-4946/

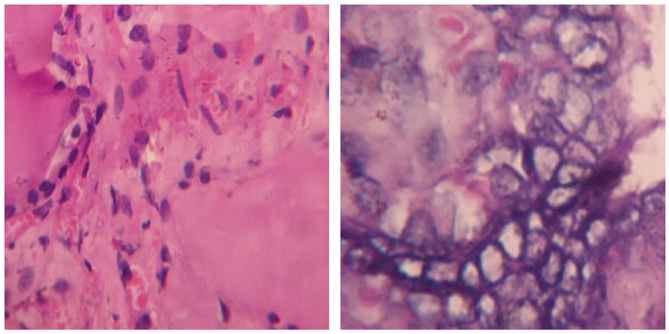

Fig. 1. Sample images from the dataset. (a) Normal Thyroid (NT) histopathology image. (b) Papillary Thyroid Carcinoma (PTC) histopathology image.

The biological and staining properties of the nuclei present in NT and PTC images vary and hence this fact has been exploited to develop a CAD system. The centre of the orphan annie-eye nucleus does not stain owing to the lack of chromatin, but its periphery stains faintly in blue due to the thin ring of chromatin [8]. A nucleus in NT histopathology image has dense chromatin and this stains deep blue. Fig. 1 shows sample NT and PTC histopathology images from the dataset respectively. The images in Fig. 1 are cropped for better visibility.

This paper proposes a Computer Aided Diagnosis (CAD) system to classify thyroid H&E-stained histopathology images into two classes: NT or PTC. It consists of the following stages: pre-processing, segmentation, Region of Interest (ROI) extraction, feature extraction, feature reduction, rule generation and classification.

Fig. 2 shows the steps involved in the proposed system. Initially, the digital images of the biopsy samples are acquired using a microscope and a camera. The acquired input images in *RGB* colour space are given as input to the system. The pre-processing stage converts the input image into gray scale, removes noise by median filtering and enhances the contrast of the image using Contrast Limited Adaptive Histogram Equalization (CLAHE). The segmentation stage uses PSO based Otsu's multilevel thresholding to segment/partition the image into different binary images containing different regions. The binary image containing the nuclei is selected manually and is given to the ROI extraction stage. The ROI extraction stage extracts the nuclei from the image eliminating unwanted artefacts. The feature extraction stage extracts the Gray Level Co-occurrence Matrix (GLCM) texture features from the image containing the nuclei. The feature reduction stage uses VPRS based β-reduct to identify the redundant features and removes them from the list of extracted features. The rule generation stage generates decision rules and stores them in the rule base. The classification stage uses the closest-matching-rule (CMR) algorithm to classify unseen test samples. The test samples are generated from query images. The query images are subjected through all stages of the CAD system such as pre-processing, segmentation and feature extraction and are finally classified as either PTC or NT.

The rest of this paper is organized as follows. We first detail the related work in Section 2. Section 3 provides the background about the techniques used in this work. Section 4 details the various stages in the decision system. The dataset, parameter tuning, metrics used for evaluation of classifiers, experiments conducted and the results obtained are provided in Section 5. Finally, the conclusion is given in Section 6.

2. Related work

This section details about the related works in the field of histopathology image analysis.

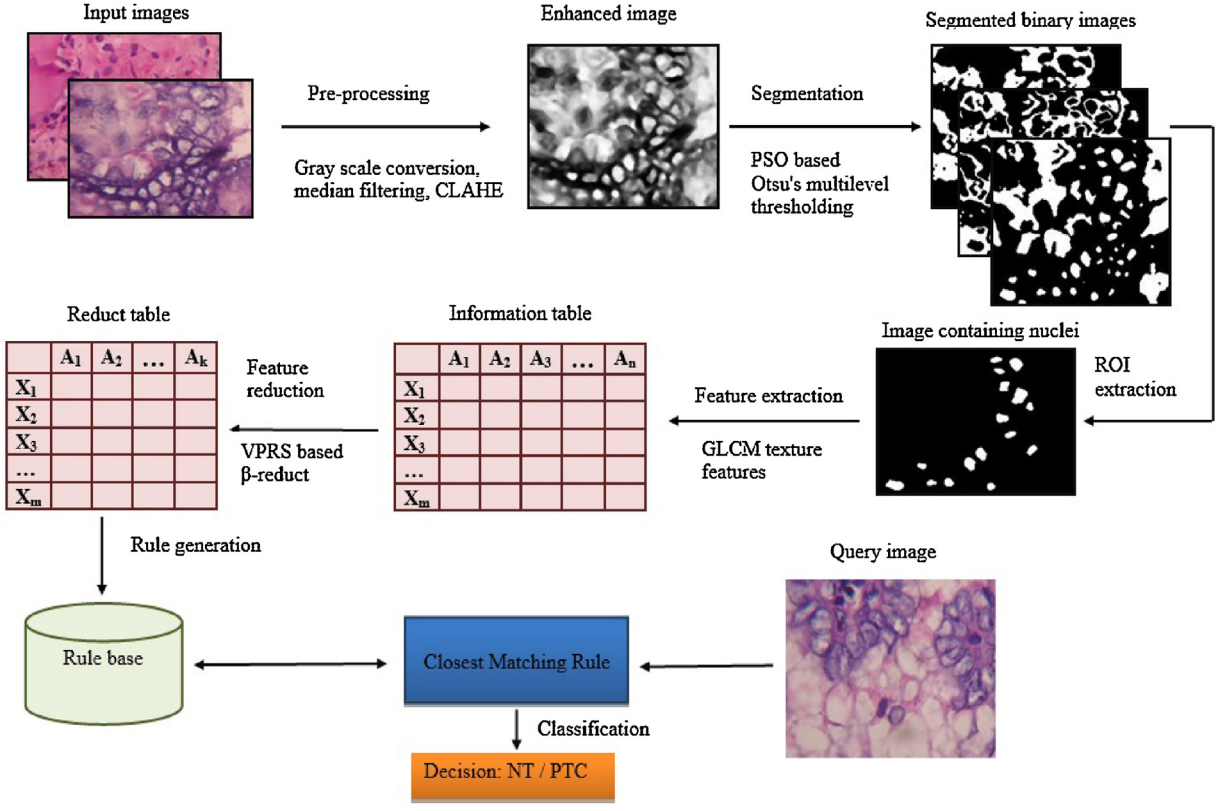

Fig. 2. Steps involved in the Computer Aided Diagnosis of thyroid H&E stained histopathology images.