Survey On Automatic Cancer Detection And Diagnosis

Anjali Vijayan[1], Seena Thomas[2]

[1] M.Tech CSE, LBSITW, Poojappura,
Thiruvananthapuram, Kearala, University, India
anjaliv.vijayan@gmail.com

[2] Assistant Professor, Dept. of IT, LBSITW, Poojappura
Thiruvananthapuram, Kerala University,India
seena.thoms@gmail.com

Abstract: *Cancer is the leading cause of death for all over the world. In case of cancer diagnosis the classification of cancerous tissue has an important role. This is a survey on automatic cancer detection techniques. Since the cause of the disease remains unknown, early detection and diagnosis is the key for cancer control, and it can increase the success of treatment, save lives and reduce cost. Many automated systems have been implemented to aid histopathologists and clinicians in cancer diagnosis and research, which have been attempted to significantly reduce the labor and subjectivity of traditional manual intervention with histology images. Generally, a automatic system consist of four stages: preprocessing, segmentation, feature extraction and selection, and classification. In this paper, the approaches used in these stages are summarized.*

Keywords: Cancer diagnosis, preprocessing, segmentation, feature extraction.

1. Introduction

Cancer is a class of diseases characterized by out of control cell growth. Cancer harms the body when altered cells divide uncontrollably to form lumps or masses of tissue called tumors. The causes of cancer are diverse, complex and only understood. Many cancers could be prevented by smoking, eating more vegetables, fruits and whole grains, eating less meat and refined carbohydrates , maintaining a healthy weight, excising, minimizing sunlight expose etc. Cancer can be deleted in a number of ways, including the presence of certain signs and symptoms, screening tests or medical imaging. Once a possible cancer is detected it is diagnosed by microscopic examination of tissue sample. People with suspected cancer are investigated with medical tests. These commonly include blood tests X-ray, CT scans and endoscopy.

The tissue diagnosis given by the pathologist indicates the type of cell that is proliferating, its histological grade, genetic abnormalities and other features of the tumor. Together, this information is useful to evaluate the prognosis of the patient and to choose the best treatment. Digital pathology provides a digital environment for the management and interpretation of pathology information that is enabled by digital slide (virtual slide). The implementation of these systems typically requires a deep analysis of biological deformation from a normal to a cancerous tissue. This deformations are typically observed in the distribution of the cells from which cancer originates, and thus in the biological structures that are formed of these calls. The degree of the deformation in these structures is an indicator of the cancer malignancy. Thus, the correct identification of the deformations and their accurate quantification are quite critical for precise modeling of cancer.

Digital pathology systems are implemented for different purpose; mainly efforts have been dedicated to tissue image classification. Conventional systems are based on the use of statistical pattern recognition techniques and also using structural pattern recognition techniques. The automated systems have major four stages. These are preprocessing, segmentation, feature extraction and classification. The statistical approaches have been proposed for feature extraction. Some of the technique uses pixel level information. While textural features can model the texture of small regions in an image well, so they lose the information of local structures, such as characteristics of glands in entire tissue image. While use of component level information, extract their features by identifying histological components. This also does not consider the organization of the components in feature extraction.

To model the organization characteristics, it has been proposed to use structural approaches, constructing a graph on its tissue components and extract global features on the constructed graph. These global features are extracted over the entire graph. But it cannot retain all the relations among the graph nodes. This paper study the different techniques included in statistical and structural pattern recognition.

2. BACKGROUND ON AUTOMATED CANCER DETECTION

The automated cancer diagnosis is based on extracting information from the tissue images of and examining this information by using either statistical analysis or structural analysis. The cancer diagnosis consists of four main computational steps: preprocessing, segmentation, feature extraction, and classification. The aim of the preprocessing step is to eliminate the background noise and improve the image quality for the purpose of determining the focal area in the image. The preprocessing becomes the most important

yet difficult step for a successful feature extraction and diagnosis. After preprocessing the image, features are extracted for further accuracy of classification of lesion. The aim of the classification step is to distinguish benignity and malignancy or to classify different malignancy levels by making use of extracted features. This step uses either statistical or structural analysis of the features and machine learning algorithms to reach a decision. An overview of these three steps is given in figure. 1.
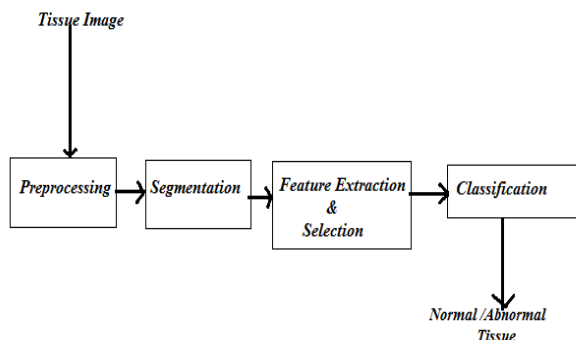

Figure . 1. 1. Overview of the computational step

3. LITRATURE REVIEW

In this section, the existing work dealing with different type of cancer detection and diagnosis is discussed. In early days automated systems are typically based on statistical patter recognition techniques.

A. Tabesh, M. Teverovskiy, H. Y. Pang, V. P Kumar, D. Verbel, A Kotsianti, and O. Saidi (2007) presented a paper which addresses the multi-feature prostate cancer diagnosis and gleasson grading of histological images[5].Their primary contribution of that paper is to aggregate color, texture, and morphometric cues at the global and histological object levels for classification. They presented a comparison of the performance of Gaussian, *k*NN, and SVM classifiers on the classification tasks. Finally they introduced fractal code features for medical image textural analysis.

Claudia Wittke, Johannes Mayer, and Franz Schweiggert (2007) introduced a new system on the classification of prostate carcinoma with methods from spatial ststistics[4].They developed automatic methods for Gleason grading based on combinations of different morphological characteristics. Among these, certain characteristics were visually selected on the basis of 1-D and 2-D plots made for training set.

The statistical pattern recognitions based on the uses of either pixel level or component level informations. But they do not consider the organization of the components in feature extraction. So model the organizational characteristics, it has been proposed to use structural approaches that constructing a graph on tissue components in an image and extracting global features on the constructed graph. The existing work on structural approach is discussed on following section.

Cigdem Gunduz- Demir, Melih Kandemir, Akif Burak Tosun, Cenk Sokmensuer(2010) developed a new approach that is automatic segmentation of colon glands using object-graphs[3]. They introduced an object-based approach for the purpose of gland segmentation. This approach decomposes the tissue image into set of primitive objects. The proposed algorithm provides an infrastructure for further analysis of tissue images that include glandular structures. This infrastructure allows us to locate gland on the tissue image and to understand whether or not a gland deviated from its normal structure.

Dogan Altunbay, Celal Cigir, Cenk Sokmensuer and Cigdem Gunduz-Demir(2010) tackled the problem of automated cancer detection by Color graph for automated cancer diagnosis and grading[2]. They report a new tissue representation in which the locations of different tissue components are considered. In this type of representation, a color graph is generated constructing a Delaunay triangulation on nuclear, stromal, and luminal tissue components and coloring the triangle edges based on the component type of their endpoints. Three global graph properties are used and they also compare the color graph approach with the intensity- based, textural, and colorless structural approaches.

Erdem Ozdemir, Cigdem Gunduz-Demir(2013) are develop a new approach, a hybrid classification model for digital pathology using structural and statistical pattern recognition[1].They proposed a effective hybrid model that use both structural and statistical pattern recognition techniques to locate and characterize the biological structure in tissue image. This hybrid model quantifies the located key regions with two different types of feature extracted using structural and statistical approaches.

3.1 Preprocessing

The major limitations of images are the low contrast and interference with speckle. So enhance the image without destroying the important features of the image for diagnosis. The main aim of the preprocessing step is to determine the focal areas in the image[6]. Due to a considerable amount of noise that arises from the staining process, it is usually necessary to reduce the noise prior to the focal area identification. In the case of tissue level feature extraction, the preprocessing step thresholds the image to identify the focal areas by eliminating the noisy regions and those with little content. Many methods enhance the image and remove speckle at the same time. The main preprocessing techniques are given in the following table 1.1.

3.2 Segmentation

The segmentation method should be chosen depending on the type of the feature to be extracted. Segmentation is a partition of the image *I* into non-overlapping regions[7]. The goal for the segmentation is to locate suspicious areas to assist pathologists in diagnosis. The main segmentation techniques are given in the table 1.2.

Table 1.1. Preprocessing Techniques

Methods	Description	Advantages	Disadvantages

Methods	Description	Advantages	Disadvantages
Filtering Techniques	Use moving window to filter with the image to reduce noise	Fast Simple	Difficult to separate noise from single scale representation
Wavelet approaches	Transform image to wavelet domain and remove noise by modifying wavelet coefficients	Statistics of the signals are simplified	Increases time complexity due to DWT AND IDWT computations.
Compounding approaches	Average image obtained by varying scanning frequency or view angle	Signal and noise processed at different scales orientation simple	Hardware support is necessary. High time complexity by registration and reconstruction

Table 1.2. Segmentation Techniques

Methods	Descriptions	Advantages	Disadvantages
Histogram thresholding method	Segment the image by using threshold value	Simple Fast	For images with non-bimodel histograms have no good result
MRF	Labeling image regions based on the corrected intensity statistics	Precise Accurate	Complex and time consuming for many iteration
NN	Segmentation regarded as a classification task	Extract the contours of tumors automatically	How to select the training set is problematic, training is time consuming and depending on database

3.2 Feature Extraction and Selection

In cancer detection and classification important step is feature extraction and selection. An optimum feature set should have effective and discriminating features, How to extract useful features and make a good selection of the features is a crucial for automated systems. Extraction and selection of effective features is a necessary step, so the general guidelines for selecting features mainly include four considerations: discrimination, reliability, independence and optimality[8]. The goal of feature extraction and selection is to maximize the discriminating the performance of the feature group. The main techniques are given in table 1.3

Table 1.3. Features

Feature Category	Feature description
Texture feature	Variance, auto- correlation or average contrast GLD matrix based features: correlation, energy ,entropy, inertia and homogeneity Standard deviation and contrast of gray value
Morphological features	Spiculation Depth to width ratio Margin sharpness Angular characteristics Area of lesion
Descriptor features	Speculated margin Shape(round, oval, irregular) Distortion, echogenic halo Duct extension

Feature selection: Algorithm for feature selection can be categorized into two classes: wrapper and filter. Filter approach selects features using a preprocessing step and does not use induction algorithms. On wrapper approach uses induction algorithm as part of evaluation function. Wrapper approach has advantages over filter approach, especially for complex feature data set.

3.4 Classification

After the features have been extracted and selected, they are the input into a classifier to categorize the image into normal/abnormal tissue[9]. The classification methods are given below table 1.4

Table 1.4. Classification Methods

Classifier	Feature used	Advantage	Disadvantage
Linear Classifier	Textural features, Morphological features Descriptor features	Simple, Effective for linearly separable data	Poor performance for non-linearly separable data Poor adaptability for complex problem
ANNs	Textural features Morphological features	Robustness no rule or explicit expression is needed and widely applicable	Long training time Initial value dependent Unrepeatable
Decision tree	Textural features	Low complexity	Accuracy fully depends on the design of the tree and the features

Support vector machines	Textural features Morphological features	Training process is faster than NN's Good performance Repeatable training process	Training data should be labeled Parameter dependent

5. CONCLUSION

In this review, we studied different steps to automatically analyze histopathological images for cancer diagnosis. This automated analysis assists pathologist diagnosis and lessen their time for reviewing large number of tissue slide per day. For computer aided analysis there are mainly four steps: Pre-processing, Segmentation, Feature Extraction and Classification. The algorithms developed for automated analysis and evaluation of histology images assists the pathologists in disease diagnosis and also reduces human error.

References

[1] Erdem Ozdemir, Cigdem Gunduz-Demir, "A hybrid classification model for digital pathology using structural and statistical pattern recognition," *IEEE Trans. Knowledge Medical Imaging.,* vol. 32, no. 2, pp. 474-483, Feb 2013.

[2] D. Altunbay, C. Cigir, C. Sokmensuer, and C. Gunduz-Demir, "Color Graphs for automated cancer diagnosis and grading," *IEEE Trans. Biomed. Eng.*, vol. 57, no. 3, pp. 665–674, Mar. 2010.

[3] Cigdem Gunduz- Demir, Melih Kandemir, Akif Burak Tosun, Cenk Sokmensuer, "automatic segmentation of colon glands using object- graphs," Elsevier Medical Image Analysis,vol. 14pp.1-12,oct 2010.

[4] C. Wittke, J. Mayer, and F. Schweiggert, "On the classification of prostate carcinoma with methods from spatial statistics," *IEEE Trans.Inf. Technol. Biomed.*, vol. 11, no. 4, pp. 406–414, Jul. 2007.

[5] A. Tabesh, M. Teverovskiy, H. Y. Pang, V. P. Kumar, D. Verbel, A. Kotsianti, and O. Saidi, "Multifeature prostate cancer diagnosis and Gleason grading of histological images," *IEEE Trans.Med. Imag,* vol. 26, no. 10, pp. 1366–1378, Oct. 2007.

[6] A.N. Esgiar, R.N.G. Naguib, B.S. Sharif, M.K. Bennett, A. Murray, Microscopic image analysis for quantitative measurement and feature identification of normal and cancerous colonic mucosa, IEEE T. Inf. Technol. B. 2(1998) 197-203.

[7] H.D. Cheng, L. Hu, J. Tian, L. Sun, A novel Markov random field segmentation algorithm and its application to breast ultrasound image analysis, in: The Sixth International Conference on Computer Vision, Pattern Recognition and Image Processing, Salt Lake City, USA, 2005.

[8] P.S.Rodrigues,Anewmethodologybasedon q-entropy forbreastlesion classification in 3D ultrasoundimages ,in: Proceedings of the 28th *IEEE EMBS Annual InternationalConference*,2006,pp.1048–1051.

[9] B. Sahiner, Malignant and benign breast masses on 3D US volumetric images: effect of computer-aided diagnosis on radiologist accuracy, Radiology 242 (3) (2007) 716–724.