Multispectral imaging and machine learning for automated cancer diagnosis

Somaya Al Maadeed, Suchithra Kunhoth
Department of Computer science and Engineering
Qatar University
Doha, Qatar

Ahmed Bouridane, Remy Peyret
Department of Computer Science and Digital Technologies
Northumbria University
UK

***Abstract*—Advancing technologies in the current era paved a lot to break the hurdles in medical diagnostic field. When cancer turned out to be the most common and dangerous disease of the age, novel diagnostic methodologies were introduced to enable early detection and hence save numerous lives. Accomplishment of various automatic and semi-automatic approaches in the diagnosis has proved its sufficient impetus to improve diagnostic speed and accuracy. A wide range of image processing based tools are currently available as a part of automatic cancer detection systems. Different imaging modalities have been utilized for extracting the suspected patient information, where the multispectral imaging has emerged as an efficient means for capturing the entire range of spectral and spatial data. In this paper, we review the current multispectral imaging based methods for automatic diagnosis of major types of cancer and discuss the limitations which are yet to be overcome, so as to improve the existing systems.**

Keywords—Cancer detection; automatic; multispectral; hyperspectral; infrared imaging

I. INTRODUCTION

The number of persons affected by one or other type of cancer is increasing day by day and hence the death toll also. An earlier and a prompt diagnosis of the cancer would help to avoid the casualties to a greater extent. With the advent of modern highly sophisticated technologies, a wide variety of tools are available for the detection of the cancerous tissues in our body. Cancer may go through several stages from the initial localized ones such as tumors to the metastasis phase where it spreads over a distant area. The diagnostic tools capable of accurately determining the stage would aid the doctor to initiate an effective treatment. Specific screening tests are available for majority of cancer types, which can identify ahead of the onset of specific symptoms [1]. Once the screening test indicates a positive result, further diagnostic tests are prescribed to determine the severity and hence to start an ideal treatment methodology.

In this paper, we have given an overview of multispectral imaging for automation of cancer diagnosis in section 2. Section 3 provides a review of related researches in 5 major types of cancer detection which is followed by an analysis in section 4. Finally, concludes the paper in section 5.

II. MULTISPECTRAL IMAGING FOR CANCER DIAGNOSTIC AUTOMATION

Although there exist different techniques such as x-rays, ultrasound etc., it seems that biopsy proves to be the ultimate method for cancer diagnosis. As far as manual inspection of biopsy slides are concerned, the tissues are visually inspected under a microscope by the pathologist. A fine inspection of the cellular structures involves a staining process of the slides; normally the haematoxylin and eosin stain (H&E staining). Computer aided analyses of the images are very helpful to avoid the radiologist/pathologist dependency on the diagnostic inspection [2]. As far as a pathologist is concerned, there would be a number of tissue samples that has to be looked under the microscope for each patient. This tedious and monotonous task can produce erroneous results which can be false positives or false negatives. So the automated analysis can improve the detection accuracy where the inter- operator variation problem is solved. A typical spectral imaging setup deployed for pathological slide analysis is demonstrated in Fig. 1.

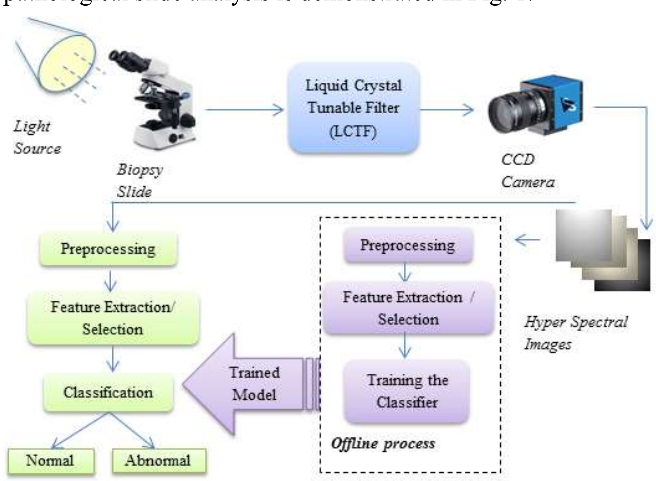

Figure 1. A multispectral Imaging Based Cancer Detection System

In automated procedures, images of the biopsy slides are captured and then suitable features are extracted from them. These features and hence the images are then classified by machine learning algorithms . Earlier approaches made use of the standard RGB color format imaging. Since the overall characterization of objects is limited here by the use of only three color channels, multispectral methods seems to be very useful. The additional information coming from the enormous spectral bands in the spectral imaging techniques have proved its efficiency in variety of fields including medical.

III. MACHINE LEARNING APPROACHES WITH MULTISPECTRAL IMAGING TECHNIQUES

Machine learning, a part of computer science exhibits the property of getting trained by a set of examples. This serves it to handle similar future data samples in an efficient manner, to suit a wide variety of applications. Along with the rising popularity of machine learning, it is being equipped with a number of capabilities including data prediction, classification etc. In medical field, it will enable the automation of diagnostic procedures which in turn leads to a rapid and accurate system. Machine learning algorithms like artificial neural networks, support vector machines (SVM), random forests(RF) and the recent deep learning technologies form the basis of classification stage in these diagnostic applications.

A lot of researches are currently taking place for the development of imaging methodologies as well as in automatic inspection of the cancerous cells using specialized image processing and computer vision techniques. Among around 200 different types of cancer, a review of the researches taking place in the diagnostic field of the most common types of cancers [3] is detailed in the upcoming sections.

A. Cervical Cancer

This is one of the major life threatening cancers affecting the female population. Pap smear tests are known to be the current widely used technique for the precancerous screening of cervix. In papsmear, sample cells are taken from the cervix and smeared onto a glass slide. These cells are then stained and sent for screening where the normal and malignant cells are identified by the pathologist. Cervical intraepithelial neoplasia, dysplasia and squamous intraepithelial lesions are some precancerous types whereas the squamous cell carcinoma and adenocarcinoma are the major cancerous categories of cervical cancer [4]. Zhang and Liu have developed an automatic cervical cancer detection system from multispectral microscopic thin Pap smear images using micro-interferometric spectral imaging setup [5]. Both the local multispectral and textural properties are utilized along with a proposed SVM based feature screening technique.

B. Breast Cancer

Akin cervical cancer, breast cancer is also a fatal disease among women. In terms of the estimated new cases, breast cancer stands first among all other cancer types in women[6]. According to the TNM staging system (T describes size of tumor, N describes the lymph nodes involved, M the distant metastasis), ductal carcinoma in situ, lobular carcinoma in situ are the pre-cancerous conditions. In the next stage, it will spread within the breast and the regional lymph nodes of armpits. An analysis of multispectral imagery of H&E stained breast tissue slides in [7] have proven that it has only slightly enhanced the classification of nuclei compared to the standard RGB imagery. They have acquired multispectral images with 29 spectral bands, spaced 10 nm within the range of 420 to 700 nm, for the study.

C. Colon Cancer

Colon cancer causes the death of around half a million people every year. It is also known as colorectal cancer, a malignant disease of the colon or rectum, and is the third leading cause of cancer death. The disease begins as polyps and end up in the invasive adenocarcinoma which is the malignant condition [6]. Most colorectal cancers begin as polyps and it would be ideal if detection is possible at this stage itself. K.M Rajpoot and N.M Rajpoot [8] have accomplished the classification of normal and malignant tissue cells using multiscale morphological features from biopsy images captured in 20 spectral bands within the range 450-640 nm. Multiscale morphological features were used to train the classifier which performed well to distinguish the cells. Masood et al. tried using morphological as well as grey level co-occurrence features from the hyperspectral data for the tissue cell classification in [9]. Experiments were carried out using the LDA (Linear Discriminant Analysis) and SVM for classification. Later they have investigated and found that a single spectral band within the range of 400-800 nm have the sufficient textural features for the biopsy classification [10]. Although PCA (Principal component analysis), 2Dimensional PCA were tried out, the textural features with an SVM proved to be very efficient. A spatial analysis of the same single spectral band with circular local binary pattern features was done in their further work [11]. Maggioni et al. [12] analyzed colon biopsy samples to discriminate among tissue features: gland nuclei, gland cytoplasm and lamina propria/lumens. Nuclei was extracted from them and then assigned the nuclei to be under three classes, normal, precancerous or cancerous based on the spectral features. But the accuracy had a dependency on the number of nuclei extracted. Chaddad et al. [13] have proposed an improved version of the snake algorithm for segmentation and the extraction of Haralick texture features in the multispectral segmented images. 16 spectral bands in the range of 500-650 nm was utilized to achieve an efficient classification of cancer cells of type Carcinoma (Ca), Intraepithelial Neoplasia (IN) and Benign Hyperplasia (BH). Another hyperspectral method for colon cancer detection proposed by Abkari et al. [14], was applied on the pathological slides of lungs and lymph node tissues from mice. The spectral signatures from tissues trained using SVM achieved detection in lymph nodes with a specificity of 98.3% and sensitivity of 96.2%.

D. Lung Cancer

Lung cancer holds the first position in terms of mortality [3]. Non-small cell lung cancer is the most common type compared to the small cell lung cancer. Adenocarcinoma, Squamous cell carcinoma and large cell carcinomas are the different types of tumors coming in the non-small cell type lung cancer. The hyperspectral approach followed by Akbari et al. [14] could detect the metastatic cancer in lung histologic tissues from mice with a specificity of 97.7% and sensitivity of 92.6%. Reference [15] has employed multispectral imaging of the TCCP (Tetrakis Carboxy Phenyl Porphine) stained sputum sample for the detection of lung cancer. Images taken at 650nm and 660nm were used to extract the shape based, intensity based, wavelet based and Gabor filter based features. Random Forest was the one that could yield a higher accuracy among the various experimented machine learning techniques.

E. *Prostate Cancer*

It is a major type of cancer with a higher probability in men after 50 years of age [4]. The tissues of the prostate gland in the male reproductive system found below the bladder are affected as a result. Although many tests such as PSA (Prostate-Specific Antigen) blood test, digital rectal examination are available, the most recommended one is the biopsy. Benign prostatic hyperplasia is a benign state. Prostatic intraepithelial neoplasia refers to a precursor and prostate adenocarcinoma is the cancerous stage. An earlier work which has recognized the significance of multispectral imaging for the grading of prostate neoplasia is [16]. Features extracted from the 16 spectral bands led to the discrimination between normal and 3 different grades of cancerous tissue. Analysis by varying the number of spectral bands emphasized the contribution of spectral signatures for classification. Drawbacks of PCA based feature selection adopted here was overcome with the tabu search in [17] and round robin tabu search in [18]. In the later paper [19], their different approaches towards classification of prostatic tissues using multispectral imagery was compared. A Round-Robin classification algorithm using a sequential forward selection/nearest neighbor classifier was proposed in [20] with an intention to improve the existing classification accuracy. And they were able to achieve a best classification accuracy of 99.9%. In [21], the concept of gray level co-occurrence matrix is extended for texture characterization in multiband images and hence to discriminate the healthy and pathological tissues. Texture joint information between spectral bands is also taken into consideration. Moreover, a band selection technique was employed which selects the best relevant bands of multispectral prostate cancer database.

A detailed comparison of the described multispectral image based cancer detection techniques is given in table 1.

TABLE 1. COMPARISON OF VARIOUS MULTISPECTRAL IMAGE BASED CANCER DETECTION TECHNIQUES

Ref	**Modality**	**Dataset**			**Evaluation (%)**
		Acquisition	***Source***	***Cardinality***	
[5]	400-690 nm (52 spectral bands)	MircoInterferometeric spectral imaging (Carnegie Mellon University)		40 images with 149 cells	TPR= 98, FPR= 1
[8]	450-640 nm (20 bands)	Yale University (1024X1024 pixels)		11 hyperspectral image cubes	Classification Accuracy: 99.72, Specificity: 99.62
[9]	450-850 nm	CCD (charge coupled device) Camera, Tuned Light	Yale University	10 biopsy slides	Accuracy:100
[10]	400–800 nm (128 bands)	CCD Camera, Tuned Light Source(448X691)		10 biopsy samples	Accuracy:91.14
[11]	440-700 nm(128 bands)	CCD Camera, Tuned Light Source(491X652)		32 biopsy samples	Accuracy:90
[12]	440-700 nm	CCD camera (Sensovation)		59 biopsy samples	Accuracy:97.1
[14]	450-950 nm	CRi Imaging system (1.4 MegaPixel)	Emory University &Georgia Institute of Technology	6 slides	Specificity: 98.3, Sensitivity:96.2
[15]	650nm, 660nm	Nuance Fx system	ATCC, Cureline Inc. USA	1199 cells	Accuracy:97.29 (650nm), 97.78(660nm)
[20]	500- 650 nm (16 bands)	LCTF multispectral filter, CCD Camera (128X128)	University of Ancona (Italy).	592 images	Accuracy:99.9
[21]	500-650 nm (16 spectral bands)	LCTF, CCD Camera	Queen’s University of Belfast	624 images	Accuracy:97

IV. Discussion

The ultimate aim of hyper/ multispectral imaging of biopsy slides is to achieve an efficient demarcation between the normal and abnormal cells in automated diagnostic systems. As a higher computational load is accompanied with the hyperspectral image processing, they should be capable of surpassing the efficacy of RGB imaging.

Masood and Rajpoot's [22] experiments on colon tissue samples indicate that textural analysis on a single band can achieve comparable classification to 3D spectral spatial analysis. However, we have seen in the previous section that most multispectral and hyperspectral imaging methods have utilized only the visible part of light spectrum. This may not be much adequate to reveal the complete tissue characteristics, as near-Infrared / mid-Infrared spectrum remains unutilized. The most challenging problem with hyperspectral data processing is the curse of dimensionality. PCA and ICA [9] are the commonly adopted feature selection strategies to get through this. Acquisition of medical data sets with high spatial & spectral resolution can be challenging. Capturing the extremely fine variations both spectrally and spatially paves the development of an ideal automated system for the classification of the various cell types. Hyper spectral imaging provides imaging in invisible bands such as infrared, thus offering the capability to reveal information that cannot be seen by the naked eye. But the full spectrum ranging from ultraviolet to near and the mid infrared have not been completely exploited in the current spectral imaging systems in the cancer diagnostic field. Extraction of relevant information from the high dimensional data cubes calls for an efficient feature extraction and selection algorithm. The time constraint due to the increased complexity of the hyperspectral image processing should be overcome by the selection of the appropriate spectral bands from the redundant data.

Table 2 gives a brief overview of the pros and cons of existing methodologies being used in the field of cancer detection, based on multispectral imaging and machine learning. It can be seen that most of the works involve experiments on a small number of images except for one or two studies. This can be due to the unavailability of the multispectral dataset, which demands complex acquisition.

TABLE 2. EVALUATION OF EXISTING APPROACHES

Ref	Pros	Cons
[5]	Cells are segmented prior to feature extraction, which may lead to better classification. Also SVM based feature screening is deployed. A good number of bands in visible range are used and a reasonable accuracy is achieved	Infrared multispectral bands are not included
[8]	Dimensionality reduction is applied which tends to reduce the computational complexity. Morphological features are used, in contrast to the common texture descriptors found in most of the cancer detection researches. High accuracy is obtained	Relatively low number of images are used in the experiments, which is too low to validate an algorithm which incorporates machine learning. Infrared multispectral bands are not used
[9]	Dimensionality reduction and cell segmentation are achieved by means of Independent Component Analysis (ICA) and k-means clustering. These two steps facilitates reducing the huge redundant information in the bands of multispectral imagery. Both morphological and texture features are extracted.	Even though 100% accuracy is mentioned, the experiment comprises only a small set of slides. Also, infrared multispectral bands are not utilized
[10]	Multispectral imagery contains data from a huge number of bands (128 bands) . Textural features and SVM classifier are used	Experiments involve only 10 biopsy slides. Usage of only visible range wavelength bands
[11]	Multispectral imagery contains data from a huge number of bands (128 bands) . Textural (circular local binary pattern) features and SVM classifier are used. Feature selection is also adopted and a reasonably good accuracy is achieved	Infrared multispectral bands are not utilized
[12]	Spectral features are used and a high accuracy is achieved	Infrared multispectral bands are not utilized
[14]	Library of spectral signatures along with SVM classifier. Accuracy is reasonably high	Only 6 slides are used in the experiments. Usage of only visible range wavelength multispectral bands
[15]	A simple thresholding along with multiple type of features (Intensity, shape and texture). Classification based on Random forest. Dataset is made of 1199 cells which is relatively a large number.	Only 2 bands from visible range are used. Infrared multispectral bands are not utilized
[20]	Texture and Structural features are extracted from the multispectral images. Algorithm is based on round-Robin classification algorithm using sequential forward selection/nearest neighbor classifier. The size of dataset and resulting accuracy is comparatively good	Multispectral imaging in the infrared range is not utilized
[21]	A band selection technique is implemented which helps to proceed with only relevant bands in the multispectral imagery. This in turn reduce the computational complexity and may enhance the accuracy by removing the unwanted redundant bands. A high accuracy is yielded with the proposed approach, which is experimented on a good number of image samples	Multispectral imaging in the infrared range is not utilized

A small dataset is not sufficient to quantify the success rate of specific algorithms which may need extensive training for learning purpose. All the methods have utilized only the spectral bands in visible range. Although multispectral imaging offers image acquisition capability in infrared bands, it is not utilized in any of the research.

V. CONCLUSION

We have given here a brief review of the major researches taking place in the automatic detection of cancer based on multispectral imaging and machine learning. Spectral Imaging, which combines both imaging and spectroscopy tend to emerge as a prominent method in the cancer diagnostic tool automation. Multispectral and hyperspectral approaches have already been used for the both the in vivo imaging and the imaging of the pathology slides. Majority of the works as we have seen here have utilized the spectral bands within the visible part of the electromagnetic spectrum. Advances in the cancer detection field is expected to be achieved with techniques that incorporate the spectral and spatial information from the wavelength bands over the invisible spectrum such as near infrared, mid infrared etc.

ACKNOWLEDGMENT

This publication was made possible using a grant from the Qatar National Research Fund through National Priority Research Program (NPRP) No. 6-249-1-053. The contents of this publication are solely the responsibility of the authors and do not necessarily represent the official views of the Qatar National Research Fund or Qatar University.

REFERENCES

[1] R. A. Smith et al, "Cancer screening in the United States, 2015: A review of current American Cancer Society guidelines and current issues in cancer screening," CA: A Cancer Journal for Clinicians, 2014.

[2] J. K. Turner, G. T. Williams, M. Morgan, M. Wright, S. Dolwani, "Interobserver agreement in the reporting of colorectal polyp pathology among bowel cancer screening pathologists in Wales," Histopathology, vol. 62, pp. 916-924, 2013.

[3] R. Siegel et al, "Cancer statistics, 2014," CA: A Cancer Journal for Clinicians, vol. 64, pp. 9-29, 2014.

[4] American Cancer Society (2015 October 15), Colon Cancer Overview[Online]. Available: http://www.cancer.org/acs/groups/cid/documents/webcontent/003096-pdf.pdf.

[5] J. Zhang and Y. Liu, "Cervical cancer detection using SVM based feature screening," in Medical Image Computing and Computer-Assisted Intervention–MICCAI 2004Anonymous Springer, 2004, pp. 873-880.

[6] National Cancer Institute: http://www.cancer.gov/. Last Accessed: 4th February

[7] L. E. Boucheron, Z. Bi, N.R. Harvey, B.S. Manjunath and D.L. Rimm, "Utility of multispectral imaging for nuclear classification of routine clinical histopathology imagery," BMC Cell Biol., vol. 8 Suppl 1, pp. S8, Jul 10, 2007.

[8] K. Rajpoot and N. Rajpoot, "SVM optimization for hyperspectral colon tissue cell classification," in Medical Image Computing and Computer-Assisted Intervention–MICCAI 2004Anonymous Springer, 2004, pp. 829-837.

[9] K. Masood , N.M. Rajpoot, H.A. Qureshi and K. Rajpoot, "Co-occurrence and morphological analysis for colon tissue biopsy classification," 2006.

[10] K. Masood and N. M. Rajpoot, "Classification of colon biopsy samples by spatial analysis of a single spectral band from its hyperspectral cube," 2007.

[11] K. Masood and N. Rajpoot, "Texture based classification of hyperspectral colon biopsy samples using CLBP," in Biomedical Imaging: From Nano to Macro, 2009. ISBI'09. IEEE International Symposium On, 2009, pp. 1011-1014.

[12] M. Maggioni et al, "Hyperspectral microscopic analysis of normal, benign and carcinoma microarray tissue sections," in Biomedical Optics 2006, 2006, pp. 60910I-60910I-9.

[13] A. Chaddad, C. Tanougast, A. Dandache, A. Al Houseini, and A. Bouridane, "Improving of colon cancer cells detection based on haralick's features on segmented histopathological images," in Computer Applications and Industrial Electronics (ICCAIE), 2011 IEEE International Conference On, 2011, pp. 87-90.

[14] H. Akbari et al, "Detection of cancer metastasis using a novel macroscopic hyperspectral method," in SPIE Medical Imaging, 2012, pp. 831711-831711-7.

[15] K. Kancherla and S. Mukkamala, "Lung cancer detection using labeled sputum sample: Multi spectrum approach," in International Conference on Industrial, Engineering and Other Applications of Applied Intelligent Systems, 2011, pp. 446-458.

[16] M. Roula, J. Diamond, A. Bouridane, P. Miller and A. Amira, "A multispectral computer vision system for automatic grading of prostatic neoplasia," in Biomedical Imaging, 2002. Proceedings. 2002 IEEE International Symposium On, 2002, pp. 193-196.

[17] M. A. Tahir, A. Bouridane, F. Kurugollu, and A. Amira, "A novel prostate cancer classification technique using intermediate memory tabu search," EURASIP Journal on Applied Signal Processing, vol. 2005, pp. 2241-2249, 2005.

[18] M. A. Tahir and A. Bouridane, "Novel round-robin tabu search algorithm for prostate cancer classification and diagnosis using multispectral imagery," Information Technology in Biomedicine, IEEE Transactions On, vol. 10, pp. 782-793, 2006.

[19] M. A. Tahir, A. Bouridane and M. A. Roula, "Prostate Cancer Classification Using Multispectral Imagery and Metaheuristics," Computational Intelligence in Medical Imaging: Techniques and Applications, pp. 139, 2009.

[20] S. Bouatmane , M.A. Roula, A. Bouridane, and S. Al-Maadeed, "Round-Robin sequential forward selection algorithm for prostate cancer classification and diagnosis using multispectral imagery," Mach Vision Appl, vol. 22, pp. 865-878, 2011.

[21] R. Khelifi, M. Adel and S. Bourennane, "Multispectral texture characterization: application to computer aided diagnosis on prostatic tissue images," EURASIP Journal on Advances in Signal Processing, vol. 2012, pp. 1-13, 2012.

[22] K. Masood and N. M. Rajpoot, "Spatial analysis for colon biopsy classification from hyperspectral imagery," Annals of the BMVA, vol. 2008, pp. 1-16, 2008.