Early Detection of Lung Cancer from CT Scan Images Using Binarization Technique

Fatema Tuj Johora, Mehdi Hassan Jony, Parvin Khatun and Humayan Kabir Rana

October 1, 2018

Early Detection of Lung Cancer from CT Scan Images Using Binarization Technique

Fatema Tuj Johora[1], **Mehdi Hassan Jony**[2], **Parvin Khatun**[3] **and Humayan Kabir Rana**[4]
Department of Computer Science and Engineering, Green University of Bangladesh
[1]*fatema@cse.green.edu.bd,* [2]*shoyshab@gmail.com,* [3]*parvin.do.cse@green.edu.bd and* [4]*humayan@cse.green.edu.bd*

***Abstract*— Lung cancer is one of the most dangerous diseases and prediction of cancer is the most challenging problem nowadays. Most of the cancer cells are overlapped with each other. It is hard to detect the cell but also essential to identify the presence of cancer cells in the early stage unless unable to prevent. Early detection of lung cancer can reduce 14-49% of the death rate. An extensive review for the detection of lung cancer by the former researcher using image processing techniques is presented. Here, We work with CT scan images which are more efficient then X-ray. In this paper, MATLAB is used for processing some images with image enhancement and image segmentation procedures. In our proposed method, Binarization technique is used to alter binary image and then compare it with a threshold value to detect abnormal lung of patients. The performance of the proposed system shows satisfactory results.**

Keywords: Lung Cancer Detection, Gabor Filter, Computed Tomography (CT), Binarization, Feature Extraction.

I. INTRODUCTION

Lung cancer is the maximum mainstream cancer that cannot be neglected. It is the second most regularly analyzed cancer in men and women. An anticipated 234,030 new instances of lung cancer are normal in 2018, representing around 14% of all cancer analyze. The rate has been declining since the mid-1980s in men, however just since the mid-2000s in women. Gender orientation contrasts reflect authentic examples of smoking take-up and end in the course of recent decades. From 2005 to 2014, lung cancer frequency rates decreased by 2.5% per year in men and 1.2% per year in women. Lung cancer represents a larger number of deaths than some other cancer in men and women. An expected 154,050 deaths are relied upon to happen in 2018, representing around 1 of every 4 cancer death. The lung cancer death rate has decayed by 45% since 1990 in men and by 19% since 2002 in women due to reductions in smoking, with the pace of decline quickening over the past decade; from 2011 to 2015, the rate decremented by 3.8% per year in men and by 2.3% per year in women. [1]. Identifying of lung cancer are often done through its symptoms embrace a cough, coughing of blood, fatigue, unexplained weight loss, shortness of breath etc. CT (CT) is that the most efficacious technique of lung nodule detection for its competency to make three-dimensional (3D) pictures of the chest, leading to more immensely colossal resolution of nodules and tumor pathology. A CT image by a computer process to avail lung nodule medicine has been widely utilized in the clinic. the method of computer availed diagnosis (CAD) of lung cancer are often divided into a detection system (often abbreviated as CADe) and diagnostic system (often abbreviated as CADx). The forsaken system divides the candidate nodules kenned within the precedent step into nodules or non-nodules (i.e., traditional anatomic structures). The goal of the CADx system is to relegate detected nodules into benign and malignant nodules [2]. Since the likelihood of malignancy is proximately associated with the geometric size, shape, and look, CADx will distinguish the benign and malignant pulmonary nodules by the efficacious options like texture, shape, and rate of magnification. Thus, the prosperity of a culled CADx system is often quantified in terms of precision of identification, haste, and automation level [3].

In this study, MATLAB has been used through every procedure made. Image enhancement, segmentation, and feature extraction procedures have been applied by MATLAB. The paper is aiming to get the more accurate results by using various enhancement and segmentation techniques and introducing a new feature extraction process which is much effective and time-consuming to detect affected cell of lung cancer.

II. RELATED WORKS

Various initiatives are frequently developed since the middle of the 18th century for detecting cancer S. Long[4]. Image processing approaches brought huge benefit for identifying Lung Cancer at an early stage. According to this paper discussing Magnetic resonance imaging (MRI), CT scan and image processing where is using image enhancement, image segmentation and related to works. Magnetic resonance imaging (MRI), which produces computerized images of the body that are used to diagnose several different cancers, was developed by Paul Lautenberg and Peter Mansfield [5]. The imaging is based on the different signals obtained from different tissue types that are subjected to a high magnetic field. In 2003 the developers won the Nobel Prize in medicine for their work.

Radiologists have interior relied on examining images from chest radiography and PET scans to detect lung cancer [6]. However, advancements in computed tomography (CT) in the 21st century made it a more auspicious implement in both resolution and speed [7]. The manual detection of solid and sub-solid pulmonary lesions in thoracic CT scans is quite error-prone, with a categorically high erroneous-negative rate for detecting minuscule nodules due to, for example, their size, density, location, and conspicuousness. To ameliorate the hand annotation of nodules, medical experts expand beyond an axial scan mode and rely on other techniques such as maximum intensity projections and 3D volume renderings [8]. Maximum intensity projection (MIP) is a volume rendering technique for 3D images that projects voxels with a maximum intensity of the parallel rays from a given viewpoint onto the plane [9]. This technique makes it more facile to detect denser objects like nodules since maximum projections will be concentrated in a particular area, whereas other structures like thin blood vessels will have maximum intensities more distributed throughout the lung/image.

In [10] the authors largely fixate on adequately great amelioration in contrast of masses along with the suppression of background tissues is acquired by processing the parameters of the proposed transformation function in the definite range. The manual analysis of the sputum samples is time-consuming, erroneous and needs concentrated trained person to evade diagnostic mistakes. The segmentation results will be utilized as a base for a Computer Availed Diagnosis system for early detection of cancer, which increases the chances of the patient survival. As well as Gabor filter is proposed by authors for enhancement of medical images. It is a good enhancement to implement for medical images.

Thus, Researchers are becoming more and more concerned with the elaboration of automated CAD systems for lung cancer. Many publications proposed different automated nodule recognition systems utilizing image processing, and including, different techniques for segmentation, feature extraction, and relegation. Moreover, the detection processes work in the middle stage of the lung cancer, which is very hazardous for the affected patient [11].

III. METHODOLOGY

The methodology focuses to obtain a more accurate result. In the proposed method, it is enhancing the contrast of the input image through the pre-processing method. It is done by first converting the input image to a grayscale image. Images are modified with the Gabor filter to enhance the image contrast. Frequency and orientation representations of Gabor filters are homogeneous to those of the human visual system, and they have been found to be concretely agreeable for texture representation and discrimination [12].

In image processing, segmentation is moreover very important part, in this section ideally enhanced images is segmented into two or more sub-segments that will be the more easier representation of the images for further extracting the elements. Segmentation could be utilized for object recognition, occlusion boundary estimation within motion or binaural systems. The goal of image segmentation is to cluster pixels into relevant image regions, corresponding to individual surfaces, objects, or natural components of objects. We mainly follow an approach like a threshold. In feature extraction, a pixel comparing process is proposed in this paper.

The first stage of methodology starts with taking a collection of CT images (normal and abnormal) from the available Database from IMBA Home (VIA-ELCAP Public Access) [12].

We can see the various steps of lung cancer detection in the following fig 1.

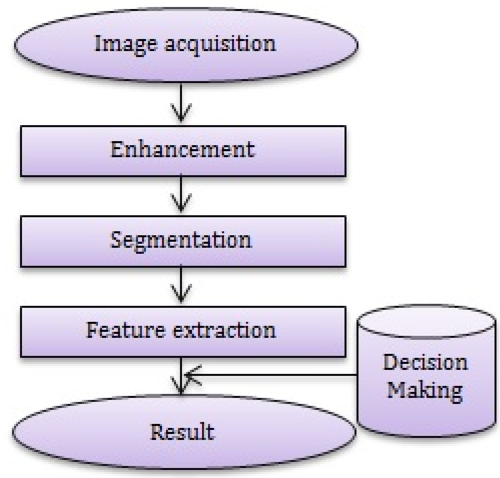

Fig. 1. Lang Cancer Detection Steps

A. Image Acquisition

An image is an artifact that delimitate visual perception that has a similar appearance to some subjects usually a physical object or a person. The first step of pre-processing is collected capture image of the human lung to create a visual representation of some organs and tissues.

An image can be captured by the optical device include cameras mirrors, lenses, telescopes, microscopes etc.

There have several medical image capture modalities such as, Tactile imaging, Magnetic resonance imaging (MRI), Computed tomography (CT), d) Positron emission tomography (PET), e) Single-photon emission computed tomography (SPECT), etc

B. Image Enhancement

The image pre-processing stage commences with image enhancement. The aim of image enhancement is to modernize the interpretability or cognizance of information included in the image for human viewers or to provide better input for other automated image processing techniques. Image enhancement techniques can be divided into two broad categories: Spatial domain methods and frequency domain methods [13].

1) Spatial domain techniques- which operate directly on pixels.
2) Frequency domain techniques- which operate on the Fourier transform of an image.

The various image enhancement techniques can be characterized as spatial domain methods and frequency domain methods. This includes smoothing of image and removal of noises, blurring etc. There are many ways to enhance the image but Gabor filter was found to be suitable for both the CT and MRI images. In the image enhancement stage, we utilized the following Gabor filter techniques.

Gabor Filter: A Gabor filter is a linear filter whose impulse replication is defined by a harmonic function multiplied by a Gaussian function. Because of the multiplication-convolution property, the Fourier transform of a Gabor filters impulse replication is the convolution of the Fourier transform of the harmonic function and the Fourier transform of the Gaussian function [14]. It fundamentally analyses whether there is any concrete frequency content in the image in concrete directions in a localized region around the point or region of analysis. A set of Gabor filters with different frequencies and orientations may be helpful for extracting useful features from an image. In the discrete domain, 2D Gabor filters are given by,

$$G_c[i,j] = Be^{\frac{i^2+j^2}{2\alpha^2}}\cos(2\Pi f(i\cos\theta+j\sin\theta))........(1)$$
$$G_s[i,j] = Ce^{\frac{i^2+j^2}{2\alpha^2}}\sin(2\Pi f(i\cos\theta+j\sin\theta))........(2)$$

where B and C are normalizing factors to be determined. 2-D Gabor filters have rich applications in image processing, especially in feature extraction for texture analysis and segmentation. f defines the frequency being looked for in the texture. By varying θ , we can look for texture oriented in a particular direction. By varying α , we change the support of the basis or the size of the image region being analyzed.

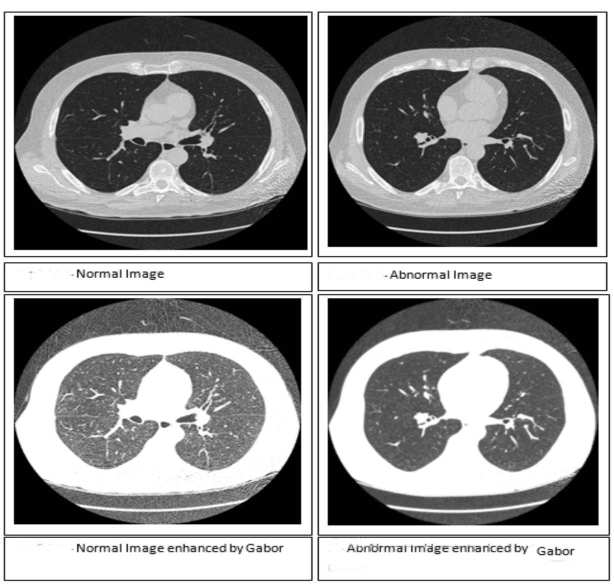

Fig. 2. Enhanced Image by Using Gabor Filter

C. Image Segmentation

Image segmentation is to facilitate the representation of an image into something that is more meaningful and easier to analyze. Image segmentation is used to allocate objects and boundaries. It is the process of assigning a label to every pixel in an image such that pixels with the same label share certain visual characteristics [15]. The result of image segmentation is a set of segments that collectively cover up the entire image or a set of contours extracted from the image. Each pixel of the image in a region is similar with respect to some characteristic or computed property, such as color, intensity, or texture, adjacent regions are significantly different with respect to the same characteristics [16]. In this research, we used Threshold and Watershed image segmentation approaches:

1) Marker-Controlled Watershed Approach: The Watershed transformation considers the gradient enormity of an image as a topographic surface. Pixels having the highest gradient enormity intensities (GMIs) match up to watershed lines, which represent the region boundaries. Water placed on any pixel roofed by a flow downwards to a common local intensity minimum (LMI). Pixels draining to a common minimum form a catchment basin, which signify the regions.

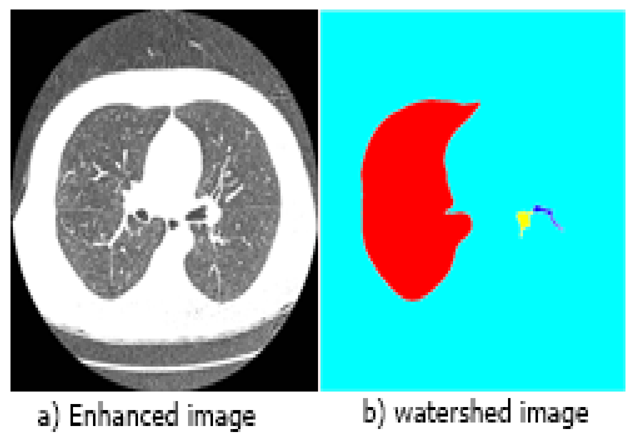

Fig. 3. Segmented Image Using Marker-Controlled Watershed

2) Threshold: The simplest method of image segmentation is called the threshold method. This method is based on a clip-level to turn a grayscale image into a binary image. The simplest property that pixels in a region can apportion is intensity. So, a natural way to segment such regions is through the threshold, the disseverment of light and dark regions. We use here Otsu's threshold method. Otsu's method is based on threshold selection by numerical criteria. This method gives satisfactory results for bimodal histogram images[17]. Threshold values based on this method will be between 0 and 1, after achieving the threshold value; the image will be segmented based on it.

D. Feature Extraction

In image processing, the Image features extraction stage is very vital image processing techniques which utilizing algorithms and techniques to detect and isolate portions or shapes (features) of an image [18]. Once the segmentation

Algorithm 1 Otsu's Threshold algorithm

procedure OTSU'S THRESHOLD
  Calculate histogram
  Calculate possibilities of every intensity section
  initial weight $h_i(0)$ & mean $p_i(0)$
  Increasing the threshold value from $t = 1, 2...$ for maximize the intensity
  Update $h_i(0)$ & $p_i(0)$
  Calculate variance $\alpha_2^b(t)$
  Achieving the threshold the maximum variance $\alpha_2^b(t)$.

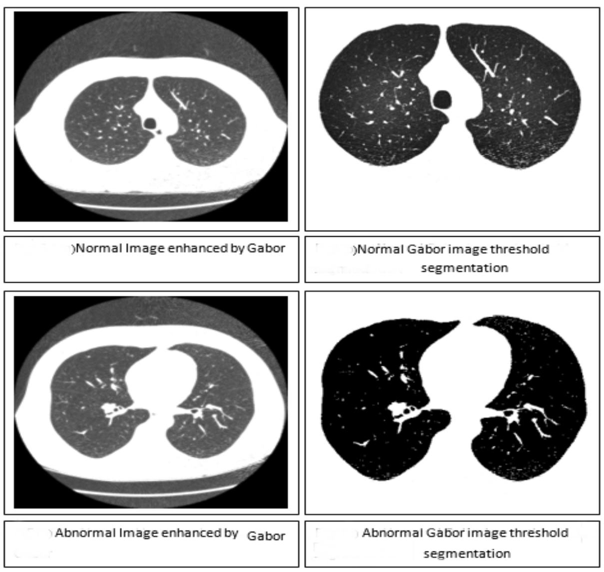

Fig. 4. Enhanced Image using Threshold Process

is accomplished on the lung region, the features can be obtained from it and the diagnosis rule can be designed to precisely detect the cancer nodules in the lungs. This diagnosis rules can eliminate detection of cancer nodules resulted in segmentation and provides a better diagnosis [19-21]. In this stage to obtain the general features of the enhanced segmented image using Binarization.

Binarization: Binarization is a process where comparing the values of the white pixel or black pixel with a comparative value. This paper is using Otsu's threshold for segmentation. After segmentation, the images have divided into black and white pixel only. Now after verifying normal lungs with a threshold, the number of the black pixel are 20148.48, is the threshold value for detection normal lung. If the testing images pixel is lower than this value it is called affected lung otherwise normal lung.

IV. RESULT AND DISCUSSION

The new technique of binarization has been presented is more effective than threshold value binarization.
Success Rate:
True positives (TP): These refer to the positive tuples that

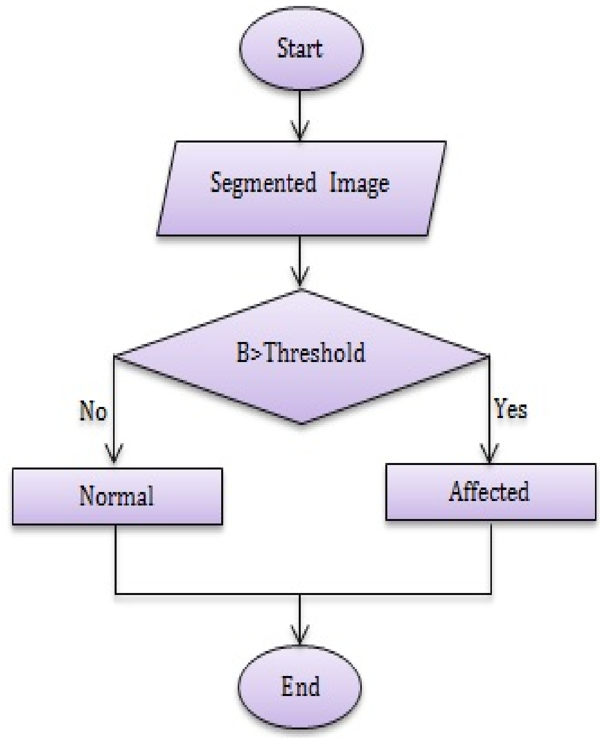

Fig. 5. Flow Diagram of Binarization Processing

were correctly labeled by the classifier. Let TP be the number of true positives.
True negatives (TN): These are the negative tuples that were correctly labeled by the classifier. Let TN be the number of true negatives.
False positives (FP): These are the negative tuples that were incorrectly labeled as positive (e.g., tuples of class buy computer = no for which the classifier predicted buys computer= yes). Let FP be the number of false positives.
False negatives (FN): These are the positive tuples that were mislabeled as negative (e.g., tuples of class buy computer = yes for which the classifier predicted buys computer= no). Let FN be the number of false negatives [21].
True positive rate = TP/(TP+FN)
True negative rate = TN/(TN+FP)
(Total) Accuracy = (TP+TN)/(TP+FN+TN+FP) positive predictive value = TP/(TP+FP)
Negative predictive value= TP/(TN+FN)
Now the dataset success rate is being compared with ISSN 1583-1078 which has same dataset with threshold binarization process. Patient one is the result with same patients' MRI analysis report is done by this paper and subject two is the result with same patient done by the following paper.

V. CONCLUSIONS

An image improvement technique is just beginning for earlier disease detection and treatment stage. Image quality and accuracy are the core factors of this research, image quality assessment as well as the enhancement stage where

TABLE I

DATA SET COMPARISON TEST

Patient	True Acceptance Rate	False Acceptance Rate	Success Rate
1	97.73%	80%	96%
1[12]	92.86%	7.14%	81.625%
2	96.23%	83%	95.12%
2[12]	91.47%	14.12%	82.03%
3	97.14%	79.23%	95.63%
3[12]	93.42%	29.34%	83.63%
4	95.66%	75.23%	95.71%
4[12]	89.93%	21.23%	81.21%

were adopted on low pre-processing techniques based on Gabor filter within Gaussian rules. Based on related works, its clear that researchers mainly focus on CAD system for lung cancer. But our proposed technique is efficient for segmentation principles to be a region of interest foundation for feature extraction obtaining. The proposed procedure gives very promising results compared with other used techniques. Relying on general features, a normality comparison is made. The main detected features for accurate images comparison are pixels percentage and mask-labeling with high accuracy and robust operation.

REFERENCES

[1] A. Staff, " *Cancer Facts & Figures 2018,*" Atlanta: American Cancer Society, Cancer, pp. 19-20, 2018.

[2] L. Sali, S. Delsanto, D. Sacchetto, L. Correale, M. Falchini, A. Ferraris, G. Gandini, G. Grazzini, F. Iafrate, G. Iussich, L. Morra, A. Laghi, M. Mascalchi and D. Regge, "*Computer-based self-training for CT colonography with and without CAD*", European Radiology, 2018.

[3] Q. Song, L. Zhao, X. Luo and X. Dou, "*Using Deep Learning for Classification of Lung Nodules on Computed Tomography Images*", Journal of Healthcare Engineering, vol. 2017, pp. 1-7, 2017.

[4] S. Long and M. Cohen, "*Classics in cytology VI: The early cytologic discoveries of Lionel S. Beale*", Diagnostic Cytopathology, vol. 9, no. 5, pp. 595-598, 1993.

[5] T. Jones, D. Duquette, M. Underhill, C. Ming, K. Mendelsohn-Victor, B. Anderson, K. Milliron, G. Copeland, N. Janz, L. Northouse, S. Duffy, S. Merajver and M. Katapodi, "*Surveillance for cancer recurrence in long-term young breast cancer survivors randomly selected from a statewide cancer registry*", Breast Cancer Research and Treatment, vol. 169, no. 1, pp. 141-152, 2018.6, 1993.

[6] H. Soda, S. Kohno, M. Oka and H. Tomita, "*Limitation of annual screening chest radiography for the diagnosis of lung cancer. A retrospective study*", Cancer, vol. 72, no. 8, pp. 2341-2346, 1993.

[7] S. Sone, F. Li, Z. Yang, S. Takashima, Y. Maruyama, M. Hasegawa, J. Wang, S. Kawakami and T. Honda, "*Characteristics of small lung cancers invisible on conventional chest radiography and detected by population based screening using spiral CT.*", The British Journal of Radiology, vol. 73, no. 866, pp. 137-145, 2000.

[8] I. Valente, P. Cortez, E. Neto, J. Soares, V. de Albuquerque and J. Tavares, "*Automatic 3D pulmonary nodule detection in CT images: A survey*", Computer Methods and Programs in Biomedicine, vol. 124, pp. 91-107, 2016.

[9] S. Napel, M. Marks, G. Rubin, M. Dake, C. McDonnell, S. Song, D. Enzmann and R. Jeffrey, "*CT angiography with spiral CT and maximum intensity projection.*", Radiology, vol. 185, no. 2, pp. 607-610, 1992.

[10] H. Chien and D. Mackay, "*How much complexity is needed to simulate watershed streamflow and water quality? A test combining time series and hydrological models*", Hydrological Processes, vol. 28, no. 22, pp. 5624-5636, 2013.

[11] P. P.Sundari, "*Analysis and Classification of Skin Lesions Using 3D Volume Reconstruction*", IOSR Journal of Computer Engineering, vol. 10, no. 2, pp. 35-40, 2013.

[12] M. S. AL-TARAWNEH, "*Lung Cancer Detection Using Image Processing Techniques*", Leonardo Electronic Journal of Practices and Technologies, no. 1583-1078, pp. 147-158, 2012.

[13] H. Rutika r, "*Automated detection and diagnosis from lungs CT scan images*", International Journal of Emerging Technologies and Innovative Research (www.jetir.org — UGC and issn Approved), ISSN:2349-5162,, vol. 2, no. 3, pp. pp785-787, 2015.

[14] S. Avinash, K. Manjunath and S. Kumar, "*An improved image processing analysis for the detection of lung cancer using Gabor filters and watershed segmentation technique*", 2016 International Conference on Inventive Computation Technologies (ICICT), 2016.

[15] M. Firmino, G. Angelo, H. Morais, M. Dantas and R. Valentim, "*Computer-aided detection (CADe) and diagnosis (CADx) system for lung cancer with likelihood of malignancy*", BioMedical Engineering OnLine, vol. 15, no. 1, 2016.

[16] H. Ren, Y. Zhou and M. Zhu, "*Tree Image Segmentation Based on an Improved Two-Dimensional Otsu Algorithm*", International Journal of Hybrid Information Technology, vol. 9, no. 9, pp. 199-210, 2016.

[17] H. Ng, "*Automatic thresholding for defect detection*", Pattern Recognition Letters, vol. 27, no. 14, pp. 1644-1649, 2006.

[18] Rana HK., Azam MS. and Akhtar MR. "*Iris Recognition System Using PCA Based on DWT*", SM Journal of Bio-metrics & Bio-statistics, vol. 2, no. 3, 2017.

[19] Y. Yuan, J. Wang, B. Li and M. Meng, "*Saliency Based Ulcer Detection for Wireless Capsule Endoscopy Diagnosis*", IEEE Transactions on Medical Imaging, vol. 34, no. 10, pp. 2046-2057, 2015.

[20] X. Jia and M. Meng, "*A deep convolutional neural network for bleeding detection in Wireless Capsule Endoscopy images*", 2016 38th Annual International Conference of the IEEE Engineering in Medicine and Biology Society (EMBC), 2016.

[21] S. Suman, F. hussin, A. Malik, K. pogorelov, M. Riegler, S. Ho, I. Hilmi and K. Goh, "*Detection and Classification of Bleeding Region in WCE Images using Color Feature*", Proceedings of the 15th International Workshop on Content-Based Multimedia Indexing - CBMI '17, 2017.