Automated Detection of Polyps in CT Colonography images using Deep Learning Algorithms in Colon Cancer Diagnosis

Akshay M Godkhindi
Department of Information Science & Engineering.
Ramaiah Institute of Technology
Bengaluru, India
msritakshay@gmail.com

Rajaram M. Gowda
Department of Information Science & Engineering.
Ramaiah Institute of Technology
Bengaluru, India
raj.gowda@gmail.com

***Abstract*— Colon cancer is cancer that is present on the inner side of colon walls or the rectum walls in the large intestine. Most of these types of cancer begin as abnormal growth of tissue called as polyp. Colonography uses low dose radiation Computed tomography (CT) scanning to obtain an interior view of the colon making use of special x-ray machine to view the large intestine for cancer and abnormal growths known as polyps. Radiologists examine these images to find polyp like structure using computer tools. As CT Colonography image contain noise such as lungs, small intestine, instruments during image capturing; segmenting colon from noise is the key task. Polyp occurrence can be detected mainly using shape feature; eliminating shapes similar to polyp is challenging. Hence, to tackle above issues, Image processing techniques are used by applying deep learning algorithm – Convolution Neural Network (CNN) and the results are compared with classical machine learning algorithm. In proposed method, each image is pre-processed to filter air filled dark region that includes colon, lungs etc. Next, each pre-processed CT Image separated into fixed number of blocks. Using pre-trained CNN, each block of ROI is classified as Type 1 (Usually Ascending and descending colon), Type 2 (Usually Traversal and sigmoidal colon) and Type3 (Noise such as lungs, instruments) to segment colon blocks by eliminating noise. Classified Blocks is further diagnosed for polyp like structure using pre-trained CNN by classifying each colon block as normal or abnormal. The experiment is setup with classical machine learning algorithms - Random Forest (RF) and k-nearest neighbor (KNN) by extracting texture feature - Local binary pattern (LBP) and shape feature - Histogram oriented gradient (HOG) for comparison. The experiment results showed the accuracy of proposed method for colon segmentation using CNN (87%) outperforms RF (85%) and KNN (83%). In, addition, the polyp detection accuracy of CNN (88%) is better than Random forest (85%) and KNN (80%). Hence, in the proposed method, there is significant accuracy improvement using deep learning algorithm compared to classical machine learning algorithms. It also provide baseline for automated colon cancer diagnosis using Deep learning algorithms for further research.**

Keywords— Colon Cancer, CT Colonography, Convolution neural network.

I. INTRODUCTION

Colon cancer is the third leading cause of cancer death in both men and women in the United States and also the third most commonly diagnosed cancer. The American Cancer Society (ACS) estimates that 136,830 people diagnosed with colorectal cancer and 50,310 people die from the disease in 2014. By applying existing knowledge about cancer prevention, increasing the use of recommended screening tests, accurate diagnosis and ensuring that all patients receive timely, standard treatment all major type of cancers and deaths could be prevented.

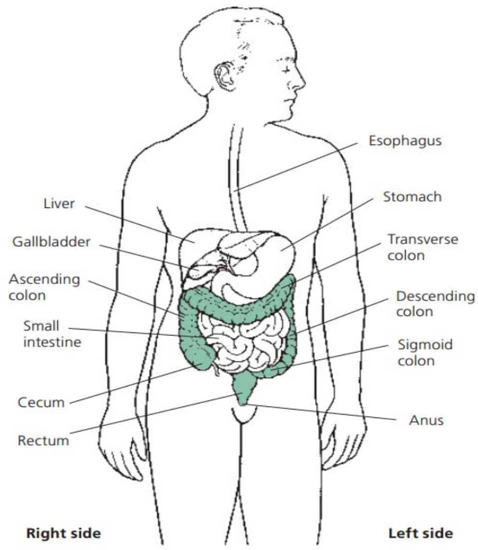

Figure 1 Anatomy of the Colon & Rectum [2]

The colon is a muscular tube about 5 feet long and is longest part of the large intestine. Main function of the colon is to absorb water and mineral nutrients from the food matter. Waste left from this process passes into the rectum [3]. There are four section of colon [2]. There are four sections of colon, viz., the first section - ascending colon begins with a pouch called the cecum; the second section - transverse colon crossing the body from the right to the left side, the third section - descending colon descending on the left side and the fourth section – sigmoid colon joining the rectum, which connects to the anus (see Figure 1).

978-1-5386-1887-5/17/$31.00 ©2017 IEEE

A. Polyp Description

Colorectal cancer usually found in the colon or the rectum. This cancer develops more often in the colon or rectum than in the small intestine [4]. It usually takes 10 to 20 years for this type of cancer to develop slowly. In the inner lining of the colon or rectum a noncancerous growth called a polyp that develops over a period of time. Polyps are small protruding mounds that may occur throughout the intestinal system (see Figure **2**). The polyps that are located in the large bowel, or colon, are referred to as colorectal polyps [5].

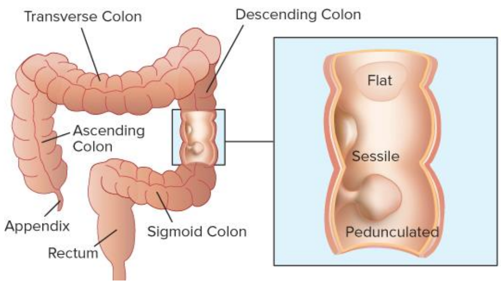

Figure 2 Polyp in Colon

The most common types of polyps are hyperplastic and adenomatous polyps [6]. Adenomatous polyps (adenomas) sometimes change into cancer and are called a pre-cancerous condition. Hyperplastic polyps and inflammatory poly are not pre-cancerous and are more commonly found. Polyps that contain cancerous cells are known as malignant polyps. Unfortunately, there is a chance of gradual development into malignancy, and, this chance is related to its size. It is generally accepted by medical fraternity that polyps with a diameter smaller than 6 mm require no further action, and whereas polyps equal to and larger than 10 mm should be removed by colonoscopy and polyps in between 6 mm and 10 mm can either be treated or removed.

B. CT Colonography.

The diagnosis of Colon cancer is done using medical test such as CT Colonography and colonoscopy. CT colonography is widely used because of less risk and other clinical factors. Computed tomography(CT) (see **Figure 3**) is a diagnostic medical test that, like traditional x-rays as in **Figure 4**, produces multiple images of the inside of the body [7]. CT Colonography uses low dose radiation scanning to obtain an interior view of the colon (the large intestine) that is otherwise only seen with a more invasive procedure with endoscope. The major reason for performing CT Colonography is to find abnormal growths called polyp in their early stages, so that they can be removed before cancer has a chance to spread.

Figure 3 CT Scanning Device [8]

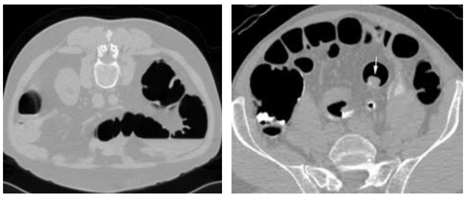

Figure 4 CT Colonography DICOM Images

C. Key Issues

Accurate detection of any Computer aided diagnosis (CAD) plays a major role in radiological diagnosis. At present the detection of abnormality via CT often occurs using manual prospective visual inspection of every image slice which may be in thousands. This is very time consuming complex process and prone to manual error. As radiologist experts are needed to detect polyps, it's very hard for people in remote areas (rural areas) in developing and underdeveloped countries to connect the experts every time. Due to lack of experts and advanced automated tool, the pre-cancerous and cancerous cells goes undetected in particular time span and ultimately causes death of patient. Other issues are, as CT Colonography image contain noise such as lungs, small intestine and instruments during image capturing, segmenting colon form noise is key task [10]. And also polyp occurrence can be detected mainly using shape feature eliminating shapes similar to polyp is challenging. Another issue, specific to automated detection, is false detections on either the ileocecal valve, or the rectal tube, which both have characteristics similar to polyps. Many of present CADe result is low sensitivity and specificity levels due to which they have not been integrated into clinical practice.

Hence there is need of advanced automated CAD tool to segment colon, detect polyps, access it risk, and look for cancerous growth using CT Colonography images by exploring the advanced technologies such as deep learning algorithms and reduce inspection time, without sacrificing detection sensitivity and specificity. Recently, the large amount of annotated training set and advancement in machine leaning in medical imaging, it is feasible to train convolution neural network with available computer configuration and technology.

Our contribution in this work is as follows,

1. A fully automated method to diagnose colon cancer using CT colonography images.
2. Pre-processing of CT images to remove most of noise and unwanted regions and filter only air filled dark regions which include colon, lung, small intestine etc.
3. With help of convolutional neural networks we segment colon from other air filled regions such as lungs, and small intestine which helps to reduce false positive.
4. The polyp can appear anywhere in the colon and have any kind of shape, size, and contrast we use CNN to classify polyp from normal colon blocks.
5. As only few researchers have proposed advantages of deep learning in detection of abnormality in Medical images diagnosis our proposed model provides baseline for further research in this direction.

II. LITERATURE REVIEW

The American Cancer Society's estimates for the number of colorectal cancer cases in the United States for 2016 with 95,270 new cases of colon cancer and 39,220 new cases of rectal cancer 0.

In our previously published survey paper [9], we have reviewed various paper from different journals and conference and highlighted the work of various researchers and contribution in colon cancer diagnosis. From few papers we noted preprocessing and colon segmentation is the main step before polyp detection. The different dataset used by researcher and to note The Cancer Imaging Achieve (TCIA) [11] provides large dataset of CT colonography images with ground truth information and it is publically available for research. As far as our literature survey in [9] only few researchers have explored using deep learning algorithms in colon cancer diagnosis.

III. MATERIAL AND METHODS

In this section we describe methods used in details describes the basic steps involved in medical diagnosis. We have provides graphical user interface to select patient data, select positions of patient in which patient images are captured (supine and prone) [13]. Further steps in explained in sub sections.

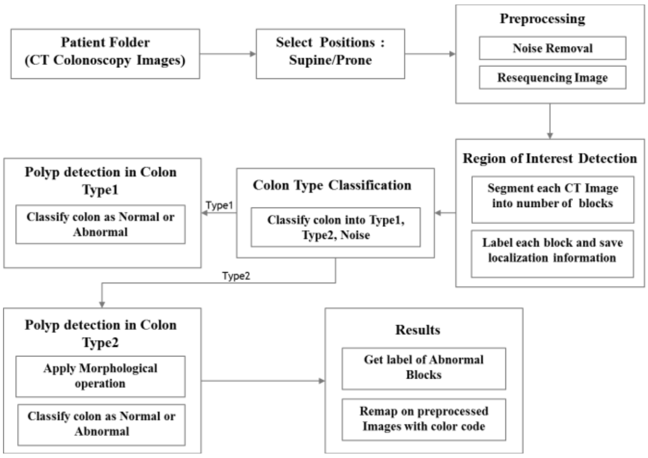

Figure 5 System architecture of proposed method

A. Preprocessing

In this step, as CT colonography is done by inserting low dose carbon from rectum to highlight colon walls. We need to select only regions filled with air. In Figure **6** we can see the dark

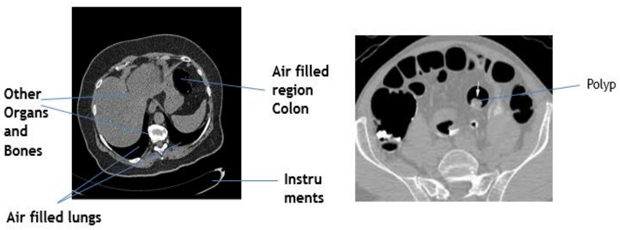

Figure 6 Labeled CT Colonography Image

region clearly visible and distinguish from other organs [14]. We threshold using fixed values and filter only air filled regions from other [15]. In Figure **7** we can results after processing and corresponding three dimension view.

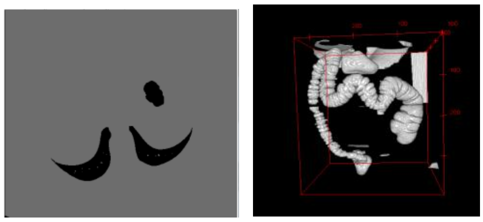

Figure 7 CT Image after Preprocessing

B. Region of Interest Detection

As each CT image size has set of fixed blocks which is air filled regions we segment it into number of blocks by converting each slice into binary image and we label location each block and save it for further processing. These block contains colon, lungs, instruments and other noise (see Figure **8**).

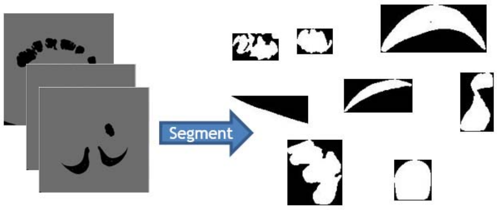

Figure 8 Segmenting CT Images to Blocks

C. Convolution neural network

Convolutional Neural Networks (CNN) has been successfully applied in various medical image diagnosis such as brain tumor diagnosis using MRI images [16]. CNN are made up of neurons that have learnable weights and biases (see Figure **9**). Each neuron receives some inputs, performs a dot product and optionally follows it with a non-linearity [17]. The whole network still expresses a single differentiable score function: from the raw image pixels on one end to class scores at the other.

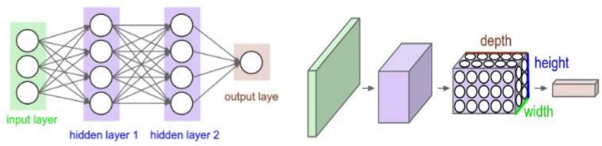

Figure 9 A basic 3 layer Neural Network, Right: Three Dimension Convolution neural network [17]

Two main processes in CNN are convolution and Pooling [18]. Convolution process use a trainable filter Fx, deconvolution of the input image we get feature map, then add a bias bx, we can get convolution layer Cx. A pooling process is n pixels of each neighborhood through steps, become a pixel, and then by scalar weighting $Wx+1$ weighted, add bias $bx+1$, and then by an activation function, produce a narrow n times feature map $Sx+1$ (see Figure **10**).

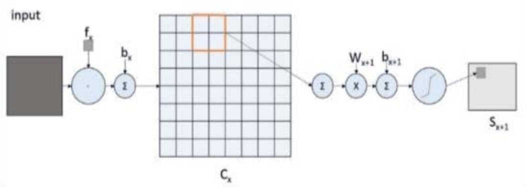

Figure 10 Main process in CNN [18]

Three basic Layers used to build ConvNets are Convolutional Layer, Pooling Layer, and Fully-Connected Layer [17]. INPUT layer will hold the raw pixel values of the image .CONV layer will compute the output of neurons that are connected to local regions in the input, each computing a dot product between their weights and a small region they are connected to in the input volume. RELU layer will apply an elementwise activation function, such as the $max(0, x)$, $mean(0, x)$ thresholding at zero. POOL layer will perform a down sampling operation along the spatial dimensions (width, height) Fully-Connected (FC) layer will compute the class scores, resulting in volume of size, it Usually contains Number of class as output. Figure 11 shows the layer details in CNN with input [60x60] , RELU activation function will filter input image into feature map of [56x56] , with kernel size [5x5], in pooling layer mean function will reduce image dimension [28x28]. Similarly series of layer is setup and finally fully connected layer with tanh activation function and softmax function [19] which has number of classes as output.

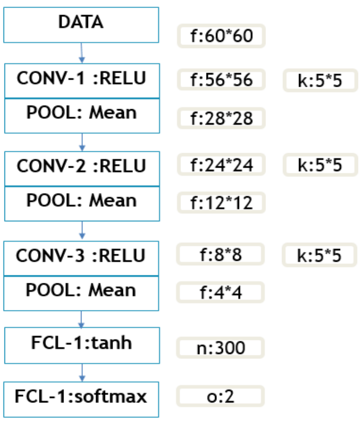

Figure 11 Convolution neural network Setup

D. *Feature Extraction for Classical Machine Learning Algorithms*

Texture - Local Binary Pattern (LBP) [20] is a type of visual descriptor used for classification in computer vision. LBP looks at points surrounding a central point and tests whether the surrounding points are greater than or less than the central point (i.e. gives a binary result). In a local neighbourhood of an input image, given a pixel (x_c, y_c)which is surrounded by 8 neighbors, we can calculate its LBP value by using Equation (1)

$$LBP(x_c, y_c) = \sum_{p=0}^{7} s(i_p - i_c)2^p \quad (1)$$

where i_c indicates the grayscale value of the center pixel(x_c, y_c); i_p; corresponds to the grayscale value of the pth neighbor, $s(x)$ is a sign function where $s(x) = 1;$ $if\ x \geq 0;\ else, s(x) = 0.$ We extract Texture [25] LBP map of each block from previous stages and reduce it into 10 bin histogram and save vector as feature set.

Histogram of Oriented Gradients (HOG) [21] is used to capture edge and shape information, we calculate the gradient map, which is shown to be complementary to the color feature. The essential thought behind the histogram of oriented gradients descriptor is that local object appearance and shape within an image can be described by the distribution of intensity gradients or edge directions. As polyp classification mainly depend on shape feature we extract HOG map for each block and reduce into 16 bin histogram and save vector as feature set.

E. *Classical Machine Learning Algorithms*

We use two classical machine learning algorithms - Random forest (RF) and k nearest neighbor (kNN) to classify colon type and also to detect polyp in each type. Main reasons of using these algorithm is to compare results of CNN and measure performance on same dataset.

Random Forest is an increasingly popular machine learning method [22]. It builds an ensemble of many decision trees trained separately on a bootstrapped sample set of the original data. Each decision tree grows by randomly selecting a subset of candidate attributes for splitting at each node.

k-Nearest Neighbors (kNN) [23] is one of the simplest classifiers, which classifies a new instance by a majority vote of its k nearest neighbors. In this paper, we use the Euclidean distance metric to find the k nearest neighbors. We set value of k =1 and step wise increment.

IV. EXPERIMENTS

A. *Datasets*

We have collected data from "The cancer imaging archive" (TCIA) [11] with ground truth information. There are presently 825 cases in this collection with XLS sheets that provide polyp descriptions and their location within the colon segments. . The dataset contains CT colonography images obtained from CT devices which are DICOM series images taken in both positions supine and prone. Each image has dimension of approximately 525* 525 pixels. As we have setup supervised machine learning algorithms we have prepared dataset manually with help of ground truth provided. Firstly we classify each block into colon type 1(usually ascending/ descending colon), colon type 2 (usually Transverse and sigmoidal colon) and type 3 - noise as shown in Figure 12. Secondly we detect polyp in colon type as in Figure 13. We use morphological operations such as dilate and erode in order to eliminate false positive in Colon type 2.

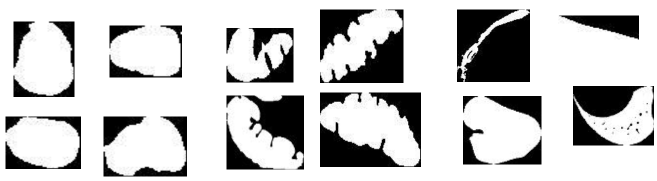

Figure 12 Colon Type1, Type2, and Noise Blocks

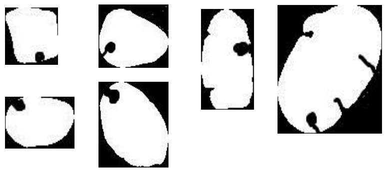

Figure 13 Polyp in type 1 and type 2 colon blocks

Table 1 gives the details of dataset which has colon blocks used for considered for experiment.

Table 1 Dataset description

Image Blocks	Number of Blocks
ColonType1	500
ColonType2	400
Noise	400
Abnormal - Polyp	245
Normal	356

B. Training and Testing

The proposed method and CNN is using MATLAB. Manually annotated colon blocks is stored in database, first we train CNN with labeled blocks with layer configuration as in Figure 11, and save CNN trained model in .MAT format. We have selected thirty percent of dataset for testing. Similarly we detect polyp in each type of colon. With extracted feature LBP-HOG we train RF and kNN using open source tool WEKA [24] and using 10 Fold cross validation results are obtained. We compare these results with CNN. The setup is shown in Figure 14.

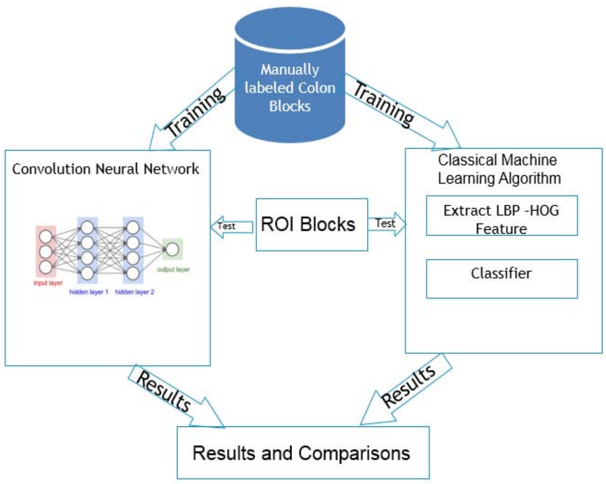

Figure 14 Supervised machine learning Setup

V. RESULTS

The Measures used evaluate performance of algorithms on selected dataset is using equation (2), (3) and (4).

$$\textbf{Accuracy} = \frac{\textbf{Number of correctly classified}}{\textbf{Total Number of Test Sample}} * \textbf{100} \quad (2)$$

$$\textbf{Sensitivity} = \frac{\textbf{True Positives}}{\textbf{True Positives+False Negative}} * \textbf{100} \quad (3)$$

$$\textbf{Specificity} = \frac{\textbf{True Negative}}{\textbf{True Negative+False Positive}} * \textbf{100} \quad (4)$$

Where True Positive are correctly identified as polyp, false positives are incorrectly identified as polyp, False Negative are incorrectly denied and True Negative are correctly identified as normal. After training the labeled blocks on CNN we save the model, test it on test data and the error obtained after several iterations is shown Figure **15** for colon type classification and Figure **16** for polyp detection. For Random Forest and K Nearest neighbor we use weka tool to train and test the LBP-HOG feature set extracted and results are saved.

Table 2 describe the preliminary results of CNN, RF, and kNN, We have also used other machine learning algorithms such as Support Vector Machine (SVM), Logistic regression, as accuracy as less than 75% we have excluded in results. From results we can infer CNN outperforms RF and kNN both in Colon Type classification and Polyp detection for our prepared dataset. In [12] Sensitivity of CT Colonography for Detection of Large Adenomas is 90%, Per-polyp sensitivity for large adenomas or cancers was 0.84 ± 0.04. Per-patient sensitivity estimates in detecting patients with adenomas ≥ 6 mm, was 78%. Our proposed algorithm per polyp sensitivity is 89.77% for polyp ≥ 6 mm.

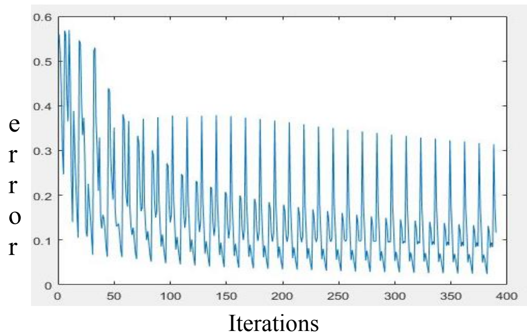

Figure 15 Test Error in Colon Type Detection

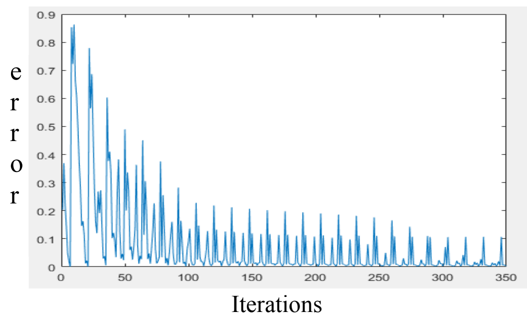

Figure 16 Test Error for polyp detection

Table 2 Comparison of Algorithm results

Algorithm	Colon Type Detection	Polyp Detection In Colon		
	Accuracy (%)	Sensitivity (%)	Specificity (%)	Accuracy (%)
Convolution Neural Network	87.03	88.77	87.35	88.56
Random Forest	84.76	80.06	89.68	85.37
K Nearest Neighbor	82.92	83.15	88.88	80.30

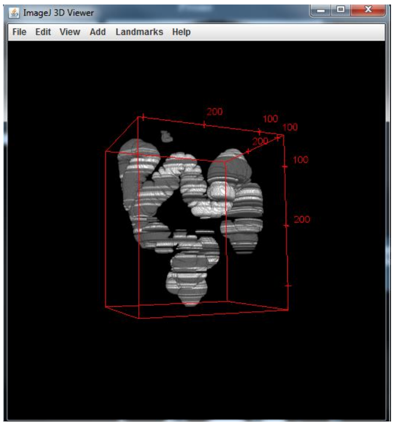

Figure 17 Colon Diagnosis Results - Abnormal Region (White Color)

The Figure **17** Colon Diagnosis Results - Abnormal Region (White Color)Figure 17 shows the end result of abnormal patient of the proposed method with abnormal region marked with white color and normal with gray color. Figure 18 is the colon which is completely normal.

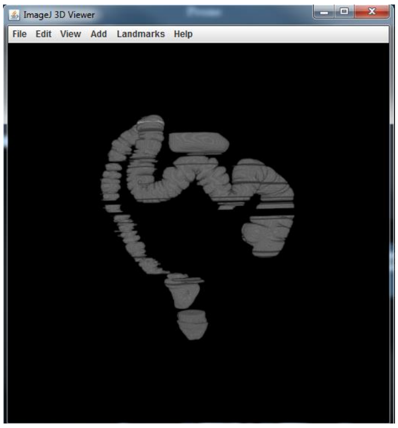

Figure 18 Colon Diagnosis Results - Normal Region (Grey Color)

VI. CONCLUSION AND FUTURE SCOPE

In this paper a fully automated colon cancer diagnosis with CT colonography images using image processing and deep learning algorithms is designed and implemented. CT Colonography image contain noise such as lungs, small intestine, instruments during image capturing; segmenting colon from noise was the key issue. Another challenge was to eliminate shapes similar to polyp, as its occurrence is detected mainly using shape feature. In step wise approach a pre-processing on CT images is performed and noise is eliminated so that overall accuracy is improved. The deep learning algorithm - convolution neural network is explored and used for classification at various levels. Classifying colon type as Type 1, Type 2 and Type 3 helped to eliminate additional noise. Type 2 colon blocks are prone to have high false positives, so additional Image processing is performed before polyp detection. Experiment is carried on same dataset to extract LBP-HOG feature, then train and test on Random forest and k–nearest neighbor. Results obtained are used for comparison. The experiment results showed the accuracy proposed method for colon segmentation using CNN (87%) outperforms RF (85%) and KNN (83%). In addition, the polyp detection accuracy of CNN (88%) is better than Random forest (85%) and KNN (80%), also Sensitivity of polyp detection using CNN (88%) is quite higher than RF (80%) and KNN (83%). Hence, conclusion of experiment results is- CNN classification performance in both colon segmentation and polyp detection is better than RF and KNN. In this paper the preliminary results of low level classification of CNN, RF and KNN on relatively medium size dataset is obtained. Future scope of this project is to train these algorithm on large dataset and also patient level accuracy on different dataset can be performed. For reducing the execution time graphical processing unit (GPU) can be incorporated in CNN processing. The approach used in project will provide baseline for further researches in colon cancer diagnosis using deep learning algorithms. The proposed method can also be explored in other medical imaging diagnosis such as Lung cancer, Breast cancer and Brain tumor Detection. The accurate detection of pre-cancer growth using automated tools will help the patient to get appropriate treatment well within time, as most of cancer is curable only if it is detected in early stages

References

[1] Alteri, R., J. Kramer, and S. Simpson. "Colorectal Cancer Facts and Figures 2014-2016." Atlanta: American Cancer Society (2014): 1-30.
[2] "Fundamentals of Human Nutrition/Absorption."- Wikibooks, Web. 02 May 2017. https://en.wikibooks.org/wiki/Fundamentals_of_Human_ Nutrition/Absorption
[3] Large Intestine Diagram - IvyRose Holistic : https://http://www.ivyroses.com/HumanBody/Digestion/Large-Intestine-diagram.php
[4] Haggar, Fatima A., and Robin P. Boushey. "Colorectal cancer epidemiology: incidence, mortality, survival, and

risk factors." Clinics in colon and rectal surgery 22.04 (2009): 191-197.

[5] Polyps of the Colon and Rectum : http://www.msdmanuals.com/professional/gastrointestinal-disorders/tumors-of-the-gi-tract/polyps-of-the-colon-and-rectum

[6] Patient education: Colon polyps (Beyond the Basics) https://www.uptodate.com/contents/colon-polyps-beyond-the-basics

[7] CT Colonography: https://www.radiologyinfo.org/en/info.cfm?pg=ct_colo

[8] CT Device: https://info.blockimaging.com/whats-the-best-ct-scanner-for-you

[9] Godkhindi, Akshay., Dr Dayananda, and Dr Sowmyarani. (2016). "A Literature Survey on Computer-Aided Diagnosis in Detection and Classification of Polyp in Colon Cancer using CT Colonography". International Journal of Advanced Research in Computer and Communication Engineering Vol. 5, Special Issue 2, October 2016" IJARCCE - ISSN (Online) 2278-1021.

[10] Ma, Ming, et al. "Random forest based computer-aided detection of polyps in CT colonography." Nuclear Science Symposium and Medical Imaging Conference (NSS/MIC), 2014 IEEE. IEEE, 2014.

[11] Smith K, Clark K, Bennett W, Nolan T, Kirby J, Wolfsberger M, Moulton J, Vendt B, Freymann J. (2015). Data From CT_COLONOGRAPHY. The Cancer Imaging Archive. http://doi.org/10.7937/K9/TCIA.2015.NWTESAY1

[12] Johnson, C. D., Chen, M.-H., Toledano, A. Y., Heiken, J. P., Dachman, A., Kuo, M. D., … Limburg, P. J. (2008, September 18). Accuracy of CT Colonography for Detection of Large Adenomas and Cancers. New England Journal of Medicine. New England Journal of Medicine.

[13] Chen, Sloane C., et al. "CT colonography: value of scanning in both the supine and prone positions." AJR. American journal of roentgenology 172.3 (1999): 595-599.

[14] Yang, Xiaoyun, Xujiong Ye, and Greg Slabaugh. "Multilabel region classification and semantic linking for colon segmentation in CT colonography." IEEE Transactions on Biomedical Engineering 62.3 (2015): 948-959.

[15] Manjunath, K. N., K. Gopalakrishna Prabhu, and P. C. Siddalingaswamy. "A knowledge based approach for colon segmentation in CT colonography images." Signal and Image Processing Applications (ICSIPA), 2015 IEEE International Conference on. IEEE, 2015.

[16] Havaei, Mohammad, et al. "Brain tumor segmentation with deep neural networks." Medical image analysis 35 (2017): 18-31.

[17] Convolutional Neural Networks for Visual Recognition : http://cs231n.github.io/convolutional-networks/

[18] Liu, Tianyi, et al. "Implementation of Training Convolutional Neural Networks." arXiv preprint arXiv: 1506.01195 (2015).

[19] A Beginner's Guide To Understanding Convolutional Neural Networks : https://adeshpande3.github.io/adeshpande3.github.io/A-Beginner's-Guide-To-Understanding-Convolutional-Neural-Networks/

[20] Pietikäinen, Matti, et al. Local binary patterns for still images. Springer London, 2011.

[21] "Histogram of Oriented Gradients (HOG) Descriptor" Intel : https://software.intel.com/en-us/node/529070

[22] Breiman, Leo. "Random forests." Machine learning 45.1 (2001): 5-32.

[23] Peterson, Leif E. "K-nearest neighbor." Scholarpedia 4.2 (2009): 1883.

[24] Holmes, Geoffrey, Andrew Donkin, and Ian H. Witten. "Weka: A machine learning workbench." Intelligent Information Systems, 1994. Proceedings of the 1994 Second Australian and New Zealand Conference on. IEEE, 1994.

[25] Hu, Yifan, et al. "Texture feature extraction and analysis for polyp differentiation via computed tomography colonography." IEEE transactions on medical imaging 35.6 (2016): 1522-1531.