# Novel Adaptive Statistical Method-CNN Synergism Based Two-Step WCE Image Segmentation

Shibarjun Mandal

Registration No: 137292 of 2016-2017

Examination Roll No: M6IAR19002

Under the guidance of

Prof. Sheli Sinha Chaudhuri

DEPARTMENT OF ELECTRONICS AND TELECOMMUNICATION ENGINEERING

Jadavpur University

May 2019

# LIST OF PUBLICATIONS

1. Survey on polyp and ulcer detection techniques in WCE images" Shibarjun Mandal[1], Mou Adhikari[2], Sriparna Banerjee[3], Abhijit Mondal[4], Sheli Sinha Chaudhuri[5]. International Conference on Intelligent Computing and Sustainable System (ICICSS 2018), IEEE Xplore.

2. Novel Adaptive Statistical Method-CNN Synergism Based Two-Step WCE Image Segmentation" Shibarjun Mandal[1], Mou Adhikari[2], Sriparna Banerjee[3], Sheli Sinha Chaudhuri[4]. International Conference on Intelligent Computing and Control Systems (ICICCS 2019), Scopus indexed- IEEE Xplore.

# CONTENTS

# LIST OF FIGURES

**5. CNN Architecture and Classification**

# LIST OF TABELS

# ABSTRACT

Accurate automatic segmentation of anomalies i.e. ulcers, bleeding regions, polyps, etc. from Wireless Endoscopy Images (WCE) are absolutely essential for proper diagnosis of diseases. Using a single segmentation method for segmenting all these above-mentioned anomalies is a very challenging task due to their different shapes and characteristics. Mostly, existing WCE anomaly segmentation methods are dedicated to only one type of anomaly, but the method proposed in this thesis is capable of segmenting four types of anomalies (bleeding regions, small bowel erosions and ulcers, tapeworm and vascular ectasia) from WCE images despite their varied shapes and characteristics. Here a novel adaptive statistical method-Convolution Neural Network (CNN) synergism-based two-step segmentation procedure is proposed. Initially, the labeled regions from the acquired WCE images are removed using a polygonal mask and then converted them from the RGB to HSV color space. Here the HSV color space is chosen to carry out the entire segmentation procedure due to its resemblance with the human vision system. In the first step of the segmentation method the Region of Interest (ROI) from WCE images were extracted using statistical measures and by manipulating the intensities of pixels belonging to the ROI and NON-ROI regions accordingly using two adaptive reinforcement constants. The segmented ROI is obtained from this step itself, but it sometimes gives erroneous results. To overcome this shortcoming, the output images were classified at this stage using AlexNet, GoogLeNet, InceptionV3, and SVM trained with Bag of-Features to obtain accurate results. This stage finally classifies those segmented images into ROI and NON-ROI images. Finally, a segmentation accuracy of 95.49% is obtained using InceptionV3 and also verified the reliability of the method from professional experts.

# 1. Introduction

# 1. Introduction

## 1.1 Endoscopy

Endoscopy is a non-surgical procedure used in medicine to visually examine a person's digestive tract. The endoscopy procedure uses a long, thin tube with a tiny camera at the end i.e. endoscope, inserted directly into the body to observe an internal organ or tissues in detail, unlike many other medical imaging techniques as shown in Fig. 1. Endoscopy was developed for the region in our body where hollow organs were connected to the exterior via natural orifices like the urethra, the vagina, the rectum, the ear canal, and the throat and pharynx.
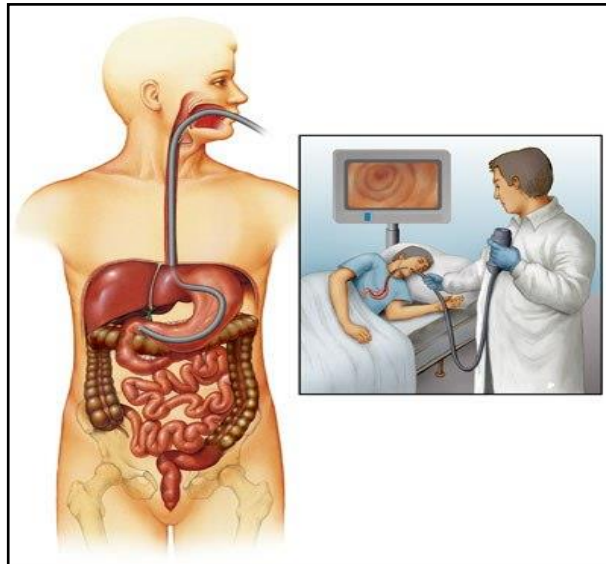


*Fig. 1. Endoscopy*
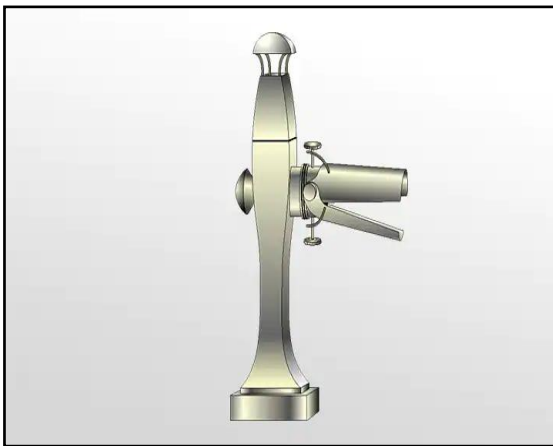


*Fig. 2. Lichteiter invented by Bozzini*

*Fig. 3. Endoscope invented by Desormaux*

The first endoscope was developed in 1806 by Philipp Bozzini in Mainz [1] to examine the canals and cavities of the human body through the tube he created known as Lichtleiter (light guiding instrument) as shown in Fig.2. Next, in 1853, Antoine Jean Desormeaux of France developed an instrument which was named "endoscope" as shown in Fig.3, specially designed to examine the urinary tract and the bladder. It was the first time this term "endoscope" was used in history. After this stage the contribution and invention on endoscopy with particular focus on the engineering concepts and subsequent applications in the medical field are enormous. So few numbers of well-regarded and generally accepted important contribution and its inventors in the early phase of endoscopy and mostly on the technology of the more recent history of endoscopy are summarized below in Table 1.

| Year | Important person(s) or entity | Contribution |
|---|---|---|
| 1806 | Phillip Bozzini | The "Lichleiter", candle illuminated in a container reflecting light by an angled mirror |
| 1853 | Antonin Desmoreaux | First used the term "endoscopy", used kerosene lamp with paraffin flame and 45-degree angled mirror |
| 1858 | Nepomuk Czermak | First endoscopic image |
| 1867 | Julius Bruck | Platinum filament surrounded by water within a glass cooling system for better illumination |
| 1878 | Joseph Swan | Invented lamp produced within a vacuum that was neither hot in temperature nor able to burn out |
| 1879 | Max Nitze | Created cystoscope with water-cooled platinum filament lamp and series of lenses in a metal tube |
| 1879 | Thomas Edison | Invention of incandescent light bulb |
| 1883 | David Newman | Replaced platinum wires with incandescent light bulb on endoscope |
| 1901 | Georg Kelling | First laparoscopy in a dog |
| 1910 | Hans Christian Jacobaeus | First published laparoscopy in a patient |
| 1924 | Richard Zollikofer | First use of $CO_2$ for insufflation |
| 1948 | Harold Hopkins | Developed zoom lens |
| 1953-1954 | Harold Hopkins and Narinder Singh Kapany | Created a bundle of glass fibers to transmit images. Named their instrument the fiberscope |
| 1956 | Basil Hirschowitz and Larry Curtis | Replaced Hopkins'glass fibers with better flexible optical fiber material and glass coating |
| 1959 | Harold Hopkins | Developed rod-lens system |
| 1966 | Harold Hopkins and Karl Storz | Designed new rigid endoscope with self-focusing lens |
| 1969 | Willard Boyle and George Smith | Invented CCD |
| 1971 | Hiromi Shinya and William Wolff | First polypectomy during colonoscopy |
| 1976 | Bryce Bayer | Bayer filter to create the color CCD |
| 1985 | Erich Muhe | First laparoscopic cholecystectomy |
| 1993 | Becker et al. | First report of 3D endoscopic system |
| 1996 | Visionsense Corp. | Development of Visionsense 3D endoscopic system |
| 2000 | FDA | First systems for general robotic surgery approved for use in humans |

*Table 1: Important milestone in the history of Endoscopy (adapted from [1])*

Endoscopy is useful for investigating many sites within the human body like the following areas are [2]:

- Gastrointestinal tract: "esophagogastroduodenoscopy" for esophagus, stomach, and duodenum, "enteroscopy" for the small intestine, "colonoscopy, sigmoidoscopy" for large intestine/colon, "anoscopy" for the anus and "rectoscopy" for bile duct, rectum.
- Respiratory tract: "rhinoscopy" for Nose, "bronchoscopy" for the lower respiratory tract.
- Ear: Otoscopy
- Urinary tract: Cystoscopy
- Female reproductive tract "gynoscopy": "colposcopy" for Cervix, "hysteroscopy" for uterus, "falloposcopy" for fallopian tubes.
- Through a small incision: "laparoscopy" for Abdominal or pelvic cavity, "arthroscopy" for the interior of a joint, "thoracoscopy and mediastinoscopy" for organs of the chest.

The intestine, the upper intestine, food pipe, stomach and a small part of the intestine are visible using an endoscope from the top end, whereas the large intestine can be seen from a colonoscope which is from the bottom end. But it still leaves about 20 feet of the small intestine that cannot examine because the endoscope doesn't reach there. So, the idea of capsule endoscopy was introduced where the patient has to swallow a capsule contains a camera which transmits images of the whole gastrointestinal (GI) tract to the recording device on the patient's body.

Wireless capsule endoscopy [3] is a technology developed for the endoscopic exploration of the entire GI tract. Initially, capsule model endoscope was developed by Given Imaging and approved by the FDA (Food and Drug Administration) in 2001, called the M2A Plus capsule and then remarketed as the PillCam SB. The dimension of Pillcam SB was 11mm × 26mm and weighed 3.7g, as shown in Fig. 4. This capsule covers a 140-degree field of view, 1:8 magnifications, one

to thirty mm depth of view, and a minimum size of detection of about 0.1 mm. The battery inside the capsule lasts around six to eight hours and transmission occurs at two frames per second of $256 \times 256$ pixels color images. Over these subsequent years, this technology has been refined to provide increased battery life, superior resolution, and capabilities to view different parts of the GI tract. At this time there are few companies that manufacture small bowel WCE systems approved by the FDA, namely, PillCam SB2, Endocapsule, Olympus America and MiroCam, as shown in Fig. 5 Table 2.

| WCE company | Size, mm | Weight, g | Field of view | Images/sec | Battery life | Resolution, pixels |
|---|---|---|---|---|---|---|
| EndoCapsule; Olympus America, Inc, Center Valley, Pennsylvania | 11×26 | 3.5 | 145 | 2 | 8 hours | 512×512 |
| PillCam SB2; Given Imaging, Ltd, Yoqneam, Israel | 11×26 | 2.8 | 156 | 2 | 8 hours | 256×256 |
| PillCam SB2EX; Given Imaging | 11×26 | 2.8 | 156 | 2 | 12 hours | 256×256 |
| MiroCam; Intromedic Co Ltd, Seoul, Korea | 11×24 | 3.3 | 170 | 3 | 11 hours | 320×320 |
| PillCam ESO2; Given Imaging | 11×26 | <4 | 169 | 18 | 8 hours | 256×256 |
| CapsoCam Plus, CapsoVision, Inc. | 11×31 | 4 | 360 | 20 | 15 hours | 221×884 |

*Table 2. FDA- approved wireless capsule systems and specifications*

The wireless capsule system consists of three components, (1) capsule endoscope, (2) a sensing system with sensing pads or a sensing belt to attach to the patient, a data recorder, and a battery pack, and (3) a personal computer workstation with proprietary software, as shown in Fig. 6. Generally, all capsule endoscopes have similar components: a disposable plastic capsule, a complementary metal oxide semiconductor or high-resolution charge-coupled device image capture system, a compact lens, white-light-emitting diode illumination sources, and an internal battery source. The mode of data transmission is via ultra-high frequency band radio telemetry (PillCam, EndoCapsule), human body communications (MiroCam), an onboard storage system (CapsoCam Plus).



*Fig. 4. Wireless Capsule Endoscope*

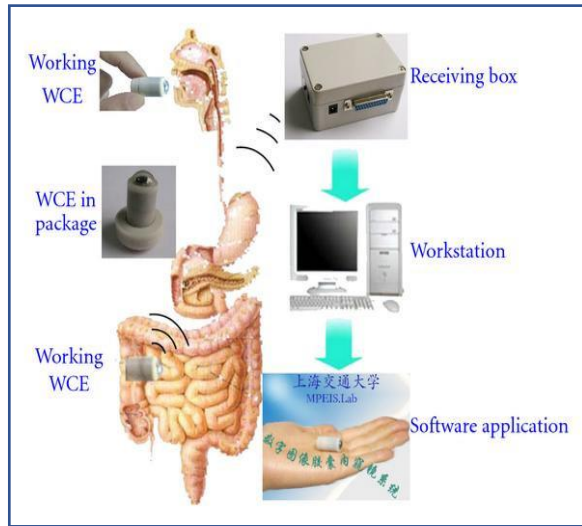Fig. 5. The available wireless capsule endoscope



Fig. 6. Pictorial representation of the WCE method

Capsule endoscopy helps the doctor to evaluate the small intestine since this part of the bowel cannot be reached by traditional upper endoscopy or by colonoscopy. The most common reason for doing capsule endoscopy is to find the cause of bleeding from the small intestine. It is also very useful for detecting polyps, inflammatory bowel disease (Crohn's disease), ulcers, and tumors of the small intestine. Wireless capsule endoscopy is now accepted as the most effective method of examining a patient's entire small intestine. The non-digestible capsule is swallowed and propelled through the food tract by normal peristalsis (the motion that propels food through the intestine). One end of the capsule contains an optical dome with white light emitting diodes (LEDs) as shown in Fig. 4, that illuminate the internal surface of the food tract, and the camera takes pictures of the illuminated region. These images are relayed via a transmitter in the capsule to a small receiver/data-recorder worn by the patient on a belt or stored in the onboard storage system. Batteries in the capsule provide power for around 8 hours, during which time the patient is free to carry out normal activities. At the end of the 8 hours, the data-recorder is removed and the image data uploaded to a workstation/PC computer for later viewing. The stored data consists of approximately 80,000-12,00,000 high-quality images and among them around 40,000 of the images are useful. It is viewed as a color video sequence using special software provided by the manufacturers. The software allows the examiner to travel along and study closely a large region of the interior structure of the GI tract.

An important factor with regards to using the WCE system is that viewing and analyzing one patient's data costs an experienced medical clinician about 2 hours on average. Furthermore, anomalies in the GI tract may be present in only one or two frames of the video, so it might get missed by the physicians due to oversight sometimes. Moreover, there may be some anomalies that cannot be detected by naked eyes because of their size and distribution and in addition, different clinicians may have different findings when they come to the same image. All these problems motivate researchers to develop reliable and uniform assisting methods to reduce the burden of physicians. However, it should be admitted that this goal is very challenging because the true features associated with diseases are not exactly known. Furthermore, different anomalies lead to different diseases, which may have totally different symptoms in the digestive tract, and even the same disease may show great varieties in appearances, shapes, and sizes.

## 1.2 Anomalies in Gastrointestinal (GI) Tract

The standard definition of anomaly is, the abnormality inconsistent with or deviating from what is usual, normal or expected. Based on the standard definition, anomalies in the GI tract are disorders or diseases occurring in the esophagus, stomach, large and small intestines, liver, pancreas and the gallbladder. as shown in Fig. 7 shows some abnormal GI tract frames compared to a normal frame.

4
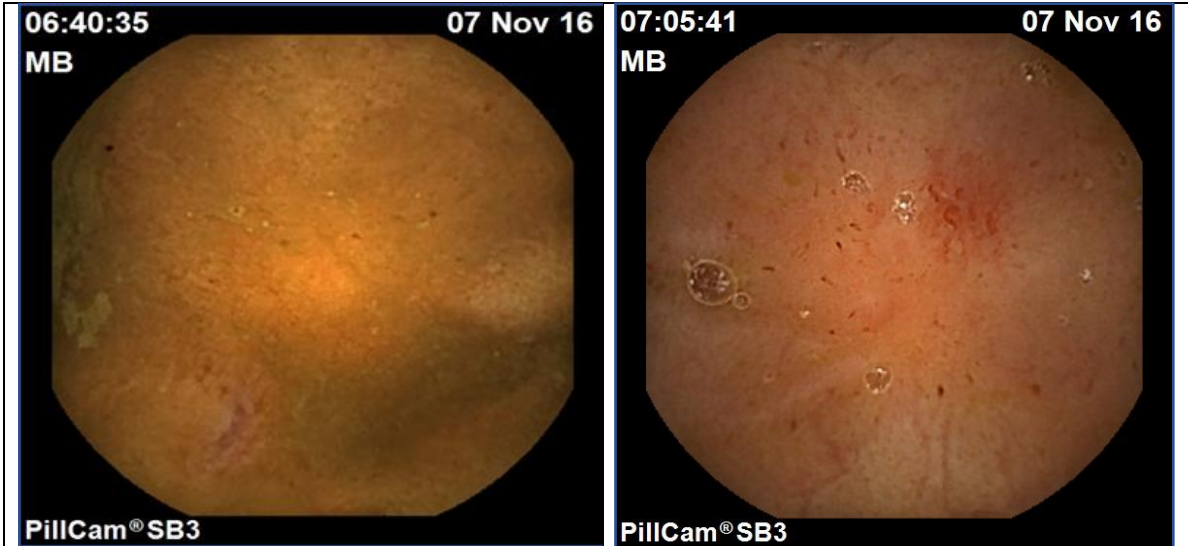
| | |
|---|---|
| Normal Frame | Frame with Erosion and Ulcer |
| Normal Frame | Frame with Tapeworm |
| Normal Frame | Frame with Vascular Ectasia |

| Normal Frame | Frame with Bleeding region |
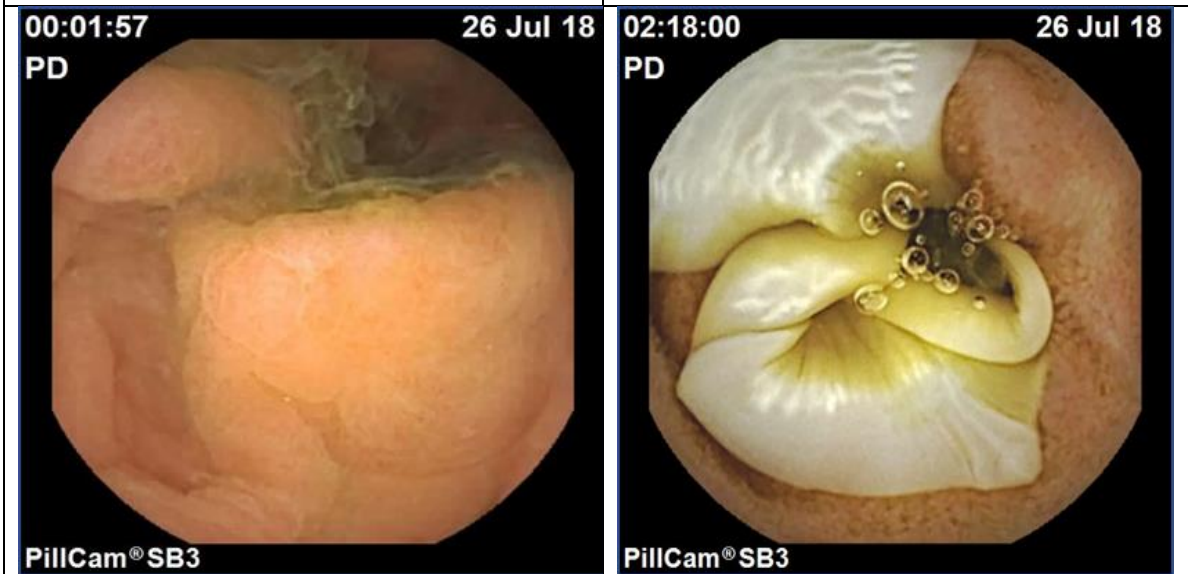
*Fig. 7 Abnormal GI tract frames compared to a normal frame*

GI tract disorders and diseases include unexpected bleeding in the digestive tract, inflammatory bowel diseases such as Crohn's disease, Vascular Ectasia, Lymphangiectasia, Lymphoid Hyperplasia, Tumour, Polyp, Ulcer, presence of Tapeworm, Hookworm, Lesion and many more. In this particular t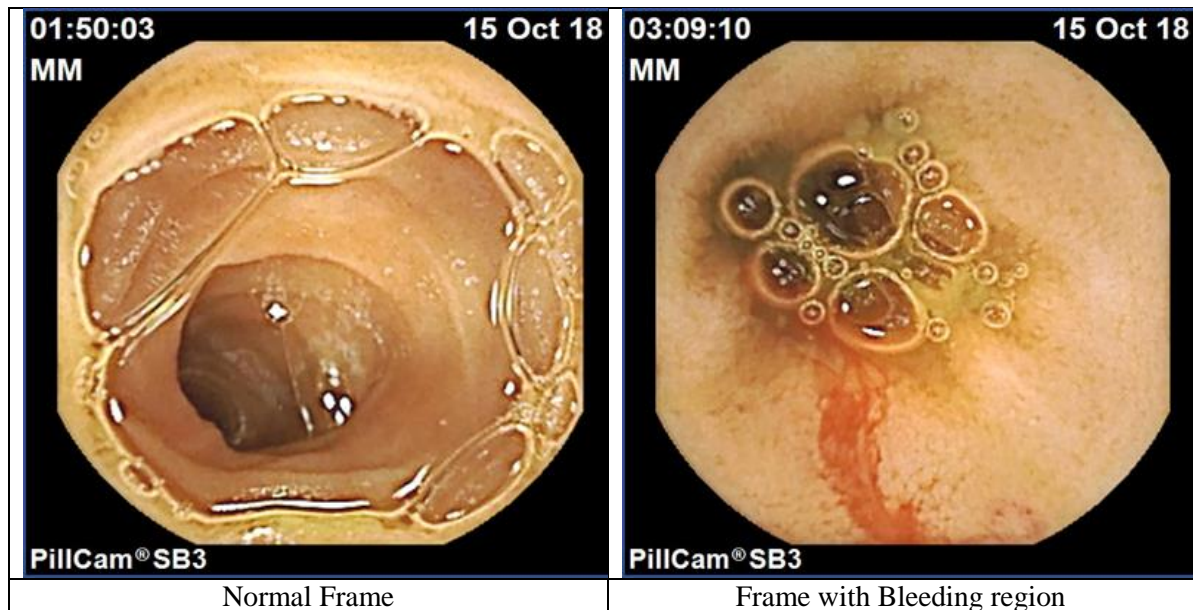hesis, four-five types of diseases, i.e. unexpected Bleeding regions, Small Ulcers and Crohn's diseases, Tapeworm, and Vascular Ectasia is been explored. So, I am going to discuss these four-five types of disorder and diseases only, which will help the reader to get a basic idea of what they are dealing with.

Unexpected Bleeding Regions [4] - Bleeding is considered as the indicator of most anomalies like ulcers, vascular regions, inflammatory regions, and tumors. Bleeding in the GI is a symptom of a problem rather than a disease itself. The cause of the bleeding may not be serious, but it's important to find the source of this symptom. Bleeding can come from one or more of GI tract areas, for example, a small area such as an ulcer on the lining of the stomach or from a broader problem such as inflammation of the colon. As shown in Fig. 8 shows a few bleeding regions frames in the GI tract.
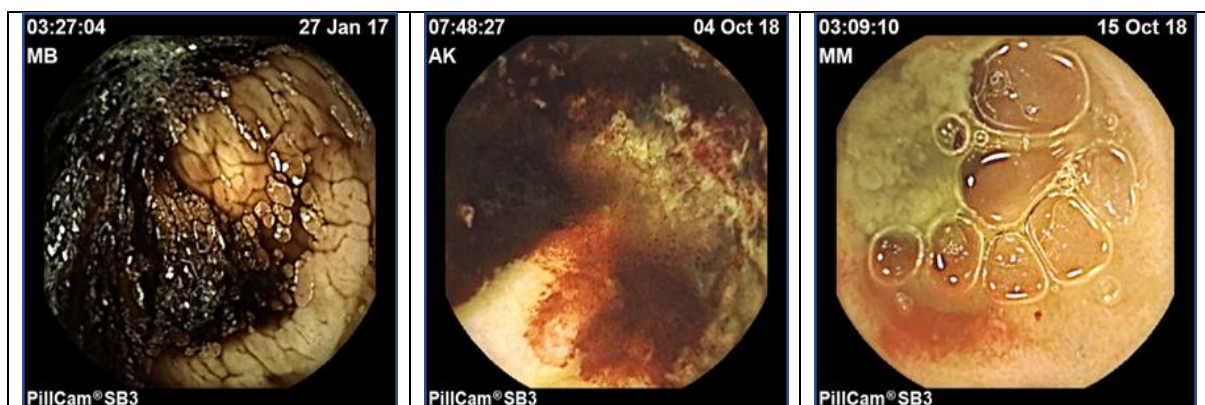


*Fig. 8. Frames with bleeding regions*

Small Ulcers and Crohn's Diseases [5]- Crohn's disease is an inflammatory bowel disease. It causes inflammation in the digestive tract, which can lead to abdominal pain, severe diarrhea, fatigue, weight loss, and malnutrition. Inflammation caused by Crohn's disease affects different areas of the digestive

tract in different people. The inflammation caused by Crohn's disease can spread deep into the layers of affected bowel tissues. Crohn's disease is both painful and debilitating and sometimes may lead to life-threatening complications. The most common areas get affected by Crohn's disease are the last part of the small intestine and the colon. An ulcer is sore or breaks in the lining of any part within the digestive tract that contains concentrated gastric juice. Chronic inflammation lead to open sores i.e. ulcers anywhere in your digestive tract, including your mouth, anus, and in the genital area. Fig. 9 shows a few small ulcers and Crohn's affected regions frames in the GI tract.



*Fig. 9.Frames with Small Ulcers and Crohn's affected regions*

Tapeworm [6] - Tapeworms are flat, segmented worms that live in the intestines. A person gets infected with these parasites on drinking contaminated water and food. Eating undercooked meat from infected animals is the main cause of tapeworm infection in people. Although tapeworms in humans usually cause few symptoms and easily get treated, they can sometimes cause serious, life-threatening problems. Tapeworm causes symptoms such as weakness, diarrhea, abdominal pain, fatigue, nausea. Fig. 10 shows a few tapeworms affected regions frames in the GI tract.



*Fig. 10. Frames with tapeworms affected regions frames*

Vascular Ectasia - Vascular ectasias in GI tract is relatively rare cause of recurrent and sometimes severe blood loss which include arterio-venous malformations as angiodysplasia and Dieulafoy's lesion, venous ectasias (multiple phlebectasias and haemorroids), teleangiectasias which can be associated with hereditary hemorrhagic teleangiectasia (HHT), Turner's syndrome and systemic sclerosis, haemangioma's, angiosarcoma's and disorders of connective tissue affecting blood vessels as pseudoxanthoma elasticum and Ehlers-Danlos's disease. As a group, they are relatively rare lesions that may be a major source of upper and lower gastrointestinal bleeding. Clinical presentation is variable, ranging from acute or recurrent bleeding that may be life-threatening to asymptomatic cases over iron deficiency anaemia. Fig. 11 shows few vascular ectasias affected regions frames in the GI tract.

*Fig. 11. Frames with vascular ectasias affected regions*

It's clearly visible from the above images, that every disorder and disease have different properties, different structure, and different origin. As it was mentioned earlier in section 1.1 that the capsule endoscopy provides around 80,000 to 1,20,000 images per patients, examining each image thoroughly to find anomalies is a hectic job for any physician. This being a challenge taken by many researchers and scientists, they have worked and developed image processing techniques to deal with this problem. Till date and going on there are numerous algorithms, techniques on image processing are contributed by different scientists and researchers.

## 1.3 Digital Image Processing

Image processing is a method of performing some operations on an image, in order to get an enhanced image and to extract some useful information from it [7]. It is a type of signal processing in which input image is converted into digital form and output may be image or characteristics/features associated with that image. It is among rapidly growing technologies today and it forms core research area within engineering and computer science disciplines too. Image processing basically includes three following steps:

- Importing the images via image acquisition tools
- Manipulating and analyzing the images
- Output in which results can be altered image or report that is based on image analysis.

The detail steps and phases of digital image processing are mentioned in the flowchart in Fig. 12
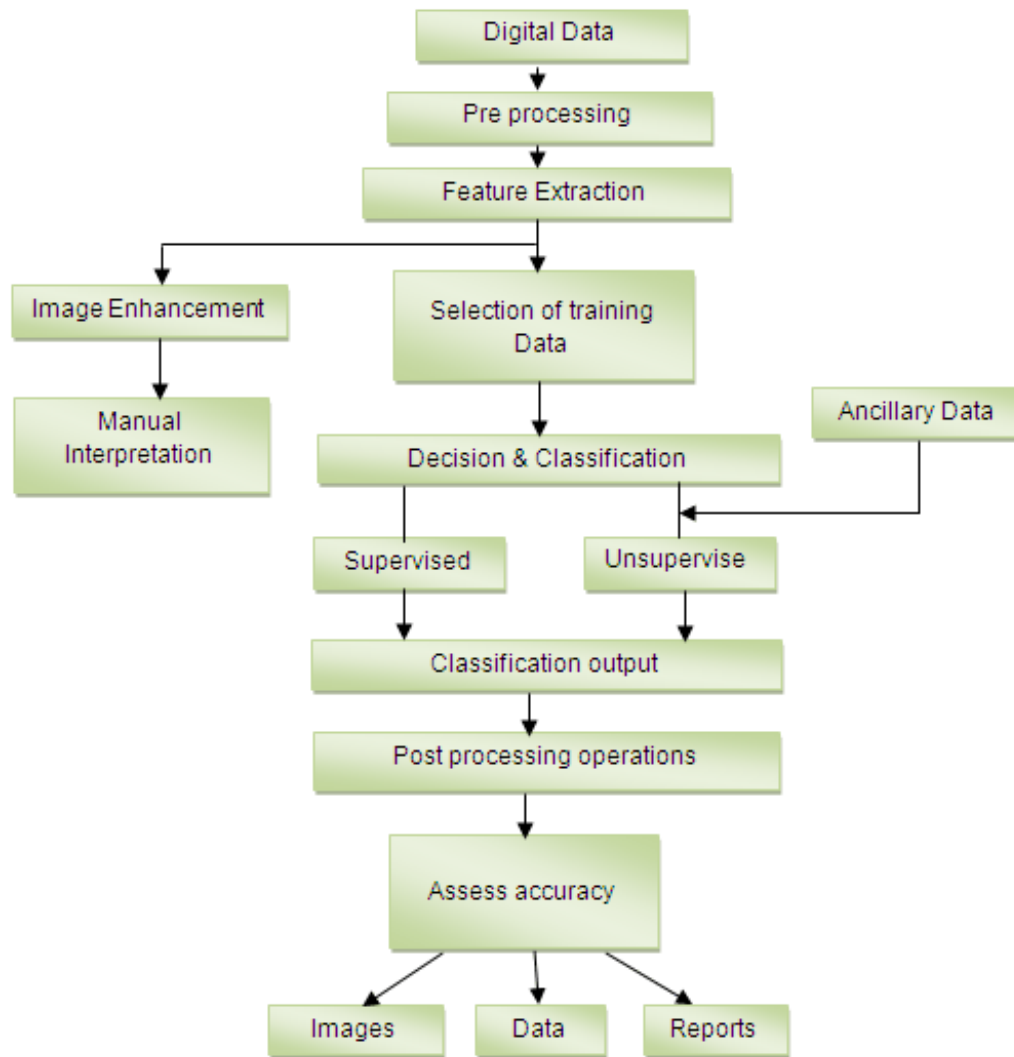
*Fig. 12. The workflow of a digital image processing*

Image processing in the field of medical application has gained a lot of popularity due to its rapid progress in computerized medical image reconstruction and the associated developments in analysis methods and computer-aided diagnosis has boosted medical imaging into one of the most important sub-fields in scientific imaging. Ultrasound, MRI, CT-Scan, PET Scan, Endoscopy are few medical techniques used by the physicians for visualization of the internal structure of the human body without any surgery.

Biomedical imaging [8] concentrates on the capture of images from both diagnostic and therapeutic view respectively. Snapshots in vivo physiology and physiological processes are garnered through advanced sensors and computer technology. Biomedical imaging technologies utilize either x-rays (CT scans), magnetism (MRI), sound (ultrasound), radioactive pharmaceuticals (nuclear medicine: SPECT, PET) or light (endoscopy, OCT) to assess the current condition of an organ or tissue and can monitor a patient over time for diagnostic and treatment evaluation.

### 1.4 Wireless Capsule Endoscopy Image Processing

A computer-aided diagnosis system for capsule endoscopy can be clinically very useful, which can differentiate abnormal tissue from health structure and provide correlation information among the images [9]. As mentioned in section 1.3 it is a common practice that each image is pre-processed to enhance the accuracy of feature extraction, followed by feature refinement. Next, the output of the

feature refinement is in a concise form of the image abstraction for the final classification, and the classifiers may be artificial intelligent based or rule-based. In general, artificial-intelligent-based classifiers need to be trained before use, whereas rule-based classifiers utilize "If-then" rules for classification which requires no training. The general process of a typical WCE image vision analysis system is illustrated as a flowchart in Fig. 13
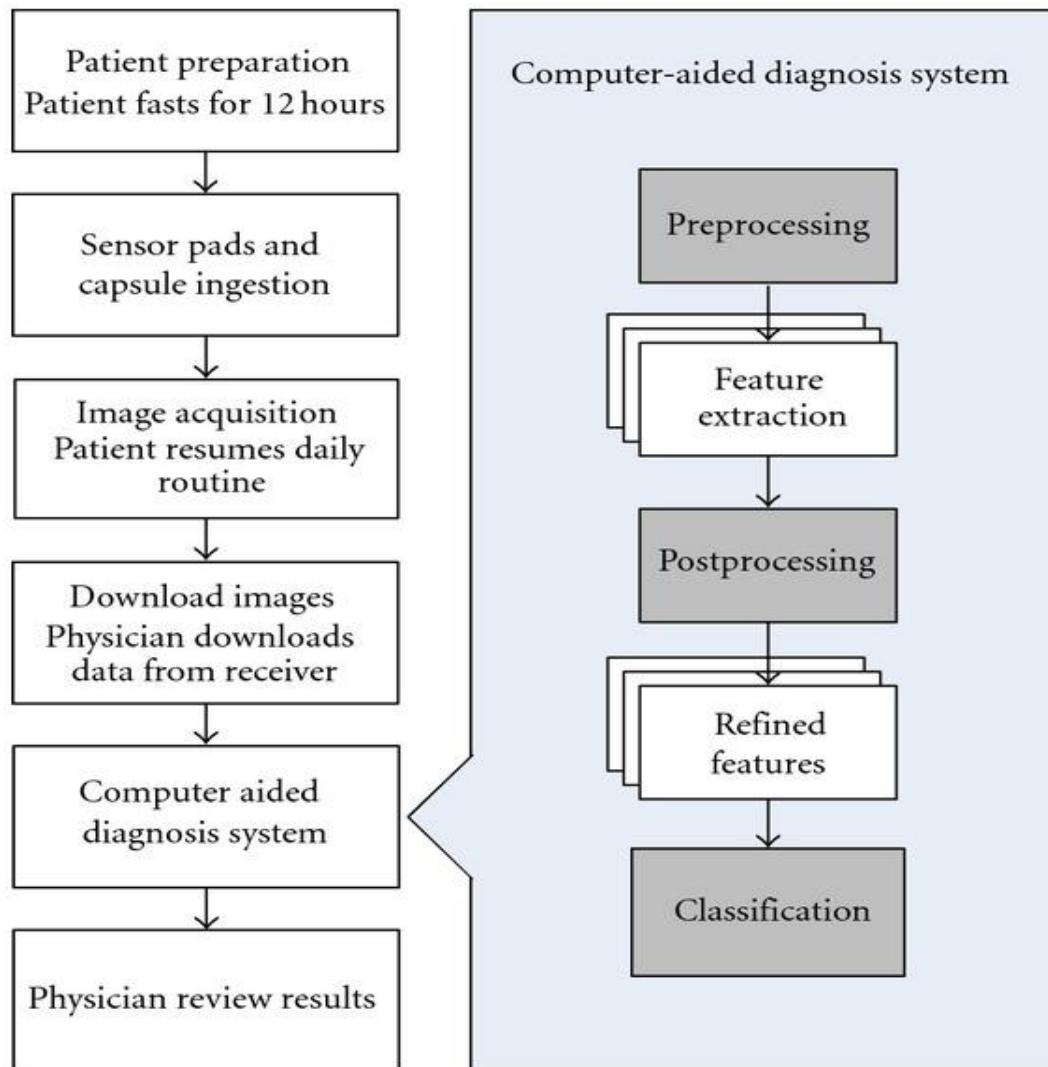


*Fig. 13. The workflow of a typical WCE image vision analysis system*

A color image of wireless capsule endoscopy is a snapshot of the digestive tract at a given time. However, in a computer-aided diagnosis system, the image content semantically needs to be translated in numerical ways for interpretation. There are several ways to represent the numerical form of an image which is termed as image abstraction. Among WCE applications, there are three popular features for image abstraction: (1) color, (2) texture, and (3) shape features. Colour images produced by capsule endoscopy contain useful color information and hence it can be used as an effective clue to suggest the topographic location of the current image. In WCE applications, a unique texture pattern called "villi" is used to distinguish the small intestine from other organs. In addition, abnormality in capsule endoscopy video is discriminated by comparing the texture patterns between normal and abnormal mucosa regions, making texture pattern a popular feature for image abstraction. Shape feature is another common abstraction approach used for machine vision applications. Object shapes

can provide strong clues to object identity, and humans recognize objects solely on their shapes. Fig. 14 shows typical images taken from each organ.
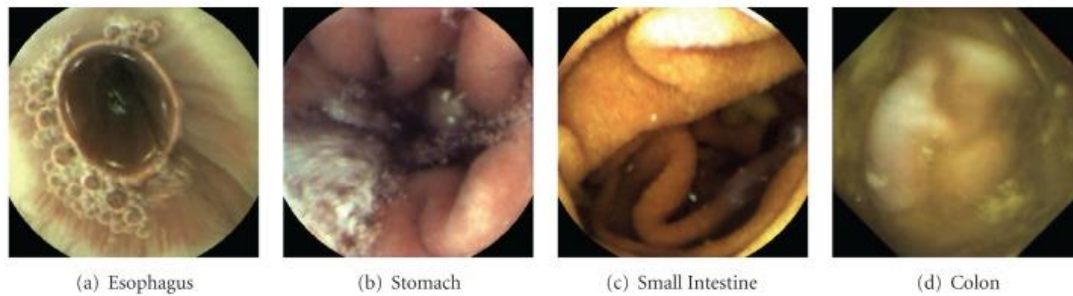


(a) Esophagus      (b) Stomach      (c) Small Intestine      (d) Colon

*Fig. 14*

## 1.5 Aim of this Thesis

In this thesis, the aim is to build an method which segments out the anomalous regions and classifies anomalies i.e. the Region of Interest (ROI) and non-anomalies i.e. Non-Region of Interest (Non-ROI) regions from wireless capsule endoscopy videos. For segmentation, a two-step novel adaptive statistical operation CNN synergism-based segmentation is used, which detects and segment anomalies in WCE images. In this process, HSV color space is chosen to carry out the entire segmentation procedure due to its uncanny resemblance with the human vision system. By the means of statistical operation, the luminance of the input image is reduced and only the ROI get enhanced which helps in segmentation. But this step also segments Non-Region of Interest (Non-ROI) regions occasionally. In order to prevent this type of erroneous segmentation, further classification of the above-segmented images is carried out using AlexNet, GoogLeNet, InceptionV3 and SVM trained with Bag-of-Features into two classes namely ROI and Non-ROI to obtain the accurate ROI images. The steps are illustrated by the flowchart in Fig. 15 below.
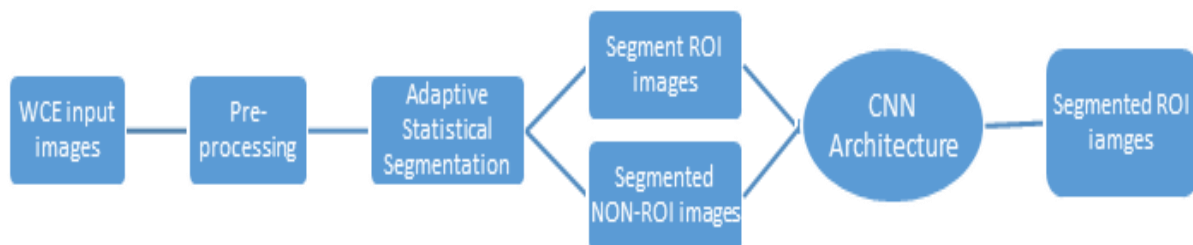


*Fig. 15 Flowchart of the proposed algorithm*

# 2. LITERATURE SURVEY

## 2. Literature Survey

In India, doctors prefer WCE over traditional endoscopy diagnosis (TED), as TED cannot reach deep into the internal abdominal organs. Moreover, TED is invasive and completely dependent on the insertion of long flexible pipe through the mouth or the anus part of the human body which causes severe pain, whereas WCE retrieves the images from the GI tract without any notable discomfort or pain. Now as discussed in section 1.2 about the types of anomalies in GI tract and there different properties with respect to shape, size, texture, it is a challenging task for every researcher in the field of biomedical image processing to successfully segment or classify ROI and Non-ROI regions in an fast automated way, computer-aided diagnosis. Image segmentation is an important yet very fascinating task in medical image analysis. Proper results can be produced with fully automatic segmentation methods; however, accuracy can vary statistically and significantly in accuracy for an entirely autonomous process. Even the basic segmentation in the field of medical imaging vary between physicians. Henceforth, interactive segmentation techniques have increased in popularity and different techniques have been proposed by researchers to segment ROI regions from WCE images to make the diagnosis process easier for the physicians.

In paper [10] the authors analyzed bleeding patterns from five WCE videos using multiple features. But this method performs various missed detection of small bleeding spots, small blood flow patterns. An automatic method for segmenting the reddish lesion is proposed in [11] which is based on the central moments of the probability distribution of intensities at arbitrary locations within an image across a continuously varying range of the scale. Polyp detection through WCE suffers from various types of challenges due to its unique protruding shape. In [12] watershed segmentation is used in combination with the novel Gabor texture features based initial marker selection method, proposed by the authors and K-means clustering method to outline the polyp regions. A block-based segmentation algorithm is used in [13] for bleeding detection using average saturation of the HSI model and skewness and kurtosis of uniform local binary patterns histogram as features. To locate the presence of peptic ulcers, image segmentation with thresholding, watershed segmentation, and morphological operators are used by the authors in [14]. This algorithm is limited to a few images since the threshold values differ for every image. Fiori et. al proposed a colon segmentation technique enhancing polyps based on texture and geometrical features[15]. A new approach that automatically and rapidly singles out the blood regions from a WCE frame was presented by [16]. It was the first study to report the application of a deep learning strategy for the segmentation of blood regions in WCE images. The key contribution of this work was to adapt the fully convolutional networks (FCNs) for bleeding region segmentation, which performed a high-level detection for both active and inactive bleeding regions. MAP (Maximum a Posteriori) approach with Markov Random Fields is proposed by P. Vieira et al. [17] to segment angiodysplasias using different color spaces. But segmentation in this proposed algorithm works better when the lesions presented higher probabilities of bleeding. Another approach was proposed by the same author in [18] for segmenting small bowel tumor tissues using the MAP algorithm where the image is segmented into two regions and taking the most intense as the ROI but due to some camera viewing point sometimes segmentation failure occurs, so to deal with this situation a double segmentation procedure, into two and three classes is proposed. The right model was chosen by using the Maximum Likelihood (ML) estimation. In the cases where a three Gaussian model is more likely than a two Gaussian model the most intense class is discarded and the ROI was assumed to be the second most intense region.

Graph Cut techniques and Grab Cut techniques are very well-known techniques in the field of segmenting ROI. A combinatorial optimization technique, was used by Boykov and Jolly [19], [20] and [21] for the task of segmenting the images. The image was treated as a graph and each pixel of the image is considered as a graph node. The globally optimal pixel labeling with respect to the

mentioned cost function for the object of interest and backgrounds are calculated in an efficient manner by max- flow/min-cut algorithms. This technique was applied in a proficient manner to N-dimensional images [22]. If the intended object and background seed pixels are given, the rest of the pixels are automatically labeled. Rother et al. [23] broaden the graph-cut technique by the introduction of repeated segmentation scheme that uses graph-cut for intermediate steps. In this process, the user draws bounding boxes (rectangle in shape) around the region of interest, which gives the first approximation of the final object/ background labeling. Then, each looping step assembles color statistics according to the current segmentation, re-weights the image graph and then the graph-cut technique was applied to calculate the newly refined segmentation. After the iterations come to a halt, the results of the segmentation were fine-tuned by mentioning additional seeds, like the original graph-cut. Lumen-oriented image segmentation algorithm has been developed and proposed in [24] where individual regions corresponding to specific homogenous features are segmented. In [25] an interactive grow cut segmentation for capsule endoscopy image is discussed, which uses an iterative process to segment the image as gray matter and white matter with less human interventions. Grow cut based segmentation proves to be a better result than graph cut and grab cut.

The SuperPixel segmentation method is a new emerging segmentation method being used in various image processing techniques especially bio-medical image processing. The concept of superpixel is to group connected pixels with similar colors or gray levels. Superpixel segmentation divides an image into hundreds of non-overlapping superpixels. A superpixel segmentation, in particular, using SLIC (simple linear iterative clustering) implementation is proposed by boschetto et al. [26] for automatic segmentation of villi. Automatic villi detection lead to automatic quantitative analysis for staging and grading diseases such as celiac disease and irritable bowel disease. Different SLIC superpixel segmentation approach is proposed in [27] for the detection of polyps. In this process, different SLIC superpixel numbers are examined to find the highest sensitivity for detection of polyps this segmentation process is a promising one and an improved version of their previous studies. The authors in [28] used color based and contour-based segmentation methods on bleeding, lesion and reflux detections where the bleeding cases achieve the highest detection rate in comparison to lesion and reflux. This limits the possibility of detecting any other abnormalities. In fact, most of the algorithms used in literature are limited to one or two specific diseases whereas the proposed algorithm can segment four to five types of anomalies.

The two-step segmentation method proposed in this thesis is simple yet very efficient. This method will eases the job of the expert's physicians as it automatically segments the ROI (anomalies) regions from WCE video frames. It produces mostly accurate results and the reliability of the method is validated from the expert's physicians from Medica Superspeciality Hospital, Kokata, India.

# 3. Pre-Processing And Colour Features

# 3. Pre-Processing and Colour Features

## 3.1 Pre-Processing

Image pre-processing is the operations applied on input images at the lowest level of abstraction whose aim is an improvement of the image data that suppresses undesired distortions or enhances some image features important for further processing [29]. It does not increase image information content but its methods use the considerable redundancy in images. The neighboring pixels corresponding to one object in real images have similar brightness value. If a distorted pixel is picked out from the image, it can be restored as an average value of neighboring pixels. Image pre-processing methods are classified into four categories, according to the size of the pixel neighborhood that is used for the calculation of new pixel brightness. The following categories are:

- Pixel brightness transformation
- Geometric transformation
- Local pre-processing
- Image restoration

Which can be further elaborated into image cropping and filtering; intensity adjustment and histogram equalization; brightness thresholding; clearing areas of a binary image; detecting edges.

Image cropping focuses on the relevant part of an input image. Image filtering uses a small neighborhood of a particular pixel in the input image to produce a new intensity value in the output image. Smoothing methods help to reduce noise or other small fluctuations in the image. The grayscale transformation aims to modify brightness without regard to the position in the image. Intensity adjustment and histogram equalization are used to enhance the contrast in the image. Brightness thresholding deals with grayscale transformation whose result is a binary image.

Previous studies have found that a large number of false alarms were generated while using the existing algorithms for WCE segmentation due to the presence of peripheral black regions on the input image. Generally, in most of the WCE frames, the center circular or octagonal shaped portion of the WCE image frame contains the vital and actual information. The surrounding black regions pixels with labels contain no useful information regarding the nature of the WCE image frame. Therefore, incorporation of these black regions in anomalies segmentation process would surely degrade the performance of the proposed method. To deal with this situation a polygon-shaped mask to remove the surrounding labeled peripheral regions as shown in Fig. 16.
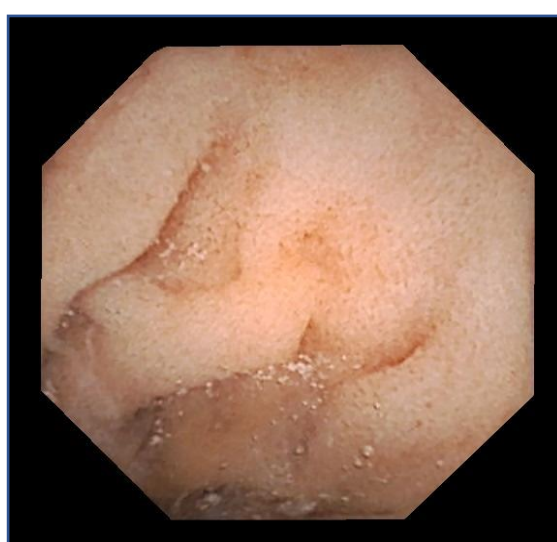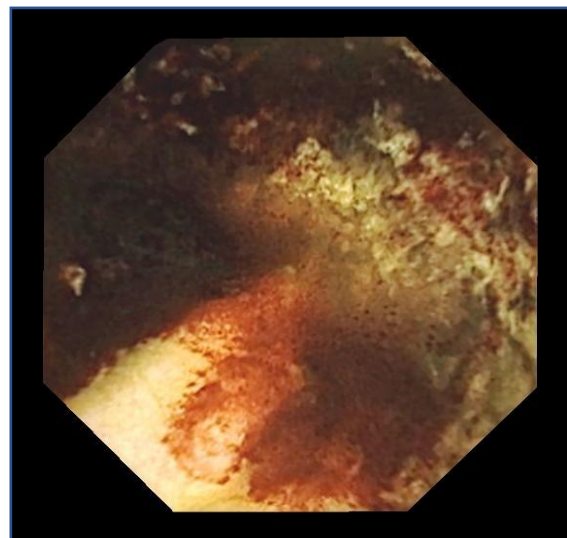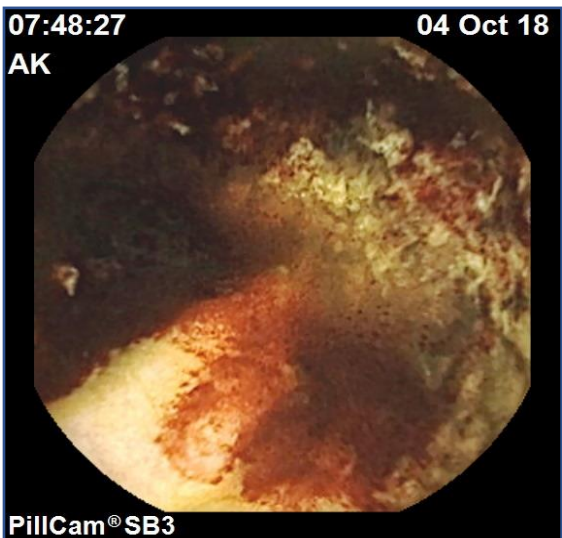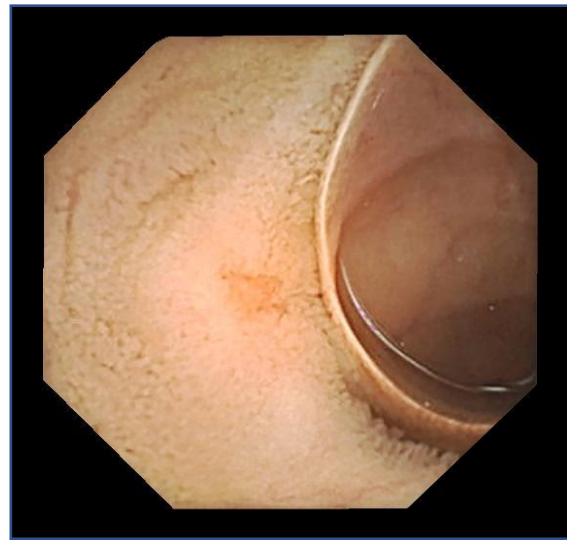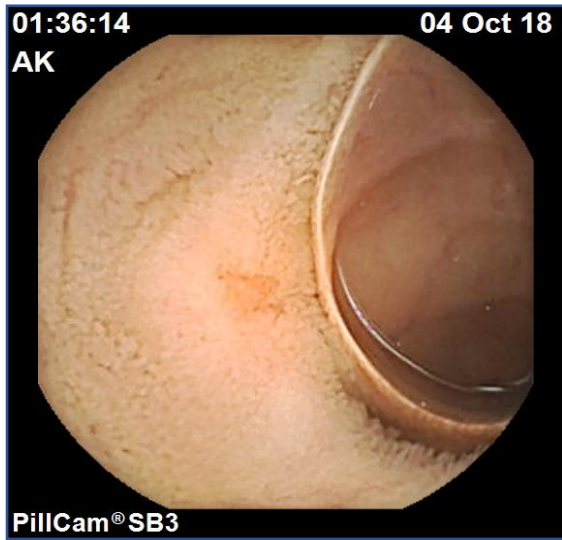
*Fig. 16. Polygon shape masked preprocessed images.*

**3.2 RGB color space**

Range of colors are created by the primary colors of pigment and these colors then define a specific color space. Color space is also known as the color model (or color system), which states an abstract mathematical model that simply describes the range of colors as tuples of numbers. In other words, color space is an elaboration of the coordinate system and sub-space. Each color in the system is represented by a single dot. A color space is a useful method for users to understand the color capabilities for a particular digital device or file. It represents that, what a camera can see, a monitor can display or a printer can print, and etc. There are a variety of color spaces, namely RGB, CMY, HSV, HIS. I will talk about RGB color space in this section.

RGB (R=Red, G=Green, B=Blue) is a kind of color space which uses red, green and blue to define a color model in detail [30]. An RGB color space is simply interpreted as "all possible colors" which can be made from three colors for red, green and blue as shown in Fig. 17. With such conception, each pixel of an image is assigned a range of 0 to 255 intensity values of RGB components and using only these three colors, there can be 16,777,216 colors on the screen by different mixing ratios.
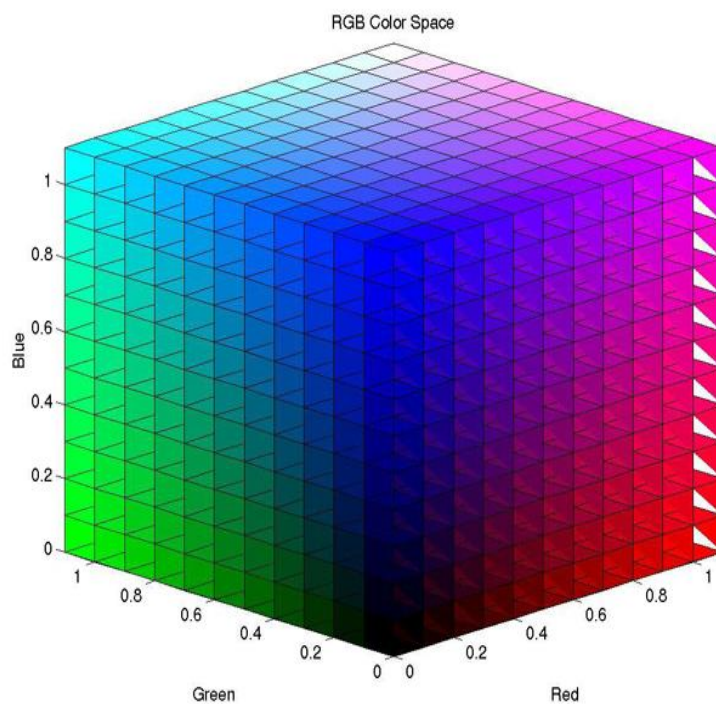


*Fig. 17. RGB Color Space*

RGB model states, that each color image is actually formed by three different images. Red image, Green image, and Blue image. A normal grayscale image is defined by only one matrix, but a color image is actually composed of three different matrices.

One color image matrix = red matrix + green matrix + blue matrix

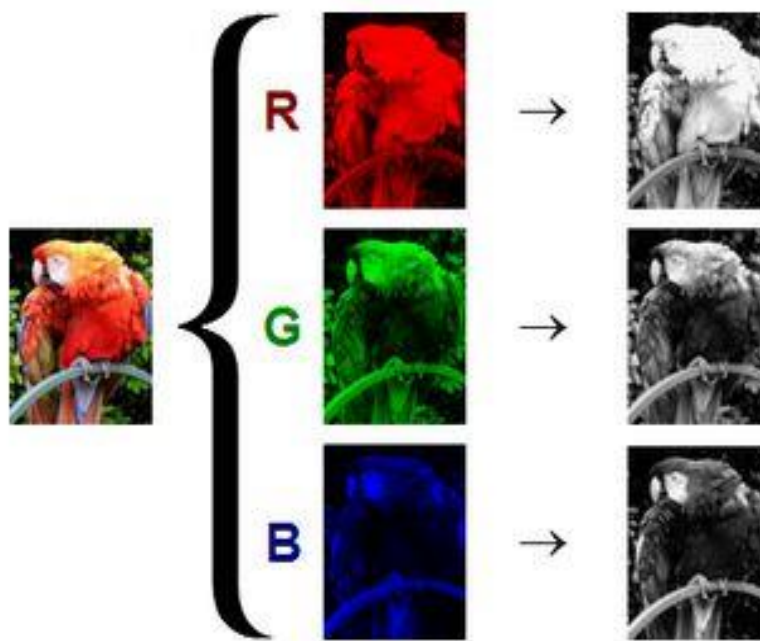This can be illustrated by Fig. 18 given next page

*Fig. 18 RGB channel representation*

RGB color space contains both color and intensity information but the main disadvantage of RGB color space-based representation of images is that a change in the intensity leads to change in all the three components [31]. In order to overcome this problem and due to the resemblance with the human vision system, the use of HSV color space for anomalies detection is preferred. Hue Saturation Value (HSV) can separate image luminance from color information and this makes it easier on working with the luminance of the image/frame.

### 3.3 HSV color space

HSV (hue, saturation, value) is an alternative representation of the RGB color model, designed by computer graphics researchers in the 1970s to more closely align with the way human vision perceives color-making attributes [32]. The HSV color space consists of three components: hue, saturation, and value. 'Value' can be sometimes substituted with 'brightness' and then it is known as HSB. HSV is a cylindrical geometries as shown in Fig. 19, with hue as their angular dimension, starting at red primary at 0°, passing through the green primary at 120° and the blue primary at 240°, and then wrapping back to red at 360°. In every geometry, the central vertical axis comprises the neutral, achromatic, or gray colors, ranging from black at lightness 0 or value 0, the bottom, to white at lightness 1 or value 1, the top.
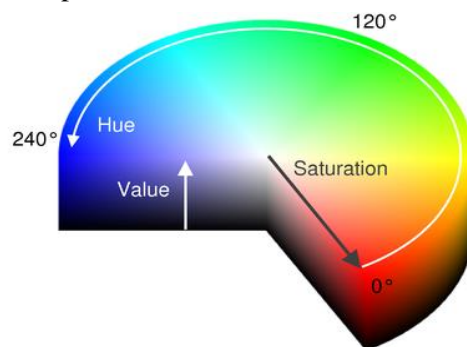


*Fig. 19 Cylindrical geometries*

Hue is the color portion of the color model, which is expressed as a number from 0 to 360 degrees:

| Colour | Angle |
|---|---|
| Red | 0-60 |
| Yellow | 60-120 |
| Green | 120-180 |
| Cyan | 180-240 |
| Blue | 240-300 |
| Magenta | 300-360 |

Saturation is indicated as the range of grey in the color space, it ranges from 0 to 100%. The value is calculated from 0 to 1 when the value is '0,' the color is grey and when the value is '1,' the color is a primary color. The faded color is due to the lower saturation level, which indicates the color contains more grey.

Value is considered as the brightness of the color and varies with color saturation, it ranges from 0 to 100%. When the value is '0' the color space will become totally black. With the increase in the value, the color space brightness up and gradually shows various colors.

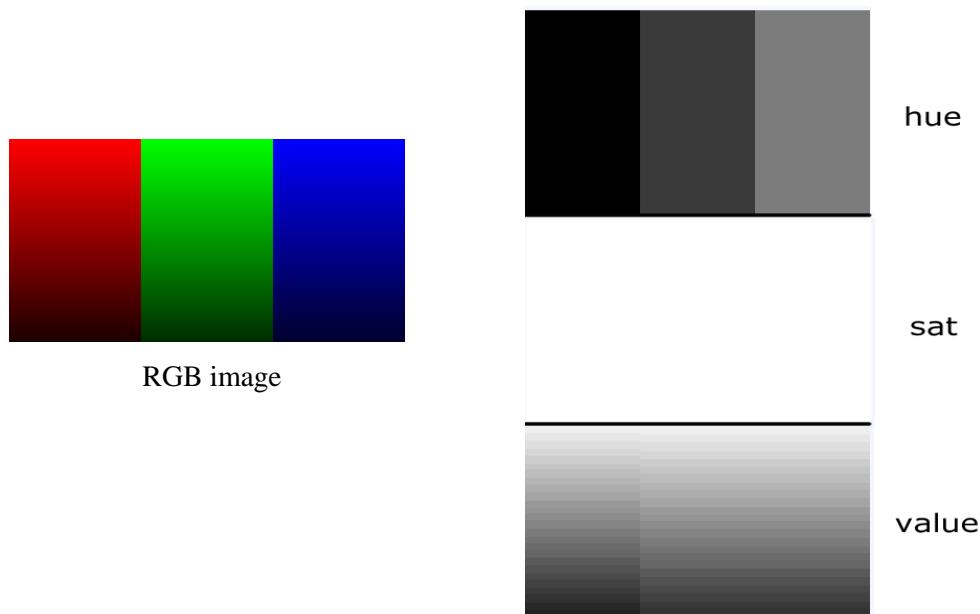Fig. 20 below might help to clarify what each channel represents.



RGB image

*Fig 20 RGB to HSV representation*

The HSV color space is widely used for generating high-quality computer graphics. In simple terms, it can be used to select various different colors needed for a particular picture. An HSV color wheel helps to select the desired color, the user can select a particular color needed for the picture from the color wheel. It represents the color according to human perception.

Sometimes the HSV model is illustrated as a cylindrical or conical object Fig. 21. When it is represented as a conical object, the hue is indicated by the circular part of the cone. The cone usually represents the three-dimensional form. The saturation is calculated using the radius of the cone and the value is calculated from the height of the cone. A hexagonal cone is also used to represent the HSV model. The conical model is used to represent the HSV color space in a single object. Due to the two-dimensional nature of computer interfaces, the conical models of HSV are best suited for selecting colors in computer graphics.
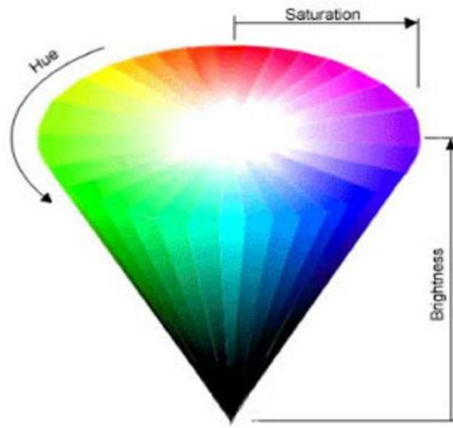
*Fig. 21 Conical geometries*

The HSV color space is also represented as a wheel which helps the user to pick the desired color Fig. 22. Hue is represented by the circle in the wheel, and a separate triangle is used to represent saturation and value. The horizontal axis of the triangle represents value and the vertical axis indicates saturation. When a particular color is needed for the picture, first the color from hue (the circular region), and then from the vertical angle of the triangle the desired saturation is selected. For brightness, the desired value from the horizontal angle of the triangle is selected.
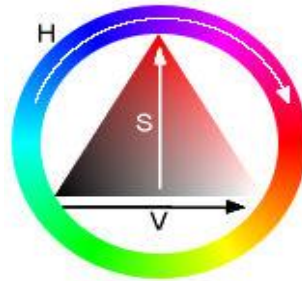


*Fig. 22. HSV color space as a wheel*

**3.4 Colour Conversion**

Color vision can be processed using RGB color space or HSV color space. RGB color space describes colors in terms of the amount of red, green, and blue present. HSV color space describes colors in terms of Hue, Saturation, and Value. In situations where color description plays an integral role, the HSV color model is often preferred over the RGB model. The HSV model describes colors similarly to how the human eye tends to perceive color. RGB defines color in terms of a combination of primary colors, whereas, HSV describes color using more familiar comparisons such as color, vibrancy, and brightness.

In this thesis the approach is to enhance the region of interest (ROI) for which V and S channel are the most suitable one since, V channel deals with the chromatic notion of intensity i.e. the lower the value of colors, more similar to black and higher the value of colors, more similar to color itself. And S channel deals with the amount of white light mixed with hue. To convert an image from RGB plane to HSV plane following mathematical calculation takes place.

The input RGB image is initially normalized to limit the range of input intensities within (0 1) [33], [34]. Then the normalized image is converted from RGB color space to HSV color space using the following equations:

$$R' = \frac{R}{255}; G' = \frac{G}{255}; B' = \frac{B}{255};$$

$$C_{max} = max\ (R', G', B') \tag{1}$$

$$C_{min} = min\ (R', G', B') \tag{2}$$

$$\Delta = C_{max} - C_{min} \tag{3}$$

To get the hue (H) value, the largest of the R, G, B, $C_{max}$, equation (1) values is considered. The smallest two are subtracted off, and divided by the difference between the largest and the smallest. To normalize the hue either 0, 2, or 4 is added, hence the resulting H is any real number. However, any arbitrary number below 0, and above 6 is considered redundant, and derive a value H mod 6, or if H is negative, then (H mod 6) + 6, equation (4).

$$\text{Hue Calculation: } \begin{cases} 0° & \Delta = 0 \\ 60° \times \left(\frac{G'-B'}{\Delta} mod\ 6\right) & ,C_{max} = R' \\ 60° \times \left(\frac{B'-R'}{\Delta} + 2\right) & ,C_{max} = G' \\ 60° \times \left(\frac{R'-G'}{\Delta} + 4\right) & ,C_{max} = B' \end{cases} \tag{4}$$

The saturation, S, is the difference between the largest and smallest color channel values, divided by the brightness, V. If V is 0, then the resulting saturation is 0, equation (5).

$$\text{Saturation Calculation: } \begin{cases} 0 & ,C_{max} = 0 \\ \frac{\Delta}{Cmax} & ,C_{max} \neq 0 \end{cases} \tag{5}$$
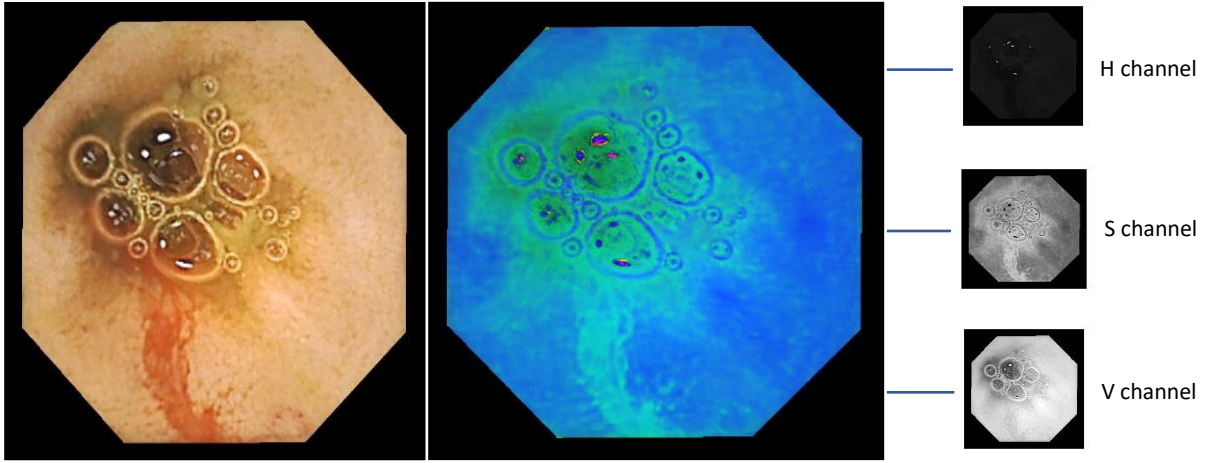
The brightness, V, is based on the brightest color channel, $C_{max}$, equation (6)

Value Calculation: $V = C_{max}$ (6)



(a)     RGB color space                    HSV color space

H channel

S channel

V channel
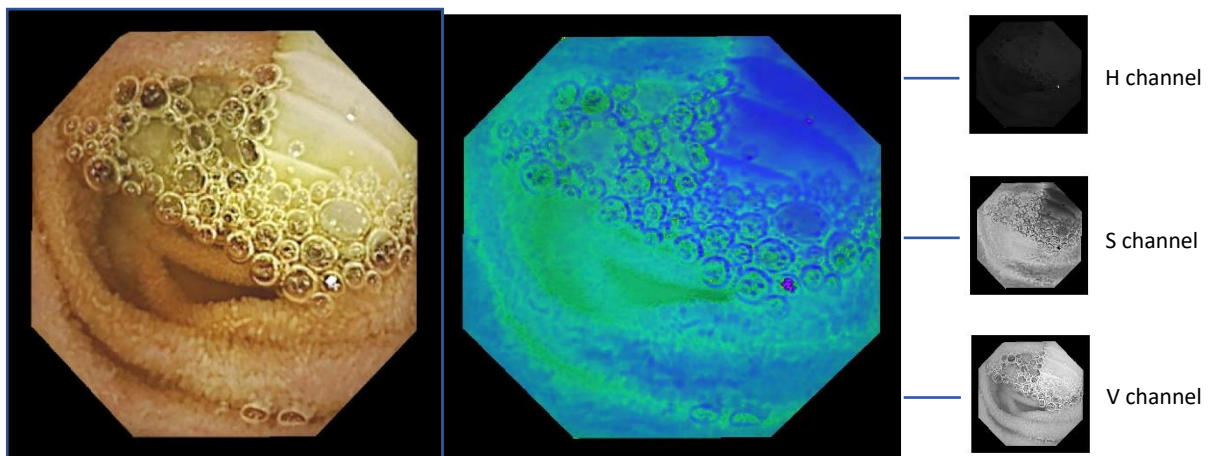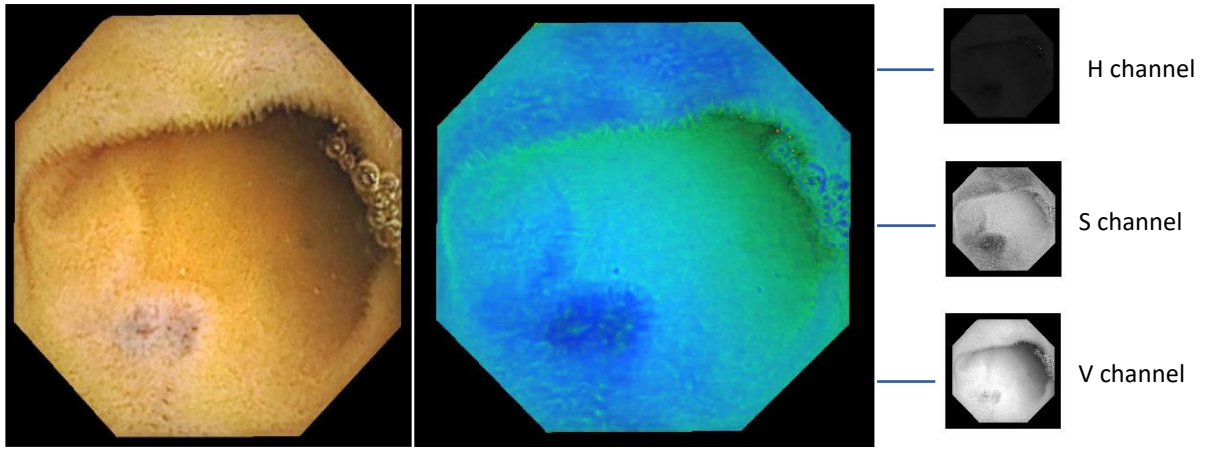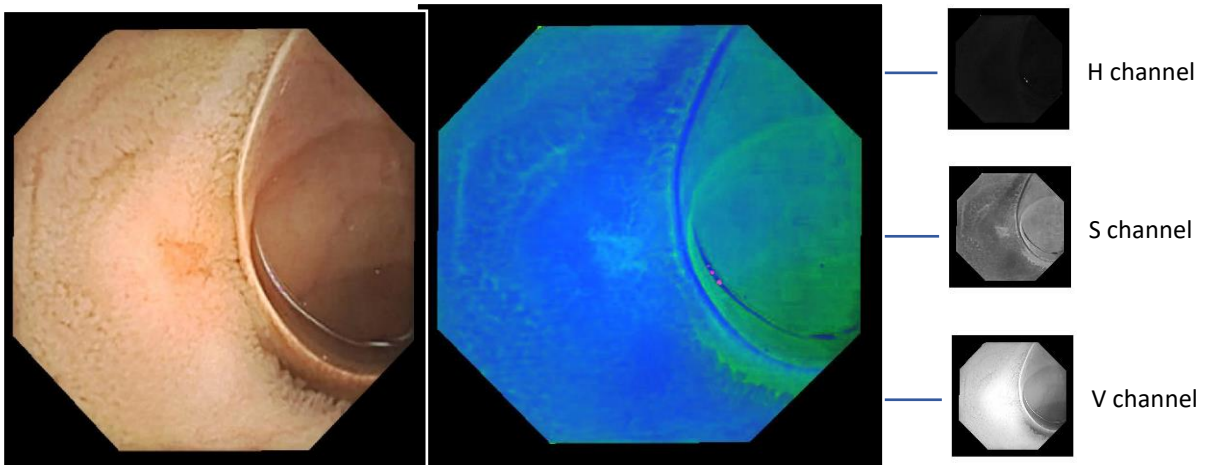
(b)    RGB color space                    HSV color space

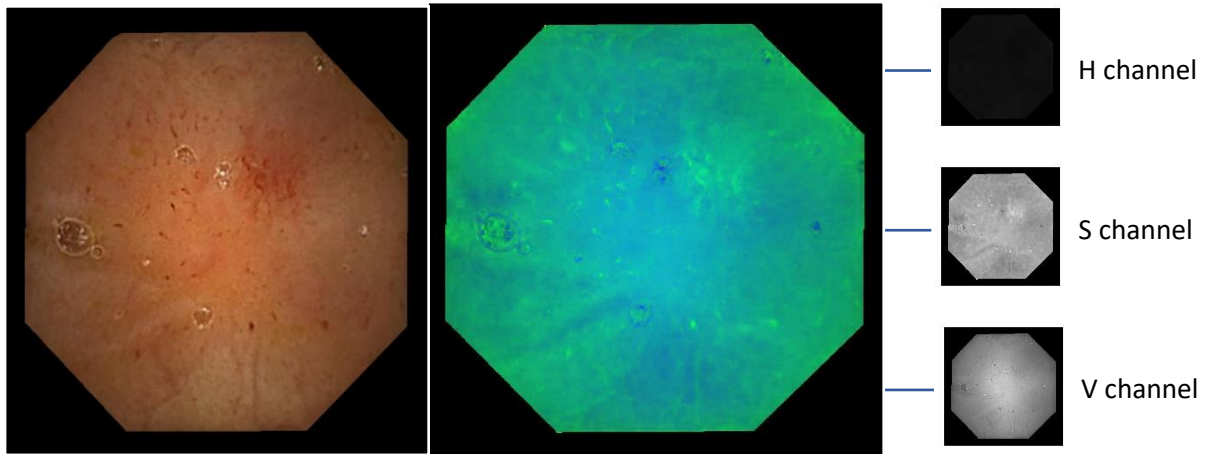

(c)    RGB color space                    HSV color space



(d)    RGB color space                    HSV color space

21

(e)  RGB color space       HSV color space

*Fig. 23 shows the pictorial representation of input RGB image and its converted HSV color space image with individual H, S and V channel. All the input RGB images belong to different ROI regions extracted from self-acquired WCE videos from Medica Superspeciality Hospital, Kolkata, India. (a) Tapeworm, (b) Bleeding regions, (c) Vascular Ectasia, (d) Crohn's disease, (e) Small erosion and Ulcer.*

# 4. ADAPTIVE STATISTICAL SEGMENTATION METHOD

# 4. Adaptive Statistical Segmentation Method

The division of an image into meaningful structures, image segmentation, is often an essential step in image analysis, object representation, visualization, and many other image processing tasks [35]. In this section, I focussed on how to analyze and represent an object i.e. anomalies. A great sort of segmentation strategies has been projected within the past decades, and some categorization is important to present the strategies properly here. A disjunct categorization doesn't appear to be potential although, as a result of even two terribly totally different segmentation approaches could share properties that defy singular categorization1. The categorization bestowed during this section is thus rather a categorization regarding the emphasis of an approach than a strict division.

Few segmentations techniques can be categorized as follows:

- Threshold-based segmentation: Histogram thresholding and slicing techniques can use to segment the image. They'll be applied on to a picture, however, can even be combined with pre- and post-processing techniques.
- Edge-based segmentation: With this system, detected edges in a picture are assumed to represent object boundaries, and want to determine these objects.
- Region-based segmentation: Where an edge-based technique will attempt to find the object boundaries and locating the object itself by filling them in, a region-based technique takes the opposite approach, by (e.g.) starting in the middle of an object and then "growing" outward until it meets the object boundaries.
- Clustering techniques: Clustering techniques are sometimes used as a synonym for segmentation techniques, it can be used to denote techniques that are primarily used in exploratory data analysis of high-dimensional measurement patterns. In this context, clustering methods aim to group together patterns that are similar in some respect. This technique is very similar to what I generally do while segmenting an image, and indeed some clustering techniques can readily be applied for image segmentation.
- Matching: When I tend to recognize what an object I want to spot in a picture (approximately) seems like, I can use this knowledge to locate the object in an image. This approach to segmentation is called matching.

Perfect image segmentation –i.e., each pixel is assigned to the correct object segment– is a goal that cannot be usually achieved. Indeed, because of the way a digital image is acquired from moving WCE, this may be impossible, since a pixel may straddle the "real" boundary of objects such that it partially belongs to two (or even more) objects. Most methods presented in literature – indeed most current segmentation methods– only attempt to assign a particular type of anomalies in WCE frames. Perfect image segmentation is also often not reached because of the occurrence of oversegmentation and undersegmentation. In the oversegmentation case, pixels belonging to the same object are classified as belonging to different segments. A single object is represented by two or more segments. In the undersegmentation case, the opposite happens: pixels belonging to different objects are classified as belonging to the same object.

The segmentation method proposed in this thesis is an adaptive segmentation technique based on statistical methods, the term adaptive implies that this segmentation method can get adapted to four types of anomalies present in WCE frames. The WCE videos I acquired from the Medica Superspeciality Hospital, Kolkata consists types of anomalies namely- Tapeworm, Small Erosions, Ulcers, Crohn's disease, and Vascular Ectasia and that's the reason the segmentation technique is limited to four to five types of anomalies. These anomaly regions are found by taking a statistical measurement on the V and S channels of the HSV color space.

To understand the proposed segmentation algorithm, let us assume that the dimension of the input image in RGB color space is $m \times n \times 3$ and after conversion from RGB to HSV color space

using equation (1)-(6) in section 3.4, the dimension of the generated HSV image is $m \times n \times 3$. Next, splitting the HSV image to H, S and V channels, the dimension of each channel will be $m \times n$.

Let us assume the intensity value of any pixel in HSV image say (i, j) be $I(i,j)$. The entire method is explained using 2×2 patches from V and S channels respectively considering four adjacent pixels denoted by *a, b, c, d* as depicted in Fig. 24

.

| $I(i,j)$ | $I(i+1,j)$ |
|---|---|
| *(a)* | *(b)* |
| $I(i,j+1)$ | $I(i+1,j+1)$ |
| *(c)* | *(d)* |

*Fig. 24 Pictorial representation of the 2x2 patch*

After splitting the generated HSV image into H, S, and V channels respectively, I can represent the following channels as:

For H channel $\rightarrow a' = I(i,j), b' = I(i,j+1), c' = I(i+1,j), d' = I(i+1,j+1)$

For S channel $\rightarrow a'' = I(i,j), b'' = I(i,j+1), c'' = I(i+1,j), d'' = I(i+1,j+1)$

For V channel $\rightarrow a''' = I(i,j), b''' = I(i,j+1), c''' = I(i+1,j), d''' = I(i+1,j+1)$

The adaptive statistical method begins by taking the mean of all the pixel values in the array of V channel and quantifying with a constant factor *k,* where *k* is introduced as a reinforcing factor which strengthens the mean value as illustrated in equation (7). In the mean calculation, only the pixels with non-zero intensities is considered since it is observed that pixels having zero intensities mostly represent the surrounding black regions which are termed as artifacts in the proposed method. This obtained mean value (Mean of V channel) is then subtracted from the array of pixel values in the V channel, which generates a New V channel as illustrated in equation (8) and Fig. 26. It is being thoroughly observed that region with anomalies shares high pixel values in the V channel. Therefore, after subtraction, pixel values greater than the Mean of V channel implies positive pixel values and pixel values less than the Mean of V channel implies negative pixel values as shown in Fig. 25, and as the properties of the V channel implies the lower the pixel values of all the colors, similar to black and higher the pixel values of all the color, similar to the color itself.

Considering the example of pixel representation mentioned above, the discussed adaptive statistical method can be represented as follows,

$$Mean\ of\ V\ channel\ = k \times \left( \frac{a''' + b''' + c''' + d'''}{4} \right) \tag{7}$$

$$New\ V\ channel = (V\ channel - Mean\ of\ V\ channel) \tag{8}$$

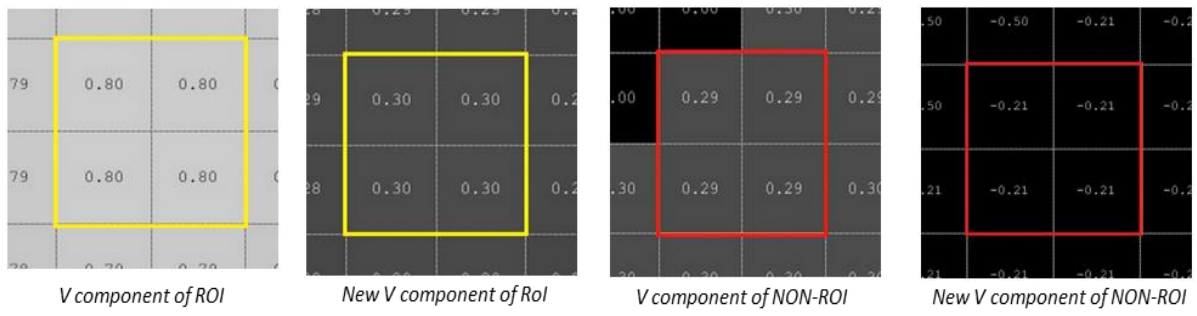*Fig. 25. Figure depicting the influence of mean value on the proposed segmentation method*

| | *V channel* | *New V channel* |
|---|---|---|
| Tapeworm | | |
| Bleeding Region | | |

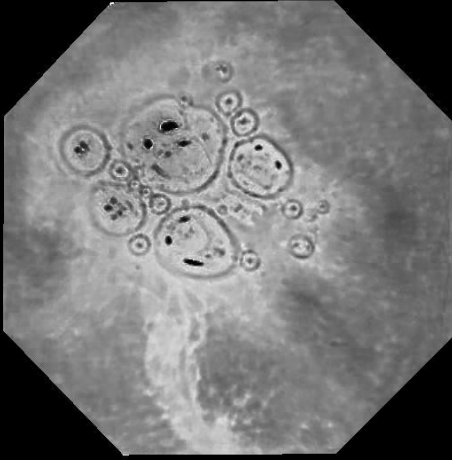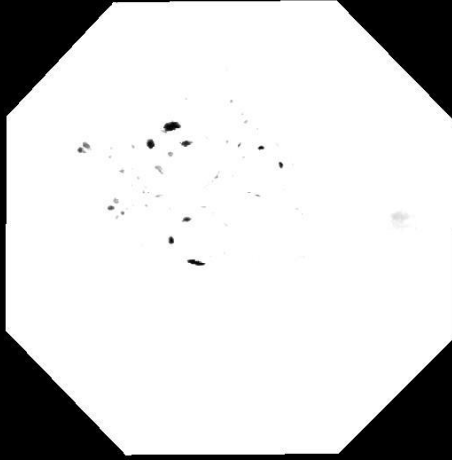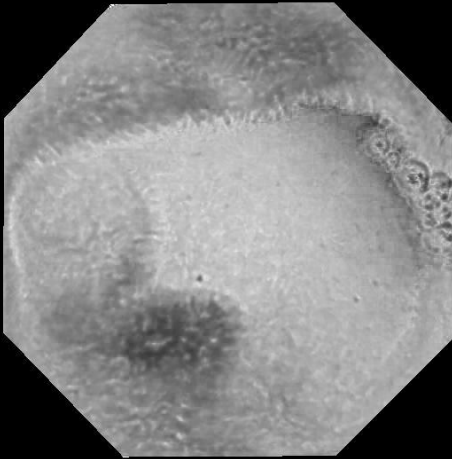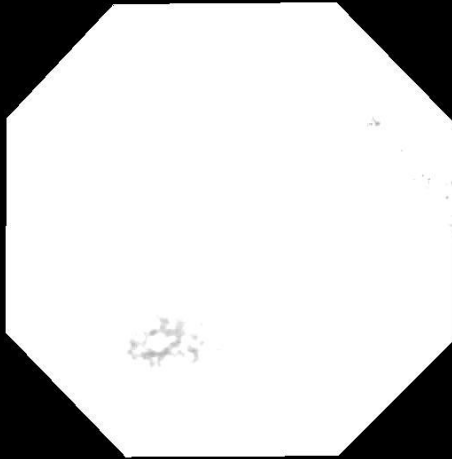| | |
|---|---|
| Vascular Ectasia |  |
| Small Ulcer |  |
| Crohn's Disease |  |
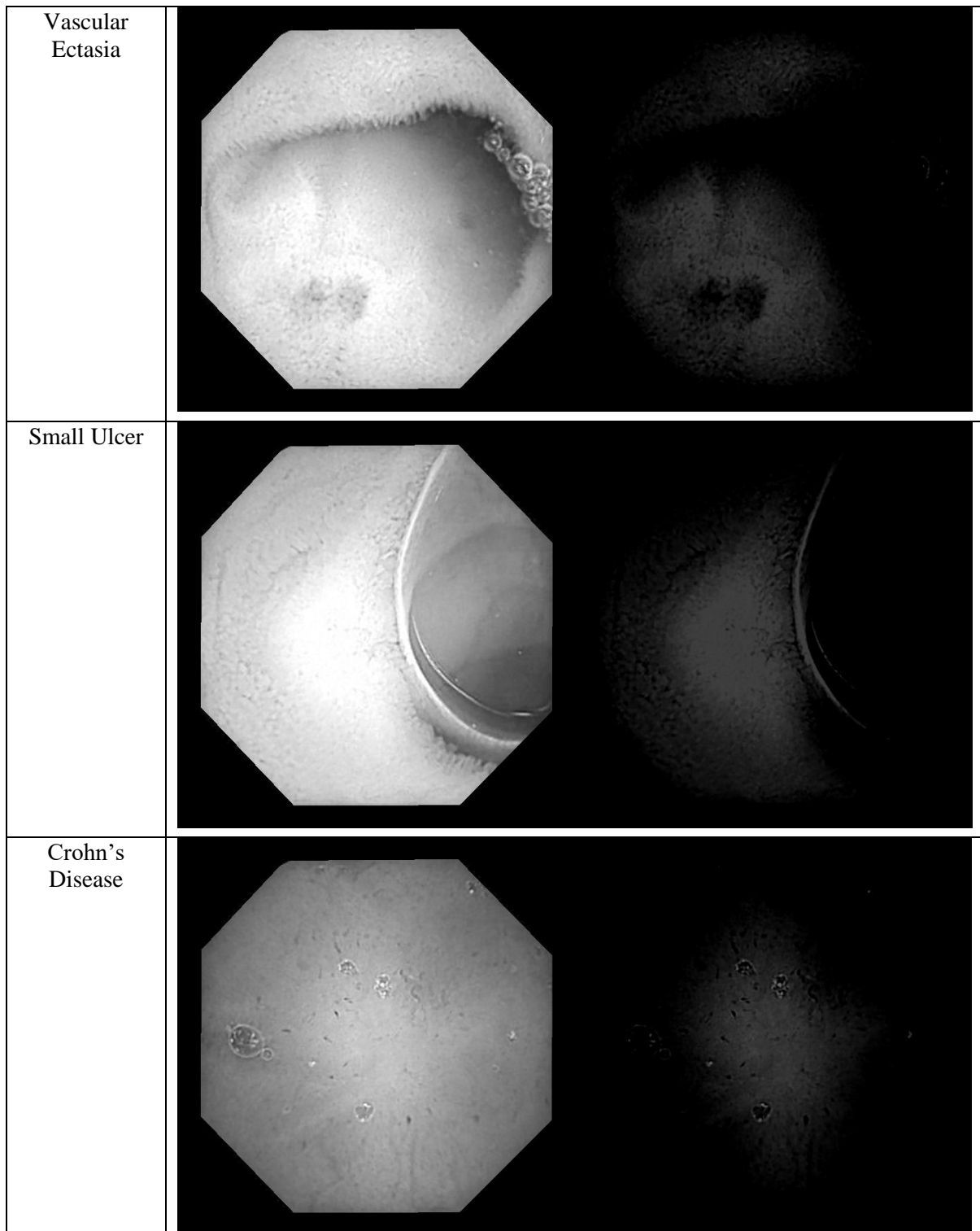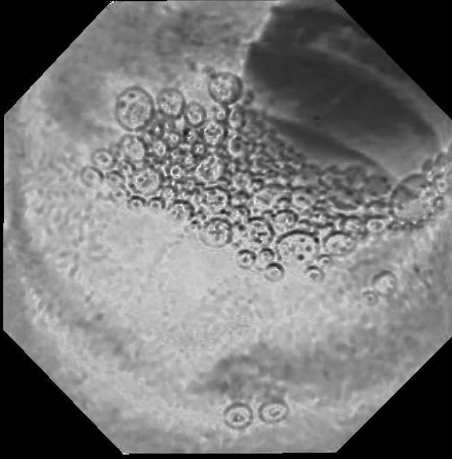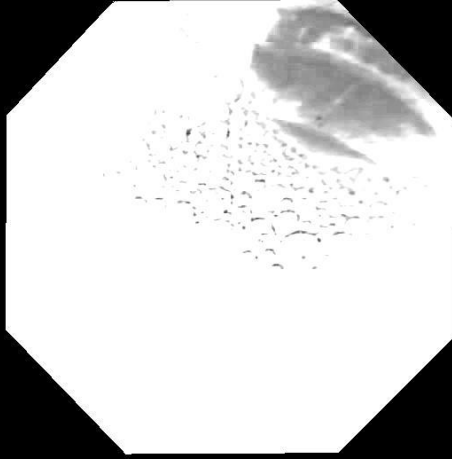
*Fig.26. Pictorial representation of V channel and New V channel of few anomalus regions*

Next, to enhance the anomalous region, a New S channel is generated by quantifying a constant factor $l$ with the pixel values of the S channel as illustrated in equation (9) Fig. 27.

$$New\ S\ channel\ =\ l \times [a''\ b''\ ;\ c''\ d'']$$ (9)

| | *S channel* | *New S channel* |
|---|---|---|
| Tapeworm | | |
| Bleeding Region | | |
| Vascular Ectasia | | |

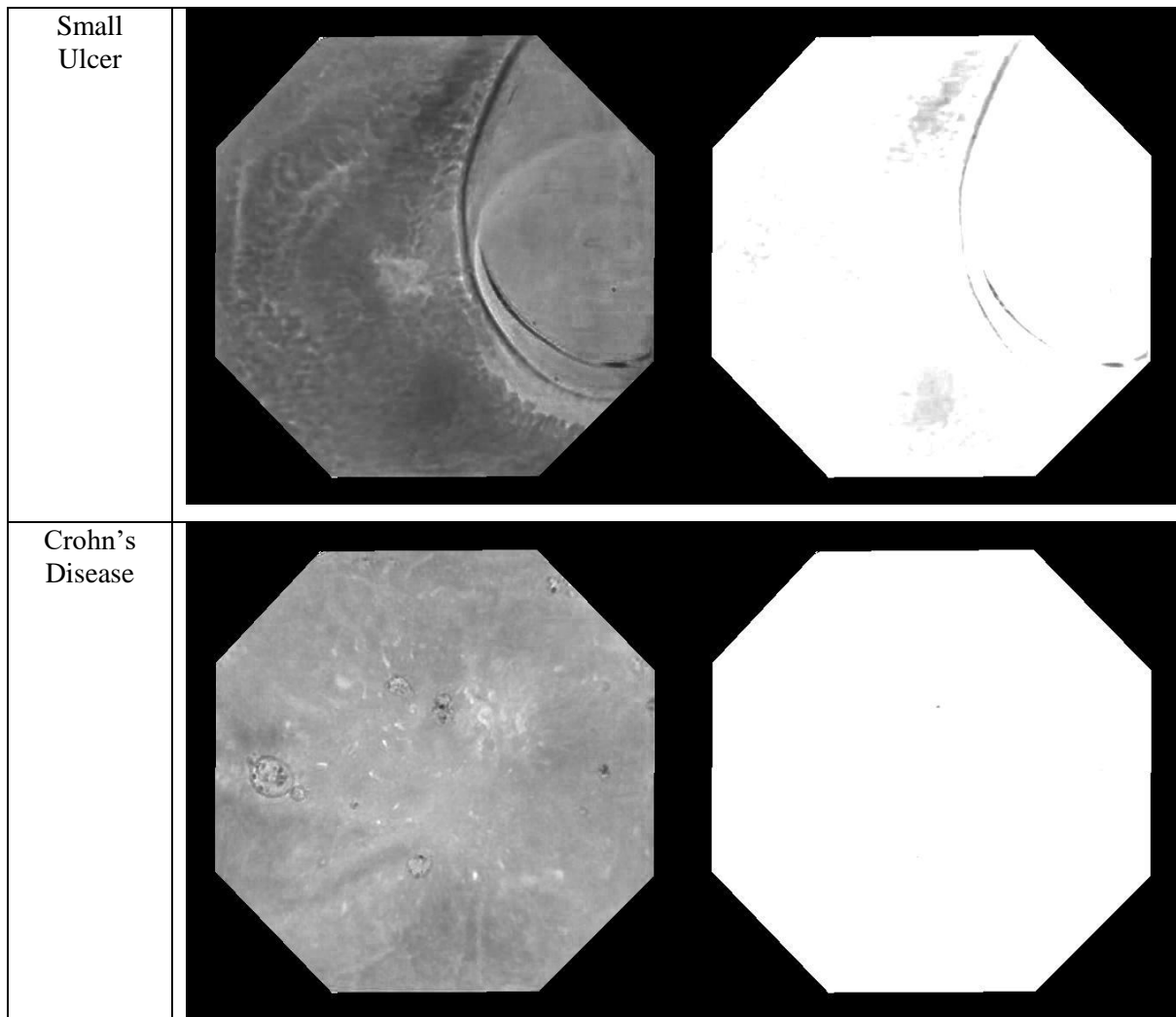| | | |
|---|---|---|
| Small Ulcer |  | |
| Crohn's Disease |  | |

*Fig. 27 Pictorial representation of the S channel and New S channel of few anomalous regions*

Based on the observation and experimental results the constant factors $k$ and $l$ above are set to $k = 1.5 \,\&\, l = 3$ which yield optimal performance in the proposed method. A new RGB image is formed by concatenating the above New V component and New S component with the original H component to form a new HSV color space and converting back to RGB color space. To make a significant segmentation two morphological operations are carried out on the binary image of the newly obtained RGB image, first closing and then filling holes.
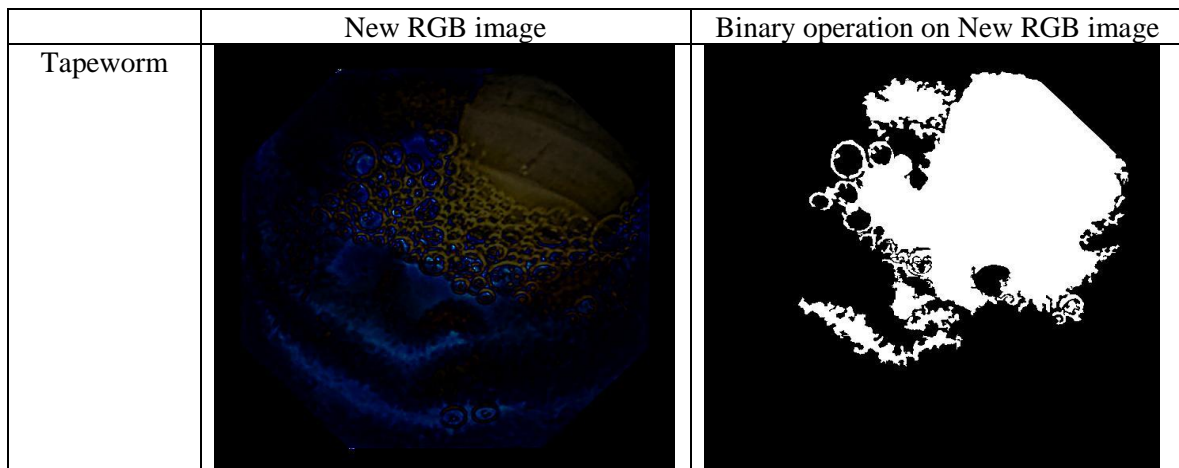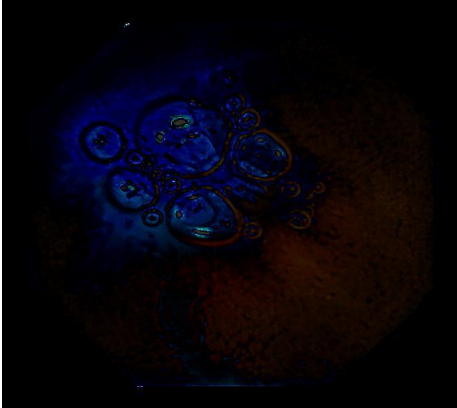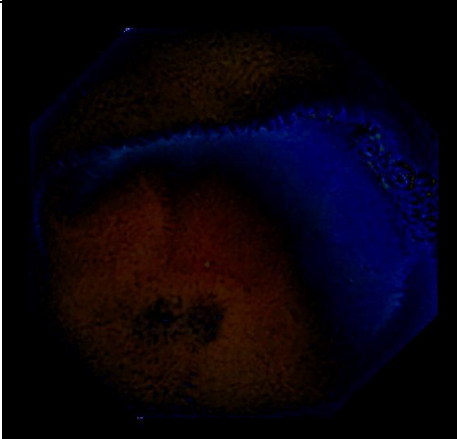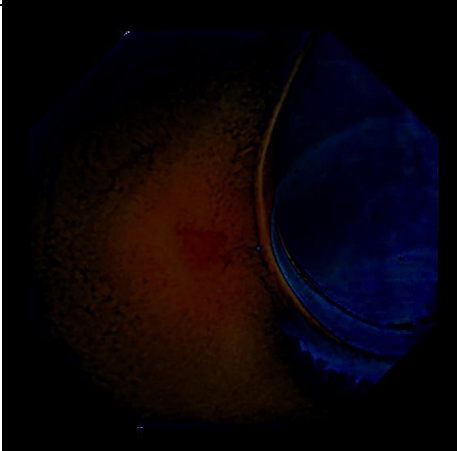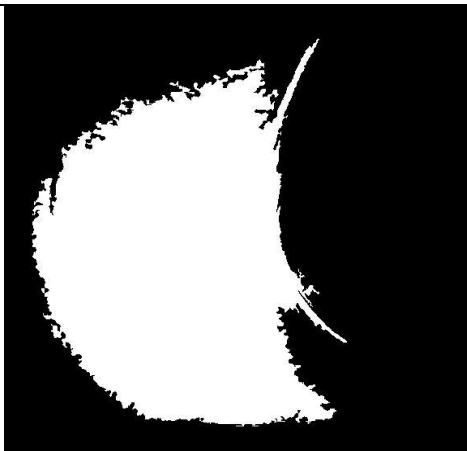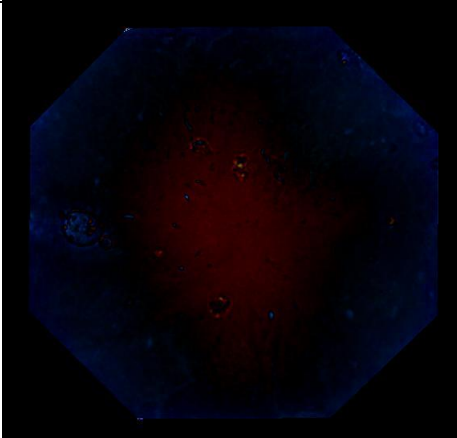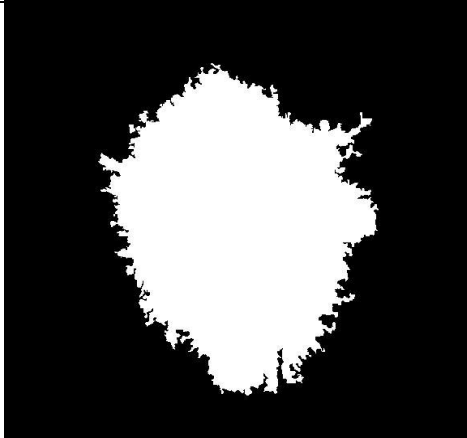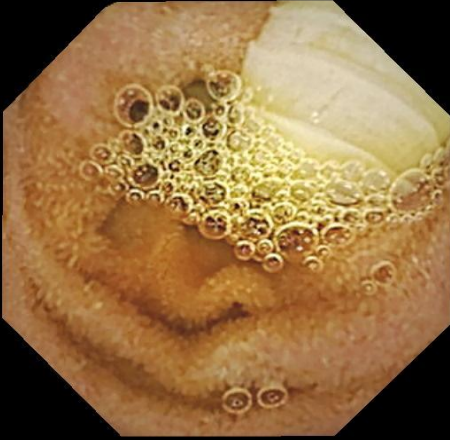
| | New RGB image | Binary operation on New RGB image |
|---|---|---|
| Tapeworm |  |  |

| | | |
|---|---|---|
| Bleeding Region | | |
| Vascular Ectasia | | |
| Small Ulcer | | |
| Crohn's Disease | | |

*Fig. 28 Pictorial representation of New RGB image and Binary operated New RGB image*

After this process, an image with the anomalous region is segmented out, keeping other regions black as shown in Fig. 29.

| | Input RGB image | Segmented Output RGB image |
|---|---|---|
| Tapeworm |  |  |
| Bleeding Region |  |  |
| Vascular Ectasia |  |  |

| | | |
|---|---|---|
| Small Ulcer |  | |
| Crohn's Disease |  | |
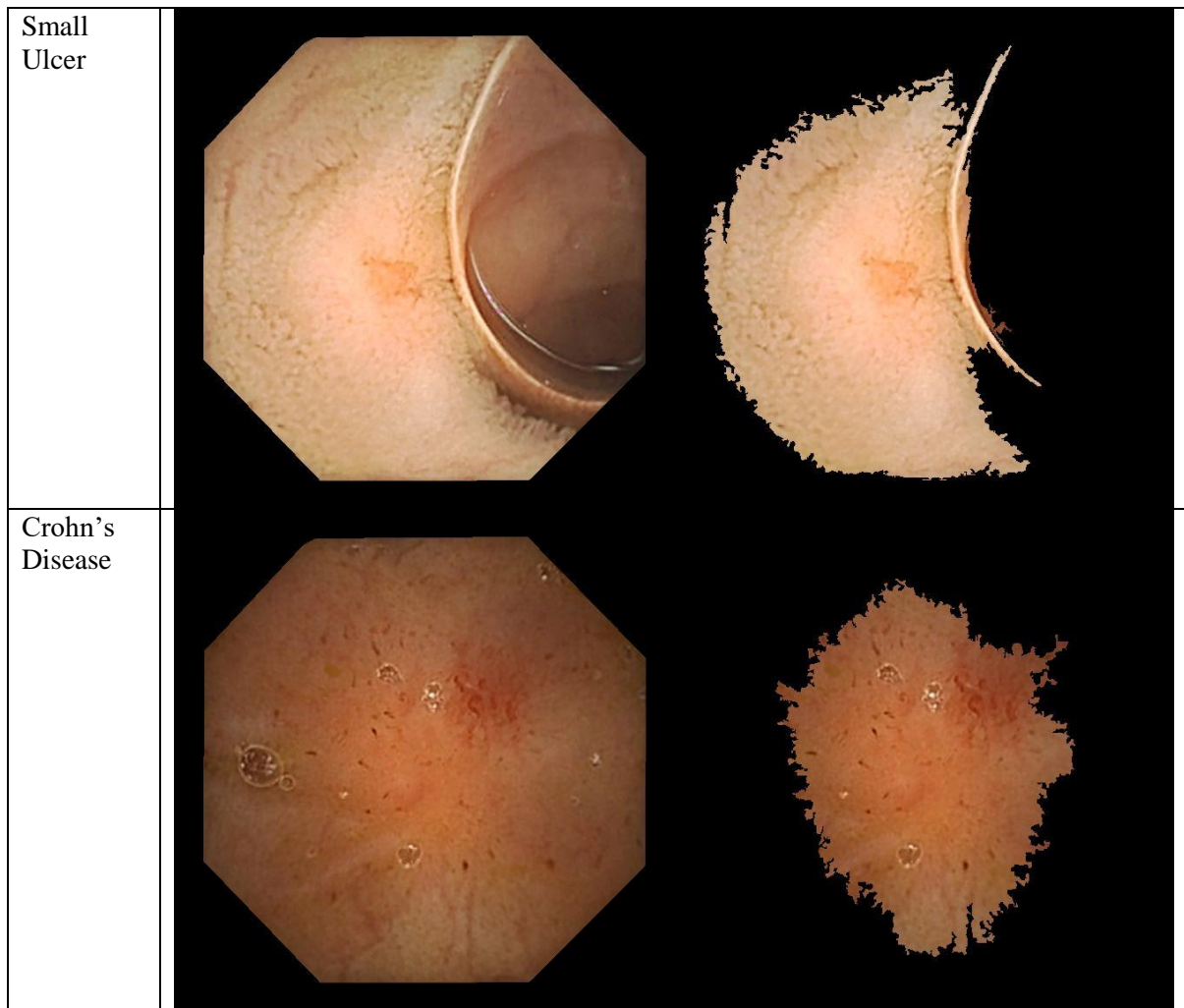
*Fig. 29 Pictorial representation of Input RGB image and Segmented Output RGB image*

This proposed segmentation algorithm is applied to the dataset which consists of ulcers, vascular ectasia, erosions, tapeworm, and bleeding regions, where it successfully segmented the above anomalous regions with an accuracy of 90.20%. The accuracy is determined by verifying its ground truth under supervision of an expert associated with the Medica Superspeciality Hospital, Kolkata, India.

# 5. CLASSIFICATION AND CNN ARCHITECTURE

# 5. CNN Architecture and Classification

Classification between the objects seems to be an easy task for humans but it has proved to be a complex task for machines[36]. Within the field of computer vision, classification is an important task. Image classification refers to the labeling of pictures into one in every variety of predefined classes. In general, the classification includes image sensors, image pre-processing, object detection, object segmentation, feature extraction and object classification[37]. The classification system consists of a database that consists of predefined patterns that will be compared with the detected object to classify into the proper category. Image classification is an important and challenging task in various application domains, including biomedical imaging, video surveillance, biometry, vehicle navigation, robot navigation, industrial visual inspection, and remote sensing.

WCE image data are produced by imaging modalities. The issue of this field is how to extract the image and classify into a similar pattern by the extraction result. Then identify and understand which parts of the GI tract human body are affected by the specific disease from image classification result.

There are three stages of Medical Image analysis tasks, which include (1) feature extraction and representation, (2) feature selection that will be used for classification, and (3) feature and image classification[37]. Moreover, in computer-aided-diagnosis image classification needs to play the most important role. Furthermore, medical image classification has three main steps: pre-processing, feature extraction and classification. After the pre-processing step, it needs to extract features of interest part from the image for further analysis. The purpose of the pattern classification system is to map the input variables (such as record information or image data) become the output variables to represent one specific category (with a disease or with no a disease class)[38]. Image classification could be a massive challenge on image analysis tasks, particularly the selection of methods and techniques in exploiting the result of image processing and pattern recognition, classification methods, subsequently validating the image classification result into medical expert knowledge[38]. The main objective of medical images classification isn't solely to succeed in high accuracy but also to identify which parts of the human body are infected by the disease.

WCE image is considered as a representation of anomalies with specific properties that are employed in image processing. The medical image analysis tasks consist of Feature extraction and representation, Feature selection that will be used for classification, and Feature and image classification.

Feature Extraction and Representation: Features are an important measurement for WCE image understanding, especially the feature representation of the segmented region that is used for anomalies classification and analysis. The techniques for Feature extraction and representation includes- (1) Statistical Pixel-Level (SPL) Features, (2) Shape Feature, (3) Texture Features, (4) Relational Features. These features provide information about the relational and hierarchical structure of the regions related to a single anomaly or a group of anomalies.

Feature Selection for Classification: Feature selection is used to discover the important features that are most appropriate for the classification task. Selection of correlated features for dimensionality reduction in the classification task will improve the computational efficiency and classification performance. The final set of features will be determined through data correlation, clustering, and analysis algorithms to explore similarity patterns in the training data. Feature selection for classification techniques includes- (1) Linear Discriminant Analysis (LDA), (2) Principal Component Analysis (PCA), (3) GA (Genetic Algorithms)-Based Optimization.

Feature and Image Classification: The selected features of image representation that are produced from feature selection, are used in object recognition and characterization. In the WCE imaging analysis, features and measurements can also be used for region segmentation to extract meaningful structures, subsequently, interpret the result using knowledge-based model and classification methods. Feature and image classification techniques include (1) Statistical Classification Methods, (2) Rule-Based Systems, (3) Neural Network Classifiers, (4) Support Vector Machine (SVM) for Classification. A typical classification system showed in Fig.30

Input Data

Classification

| Feature Selection/ Extraction | → | Inter-pattern Similarity | → | Clustering and/or Classifier | → |

*Fig. 30 A classification system*

In the field of image classification, several neural networks, deep CNN architectures have been proposed, like AlexNet, GoogleNet, ResNet, VGGNet, InceptionV3, etc. In this thesis, three classic CNN architectures are explored i.e. AlexNet, GoogleNet, InceptionV3 and an SVM classifier using Bag-of-Features model for image categorization. According to the experimental results (enclosed in section IV), the InceptionV3 provided the highest accuracy for the classification task compared to the other mentioned CNN and SVM classifier. This network is also computationally efficient for the specific image problem because this network has the capability of learning rich feature representation for a wide range of images.

**5.1 CNN Architecture**

An artificial neural network is ideal for feature classification, object recognition and image interpretation namely back-propagation, radial basis function, associative memories, and self-organizing feature maps. At time fuzzy system-based approaches have been applied in artificial neural networks for better classification and generalization result.

In a general Neural Network, there are three types of layers [39], Fig 31:

Input Layers: It is the layer in which the input is presented to the model. The number of neurons in this layer is equal to the total number of features in the data (number of pixels in case of an image).

Hidden Layer: The input from the Input layer is then feed into the hidden layer. There are many hidden layers depending upon the model and data size. Each hidden layer can have different numbers of neurons which are generally greater than the number of features. The output from every layer is computed by matrix multiplication of output of the previous layer with learnable weights of that layer and then by the addition of learnable biases followed by activation function which makes the network nonlinear.

Output Layer: The output from the hidden layer is then fed into a logistic function like sigmoid or softmax which converts the output of each class into probability score of each class.
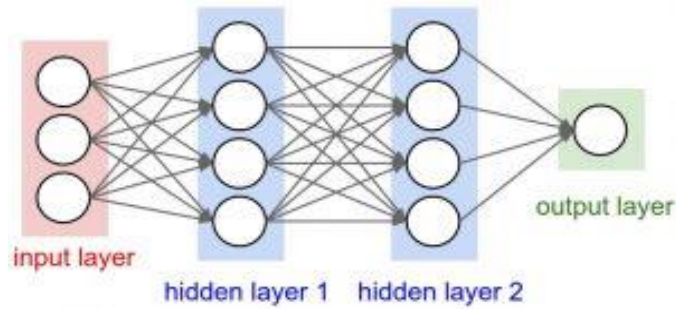
*Fig. 31 Typical Neural Network*

In neural networks, Convolutional neural network [40] (ConvNets or CNN) are a special kind of multi-layer neural networks, designed to recognize visual patterns directly from pixel images with minimal preprocessing.

To train and test, Deep learning CNN models, each input image will pass through a series of convolution layers with filters (Kernals), Pooling, fully connected layers (FC) and apply Softmax function to classify an object with probabilistic values between 0 and 1. Fig. 32 is a complete flow of CNN to process an input image and classifies the objects based on values.
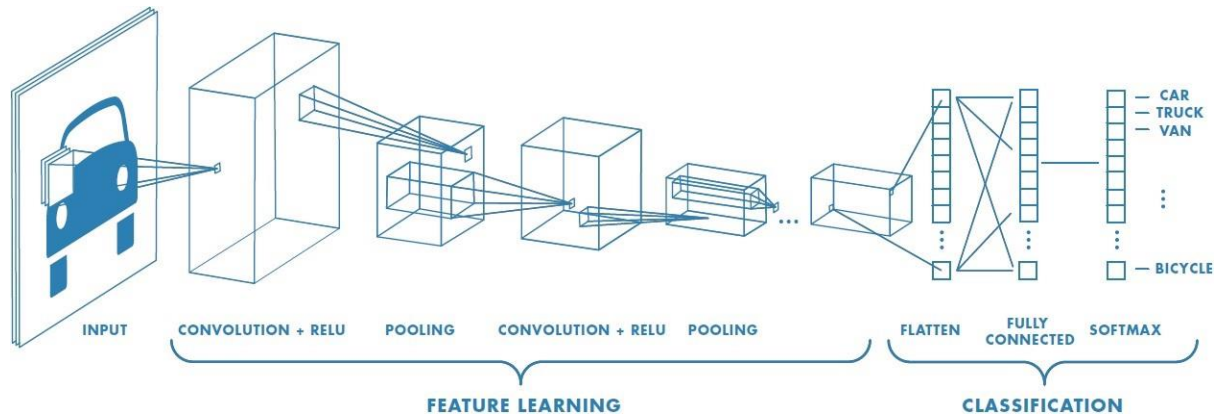


*Fig. 32. The workflow of a CNN model*

Input Layer: This layer holds the raw input of image with width 32, height 32 and depth 3.

Convolution Layer: This layer computes the output volume by the computing dot product between all filters and image patch. Suppose I use total 12 filters for this layer I'll get output volume of dimension 32 x 32 x 12.

Activation Function Layer: This layer will apply the element-wise activation function to the output of the convolution layer. Some common activation functions are RELU: $max(0, x)$, $Sigmoid$: $1/(1 + e^\wedge - x)$, Tanh, Leaky RELU, etc. The volume remains unchanged hence output volume will have dimension 32 x 32 x 12.

Pool Layer: This layer is periodically within the covnets and its main function is to scale back the dimensions volume which makes the computation fast reduces memory and also prevents from overfitting. Two common sorts of pooling layers are max pooling and average pooling. If I use a max pool with 2 x 2 filters and stride 2, the resultant volume will be of dimension 16x16x12.

Fully-Connected Layer: This layer is regular neural network layer which takes input from the previous layer and computes the class scores and outputs the 1-D array of size equal to the number of classes.

There is numerous pretrained CNN architecture available online, but as mentioned in the previous section, the CNN architectures I explored are AlexNet, GoogleNet, InceptionV3 so I will be focusing on these three types of pre-trained neural network.

AlexNet[41]: AlexNet is one of the most basic and classic deep CNN networks, which is composed of several effective processing techniques including ReLU, training on GPUs, overlap pooling and local response normalization (LRN). AlexNet is 8 layers deep with 5 convolution layer and 3 fully connected layer, 3 max-pooling layers, and 1 softmax layer. Each convolutional operation is followed by a ReLU layer and each fully connected layer is followed by a dropout layer. The input dimension of images is 227-by-227. Fig. 33 shows a representation of AlexNet CNN architecture layers.
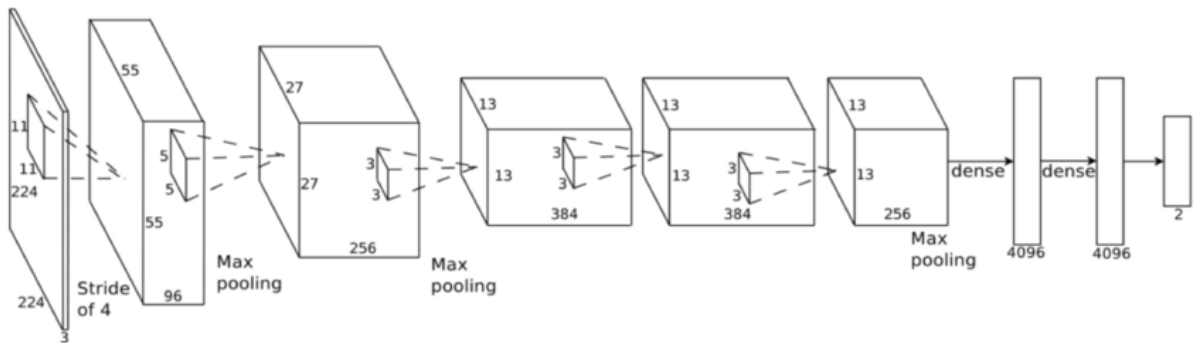


*Fig. 33. Architecture layers of AlexNet*

GoogleNet [42]: GoogleNet is composed of 22 convolutional layers, including 9 inception layers. The inception layer is of depth 2 which has 3 different sizes of kernel filters 5×5, 3×3, 1×1 for convolution and one 3×3 filter for pooling. The input dimension of images is 224-by-224. GoogleNet controls the computation by adding a bottleneck layer by using a 1x1 convolutional filter, before employing large size kernels. It used sparse connections to overcome the problem of redundant information and reduced cost by omitting channels that were not relevant. Furthermore, connections density get reduced by using global average pooling at the last layer, instead of using a fully connected layer. Fig. 34 shows a representation of GoogleNet CNN architecture layers.



*Fig. 34. Architecture Layers of GoogleNet*

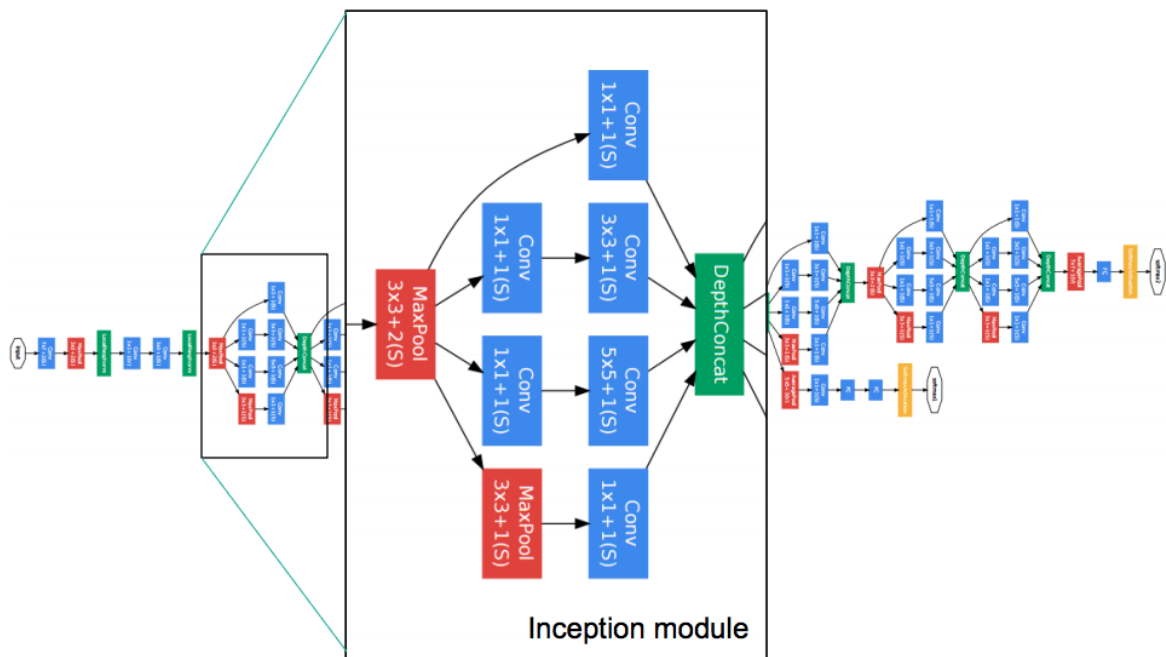Inception-V3[43]: The network is 48 layers' deep. This CNN architecture model consists of symmetric and asymmetric building blocks, (convolutions, average pooling, max pooling, concats, dropouts, and fully connected layers). Batch-norm is applied to activation inputs and used extensively throughout the model. The loss is computed via softmax and the network has an image input size of 299-by-299. The idea of Inception-V3 was to reduce the computational cost of deeper Nets without affecting the generalization. In Inception-V3, the 1x1 convolutional operation was used, which maps the input data into 3 or 4 separate spaces that are smaller than the original input space, and then maps all correlations in these smaller 3D spaces, via regular 3x3 or 5x5 convolutions. Fig. 35 shows a representation of InceptionV3 CNN architecture layers.
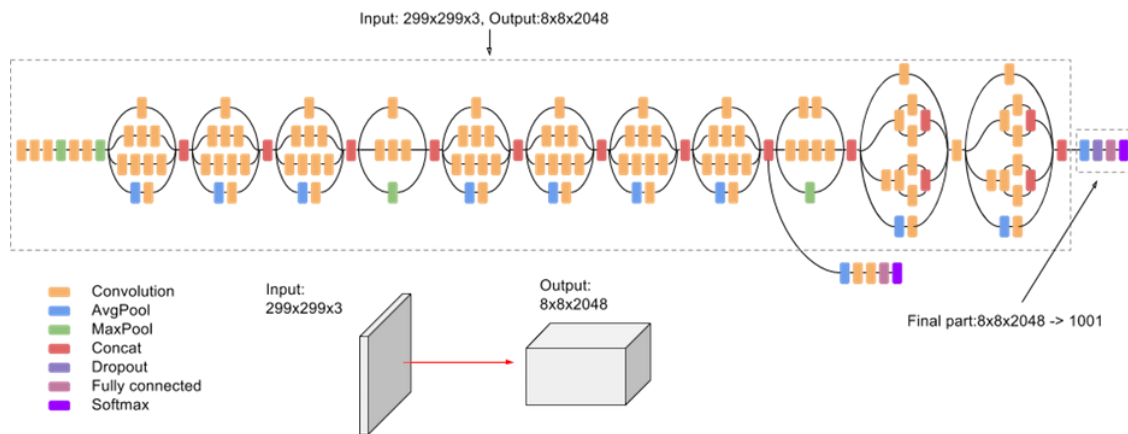


*Fig. 35. Architecture Layers of Inception-V3*

## 5.2 Support Vector Machine (SVM)

A support vector machine builds a hyperplane or set of hyperplanes in a high- or infinite- dimensional space, used for classification. Good separation is achieved by the hyperplane that has the largest distance to the nearest training data point of any class (functional margin), generally larger the margin lowers the generalization error of the classifier. SVM uses Non-parametric with binary classifier approach and may handle a lot of input file terribly with efficiency. Performance and accuracy depend on the hyperplane choice and kernel parameter. Fig. 37 shows a pictorial example of a classic SVM classifier.



*Fig. 36. SVM Classifier*

## 5.3 Fine Tuning

Image classification using the pre-trained network is quite popular as it requires fewer input data compared to millions of data required for training the network from scratch. Moreover, pre-trained networks are also used as feature extractors, as they can extra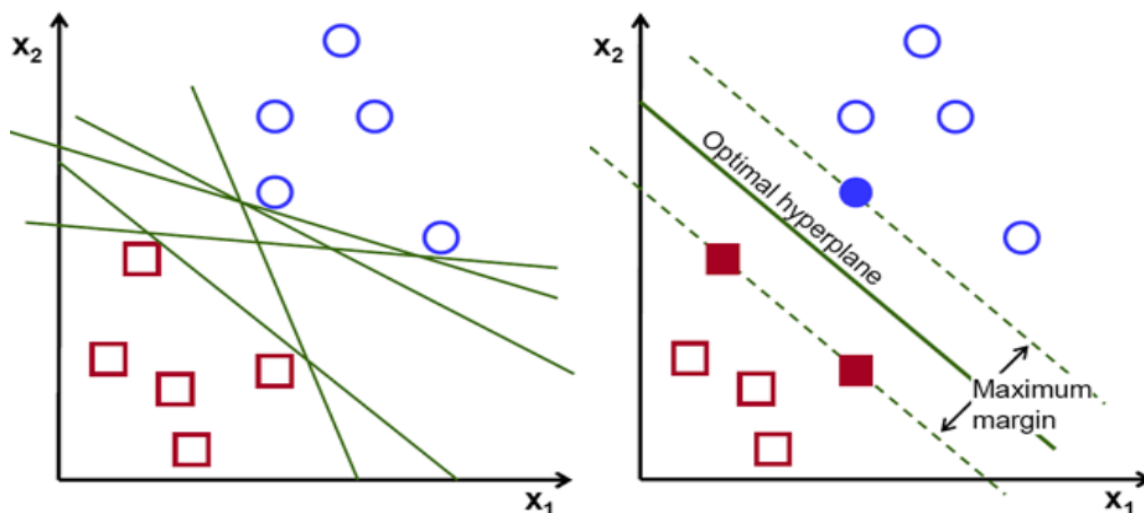ct both low-level and high-level features from images effectively. Most of the pre-trained networks mentioned are trained using the ImageNet database. These networks are made to be trained on more than a million images and classify them into 1000 object categories. In our situation, the categories are divided into two classes i.e. ROI and NON-ROI, so fine tuning the network according to the objective is necessary. Convolutional layers of the networks are used to extract image features that the last learnable layer and the final classification layer use to classify the input image. In most of the trained neural network, these two layers, 'loss3-classifier' and 'output' contain information on how to combine the features in order to predict class probabilities, a loss value and predicted labels. So, to retrain a pre-trained network for new images classification, I replaced these two layers with new layers adapted to the new dataset. Finally, I selected an optimal MiniBatchSize, MaxEpochs and learning rate to accelerate the training and better handling of the memory.

## 5.4 Data Augmentation

Image data augmentation [44] is a technique which will be accustomed by artificial means expand the scale of a training dataset by making modified versions of images within the dataset. Training deep learning neural network models on more data can result in more skillful models, and the augmentation techniques can create variations of the images that can improve the ability of the fit models to generalize what they have learned to new images

In order to train the networks accurately, the dataset has been augmented to automatically resize the training images so that it fulfills the criteria of the input layer of the pre-trained networks. Also, augmentation operations are performed on the training images like random flipping along the vertical axis and random translation up to 30 pixels horizontally and vertically. Data augmentation prevents the network from overfitting and memorizing the exact details of the training images.

# 6. RESULTS

# 6. Results

## 6.1 Database Preparation

To evaluate the performance of the proposed method I constructed the dataset consisting of 2100 segmented anomalous regions (Segmented ROI) and 2000 segmented non-anomalous regions (Segmented NON-ROI). These images were extracted from different patient's WCE images consisting ulcers, vascular ectasia, tapeworm, small bowel erosions, gastrointestinal bleedings and each of them were manually annotated by an expert physician from Medica Superspeciality Hospital, Kolkata, India. Images with anomalous regions are labeled as ROI and others as NON-ROI.

During the operation I randomly sampled 70% of the dataset into a training dataset, to train the classifier and rest 30% to test dataset for testing the well-trained model.

## 6.2 CNN Architecture and SVM

I optimized 3 CNN networks i.e. AlexNet, GoogleNet, InceptionV3, on the augmented training dataset and tested them on the test dataset. AlexNet achieved an accuracy of 85.2%, GoogleNet with 92% whereas InceptionV3 has achieved an accuracy of 95.49%. For comparison, I employed a Bag of Features technique to train an SVM classifier with the same experimental settings. Finally, it achieves an accuracy of 72.34%, demonstrating the necessity of using deep learning techniques. From Table 3, it can be noticeable that Inception V3 gives the best performance in this application using a lesser number of features.

| No | PERFORMANCE ANALYSIS | | |
|----|------------------------------------|-----------------|-----------|
|    | *Network* | *No. of features* | *Accuracy* |
| 1  | SVM trained with Bag- of - Features | 7375872 | 72.34% |
| 2  | AlexNet | 4096 | 85.19% |
| 3  | GoogleNet | 1024 | 92.00% |
| 4  | **InceptionV3** | **2048** | **95.49%** |

Table- 3 Comparative Analysis of Performances of Networks

To understand the performance of the classifiers model, a respective confusion matrix is shown in Fig. 38. In this figure, the first two green diagonal cells show the number and percentage of correct classification, whereas red cells show the number and percentage of incorrect classification. The column on the far right of the matrix shows the percentages of all the data predicted to belong to each class that is correctly and incorrectly classified. These metrics are called the precision or positive predictive value and false discovery rate, respectively. The row at the bottom of the plot shows the percentages of all the data belonging to each class that is correctly and incorrectly classified. These metrics are often called the recall or true positive rate and false negative rate, respectively. The cell in the bottom right of the plot shows the overall accuracy on the yellow cell.

## Confusion Matrix for AlexNet

| Output Class | | Target Class | | |
|---|---|---|---|---|
| | | NON_ROI | ROI | |
| | NON_ROI | **521**<br>41.3% | **78**<br>6.2% | 87.0%<br>13.0% |
| | ROI | **109**<br>8.6% | **555**<br>43.9% | 83.6%<br>16.4% |
| | | 82.7%<br>17.3% | 87.7%<br>12.3% | 85.2%<br>14.8% |

## Confusion Matrix for GoogleNet

| Output Class | | Target Class | | |
|---|---|---|---|---|
| | | NON_ROI | ROI | |
| | NON_ROI | **555**<br>43.9% | **26**<br>2.1% | 95.5%<br>4.5% |
| | ROI | **75**<br>5.9% | **607**<br>48.1% | 89.0%<br>11.0% |
| | | 88.1%<br>11.9% | 95.9%<br>4.1% | 92.0%<br>8.0% |

## Confusion Matrix for Inception-V3

| Output Class | | Target Class | | |
|---|---|---|---|---|
| | | NON_ROI | ROI | |
| | NON_ROI | **579**<br>45.8% | **6**<br>0.5% | 99.0%<br>1.0% |
| | ROI | **51**<br>4.0% | **627**<br>49.6% | 92.5%<br>7.5% |
| | | 91.9%<br>8.1% | 99.1%<br>0.9% | 95.5%<br>4.5% |

| Confusion Matrix for SVM | | | |
|---|---|---|---|
| | Target Class | | |
| | **NON_ROI** | **ROI** | |
| **NON_ROI** | **385**<br>30.5% | **246**<br>19.5% | **61.0%**<br>39.0% |
| **ROI** | **101**<br>8.0% | **530**<br>42.0% | **84.0%**<br>16.0% |
| | **79.21%**<br>20.7% | **68.3%**<br>31.70% | **72.0%**<br>28.0% |

*Output Class* appears as a row label on the left side spanning the data rows.

## 6.3 Results

The proposed method has obtained a segmentation accuracy of 90.20% at the completion of the first stage of the proposed algorithm. But at the end of this stage, it is observed that some NON-ROI regions are erroneously getting segmented as ROI regions, which reduces the overall segmentation accuracy of the method. Hence, to overcome this limitation, the second step of the algorithm is introduced. In this step, the segmented images obtained from the first step are classified into ROI and NON-ROI images using three fine-tuned, pre-trained CNN's namely AlexNet, GoogLeNet, and Inception V3 respectively. An SVM classifier trained with the Bag-of features is also used to classify those images into ROI and NON-ROI images. The accuracy obtained for each case is given in Table 3. From Table 3, it is noticeable that after the introduction of the second step, the segmentation accuracy of the proposed method increases by 1.4% and 4.89% by using GoogLeNet and Inception V3 respectively and decreases by 5.41% and 18.26% by using of AlexNet and SVM trained with Bag-of features respectively compared to the accuracy obtained after the first step. As Inception V3 gives the best performance compared to others, it has been chosen to carry out segmentation in combination with the novel adaptive statistical method over others. That's why Inception V3 gets highlighted in the performance analysis table, Table 3. Fig. 37 on the next page will illustrate the successful classifier results of random WCE images with segmented ROI and Non-ROI regions.
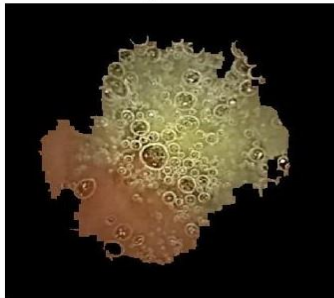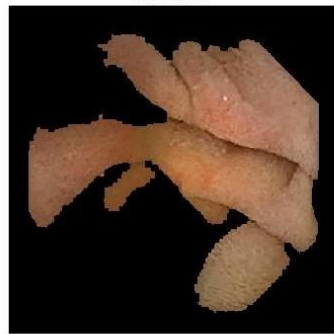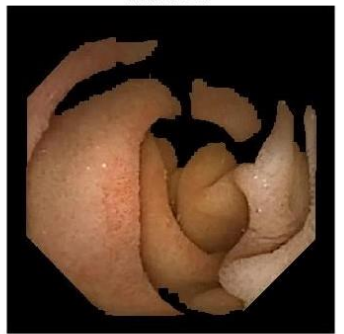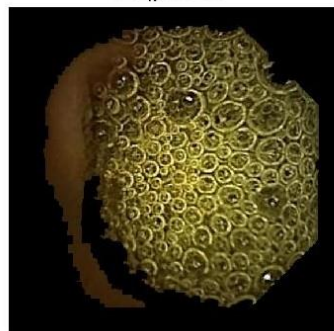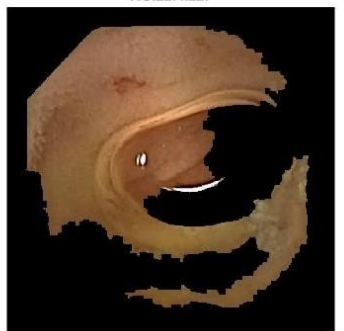
ROI227x227
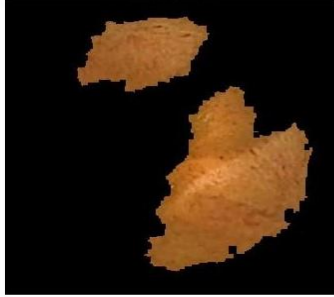

NON_ROI227x227
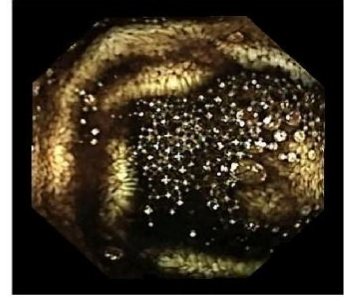

NON_ROI227x227
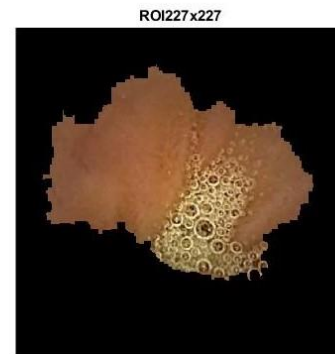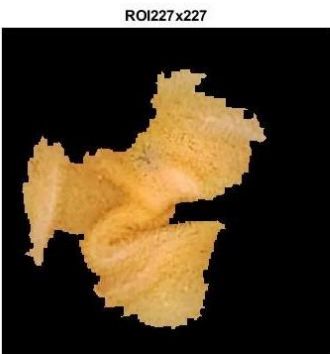

ROI227x227


ROI227x227

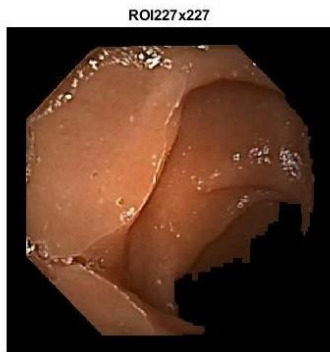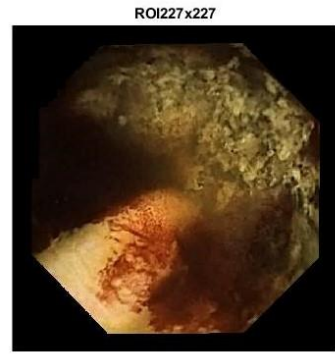
ROI227x227


NON_ROI227x227


ROI227x227

41

*Fig. 37 Successful classification of random WCE segmented images*

# REFERENCES

# References

[1]     P. C. De Groen, "History of the Endoscope [Scanning Our Past]," *Proc. IEEE*, vol. 105, no. 10, pp. 1987–1995, 2017.

[2]     "Endoscopy: Types, preparation, procedure, and risks." [Online]. Available: https://www.medicalnewstoday.com/articles/153737.php. [Accessed: 30-Apr-2019].

[3]     C. S. Pitchumoni and N. G. Gidwaney, "Wireless capsule endoscopy," *Geriatr. Gastroenterol.*, pp. 221–226, 2012.

[4]     "Bleeding in Digestive Tract: Why It Happens &amp; How To Treat It." [Online]. Available: https://www.webmd.com/digestive-disorders/bleeding-digestive-tract#1. [Accessed: 02-May-2019].

[5]     "Crohn's disease - Symptoms and causes - Mayo Clinic." [Online]. Available: https://www.mayoclinic.org/diseases-conditions/crohns-disease/symptoms-causes/syc-20353304. [Accessed: 02-May-2019].

[6]     "Tapeworms in Humans: Causes, Symptoms, and Treatments." [Online]. Available: https://www.webmd.com/digestive-disorders/tapeworms-in-humans#1. [Accessed: 02-May-2019].

[7]     "What is Image Processing : Tutorial with Introduction, Basics, Types &amp; Applications." [Online]. Available: https://www.engineersgarage.com/articles/image-processing-tutorial-applications. [Accessed: 03-May-2019].

[8]     "Medical Image Processing : A Review," 2016.

[9]     M. Liedlgruber and A. Uhl, "Endoscopic image processing - an overview," pp. 707–712, 2015.

[10]    P. Y. Lau and P. L. Correia, "Detection of bleeding patterns in WCE video using multiple features," *Annu. Int. Conf. IEEE Eng. Med. Biol. - Proc.*, pp. 5601–5604, 2007.

[11]    H. Vu, T. Echigo, Y. Imura, Y. Yanagawa, and Y. Yagi, "Segmenting reddish lesions in capsule endoscopy images using a gastrointestinal color space," *Proc. - Int. Conf. Pattern Recognit.*, pp. 3263–3268, 2014.

[12]    W. Sdshu *et al.*, "6Dh +Zdqj Dqg 0 (Puh &Hohel," vol. 7, pp. 678–681, 2010.

[13]    E. Tuba, M. Tuba, and R. Jovanovic, "An algorithm for automated segmentation for bleeding detection in endoscopic images," *Proc. Int. Jt. Conf. Neural Networks*, vol. 2017-May, pp. 4579–4586, 2017.

[14]    S. Khan and N. S. D. Shree, "Detection Of Peptic Ulcers based on Thresholding and Watershed segmentation," 2015.

[15]    M. FIORI, P. MUSÉ, and G. SAPIRO, "a Complete System for Candidate Polyps Detection in Virtual Colonoscopy," *Int. J. Pattern Recognit. Artif. Intell.*, vol. 28, no. 07, p. 1460014, 2014.

[16]    X. Jia and M. Q.-H. Meng, "A study on automated segmentation of blood regions in Wireless Capsule Endoscopy images using fully convolutional networks," *Biomed. Imaging (ISBI 2017), 2017 IEEE 14th Int. Symp.*, pp. 179–182, 2017.

[17]    P. M. Vieira, B. Goncalves, C. R. Gonçalves, and C. S. Lima, "Segmentation of angiodysplasia lesions in WCE images using a MAP approach with Markov Random Fields," *Proc. Annu. Int. Conf. IEEE Eng. Med. Biol. Soc. EMBS*, vol. 2016-Octob, pp. 1184–1187, 2016.

[18]    P. Vieira *et al.*, "Segmentation of small bowel tumor tissue in capsule endoscopy images by using the MAP algorithm," *Proc. Annu. Int. Conf. IEEE Eng. Med. Biol. Soc. EMBS*, vol. i, pp.

4010–4013, 2012.

[19] A. A. Voronenko, "Testing disjunction as a read-once function in an arbitrary unrepeated basis," *Moscow Univ. Comput. Math. Cybern.*, vol. 32, no. 4, pp. 239–240, 2008.

[20] Y. Boykov and M.-P. Jolly, "Interactive Organ Segmentation Using Graph Cuts," pp. 276–286, 2011.

[21] C. Rother, V. Kolmogorov, and A. Blake, "'GrabCut': interactive foreground extraction using iterated graph cuts," *Acm Tg*, vol. 23, no. 3, p. 309, 2004.

[22] L. Grady, "Random walks for image segmentation," *IEEE Trans. Pattern Anal. Mach. Intell.*, vol. 28, no. 11, pp. 1768–1783, 2006.

[23] J. Freixenet, X. Muñoz, D. Raba, J. Martí, and X. Cufí, "Yet Another Survey on Image Segmentation: Region and Boundary Information Integration," pp. 408–422, 2007.

[24] P. Wang, S. M. Krishnan, Y. Huang, and N. Srinivasan, "An Adaptive Segmentation Technique for Clinical Endoscopic Image Processing," *Proc. Annu. Int. Conf. IEEE Eng. Med. Biol. Soc.*, pp. 1084–1085, 2002.

[25] P. Shanmugasundaram and N. Santhiyakumari, "Interactive segmentation of capsule endoscopy images using grow Cut method," *Proc. - 2014 6th Int. Conf. Comput. Intell. Commun. Networks, CICN 2014*, pp. 190–192, 2014.

[26] D. Boschetto, H. Mirzaei, R. W. L. Leong, and E. Grisan, "Superpixel-based automatic segmentation of villi in confocal endomicroscopy," *3rd IEEE EMBS Int. Conf. Biomed. Heal. Informatics, BHI 2016*, pp. 168–171, 2016.

[27] O. H. Maghsoudi, "Superpixel Based Segmentation and Classification of Polyps in Wireless Capsule Endoscopy," pp. 19–22, 2017.

[28] T. K. Kho, K. S. Sim, and F. F. Ting, "Gastrointestinal endoscopy colour-based image processing technique for bleeding, lesion and reflux," *Proc. 2016 Int. Conf. Robot. Autom. Sci. ICORAS 2016*, 2017.

[29] M. Sonka, V. Hlavac, and R. Boyle, "Image pre-processing," *Image Process. Anal. Mach. Vis.*, pp. 56–111, 1993.

[30] D. Pascale, "A Review of RGB Color Spaces," *BabelColor*, p. 35, 2003.

[31] M. Tkalčič and J. F. Tasič, "Colour spaces - Perceptual, historical and applicational background," *IEEE Reg. 8 EUROCON 2003 Comput. as a Tool - Proc.*, vol. A, pp. 304–308, 2003.

[32] "What Is the HSV (Hue, Saturation, Value) Color Model?" [Online]. Available: https://www.lifewire.com/what-is-hsv-in-design-1078068. [Accessed: 04-May-2019].

[33] "RGB to HSV conversion | color conversion." [Online]. Available: https://www.rapidtables.com/convert/color/rgb-to-hsv.html. [Accessed: 04-May-2019].

[34] B. Robot, "Hsv_Writeup," pp. 1–2, 2005.

[35] C. Nikou, "Chapter_10_I[1] C. Nikou, 'Chapter_10_Image_Segmentation,' 2008.mage_Segmentation," 2008.

[36] D. P. Lusch, "Digital Image Classification □," no. October, 2015.

[37] P. Kamavisdar, S. Saluja, and S. Agrawal, "A survey on image classification approaches and techniques," *nternational J. Adv. Res. Comput. Commun. Eng.*, vol. 2, no. 1, pp. 1005–1009, 2013.

[38]     H. H. and D.-S. K. 6 Atam P Dhawan, *PRINCIPLES AND ADVANCED METHODS IN MEDICAL IMAGING AND IMAGE ANALYSIS*, vol. ثق قفثق, no. نثق نقثقثق. World Scientific Publishing Co. Pte. Ltd. 5 Toh Tuck Link, Singapore 596224 USA office: 27 Warren Street, Suite 401-402, Hackensack, NJ 07601 UK office: 57 Shelton Street, Covent Garden, London WC2H 9HE British, 2008.

[39]     "Buzzword : CNN," 2018.

[40]     A. Mishra and H. Cheng, "Advanced CNN Architectures," 2017.

[41]     A. Krizhevsky, I. Sutskever, and G. E. Hinton, "2012 AlexNet," *Adv. Neural Inf. Process. Syst.*, pp. 1–9, 2012.

[42]     G. Smoluk, "Google net," *Mod. Plast.*, vol. 57, no. 3, pp. 62–63, 1980.

[43]     A. Khan, A. Sohail, U. Zahoora, and A. S. Qureshi, "A Survey of the Recent Architectures of Deep Convolutional Neural Networks," pp. 1–62, 2019.

[44]     L. Perez and J. Wang, "The Effectiveness of Data Augmentation in Image Classification using Deep Learning," 2017.

# FUTURE SCOPE OF WORK

## Future Scope Of Work

In this thesis, a new approach on segmentation, combining adaptive statistical method and CNN is presented. The segmentation of anomalies from WCE images can be achieved at the first stage of the proposed algorithm itself using statistical operations. As it gives erroneous results sometimes, the second stage of the algorithm is introduced, where it classified the segmented images obtained from the first step to ROI and NON-ROI classes using three CNN and a binary SVM classifier trained with Bag-of-Features respectively. Experimental results obtained using the proposed algorithm proves its' efficiency over other state-of-the-art methods. The reliability of this algorithm is also verified by professional experts. Currently, our algorithm is capable of segmenting four to five types of anomalies but in the near future, it will focus on further up gradation of the algorithm, so that it can work for all types of anomalies.