Automatic Bleeding Detection in Wireless Capsule Endoscopy Images

Thesis submitted in partial fulfillment of the requirements for the award of degree of

Master of Engineering
in
Computer Science and Engineering

Submitted By
Apoorva Singh
(Roll No. - 801732003)

Under the supervision of:
Dr. Husanbir Singh Pannu
Assistant Professor

COMPUTER SCIENCE AND ENGINEERING DEPARTMENT
THAPAR INSTITUTE OF ENGINEERING AND TECHNOLOGY
PATIALA – 147004

August 2019

CERTIFICATE

I hereby certify that the work which is being presented in the thesis entitled, **"***Automatic Bleeding Detection in Wireless Capsule Endoscopy Images***"**, in partial fulfillment of the requirements for the award of degree of Master of Engineering in *Computer Science and Engineering* submitted in Computer Science and Engineering Department of Thapar Institute of Engineering and Technology, Patiala, is an authentic record of my own work carried out under the supervision of Dr. Husanbir Singh Pannu and refers other researcher's work which are duly listed in the reference section.

The matter presented in the thesis has not been submitted for the award of any other degree of this or any other University.

Signature: Apoorva Singh

Place: Patiala 801732003

Date: August 2019

This is to certify that the above statement made by the candidate is correct and true to the best of my knowledge.

Signature:

Dr. Husanbir Singh Pannu, Assistant Professor

Computer Science and Engineering Department

Thapar Institute of Engineering and Technology Patiala (Punjab), India

ACKNOWLEDGEMENT

No volume of words is enough to express my gratitude towards my guide, **Dr**. **Husanbir Singh Pannu**. I thank him for his time, patience, discussions and valuable comments. He has been concerned and has aided for all material essential for the preparation of this thesis report. I am equally grateful to **Dr. Maninder Singh**, Head of Computer Science and Engineering Department and **Dr. Rinkle Aggarwal**, P.G. Coordinator, for motivation and inspiration that triggered me for the thesis work. I also want to express my gratitude to **Dr. S.S. Bhatia**, Dean of Academic Affairs, for making provisions of infrastructure such as library facilities, computer labs equipped with net facilities, immensely useful for the learners to equip themselves with the latest in the field. I am also thankful to the entire faculty and staff members of the Computer Science and Engineering Department for their help, cooperation and affection, which made my stay at **Thapar Institute of Engineering & Technology** memorable. I express my gratitude to **Dr. Dinesh Singhal** of the **PSRI hospital** (New Delhi, India) for providing me with the dataset to test my thesis on.

I am highly thankful to my sister, **Archna Singh** for her immense support without which I would not have been able to pursue my post-graduation. I also thank my friends, **Reaya Grewal** and **Namrata Dimari** for assisting me as per their abilities, in whatever manner possible throughout the process of research. Last but not the least, I would like to express my very profound gratitude to my parents and Almighty for showing me the right direction. This accomplishment would not have been possible without them.

ABSTRACT

Image segmentation in medical images is performed to extract valuable information from the images by concentrating on the region of interest. Mostly, the number of medical images generated from a diagnosis is large and not ideal to treat with traditional ways of segmentation using machine learning models due to their numerous and complex features. To obtain crucial features from this large set of images, deep learning is a good choice over traditional machine learning algorithms. Wireless capsule endoscopy images comprise normal and sick frames and often suffers with a big data imbalance ratio which is sometimes 1000:1 for normal and sick classes. They are also special type of confounding images due to movement of the (capsule) camera, organs and variations in luminance to capture the site texture inside the body. So, we have proposed an automatic deep learning model based to detect bleeding frames out of the WCE images. The proposed model is based on Convolutional Neural Network (CNN) and its performance is compared with state-of-the-art methods including Logistic Regression, Support Vector Machine, Artificial Neural Network and Random Forest. The proposed model reduces the computational burden by offering the automatic feature extraction. It has promising accuracy with an F1 score of 0.76.

Table of Contents

List of Figures

List of Tables

List of Abbreviations

ANN	Artificial Neural Network
AUC	Area Under Curve
BN	Batch Normalization
CE	Capsule Endoscopy
CIE	International Commission on Illumination
CNN	Convolutional Neural Network
FC	Fully Connected
FN	False Negative
FP	False Positive
GERD	Gastroesophageal Reflux Disease
GI	Gastrointestinal
GLCM	Gray Level Co-Occurrence Matrix
HOG	Histogram of Oriented Gradients
HPF	High Pass Filter
HRPF	High Reconstruction Pass Filter
HSI	Hue, Saturation, Intensity
HSV	Hue-Saturation-Value
IBD	Inflammatory Bowel Disease
KNN	K-Nearest Neighbors
LBP	Local Binary Patterns
LPF	Low Pass Filter
LRPF	Low Reconstruction Pass Filter
ML	Machine Learning
NN	Neural Network
PR	Precision-Recall
ReLU	Rectified Linear Unit
RGB	Red, Green, Blue
ROC	Receiver Operating Characteristic
SGDM	Stochastic Gradient Descent with Momentum
SIFT	Scale Invariant Feature Transform

SVM	Support Vector Machine
SWT	Stationary Wavelet Transform
TN	True Negative
TNR	True Negative Rate
TP	True Positive
TPR	True Positive Rate
WCE	Wireless Capsule Endoscopy

CHAPTER 1: Introduction

1.1 Background: Gastroenterology

Gastric cancer is the fifth most common cancer worldwide and seventh most prevalent in accordance to the GLOBOCAN 2018 as shown in Figure 1.1. For males, 10% of the total cancers is Colorectal cancer and that makes it the third most prevalent cancer in males with cases of 663,000 all over the world. Whereas, it is the other most frequent cancer in females with cases of 570,000, which is 9.4% of the total cases globally [1,2].

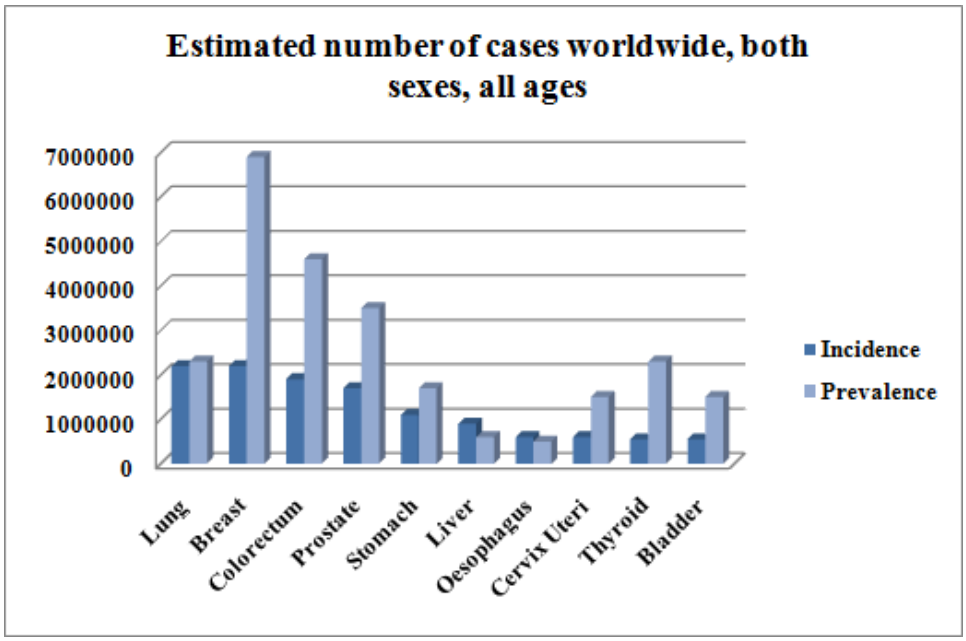


Figure 1.1: The estimated incidence and prevalence of worldwide cancer cases [1,2]

Persistent discomfort in the abdomen is a flag that something is wrong. Other nagging signs are gas, blood in stools, bloating, diarrhea, heartburn, prolonged inflammation in the digestive tract, etc. These are the reasons to go consult a gastroenterologist. Gastroenterology deals with the digestive system and its dysfunctions. The digestive tract or gastrointestinal (GI) tract is an organ system that begins from the mouth and continues up to the anus. GI tract comprises of several organs such as digestive canal, throat or esophagus, liver, duodenum, bile ducts, pancreas, gallbladder, small intestine, and large intestine, colon, rectum. Figure 1.2 shows the GI tract and its major organs.

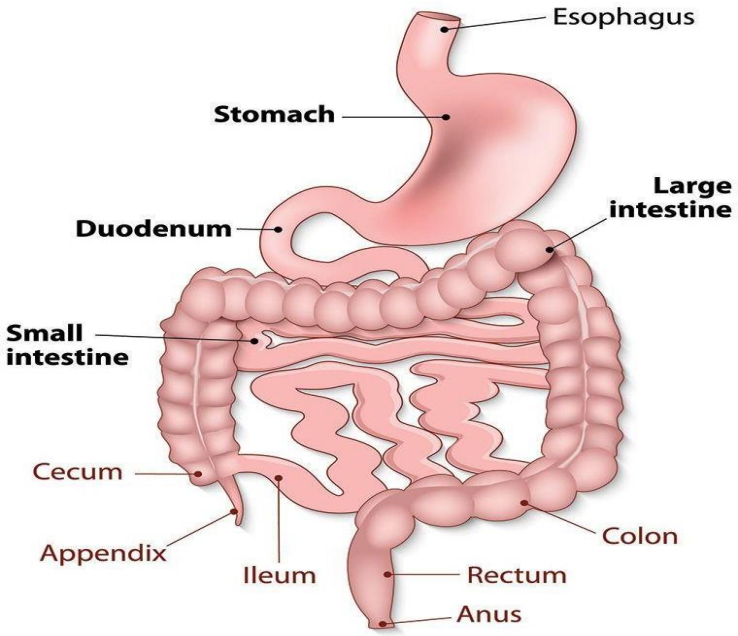


Figure 1.2: Major body parts in the gastrointestinal tract

Gastroenterology is the study of every organ's functionality in the digestive system while concentrating on multiple conditions that may strike them [3]. The study aids an absolute understanding of the digestion method, consumption of nutrient, and excreta. It also addresses the complications that may harm these organs including polyps, ulcers, cancer, and gastroesophageal reflux. Physicians practicing in this domain are addressed as gastroenterologists.

1.1.1 GI Tract and anomalies

Figure 1.3 shows the flow of food in the digestive tract. Food enters the oral cavity through our mouth and moves to pharynx after chewing. Pharynx enables swallowing of the chewed food to the esophagus from where it is passed to the stomach to get mix with the digestive juices to break down food. Trouble in swallowing the food indicates a problem in the esophagus. The esophagus, chest, and stomach are also affected when a person experience prolonged flaming and frequent acid reflux (GERD). This mixture of partially digested food and digestive juices is passed to the upper part of the small intestine, duodenum. In the duodenum part, the food gets mixed with enzymes from the pancreas, intestine and bile acids from the liver,

gallbladder for further breakdown. The walls of the jejunum absorb the products of digestion, nutrients from the digested food into the bloodstream. The nutrients that are not absorbed by the jejunum are absorbed by the ileum. The small intestine part of the digestive system can develop polyp and tumor cells in the walls which can lead to cancer. Inflammatory bowel disease is also an anomaly of the small intestine. The remaining products of the digested food are passed to the large intestine.In the large intestine, the cecum breaks down the cellulose present in the food products received from the ileum and the colon absorbs the water, fluids, and salts and passes the remaining waste to the rectum. Rectum stores the waste until the body is ready to excrete the waste through the anus.

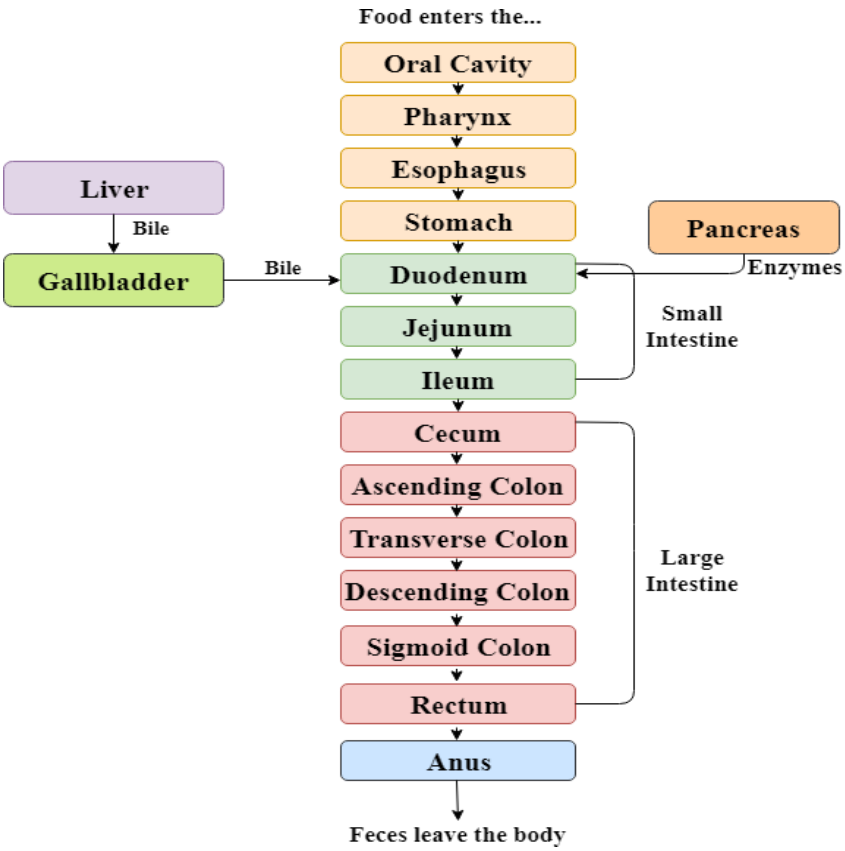


Figure 1.3: Details of human gastrointestinal (GI) tract. [4]

In few states of India such as Tamil Nadu, Assam, Kerala, and Karnataka, the malignant tumors or cancers are most commonly present in the squamous cells of the esophageal carcinoma section of the digestive tract [5]. Some of the abnormalities in the GI tract are:

- Constipation: A person suffers from constipation when their bowel movement is not frequent and causes discomfort in passing through.
- Diarrhea: Diarrhea is the disorder when a person takes more than two or three sloppy and watery bowels in a day.
- Prolonged abdomen pain and distress, upset stomach, vomiting, and nausea are also related to dysfunction of the digestive system.
- Bleeding in the digestive tract isdue to other anomalies like colon cancer, esophageal cancer, polyps, ulcers, GERD etc. Blood spots in stool, excrete waste or vomit is a manifestation of dysfunction in the digestive tract. It can be life-threatening depending upon the level of bleeding. The bleeding level can vary from moderate to critical. Advanced imaging technology, if necessitated, can usually determine the origin of the bleeding. Remedy depends on the root of the bleeding. In our thesis, we will be focusing on the detection of this anomaly. The proposed system is to detect the presence of bleeding in the digestive tract.
- Cancer (e.g., intestine cancer, liver cancer, gastric cancer, colorectal cancer, pancreatic cancer). Colon cancer customarily is formed of unnatural growths, polyps, tumors on the rectum or colon. Detecting polyp growths at the beginning allows doctors to help remove them before they hold a chance to grow cancerous. Immediate discovery and diagnosis of colon cancer are helpful.
- Trouble swallowing indicates a problem in the esophagus.
- Diverticular disorder and other conditions of the colon. For example, colitis, polyps, Crohn's disease, irritable bowel syndrome.
- Gallbladder infection: Persistent chest pain, vomiting or nausea are the symptoms of problem in gallbladder.
- Gastroesophageal reflux disease (GERD) and Heartburn occurs when a person experiences prolonged flaming in the chest or esophagus and frequent acid reflux.

- Heartburn is pain or flaming in the chest or esophagus. It is the pipe that connects the mouth to the abdomen. Heartburn happens when stomach acid backs up into the throat. Many people experience heartburn once in a while. It usually happens after meals or at night-time. Symptoms that happen more than double a week could be a warning of a more pressing problem: Gastroesophageal Reflux Disease (GERD), also called acid reflux.

- GERD transpires when acid from the belly burns the wall of the esophagus. GERD is more usual amidst people who are overweight or do smoking. Pregnant women also are prone to get the disease. Medication is available to operate GERD. Critical cases may require surgery. Untreated GERD can induce more severe problems, including recurrent inflammation in the esophagus and breathing difficulties.

- Hemorrhoids is caused due to the swelling in the veins in rectum.

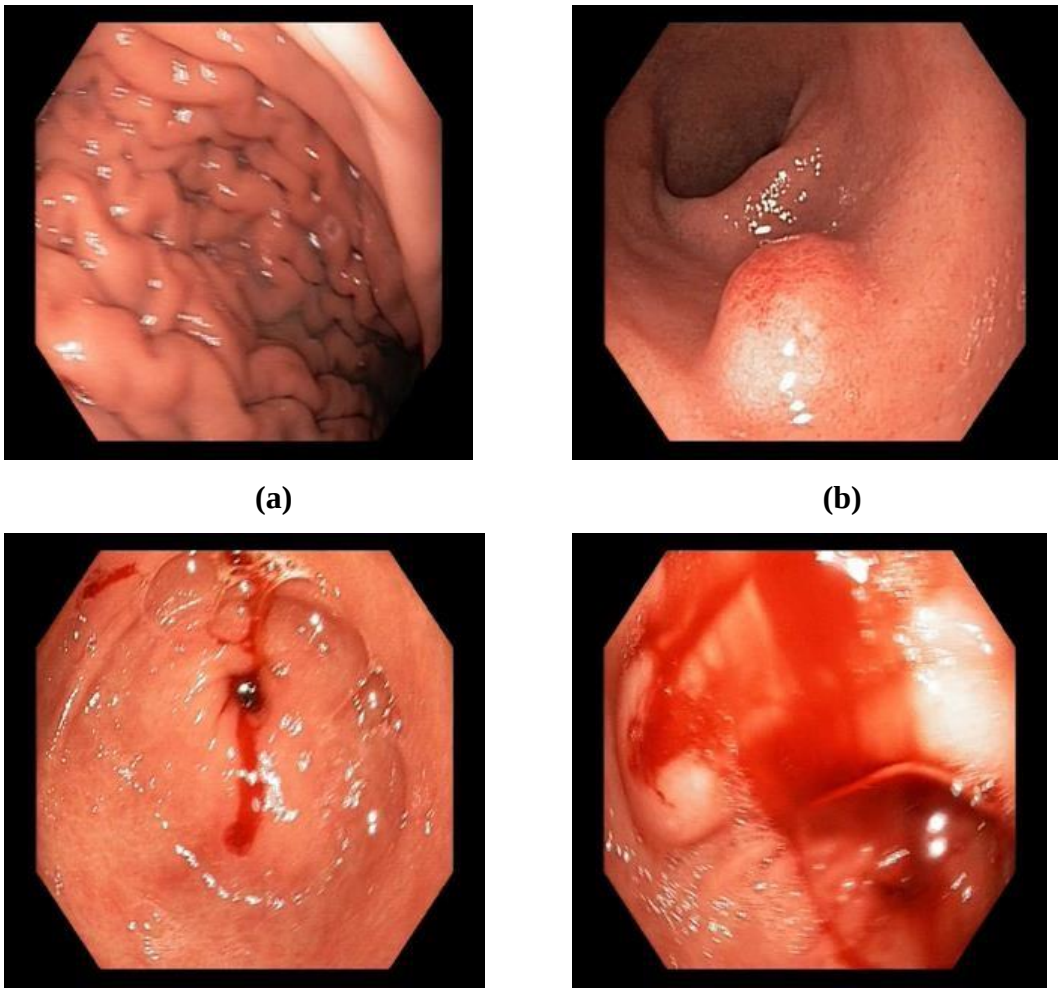
- Inflammatory bowel disease (e.g., gastritis) is the persistent inflammation in the gut causing intestines to get irritated and swollen. Seldom, the immune system of our body confuses food and different substances in the GI tract for threatening germs. It fights against muscular and healthy tissue by delusion. This results in inflammatory bowel disease (IBD). IBD encompasses various dysfunctions that trigger persistent inflammation in the gut. Due to this, the small and large intestines grow irritated and swell, provoking severe stomachache, diarrhea, and bleeding in the rectum, besides, symptoms that appear irrelevant, such as lethargy, joint agony, and fever.

- The most general IBDs are Crohn's disease and ulcerative colitis. Ulcerative colitis attacks the large intestine. Crohn's disease strikes wherever along the GI tract. Endoscopy, blood tests, X-rays, stool specimens, and CT scans assist physicians to diagnose IBD. That entails inspecting the lining of the intestines using a scope with a functional camera. Treatment may involve medicine, life-style alterations, measures to avoid stress, and surgery.

- Liver illness (e.g., hepatitis, jaundice) is inflammation in the liver.

- Malabsorption ailments (e.g., celiac disease, lactose intolerance) makes the immune system of a body to fight against the gluten or nutrients from the food.

Some of the abnormalities of the GI tract are shown in the figure 1.3 with the example of normal image of inside of GI tract without any disorder's symptoms. These images in figure 1.3 are extracted from the videos collected from PSRI hospital (New Delhi, India). The seven videos collected are from seven patients, each video being 5 to 7 minutes long with extension as Audio Video Interleave (avi).



(a) **(b)**

(c) **(d)**

Figure 1.4: Capsule endoscopy images collected from PSRI hospital, New Delhi [6]; (a) Image with no abnormality, (b) Polyp, (c) Bleeding in digestive tract, (d) Severe bleeding.

1.1.2 Bleeding in the GI tract

Many disorders can induce bleeding in the digestive system. A doctor can try to determine the issue of the bleeding by locating its source. Causes of bleeding in the GI tract can be:

- Angiodysplasias: severe bleeding from the small blood vessels found underneath the internal intestinal wall causing blood loss resulting in anemia.
- Colon cancer and polyps: bleeding from the tumor or polyps in the small intestines.
- Ulcerative colitis: bleeding from the ulcers in the rectum.
- Diverticular disease: small sacks or cysts on the colon wall.
- Esophageal cancer or chronic liver conditions.
- Acid reflux or GERD.
- Gastritis, Crohn's disease
- Hemorrhoids: Constipation and strain during bowel movements.
- Peptic ulcers can rip the mucosa resulting in bleeding.

1.1.3 Fields of gastroenterology

Gastroenterology incorporates several diagnostic and remedial procedures including endoscopy, endoscopic ultrasound, endoscopic retrograde cholangiopancreatography, advanced endoscopy and liver biopsy, inflammatory bowel disease, colonoscopy, transplant hepatology, sigmoidoscopy, motility.

- Endoscopy: Endoscopy is the study of the analysis of the upper and lower section of the GI tract.
- Endoscopic ultrasound: It is same as the endoscopy. Ultrasounds is done to explore the upper and lower GI tract, along with other internal organs.

- Endoscopic retrograde cholangiopancreatography: It is performed to recognize tumors, scar tissue in the bile duct region, and gallstones.
- Advanced endoscopy: Advanced or Capsule endoscopy (CE) is performed to check the small intestine using a capsule camera.
- Liver biopsy: Liver biopsy is carried out to estimate fibrosis and inflammation.
- Inflammatory bowel disease: endoscopy for Crohn's disease and ulcerative colitis.
- Colonoscopy: Colonoscopy is done to discover colon cancer or polyps.
- Hepatology: Hepatology deals with the study of the liver, biliary tree, and pancreas.
- Sigmoidoscopy: Sigmoidoscopy is done to assess blood loss or discomfort in the bowel.
- Proctology: Proctology incorporates the areas of anus and rectum complications.

1.2 Endoscopy

An endoscopy in plain terms means looking inside. It is practiced examining the insides of the body. The endoscopy process utilizes an endoscope to inspect the depths of an organ of the body. Unlike several other medical imaging methods, endoscopes are interpolated straight into the organ. There are various kinds of endoscopes. Depending on the section in the body and nature of the procedure, an endoscopy may be done by a surgeon or a doctor. Most often the endoscopy is used to indicate to an inspection of the upper part of the GI tract, known as an esophagogastroduodenoscopy. Endoscopy is practiced examining the symptoms in the digestive tract including queasiness, vomiting, abdominal distress, trouble swallowing, and gastrointestinal bleeding. It is further utilized in diagnosis, generally by conducting a biopsy to probe for conditions such as bleeding, anemia, irritation, and tumors of the digestive tract. The process may also be used for stretching a narrow esophagus, cutting off a polyp, eliminating a foreign article.

1.3 Wireless Capsule Endoscopy (WCE)

Wireless capsule endoscopy (WCE) is a non-invasive diagnostic endoscopy to envision irregularities in the gastrointestinal (GI) tract. It is also helpful in small bowel examination. This technology has eliminated the patient's distress which was a problem in conventional endoscopy. The personal swallows a plastic capsule which carries a digital camera to record video, signal transmitter, LED light source, and battery. WCE technique was first introduced in 2000 [7]. The capsule progresses into the GI tract by peristaltic contractions, takes the images. The captured images are broadcasted in real-time to an external console that is worn by the person. The images are copied and saved in the console memory and then are uploaded to a computer machine for visual examination or automated analysis shortly after the monitoring has been finished. Currently available capsules take two frames per second on average, which create thousands of pictures (approximately up to 50,000) and more than 8 hours of video per case [8]. This significant amount of data demands a substantial amount of time to examine these images or video for anomalies by pharmaceutical experts in category gastroenterology. As shown in figure 1.4, eight sensors are fixed adhesively on the stomach of the patient [9]. These sensors will transfer to a data-recorder belt that is to be worn by the patient around their waist.

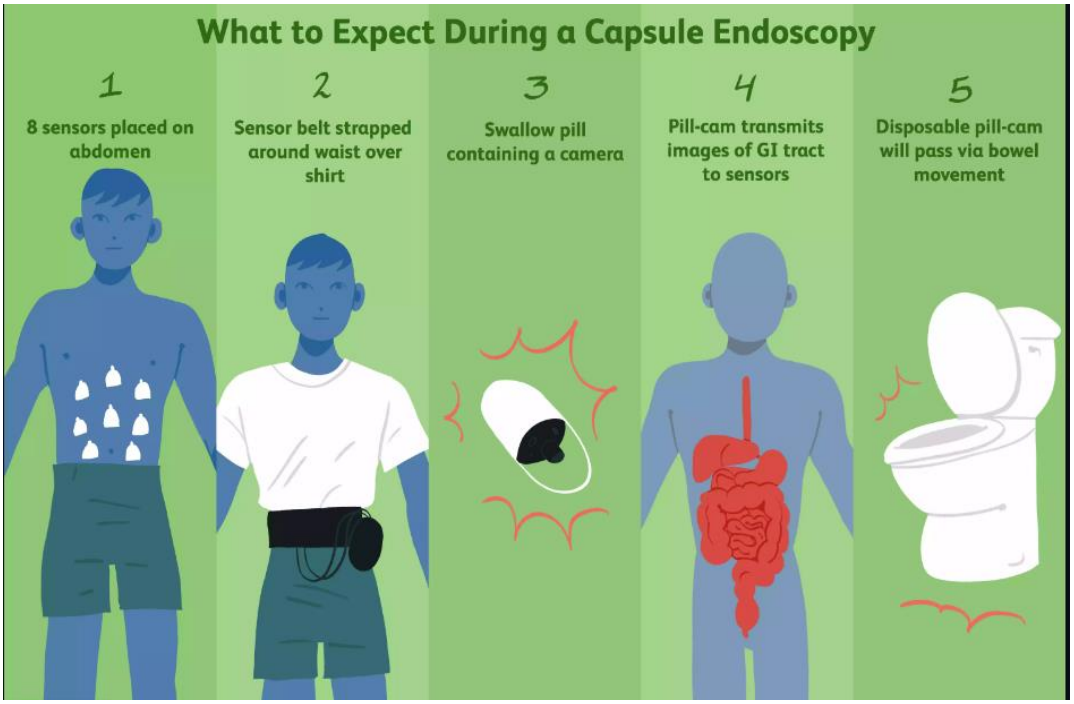


Figure 1.5: Wireless capsule endoscopy (WCE) procedure [9]

1.3.1 Why capsule endoscopy is preferred

Capsule endoscopy has won the approval of the doctors as well as the patients. The chief aspect of CE is the wireless manner, empowering it to be extremely well-tolerated. It is safe, painless, non-invasive, and doesn't demand the patient to be on anesthesia [10-12]. CE is preferable over the conventional methods such as simple endoscopy that used a wired endoscope, endoscopic ultrasounds, X-ray, etc. Capsule endoscopy is very convenient as children and elderly people can also swallow the pill without any difficulty. Furthermore, it enables the doctor to envision the intact stretch of the small intestine, not just the opening portion of it. The small intestine is one of the complex organs to diagnose and heal without conducting surgery. CE supports physicians to see inside the regions of our body that are not readily reached with traditional endoscopy. Conventional radiological approaches have been available for examination, but these methods are tedious and time-consuming. Simple endoscopy or colonoscopy sometimes fail to reveal the source of the bleed and may not be valid in recognizing small tumors and several other anomalies. Wireless capsule endoscopy presents more reliable imaging of the wall of the small intestine than any other methods. It also is very reliable in revealing the reasons for gastric pain, diarrhea, and intestinal blood loss resulting in anemia and diagnosing conditions of the small intestine, like, Crohn's disease. Figure 1.5 represents a typical pill camera that can be used in WCE procedure.

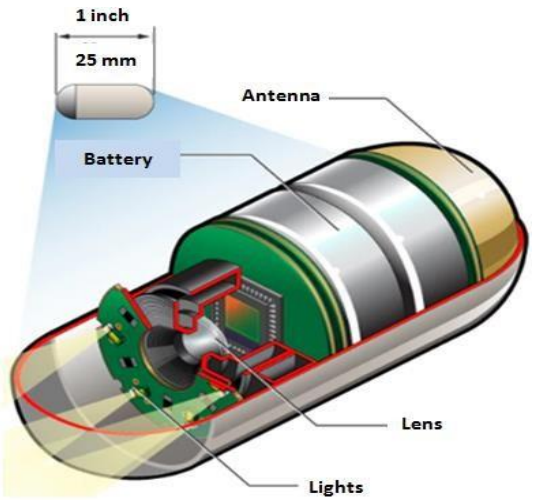


Figure 1.6: A typical capsule camera [13]

1.3.2 Risks of WCE

Wireless capsule endoscopy is usually a safe and reliable technique that involves very few risks. In the natural case, the capsule moves out of the system after the bowel movement. It does not need to be retrieved. Rarely, the capsule can be stuck in the digestive region. The risk is greater in people who are suffering from intestinal clogging or have Crohn's disease. The capsule can be recovered practicing balloon enteroscopy, which requires implanting a deep, slim tube adorned with a camera downed inside the esophagus or up into the rectum. CE can exclusively be practiced for visible diagnoses, and not for treatment. However, simple endoscopy, polypectomy, and colonoscopy can be done to eliminate polyps.

1.3.3 Results of WCE

The video obtained by the pill-sized camera used in WCE is carefully observed by the doctor for irregularities inside the digestive system. It may need a few days or even weeks to get the outcomes of the wireless capsule endoscopy. While some issues can be immediately discerned, like bleeding or contractions, some may be obscure like inactive bleeding, blood specks that hard to detect.

1.3.4 Types of disorders diagnosed with WCE

Wireless capsule endoscopy progresses to grow better as it has reformed diagnosis completely by providing a fine-tuned, comfortable, and non-invasive ways of monitoring the core of the small intestine. Some anomalies of the small intestine that are diagnosed by WCE include:

- Angiodysplasias: severe bleeding from the small blood vessels found underneath the internal intestinal wall causing blood loss resulting in anemia.
- Small intestinal polyps and tumors such as carcinoid tumor, and lymphoma, and colon cancer.
- Angioectasias: bleeding from tiny vascular tumors or angioectasias, lesions.
- Inflammatory bowel disease or Crohn's disease.

1.3.5 Accuracy

The exactness of WCE can be differed by the purpose of the examination and the equipment used. CE is deemed superior to diagnose the IBD in the small intestine or Crohn's disease, as it detects inflammatory tumors earlier as contrasted to all other methods. Furthermore, CE is more specific in identifying celiac disease, even though a biopsy is nevertheless required for an absolute diagnosis.

1.3.6 Capsule camera

There are several pill cameras available in the market such as PillCam SB3 (originally known as M2A) by Given Imaging Ltd., MiroCam by IntroMedic company, EndoCapsule by Olympus America, OMOM@Jishan by Science and Technology, CapsoCam Plus by CapsoVision, etc. [14]. M2A was the first capsule-sized camera ensemble to introduced into the capsule endoscopy field [7]. Table 1.1 shows the comparison between these pill-sized cameras available in the market for evaluating the performance, affordability and approvals. This table presents the difference in these pill cameras, how they differ in size, weight, durability, resolution.

Table 1.1: Comparison of pill cameras available in the market [15]

Capsule	**PillCam SB 3 Given Imaging**	**EndoCapsuleOlympus America**	**MicroCamIntromedicCompany**	**OMOM Jinshan Science and Technology**
Size	Length: 26.2 mm Diameter: 11.4 mm	Length: 26 mm Diameter: 11mm	Length: 24.5 mm Diameter: 10.8 mm	Length: 27.9 mm Diameter: 13 mm
Weight	3.00 g	3.50 g	3.25-4.70 g	6.00 g
Battery Life	8 h or longer	8 h or longer	11 h or longer	6-8 h or longer

Resolution	340 x 340	512 x 512	320 x 320	640 x 640

1.4 Machine learning in wireless capsule endoscopy

The manual review of WCE video is not time-efficient, with an average reading time of 45–120 minutes approximately [8,16]. Based on the encounters, it has been extremely advised to study the video in a single uninterrupted sitting. So, an automatic algorithm that can provide the frames with abnormalities from the thousands of frames from the WCE video to avoid the tedious and time-consuming work of manual examination is available with the PillCam, known as RAPID algorithm.

1.4.1 RAPID algorithm by PillCam

RAPID software is available with the PillCam kit to automatically detect the bleeding frames out of all the frames from video captured by the camera [17]. The efficiency obtained by this software is not satisfactory [18-20]. The miss-rate in detection of abnormal frames is unacceptably high [21]. And it may also skip the frames with inactive bleeding or frames with blood spots of very small sizes. As the results obtained by PillCam's RAPID software algorithm are not efficient, there is a need for a better algorithm to detect the anomalies in WCE frames. Since bleeding in the digestive tract is a very common condition, large number of papers has worked on this topic. The principal problem is the blood specks and spots not having any standard texture and frame, and the blood shade may fluctuate extensively from bright red to deep red to brownish, which makes the blood challenging to differentiate from other articles and the abdominal wall. This heterogeneity in color depends on the sorts of condition, time of the bleeding, capsule location.

In this research, we focus on bleeding detection in the wireless-capsule endoscopic videos. Our goal is to propose an automated technique for the detection of presumed frames that can have the appearance of bleeding. The method pre-processes the images or video, recognize and transfer speculated frames, and prepare them to be visually examined, whereas the other frames are ricocheted and not transmitted for the

examination. This may significantly lessen the evaluation time while the ultimate judgment is still left to the endoscopy experts.

1.4.2 Image segmentation

An image is a group or assemblage of several pixels. We assort the pixels with comparable traits by applying an image segmentation technique. Image segmentation is the technique in which an image is partitioned into many segments to determine shapes, objects, and edges in the image. The purpose of segmentation is to analyze or modify the description of an image into a more significant representation that can focus on the important features of the image. In the image segmentation process, each pixel in the image is labeled so that pixels with identical labels possess certain features such as contrast, color, homogeneity, intensity, energy, and texture, etc. The success of classification depends on the image segmentation technique used and its effectiveness.Some of the image segmentation techniques are thresholding, color histogram-based segmentation, edge detection, k-means clustering, etc.

- Thresholding: Thresholding based segmentation is based on a threshold value that is applied to every pixel in the image to convert RGB or a gray-scale image into a binary image. The pixels are segregated based on their intensity or energy value. In case when solely one threshold value is chosen for the whole image, it is called global thresholding. If the threshold is decided by some local characteristics of the pixels of any image region, it is called local thresholding. Thresholding regards only the intensity values of the pixels and ignores other features like contrast, homogeneity, correlation. It does not take in to account the relationships between the pixels.
- Clustering: The K-means algorithm is an unsupervised technique that is applied to split an image into *K* clusters. In this technique, K cluster centers are picked in a random manner or by a heuristic approach. Each pixel of the image is allotted to a cluster with a minimum distance between the pixel and the center of the cluster. The centers of the modified clusters are re-calculated by taking the mean of all the pixels in the corresponding cluster. The clusters are re-assigned based on the re-evaluated centers. These steps are repeated until the convergence is achieved. The distance between a pixel and a cluster center can be found by computing the absolute difference between them. This

difference can be the difference between the intensity, texture, color, homogeneity, and location of the pixel.

- Histogram-based segmentation: Histogram-based methods are very effective as opposed to other techniques for image segmentation as they traditionally demand just one traverse through the pixels. The histogram for the image is plotted using the information (features like color, intensity) of all the pixels of the image. The significant highs and lows in the histogram represent the clusters in the image.
- Edge detection segmentation: Edge detection in itself is a field of image processing. Edges in an image and region boundaries in the image are firmly linked as there is usually a sharp change in energy at the region boundaries. Due to these sudden changes in the intensity, the edge detection method is used as the foundation of another segmentation procedure. This technique usually recognizes the disconnected edges. So, closed region boundaries are required for the segmentation of an object out of an image. The points of discontinuity are organized in curved line edges by segmentation.

Since we are working with medical images, the high variability in the images is a challenge for segmenting the images. Also, medical images comprise of valuable information that is to be preserved during segmentation. The important features are extracted from the medical images using an effective feature extraction technique.

1.4.3 Feature Extraction

Feature extraction is a vital step for reducing the dimension of data and extracting meaningful characteristics from it while effectively rendering the significant features of the data or images with adequate memory. Some common feature extraction methods are Histogram of oriented gradients (HOG), Gray Level Co-Occurrence Matrix (GLCM), Scale Invariant Feature Transform (SIFT), Color histograms.

- HOG: Histogram of Oriented Gradients is a technique for describing features for the purpose of detection of objects in images, which supports the fields of computer vision and image processing. An abrupt shift in color is known as a gradient. The change from a darkened tint to a lighter shade is a positive gradient and vice-versa. HOG counts these incidents of gradients in the region

of interest in a grayscale image and makes a set of corresponding pixels with gradients to derive the features.

- GLCM: Gray Level Co-Occurrence Matrix (GLCM) or Gray Level Spatial Dependence Matrix is a statistical way to represent the texture related attributes of the images. The statistics are derived after GLCM is constructed from a gray-scale image. The texture features are derived by estimating the incidence of specific pairs of spatially related pixels and GLCM extracts meaningful data out of them in terms of textural traits.

- SIFT: Scale Invariant Feature Transform is a way to detect and describe prominent, stable, and local characteristics in the images using the detectors and descriptors. These characteristics are invariant to minor adjustments in the image, scale of the image, illumination difference, image rotation, etc. A set of such local features illustrates the ideal region of interest in the image.

- Color Histogram: Color histogram corresponding to an image represents the color composition of that image to provide an idea for the distribution of the colors appearing in the image. The histogram can be used to extract the color features of the image into a vector.

When dealing with complex and large datasets like medical data, the extraction of crucial features from the numerous and complicated features of the data is challenging as analysis with a large set of features usually needs a bulk of memory and high computational power. Also, it may result in the overfitting of the classification model. Deep learning models like artificial neural network, convolutional neural network, principal component analysis (PCA) provide better feature extraction on a large set of images with the imbalanced class distribution. A Convolutional neural network offers automatic feature extraction that reduces human intervention in the process of medical image classification.

1.4.4 Deep learning

Deep learning is a subdivision of machine learning involving neural networks that work similarly to the human brain and are capable of learning from unstructured data like images. Deep learning is preferred over traditional machine learning while working on a complex dataset like image dataset, or simply a large-sized dataset [22].

The main benefit of deep learning is that the low-level features of an image (like edges or textures) are compared and connected to the higher-level features (like shapes, objects) automatically and autonomously by the model. In traditional machine learning algorithms like SVM, Linear regression, Random Forest, the features have to be extracted from the images manually. Whereas, in deep learning like CNN, features are learned from the raw data automatically by the network [22,23].

1.4.5 Convolutional neural network

A CNN is a feed-forward NN that is used to identifies the complicated features in the dataset. It can devise and derive features from unprocessed data automatically and can work on massive amounts of data. CNN performs remarkably good in the fields of images-analysis, pattern-detection, edge-detection, image/object recognition, powering vision in robots, for self-driving vehicles, etc. It also reduces computational burden by offering the automatic feature extraction. Figure 1.7 shows a CNN architecture.

1. **Structure of a convolutional neural network**

Layers in CNN:

- Input layer
- Output layer
- Multiple hidden layers:
    - Convolutional (Conv) layers
    - Batch normalization (BN) layers
    - ReLU layers
    - Pooling layers
    - Fully connected (FC) layers

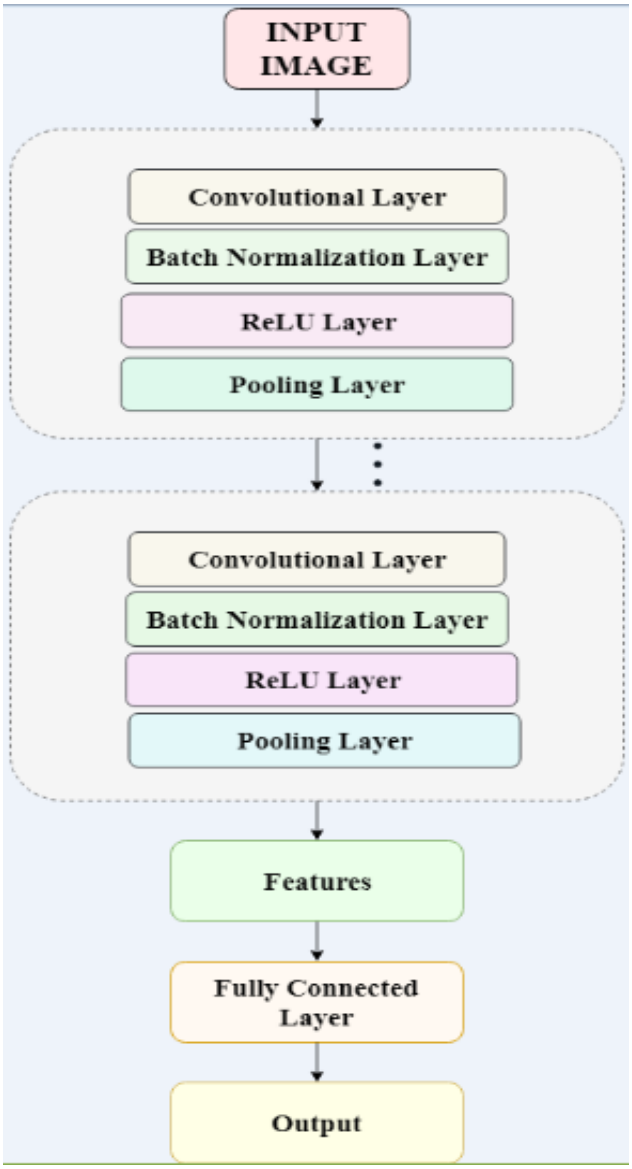


Figure 1.7: A CNN structure [24]

2. **Operation in CNN layers**

Operations in CNN layers are convolution at the convolution layer, batch normalization at the layer with same name, making the images non-liner at the ReLU layer, applying a pool function and finally getting the classified labels.

- **Convolutional layer or Conv layer**

The convolutional layer is the foundation of this deep learning network. CNN is fabricated of convolutional layers which themselves are made up of feature detectors (called "Filter"). As the title of the layer implies, a convolutional operation is implemented by Conv layer. This layer receives a primary image as input, transform it

by implementing a sliding filter and transfer it to the subsequent layer. The very first convolutional layer is accountable for recognizing low-level features. The Conv operation recognizes the high-level features of the image. The deeper the CNN goes, the better the features grow. The *stopping criteria* for deepening the network is the plot of training and testing errors. As the network is deepened, the training loss and the testing loss are plotted. As soon as the training loss stops decreasing or testing loss starts increasing, the addition of the hidden layers or neurons is stopped [25]. A filter or Convolutional kernel is a matrix with values, called weights. These weights provide a way to discover features. The kernel is glided over each portion of the image to check for the existence of the feature to be identified. The filters detect patterns like edges, objects, corners, shapes, etc.

A deeper Conv layer enables the filter to detect the patterns and specific objects (like ears, eyes, etc.) more explicitly. The number of filters is to be specified while adding a Conv layer. The proposed model of CNN receives the de-noised image as input. A filter of a size 3×3 is used that will launch and skid itself across every sub-region of the received image until it has slid over all the blocks of pixels from the intact image. This process of sliding the filter is referred to as Convolving. so, the filter is convolved across the complete image.

Convolutional operation is accomplished by calculating the dot product of the kernel and the sub-image. This action yields a real number which verifies the existence of the feature to be detected, based on the training of the network. This number gives the resultant matrix that is received from gliding the filter over the entire input image, saves the convolutions of this filter over sub-parts of the initial image and is named as the feature map (or activation map). This feature map is passed as input for the following layer. The image information loss is restricted by producing several activation maps. Each map identifies the position of specific features in the image.

- **Batch normalization (BN) layer**

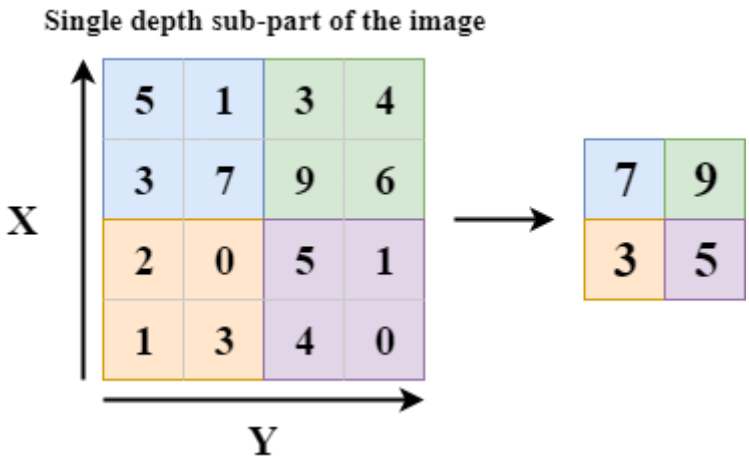
Each input image-channel across a mini-batch is normalized at the batch normalization layer. BN layers are applied between Conv layers and nonlinearities (ReLU layers) to quicken up the training of CNN and decrease the sensitivity to CNN initialization.

- **ReLU layer**

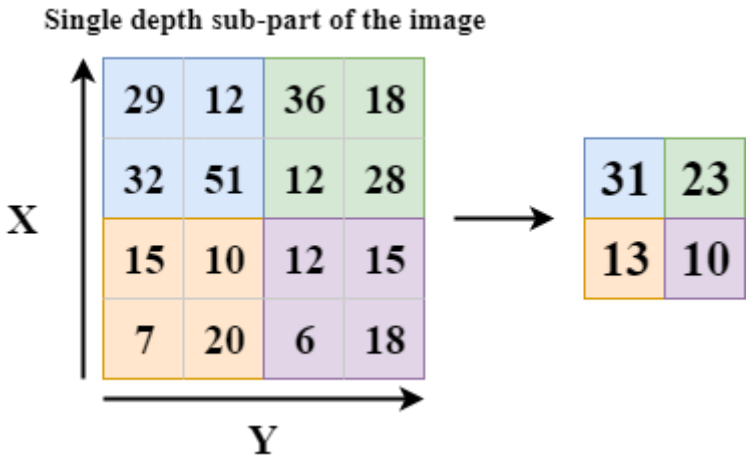
Images are composed of various objects and articles that may not be always linear to each other. So, this makes the image classification a non-linear problem. In this layer, a rectifier method is implemented to enhance non-linearity in CNN. Without using this method, the image classification will be handled as a linear problem.

- **Pooling Layer**

Pooling empowers CNN to discover features in several images regardless of the diversity in illumination in the images and complex aspects of the pictures. The foremost purpose of pooling layer is to lessen the number of parameters and computations in the network. And it is accomplished by decreasing the spatial size of the presentation gradually. Hence, it also manages the overfitting in the network. In a typical ConvNet, pooling layer is linking the progressive Conv layers. Pooling layer runs on every feature map separately and resizes the spatial size by employing a pool function. Pool function merges the outputs of neuron batches at a layer into an individual neuron at the following layer. For instance, max pooling accepts the maximum value from each of the clusters of neurons in the preceding layer. Max pooling strives to conserve the essential features while also decreasing the size of the picture. Another method is average pooling, which applies the mean value from each of the clusters of neurons in the previous layer. Figure 1.8 shows the examples of max pooling and average pooling.



(a) Max Pooling



(b) Average Pooling

Figure 1.8: Pooling operation example: (a) Example of Max pool, (b) Example of Average pool [24]

- **Flattening**

After getting the pooled featured map, flattening of this feature map is performed. Flattening requires remodeling the whole map-matrix into an individual column which is then served to the CNN for processing.

- **Fully Connected (FC) Layer**

The fully connected layer is just like the hidden layer in Artificial Neural Networks. The only difference is that it's fully connected as the name suggests. FC layer links every neuron or feature map in the prior layer to every neuron in the subsequent layer. Conv and pooling layers deliver high-level feature maps of the input frame as output. The chief objective of the FC layer is to divide the input image into several classes based on the training dataset, utilizing the feature maps. After classification, the error in prediction is computed. The error is then back propagated into the network to enhance the prediction.

- **Gradient Descent**

Gradient descent accepts the arbitrary numbers for the filters preferably to +1 or -1. It assesses the error and attempts to minimize the deviations. So, the error function is repeatedly used to find the derivative of gradient and utilizes that gradient to attain a

fresh model with reduced error function until it reaches to the limit where error cannot be decreased any further.

1.5 Thesis Organization

This thesis is structured into six chapters that comprise of an introduction, literature review, problem statement, methodology, result analysis, conclusion followed by references.

Chapter 1 includes the problem area on which the participation is done by the research work and lists the goals of the thesis. In this chapter, we explained the background of our research topic. In section 1.1, the background of gastroenterology, GI tract and its anomalies, fields of gastroenterology are discussed. Section 1.2 explains the endoscopy in brief followed by the section 1.3 to explain the procedure of wireless capsule endoscopy, its need, risks, results, and accuracy of the procedure. Section 1.4 deals with the use of machine learning in the medical imaging, importance of deep learning, image segmentation and its techniques.

Chapter 2 gives insights into the literature review of state-of-the-art methods that have been before supported in the study of bleeding detection in WCE images. The previous work done in this field is explored and a comparison analysis is carried out. The review of the related literature is done on the basis of detection of bleeding anomaly using application of CNN, using other machine learning techniques, and another anomalies detection. The comparison of the related literature is done in the section 2.2.

Chapter 3 states the problem statement and research gaps in recent related literature. This chapter explains the problem still existing and need for further research. The section 3.1 discuss the gaps in the related literature explored in the previous chapter. Section 3.2 states the research objectives of the thesis.

Chapter 4 presents the steps involved in the proposed methodology in fulfilling the desired objectives. In this chapter, we have explained the proposed system for our problem statement in section 4.1. Section 4.2 describes the dataset collected, its source and preparation, pre-processing being done on the data, followed by section 4.3 explaining the application of the proposed system using CNN, parameters tuning

to improve efficiency of the model, performance evaluation metrics. The comparison of the proposed model with other ML techniques is done in the section 4.6.

Chapter 5 features the analysis of the results performed using the proposed methodology. The results of the proposed system is presented in the section 5.1. The performance metrics derived from the confusion matrix of the network are presented. The performance of the model is compared with other ML techniques by comparing the results of all the state-of-the-art-methods.

Chapter 6 sums the conclusion of the study and gives future scope. Section 6.1 draws the conclusions from the experiments results from the previous chapter. Section 6.2 discusses the limitation of the proposed system. Section 6.3 discusses the future scope of the proposed model. The work that can be done on the proposed model or in the field of the application is mentioned.

CHAPTER 2: Literature Review

2.1 Summary on explored related literature

This chapter studies the work done in the field of Capsule Endoscopy to detect plausible abnormalities like polyp, tumor, ulcer, bleeding, etc. The State-of-the-art research done in existing related literature is the motivation for our research work. Related work done by various authors is summarized to analyze different approaches used and compared to analyze the efficiency of the related literature.

2.1.1 Review of literature based on bleeding detection using the Convolutional Neural Network (CNN)

Authors proposed a way in [26] for detection of bleeding regions in CE images employing a semantic segmentation based on deep neural network, called SegNet. CNN is trained using the SegNet layers. Bleeding regions are detected in CE images by segmenting the test images on the trained CNN. The efficiency achieved is 94.42%. In [27], Authors advised a simplified neural network (NN) for the detection of bleeding frames by performing automatic bleeding regions segmentation on the CE dataset. Fitting color channels are chosen as inputs to the NN. An multi-layer perceptrons and CNN are applied to conduct image classification individually. They decreased the number of computational operations. The performance of the recommended systems is assessed using the DICE score. The area under the receiver operating curve (AUC-ROC) is 0.97. Due to significantly fewer computations, CNN is proved to be more beneficial than multi-layer perceptron. Authors in the paper [28] presented a high-level detection system for the bleeding frames in the WCE image dataset. This system is implemented using a deep CNN that detects both active and inactive frames. Authors have designed CNN to have 8 layers. The network composes 3 convolutional layers, 3 pooling layers, and two fully connected layers. The ReLU or the rectifier function is implemented at the convolutional layers and the first FC layer to increase non-linearity in the network. Images are made of different objects that are not linear to each other. Without applying this activation function, the image classification will be treated as a linear problem, while it is in actual a non-linear one. Pooling layers implement Max-Pooling to preserve the main features while also reducing the size of the image. This helps reduce overfitting. One of the main causes

for the overfitting to occur is the too much information fed into CNN. Especially if that information is not relevant in classifying the image. SVM is more beneficial in the case where the user wants to depreciate the entropy loss for prediction. So, CNN is devised by substituting the SoftMax regression function at the secondary FC layer with an SVM classifier. But this proposed system resulted in complex computations and required a large dataset of designated images for training the CNN. So, the authors presented a way in the paper [29] that implements the deep-learning technique of CNN with handcrafted features that are obtained using a k-means clustering technique. This model is focused on detection of frames with active and inactive bleeding. This approach reduces the computational cost incurred in the training of CNN.

2.1.2 Review of literature based on bleeding detection other ML techniques

In [30], authors proposed a method for automatic feature extraction and detection of bleeding in endoscopy images. The endoscopy images are segmented using block-based segmentation. The local features are discovered using color characteristics. Different algorithms are investigated in this approach to classify the bleeding and non-bleeding images. The performance parameters are also calculated to test the effectiveness of these algorithms. The accuracy obtained by the proposed approach is 95%. In [31], this approach presented bleeding detection in WCE images based on region-based feature extraction. This method extracts features from HSI and CIE color spaces. Authors used a uniform library binary pattern to label each region. The secondary set of features can be extracted from the grayscale image. Classification of regions is done by support vector machine (SVM) into three categories, namely, non-bleeding region, bleeding region, and background. GrowCut algorithm is used for the concluding segmentation of CE images. This paper in [32] introduced a three-step algorithm for automated detection of bleeding in endoscopy images. Authors have done Key-frame extraction and edge removal as the first step of preprocessing. In the second step, they separated the bleeding images from the dataset of all the frames using the KNN classifier, applying the concept of principle color spectrum by utilizing the superpixel color histogram feature. In the last step, the segmentation of bleeding regions from the various color spaces is executed by securing a 9-D color feature vector at the superpixel feature. The accuracy attained is 0.9922. This proposed system presented in [33], focused on identifying the multiple blood specks

in the several frames captured from the WCE video. To overcome the performance degradation due to the small size of the region of interest, the authors suggested an unsupervised ML technique. It breaks the principal classification query into many confined classification queries. This technique cluster the training set using fuzzy C-means, then implement improvised KNN on the determined centers rather than the entire training dataset. Authors have also analyzed the performance and results of the proposed algorithm as opposed to the typical KNN and SVM.

In the paper [34], Naive Bayes classifier and superpixel segmentation are used for automated obscure bleeding disclosure in WCE video dataset. They used the color histogram for region discovery and feature extraction. In the final step, they adopted an improvised semi naive bayesian classifier. In [35], authors have classified bleeding and non-bleeding frames in wireless CE images utilizing color histogram of block statistics. Local feature extraction is done using a WCE image block, preferably of an individual pixel. For the contrasting color panels of RGB color space, index values are defined. So, the authors used index values to extract the color histogram. Color histogram is useful in securing distinct color texture characteristics. Feature reduction using color histogram and principal component analysis is adopted to decrease the dimension of these local features. Extracted local features that do not result in any computational strain provides the blocks with bleeding regions. Authors used a public dataset of 2350 images that renders 97.85% accuracy.

In [36], authors have worked on the system that distinguishes the bleeding images and regions from the WCE dataset. WCE images are preprocessed to convert into a color space defined by green to red pixel-ratio. These transformed images are used to obtain various analytical features from the overlying spatial blocks. These various blocks are then clustered into two clusters using K-means based clustering (unsupervised). These clusters are used to obtain cluster-based features. These features combined into a global feature is used along with differential cluster-based features to detect blood zone frames using an SVM (supervised learning classifier). The proposed system achieved 97.05% precision.

2.1.3 Review of literature based on detection of other anomalies of the digestive tract

Authors in [37] have worked on a computerized way of finding abnormalities in the endoscopy images. These abnormalities can be erosion, erythema, ulcerations, polyp, bleeding, etc. This proposed method examines images for varied textures so that, it can differentiate abnormal images from the normal ones effectively. The distribution of different textures in an endoscopy image can be captured using a textons histogram. It is done by applying a FB (filter bank) and LBP (local binary patterns). This proposed approach gives 92% recall and 91.8% specificity on WCE images.

In [38], authors have worked on a computer-aided system that uses various approaches to detect the abnormalities in the images obtained from CE. Authors have employed CNN, region recommendation, transfer learning. In the first step, they have used a CascadeProposal to recommend high-recall regions and abnormal frames. In the second step, the authors used a multi-regional combination technique to detect the regions of interest and have also operated a salient region segmentation approach to catch certain region spots. For object boundary filtration, a dense-region fusion algorithm is applied. And lastly, to increase the efficiency of the proposed model, transfer learning tactics are exercised in CNN. Authors of [39] presented a computer-aided look-behind fully CNN (LB-FCN) algorithm to automatically catch the anomalies in CE images. It uses blocks of parallel convolutional layers with varied filter dimensions to derive the multi-scale features from WCE images. All the LB linked features are combined with the features deduced from prior layers. As LB-FCN has fewer free parameter as compared to conventional CNN, it makes it much easier to train the network on smaller datasets. The AUC performance of LB-FCN achieved is 93.5%.

In [40], they have aimed at reducing the analysis time of the WCE video frames by presenting a computer-aided approach that automatically identifies the abnormal frames. They have worked on an ensemble of two SVMs that are based on HSV and RGB color spectrums. Feature selection and parameter tuning are done by using a nested cross-validation approach. For the betterment of performance, exhaustive analysis is carried out to decide the best feature sets. The dataset used comprises of

8872 WCE frames. This fusion system renders an accuracy of 95%, specificity of 95.3% and sensitivity of 94%. A CNN is suggested in [41] for the GI angioectasia detection during small bowel in CE images. Local features are extracted through deep feature extraction using an approach of segmentation of images based on semantics. Authors created a semantic segmentation-based CNN for classification of GI angioectasias. The sensitivity and specificity achieved is 100% and 96% respectively.

The work done in [42] aims for an automated way for the detection of angioectasias in WCE image dataset. This approach depends on the automatic separation of a region of importance. That region is chosen by applying a module for the task of image segmentation based on the approach of Maximum a Posteriori where a new hastened variant of the Expectation-Maximization is also advised. This proposed method attained sensitivity and specificity values of 96% and 94.08% respectively with 95.58% accuracy in a database comprising 800 WCE frames designated by two gastroenterologists.

In this paper [43], they have conducted ulcer and lesion detection and classification in WCE dataset employing two pre-trained CNN, GoogleNet and AlexNet. These two networks perform object classification to obtain ulcer and non-ulcer frames. Due to a huge number of layers in GoogleNet, AlexNet resulted in double the efficiency of GoogleNet for training. The efficiency of both networks is enhanced by tuning the parameters. It is also found in this study that higher the learning rate of the network, higher is the resulting accuracy. The learning rate of 0.0001 renders adequate results for both the networks. AlexNet attained 100% accuracy with the rate of 0.001.

Authors of [44] trained a deep CNN to distinguish ulcers and erosions in small bowel CE images automatically. This CNN is based on an single shot multiBox detector that holds 16 layers. The CNN is trained using SSD on the 5360 images. For the testing phase, 10,440 WCE images are fed to CNN, out of which, 440 are of erosions or ulcers. This system renders an accuracy of 91.5%. Whereas, in the paper [45], they have focused on decreasing the misidentification rate for a polyp in CE images. This will support the professionals in finding the most significant regions to pay consideration. Features are deduced using color wavelet and CNN. These extracted

features are then fed to a train an SVM. SVM will classify the CE frames into the polyp region and normal frames classes. They achieved 98.34% accuracy.

The study performed by [46] has focused on detecting a hookworm abnormality in wireless capsule endoscopy images. They have adopted the deep learning algorithm to recognize the tube-like pattern of hookworm. For the better activity of the classification, two neural networks are employed, edge extraction CNN and hookworm classification CNN. Both the CNNs are seamlessly integrated into the recommended system to evade edge feature caching. The edge extraction CNN provides the tubular regions and the hookworm CNN gives the feature maps. Both the results are integrated into the pooling layers to produce an intensified feature map accentuating tubular regions. They achieved 88.5% accuracy.

2.2 Comparative Analysis

The Table 2.1 depicts the summary of the recent technique explored for the related literature with comparison of the performance of the proposed systems in respective papers.

Table 2.1: Comparative analysis of related literature

S. No.	Authors	Year of Publication	Abnormality Detection in CE dataset	Methodology used	Performance
1	[30] Nataliia Obukhova et al.	2019	Bleeding framesinKVASIR (open), 8000 images	Block-based segmentation and color characteristics	95% accuracy
2	[38] Libin Lan et al.	2019	Abnormal frames in Chongqing Jinshan (open access), 7381 images	CNN, region proposal, transfer learning	72.3% mean Average Precision
3	[31] Eva tuba et al.	2019	Bleeding framesinF. Dee	Texture and color features	0.85 Dice similarity

			ba, “Bleeding images and corresponding ground truth of CE images”, 50 images	(HSI, CIE ULBP), Grow Cut	coefficient, 0.092 misclassification error
4	[27] M. Hajabdollahi et al.	2019	Bleeding regionsinF. Deeba, “Bleeding images and corresponding ground truth of CEimages”	Multi-layer perceptron, CNN	AUC-ROC 0.97, DICE score For CNN 0.869, For multi-layer perceptron 0.831
5	[42] Pedro M. Vieira et al.	2019	Small bowel angioectasiasin KID (public), 27 images, and PillCam, MiroCam in Hospital of Braga (Portugal), 300 frames	Maximum a Posteriori, Expectation-Maximization	Sensitivity 96%, Specificity 94.08%, Accuracy 95.58%
6	[43] Haya Alaskar et al.	2019	Ulcer in Dr. Khoroo’sMedical Clinic (Online available), 1875 images	AlexNet and GoogleNet CNN	100% accuracy with learning rate 0.0001
7	[44]	2019	Erosions and	Deep CNN	91.5%

	Tomonori Aoki et al.		Ulcer frames in The University of Tokyo Hospital, Japan, 15800 images	with a Single Shot Multibox Detector	accuracy
8	[37] Ruwan Nawarathna et al.	2019	Mucosal abnormality in MiroCam WCE images	Filter bank, local binary patterns, Textons histogram	Recall 92%, Specificity 91.8%
9	[32] Xiao Jia et al.	2018	Bleeding frames in 1000 WCE images from random patients.	Superpixel-color histogram, KNN	0.9922 accuracy
10	[26] Tonmoy Ghosh et al.	2018	Bleeding zones in "Kid: Koulaouzidis-iakovidis database for capsule Endoscopy", 335 images	Semantic segmentation, SegNet, CNN	94.42% accuracy
11	[33] Ouiem Bchir et al.	2018	Multiple bleeding frames in Imaging PillCam, 1275 frames	Fuzzy C-means clustering, KNN	90.92% accuracy
12	[41] Romain Leenhardt et al.	2018	GI Angiectasia during small bowel in 6360 frames from	CNN-based semantic segmentation	Sensitivity 100%, Specificity 96%

			pre-med students		
13	[39] Dimitrios E. Diamantis et al.	2018	GI Abnormalities in Endovis challenge, 10,000 images, and KID, 2352 images	Look-Behind Fully CNN (LB-FCN)	AUC 93.5%
14	[35] Tonmoy Ghosh et al.	2018	Bleeding frames in CE (Online), 2350 images	Color Histogram of Block Statistics	97.85% accuracy
15	[46] Jun-Yan He et al.	2018	Hookworm frames in West China Hospital, 440K images	CNNs (Edge extraction and Hookworm classification network)	88.5% accuracy
16	[34] P. Sivakumar et al.	2017	Bleeding frames	Superpixel segmentation, Semi-Naïve Bayesian classifier	N/A
17	[45] MustainBillah et al.	2017	Polyp frames in Endoscopic Vision Challenge, more than 14,000 images	Color wavelet, CNN and SVM	Accuracy 98.34%, Sensitivity 98.67%, Specificity 98.23%
18	[40] Farah Deeba et al.	2017	Bleeding frames in PillCam SB1	SVM ensemble and exhaustive feature	Accuracy 95%, Specificity

			and PillCam S B2, 8872 images	selection	95.3%, Sensitivity 94%
19	[36] Tonmoy Ghosh et al.	2017	Bleeding frames in The capsule endoscopy website (public), 2350 frames from 32 WCE videos	Cluster based statistical featu re extraction	97.05% precision
20	[29] Xiao Jia et al.	2017	Bleeding frames in 1500 WCE images	Handcrafted features based CNN	F1 score 0.9285
21	[28] Xiao Jia et al.	2016	Bleeding frames in 10,000 WCE images	Deep CNN with SVM	Recall 99.2%, F1 Score 99.5%

Most of the investigated modern techniques based on this comparison have worked on the automated detection of abnormalities seen in the GI Tract Endoscopy. Our research concentrates on the automatic bleeding detection in capsule endoscopy videos using a convolutional neural network. Literature review organization in terms of techniques has been shown in Figure 2.1 while highlighting the CNN based approaches.

We are using CNN in our research for the detection of the bleeding frames in WCE images as it offers a wide range of benefits over other ML techniques. CNN is fast, efficient, and it needs limited preprocessing of the images [39, 43, 44]. CNN derives the features from images automatically, which results in a reduced computational burden. Features are learned during the training of the network on the image dataset. One more benefit of using CNN is that only the number of filters and the filter size is required to be defined, whereas the values of the filters are determined by CNN

automatically during the training phase. Unlike most of the other ML techniques, object detection in images is carried out by CNN regardless of the location of the object to be recognized. Pooling feature of CNN also prevents overfitting of the network.

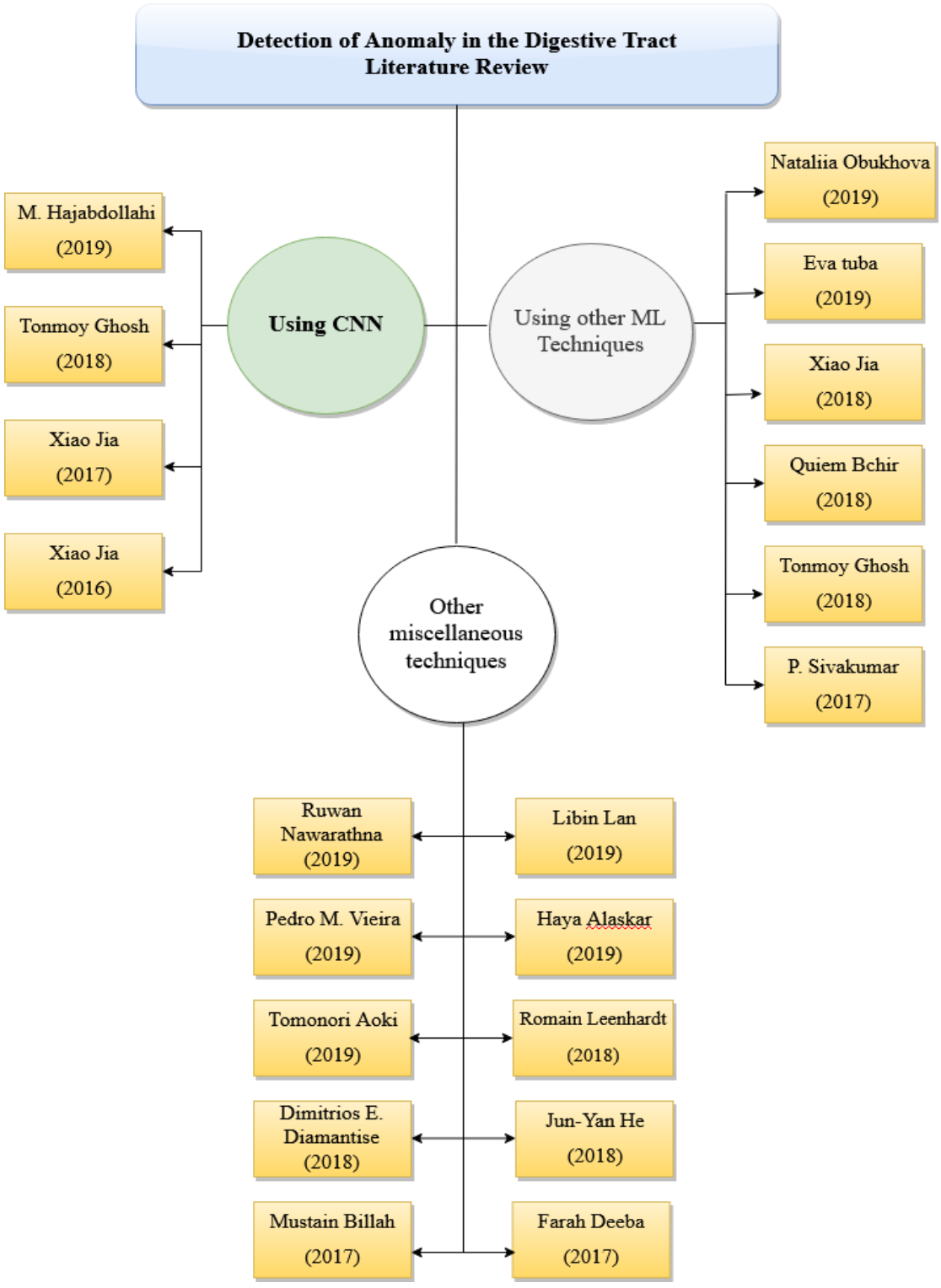


Figure 2.1: Literature review organization

CHAPTER 3: Problem Statement

3.1 Problem definition

Many researchers have focused on the problem of detecting the presence of bleeding in the digestive tract by extracting the bleeding frames from the WCE images. Manual review of the large dataset of the wireless capsule endoscopy video is time-consuming and can be avoided by using a computer-aided program for the detection. Many state-of-the-art papers have worked on automatic detection of bleeding in WCE frames using various techniques such as image processing, machine learning, etc. In our research, we are focusing on to present an approach that reduces the preprocessing and manual work to minimal by automatically extracting the features from images and focuses on the reduction in computational cost.

The principal problem is the blood spots and patches not having any standard texture and frame, and the blood shade may fluctuate extensively from bright red to deep red to brownish, which makes it challenging to differentiate blood from other articles and the abdominal wall. This heterogeneity in color depends on the sorts of condition, time of the bleeding, capsule location. The manual review of WCE video also requires an average reading time of 45–120 minutes approximately [8,16]. Also, important frames can be missed in a matter of seconds by the human eye. Based on the encounters, it has been extremely advised to study the video in a single uninterrupted sitting. While some issues can be immediately discerned, like bleeding or contractions, some may be obscure like inactive bleeding, blood specks that hard to detect [47]. By using machine learning, images with bleeding can be detected from a very large number of frames. It saves a lot of time and labor and it can also catch obscure deformities that are not evident to the naked eye. Henceforth, delivering better accuracy in detection of bleeding frames and contributing to the state-of-the-art. In paper [48], authors stated that in spite of various available ways to detect anomalies in GI tract by using artificial intelligence with WCE images, the problem still persists as many of these methods are either too complex computationally or too expensive.

3.2 Gap analysis

Based on recent existing researches, the underlying limitations are found, which if not taken into account, can result in misdetection rates and poor efficiency.

- In [31], authors stated that using only saturation-based color feature in their previous work [49] resulted in incorrect segmentation of several CE frames due to the shortage of data.
- Working on an inadequate number of features extracted from CE frames led to inaccurate results [31].
- In paper [32], it is stated that pixel-based approach used in [50] underwent many computations resulting into high cost despite the adequate efficiency.
- Patch based and superpixel based SVM delivered fewer complex computations but lower precision [32].
- Performance of approaches based solely on color spaces is not satisfying [33].
- High accuracy detection is missing in the Cluster-based methods [33].

3.3 Research objectives

In this research, we focus on bleeding detection in the wireless-capsule endoscopic videos. Our goal is to propose an automated technique for the detection of presumed frames that can have the appearance of bleeding. The method pre-processes the images or video, recognize and transfer speculated frames, and prepare them to be visually examined, whereas the other frames are ricocheted and not transmitted for the examination. This may significantly lessen the evaluation time while the ultimate judgment is still left to the endoscopy experts. Research objectives of our thesis includes the research on the state-of-the-art methods for the chosen problem area and proposing a solution for the same. The research objectives of our thesis are summarized as follows:

- To search and analyze the previously available bleeding detection systems for the WCE videos and to study their limitations.

- To explore the challenges in the content retrieval in the medical images due to three factors (a) movement of camera contained in the capsule (b) organ movements (c) luminance is not ideal due to internal body pictures with limited camera movement (d) no control over the movement of the camera and the organs.
- To deal with noise and blur involved in the images along with and imbalance ratio of healthy > sick image classes present in the dataset.
- To propose a computer-assisted binary classification technique for detecting and classifying the frames from WCE images dataset as bleeding and non-bleeding images by making use of machine learning algorithm, convolutional neural network (CNN) which is very reliable in extracting valuable features and detecting patterns, edges, objects, specific color space, etc.
- To validate and test the stated procedure using different best performance parameters for quantitative analysis.

CHAPTER 4: Methodology

A computerized system for bleeding detection in WCE images is proposed to catch the presence of a threat in the digestive tract that caused the bleeding. A dataset of WCE videos with frames holding both bleeding and non-bleeding frames is collected (from PSRI hospital, New Delhi) and used for the proposed approach. At very first, we extracted images from the collected WCE videos using a VLC media player at a rate of 2 frames per second. We have used VLC software for the extraction as it is very simple easy to use. One can set the properties of the images to be extracted like dimensions, bit depth, frame extraction rate, etc. in the preferences tab of the VLC software.

At an initial phase, image preprocessing steps are done to enhance the performance of the algorithm to which the processed images are being fed. The first step of the pre-processing is to *resize* all the images to the same size, in case the acquired images have different sizes. It is done as the deep learning algorithms need the images to be of the same size to ensure the efficient implementation of the operations being carried out. In the second step, the resized images are treated to a *de-noising* algorithm to soften the images, to remove any noise and, unwanted distortions present in the images. The noise present in the images degrade the algorithm's performance and increase complexity and computations [51]. The essential quality of a reliable image de-noising technique is that it should effectively eliminate noise as much as plausible as well as conserve edges and information [52]. We have used the stationary discrete wavelet transform (SWT) for de-noising process in our work. The stationary discrete wavelet transform is a wavelet transform tool that enables to examine the statistically non-predictable signals, particularly at those regions that have discontinuities [53]. Since images have discontinuities at the edges, they can be presented spatially in a multi-resolution manner by using the SWT technique [54]. These RGB de-noised images are then classified using supervised deep learning.

As detecting bleeding frames is a binary classification problem. Supervised learning has been chosen because labeled dataset has been obtained from known gastroenterologists for model training. Thus, we can also evaluate its performance based on the correct classifications and misdetection rate as given by the labeled dataset. It is conducted by employing a Convolutional Neural Network (CNN)

architecture to classify the frames into bleeding and non-bleeding to detect the presence of an anomaly in the digestive system as it offers a wide range of benefits over other ML techniques. It is fast, efficient, and it needs very limited preprocessing of the images [39, 43, 44] as it derives the features from images automatically, which results in a reduced human intervention. Features are learned during the training of the network on the image dataset. Another benefit of using CNN is that only the number of filters and the filter size is required to be defined, whereas the values of the filters are determined by CNN automatically during the training phase. Unlike most of the other ML techniques, object detection in images is carried out by CNN regardless of the location of the object to be recognized. Pooling feature of CNN also prevents overfitting of the network. CNN carry out automatic segmentation of the images for classification.

4.1 Proposed approach

The proposed method as shown in figure 4.2 works on preprocessing of WCE images by applying a de-noising algorithm and color segmentation on them and implementing supervised deep learning for classification.

- WCE video dataset is collected from PSRI hospital (New Delhi) and fed to the VLC software to extract images. Figure 4.1 shows the snapshot of the extracted image dataset.

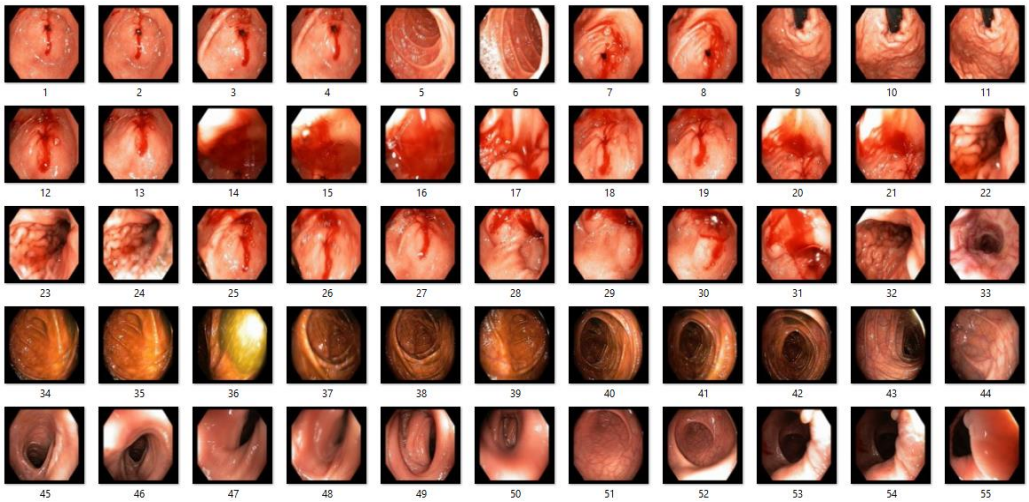


Figure 4.1: Snapshot of the WCE dataset collected from PSRI hospital [6]

- The number of WCE images extracted for the dataset is 2,621 with 505 bleeding and 2,116 normal frames.

- All the images extracted are resized to prepare them to feed to our proposed model. High-resolution images involve more computations and higher memory specifications. If the input is a scaled-down variant of the bigger images, then determining key features in the initial layers will be easier for the network. For higher resolution images, these principal features might be discovered at the subsequent layers of the network [55]. So, we have resized our WCE images to a size of 100×100 pixels for a scaled-down CNN.
- Stationary Discrete Wavelet Transform tool in MATLAB (SWT) is used as a de-noising algorithm to smoothen the images, to remove any noise and, undesired distortions present in the images.
- These processed images are then fed to the convolutional neural network (CNN). The complexity of the model depends upon the complexity of the data. We can start by adding only one hidden layer in the network with neurons and then check the test accuracy of the network using cross-validation. And then deepen the network by adding more layers and neurons. The proper way to decide the number of hidden layers and neuron is discussed in section 4.4 in detail.
- CNN works by automatically extracting features from the image using the training set of WCE images.
- CNN parameters and options are tuned to obtain better performance. The methods for tuning the options of the CNN is discussed in section 4.4 in detail.
- The trained network is applied to the test set of images for the classification after it met the criteria for validation.
- The accuracy, sensitivity, and specificity of the network are measured from confusion matrix, Precision-Recall graph curve is also plotted.
- Performance of the proposed model is compared with state-of-the-art methods such as Linear, SVM, ANN and random forest algorithms.

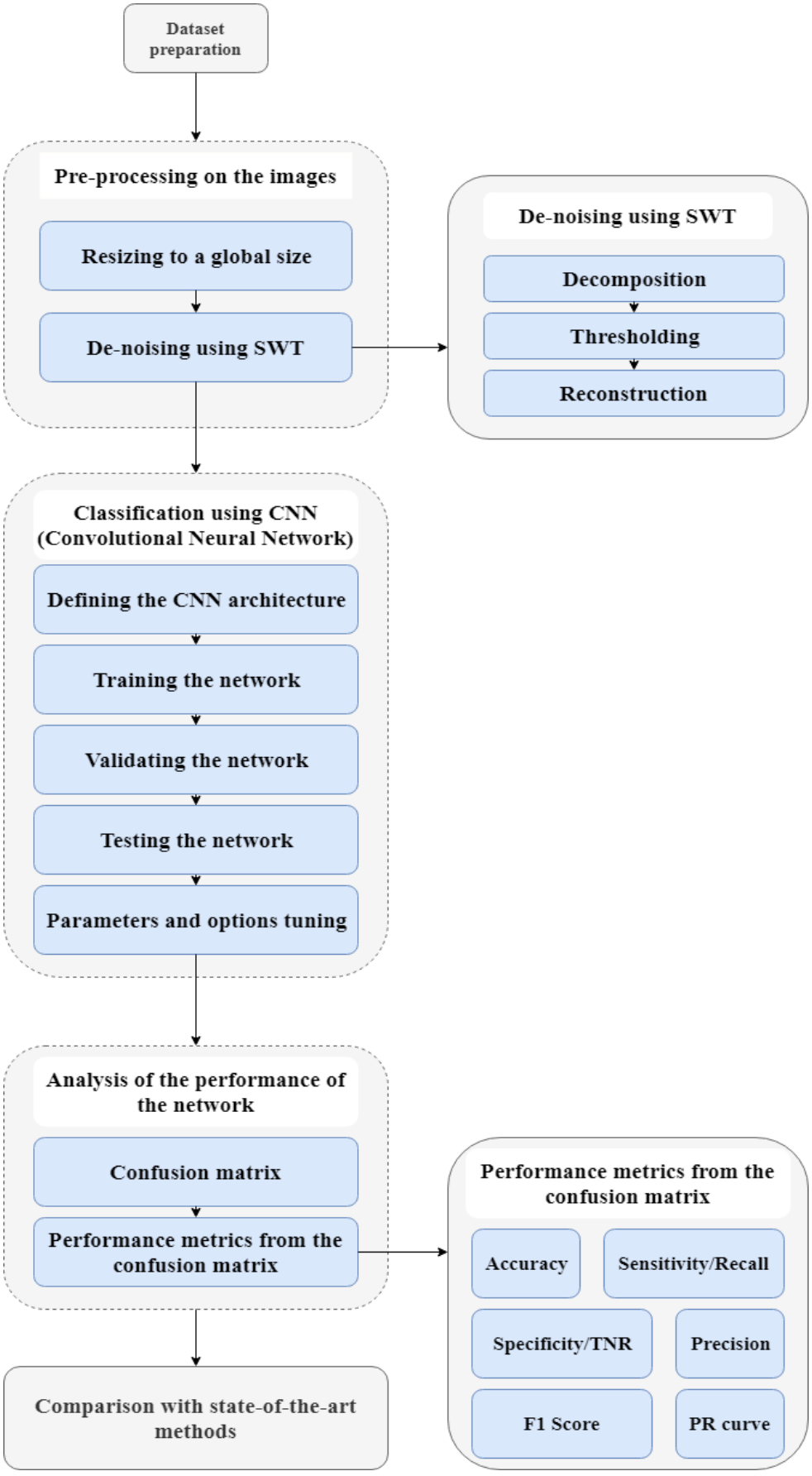


Figure 4.2: Methodology of the proposed system

4.2. Dataset preparation

The WCE dataset used in this research is collected from the PSRI institute of gastroenterology (Pushpawati Singhania Research Institute, Delhi, India). A set of 2621 WCE images is extracted from the video using VLC software at a rate of 2 frames per second. This dataset comprises of 505 bleeding frames and 2,116 normal frames. The WCE images have color homogeneity problem as the color to be detected has a wide range of shades. The blood shade may fluctuate extensively from bright red to deep red to brownish, also containing redness of the normal skin tissues.

4.3 Pre-processing on the images

After acquiring the WCE image dataset, image preprocessing steps are done to enhance the performance of the model to which the processed images are being fed. Pre-processing steps of image processing techniques are applied to the acquired image dataset for getting more enhanced images which are then fed to the network for binary classification. The steps for preprocessing are as follows:

4.3.1 Resizing to a global size

The first step of the pre-processing is to resize all the images to the same size, in case the acquired images have different sizes. It is done as the deep learning algorithms need the images to be of the same size to ensure the efficient implementation of the network. High-resolution images involve more computations and higher memory specifications. If the input is a scaled-down variant of the bigger images, then determining key features in the initial layers will be easier for the network. For higher resolution images, these principal features might be discovered at the subsequent layers of the network [55]. So, beforehand performing the next steps of the proposed system, the CE images are resized into a size of 100×100 pixels to make all the images of equal dimensions. We have chosen 100×100 pixels size for a scaled-down CNN.

4.3.2 De-noising using SWT

In our thesis, we have implemented stationary wavelet transform (SWT) for noise removal. The stationary discrete wavelet transform is a wavelet transform tool that enables to examine the statistically non-predictable signals, particularly at the region

of discontinuities [53]. Since images have discontinuities at the edges, they can be presented spatially in a multi-resolution manner by using the SWT technique [54]. SWT is a wavelet transform that is used to transform signals or images to derive valuable information for the analysis as well as saves the computational cost and reduce required memory space. As we are working with medical images, the loss in information can adversely affect the result. So, SWT has been chosen for de-noising to secure the medical information in the images while undergoing the decomposition and de-noising. SWT de-noising requires decomposition, level thresholding, and inverse transforming for reconstructing the image. For the step of decomposition, SWT algorithm decomposes an image into coefficients to get details about the image like contrast, correlation, energy, homogeneity, entropy, etc. It is done by a choosing a wavelet that decomposes an image to get desired results. There are many wavelets that can be applied for decomposition like Haar, Daubechies, SymN, etc. We have chosen Daubechies for our work as it best proven in paper [56] for the image decomposition and de-noising. Daubechies wavelet being an orthogonal wavelet does not unnecessarily color the white noise, preserves the energy and relatively offer longer support [57]. The SWT decomposition of the image results into four sub-images to get the coefficients. These sub-images are obtained from different applications of vertical and horizontal filters that are Low-pass filter (LPF) and High-pass filter (HPF). The resultant four sub-images of varying contrast, orientations, sharpness, and resolutions are called as approximations (average component) and detail components (horizontal, vertical and diagonal). The approximation coefficient is the low-resolution components and is the result of the LPF (averaging filter) of the SWT. Whereas, the detail coefficients are the output of the HPF (difference filter) of the SWT and are considered as the high-resolution components. The SWT involves up-sampling performed by of the LPF and HPF instead of down-sampling as done in discrete wavelet transform. Up-sampling involves transforming the filters by embedding zeros between the coefficients of the filters which enables the SWT to preserve the coefficients at each level of transformation. In this way, the SWT provides us with coefficients at each level over the similar frequency ranges and same as original in length. It means that the produced coefficients are half in resolution but are same in size as the input image. Whereas, in discrete wavelet transform, the coefficients are halved in size with each successive level as it involves down-sampling [58]. During the SWT decomposition of the images, some of the obtained

coefficients contain important details (high-resolution sub-images). Very small details are disregarded without essentially altering the key features of the images. The purpose of thresholding is to replace all high-resolution sub-image coefficients that are smaller than a selective threshold with zero. These coefficients are utilized in the reconstruction of the image by using an inverse wavelet transformation. The effectiveness of SWT de-noising depends on the threshold value [59]. We have chosen global thresholding by employing a singular threshold value to every detail coefficient in all wavelet levels. The values higher than the threshold value are considered as data or details whereas smaller values are deemed as noise. The threshold limit must be implemented in one of the two modes, a soft or hard thresholding mode. As stated in [60], Soft thresholding provides a more visually pleasing image and decreases the sudden sharp transitions that happen in hard thresholding. Therefore, soft thresholding is chosen over hard thresholding. After de-noising the images with these threshold limits, we apply an inverse transform to reconstruct the original image from the sub-images but without noise. The reconstruction is the reversed course of decomposition or inverse wavelet transform. The modified approximation and detail coefficients at all levels are up sampled and moved through the low pass reconstruction filter (LPRF) and high pass reconstruction filters (HPRFs) and then added to obtain a level reconstructed image. To obtain the original image, this manner is maintained through the corresponding number of levels as in the process of decomposition of the original image [61].

4.4 Classification using CNN (Convolutional Neural Network)

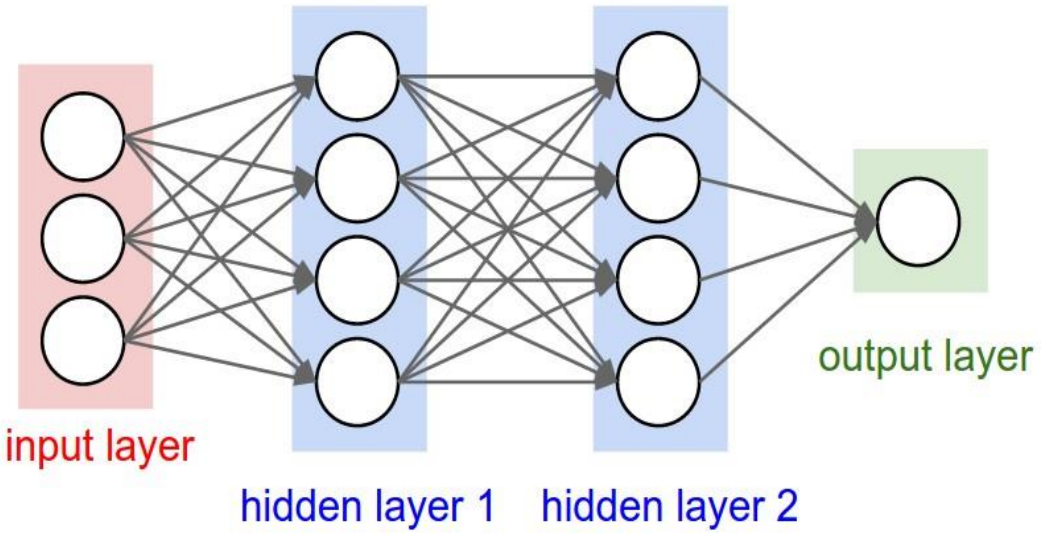
Classification of frames into bleeding and non-bleeding category is done by employing a convolutional neural network (CNN). De-noised images are fed into ConvNet (CNN) for detection of the bleeding in the processed set of images.

4.4.1 Defining the CNN architecture

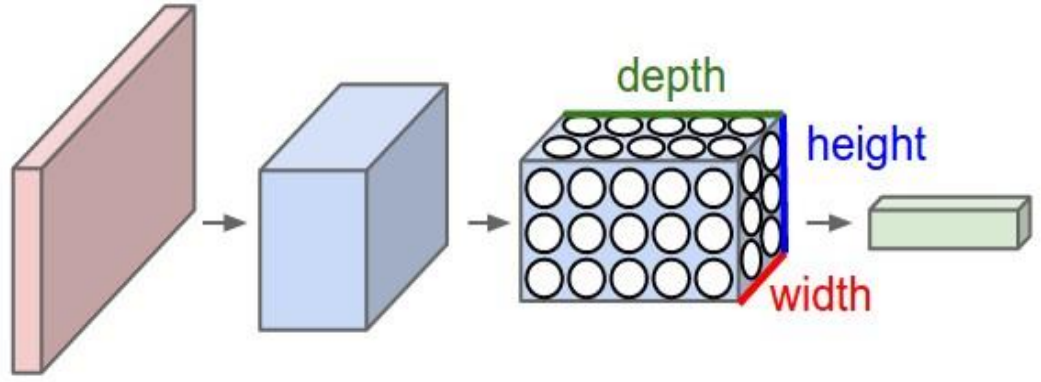
Neurons arrangement in CNN: ConvNet takes in images as three-dimensional objects. The neurons in the layers of CNN are organized in 3 dimensions: width, height, depth. Depth is same as the number of color channels in the input image. The arrangement of neurons in a neural network and in a CNN is depicted in the figure 4.3. Determining the optimal number of hidden layers and neurons that are

incorporated into the network is a crucial part. Underfitting arises as a result of using too few neurons in the layers to sufficiently recognize the signals in a complex data set. Whereas, using too many neurons can give rise to numerous issues. First, it can amount to overfitting in the network. Overfitting happens when the network has excessive data processing capability such that the confined amount of data comprised in the training set is not sufficient to train all of the neurons in the hidden layers. Another issue that can arise is the increased time taken to train the network with an overly huge number of neurons in the layers. Some adjustment needs to be made between an abundant and inadequate number of neurons in the hidden layers. We can start by adding only one hidden layer in the network with neurons and then check the test accuracy of the network using cross-validation. And to optimize the network, we deepen the network so that it can deal with much more complex data and avoid underfitting. According to Heaton research in [62], more than 2 hidden layers in the network generally result in enabling the model to learn complex representations and learn features automatically. We can try to optimize the performance by adding more neurons in the existing hidden layers or by adding new hidden layers. But adding too many neurons result in overfitting of the network. So, it is a hit-and-trial method to choose the number of hidden layers and neurons for the network. Heaton research [62] suggests that we can use a thumb rule instead of hit-and-trial which can be time-consuming and laborious. Heaton thumb rule states that for defining a satisfactory number of neurons to organize in the hidden layers, the following rules can be helpful:

- The number of neurons should fall within the range of the size of the input layer and the output layer.
- It shall be 2/3 the size of the input layer, as well as the size of the output layer.
- It needs to be less than twice the size of the input layer.



(a)



(b)

Figure 4.3: Neurons arrangement in networks: (a) a typical NN, (b) a CNN [24]

The very first layer in the network is an Image Input Layer that receives the de-noised images that are fed to the network as input. The size of the de-noised image (100×100×3, 3 is for the color channels) is specified as input size. We have incorporated 3 Conv layers, 2 pooling layers and, 1 fully connected layer. Due to small input image patch, 3 layers of convolution is enough, two pooling layers to be in between convolutional layers and one fully connected layer for binary classification. A batch normalization layer is always introduced after every Conv layer, followed by a ReLU layer to maintain the non-linearity of the image. Non-linear features of an image are the changes and shifts in pixels, the edges, borders, various colors, etc. Linear images do not appear normal to the human eye as they lack brightness and above-mentioned features. Conventional convolutional layers are linear in nature. This linearity limits their expressiveness. To overcome this limitation,

various non-linear activation functions have been incorporated in CNN. A combination of linear inputs cannot generate a non-linear output. So, without a non-linear activation function, the network will act like a single-layer perceptron network as accumulating all the layers of the network will follow into a linear function only [63]. As the second layer, a 2-D convolutional layer is defined with eight 3x3 convolutions with stride 1 and the same padding as input images. The pooling layer is applied with the arguments of 2x2 max pooling, the stride of 2 and same padding. One fully connected layer for the bleeding and non-bleeding classes is defined and followed by a soft-max and classification layer to apply soft-max function and evaluate the cross-entropy loss.

4.4.2 Training the network

Dataset is split into the training, validation and testing sets using the proportion of 70%, 15%, and 15% respectively. So, CNN is trained on the training set obtained by randomized 70% splitting while regulating the weights on the network for less training error.

4.4.3 Validating the network

Validation set is used to minimize overfitting in the network. It is to ensure that any increment in accuracy over the training dataset results into an increment in accuracy over a dataset that has not been yet exposed to the network or the network has not been trained on it and that is validation dataset.

4.4.4 Testing the network

The trained network is then run on the test dataset of images for the classification of the bleeding and non-bleeding frames and to evaluate the performance and prediction robustness of the network.

4.4.5 Parameters and options tuning

For better performance of the network, parameter and options are tuned to better values. Stochastic gradient descent with momentum (SGDM) is used as a solver with mini-batches of sizes 10. The total number of training samples in a batch is the batch size. Too small batch size results into gradient descent not being smooth, slow learning of the model and error may oscillate too much, whereas too high batch size

results into the longer time required to do one training iteration with relatively small results [64]. SGDM is one of the best optimization algorithms as it helps in preventing oscillations. The number of epochs is the number of passes through the whole training set while training the network. Using a large number of epochs can result in the overfitting of the network and using a very small number of epochs result in an underfit network [25]. Early stopping process enables us to use a large number of epochs but stops training the network as soon as a validation criterion is met. A smaller learning rate decreases the speed of the learning in the network, but it enables the network to converge smoothly [65]. First, we chose a small learning rate varying between 0.00001 to 1 and then we check the performance of our network. To improve performance, a smaller learning rate is used [65]. The maximum number of epochs used in the proposed network is 6 with a learning rate of 0.00001 for better training after checking the performance of the network with different hit-and-trial values for these parameters.

4.5 Analysis of the performance of the network

The various performance metrics that have been used for the evaluation of the efficiency of the proposed system.

4.5.1 Confusion Matrix

The exactness and reliability of a system are calculated through a confusion matrix which is also termed as an error-matrix. This matrix supports in getting a clear idea of the performance of the system by analyzing the misclassification rate and accuracy. In figure 4.4, we have shown a general confusion matrix for binary classification.

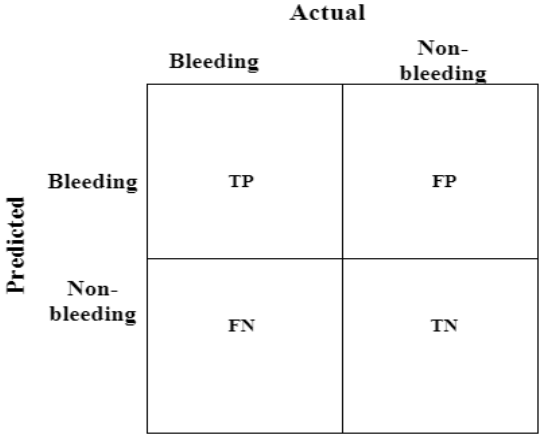


Figure 4.4: A confusion matrix

True Positive (TP): When the predicted and actual classes are identical, true class (1). For instance, A frame that has bleeding present in it is getting predicted in the bleeding class.

True Negative (TN): When an element is predicted to be in False class (0) when it belongs in false class (0) in actual too. For instance, a normal image without any bleeding getting predicted in non-bleeding class.

False Positive (FP): When the system predicts an element to be in true class (1) but in actual it does not. Example: A non-bleeding frame getting predicted in the bleeding class.

False Negative (FN): When the system predicts that an element does not belong to a false class (0) but in actual it does. Example: A bleeding frame is predicted by the model to be a normal frame.

In our research, the WCE image dataset is slightly unbalanced as it comprises of a small ratio of bleeding frames as compared to normal frames. Accuracy alone is not considered a good evaluation metric in cases of unbalanced data classification. Both false negatives (FN) and false positives (FP) are important in medical image classification. In the case of endoscopy images, the cost of false negatives is as important as the false positives [66]. The damage of a bleeding frame to not get detected is worse than the damage of detecting a normal frame as bleeding as all the frames predicted as bleeding will be observed by the physician for the final judgment but the frames predicted as normal frames will probably be overlooked, but also we do not want a large number of false positives in our prediction as it will reduce the efficiency of the network.

4.5.2 Performance metrics from the confusion matrix

Several performance parameters are derived from the confusion matrix such as accuracy, sensitivity, specificity, and others.

- **Accuracy**

  Accuracy refers to the total number of correct classifications done by the network out of the total number of classifications made. As shown in equation (1), accuracy is the rate of true predictions by all the true and false predictions combined.

$$\textbf{Accuracy} = (TP + TN) / (TP + FP + FN + TN) \quad (1)$$

- **Sensitivity or Recall**

  Sensitivity is the rate of accurately predicted positives to genuine positives. Recall gives us an idea about a model's performance proportionate to false negatives. As shown in equation (2), recall focuses on catching all the frames that have "bleeding" with the prediction as "bleeding", not just concerning catching frames correctly. For medical image classification problems, high sensitivity is preferred as it indicated high true positive value and low number of false negatives.

$$\textbf{Sensitivity or Recall} = TP / (TP + FN) \quad (2)$$

- **Specificity or True Negative rate (TNR)**

  Specificity is the rate of accurately predicted negatives to the actual negatives. Specificity is the exact reverse of sensitivity. Equation (3) shows the specificity derived from a confusion matrix.

$$\textbf{Specificity} = TN / (TN + FP) \quad (3)$$

- **Precision**

  Precision is the rate of accurately predicted positives to all the predicted positives. In equation (4), precision shows the proportion of the frames that are detected as having the presence of bleeding, actually had bleeding.

$$\textbf{Precision} = TP / (TP + FP) \quad (4)$$

Recall provides us an idea about a network's performance concerning false negatives, the frames that the network missed, and the Precision provides us with the idea of its performance concerning false positives, the frames that were predicted. Precision is about predicting frames correctly, whereas Recall is about prediction all the positive frames correctly [67]. So, for minimizing false negatives, we have to focus on getting Recall as best as possible with a decent and acceptable Precision value. The values of both Precision and Recall can be monitored by a single value performance metric called as F1 score.

- **F1 Score**

  To consider the role of both precision and recall, the F1 score is computed. As presented in equation (5), F1 score is simply the harmonic mean of precision and recall. In the case of unbalanced class distribution in the dataset, F1 score is a better evaluation metric than accuracy. Low value of F1 score indicates a problem when one of the Precision and Recall has a low value. In that case, F1 score is closer to the smaller value than the bigger value out of these two.

$$\textbf{F1 Score} = (2 * \textbf{Precision} * \textbf{Recall}) / (\textbf{Precision} + \textbf{Recall}) \quad (5)$$

- **Precision-Recall curve**

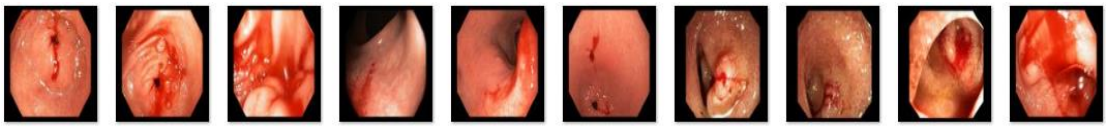
  To demonstrate the trade-off in precision and recall, the precision-recall (PR) curve is plotted. In an image dataset with imbalanced class distribution, the PR curve is a better evaluation metric than accuracy. Thus, PR curve gives a more informational depiction of the performance of the network with unbalanced dataset [68,69]. The area under the PR curve varies from 0 to 1 and also gives an idea about the network's performance. The closer AUC is to 1, the better the network [70]. The closer the curve is to the top-right edge, the more reliable the system. Henceforth, a greater area under the curve (AUC) symbolizes that the system has higher precision and higher recall [69].

4.6 Comparison with state-of-the-art methods

For the evaluation of the proposed model, we have compared the performance of our model with state-of-the-art methods like Logistic Regression model, SVM, artificial neural network (ANN) and Random Forest. These models accept the data of color and texture features as input datasets with the labels for all WCE images obtained from the PSRI hospital, New Delhi. These color and texture features can be extracted from the color histogram and co-occurrence matrix of an individual image respectively. The models are trained, validated and tested on the data to get the evaluation metrics for the respective model. The ratio of train, validation and test data is set as same as our model that is 70%, 15%, and 15% respectively. The performances of these models are then compared with the proposed model.

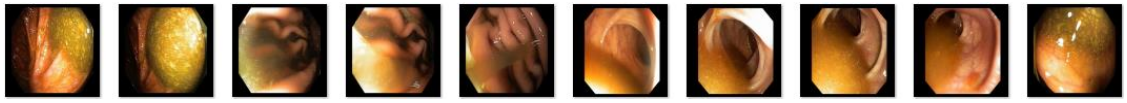
CHAPTER 5: Experimental Results

In this research, we have used MATLAB 2018a for implementing the proposed model. MATLAB offers a good data visualization and it also offers a large number of toolboxes and applications for the image dataset processing and plotting with easy deployment. We have preprocessed the WCE images and trained a convolutional neural network having processor Intel(R) Core (TM) i3-3217U CPU @ 1.80 GHz 1.80 GHz with 4 GB RAM and 64-bit operating system, x-64 based processor. The WCE dataset of 2,621 images are collected from the PSRI institute of gastroenterology (Pushpawati Singhania Research Institute, Delhi, India) comprises of 505 bleeding frames and 2,116 normal frames.

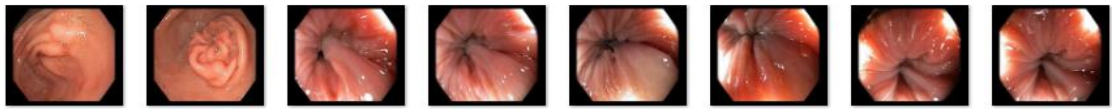


(a) Bleeding

(b) Polyp



(c) Turbid



(d) Wrinkles and Blobs

(e) Wall

Figure 5.1: WCE images from PSRI hospital [6]: (a) Bleeding frames; (b) Polyp; (c) Turbid; (d) Wrinkles and blobs; (e) Wall

In figure 5.1, the sample images from the WCE image dataset collected from PSRI hospital [6] are shown. These sample images show the frames with anomalies like

bleeding, polyp, wrinkles and contractions. The images with turbid present in the system and images of walls of digestive system are also depicted in the figure 5.1.In our research, we focused on extracting the frames that have bleeding present in them from a large number of wireless-capsule endoscopic images. Bleeding in the digestive tract can be a symptom of a severe disorder of the GI system.

5.1 Results of classification by CNN

We implemented a convolutional neural network on preprocessed WCE images in our proposed system to classify them into the bleeding and non-bleeding frames. The performance of proposed model is compared with state-of-the-art methods like linear discriminant model, SVM, ANN, random forest. The performances of the models are evaluated from metrics derived from confusion matrix like accuracy, specificity, sensitivity, precision, F1 score, etc.

5.1.1 Confusion matrix

Confusion matrix of the proposed system is plotted in figure 5.2 to calculate the evaluation metrics. The metrics like accuracy, sensitivity (recall), specificity, precision, F1 score are evaluated using the values of TP, FP, TN, FN from the confusion matrix. In figure 5.2, we can see that the false negatives are 28 and false positives are 4. The test accuracy obtained is 91.9%.

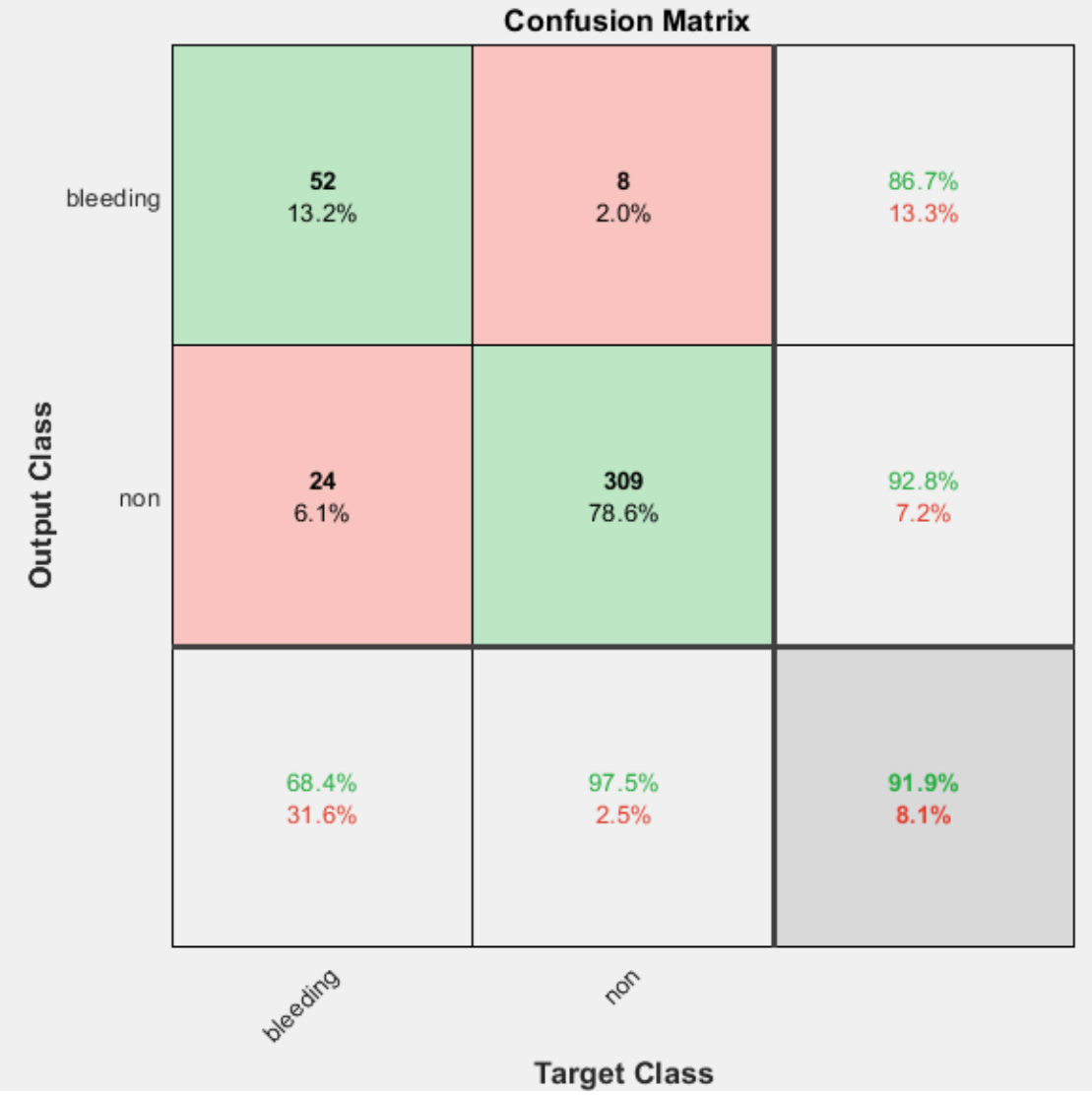


Figure 5.2: Confusion matrix for CNN model

The values taken from confusion plot are the TP = 52, TN = 309, FP = 8, FN = 24, sensitivity (recall or TPR) = 68.5%, specificity (TNR) = 97.5%, precision = 86.7%. Precision and recall are used to calculate F1 score. As we can see in Table 5.1, test accuracy is better than validation accuracy, so it can be concluded that network is not overfitted. Recall has slightly low value and precision is high. F1 score falls in between both the recall and precision. The recall is affected due to the high number of false negatives.

Table 5.1: Evaluation metrics of the model CNN on WCE dataset

S. No.	Evaluation metric	Value
1	Validation Accuracy	90.84%
2	Test Accuracy	91.92%
3	Sensitivity/Recall	68.42%
4	Specificity	97.48%
5	Precision	86.67%
6	F1 Score	0.7647

5.1.2 Precision-Recall curve

The precision-recall curve is plotted using the results from confusion plot in figure 5.3. In a dataset with unbalanced class distribution, PR curve serves to be a more reliable performance evaluation metric than the accuracy of the network. PR curve gives a more informational depiction of the performance of the network with unbalanced dataset [69,70]. The area under the PR curve varies from 0 to 1 and also gives an idea about the network's performance. The closer AUC is to 1, the better the network [70]. As presented by the figure 5.3, the curve depicts the trade-off between precision and recall of the network. The area under the PR curve (AUC-PR) is calculated to be 0.82. The value of AUC-PR is affected by the low recall value in our experiments. PR Curve depicts the trade-off between precision and recall values of the model.

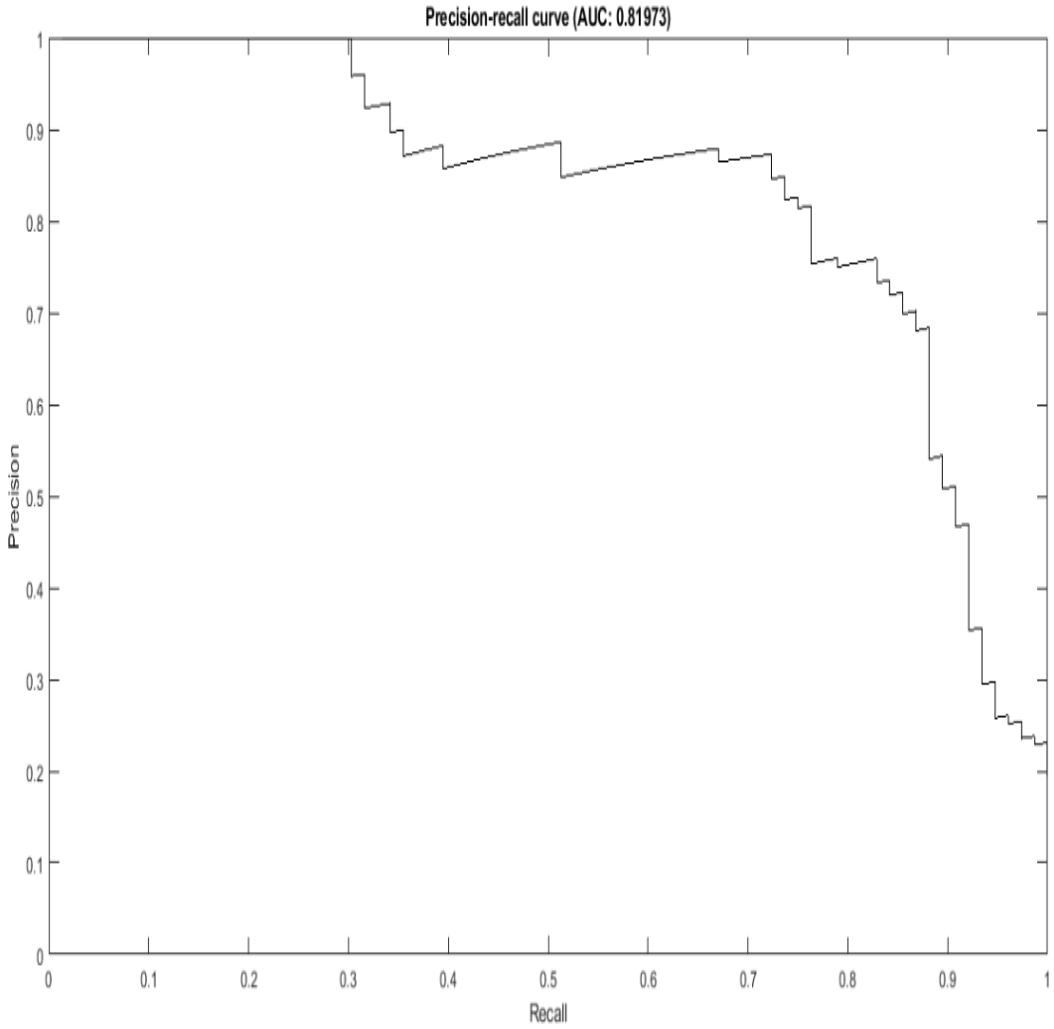


Figure 5.3: Precision-Recall curve of CNN model

5.2 Comparison of the proposed model with state-of-the-art methods

In Table 5.2, we have compared the performance of our model with the other models taken into consideration based on these evaluation metrics. Due to the fact of WCE image dataset being imbalanced in nature, the performance evaluation of the models cannot rely on just accuracy and therefore, more important metrics are calculated. As we can see in the Table 5.2, the F1 score of SVM and Random forest is close to the proposed model. But, these two models require manual feature extraction, segmentation, thresholding, etc. Whereas, the proposed model does not rely on handcrafted features. So, it reduces the human intervention and computation burden.

Table 5.2: Comparing performances of models based on the evaluation metrics

Sr.	ML Model	Accuracy	Sens.	Spec.	Precision	F1 Score
1	CNN	**91.92%**	68.42%	**97.48%**	86.67%	**0.76**
2	Logistic Regression Model	89.72%	82.27%	90.84%	58.92%	0.68
3	SVM	91.91%	92.19%	91.86%	63.29%	0.75
4	ANN	89.43%	56.34%	97.21%	82.79%	0.67
5	Random Forest	91.11%	87.95%	91.61%	62.30%	0.73

Hence, the proposed system with CNN is applied to detect the bleeding frames in the WCE images. The performance of the proposed system on WCE images is slightly better (with F1 score of 0.76) than the state-of-the-art methods as compared in above Table 5.2. The performance of a classifier is not always evaluated just by the accuracy. Other important metrics like precision, recall and F1 score are also evaluated in order to check the predictive power of the model. Moreover, the computational burden and human intervention is also decreased in the proposed model as there is no requirement for manual feature extraction as it is in other above stated models. The manual feature extraction by using color histogram and co-occurrence matrix increases the computations for the model. Whereas, using CNN ensures less human intervention and computations because of the automatic feature extraction functionality of the model.

CHAPTER 6: Conclusion

6.1 Conclusions

The proposed bleeding detection model checks for the bleeding frames in the wireless capsule endoscopy dataset. We have proposed a convolutional neural network for the detection to be done as the network is being fed with the de-noised WCE images. The main tasks accomplished in the research are listed below:

- For the dataset to be used in the classification, the WCE video dataset is collected from PSRI hospital (New Delhi, India).
- WCE images are extracted from the collected video using VLC software and are pre-processed to get the equal-sized, de-noised images to make it easier for the model to work on them.
- A convolutional neural network is implemented for the classification of WCE frames into the bleeding and non-bleeding categories.
- The performance of the proposed model is compared with other state-of-the-art models on the basis of evaluation parameters.

6.2 Challenges and Limitations

The proposed approach involves multiple challenges and assumptions. Challenges involve (a) movements of the camera and organs (b) limited quality of luminance to capture the site texture for diagnose inside the body. Assumptions involve (a) labeled data must be available to train the model (b) de-noising must be done for image clarity (c) data class ratio should be balanced for better classifier accuracy (d) parameter tuning for CNN is also an open question (e) dataset sample used for the case study is quite small to capture most of the image attributes for real world applications (f)system setup for this automatic method and its maintained is also an overhead.

6.3 Future Scope

The future scope of the proposed model of CNN is listed below.

- Bleeding in the digestive tract is a huge threat to the health of a person as the disorders in the digestive system can be fatal. For example, ulcers, colon cancers, etc.

- The proposed system can be trained to detect the other abnormalities of the digestive tract such as polyps, ulcers, tumors, hookworm disease, and many others, based on the features that can be extracted from these disorders.
- The performance and exactness of the proposed system can be enhanced by using a larger dataset for the training, validation, and testing of the network, and try to increase the recall value for the network.
- An ensemble of CNN can be implemented to improve the classification and results of the network can be proposed. A way to try to reduce the number of false negatives and false positives by adjusting the thresholds or by balancing the data can be followed.

References

[1] Bray, F., Ferlay, J., Soerjomataram, I., Siegel, R. L., Torre, L. A., & Jemal, A. (2018). Global cancer statistics 2018: GLOBOCAN estimates of incidence and mortality worldwide for 36 cancers in 185 countries. *CA: a cancer journal for clinicians*, *68*(6), 394-424.

[2] Rawla, P., & Barsouk, A. (2019). Epidemiology of gastric cancer: global trends, risk factors and prevention. *Przeglad gastroenterologiczny*, *14*(1), 26.

[3] Burridge, A. L. (2003). Ancient Egyptian medicine. *JAMA*, *290*(6), 826-827.

[4] Gillespie, M. P. (n.d.). The Language of Medicine - ppt download. Retrieved from https://slideplayer.com/slide/2374097/

[5] Chitra, S., Ashok, L., Anand, L., Srinivasan, V., & Jayanthi, V. (2004). Risk factors for esophageal cancer in Coimbatore, southern India: a hospital-based case-control study. *Indian journal of gastroenterology*, *23*(1), 19-21.

[6] (n.d.). Retrieved from https://www.psrihospital.com/

[7] Iddan, G., Meron, G., Glukhovsky, A., & Swain, P. (2000). Wireless capsule endoscopy. *Nature*, *405*(6785), 417.

[8] Delvaux, M., & Gay, G. (2008). Capsule endoscopy: technique and indications. *Best Practice & Research Clinical Gastroenterology*, *22*(5), 813-837.

[9] Anderson, J., & Roberts, E. (2019, June 24). How Capsule Endoscopy Is Used to Diagnose Digestive Disease. Retrieved from https://www.verywellhealth.com/capsule-endoscopy-for-celiac-disease-diagnosis-562692

[10] McAlindon, M. E., Ching, H. L., Yung, D., Sidhu, R., & Koulaouzidis, A. (2016). Capsule endoscopy of the small bowel. *Annals of translational medicine*, *4*(19).

[11] Lapalus, M. G., Dumortier, J., Fumex, F., Roman, S., Lot, M., Prost, B., ... & Ponchon, T. (2006). Esophageal capsule endoscopy versus

esophagogastroduodenoscopy for evaluating portal hypertension: a prospective comparative study of performance and tolerance. *Endoscopy*, *38*(01), 36-41.

[12] Irvine, A. J., Sanders, D. S., Hopper, A., Kurien, M., & Sidhu, R. (2016). How does tolerability of double balloon enteroscopy compare to other forms of endoscopy?. *Frontline Gastroenterology*, *7*(1), 41-46.

[13] (2019, July 16). Capsule endoscopy. Retrieved from https://www.mayoclinic.org/tests-procedures/capsule-endoscopy/about/pac-20393366.

[14] Cave, D. (2002). Wireless video capsule endoscopy. *Clin Perspect Gastroenterol*, *5*(4), 203.

[15] Van de Bruaene, C., De Looze, D., & Hindryckx, P. (2015). Small bowel capsule endoscopy: Where are we after almost 15 years of use?. *World journal of gastrointestinal endoscopy*, *7*(1), 13.

[16] Swain, P., & Fritscher-Ravens, A. (2004). Role of video endoscopy in managing small bowel disease. *Gut*, *53*(12), 1866-1875.

[17] Liangpunsakul, S., Mays, L., & Rex, D. K. (2003). Performance of Given suspected blood indicator. *The American journal of gastroenterology*, *98*(12), 2676.

[18] D'Halluin, P. N., Delvaux, M., Lapalus, M. G., Sacher-Huvelin, S., Soussan, E. B., Heyries, L., ... & Heresbach, D. (2005). Does the "Suspected Blood Indicator" improve the detection of bleeding lesions by capsule endoscopy?. *Gastrointestinal endoscopy*, *61*(2), 243-249.

[19] Signorelli, C., Villa, F., Rondonotti, E., Abbiati, C., Beccari, G., & de Franchis, R. (2005). Sensitivity and specificity of the suspected blood identification system in video capsule enteroscopy. *Endoscopy*, *37*(12), 1170-1173.

[20] Buscaglia, J. M., Giday, S. A., Kantsevoy, S. V., Clarke, J. O., Magno, P., Yong, E., & Mullin, G. E. (2008). Performance characteristics of the suspected blood indicator feature in capsule endoscopy according to indication for study. *Clinical gastroenterology and hepatology*, *6*(3), 298-301.

[21] Westerhof, J., Koornstra, J. J., & Weersma, R. K. (2009). Can we reduce capsule endoscopy reading times?. *Gastrointestinal endoscopy*, *69*(3), 497-502.

[22] Mahapatra, S. (2019, January 22). Why Deep Learning over Traditional Machine Learning? Retrieved from https://towardsdatascience.com/why-deep-learning-is-needed-over-traditional-machine-learning-1b6a99177063.

[23] Rubeaux, M. (n.d.). AI in Medical Imaging. Retrieved from https://blog.keosys.com/ai-in-medical-imaging.

[24] Johnson, J., &Karpathy, A. (n.d.). Convolutional Neural Networks for Visual Recognition. Retrieved from http://cs231n.github.io/convolutional-networks/.

[25] Brownlee, J. (2019, August 6). Use Early Stopping to Halt the Training of Neural Networks At the Right Time. Retrieved from https://machinelearningmastery.com/how-to-stop-training-deep-neural-networks-at-the-right-time-using-early-stopping/.

[26] Ghosh, T., Li, L., &Chakareski, J. (2018, October). Effective Deep Learning for Semantic Segmentation Based Bleeding Zone Detection in Capsule Endoscopy Images. In *2018 25th IEEE International Conference on Image Processing (ICIP)* (pp. 3034-3038). IEEE.

[27] Hajabdollahi, M., Esfandiarpoor, R., Soroushmehr, S. M., Karimi, N., Samavi, S., &Najarian, K. (2018). Segmentation of bleeding regions in wireless capsule endoscopy images an approach for inside capsule video summarization. *arXiv preprint arXiv:1802.07788*.

[28] Jia, X., & Meng, M. Q. H. (2016, August). A deep convolutional neural network for bleeding detection in wireless capsule endoscopy images. In *2016 38th Annual International Conference of the IEEE Engineering in Medicine and Biology Society (EMBC)* (pp. 639-642). IEEE.

[29] Jia, X., & Meng, M. Q. H. (2017, July). Gastrointestinal bleeding detection in wireless capsule endoscopy images using handcrafted and CNN features. In *2017 39th Annual International Conference of the IEEE Engineering in Medicine and Biology Society (EMBC)* (pp. 3154-3157). IEEE.

[30] Obukhova, N., Motyko, A., Timofeev, B., & Pozdeev, A. (2019, April). Method of Endoscopic Images Analysis for Automatic Bleeding Detection and Segmentation. In *2019 24th Conference of Open Innovations Association (FRUCT)* (pp. 285-290). IEEE.

[31] Tuba, E., Tomic, S., Beko, M., Zivkovic, D., & Tuba, M. (2018, November). Bleeding Detection in Wireless Capsule Endoscopy Images Using Texture and Color Features. In *2018 26th Telecommunications Forum (TELFOR)* (pp. 1-4). IEEE.

[32] Xing, X., Jia, X., & Meng, M. H. (2018, July). Bleeding Detection in Wireless Capsule Endoscopy Image Video Using Superpixel-Color Histogram and a Subspace KNN Classifier. In *2018 40th Annual International Conference of the IEEE Engineering in Medicine and Biology Society (EMBC)* (pp. 1-4). IEEE.

[33] Bchir, O., Ismail, M. M. B., & AlZahrani, N. (2019). Multiple bleeding detection in wireless capsule endoscopy. *Signal, Image and Video Processing*, *13*(1), 121-126.

[34] Sivakumar, P., & Kumar, B. M. (2018). A novel method to detect bleeding frame and region in wireless capsule endoscopy video. *Cluster Computing*, 1-7.

[35] Ghosh, T., Fattah, S. A., & Wahid, K. A. (2018). CHOBS: Color Histogram of Block Statistics for Automatic Bleeding Detection in Wireless Capsule Endoscopy Video. *IEEE journal of translational engineering in health and medicine*, *6*, 1-12.

[36] Ghosh, T., Fattah, S. A., Wahid, K. A., Zhu, W. P., & Ahmad, M. O. (2018). Cluster based statistical feature extraction method for automatic bleeding detection in wireless capsule endoscopy video. *Computers in biology and medicine*, *94*, 41-54.

[37] Nawarathna, R., Oh, J., Muthukudage, J., Tavanapong, W., Wong, J., De Groen, P. C., & Tang, S. J. (2014). Abnormal image detection in endoscopy videos using a filter bank and local binary patterns. *Neurocomputing*, *144*, 70-91.

[38] Lan, L., Ye, C., Wang, C., & Zhou, S. (2019). Deep Convolutional Neural Networks for WCE Abnormality Detection: CNN Architecture, Region Proposal and Transfer Learning. *IEEE Access*, *7*, 30017-30032.

[39] Diamantis, D. E., Iakovidis, D. K., & Koulaouzidis, A. (2019). Look-behind fully convolutional neural network for computer-aided endoscopy. *Biomedical Signal Processing and Control*, *49*, 192-201.

[40] Deeba, F., Islam, M., Bui, F. M., & Wahid, K. A. (2018). Performance assessment of a bleeding detection algorithm for endoscopic video based on classifier fusion method and exhaustive feature selection. *Biomedical Signal Processing and Control*, *40*, 415-424.

[41] Leenhardt, R., Vasseur, P., Li, C., Saurin, J. C., Rahmi, G., Cholet, F., ... & Sacher-Huvelin, S. (2019). A neural network algorithm for detection of GI angiectasia during small-bowel capsule endoscopy. *Gastrointestinal endoscopy*, *89*(1), 189-194.

[42] Vieira, P. M., Silva, C. P., Costa, D., Vaz, I. F., Rolanda, C., & Lima, C. S. (2019). Automatic Segmentation and Detection of Small Bowel Angioectasias in WCE Images. *Annals of biomedical engineering*, *47*(6), 1446-1462.

[43] Alaskar, H., Hussain, A., Al-Aseem, N., Liatsis, P., & Al-Jumeily, D. (2019). Application of Convolutional Neural Networks for Automated Ulcer Detection in Wireless Capsule Endoscopy Images. *Sensors*, *19*(6), 1265.

[44] Aoki, T., Yamada, A., Aoyama, K., Saito, H., Tsuboi, A., Nakada, A., ... & Matsuda, T. (2019). Automatic detection of erosions and ulcerations in wireless capsule endoscopy images based on a deep convolutional neural network. *Gastrointestinal endoscopy*, *89*(2), 357-363.

[45] Billah, M., & Waheed, S. (2018). Gastrointestinal polyp detection in endoscopic images using an improved feature extraction method. *Biomedical engineering letters*, *8*(1), 69-75.

[46] He, J. Y., Wu, X., Jiang, Y. G., Peng, Q., & Jain, R. (2018). Hookworm detection in wireless capsule endoscopy images with deep learning. *IEEE Transactions on Image Processing*, *27*(5), 2379-2392.

[47] Byrne, M. F., & Donnellan, F. (2019). Artificial intelligence and capsule endoscopy: Is the truly "smart" capsule nearly here?. *Gastrointestinal endoscopy*, *89*(1), 195-197.

[48] Novozámský, A., Flusser, J., Tachecí, I., Sulík, L., Bureš, J., & Krejcar, O. (2016). Automatic blood detection in capsule endoscopy video. *Journal of biomedical optics*, *21*(12), 126007.

[49] Tuba, E., Tuba, M., & Jovanovic, R. (2017, May). An algorithm for automated segmentation for bleeding detection in endoscopic images. In *2017 International Joint Conference on Neural Networks (IJCNN)* (pp. 4579-4586). IEEE.

[50] Al-Rahayfeh, A. A., & Abuzneid, A. A. (2010). Detection of bleeding in wireless capsule endoscopy images using range ratio color. *arXiv preprint arXiv:1005.5439*.

[51] Kumar, V., & Samadhiya, A. (2019). Comparative performance analysis of image de-noising techniques. *arXiv preprint arXiv:1901.06529*.

[52] Anutam, R. (2014). Performance analysis of image de-noising with wavelet thresholding methods for different levels of decomposition. *Int. J. Multimed. Its Appl*, *6*(3).

[53] Acharya, T., & Ray, A. K. (2005). *Image processing: principles and applications*. John Wiley & Sons.

[54] Pandey, V. (2014). Analysis of image compression using wavelets. *International Journal of Computer Applications*, *103*(17).

[55] Van Noord, N., & Postma, E. (2017). Learning scale-variant and scale-invariant features for deep image classification. *Pattern Recognition*, *61*, 583-592.

[56] Luisier, F., Blu, T., Forster, B., & Unser, M. (2005, September). Which wavelet bases are the best for image denoising?. In *Wavelets XI* (Vol. 5914, p. 59140E). International Society for Optics and Photonics.

[57] Masumdar, R. E., &Karandikar, R. G. (2016). Comparative Study of Different Wavelet Transforms in Fusion of Multimodal Medical Images. *International Journal of Computer Applications*, *146*(11).

[58] Alwan, I. M. (2012). Color Image Denoising Using Stationary Wavelet Transform and Adaptive Wiener Filter. *Al-Khwarizmi Engineering Journal*, *8*(1), 18-26.

[59] Bitenc, M., Kieffer, D. S., &Khoshelham, K. (2015). EVALUATION OF WAVELET DENOISING METHODS FOR SMALL-SCALE JOINT ROUGHNESS ESTIMATION USING TERRESTRIAL LASER SCANNING. *ISPRS Annals of Photogrammetry, Remote Sensing & Spatial Information Sciences*, *2*.

[60] Anutam, R. (2014). Performance analysis of image de-noising with wavelet thresholding methods for different levels of decomposition. *Int. J. Multimed. Its Appl*, *6*(3).

[61] Mortazavi, S. H., &Shahrtash, S. M. (2008, September). Comparing de-noising performance of DWT, WPT, SWT and DT-CWT for partial discharge signals. In *2008 43rd International Universities Power Engineering Conference* (pp. 1-6). IEEE.

[62] Heaton, J. (2015). *Artificial Intelligence for Humans, Volume 3: Deep Learning and Neural Networks*. s.l.: Createspace Independent Publishing Platform.

[63] Zoumpourlis, G., Doumanoglou, A., Vretos, N., &Daras, P. (2017). Non-linear convolution filters for cnn-based learning. In *Proceedings of the IEEE International Conference on Computer Vision* (pp. 4761-4769).

[64] (n.d.). Retrieved from https://jhui.github.io/2018/02/11/How-to-start-a-deep-learning-project/.

[65] Yu, F. (n.d.). A Comprehensive guide to Fine-tuning Deep Learning Models in Keras. Retrieved from https://flyyufelix.github.io/2016/10/03/fine-tuning-in-keras-part1.html.

[66] Silva, J., Histace, A., Romain, O., Dray, X., &Granado, B. (2014). Toward embedded detection of polyps in wce images for early diagnosis of colorectal cancer. *International Journal of Computer Assisted Radiology and Surgery*, *9*(2), 283-293.

[67] Drakos, G. (2018, September 12). How to select the Right Evaluation Metric for Machine Learning Models: Part 3 Classification Metrics. Retrieved from https://towardsdatascience.com/how-to-select-the-right-evaluation-metric-for-machine-learning-models-part-3-classification-3eac420ec991.

[68] Davis, J., & Goadrich, M. (2006, June). The relationship between Precision-Recall and ROC curves. In *Proceedings of the 23rd international conference on Machine learning* (pp. 233-240). ACM.

[69] Saito, T., & Rehmsmeier, M. (2015). The precision-recall plot is more informative than the ROC plot when evaluating binary classifiers on imbalanced datasets. *PloS one*, *10*(3), e0118432.

[70] Brownlee, J. (2019, September 25). How to Use ROC Curves and Precision-Recall Curves for Classification in Python. Retrieved from https://machinelearningmastery.com/roc-curves-and-precision-recall-curves-for-classification-in-python/.

Plagiarism Report

Thesis

ORIGINALITY REPORT

10%	3%	4%	6%
SIMILARITY INDEX	INTERNET SOURCES	PUBLICATIONS	STUDENT PAPERS

PRIMARY SOURCES

1	Adam Novozámský, Jan Flusser, Ilja Tachecí, Lukáš Sulík, Jan Bureš, Ondrej Krejcar. "Automatic blood detection in capsule endoscopy video", Journal of Biomedical Optics, 2016 Publication	2%
2	www.wjgnet.com Internet Source	1%
3	Submitted to University of Wales Institute, Cardiff Student Paper	1%
4	Submitted to Bridgepoint Education Student Paper	<1%
5	Nazare Jr., Antonio C., and William Robson Schwartz. "A scalable and flexible framework for smart video surveillance", Computer Vision and Image Understanding, 2016. Publication	<1%
6	www.ijcea.com Internet Source	<1%