Segmentation and Classification of Gastroenterology Images using Saliency Mapping in CNNs

By:

Yusra Khan
(Registration No: MS-CE-20-330106)

Supervisor:
Dr. Arslan Shaukat

DEPARTMENT OF COMPUTER & SOFTWARE ENGINEERING,
COLLEGE OF ELECTRICAL & MECHANICAL ENGINEERING,
NATIONAL UNIVERSITY OF SCIENCES AND TECHNOLOGY,
ISLAMABAD

Annex A

THESIS ACCEPTANCE CERTIFICATE

Certified that final copy of MS/MPhil thesis written by NS **Yusra Khan** Registration No. 00000330106, of College of E&ME has been vetted by undersigned, found complete in all respects as per NUST Statutes/Regulations, is free of plagiarism, errors and mistakes and is accepted as partial fulfillment for award of MS/MPhil degree. It is further certified that necessary amendments as pointed out by GEC members of the scholar have also been incorporated in the thesis.

Signature : ____________________

Name of Supervisor: Dr Arslan Shaukat

Date: 27 AUG 2024

Signature of HOD: ____________________
(Dr Usman Qamar)
Date: 27 AUG 2024

Signature of Dean: Rashid
(Brig Dr Nasir Rashid)
Date: 27 AUG 2024

Segmentation and Classification of Gastroenterology Images using Saliency Mapping in CNNs

By
Yusra Khan
(Registration No: MS-CE-20-330106)

A thesis submitted to National University of Sciences and Technology, Islamabad
in partial fulfillment of the requirements for the degree of

Master of Sciences in Computer Engineering

Supervisor
Dr. Arslan Shaukat

DEPARTMENT OF COMPUTER & SOFTWARE ENGINEERING,
COLLEGE OF ELECTRICAL & MECHANICAL ENGINEERING,
NATIONAL UNIVERSITY OF SCIENCES AND TECHNOLOGY
ISLAMABAD
August 27, 2024

I want to dedicate my work to my parents, who let me choose the work that I have done, to my siblings, who always gave me positive remarks and support to continue doing my work even when things got really tough, and finally to all of my friends and family members, who believed in me.

ACKNOWLEDGMENT

On the successful completion of my thesis research work, I first and foremost thank Allah Almighty, for without His mercy and help I couldn't have achieved yet another milestone in my academic career. It was only through the strength and perseverance He gave me that I took up to this challenging task of working on cancer detection research topic "Segmentation and Classification of Gastroenterology Images using Saliency Mapping in CNNs" and completed it successfully.

I would offer my heartily thanks to my Thesis Supervisor, Dr. Arslan Shaukat under whose guidance and supervision, I completed my work with much more proficiency. My supervisor's support by any means may it be technical or motivational never failed to raise my spirits and pushed me to give my level best to this research. Thank you, for being part of my thesis work and believing in me. I would also like to thank and mention Dr. Farhan Riaz here, who was my first supervisor and gave me the topic and direction to work on initially. I also thank my GEC members, Dr. Usman Akram, and Dr. Farhan Hussain for being notable members of my research committee and understanding the purpose behind my research work.

Finally, I would like to thank my parents, siblings and friends, whose support and motivation was the oil to the lamp of my trials. They played a key role in helping me stay focused and providing any ease they could in their power to help me achieve my goal.

I feel indebted to all the sincere people I have mentioned here for supporting me to achieve my goal of successful completion of my postgraduate thesis.

ABSTRACT

This research work has been done towards the treatment of Gastrointestinal (GI) Cancer by experimenting as to how a computer aided design can help oncologists classify and segment GI Cancer using Gastroenterology (GE) Images. The concept of Attention Module by the name of Convolution Block Attention Module (CBAM) using Convolutional Neural Network (CNN) layers has been utilized to achieve saliency mapping in these images. The different layers of the CNN are tweaked and the hyper parameters are consistently checked and changed to achieve maximum accuracy. The CBAM Algorithm is divided into Channel Attention (CA) and Spatial Attention (SA) module. The Channel Attention has been implemented in the HSV Color Space for finding the best color space for segmentation purposes of the Chromo endoscopy medical images. Classification, followed by segmentation, is again implemented using CNN based model for normal and abnormal cancer images. For classification, 92% test accuracy has been achieved, while 70% Dice Coefficient has been obtained for segmentation. The results achieved are competitive as compared to the previous reported results on the GE images dataset.

Keywords—Convolutional Block Attention Module (CBAM), CNN, GI Cancer Classification, GI Cancer Segmentation, HSV.

LIST OF CONTENTS

LIST OF FIGURES

LIST OF TABLES

CHAPTER 1

INTRODUCTION

Chapter 1 gives a generic overview about the research that has been conducted. It comprises of brief explanations about the motivation behind this thesis, its introduction, its research orientation, basic working and structure of this thesis report.

1.1 INTRODUCTION TO GI CANCER AND ITS DETECTION

The lethal disease of cancer has eaten up many lives heartlessly for there is no diagnosis and medication fully developed to a level where cancer at any stage can be diagnosed and treated easily. Up till recently, the most dangerous disease which makes its victims hopeless due to the lack of proper diagnosis methodology is cancer. Globally, there were an estimated **20 million** new cases of cancer and **10 million** deaths from cancer in 2023. A well-known philosopher from America, *Ralph Waldo Emerson* once wrote:

"*The first wealth is health*".

In the race of life, the most valuable treasure one can hold is his health. It is therefore a dire need of the world to come up with better solutions and more efficient ways to ensure good health of the people of the earth so that they can earn a livelihood, get good education and enjoy life to its fullest. As cancer currently is the disease which does not have a proper cure or diagnosis methodology, therefore this thesis is directed towards finding yet another way of detection and classification of GI Cancer. Among the different types, there is blood cancer, lung cancer, skin cancer etc. This research of cancer detection, however, deals with a different type of cancer, the one which is even more difficult to be detected appropriately. The "*Gastrointestinal Cancer*" or "*GI Cancer*". The Gastrointestinal tract of human body consists of mouth, esophagus, stomach and the intestines as shown in the labelled GI tract diagram of Figure 1 [1]. The different parts of the Human GI Tract are prone to different types of diseases. For example, the mouth can get a mouth cancer, the esophagus can get barrett's disease, the stomach can get stomach ulcers, polyps and cancer, and the colon or

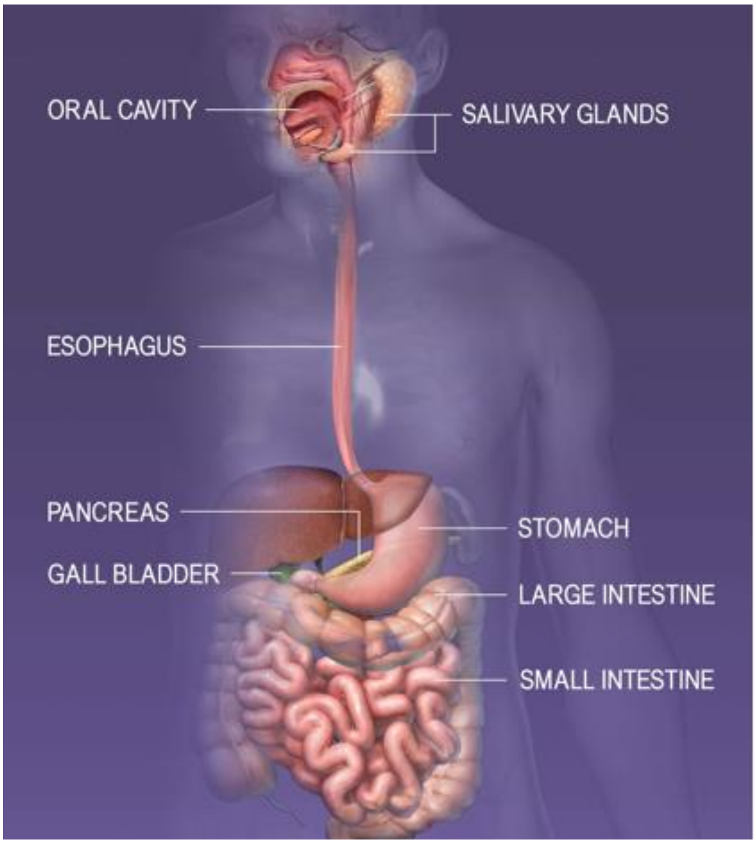

Figure 1: Human GI Tract Labelled Diagram [1]

the intestine can get the colon cancer. This shows that the field of Gastroenterology is indeed very vast and it needs different types of specialists and researchers to work on the diagnosis and prevention of the different GI Tract diseases. This thesis, however, concerns with the cancer caused by the stomach tissue lining. Gastric Cancer accounts for 783,000 deaths each year.

Now, due to the structure of the GI tract, it is quite evident that the detection and diagnosis of cancer in this region of the human body is quite difficult. For this purpose, Computer Aided Design (CAD) can help oncologists to find and ultimately cure the cancer with more accuracy. Researchers have come up with various methods and ways to achieve the goal of better diagnosing cancer in the human body. Different Image Processing techniques like saliency mapping, feature extraction based on supervised and unsupervised learning have been proposed since the beginning of the researches made in the direction of a better cancer diagnosis. This thesis, however, implements the expertise produced by the field of Explainable Artificial Intelligence or Explainable AI. Here, the work done is directly applied at the deepest levels of the machine learning or deep learning layers to achieve higher accuracy.

1.2 MOTIVATION BEHIND THE RESEARCH

It has already been explained in the previous section how the structure of GI tract makes it difficult for the oncologists to diagnose any sort of cancer in this area of the human body. During a GI exam, *Chromo endoscopy* or *Narrow Band Imaging (NBI)* is used to capture images or check at real time during the examination any area of the GI tract which is under observation. As for this research, chromo endoscopy or CE images have been used. But even with highly efficient endoscopy mechanisms, the angle of the endoscope, the shadow casted by the endoscope on the area under observation, a certain limited degree of view of the endoscope, and the limited of movement of the endoscope gives rise to multiple challenges which makes it harder for the specialist to give an accurate diagnosis 100% of the time. It is therefore necessary for computer engineers and other expert researchers to work and develop algorithms which would help specialists in making the final judgement call regarding the area affected by the cancer as well as the type and class of cancer. For this purpose, this thesis work has been carried out. Hence this detailed research is an effort to join hands with the specialists and play our role in the battle against cancer so that the world can move one more step forward to help the patients suffering from this lethal disease.

1.3 AIM AND WORKING

It is now quite clear with the little discussion that has been made so far that the basic aim of this thesis is to come up with a solution to aid the oncologists and other specialists to diagnose cancer with more accuracy based on a second opinion provided by the developed and trained algorithm in this research. There are two parts of the thesis, segmentation and classification. *Python Programming* has been done to create *CNN Algorithms* which first classify the type of cancer and then segment the cancerous region and then find the segmentation and classification accuracies. The purpose is to put the best efforts in training the algorithms to achieve the best accuracy possible for segmentation and classification. Multiple CNN based architectures have been designed, developed and trained using the provided proprietary GI Cancer Images acquired from real time GI tract exam. As mentioned earlier that the work done is actually the application of the concept of Explainable AI for detection and classification of cancer, therefore, experiments have been carried out to come up with the best arrangement and design of the CNN layers to achieve higher accuracy based on Dice Co-efficient values found by comparing the final segmented test images and their respective annotated images for segmentation and final classification accuracy for classification. The final aim is to better the classification and segmentation accuracy on the given GE images dataset by experimenting to find out the right approach via the usage of the best color space as well as by devising the best algorithm using Convolutional Block Attention Module or CBAM.

1.4 PROCEDURE FOLLOWED

A thorough and detailed procedure has been followed during the experiments carried out. First, the GI images have been acquired during a real time GI exam using an endoscope. The best and the most appropriate images were then selected by specialists for this research work. Afterwards, CNN based models have been developed to first classify the image into its type and then eventually use the dedicated algorithm for that type to segment the cancerous region.

During segmentation, the technique of Convolutional Block Attention Module (CBAM) has been utilized. It mainly consists of two modules, the first being the Channel Attention (CA) and the second being the Spatial Attention (SA). For Channel Attention, HSV Color Space has been utilized. It is found via training and testing of CNN models as to which out of the three channels i.e. Hue, Saturation and Value contribute most towards the image ROI content. That channel is utilized finally alongside the channel attention module to segment out the cancerous region. Both the modules are fully trained and tested. The final Dice Similarity Co-efficient values are also calculated for the test images against their annotated images. For the segmentation, the Channel Attention works best with almost no affect from the Spatial Attention Module. U-Net Architecture was also experimented with for segmentation.

For classification, the Spatial Attention has been finally utilized after many experiments done using other state of the art models like VGG etc. The SA works without splitting the image into channels and is thus more ideal for the classification purposes. The number of correctly classified images give the final classification accuracy. Figure 2 shows the procedure followed in this thesis research work.

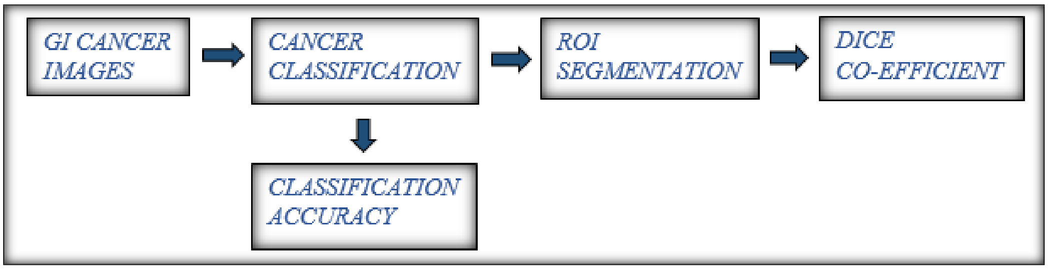

Figure 2: Procedure followed in the thesis

1.5 STRUCTURE OF THESIS REPORT

The work explained in this thesis report has been divided into multiple chapters in the following manner:

- Chapter 1 gives introduction of the work. It consists of the introductory explanations along with the discussion of motivation, aim, working and basic procedure followed in the thesis.
- Chapter 2 has a thorough literature view and background work on the thesis topic of "Segmentation and Classification of Gastroenterology Images using Saliency Mapping in CNNs".
- Chapter 3 deals with the data used and the basic explanation of the methodology applied on it.
- Chapter 4 is designed to show how the classification model is designed, developed, and trained.
- Chapter 5 is designed to show how the segmentation model is designed, developed, and trained based on the type of cancer class classified.
- Chapter 6 discusses the results acquired from the previous two chapters namely chapter 4 and chapter 5 and the conclusions derived from them.
- Chapter 7 concludes the research and talks about how the research and algorithms developed here can be utilized in other fields as well as the field of oncology.

CHAPTER 2

BACKGROUND AND RELATED WORK

Chapter 2 shows the studies made with the aim of knowing and understanding the work which has already been done in the field of classification and segmentation of GE Images.

2.1 RELATED WORK

The concept of digital image processing which is a subcategory of digital signal processing dates back to as early as 1960s. Image Processing was used by different research facilities to mainly refine images for example satellite images. Today, digital image processing due to its advantages over analogue image processing has come up with multidimensional and high end algorithms to assist in the fields of satellite imagery, videophone and especially medical imagery etc. This literature review explains the related work done prior on the development of digital image processing techniques in the field of Gastroenterology. Riaz et al. [2] developed Normalized Cuts (NCuts) combined with creaseness, edge maps and LUV color space of GE Images to work on better segmentation of the GI Cancer Images. The DSC values after experimenting with the normalized cuts for segmentation purposes have been calculated. The value 0.64 or 64%, has been considered as the objective value to be surpassed by the experimental results of segmentation done in this thesis.

On the other hand, Autocorrelation Gabor Filters (AGF) [3] focuses on relative measure between different orientations and scales thus helping in the determination of different characteristics of patterns of images such as shape, size, regularity etc. instead of working on the absolute representation of the features. It works on the classification of the GE Images. The values have been shown in tabular form in [3], but the focus should only be put on CH images. The table in [3] shows the experimental results for the different versions of the Gabor Filters. The value 0.831 or 83.1% has been considered as an objective value to be surpassed by the classification accuracy results obtained in this thesis. Before discussing further, TABLE 1 can be reviewed for an idea about the Literature Review.

Another approach to segment out images is by the use of Dermatologist Knowledge Network (DermaKNet) [4]. This was implemented on skin cancer segmentation. DermaKNet works by first creating a Binary Image Mask of the lesion using a Lesion Segmentation Network (LSN) and then generating additional augmented views v of this mask by applying rotations and crops. Each view is then fed into a Dermoscopic Structure Segmentation Network (DSSN) which segments or breaks down each view into eight dermoscopic features representing global and local structures like dots/globules, regression areas and streaks. These are fed together into a Diagnosis Network (DN). The final diagnosis is done by incorporating the metadata of patients as well like their age, gender etc. This gives different accuracy results for different skin issues. The experimental results show 91.7 average AUC for skin lesion segmentations, while a value of 65.2 as an average Specificity at 95%. But the preprocessing involved in the DermaKNet is too much and does not relate with our aim and objective much.

In [5], Fast and Robust Fuzzy C-means (FRFCM) and Simple Linear Iterative Clustering (SLIC) along with Google's AutoAugmnet technology has been utilized in the preprocessing segmentation stage of the Gastric Cancer images, which leads to a good classification accuracy and gives an area under the curve value of 0.96.

While in [6], five different GI diseases, Polyp, Barret's Esophagus (BE), suspicious, High Grade Dysplasia (HGD) and cancer, have been detected first. And for that, comparison has been done between YOLOV3, and Faster RCNN with Resnet 101 as backbone. Both work well, but for suspicious disease, YOLO V3 does not do a much better job in comparison. For segmentation, different forms of Mask R-CNN have been compared. Cascade Mask R-CNN works the best with a segmentation score of 0.6526±0.3418. Similarly, different other built-in architectures like the U-Net, DenseNet, AlexNet, attention fusion mechanisms, and simple CNN architectures have been used to classify and segment gastric issues as in [7], [8], [9], [10], and [11].

Especially in [7], the authors have done a great job in comparing studies that have already been done in the field of Gastric Cancer detection and classification. The architectures like the AlexNet, ResNet, VGG, Inception, DenseNet and Deep lab, etc. have been studied from other resources available for both classification and segmentation. The studies showed an accuracy of 77.3 – 98.7%.

The work in [8] has been done on two datasets of SEED and BOT and the results were established based on different metrics. Focal Tversky Loss (FTL) is applied to get better detection results along with a V-shaped architecture in which the attention fusion mechanism is applied at the decoder level. The SEED dataset shows DSC of 0.8265, while the BOT set shows a DSC of 0.8991. It is a very good algorithm that has been devised and tested on both the datasets, but the merging of different attention mechanisms with other V shaped architectures and the difference in an updated loss function makes it a very complicated algorithm. Similarly, the datasets are very different to the one which has been used in this research.

I need the readers to **sharply focus** on the work done in [9]. The classification and segmentation of the gastric lesions of the stomach lining has been carried out and very high performance accuracies have been acquired using different ResNet architectures with U-Net. Two different datasets have been utilized as data and the masks. But here is a twist, the researchers have come across the same challenge of defining a good value for thresholding the final images to black and white, but they omitted all the ones giving a low accuracy results. Such biasness has not been introduced in this research.

The segmentation accuracy of 71% has been achieved in [10], which is very close to the 70% DSC achieved in this research. But again, the reason is a back bone already built architecture of ResNet 50 which has been used along with Mask R-CNN and the accuracies have been calculated with a different approach of k=5 fold stratified cross-validation. If the CBAM and then too only half of the CBAM module i.e. Channel Attention without a backbone architecture, gives 70% accuracy as approximate, which could have been even more if the low threshold black and white images were not considered in the final accuracy calculation, then obviously the work done in this research is far more **efficient although it is simple** and has been applied on a very complex dataset.

In [11], only the different stages of Gastric Cancer are detected with an accuracy of 94.1% overall using a deep learning model based on the U-net architecture.

TABLE 1: LITERATURE REVIEW ON SEGMENTATION AND CLASSIFICATION OF GASTRIC CANCER

Reference	Year	Technique	Accuracy	Remarks
Riaz et al. [2]	2012	Invariant Gabor Texture Descriptors	83.1% Classification Acc.	Not ML based
Riaz et al. [3]	2013	Normalized Cuts	DSC of 64% Segmentation Acc.	Conventional DIP technique
Ivan Gonzalez-Diaz [4]	2019	DermaKNet (CNN based)	Avg. Specificity at 95% = 65.2	Too much Pre-processing involved
Lee et al. [5]	2021	Fast and Robust Fuzzy C-means (FRFCM), Simple Linear Iterative Clustering (SLIC) along with Google's AutoAugmnet	Classification accuracy with an area under the curve value of 0.96.	Many already built architectures used. Too complex.
Krenzer et al. [6]	2020	YOLOV3, and Faster RCNN with Resnet 101 for classification and Cascade Mask R-CNN	Out of 5 diseases, YOLOV3 does not do well for classifying suspicious disease For segmentation, score is 0.6526±0.3418	It works on five different diseases and has implemented about 10 architectures. Very different from our work.
Zhao et al. [7]	2022	Review study of AlexNet, ResNet, VGG, Inception, DenseNet and Deep lab	Study accuracy of 77.3 – 98.7%.	Just a review study
S. Xu at al. [8]	2022	V-shaped architecture with Focal Tversky Loss	SEED dataset DSC=0.8265 BOT set DSC = 0.8991	Different dataset and different architecture

Tani et al. [9]	2022	ResNet architectures with U-Net	90% plus classification and segmentation accuracy	Below 85% values were omitted, introducing biasness
Shibata et al. [10]	2020	Res-Net 50 along with Mask R-CNN and k=5 fold stratified cross-validation	71% segmentation accuracy	Very close to 70% in this thesis, but the approach is still not simple
Qiu et al. [11]	2021	U-net architecture	94.1% detection accuracy	Only detects different stages of cancer

2.2 BACKGROUND

One of the concepts newly developed in order to do so is the Convolutional Block Attention Module (CBAM) [12], which is also our target algorithm for this thesis. CBAM works on the image data at basic level by working on "what" and "where" to put attention in the image and give the final diagnosis using a minimal but only important data in the image. CBAM has been previously implemented for the diagnosis of the COVID-19 disease [13], for Automated ECG Classification [14], Change Detection in Radar Images [15], for commercial use in Remote Sensing Images [16], and for classification of CIFAR-100 datasets [17] etc. Figure 3 and Figure 4 from [12] show how the CBAM algorithm works.

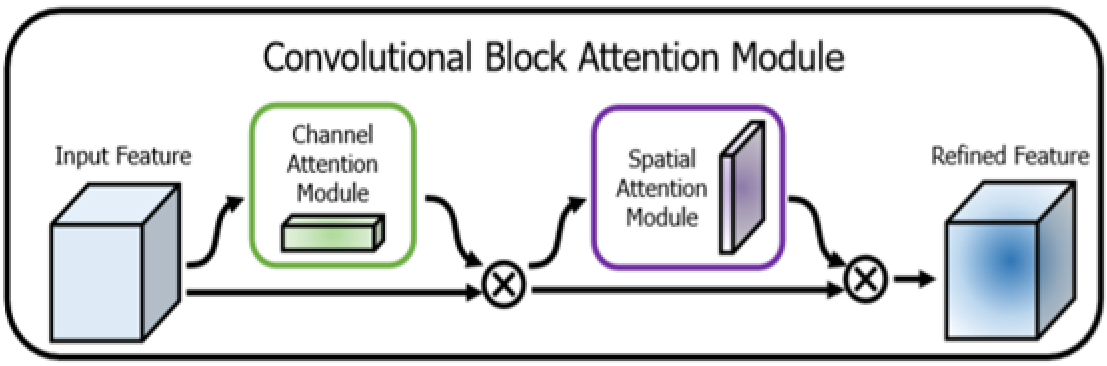

Figure 3: Illustration of how the Convolution Block Attention Module (CBAM) works [12].

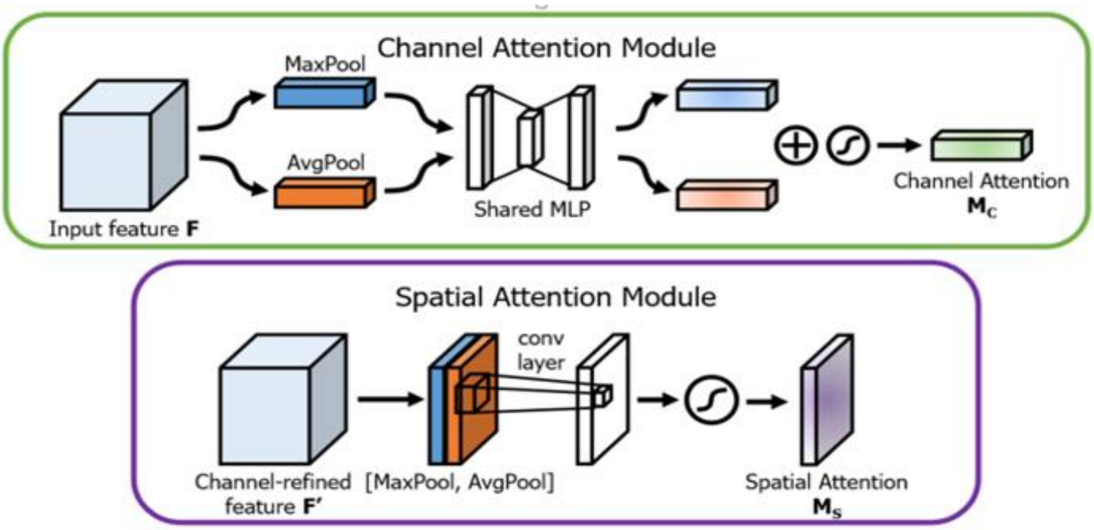

Figure 4: Illustration of how the Channel Attention Module and the Spatial Attention Module of the CBAM works [12].

The CBAM is actually made up of two different modules, the Channel Attention and the Spatial Attention. Both work in series with each other to make the image data more efficient for it perform better during classification and segmentation. Similarly, to further understand the working of the Channel and Spatial Attention, Figure 6 can be observed. Both the sub modules have their own design of CNN layers, which work best for object detection and classification as explained in [12].

CBAM has never been implemented as a stand-alone CNN architecture in the classification and segmentation of GI Cancer Images. All the work that has been mentioned in this section either works using an already built architecture which is readily trained on dataset of images or deals with attention mechanisms which are only used in combination with some other trained models. It has never been experimented whether a simple CBAM module can also work as both a classifier and segmentation network for highly complicated GE images thus reducing the execution complexity of CNNs or not. Also, images normally show better content when dived in multiple channels as happens in CBAM using a certain color space [18]. In [18], the face makeup has been done very efficiently by splitting the procedure in multiple channels of CIELab color space. This research also needs splitting of channels for the Channel Attention module. However, a different color space owing to the requirements of the data complexity will be experimented with and applied.

The algorithms and architectures used prior in the field of GASTROENTROLOGY (GE) for segmentation and classification of different GE images datasets acquired for clinical studies are as follows:

- Normalized Cuts
- Gabor Filters
- CNN
- Feature Aggregation and Attention Fusion
- U-Net manipulation.
- Google's Auto Augment.
- Mask R-CNN
- Simple Linear Iterative Clustering (SLIC)
- Fast and Robust Fuzzy c-Means (FRFCM)
- YOLO V3 and so on.

Notice that these algorithms from different studies have worked on different types of gastric issues. Also, except for the Normalized Cuts segmentation and the Gabor Filters classification studies, all the other studies have been done on a dataset different from the dataset under consideration of this research.

The CBAM architecture used prior in other fields are as follows:

- Covid – 19 diagnosis
- Automated ECG classification
- Segmentation of Expansion of Urban City Structures
- Retention of Information for Channel- Spatial Enhancement
- Remote Sensing Images via satellite imagery and so on.

Notice that the studies above have been made by using different forms of the CBAM. Some have used U-net in combination others have changed the local and global approach, while some have tweaked with the loss functions and the very CNN layers of the CBAM.

2.3 APPROACH OF THIS THESIS

This research work is focused on the application of the CBAM on Gastroenterology Images in the following manner:

- The GI Cancer images are acquired using Chromo endoscopy on different patients.
- The images are generically divided in only two classes thus raising the difficulty level for both classification and segmentation.
- The CNN models for both classification and segmentation are properly trained.
- Classification, followed by segmentation, which is based on different algorithms for the two classes is performed on the test images.
- All the work is done in HSV color space where the image channels are divided into Hue, Saturation and Value channels respectively.
- The values of 83.1% for classification and 64% for segmentation achieved on the same dataset in [3] and [2] respectively should be surpassed.

CHAPTER 3

PROPOSED METHODOLGY

Chapter 3 of this thesis report shows the data and how it has been acquired. The complexity of the GI Cancer images has been made clear in this section via showing different examples from thee two different classes of the GI Cancer images. Furthermore, how the basic algorithm has been developed to work on this data has also been shown.

3.1 INSTRUMENTS AND PROCEDURES

The data has been captured at a research center by the name of Portuguese Institute of Oncology (IPO) Porto, Portugal. The endoscope used is actually a chromoendoscope which uses chromo technology and the Methylene Blue Dye method to illuminate the insides of the stomach properly given to its complicated sac like structure. Figure 5 from [19] shows the endoscope used.

Figure 5: Olympus GIF-H180 Videogastrocopio [19]

The equipment Olympus GIF-H180 has been used at the research center by specialists. The working of how an area in the GI Tract of the human body is observed using such an instrument, consider looking at Figure 6 from [20].

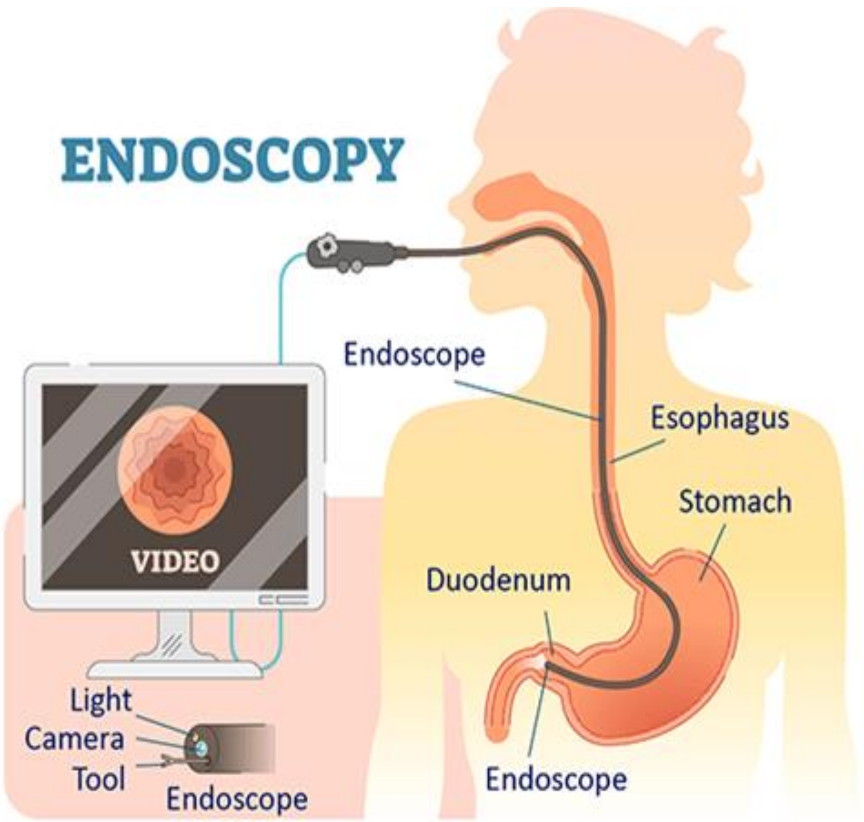

Figure 6: Procedural Diagram of carrying out observation in the stomach of a human body using endoscope [20].

The diagram above shows how difficult it is for an oncologist to actually search for a problematic area in the stomach. The light at the front of the endoscope tube lights up the stomach for a better view, but since the lining of the stomach has glossy finish, it reflects that light, thus increasing the difficulty in gaining the video of the polyps, cancer, or any other stomach ulcer present there. This very reason puts a tough responsibility on the researchers who want to come up with a computer aided algorithm to classify and segment the cancerous area inside stomach as the images are most of the time not very high quality.

3.2 DATA (GASTROINTESTINAL CANCER IMAGES)

The Olympus GIF-H180 endoscope was used to acquire the GI Cancer Images. The endoscope has a 140° field of view with four way i.e. 100° left and right, 90° down and 210° up angulation. 360 000 endoscopic images from 4 hour video tape of real time endoscopy examinations done on 28 patients were obtained [2]. The structure of the human

GI tract is not easy for viewing and diagnosing ulcers, polyps and lesions even via an endoscope. The sac like structure of stomach gives a vast field of view for an oncologist to focus on the relevant region while the limited and narrow tube like structure of intestines make it hard for the oncologists to get into and look for diseases. In this research, stomach cancer images have been utilized. In order to show how complicated and difficult the Normal and Abnormal Cancer Images are to be artificially diagnosed by a CNN architecture, please refer to Figure 7.

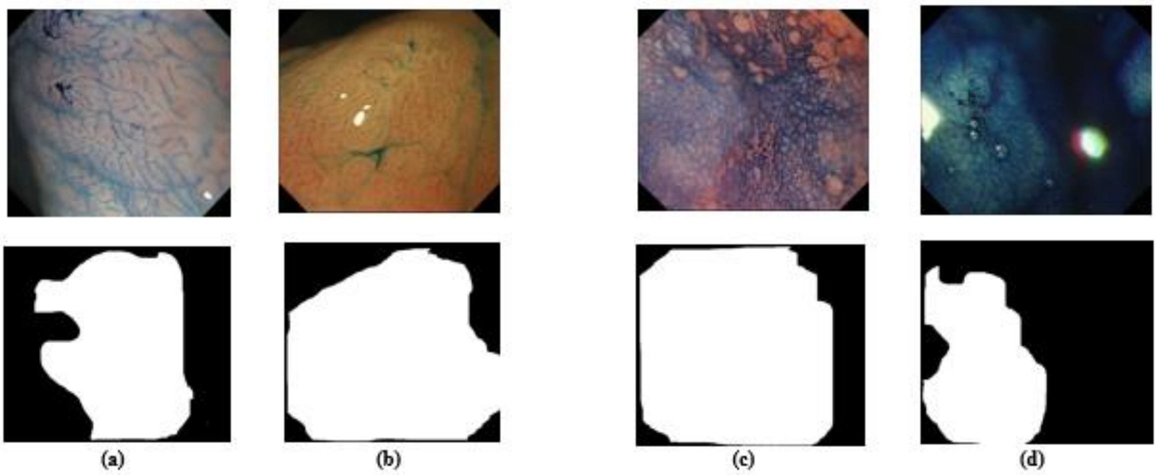

Figure 7: GI Cancer Images with their respective Annotated Ground Truth Images below them. Column (a) and (b) belong to Normal Cancer Class and show early signs of stomach cancer, while column (c) and (d) belong Abnormal Cancer Class which show later and crucial stage of stomach cancer.

Methylene Blue Dye method, which improves the visibility of the gastric mucosa under the illumination of the full visible spectrum of light in CH endoscopy has been used here. A high quality dataset of 176 images was obtained after the joint agreement and high confidence level of two physicians. The physicians have carefully annotated the relevant regions of the GI Cancer Images as well using a provided software as shown in Figure 8 from [2]. After removing the overlapping or redundant images from the dataset, 159 were finalized for experiments. The two groups of images clearly show that mostly, there is no specific target area common in them which can be discovered easily by the CBAM architecture, which ultimately increases the difficulty level of this research. The images have a dimension of 518x481.

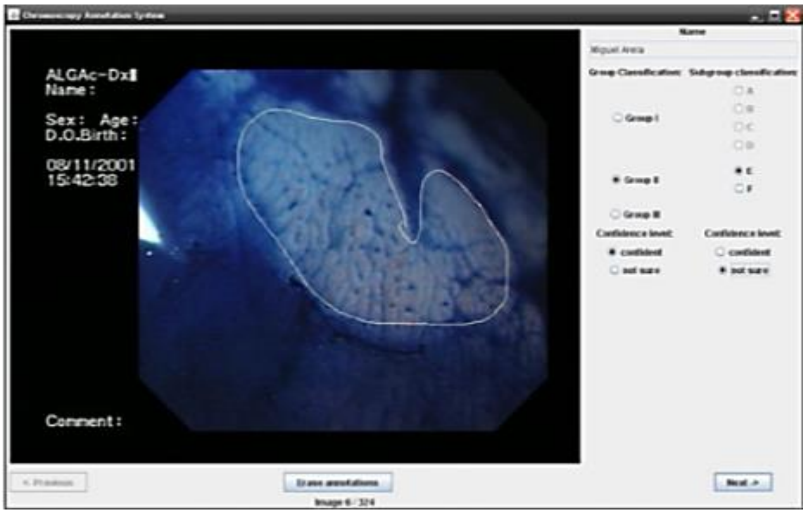

Figure 8: Manual annotation of the region of interest in CH image by an oncologist [2]

3.3 CONVOLUTION BLOCK ATTENTION MODULE IN HSV COLOR SPACE

The CBAM model utilized for both the classification and then the segmentation of the GE images is implemented and works best in the HSV or the Hue, Saturation, Value color space. It has been observed via both the literature studies and the experiments that the HSV always gives better content information in the individual grayscale channels as compared to the other color spaces.

3.3.1 HSV Color Space

The HSV color space or the Hue, Saturation and Value color space has been observed to give best results for the classification and especially the segmentation purposes. As the dataset depicts the complication that the GE images hold, it is evident that the simple RGB approach for the Channel Attention module is not good enough to cater for the underlying complications of the images. The HSV color space has been regarded as ideal when highlighting content of an image, which should be irrespective of its initial RGB color. Take a look at Figure 9 to understand the specialty of the HSV Color space.

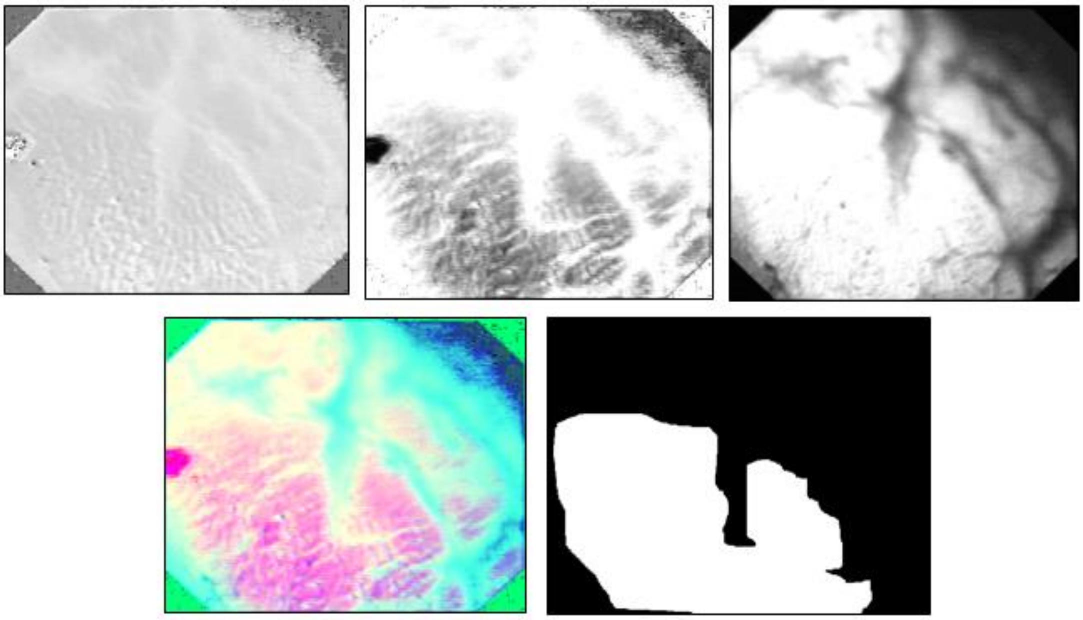

Figure 9: The top three images in the first row represent the Hue, Saturation and Value Channel images of an example Normal Cancer class image. The two images of the bottom row represent the HSV image and the ground truth marked by the oncologists of the same Normal Cancer class image.

Since the images do not have a particular color based lesion or blue vein or anything of the like to be segmented and classified, the RGB or for that matter any other color space was not deemed ideal for carrying out the segmentation and classification procedures. The above figure clearly shows why the HSV color space has been given more value over the other color spaces. The value image, which is the upper rightmost image, gives the same results as the ground truth. Similarly, the HSV image is also very close to the annotated ground truth image. These qualities and features of the HSV color space proved that it should be considered with a high confidence for the research that has been carried out for this thesis.

3.3.2 CBAM in HSV Color Space

The Convolution Block Attention Module (CBAM) is a technique that has been developed by Woo et el. [11], for the purpose of classification and object detection. Here, it has been utilized for both the classification and segmentation of the GI Cancer Images under consideration. CBAM comprises of two more modules namely, Channel Attention

(CA) and Spatial Attention (SA). The CA architecture works on the different channels of the test image, and provides with the important content which can possibly be considered the relevant region of the test image. Hence, the CA works on “what” is important in the image. Since the CA splits the 3-Channel image into 3 separate grey scale images, it makes it easier for the CNN algorithm to find out what part of the image is relevant and what is useless and redundant. On the contrary, the SA architecture works on “where” the relevant information in the Channel modified image is. The GI Cancer test image is fed into the SA module after going through CA modification. Since, it works on the 3 channel image en bloc in the 2D spatial plane as compared to the single channel approach of the CA architecture, thus it renders useful information about the position of the relevant region of the test image in Cartesian plane. It is worth noting here, that experiments have been performed on the 3 separate channels of HSV in the SA architecture as well for finding out the best possible approach for segmentation and classification of the GI Cancer Images. But merging the three channels and forming a 3 channel HSV image worked better for SA initially. It has already been discussed that the provided GI Images are very complicated to be generalized under a single algorithm for classification and segmentation. Some images have an Orange Hue, others have Blue, some have both and some a bit colorless. Given to this challenge, it is very difficult to work in the RGB Color space. After thorough research and consideration, the HSV Color space has been considered for the experimentation using CBAM. The difference in the GI Cancer images from two classes i.e. the Abnormal Class, and the Normal Class, also govern the experimentation done with CBAM layers. See Figure 10 for the flow of the algorithm of CBAM in HSV Color Space.

3.3.3 Modules of CBAM in HSV Color Space

The CA part of CBAM splits the image into 3 channels namely, Hue, Saturation, and Value of the HSV Color space. Then, the CNN layers designed for the CA architecture work deeply on the test image and give a final value which is multiplied and broadcasted with the initial input image for all the three channels separately. These modified images go through their respective SA architectures, where the SA works on the already modified CA images. But as discussed earlier, that after experimentation, 3 channel image, which is formed by merging the output of all the three channel attention HSV outputs is used and fed into the SA architecture, but later, the images were used without any modification and fed directly to SA architecture for classification. The working of the CBAM in the HSV Color space, its CA module and SA module have been shown in Figure 10, Figure 11 and Figure 12 respectively.

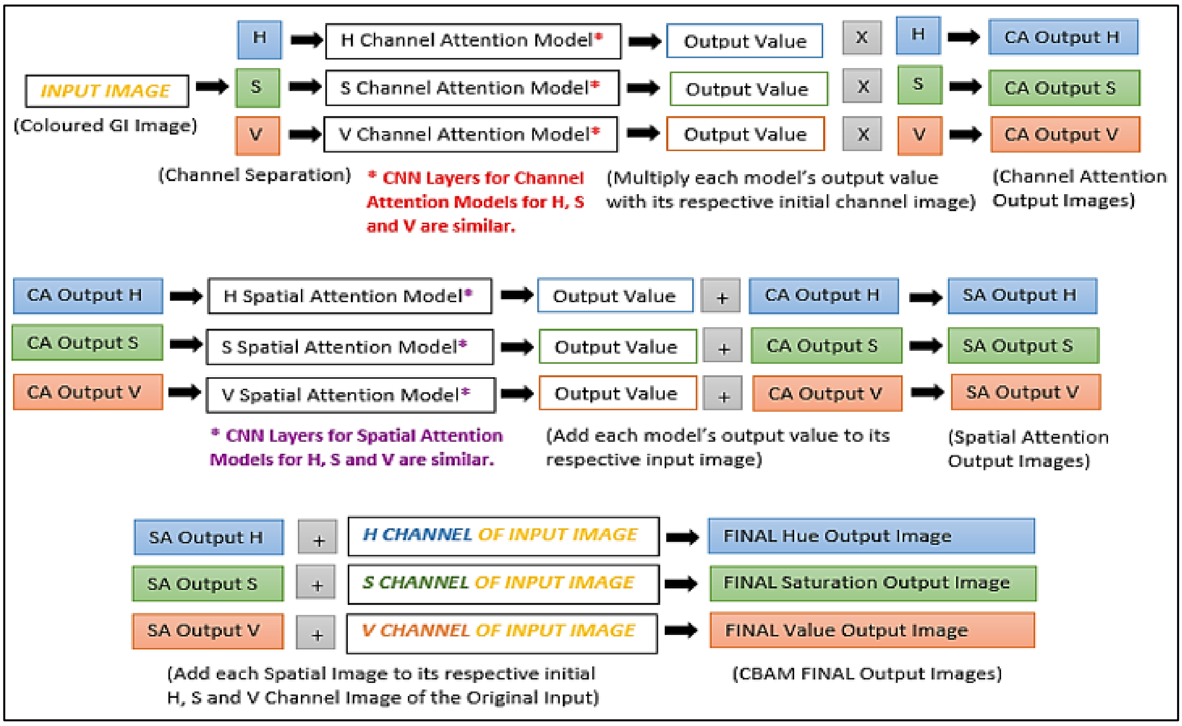

Figure 10: The general flow of CNN based algorithm for Convolution Block Attention Module

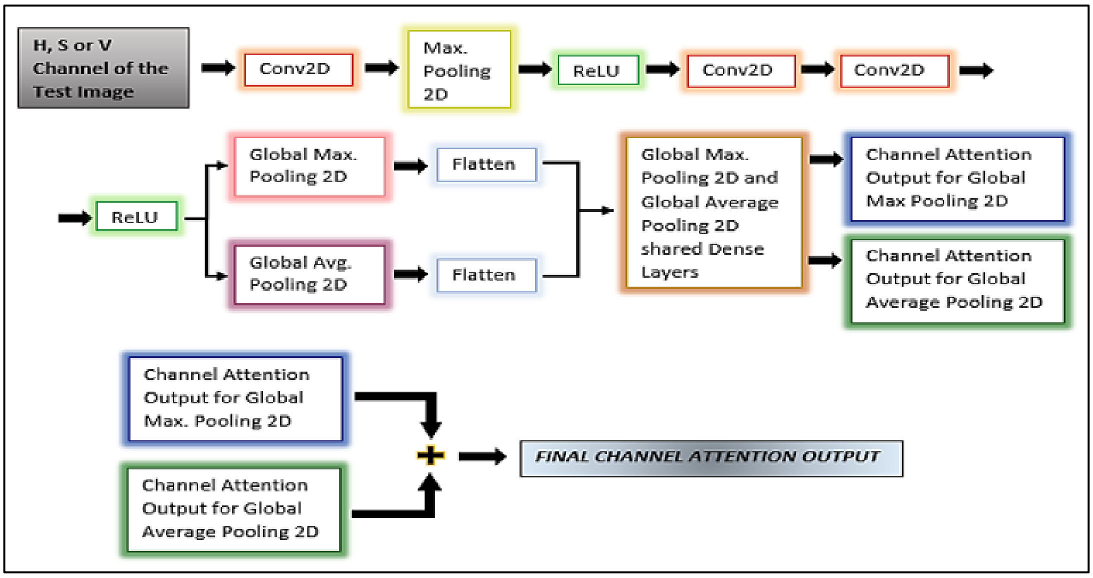

Figure 11: The general flow of CNN based algorithm for Channel Attention Module

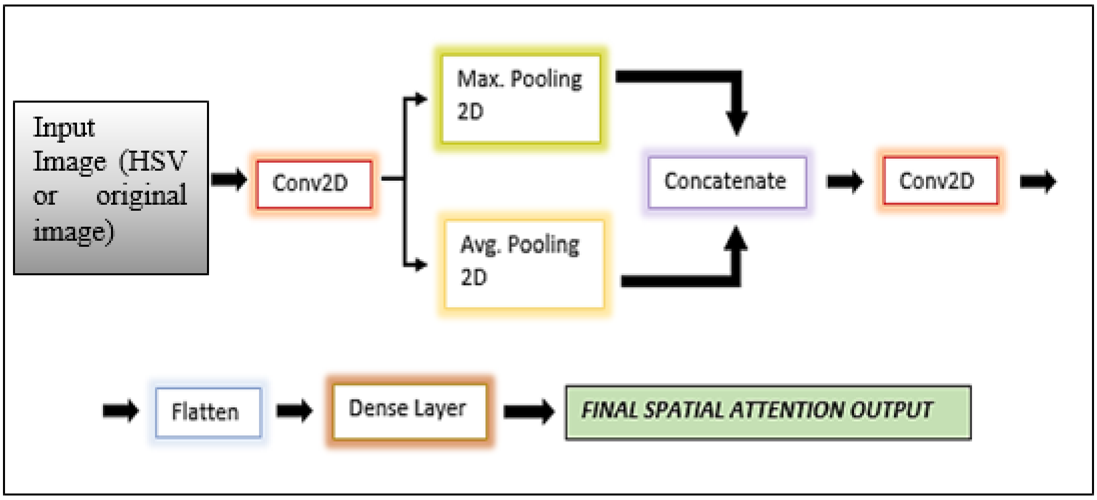

Figure 12: The general flow of CNN based algorithm for Spatial Attention Module

The figures show that the outputs of the CA models are multiplied with their respective inputs while the outputs for the SA models are added back to their respective inputs. This technique is not out of a fluke but the result of thorough experimentation on what out of multiplication and addition would work the best while broadcasting the CNN output values to the whole image for modification. The CA module leading the SA module has also been experimented with. If the SA module works on the images first and then the resultant images are fed into the CA module, the results are not commendable as compared to the results when the order is otherwise. The final form of segmentation and classification algorithm is found after experimenting with these CA and SA models.

No further complication regarding the CNN models has been added to the thesis work. The basic loss and activation functions have been utilized for training and testing of data. The hyper parameters were all continuously tested and trailed before finalizing them for the segmentation and classification results. The HSV color space has been established as a color space, which represents the vision based data in a much better manner as compared to RGB or any other color space given to its attention to the brightness, contrast, and color range available to depict an image. This behavior of the HSV color space makes it very ideal for the segmentation procedures being carried out for clinical purposes. The CA module and the SA module of the CBAM, when working in the HSV color space render very ideal outcomes for this research. It will be explained thoroughly in the upcoming chapters, how well the CA and SA modules work with the data at hand, and the behavior shown by each of them has led to them being used separately for separate purposes and requirements of this thesis. Furthermore, it will be proved using results, that CBAM in HSV is more than enough for the Segmentation and Classification of the GI Cancer images, which are being tested here in this research. After experiments, the Spatial Attention has been utilized separately on the images for classification, while the Channel Attention in the HSV Color Space has been utilized for the segmentation of the ROI in the images.

CHAPTER 4

GI CANCER CLASSIFICATION

The 4th Chapter shows how the classification model has been designed, developed and used based on the CNN architecture. It deals with classification of the GI Cancer Images into two classes of Normal Cancer and Abnormal Cancer.

4.1 CLASSIFICATION IN AI

The process of classification does the job of a human brain by understanding the content of any object, place or person and then being able to categorize them into their respective categories or types. In deep learning, models are built which are then trained on different classes of images placed in different folders in case of Image Processing tasks and after the model has learnt the features of each image type, it is then tested on an outside image to check if it can predict the correct class for this image based on the content and features of the different classes it has learnt about during its training.

In CNNs or Convolutional Neural Networks, the layers of the deep learning models created, include the layer of Convolution as its most important layer which extracts all the required information from the images using convolution masks. Hence the name Convolutional Neural Network. As for the part of Neural Network, it is the final Dense Layers of the CNN model which comprise of the hidden nodes each of which is connected to each of the other in the next layer and take a 1-dimensional feature vector coming from the Flatten Layer of and calculated from the previous layers of CNN. These final layers actually make sure the model learns every feature extracted so far via convolutional masks and refined using multiple other layers like ReLU, MaxPooling2D etc.

Many classification models are readily available on the World Wide Web to be used directly on the available data at hand, the process also known as *Transfer Learning*. These include architectures like the different versions of Inception, ResNet, and VGG etc. In this thesis, however, the CBAM architecture has been utilized. The Spatial Attention module of the CBAM architecture has been utilized for the classification purposes.

4.2 SPATIAL ATTENTION

The ***Spatial Attention*** part of the CBAM, as shown in Figure 12, takes its input initially and then performs further operations on it. The first layer of SA is the 2D Convolution layer with 32 filters of 3x3 kernel size each. After getting the initial features, 2D Max Pooling and 2D Average Pooling layers are used to get separate features. It has been already cleared, how these layers effect the features but this time these two operations are done locally and not globally to get all the possible ROIs in the whole image space possible. After getting the two types i.e. average pooled and max pooled images, they are concatenated together which blends the changes together. This creates an ideal image feature space for the next and final 2D Convolution layer with 32 filters to work on. After getting more feature refinement, the features need to be fed into a hidden network. For that, the Flatten layer is used to turn the features into a 1D vector. With the features fully ready, the final dense layer is used for the network to learn the features and output a predicted value of the test image.

4.3 CLASSIFICATION MODEL

The classification of the GI Cancer images into Normal and Abnormal Cancer classes is performed first. The predicted result decides whether to feed the test image under consideration, either to the segmentation model developed for Normal Class or Abnormal Class. This flow of work was not devised as like that initially. Segmentation, followed by classification of the segmented output technique is implemented at first. But as has been explained in the prior sections, the nature and content of the images provided is very difficult to be generalized under one segmentation algorithm for all. Even within the same class, the images differ from each other while certain images share certain similarities even when they are not from the same class. This complication exists due to the placement of these images into their respective classes based on the amount of tissue damage which has occurred in the stomach of the patient. Images revealing early signs of cancer or normal stomach ulcers have been placed in the Normal category while images with extreme tissue damage in the stomach show almost fourth stage cancer and have been placed in the Abnormal category. But there are certain images which sometimes show marginal characteristics of both the classes and are therefore challenging to be classified and segmented properly. Thus, a classification first approach, which will lead to dedicated segmentation technique separately developed for the two types of images works best. The CBAM technique applied for the classification of the images is based on the technique shown in Figure 12.

Since the classification requires to focus on every detail of every image so that CNN architecture can thoroughly learn the relevant similarities and differences within and between the two classes, thus, Spatial Attention has been implemented for classification. SA takes the whole image without splitting it into multiple channels, which helps the layers understand the images better. The SA model works on the original images without converting them to HSV color space as the CH endoscope generates them in real-time and to make sure that no amount or part of data is lost while performing classification, they are kept in their original form. The 88 training images are divided into 28 for Normal and 60 for Abnormal Class. After extreme and detailed hit and trial method, fine tuning and tweaking with the hyper parameters, the learning rate for training was set to 0.0001, the batch size was set to 4, while the number of epochs was set to 40. It is worth noting here, that all of these values have been thoroughly experimented with before finalizing them. Despite the fact, that the topic directly focuses on saliency mapping using CBAM only, VGG 16 and a custom CNN architecture were experimented with for classification purposes as well. None of them worked better than Spatial Attention architecture implemented for classification of the specific GI Cancer Images provided. Results have been provided in the RESULTS section. TABLE 2 shows the hyper parameters of the classification model.

TABLE 2: HYPERPARAMETERS FOR CLASSIFICATION MODEL

DATA SIZE	88
BATCH SIZE	4
NO. OF EPOCHS	40
LOSS FUNCTION	binary_crossentropy
LEARNING RATE	0.0001

4.3 OTHER ARCHITECTURES USED

The above part shows the approach that this thesis takes in order to get best classification results based on CBAM. But certain other models and architectures have also been utilized for the sake of comparison with the CBAM architecture. The target value from a prior work has already been discussed. The other approaches which have been experimented with are as follows:

4.3.1 VGG 16 Architecture

VGG 16 is a CNN based architecture and is utilized by researchers for different objectives in the field of AI. It is an already built architecture, which has been developed by a group of researchers at Oxford known as the Visual Geometry Group or VGG. It can be easily embedded in a python program and implemented as required. It is used for object detection, classification, segmentation etc. There are different versions of the VGG architecture and it is still being updated by the developers. Here, in this research, the version 16 has been utilized. The concept of transfer learning has been implemented here. It comprises of using a built in architecture of a programming language with the current architecture according to the requirements at hand. The final dense layers were frozen and the normal Sigmoid Layer was used with a Dense Layer of 512 and 256 before it. The VGG 16 is obviously a comparatively larger architecture and does not need to be used as a whole on a small dataset. Following figure from [21] shows the VGG 16 implementation.

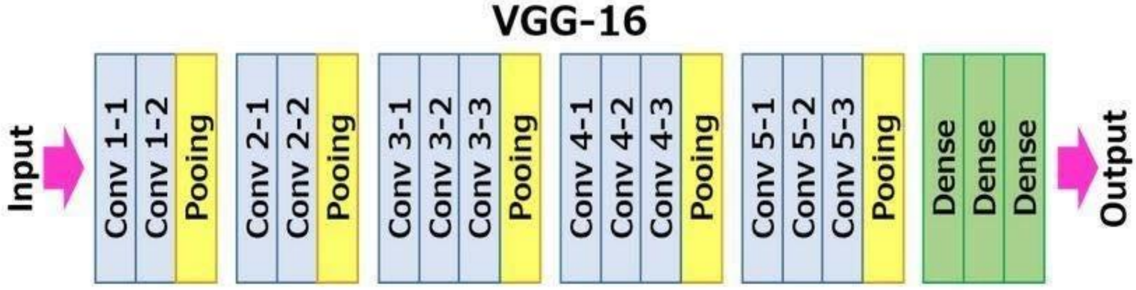

Figure 13: The VGG16 model for classification [21]

The hyper parameters have been set based on the prior research and the amount of data available. The batch size has been set to 4, given to the small dataset, as it gives the architecture room for learning. The learning rate has been set to 0.0001 after thoroughly experimenting with the different values. While the number of epochs, after trying with different values, has been set to 10. TABLE 3 shows them in a tabular form

TABLE 3: HYPERPARAMETERS FOR VGG 16 MODEL

DATA SIZE	88
BATCH SIZE	4
NO. OF EPOCHS	10
LOSS FUNCTION	binary_crossentropy
LEARNING RATE	0.0001

4.3.2 Custom CNN Architecture

A simple CNN architecture has been created for classification purposes as well. The hyper parameters have been tweaked and experimented with to set them at values which would give the best results. The convolution block comprising of a Convolution 2D layer, RELU layer and the Max Pooling layer has been utilized once. Afterwards, the parameters are sent to the dense layers with Sigmoid Layer as the final one. The hyper parameters set have been thoroughly experimented with. This design has taken the longest to get finalized for it to give the best results. A lot of tweaking and trials were carried out for it to work the best. Figure 14 shows the custom CNN architecture.

The hyper parameters have been experimented with thoroughly and only finalized when they started to give best output results. The Batch Size has been set to 4, the number of epochs has been set to 10, while the learning rate has been set to 0.01 here. This custom CNN architecture gave very pleasant results initially, but as compared to the CBAM modules, the accuracy results were a bit lower for this architecture. It is safe to say that no value for hyper parameters, loss functions or activation functions has been done so randomly. All the values for all the architectures have been researched for first and then utilized in the models. TABLE 4 shows them in tabular form.

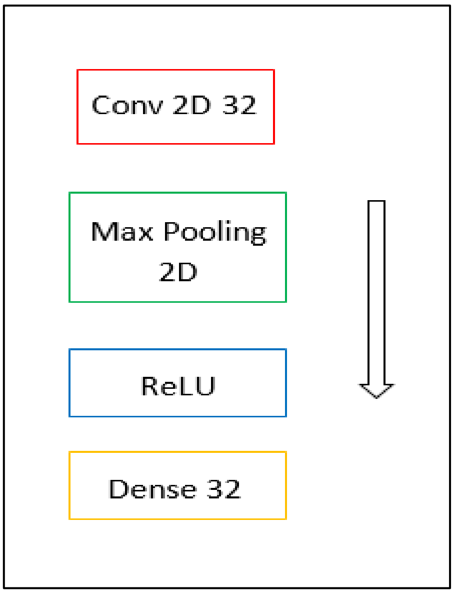

Figure 14: The custom architecture model for classification

TABLE 4: HYPERPARAMETERS FOR CUSTOM MODEL

DATA SIZE	88
BATCH SIZE	4
NO. OF EPOCHS	10
LOSS FUNCTION	binary_crossentropy
LEARNING RATE	0.01

CHAPTER 5

GI CANCER SEGMENTATION

Chapter 5 of this thesis report is a thorough explanation of how the segmentation of the GI Cancer Images has been done. This chapter gives categorical guidance regarding the data collection which includes the GI Cancer Images taken from a real time GI tract exam using a chromo endoscope, the HSV Color Space based work and the final ROI segmentation of the GI cancer images.

5.1 CHANNEL ATTENTION

The ***Channel Attention*** part of CBAM, as shown in Figure 6, comprises of a 2D Convolution layer with 32 filters with a kernel size of 3x3 each, rendering a set of images with refined features. It is followed by a 2D Max Pooling layer with a pool size of (2, 2) and in strides of (2, 2) to help the network get rid of the unrequired data. Then, to rectify the problem of unitary increase of the features learned and the problem of vanishing gradient while training the models, a ReLU Layer is used. To get more refined features, two more layers of 2D Convolution are used afterwards. Now comes the crucial point of Channel Attention where the features go through both a Global Max Pooling 2D and a Global Average Pooling 2D layer. The Max operation sharpens the image while the Average operation smooths it out. This is crucial for Channel Attention to learn the features of the Cancer Images using different approaches. But here the activation of Global Max Pooling 2D layer and Global Average Pooling 2D layer tries to learn the features by taking the maximum value out of the whole image and by taking the overall mean of the whole image respectively. This gives a rather unchanged and almost identical image for the Global Average Pooling while a low value image for the Global Max Pooling since the image has only one channel. Next, Flatten layer has been used to transform the dimension of the feature vector to a single dimensional vector before feeding it to the neurons of the upcoming and final Dense Layers of the network.

5.2 GI CANCER SEGMENTATION

The segmentation of the GI Cancer Images is performed after the classifier predicts the class of the test image. And based on that predicted class, a dedicated algorithm developed specifically based on the content of the two different types of images would be used ultimately for segmentation. Thus, the classification governs the segmentation technique of the test image. Contrary to the classification, Channel Attention works best for segmentation of the relevant region. It should be very clear that the complete process of CBAM was utilized for segmentation. Both the CA followed by the SA module is implemented on the test cases, and it was duly noted that after the images are modified using CA mechanism, the SA does not bring out any more modification in them. The final output of the SA after CA module would always create a very bright and unclear image with an output value of mostly zero. That value even if broadcasted back to the image via addition does not render any modified image. Thus, the segmentation is stopped at the end of the CA module. The images from the given two classes behave differently under a similar segmentation algorithm. It was observed that the initial technique of segmentation gave better results for the Abnormal Class, while the images form the Normal Class gave lower DC Accuracy results. After much experimentation with the HSV Channel Attention algorithm and using the input to be added back to the final results for a much clearer output, two different types of techniques were made for the two different classes. In order to visually show how these two techniques work, refer to Figure 15 for Abnormal and Figure 16 for Normal Class image segmentation.

5.2.1 Abnormal Class

For the ***Abnormal***, all three channel attentive outputs are merged to get a final HSV image. Since the Value channel carries the most important information out of all the three channels, hence, to further clarify the output, the single channel value image which is acquired by splitting the original image into hue, saturation and value channels is added to the final output. This technique was experimented with thoroughly. Hue and saturation channel images were also checked if they could play a role in the output enhancement. Similarly, the HSV output without any further modification was also compared with the ground truth as well. But only, the HSV output image added to v channel of input test image gave the best results. The Figure 9 in Chapter 3 shows the reason why the V channel has been given priority over other channels of the HSV color space. The V channel showed the results closest to the final ground truth and hence has been finalized for the final segmentation procedures. But for the Abnormal Class, only the V channel was not enough. The images required a bit of processing and thus upon further research, the final HSV image, which also gives a very close output to the ground truth, has been considered as well.

This final output is then blurred as a final task and converted to a black and white image for calculating the Dice Coefficient against the provided ground truth. The threshold value at which the HSV image is converted to a black and white image was also experimented with. Since the content of the two different types of images requires two different segmentation techniques, similarly, the threshold value at which the segmented images are converted to black and white also differs for the different images as well as the two classes but an average value is set only.

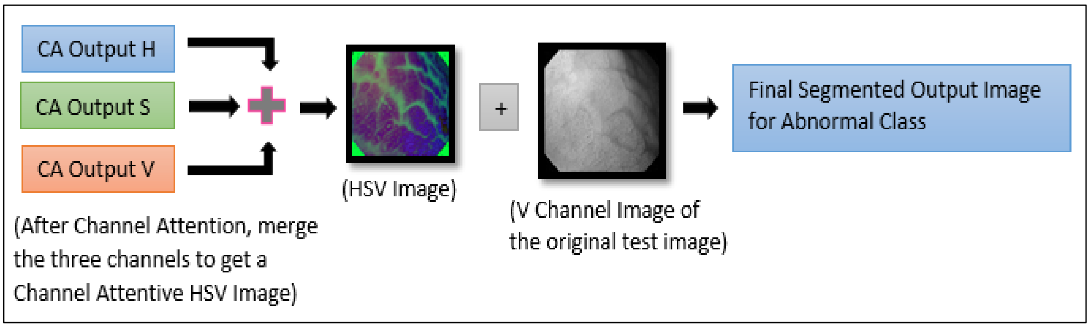

Figure 15: Flow of algorithm for Abnormal Class image segmentation

5.2.2 Normal Class

For the ***Normal***, only the Value channel attentive output has been considered for the final segmentation. Initially, the Normal Class also used the same technique as used for the Abnormal Class, but due to low accuracy of final DC values, the strategy was revitalized and necessary changes were made to get better segmented outputs for Normal Class as well. After comparing, the three channel outputs, the Value channel appeared very close and similar to the provided ground truths, so only the Value Channel Attention was finalized for the Normal Class. Here too, the final output is blurred and then converted to final black and white output. And the value at which the images are thresholded to black and white has been thoroughly experimented with.

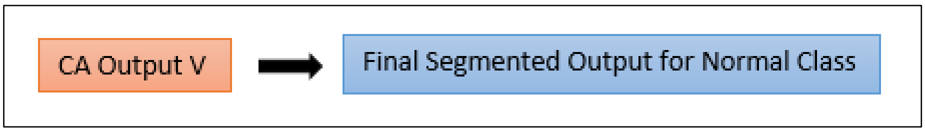

Figure 16: Flow of algorithm for Normal Class image segmentation

Out of the 88 training images, 44 were used to train the H, S and V Channel Attention models, and the rest were used as test images which gave CA outputs to be fed into Spatial Attention as inputs to train on. The Global Average and Global Max layers of the CA are separately trained and then joined at the Hidden Network layer. Both show similar training behavior with similar training accuracy and loss. Despite training and testing the SA model for all three channels, it did not play much role in increasing the accuracy of the final segmented outputs. Nevertheless, this, in itself is one of the great findings of this research. The values for the hyper parameters of the training CA models were experimented with to find the ones at which they perform the best. The number of epochs was finally set to10, learning rate to 0.0001 and the batch size was set to 4 given to the small training dataset.

TABLE 5: HYPERPARAMETERS FOR SEGMENTATION MODEL

DATA SIZE	44
BATCH SIZE	4
NO. OF EPOCHS	10
LOSS FUNCTION	binary_crossentropy
LEARNING RATE	0.0001

5.3 OTHER ARCHITECTURES USED

The above part shows the approach that this thesis takes in order to get the best classification results based on CBAM. But certain other models and architectures have also been utilized for the sake of comparison with the CBAM architecture. The target value from a prior work has already been discussed. The other approaches which have been experimented with are as follows:

5.3.1 U-Net Architecture

The U shaped Network architecture or the U-Net architecture has also been utilized here for the sake of comparison with the CA module of the CBAM model. The following screenshot shows the working of the U-Net architecture:

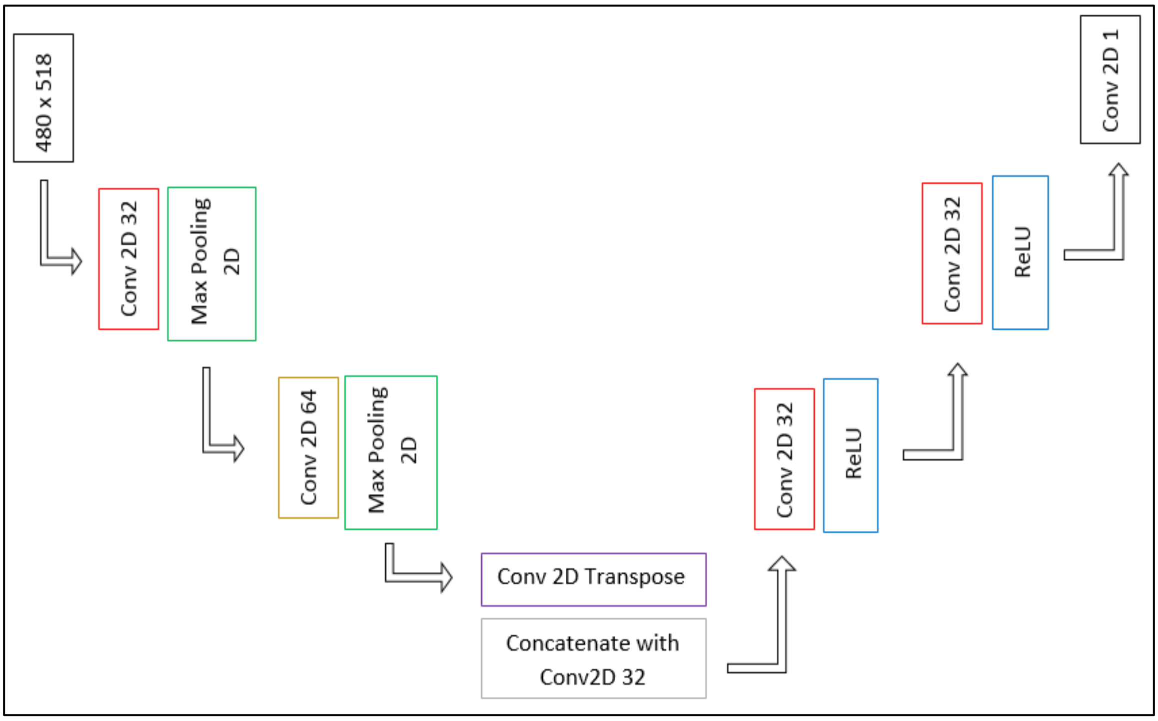

Figure 17: The U-Net architecture based CNN layers

In the above figure, it can be clearly seen that the Conv2D Transpose actually reverses the direction of the architecture and takes a U turn here. The layers following after that redevelop the image again with the segmented output. But despite experimenting and fixing the U net architecture multiple times, it failed to give a proper fully developed and segmented output. No other architecture was experimented with after wards since the Channel Attention module of the CBAM already did a great job at giving a high value of segmentation accuracy as compare to the previous work done on the same dataset. The learning rate has been set to 0.0001 with “binary cross entropy” as the loss function. TABLE 4 shows the hyper parameters that were set for the training.

TABLE 6: HYPERPARAMETERS FOR U-NET MODEL

DATA SIZE	44
BATCH SIZE	4
NO. OF EPOCHS	10
LOSS FUNCTION	binary_crossentropy
LEARNING RATE	0.0001

CHAPTER 6

EXPERIMENTAL RESULTS

Chapter 6 of the thesis report shows the final results based on the experiments and the thorough discussion carried out in the previous chapters. It consists of the most important proofs and validation that sum up the entire research made on the topic of this thesis and concludes the thesis report.

6.1 CLASSIFICATION

The results of all the trained models will be discussed here using the test dataset of the GI Cancer images. The test dataset comprises of 22 images in the Normal Class and 49 images in the Abnormal Class, making a total of 71 test images. The SA model utilized for the classification purpose generates activation map shown in Figure 18. The attention model puts its attention and focus in the relevant regions of the image. Hence, increasing the accuracy for the classification of test images.

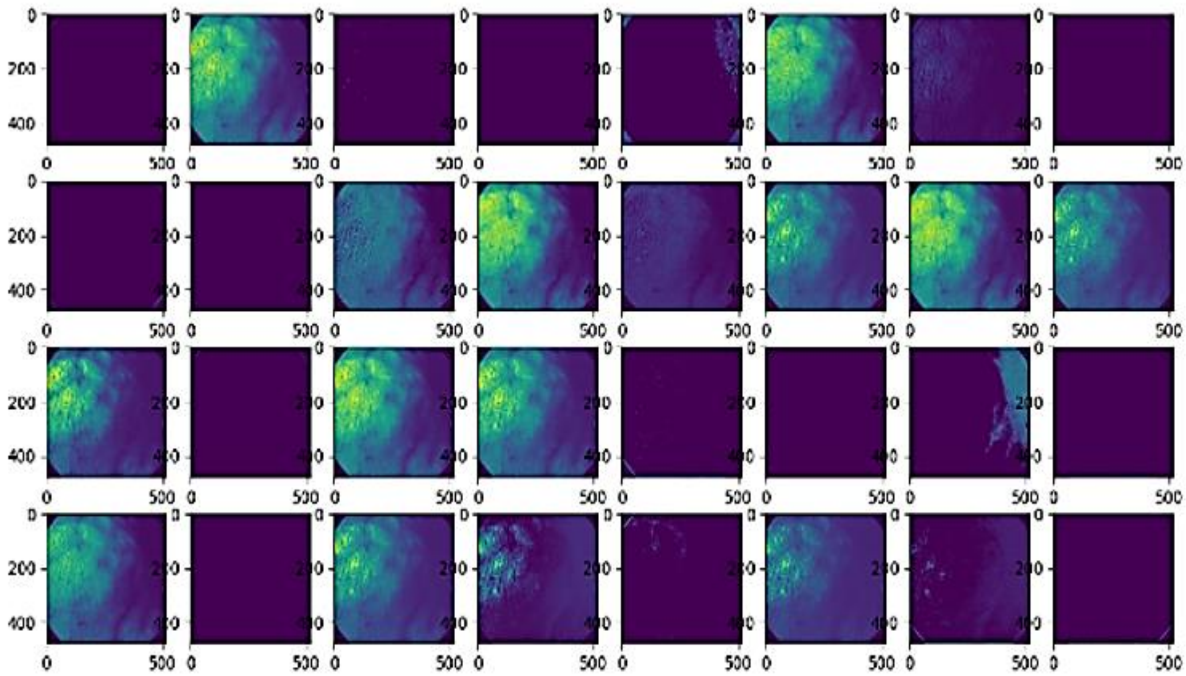

Figure 18: Activation Map of a test image for the classification model

The activation map for an Abnormal Class test image has been shown. The brighter region in the image is actually the relevant region, and the SA model does a great job in highlighting it to differentiate it from a Normal Class image. A total of 71 high quality test images gave a classification accuracy of 92%. Figure 19 shows the training accuracy and training loss of the CBAM Spatial Attention model. The graphs show how well the model has been trained. Both the model accuracy and the model loss for the training, show a very ideal behavior. The accuracy of the graph quickly increase to an ideal value of 1 after almost 12 epochs. Similarly, the loss value quickly decreases to a 0 after almost 4 epochs only. This behavior of the training algorithms show how well the models have been designed. This also proves that the choice of SA module of CBAM is an ideal choice for the classification of the data at hand.

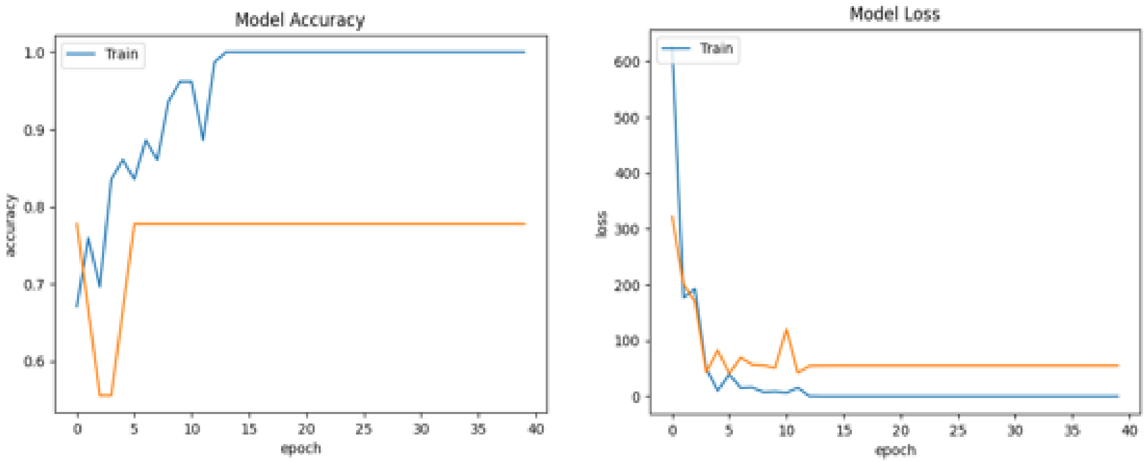

Figure 19: Training accuracy and training loss of the classification model

Before working on only high quality, non-overlapping 71 images, a total of 88 images were initially used for experimenting with certain average quality images as well. VGG 16 architecture, a custom CNN architecture and the SA architecture were all experimented with on that dataset. VGG 16 gave 81% testing accuracy, custom CNN architecture gave 77.27% testing accuracy while the Spatial Attention gave 82% testing accuracy on that testing dataset. This resulted in finalizing the SA architecture for the final phase of classification of Abnormal and Normal high quality dataset. TABLE 2 shows the summary, TABLE 3 shows comparison with the target value from [3]. Confusion matrix has been shown in TABLE 4.

TABLE 7: CLASSIFICATION ACCURACIES FOR DIFFERENT ARCHITECTURES

ARCHITECTURE	**DATASIZE**	**ACCURACY**
Custom CNN	88	77.27%
CBAM, Spatial Att.	88	82%
VGG 16	88	81%
CBAM, Spatial Att.	**71**	**92%**

TABLE 8: CLASSIFICATION COMPARISON WITH THE TARGET VALUE

ARCHITECTURE	**ACCURACY**
Gabor Filters (AHT) [2]	83.1% (Target)
CBAM, Spatial Att.	**92%**

TABLE 9: CONFUSION MATRIX OF THE CLASSIFICATION RESULTS

	Normal	**Abnormal**
Normal	**17**	**5**
Abnormal	**1**	**48**

6.2 SEGMENTATION

For the segmentation results, the same 71 test images dataset has been used. Some of the predicted outputs alongside their original image and respective ground truths have been shown in Figure 20. In comparison to the Channel Attention architecture, the U-Net architecture was also experimented with, but it failed to reconstruct the input images properly with same size and segmented content for comparison with the ground truths. It has a limitation while working asymmetric images. It is very important to mention here that for the threshold value at which the output images are converted to black and white (b & w) for the calculation of Dice Similarity Co-efficient with the ground truths, an average value

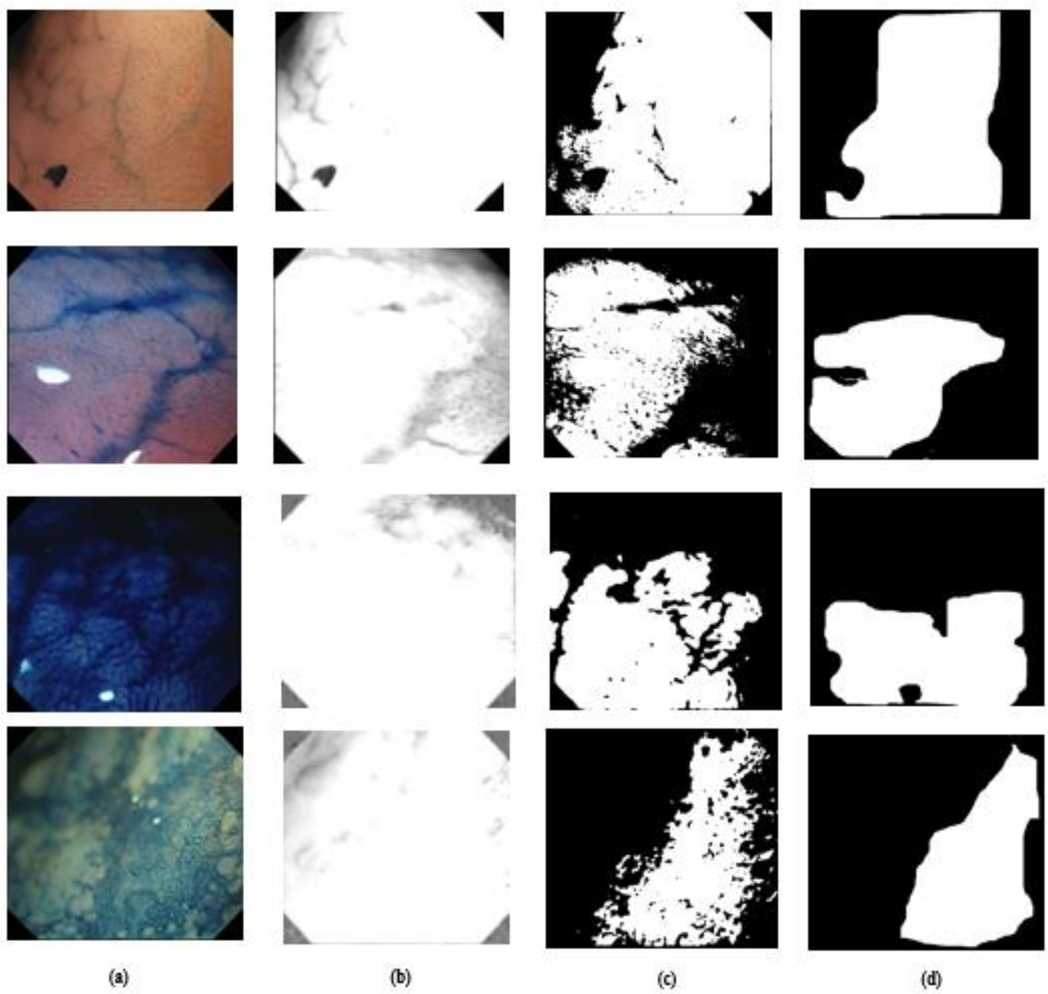

Figure 20: Column (a) represents the input test images, column (b) represents the segmented outputs before black and white thresholding, column (c) represents final segmented black and white outputs, and column (d) represents the ground truths provided by the oncologists. The first two rows belong to the Normal Class and the last two rows belong to the Abnormal Class

at which most of the dataset gave a relatively good result was selected. Certain images gave a better b & w segmented output at a particular threshold value as compared to the other images, but after testing with multiple values, an average value which works for most was selected. For the Abnormal Class, a value of 300 worked for most of the images, and for the Normal Class, a value of 270 worked for most of the images. Keep in mind, that a great amount of addition and multiplication of the image data took place during the segmentation procedure, thus the values ranged from zero to very high numbers in the image content. Also, the HSV color space has the Hue color portion in which numbers range from 0 to 360 degrees. The average value for DC achieved for the 71 GI Cancer images during segmentation is 70%. If the threshold for black and white is not kept constant for a particular class, this segmentation accuracy can increase manifold. This can be taken up as research topic for the future as well. The segmented outputs in column (c) of Fig. 10 show that the technique of thresholding plays a huge role in producing the segmented relevant region of the test images. As the threshold value increases or decreases, it dictates which parts of the grayscale image would make it to the final segmented output. It is therefore necessary to understand that some of these outputs could have shown a much better results if the threshold value was changed for them. Since the aim is to have a single value for the whole class, therefore, certain outputs are compromised by that set threshold value. Nevertheless, a good round about value of 70% for segmentation and 92% for classification has still been achieved by the CBAM algorithm comprising of the Channel and Spatial Attention modules. The summary of the comparison of accuracies with the target value from [2] has been shown down below in the tabular form.

TABLE 10: SEGMENTATION COMPARISON WITH THE TARGET VALUE

ARCHITECTURE	**DSC ACCURACY**
Normalized Cuts [1]	64% (Target)
CBAM, Channel Att.	**70%**

The Tables 7, 8, 9 and 10, give a good summary about the output results for both the classification and segmentation. It is evident that given to the detail oriented nature of the algorithm devised in this thesis based on the CBAM module, the results are much better compared to the prior or the present other architectures. The CBAM has been implemented as a stand-alone architecture in this research. For the most literature study that has been performed on this topic, even if CBAM has been used, it has been done so with the help of another backbone architecture. All the results found based on CBAM in other papers are not just CBAM based, and for this reason, this research is quite novel and the approach devised can be deemed as a positive risk taken confidently in order to meet the needs. The success of this algorithm shows that it is not always necessary to come up with complex architectures to render good results. Also, given to the complexity of the data at hand, it is quite clear that CBAM tweaked according the requirements and implemented in the HSV color space, has done a great job in classifying and segmenting the right outputs.

CHAPTER 7

CONCLUSIONS AND FUTURE WORK

Chapter 7 of the thesis concludes all the work done and gives a brief summary of the previous chapters. It also gives an insight about the future work which can be done in this very field. Both the conclusive remarks and the future experiments possible based off of this same research have been given in this chapter.

7.1 CONCLUSIONS

This overall concludes the work done for research conducted to find out CNN based techniques based on Attention Mechanism to classify and then segment out the ROI in the Normal Cancer, which is the early stage stomach cancer or ulcers and the Abnormal Cancer, which is the later stage cancer with crucial tissue damage in the stomach lining. CH endoscopy images were used for this research and the images were obtained by using the Methylene Blue Dye method for showing clearer images during procedure. It can be confidently established that the objectives of the thesis have been successfully met and a Computer Aided GI Cancer Classification and Segmentation method has been devised for oncologists to use as a helping tool during their procedures. A very high classification accuracy of approximately 92%, as compared to the previous 83.1% accuracy, has been achieved for the classification of the GE images dataset. While a very high DSC Accuracy value of approximately 70% as compared to the 64% of the previous research done on the same data has been achieved for segmentation. The HSV color space has been thoroughly discussed as well. And it has been shown via figures, how the HSV color space gives very close and ideal outputs to the ground truths annotated by the oncologists. Other different models and architectures which have been used on the same dataset and gave a comparative analysis about how much the performance of CBAM as a standalone module is even still better than those models. Even an architecture like a VGG 16 architecture cannot surpass our CBAM based HSV color spaced segmentation and classification results. The authors of this research are very satisfied with the results they have achieved. The research is complete and it proves, given to the shortcomings of the prior work done, that the algorithms designed need not be very complex. A simple approach, which has been devised technically and has a detail oriented approach towards the data at hand, will always deliver good results.

7.2 RESEARCH GAPS COVERED

Here are some of the reasons, why the research conducted for this topic is much more efficient and to the point as compared to the other work done:

- ✓ No back bone architecture has been used as a support.
- ✓ The algorithms used in this research are very simple to understand.
- ✓ No complexity has been introduced in the models.
- ✓ The models were thoroughly experimented with and the hyper parameters were tweaked enough to give good results.
- ✓ No biasness has been introduced in the results.
- ✓ All the data whether it gives a low DSC or a high DSC has been considered.
- ✓ No unnecessary comparisons with irrelevant data and research has been done.
- ✓ All the results of the experiments are reproducible and correct.
- ✓ No CNN based model, which has been used in this research, has been used without prior training.
- ✓ The targeted work in [1] and [2] has used lesser number of images i.e.142, for its research, while from the same dataset, 159 GI Cancer images have been used. This proves more generalization and better risk taking.
- ✓ This research still delivered better results.

7.3 FUTURE WORK

The Chapter 2 of literature review already shows that the CBAM architecture has not only been implemented in this research of segmenting and classifying the GE Cancer images but it has also been implemented before I other fields of studies. Some of them have already been discussed. They are Satellite Imagery, Urban City projects, Covid-19 diagnosis, Automated ECG classification etc. As has been discussed earlier that the thresholding of the grayscale outputs to the black and white outputs for comparison with the ground truths and calculation of DSC, is a very crucial process of the whole procedure of segmentation. The effect of the thresholding played a major role in deciding the final output accuracies. This can be taken up as a new research topic and a novel approach to thresholding can be introduced in the industry. Similarly, the CBAM based model can work very affectively for the surveillance projects. The idea can be implemented in the field of Military Engineering for say, enemy surveillance and enemy recognition systems etc. it can also be implemented as an object detection algorithm for detecting enemy missile systems etc. in a particular area. Although, the implementation in the medical field has been shown in this research, but other than Gastroenterology, it can be implemented in the department of cardiovascular disease detection, and the very channel attentive and detail oriented nature of this research can even be implemented in the field of Physiotherapy as well. The implementations of this thesis can be numerous when counted, however, the current work that has been done so far makes it a successful candidate for future works as well.

REFRENCES

[1] Body, V. (n.d.). Digestive System | Learn Anatomy. https://www.visiblebody.com/learn/digestive

[2] F. Riaz, F. B. Silva, M. Ribeiro, and M. Coimbra, "Impact of visual features on the segmentation of gastroenterology images using normalized cuts," IEEE Transactions on Bio-medical Engineering/IEEE Transactions on Biomedical Engineering, vol. 60, no. 5, pp. 1191–1201, May 2013.

[3] F. Riaz, F. B. Silva, M. D. Ribeiro and M. T. Coimbra, "Invariant Gabor Texture Descriptors for Classification of Gastroenterology Images," in IEEE Transactions on Biomedical Engineering, vol. 59, no. 10, pp. 2893-2904, Oct. 2012.

[4] I. González-Díaz, "DermaKNet: Incorporating the knowledge of dermatologists to convolutional neural networks for skin lesion diagnosis," IEEE Journal of Biomedical and Health Informatics, vol. 23, no. 2, pp. 547–559, Mar. 2019.

[5] S.-A. Lee, H. C. Cho, and H.-C. Cho, "A novel approach for increased convolutional neural network performance in Gastric Cancer classification using endoscopic images," IEEE Access, vol. 9, pp. 51847–51854, Jan. 2021.

[6] Krenzer, Adrian, Amar Hekalo and Frank Puppe. "Endoscopic Detection And Segmentation Of Gastroenterological Diseases With Deep Convolutional Neural Networks." EndoCV@ISBI (2020).

[7] Zhao, Y., Hu, B., Wang, Y., Yin, X., Jiang, Y., and Zhu, X, "Identification of gastric cancer with convolutional neural networks: a systematic review." Multimedia tools and applications vol. 81-8, 2022.

[8] S. Xu, L. Wang, Y. Liu, J. Wu, and H. Huang, "GCLDNet: Gastric cancer lesion detection network combining level feature aggregation and attention feature fusion," Frontiers in Oncology, vol. 12, Aug. 2022, doi: 10.3389/fonc.2022.901475.

[9] L. F. K. Tani, M. Y. K. Tani, and B. Kadri, "Gas-Net: A deep neural network for gastric tumor semantic segmentation," AIMS Bioengineering, vol. 9, no. 3, pp. 266–282, Jan. 2022.

[10] T. Shibata, A. Teramoto, H. Yamada, N. Ohmiya, K. Saito, and H. Fujita, "Automated Detection and Segmentation of Early Gastric Cancer from Endoscopic Images Using Mask R-CNN," Applied Sciences, vol. 10, no. 11, p. 3842, May 2020.

[11] W. Qiu, J. Xie, Y. Shen, J. Xu, and J. Liang, “Endoscopic image recognition method of gastric cancer based on deep learning model,” Expert Systems, vol. 39, no. 3, Aug. 2021.

[12] S. Woo, J. Park, J.-Y. Lee, and I. S. Kweon, “CBAM: Convolutional Block Attention Module,” arXiv (Cornell University), Jan. 2018.

[13] Y. Zhang, X. Zhang, and W. Zhu, “ANC: Attention Network for COVID-19 Explainable diagnosis based on Convolutional Block Attention Module,” Computer Modeling in Engineering & Sciences, vol. 127, no. 3, pp. 1037–1058, Jan. 2021.

[14] J. Wang et al., “Automated ECG classification using a non-local convolutional block attention module,” Computer Methods and Programs in Biomedicine, vol. 203, p. 106006, May 2021.

[15] D. Wang, F. Gao, J. Dong and S. Wang, "Change Detection in Synthetic Aperture Radar Images based on Convolutional Block Attention Module," 2019 10th International Workshop on the Analysis of Multitemporal Remote Sensing Images (MultiTemp), Shanghai, China, 2019,pp. 1-4.

[16] H. Li, K. Qiu, L. Chen, X. Mei, H. Liang, and C. Tao, “SCATTNET: Semantic Segmentation Network with spatial and channel Attention Mechanism for High-Resolution Remote Sensing Images,” IEEE Geoscience and Remote Sensing Letters, vol. 18, no. 5, pp. 905–909, May 2021.

[17] Y. Liu, Z. Shao, and N. Hoffmann, “Global Attention Mechanism: Retain information to enhance Channel-Spatial interactions,” arXiv (Cornell University), Jan. 2021.

[18] D. Guo and T. Sim, “Digital face makeup by example,” 2009 IEEE Conference on Computer Vision and Pattern Recognition, Jun. 2009.

[19] Videogastroscopio Olympus GIF-H180. (n.d.). https://advancemedical.com.mx/productos/videogastroscopio_olympus_gif-h180/

[20] Upper GI endoscopy. (2021, December 6). Johns Hopkins Medicine. https://www.hopkinsmedicine.org/health/treatment-tests-and-therapies/upper-gi-endoscopy

[21] Great Learning. (2022, January 5). Everything you need to know about VGG16 - Great Learning - Medium. Medium. https://medium.com/@mygreatlearning/everything-you-need-to-know-about-vgg16-7315defb5918