Detecting Crohn's Disease from High Resolution Endoscopy Videos: The Thick Data Approach

by

Darien Sawyer

HBSc Bioinformatics, Lakehead University, 2019

Thesis Submitted in Partial Fulfillment of the
Requirements for the Degree of
Master of Computer Science

in the
Department of Computer Science

LAKEHEAD UNIVERSITY
Winter 2023

Declaration of Committee

Name:	**Darien Sawyer**
Degree:	**Master of Computer Science**
Title:	**Detecting Crohn's Disease from High Resolution Endoscopy Videos: The Thick Data Approach**
Committee:	**Chair: Jinan Fiaidhi** Professor, Department of Computer Science
	Sabah Mohammed Co-Supervisor Professor, Department of Computer Science
	Arnold Kim External Examiner NOSM University

Abstract

Detecting diseases in high resolution endoscopy videos can be done in several ways depending on the methodology for detection. One such method that has been a hot topic in the field of medical technology research is the implementation of machine learning techniques to aid in the diagnosis of networks. While, this has been studied extensively with traditional machine learning methods and more recently neural networks, major issues persist in their implementation in everyday health. Among the largest issues is the size of the training data needed to make accurate prediction, as well as the inability to generalize the networks to several disease. We address these issues with a novel approach to detecting Inflammatory bowel diseases, specifically Crohn's disease in endoscopy videos. We use thick data analytics to teach a network to detect heuristics of a disease, not to simply make classifications from images. Using heuristic annotations like bounding boxes and segmentation masks, we train a Siamese neural network to detect video frames for ulcers, polyps, erosions, and erythema with accuracies as high as 87.5% for polyps and 77.5% for ulcers. We then implement this network in a protype frontend that physicians can use to upload videos and receive the processed images in an interactive format. We also pontificate as to how our approach and prototype can be expanded to several diseases with learning of more heuristics.

Keywords: Crohn's Disease, Thick Data Analytics, Siamese Neural Network

Acknowledgements

First and foremost, I would like to express my deepest gratitude to Dr. Jinan Fiaidhi for her outstanding support throughout each step of this process. Even through the limitations of the pandemic, Dr. Fiaidhi was always available at a moments notice for any troubles I may have ran into when working on or writing this thesis, and its completion could not be done without her insight and direction.

I would also like to thank my co-supervisor and internal examiner, Dr. Sabah Mohammed, for his insights during our two years working on this paper, and his critiques of the final thesis. His comments and suggestions to the direction of my work were invaluable additions to Dr. Fiaidhi's critiques whenever we met, and always encouraged reaching out if I ever needed any support.

I am also thankful to the external examiner of this paper, Dr. Arnold Kim, for agreeing to critique the final thesis and defence presentation on short notice, as well as for providing a critical physician's point-of-view to the advantages and shortcomings of the work done for this paper.

Special thanks to my committee as a whole, not only their help on my thesis, but professionally as well through the co-op program. Through the opportunities provided by my committee, I am graduating with more writing and professional programming experience than I could have ever hoped for.

Finally, I must express my special thanks to my fiancée, Rosa, for her unwavering support of my work through the long days and nights it took to complete. Thank you, my love.

Table of Contents

List of Tables

List of Figures

List of Acronyms

IBD	Inflammatory bowel disease
CNN	Convolutional neural network
SNN	Siamese neural network

Chapter 1.

Introduction

1.1. Neural Networks in Medicine

Detecting diseases in high resolution endoscopy videos can be done in several ways depending on the methodology for detection. One such method that has been a hot topic in the field of medical technology research is the implementation of machine learning techniques to aid in the diagnosis of networks, or make diagnosis directly from medical data. One method of machine learning to make diagnoses is by using a neural network. In this case, the detection of a disease is done by the network learning from training data what a positive or negative dataset looks like. In traditional convolutional neural networks, this is done by feeding the network a large number of images and data for substantial learning to be done. This method has several drawbacks however, the main issue being that neural networks require incredibly large amounts of data to train well enough to make reliably accurate predictions. This is because a neural network makes predictions using mathematical computations based on weights and biases updated during training. In contrast, when a physician is looking at endoscopy videos to diagnose a disease, they are not necessarily making mathematical computations, but instead look for indicators that are indicative of a certain disease. We refer to these indicators as heuristics. For example, Ulcerative Colitis is detected by a physician looking for ulcers and cobblestone patterns along the lining of the colon and intestinal tissue. The ulcers and these patterns are the heuristics the physician is looking for during diagnosis. In contrast, a neural network is not necessarily aware that these are the characteristics of a disease like UC. It only looks to make a prediction of diagnosis based on mathematical computations and weights adjusted during training. The practice of adding these heuristics to training data to increase the knowledge gain is a new data enhancing technique we refer to as thick data analytics. We believe the incorporation of heuristics into the training of our neural network will improve prediction accuracies beyond what has been reported without data thickening techniques.

1.2. Identifying Crohn's Disease

The focus of our work is on endoscopy videos and developing a diagnosis tool to make frame-level classifications for diagnoses. In recent years, convolutional neural networks have become the dominant machine learning tool for image processing and analysis due to their ability to reduce computational power needed for classification, and the wide variety of neural network architectures that can be created and customized for any classification task. Convolutional neural networks are often faster for image analysis when compared to earlier machine learning methods because the weights and biases learned during convolutional network training are done automatically. An additional benefit to convolutional networks in image processing is that with several convolutional layers, an image is reduced in size without losing any important feature information. Another important aspect of convolutional networks that reduces spatial size of the image matrix is pooling layers which take the maximum value of a feature in a given pooling space. This reduced the noise or unimportant information in an image while reducing its size further, making for faster analysis. The reduced feature matrix is then used in what are called fully connected layers which perform mathematical computations on the features to output a final classification. The final step in neural networks is the learning step. During training, the classification of an image is given once a classification is made by the network. Given how far or close the output was to the actual classification, the network then propagates this information back through the network, and the weights and biases of the computations are adjusted to make a better prediction on the next training sample. This is performed over many iterations of the data, also known as epochs, and the final network is tested for accuracy on a dataset it has not been given yet. Another advantage of neural networks is the ability to use pre-trained networks through python libraries such as Keras or Tensorflow. The networks available through these libraries are often pre-trained on thousands of images and proven to be very accurate in classification tasks. Many popular architectures such as AlexNet, ResNet, and Inception V3 are readily available, leading to the explosion of research in this field. While these networks have shown excellent capabilities in identifying everyday objects, they have not become widely used in a medical setting due to their focus being on classifying everyday objects. Classifying diseases via endoscopy videos is not within the scope of these pretrained networks, so we will be creating a custom neural network and training it ourselves.

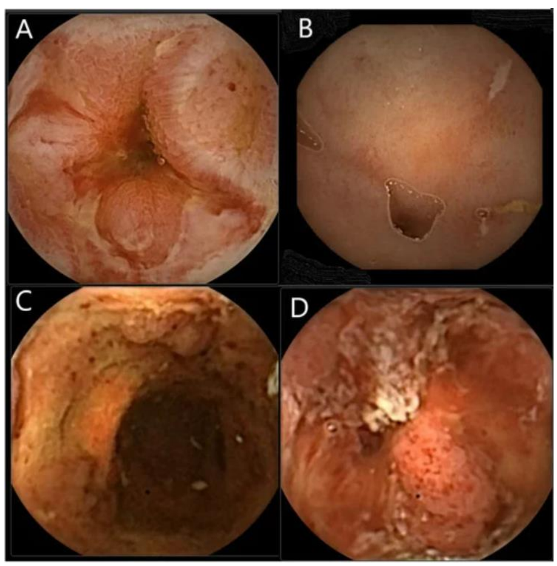

Figure 1: Capsule colonoscopy video frames of patients with IBDs. (A) Crohn's disease with inflammation, (B) Crohn's disease with linear ulcers, (C, D) Ulcerative colitis. Taken from Halder et al. [2]

Inflammatory bowel diseases (IBD) such as Crohn's Disease and Ulcerative colitis are chronic immune-mediated diseases characterized by inflammation of the gastrointestinal tract, and result in stress and decreased quality of life for those afflicted [1]. IBDs are diagnosed using a combination of clinical, endoscopic, and microscopical investigations, with endoscopic techniques being exceptionally helpful in differentiating between different IBDs, as they have characteristic architectural differences that must be identified to differentiate between them. For example, Crohn's disease and ulcerative colitis can both be characterized by the appearance of ulcers, erosions, and erythema. Examples of these two diseases as captured by capsule colonoscopy is shown in figure 1. While the heuristics of these diseases may be similar, Crohn's disease is differentiated by the discontinuous appearance of lesions throughout the intestinal and colon tract, as well as being more commonly found in the terminal ileum of the small intestine [2], an area that capsule colonoscopy has the advantage of capturing compared to other video colonoscopy methods that may not be able to go that far down the intestinal tract. These are the heuristics we will be using to identify Crohn's disease in endoscopy videos and differentiate it from a possible ulcerative colitis misclassification. The neural network used to identify these heuristics will be implemented in a physician frontend prototype to aid in diagnosis.

1.3. Drawbacks of Current Neural Network Approaches

There is a large interest in the medical research community to create a reliable and accurate automatic diagnosis tool that can take medical images and diagnose them efficiently. Over the last several years, with the advancement in image processing and automatic classification capabilities, hospitals and medical companies have been compiling large amounts of endoscopic videos and images in hopes of using the data to train and test a variety of machine learning techniques in automatic diagnosis. These techniques in the early stages of this research included support vector machine, k-nearest neighbor, and linear discriminant analysis [3,4,5], but have since fallen away to the noted benefits of convolutional neural networks in image processing. However, neural networks are not without their drawbacks.

There are several reasons why neural networks are not yet widely used in a medical setting. One such reason is the availability of data needed for training and testing. Neural networks generally require large amounts of training data for the network to properly learn how to classify images. This may not be an issue if you are working on classifying everyday objects thanks to the wide availability of images on the internet and compiled datasets like ImageNet, however, medical images are often difficult to get a hold of due to privacy and security issues. Medical data may also be unavailable to researchers due to the fact that the disease of interest is not one that is widely studied or publicly documented. For example, the recent Covid-19 pandemic exhibited the need for an automatic diagnosis tool in the early stages of the pandemic, but using traditional neural networks to design such a tool would require large amounts of lung and CT scans already classified for Covid-19. These types of datasets were largely unavailable in the beginning months of the pandemic when they would have been most useful.

Another reason neural networks may not be reliable in a medical setting is their reliance on big data to properly train for classification. reliance on big data has shown to fail in identifying actionable insights when it was once thought to be extremely useful. This is a common issue in social research such as customer insights through social media [6], but it also extends to the medical setting, where patient insights derived purely from social media often do not lead to useful insights. Instead, it has been found that this information becomes more relevant when qualitative information about the patient is included in the dataset and analysis.[7]. This is because, when only using big data or quantitative data, the context provided by qualitative data is stripped away and meaningful insights become difficult to derive [8]. This idea of integrating qualitative data with quantitative big data has not only been proposed in a social setting, but has also shown promising preliminary results when extended to medical images [9, 10]. It is for these reasons that we must look for additional preprocessing methods and neural network architecture to design medical diagnosis tools that can make accurate and reliable classifications on very small samples of data.

1.4. The Thick Data Approach

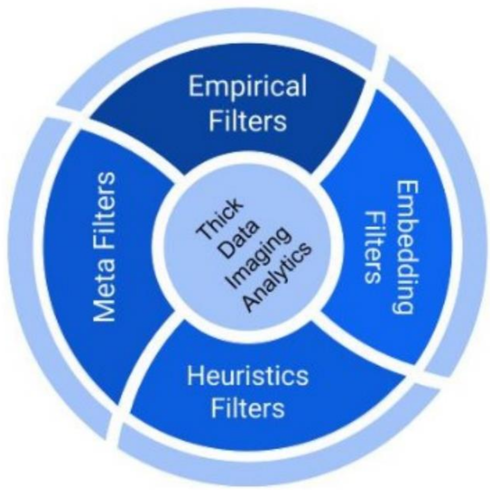

Figure 2: Data thickening techniques for images. From Taken from Fiaidhi et al. [10].

The process of adding qualitative data to quantitative data is known as thickening, or thick data analytics, and is a process coming to popularity in the data analytics world where big data has largely failed to derive meaningful insights, such as with customer sentiment or optimized workflow practices. In these circumstances, data thickening techniques can include the integration of geographic and ethnographic information, as well as adding context to a set of information by comparing it to similar examples of that set to make predictions. Thick data analytics have also been proposed to be used in a medical setting via patient sentiment analysis. techniques for thick data analytics for patient insights include social media data thickened with acquisition analytics, segmentation, retention, engagement analytics, and churn analytics [7]. In our case, data thickening techniques can be implemented by annotating images based on heuristics that a trained physician would look for when making a diagnosis. Annotations of this kind can be done using bounding boxes, segmentation masks, and key points marking. Annotation for thickening data sets can also include other filter techniques as shown in figure 2.

We aim to extend this process of thick data analytics to medical image diagnosis via neural networks. Relevant work on blending contextual data with medical images is largely neglected beyond work in feature extraction, however we propose several data thickening techniques for adding contextual data. These methods include filtering frames based on anatomical landmarks, using meta filters to apply color space augmentations, specifically contrast and center cropping, and using heuristic filters to apply bounding boxes and segmentation masks on the images based on training data annotated by a trained endoscopist. The resulting trained network and the networks used to augment the images for heuristics are then implemented in a physician frontend prototype. This prototype takes in a capsule colonoscopy video and cuts it into frames. The frames are then clustered based on similarity to each other to identify similar regions within the video, then these clustered regions of frames are fed through the trained Siamese neural network to identify for heuristics of Crohn's disease and possible misclassification markers. All information gained and calculated during processing is given back to the user in an easy to digest format so that a physician end user can make the most informed decision possible based on our diagnosis tool, a benefit of our network that is not inherent in traditional neural network approaches.

In this thesis, we will be examining the effects of thick data analytics on several types of medical images and diseases, starting with Covid-19 in CT scans, then Crohn's disease in high resolution endoscopy videos at the frame level. We aim to answer the question as to whether using thick data analytics can improve current neural network image classification techniques, with a focus on medical images, and the goal of improving overall classification accuracy while simultaneously decreasing the training sample data size need to achieve high accuracies.

Chapter 2.

Exploring Thick Data Techniques in Medical Imaging

2.1. Single and Multi Disease Classification using Neural Networks

Since the success of breakthrough image classification using neural networks, many attempts to implement neural networks for diagnosing medical images have been made. One field of medical imaging widely researched is automatic classification of CT scans and X-ray images. Neural networks have been used to try to diagnose several diseases using these types of images such as pneumonia [11], lung cancer [12], SARS [13], and brain lesions and Alzheimer’s [14]. In these studies, they achieve accuracies ranging from 76% to 88%, however, the issue is that these networks become fine tuned to the one or two diseases they are trained on. It is not tested as to whether these networks are able to differentiate between a disease that may be similar. For example, pneumonia, SARS, and Covid-19 all have similar anatomical landmarks in lung CT scans, but without being trained on all three diseases, using the neural networks in the above-mentioned studies runs the risk of misclassification due to similarities in anatomical landmarks.

To combat this issue of misclassification, several studies have been made to attempt to use neural networks to diagnose a wide range of diseases at a time. One such study was conducted by Wang, Jia, Lu and Xia where an custom neural network model was trained using the ChestX-ray14 dataset containing 112,120 images across 14 classes of thoracic diseases [15]. They managed to achieve accuracies and AUCs that outperformed leading literature in all 14 classes of disease. Another study which focuses on endoscopic images was performed by Lonseko et al. [16] where 12,147 images across 4 datasets were used for testing and training on a ResNet, and an accuracy of 93.19% was achieved. These studies create a different issue in terms of general usability in that the amount of training data used to get high accuracies is quite large. As mentioned previously, datasets for many diseases are widely unavailable, making it difficult to use these frameworks for other diseases not included in their datasets.

The body of work done specifically on identifying Crohn's disease from endoscopy images is largely based on training a deep convolutional neural network to identify Crohn's disease based on few identifiers. For example, a study conducted by Klang et al. [17] trained a deep neural network using 27,892 images. Their resulting network achieved an average accuracy of 93.5%. Another study conducted by Ruan et al. [18] performed training on a deep neural network on 49,154 images and achieved an average accuracy of 99.1%. Most papers in this topic follow the same methodology with differing neural network architecture and achieve comparable results. However, despite the great reported accuracy, an issue with these studies is that they require large sets of images to train to the high accuracy they report. Because of this, training, testing, and implementing the same network to a variety of diseases would fail due to limitations in datasets. An alternative approach which we focus on is in identifying heuristics of these diseases to aid in diagnosis rather than direct classification for a specific disease. In this regard, there are some studies which use neural networks to generalize an approach to IBD identification by classifying heuristics such as ulcers and bleeding. One such study was conducted by Sindhu and Valsan [19] which used heuristics like bleeding, polyps, and tumors to identify the presence of IBDs in general. They achieved an overall accuracy of 88.6% and used a feature extraction technique called key points detection to detect their heuristics. This was also achieved using a dataset of 1385 images. These are the results and dataset sizes we are aiming to be comparable to, however this approach again has the issue inherent with traditional neural networks. Their results and classification accuracies are output by the network with no context as to how the network reached its conclusion. For neural networks to be widely accepted in medical use, human-readable interpretations of how they reach their classification would be greatly appreciated and is an inherent benefit to our approach: the Siamese neural network. However, this study and others like it are on the right path for identifying diseases, that is, to use heuristics rather than big data.

2.2. Feature Extraction and Heuristic Identification

As we can see, research into direct classification of diseases based on images has its downfalls. As such several studies have been focusing on using neural networks not as a means of classification but for feature extraction instead. Several examples of extracting relevant information from images to enhance physician diagnosis include texture annotations of colonoscopy images [20], Improving image quality of ultrasound images [21], segmentations of areas of interest [22, 23], and bounding box annotations [24]. These studies propose several methods of using neural networks to make annotations directly on images or provide corresponding segmentation masks for the physician to make the diagnosis process more efficient. While these studies cannot provide accuracy measurements of their outputs like the studies which focus on direct classification, the ideas presented provide a more reasonable approach to implementing neural networks in medical imaging. This is because the output of suspected features of interest still requires that an expert physician verify the results during diagnosis. In this way, there is less risk of misdiagnosis by the network itself as it is just providing a diagnosis enhancing mechanism. Our method of identifying Crohn’s disease from endoscopy videos blends both diagnosis methods and feature extraction methods in a novel way.

2.3. Siamese Neural Networks and Thick Data

In order to implement our novel approach, we must use Siamese neural network architecture which has been shown to have success with basic classification tasks as well as with medical images. One of the most popular papers on Siamese neural network implementation is by Koch, Zemel and Salakhutdinov [25] who achieved an accuracy of 92% on alphabet classification of the Omniglot dataset, and since then Siamese neural networks have been used in a variety of classification tasks that require the comparison of similar black and white images, such as signature detection [26] and writing styles [27]. Since these earlier implementations, Siamese neural networks have expanded into more complex classification tasks like facial detection [28], as well as minimal experimentation on medical images like x-rays [9]. Exhibited by these papers is the Siamese neural network's unique ability to learn and make accurate classification after training on very few data samples, and in some cases like the Koch et al. [25] study, training can be conducted during testing using a technique called on-shot learning. This unique characteristic of this network architecture is well suited to leverage the advantages of thick data analytic techniques.

In the past, thick data techniques have largely been proposed and implemented in the social aspect of data analytics, such as customer insights from social media data [6] or improving the healthcare community of practice [29]. More recently, work on blending thick data analytics and medical image classification has been expanding, with many works in this field being spearheaded by Dr. Jinan Fiaidhi with a focus on using thick data techniques to identify and evaluate IBDs like Crohn's disease and ulcerative colitis. Many of these studies involve using thick data techniques to analyze severity or variability of IBDs. Our work in this thesis aims to expand upon the growing body of research concerned with identification methods using thick data techniques, mainly how we can inject physician-based heuristic identification into neural network classification to improve modern automatic diagnostic capabilities.

2.4. Identifying COVID-19 Using a Thick Data Approach

The disease we focused on at the beginning of our research was diagnosing Covid-19 from lung x ray images and CT scans. Covid-19 in these images is characterised by localized areas of white hazy spots. It is important to make the distinction between signs of Covid-19 and other pneumonia-like diseases like SARS which also can be diagnosed by the appearance of white hazy spots, but also with areas of white hazy areas covering large parts of one side of the lung. We focus more on identifying lung diseases such as Covid in images rather than detecting a distinction between lung diseases if present. Perhaps future work could involve improving the network to identify between, negative images, Covid positive images, SARS images, etc.

We use a convolutional neural network to make annotations after being trained with annotated and ground truth images. This network will be trained with a dataset that contains sets of images with the original CT-scan and the associated annotated image made by a physician. After training, the network will be able to identify areas of interest in unedited images based on how the physician has annotated the training image. To accomplish this, we use a ResNet neural network containing U-Net architecture. The U-net architecture allows us to feed in a ground truth image along with the annotated mask. The ResNet then predicts where to make annotations, and the result is compared with the annotated mask. The accuracy of the ResNet annotations is then calculated, and the predicted and actual results are then back-propagated through the U-net so that it is learned where the annotations should be based on pixel value. After training, the testing results appear to be comparable to that of the physician annotated images, although without the presence of a physician, it is not clear whether our results are 100% accurate.

We train the segmentation network with 100 Covid-19 CT-scans that were annotated with different colorings for ground-glass opacities, consolidation, and pleural effusion [30]. By training the ResNet neural network on these axial CT images and masks, we can annotate any CT scan image of the lung that we have in our data set. The result of these annotations is shown in figure 3. The dataset used to train the Siamese neural network and the pre-trained networks is a collection of CT lung scan images with 349 images being classified as positive for COVID-19, and 397 being classified as negative [31].

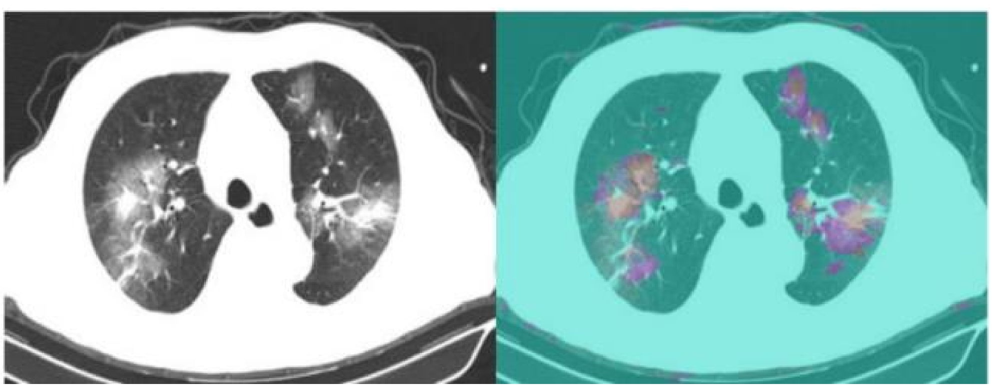

Figure 3: Segmentation of a CT Scan Image Classified as Positive for COVID19. (Left: input image. Right: output image, background is colored blue, suspected consolidation is colored purple, suspected ground-glass opacities are colored orange.)

In order to make predictions of diagnosis of lung diseases, we must pivot from the U-net ResNet to a Siamese neural network so that we can feed in two images at once. We have developed a Siamese neural network using components from the most popular image processing networks of the last decade, namely AlexNet, Inception V3, GoogLenet, and ResNet. The full architecture of our Siamese neural network for Covid-19 detection is given in figure 4.

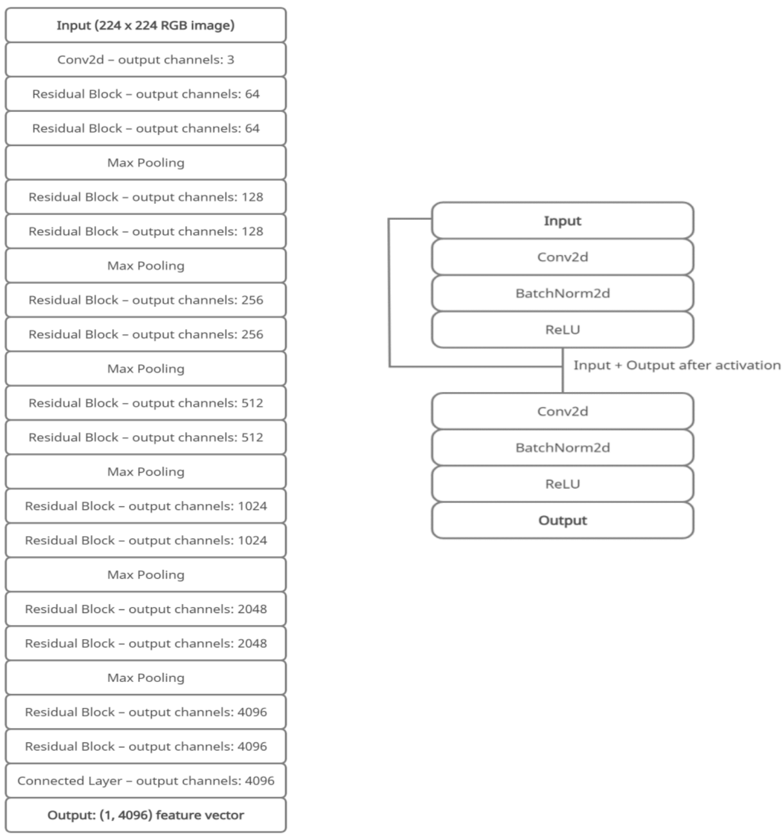

Figure 4: Base Convolutional network architecture for our Siamese neural network detecting Covid-19, and the associated residual block layer architecture.

These subsequent neural networks are then also tested against AlexNet, Inception V3, GoogLenet and ResNet using the pre-trained neural networks from the AI library Keras. These pretrained networks are then further trained with the full annotated CT scan dataset for 10 epochs, while the sequential Siamese neural network is trained using 64 pairs of images for 1 epoch. The reason we are training the Siamese network less is because part of the appeal of Siamese neural networks is the decreased need for training, therefore we are testing to see if the results reflect that notion. Testing on all networks is done using 10 random images, the only difference is that the pre-trained networks are making direct predictions of a single image, while the Siamese network is performing 2-way testing on the 10 images. The results find that AlexNet had an accuracy of 51.58%, ResNet had an accuracy of 51.47%, VGG had an accuracy of 51.68%, GoogLenet had an accuracy of 51.05%, Inception V3 had an accuracy of 51.68%, and our sequential model had an accuracy of 54.13%. From these results, we see that the Siamese architecture combined with the thick data heuristics improved accuracy by 3%, meaning that Siamese architecture takes advantage of heuristics better than traditional architecture. We also suspect that adding more heuristics can further improve accuracy.

After the testing and getting results of our ResNet annotated images, we turned our attention to more image augmentation and annotation options to explore which annotations might be most useful for thickening an image data set. Thus, we begin augmenting images that are suspected to either enhance the characteristic hazy grey areas indicating ground glass opacities or remove noise and redundant data from the images. Namely, the augmentations we test are center cropping which removes black redundant pixels from the image, contrast which increases the difference between dark and light pixels, random erasing which removes possibly redundant data (with the risk of randomly removing vital data), sharpen kernel which will better outline possible areas of interest, and bounding boxes which is another annotation of heuristics like the coloring in the first set of tests. After annotating images, testing was done using the Siamese neural network, and 5 tests were conducted per augmentation including a set of unaltered images. It was found that on average, the no filters category had an accuracy of 40%, center cropping had an accuracy of 44%, contrast had an accuracy of 68%, random erasing had an accuracy of 60%, sharpen kernel had an accuracy of 52%, and bounding boxes had an accuracy of 64%. From these results, we can conclude that CT scan images are best thickened using contrast and bounding boxes above all other augmentations, and at the very least, adding augmentations of any type performed better than classifying images with no augmentations whatsoever.

Similar work was also conducted by Fiaidhi, Zezos and Mohammed [10] to classify ulcerative colitis from endoscopic capsule video frames. The network used in this research was a custom convolutional neural network trained on ulcerative colitis frames classified by an expert physician. Instead of heuristic filters, meta filters were used during training and classification of severity, namely horizontal and vertical flips, rotations, and zooming. The accuracy of the SNN to classify the severity of a given frame was found to be 76.9%

Based on our exploration of thick data techniques, we found that Siamese neural networks take advantage of data thickening techniques when compared to traditional neural network architecture. We also found that contrast and heuristic annotations proved to add the most meaningful data thickening aspects. Based on these findings in thick data approaches, we shift the focus of this thesis to identifying Crohn’s disease from high resolution endoscopy videos using meta filters, heuristic filters, and a SNN.

Chapter 3.

Thick Data using Siamese Neural Network

3.1. Datasets

The main dataset we use for training and testing is the Kvasir-Capsule Dataset [32] which consists of 117 videos collected from examinations at hospitals in Norway, which can be used to extract 4,741,504 image frames. 47,238 frames have been labeled with bounding boxes and verified to detect anomalies from 14 different classes of findings. The labeled images we use are the images labeled for ulcers, polyps, erosions, and erythema. These images are used to train our bounding box prediction network as well as to train and test our Siamese neural network. To implement the segmentation mask heuristic filter like we have in our study focusing on Covid-19, we also utilize a similar dataset: the HyperKvasir Dataset [33]. This dataset contains 110,079 images and 374 videos collected from examinations at the Bærum Hospital in Norway. 1000 of these images from the polyp class have corresponding segmentation masks annotated by at least one experienced gastrointestinal endoscopists. These are the images we are interested in using for the color mask heuristic filter, as the Kvasir-Capsule Dataset only contains bounding box annotations.

3.2. Developing the Siamese Neural Network

Traditional convolutional neural networks have shown great success in the use of image classification and detection tasks with large training data sets due to their computational power and learning capabilities. However, CNNs fail in being able to produce reliable results when the dataset they are training from is small. For example, as exhibited in the study by Brigato and Iocch [34], this is largely due to the complexity of modern neural networks. As complexity increases, so do the trainable parameters with the neural network, requiring more data samples to effectively fine tune the needed parameters to a suitable degree. Instead, we can use Siamese neural networks to make classifications by calculating the similarity between two images. Siamese neural networks do not fall to this complexity issue with small training samples because the classification is made using an embedding of the image and a similarity calculator such as cosine similarity. Since the embedding layers of a neural network do not require as much fine tuning as the fully connected layers, we are able to take the embeddings and make reliable predictions with little training.

In order to create a machine learning process that is extremely accurate using a very small sample set, we must leverage the concept of similarity in the classification process. For example, a human can identify the difference between an image of a cat and an image of a dog even after only seeing the animals very few times. This is because a comparison of the animals together clearly outlines that they are not similar. This concept can be used in the classification of images in machine learning, to diagnose diseases when the training dataset is very small, however, to do so requires us to use a Siamese neural network rather than a traditional CNN.

Siamese neural networks are important for calculating similarity between two classes because, unlike CNNs, the product of the twin networks are features vectors rather than a direct classification result. These two vectors can then be compared using some vector distance calculation (e.g., Euclidean or Manhattan), and classifications can be made. This process of comparing vector distances to make predictions rather than relying on the network itself to make predictions makes it possible to use machine learning to classify images with very small datasets, a task that would not produce very reliable results is using a traditional CNN.

There are several loss functions that can be used to train neural networks, such as squared-error loss and contrastive loss. When using Siamese neural networks, we decided to use the Triplet-Loss function during training, given by the following equation:

$$Loss = \sum_{i=1}^{N} \left[\| f_i^a - f_i^p \|_2^2 - \| f_i^a - f_i^n \|_2^2 + \alpha \right]_+$$

This is because Siamese neural network results are composed of two vectors from three different points: the anchor image of unknown classification to the positive image and the negative image. Therefore, we have a triplet of points to learn from. In our previous study on Covid-19, we used contrastive loss to calculate the loss during training. We opt to use triplet loss in this part of our work because it has been shown to better inform the Siamese neural network of differences between a similar image and a different image, as shown in figure 5. For our purposes of comparing a positive and negative value to an anchor, the triplet loss function is clearly the best choice.

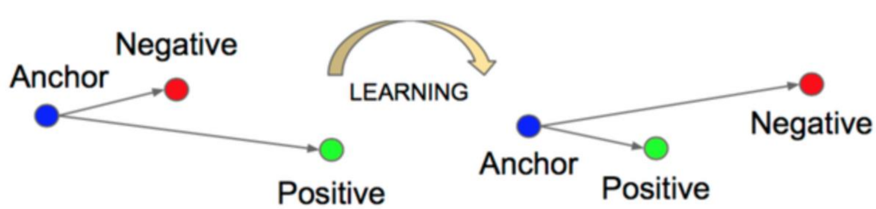

Figure 5: learning using triplet loss to adjust weights and biases so that the anchor image is moved closer in similarity to a positive image, and away from a negative. From Schroff et al. [35].

Based on the past research done in the field of thick data analytics and medical imaging, we see that adding meta filters and heuristics improve the classification accuracy when using a Siamese neural network. However, what is yet to be explored is if a Siamese neural network can distinguish from different diseases on its own. For example, the signs of ulcerative colitis and Crohn’s disease are quite similar, but there are architectural differences that can distinguish between the two diseases. Crohn’s disease can be identified by ulcers, erosions, and erythema spanning a length of the colon or intestine. Ulcerative colitis can also be identified using these heuristics, but the appearance will be more localized, and may also contain the presence of polyps, a heuristic not commonly associated with Crohn’s disease. Therefore, to develop a Siamese neural network that can accurately identify Crohn’s disease, we must add these heuristics to the images so the network can learn to distinguish between the two.

3.3. Preprocessing

3.3.1. Meta Filters

We will be using several augmentation filters to test the network’s classification capabilities and train it to identify Crohn’s disease. The first set of filters we will be using are meta filters which alter the color and size space of the images. This method was implemented in the Fiaidhi et al(2021) study and showed an improvement on the Siamese neural network's abilities to identify ulcerative colitis. We hope to achieve the same results with this method on Crohn’s disease. We will be using these filters to enhance the contrast and center crop images during preprocessing using a random number generator and python Pil image augmentation functions:

```
    from tensorflow.keras.preprocessing.image import img_to_array,
array_to_img
    from PIL import Image, ImageDraw, ImageEnhance
    if(train_with_meta_filters):
```

```
image = array_to_img(image)
enhancer = ImageEnhance.Contrast(image)
image = enhancer.enhance(3)
frac = 0.9
left = image.size[0]*((1-frac)/2)
upper = image.size[1]*((1-frac)/2)
right = image.size[0]-((1-frac)/2)*image.size[0]
bottom = image.size[1]-((1-frac)/2)*image.size[1]
image = image.crop((left, upper, right, bottom))
image = image.resize((224,224))
image = img_to_array(image)
```

We also implement several heuristic filters to train the network to distinguish between ulcerative colitis and Crohn's disease. The heuristics we are augmenting on to the images are ulcers, erosions, erythema, and polyps. This is done using two methods: bounding box annotations, and segmentation mask overlays.

3.3.2. Bounding Boxes

The bounding box augmentations are done using a pretrained VGG16 neural network that is then customized to output 4 points for the beginning and ending of the bounding box in (x1, y1, x2, y2) format. The customized network is then trained on each dataset for the four heuristics using the associated bounding box annotations from the metadata provided for each image. An unknown image is then fed through the network, and the resulting predicted bounding box is annotated on to the image using python Pil image augmentation functions:

```
# creating the model
from tensorflow.keras.applications import VGG16
vgg = VGG16(weights="imagenet", include_top=False,
input_tensor=Input(shape=(200, 200, 3)))
vgg.trainable = False
flatten = vgg.output
flatten = Flatten()(flatten)
bboxHead = Dense(128, activation="relu")(flatten)
bboxHead = Dense(64, activation="relu")(bboxHead)
bboxHead = Dense(32, activation="relu")(bboxHead)
bboxHead = Dense(4, activation="sigmoid")(bboxHead)
model = Model(inputs=vgg.input, outputs=bboxHead)


# Training the model
opt = Adam(lr=config.INIT_LR)
model.compile(loss="mse", optimizer=opt)
H = model.fit(
	trainImages, trainTargets,
	validation_data=(testImages, testTargets),
	batch_size=config.BATCH_SIZE,
	epochs=config.NUM_EPOCHS,
	verbose=1)
model.save(config.MODEL_PATH, save_format="h5")


# Making bounding box coords predictions from model
for imagePath in imagePaths:
	image = load_img(imagePath, target_size=(224, 224))
```

```
image = img_to_array(image) / 255.0

image = np.expand_dims(image, axis=0)

preds = model.predict(image)[0]

(startX, startY, endX, endY) = preds

image = cv2.imread(imagePath)

(h, w) = image.shape[:2]

startX = int(startX * 224)

startY = int(startY * 224)

endX = int(endX * 224)

endY = int(endY * 224)

cv2.rectangle(image,(startX, startY),(endX, endY),(0, 255, 0), 2)
```

3.3.3. Segmentation Masks

The segmentation mask augmentations are done using a pretrained ResNet50-UNet neural network from the python library Keras-segmentation. Segmentation mask information is provided in the Kvasir dataset, however only polyp segmentation masks are provided as opposed to the bounding box annotations that were provided for every heuristic. This works for our purposes as the goal of using the segmentation masks is to train the network to distinguish between ulcerative colitis and Crohn's disease. Since polyps are a heuristic more commonly used to diagnose ulcerative colitis, this heuristic will be used to distinguish between a possible ulcerative colitis misclassification. The ResNet50-UNet is trained on the 1000 polyps images and corresponding segmentation masks from the Kvasir dataset, then images without a segmentation mask are fed through the network to predict a corresponding mask. The mask is then overlaid on the original image, and the newly annotated overlay is used to train and test the network for polyps.

```
# Training the model
from keras_segmentation.models.unet import resnet50_unet
model = resnet50_unet(n_classes=256)
model.train(
        train_images = train_image_path,
        train_annotations = train_mask_path,
        checkpoints_path = checkpoint_path,
        epochs=5
        )
model.save('E:\Code\Final_SNN\segmentation_mask_augmentations\output\seg
mentation_model.h5', save_format="h5")


# Getting segmentation mask predictions
model = model_from_checkpoint_path(checkpoint_path)
out = model.predict_segmentation(
        inp=img,
        out_fname= out_path+"/"+filename,
        overlay_img=True
        )
```

We run several tests using each of these augmenters individually, and once with all used at the same time. Examples of the base image and the resulting augmentation from the above-described methods is shown in figure 6.

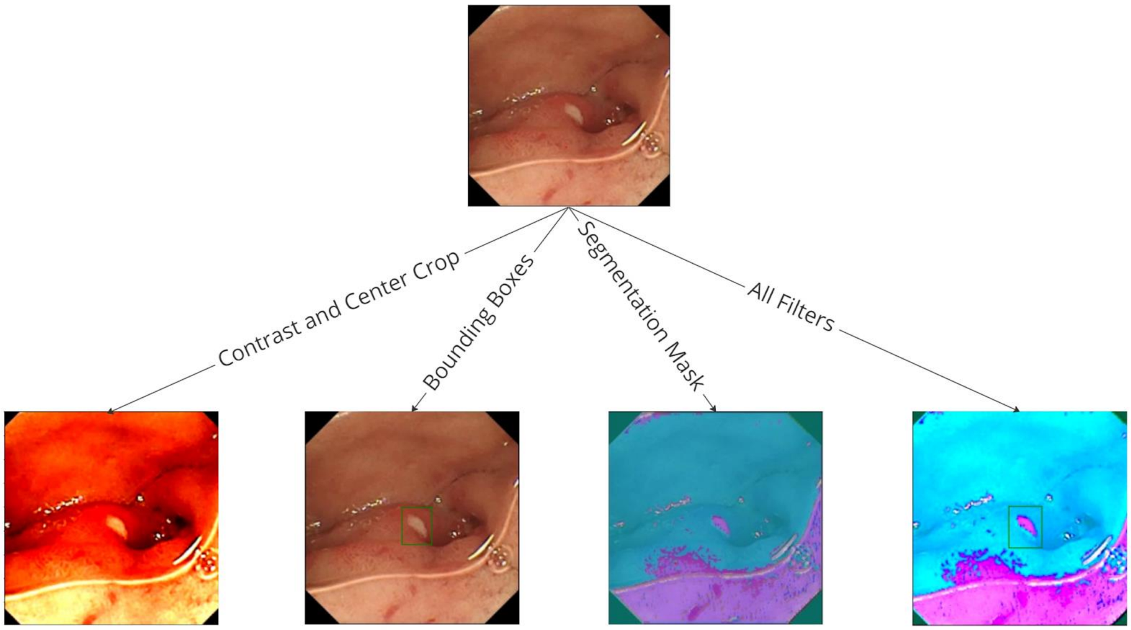

Figure 6: Meta and heuristic filters applied to an endoscopy capsule frame found positive for ulcers.

Once the networks used for the augmentations described above are properly trained, they are saved as .h5 files and subsequently loaded when needed during training and testing of the Siamese neural network. At the preprocessing stage, we create a data thickening augmentation pipeline from these networks. Within our main config python file, we specify three indicators for when we want these networks to be used. If an indicator is set to true, then the corresponding augmentation is performed on the dataset, otherwise the augmentation is skipped. This pipeline is how we analyze the performance of our SNN using individual augmentations first, then finally with all augmentations to test if the data thickening augmentations compound into useful results.

3.4. Base Network and Siamese Network Architecture

Of course, to implement these data thickening techniques, we must devise an optimized SNN architecture to learn from the applied heuristic augmentations. In our previous work on Covid-19 classification for lung CT scans, we used a custom sequential

ResNet inspired network with residual layers to incorporate features learned early on in the learning process. However, in the work done by Fiaidhi et al. [10] on ulcerative colitis, they used a sequential convolutional network to classify ulcerative colitis scores. Although they were performing different tasks, the raw accuracy of the convolutional network was higher than ours using residual layers, so during preliminary testing of Crohn's heuristics detection, we set out to test which network architecture might perform best for our study. Our early results indicate that the convolutional network not only performed slightly better than the network with residual layers, but it was also able to process the images faster, making the implementation of the convolutional neural network base SNN in a prototype later on more feasible. Additionally, early experimentation with using a Siamese neural network with a CNN base that was untrained found that, when using triplet loss, overfit for the positive or negative image as when we were testing on 200 batches of triplets and no prior training, we got accuracies of either 0% or 100% inconsistently between test runs. We therefore took inspiration from the Ulcerative colitis test where the base CNN was trained initially to identify between the different classes, then the resulting embedding layers were taken, used in a Siamese neural network, and then fine-tuned using triplet loss to distinguish between pairs of images. Our resulting final Siamese neural network using this method is displayed in figure 7 for the base network, and figure 8 for the Siamese implementation.

We train the base network to identify between the 4 classes of heuristics using the Kvasir dataset comprising 35912 images. The loss function used during training of the base network is categorical cross entropy, and RMSProp is used as the optimizer with a learning rate of 0.001.

When using the base network in the Siamese neural network, we keep all the parameters as trainable so the network can learn to distinguish between anchor, positive, and negative images using triplet loss as described above. The Siamese neural network is built using Input tensors of (224, 224) size, an output is the embedding of the images after the flatten layer of the base network. We do not use the dense layers of the base CNN so that the output embedding is not reduced to the original 4 features output. Instead, the output of the network is a tuple of the three embeddings, each of size (30976). To calculate a classification of the anchor image, cosine similarity is used. Whichever image embedding is found to be more similar to the anchor is therefore considered to be part of the same classification.

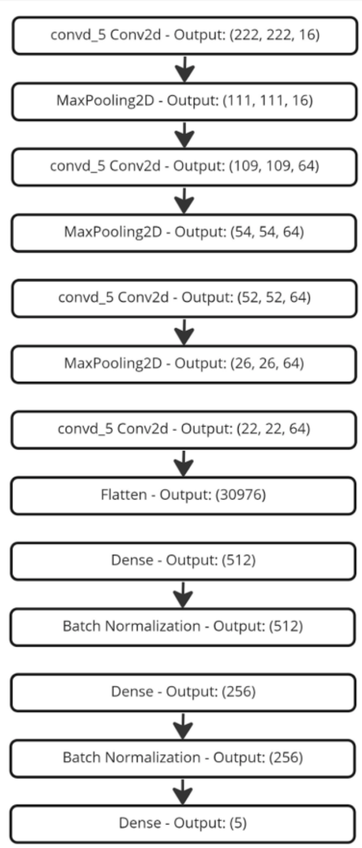

Figure 7: Base CNN architecture used in our Siamese neural network.

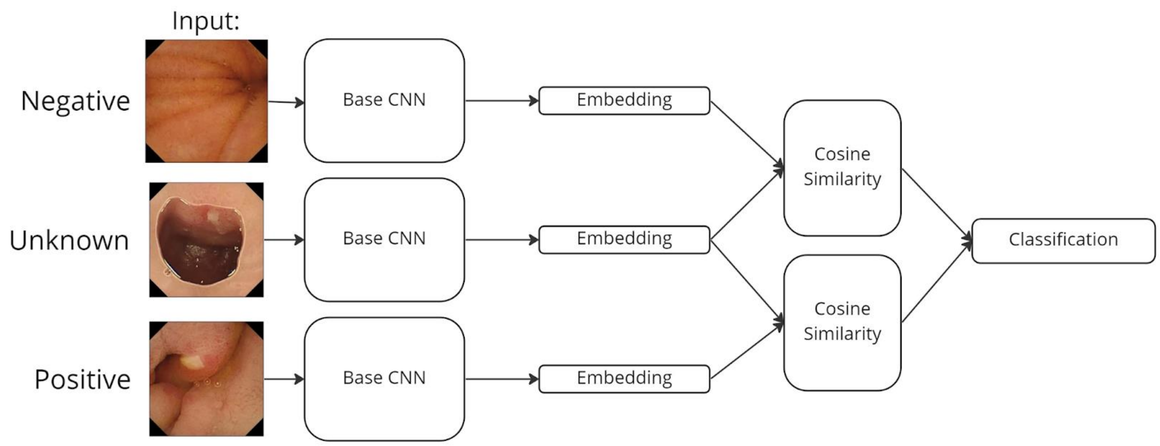

Figure 8: Siamese neural network implementation of the Base CNN taking in 3 inputs and calculating cosine similarities to reach a classification

3.5. Training and Testing Methodology

Training of the Siamese neural network is done over 10 epochs using a batch of 36 image triplets. During testing, we test 200 triplets of images containing 2 images of the same classification, and 1 of differing classification. If the 2 images from the same classification are found to be more similar than the different classification image, then that prediction is said to be correct, otherwise it is incorrect. We use 200 image triplets to test the network to ensure that enough testing triplets are used to confidently find the accuracy of a trained network, as using the traditional 80-20 training and testing split in this case would yield insignificant results in training batch sizes as low as 36 for example.

Once the network is sufficiently trained and tested, and the optimal training parameters for identifying Crohn’s disease heuristics have been found, we save the resulting network in .h5 format to be used in a physician frontend prototype to extend the network to frame-level detections of Crohn’s disease heuristics. This frontend takes in a capsule-colonoscopy video, splits the video into frames, and runs the subsequent frames through the two augmentation networks and the SNN to find if the frame is similar to images showing the heuristics of Crohn's disease. If the heuristics for polyps are found to be high along with the other heuristics, a notification is displayed to the user to indicate that the frame may be classified for ulcerative colitis as well, and further investigation is required. Our results of the SNN training and testing as well as the viability of the network at frame-level detection is explained and analyzed in detail in chapters 4 and 5. Additionally, all source code, datasets, pre-trained networks, and examples of how we came to our results are available at https://github.com/AttaboySawyer/Thesis.

Chapter 4.

Implementation of Siamese Network into Application Dashboard

4.1. Results

4.1.1. No Filters

The results of our trained networks after testing on 200 triplets of images are separated by the augmentations made to identify which augmentation is best for identifying specific heuristics, as well as to identify if there is a compounding benefit to the classification accuracy as heuristic augmentations are added. This creates 5 sets of individual test data: no augmentations, meta augmentations, bounding box annotations, segmentation masks, and all augmentations applied, as well as one fila test with the best augmentation techniques applied based on the first 5 sets of results. As mentioned, each augmentation is tested for each heuristic using the datasets from the Kvasir repository.

Table 1: Tests for heuristic Identification with No Augmentations Made on Data Set

	Heuristic			
Epoch	Ulcers	Polyps	Erythema	Erosions
1	53.50%	69.5%	50%	47.5%
2	53.50%	69%	50.5%	49%
3	56.99%	71%	50%	50.5%

4	55%	70%	47.5%	51.5%
5	56.49%	72.5%	45%	53.5%
6	58.5%	72.5%	45.5%	50.5%
7	61%	74%	46.5%	49.5%
8	62%	73%	48.5%	49.0%
9	63%	74%	48%	50%
10	62%	76%	48.5%	49.5%

After training on 36 batches of triplets over 10 epochs, we can see that ulcer detection and polyps detection were the most accurately identified by the Siamese neural network with final accuracies of 62% and 76% respectively, while erythema and erosion detection remained insufficiently accurate, at around 48.5-49.5%. Over the course of the 10 epochs, we also see that the increase in ulcer and polyp detection accuracy is directly proportional to the training on small batches, whereas erythema and erosion detection did not rise or increase with small batches training.

4.1.2. Meta Filters

Table 2: Tests for heuristic Identification with Meta Filters Augmented on Data Set

	Heuristic			
Epoch	Ulcers	Polyps	Erythema	Erosions
1	53%	70%	54%	48.5%
2	54%	69%	55.5%	48%
3	53%	71%	53%	48%

4	54.5%	72%	52.5%	46%
5	56.49%	75.5%	53%	44%
6	56%	79%	51%	43.5%
7	57.49%	79.5%	51.5%	44.5%
8	57.49%	82.5%	51%	44.5%
9	59.5%	83.5%	50%	46%
10	58.5%	84%	49.5%	45.5%

After adding meta filters for increasing contrast of the images and center cropping by 10% during the preprocessing stage, we see that again, ulcer and polyp detection accuracies were highest after 10 epochs of training, at 58.5% and 84% respectively. We note that adding the meta filters significantly increased the network's ability to detect polyps after training. Similarly, to training with no meta filters, erythema and erosion detection remained insufficient at about 50% accuracy. Another difference in training with meta filters is that ulcer accuracy did not reach as high an accuracy as it did when no filters were used.

4.1.3. Bounding Boxes

Table 3: Tests for heuristic Identification with Bounding Box Augmentations Made on Data Set.

	Heuristic			
Epoch	Ulcers	Polyps	Erythema	Erosions
1	55.5%	68.5%	53%	51%
2	56.99%	70%	52%	52%
3	59%	70.5%	49.5%	52%

4	64.5%	70%	48%	53.5%
5	64%	72.5%	49%	52.5%
6	66.5%	72%	53%	51%
7	71%	71.5%	53%	51%
8	73%	70%	55%	53%
9	75.5%	70.5%	55%	52%
10	77.5%	69%	55.5%	52.5%

After using the bounding box annotation network to detect for Crohn's disease heuristics, we see the most consistent increase in detection accuracy after training. Ulcer detection rose to a high of 77.5%, polyps rose to 69%, erythema achieved an accuracy of 55.5%, and erosions remained somewhat similar to previous tests at a high of 52.5%. Notably, the accuracies for erythema and erosion detection increased after training with bounding box annotations, whereas previously they had stayed consistently around 50%, even sometimes decreasing in accuracy with training. Although the accuracies after 10 epochs of training do not reach the maximum accuracy of 84% for polyps in the case of meta filters, bounding box annotations appear to be the most consistent method of training the Siamese neural network to identify heuristics across every heuristic we train for, thus far.

4.1.4. Segmentation Masks

Table 4: Tests for heuristic Identification with Segmentation Mask Augmentations Made on Data Set

	Heuristic			
Epoch	Ulcers	Polyps	Erythema	Erosions

1	51%	67%	45%	52%
2	51.5%	66.5%	43.5%	51.5%
3	53.5%	65.5%	43.5%	50.5%
4	52%	70.5%	44.5%	48%
5	53.5%	71%	46%	50.5%
6	55.5%	71%	47%	51%
7	57.99%	71.5%	46.5%	52%
8	58.5%	70.5%	49.5%	52%
9	62.5%	72%	50%	53.5%
10	66.5%	71%	49.5%	55.5%d

After using the segmentation mask network to apply segmentation overlays to images during preprocessing, we again see a more consistent increase in accuracy across all heuristics with using a heuristic filter as opposed to a meta filter or no filters. The overall accuracies of the segmentation mask filter tests do not reach as high accuracies as when bounding box annotations are used, as ulcers reached a high of 66.5% and polyps reached a high of 71%. Erythema detection, although again saw a consistent increase with training, failed to reach an accuracy greater than 50%, while erosions detection performed the best it has in any trial, reaching an accuracy of 55.5% after training.

4.1.5. All Filters

Table 5: Tests for heuristic Identification with Meta Filters, Bounding Box, and Segmentation Mask Augmentations Made on Data Set

	Heuristic			
Epoch	Ulcers	Polyps	Erythema	Erosions
1	47%	71%	49.5%	51%
2	47.5%	69.5%	51.5%	52%
3	46%	68%	51.5%	56%
4	48.5%	71%	54.5%	59%
5	50%	72.5%	55%	59.5%
6	49%	77.5%	54%	59%
7	53%	80.5%	53%	57.49%
8	55.5%	83%	55%	57.49%
9	55.5%	83.5%	52%	57.49%
10	55.5%	87.5%	53%	56.49%

After applying all data thickening filters on the dataset during preprocessing, we see that this results in the highest polyp detection accuracy across all tests at 87.5%. However, combining these methods of augmentation conversely decreased the maximum accuracy after training for every other heuristic, with ulcers reaching 55.5%, erythema reaching 53%, and erosions reaching 56.49%. We note that erosion detection did also reach a highest accuracy after training using these filters, about 1% higher than it's previous highest accuracy using segmentation masks, but a highest accuracy of 56.49% is negligible when ulcer and erythema detection are decreased as significantly to 55% after training.

After reviewing all the results from the heuristic and meta augmentation filters, we find that the most consistent filter for increasing accuracy across all heuristics is bounding box annotations. This method showed a high accuracy for ulcer and polyp detection, while also constantly increasing erythema and erosion detection with training. It is for these reasons that we choose to use bounding box annotations as our main heuristic filter for detecting Crohn’s disease.

4.2. Prototype Architecture

To extend the now trained Siamese neural network to frame-level detection, we propose and develop a physician frontend prototype that can take in capsule colonoscopy videos, and filter through the frames of the video to find heuristics indicative of Crohn’s disease while giving an indication for possible ulcerative colitis misclassification. When extending our Siamese neural network to frame-level detections, we can Leverage the network to implement heuristic filtering techniques that better identify Crohn's disease rather than other diseases. As mentioned, one heuristic of Crohn's disease is that the ulcers, erythema, and erosions span sections of the intestinal tract in a cobblestone-like pattern whereas other IBDs are more localized. With this in mind the first step in implementing our network to frame level detections is to cluster frames by similarity. this is done by comparing the current frame to the previous frame and the next frame. If the current frame is found to be more similar with the previous frame, then they are clustered together, otherwise if the current frame is found to be more similar with the next frame, then a new cluster is created, and the process repeats until all frames are clustered by similarity. This technique also aims to remove noise from the video by removing frames that are not similar to any others due to possible fecal matter and mucosa blocking view of the intestinal or colon lining. Once all frames are clustered and noise is removed, we passed the frames through the trained Siamese neural network to detect for heuristics of

Crohn's Disease across multiple frames. Based on the results of our meta and heuristic filters test, we use bounding boxes as the main filter to detect ulcers, polyps, erosion, and erythema. While this is the main technique used to identify heuristics of Crohn's Disease, we also provide the user with the corresponding bounding box annotations and segmentation mask overlays, as well as the calculated percentages of a possible heuristic to provide as much contextual information to the user as possible for them to make a diagnosis of the suspected frames.

Our prototype consists of two main components: the front end made using React, and the python server backend made using Flask. The user begins the classification process by uploading a capsule colonoscopy video to the React frontend. The frontend communicates with the server using HTTP, sending the uploaded video to the Flask server. The server then loads the pre-trained Siamese neural network, the pre-trained bounding box annotation Network, and the pre-trained segmentation mask annotation Network. The video is then split into frames and the frames are fed through the Siamese neural network where clustering is performed. Similarity is again calculated using cosine similarity, and any frames found to be not similar with any other surrounding frames are removed. Once all frames having clustered, the remaining frames are fed through the pretrained Siamese neural network Where bounding box annotations are used to detect ulcers, polyps, erosions, and erythema. If ulcers, erosions, and erythema are found in several frames within a cluster, then the cluster is flagged for a high chance of Crohn's Disease presence. If only one or two frames in a cluster are detected to have ulcers, erosions, and erythema, then the cluster is still flagged for a low chance presence of Crohn’s disease since a cobblestone pattern is a heuristic with Crohn’s disease, but the detection of these heuristic is still worth presenting to the user. Finally, if polyps are detected across the frames along with ulcers, erosions, and erythema, then the cluster is flagged for possible Crohn's disease presence, but with the notification that this may be ulcerative colitis since polyps are a more common heuristic of ulcerative colitis rather than Crohn's Disease. The frontend and server communication logic are shown below in figure 9

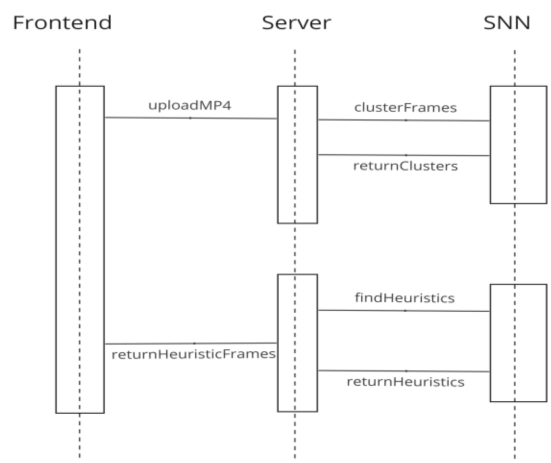

Figure 9: UML diagram of frontend, server, and SNN communication to cluster and detect heuristics at the frame level, and return detected images to the frontend.

As well, below is the pseudocode depicting how the images are clustered, analyzed for heuristics, and flagged appropriately:

```
# Splitting video into frames
all_frames_imgs = []
all_frames_ids = []
shape = (224, 224)

mp4 = request.files['myFile']
mp4.save("./temp.mp4")
vidcap = cv2.VideoCapture("./temp.mp4")
success,img = vidcap.read()
count = 1
while success:
```

```
        img = image.smart_resize(img, shape)
        img = np.asarray(img)
        img = cv2.cvtColor(img, cv2.COLOR_RGB2BGR)
        all_frames_imgs.append(img)
        all_frames_ids.append("frame " + str(count))
        success,img = vidcap.read()
        count += 1

# Load the models
snn = load_model(snn_path)
bb_model = load_model(bounding_box_augmentor_path)
seg_model = model_from_checkpoint_path(seg_mask_augmentor_path)

# Cluster images
counter = 1
img1 = all_frames_imgs[0]
img2 = all_frames_imgs[1]
img3 = all_frames_imgs[2]
for compare_image in tqdm(all_frames_imgs):
        cosine_similarity = metrics.CosineSimilarity()
        before = img1
        img2 = compare_image
        after = img3

        before = np.expand_dims(before , 0)
        img2 = np.expand_dims(img2, 0)
        after = np.expand_dims(after , 0)
```

```
            output1,output2,output3 = snn(img2,before,after)
             before_similarity = cosine_similarity(output1.numpy(),
output2.numpy())
            after_similarity = cosine_similarity(output1.numpy(),
      output3.numpy())
            if (after_similarity  > before_similarity ):
                    cluster_array.append([img2, all_frames_ids[counter],
[]])
                    cluster_collection.append([0,cluster_array])
                    cluster_array=[]
                    before = img2
                    img2 = after
                    after = all_frames_ids[counter + 1]
                    counter = counter + 1
                 else:
                    cluster_array.append([compare_image,
all_frames_ids[counter], []])
                    before = img2
                    img2 = after
                    after = all_frames_ids[counter + 1]
                    counter = counter + 1


            # Finding heuristic similarities
            def getPrediction(anchor, positive, negative):
                  emb1, emb2, emb3 = model.predict( anchor, positive,
negative)
```

```
            pos_sim = cosine_similarity(emb1.numpy(), emb2.numpy())
            neg_sim = cosine_similarity(emb1.numpy(), emb3.numpy())
            if(pos_sim > neg_sim):
                return(False, pos_sim)
            else:
                return(True, neg_sim)


        for image_cluster in cluster_collection:
              images = image_cluster[1]
              for img in images:
                    ulcer_test = getPrediction(img[0],
normal_img, ulcer_img)
            polyp_test = getPrediction(img[0], normal_img, polyp_img)
            erosion_test = getPrediction(img[0], normal_img,
erosion_img)
            erythema_test = getPrediction(img[0], normal_img,
        erythema_img)
            if(ulcer_test, polyp_test, erosion_test, erythema_test ==
        False):
                img[2] = ("No detections")
            else:
                if(max(ulcer_test ,polyp_test ,erosion_test ,
            erythema_test) == ulcer_test): img[2] = ("Ulcer
            detection", ulcer_test)
                if(max(ulcer_test ,polyp_test ,erosion_test ,
            erythema_test) == polyp_test): img[2] = ("Polyp
            detection", polyp_test)
```

```
	if(max(ulcer_test ,polyp_test ,erosion_test ,
erythema_test) == erosion_test): img[2] = (“Erosion
detection”, erosion_test)
	if(max(ulcer_test ,polyp_test ,erosion_test ,
erythema_test) == erythema_test): img[2] = (“Erythema
detection”, erythema_test)
```

4.3. Examples of Use

The home screen of the prototype is shown in figure 10 and the output of a processed video using our prototype is shown in figure 11**.** The left side of the prototype is the cluster display window where the clusters are colored based on their returned flags, and the user can go through the clusters frames side by side to verify if a cobblestone pattern is present across a section of the colon or intestinal lining. The middle window is used to analyze single frames in greater detail. By clicking on a frame from the left window, an upscaled image is revealed in the center for easier viewing. It is here that frame-level analysis is available to the user. The corresponding bounding box and segmentation mask for the frame are viewable options at the top of the window, and he calculated similarities to the analyzed heuristics are shown to the right of the image. This gives the user full insight into how the network is calculating for possible Crohn’s disease detection to enhance the diagnosis process. a feature novel to the Siamese neural network approach. Finally, the right side of the prototype frontend is for users to make annotations of their own, as well as to take notes on frames they are inspecting.

Users can draw on images directly if they notice something that was not picked up by the Siamese neural network and save the annotated frames to their computer.

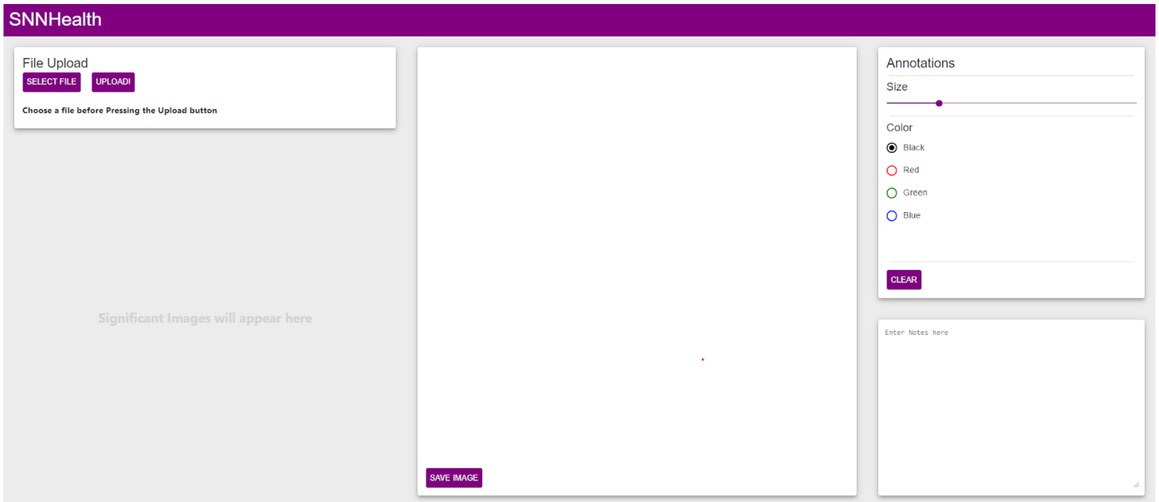

Figure 10: Prototype home screen. The user uploads a capsule colonoscopy video in mp4 format to the left, and processing is conducted when they press the UPLOAD button.

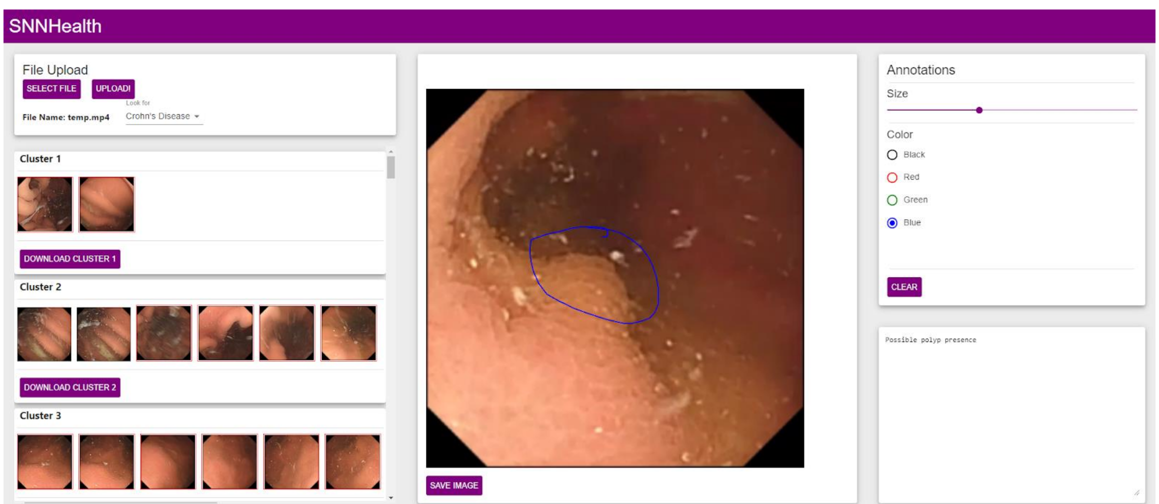

Figure 11: Prototype after processing has been conducted on a capsule colonoscopy video. clusters and images flagged for heuristics are displayed to the left, single images are brought to the middle screen when clicked from the left, and annotations to the center image can be made using tools on the right.

In the prototype implementation, we suggest that frame clustering can be used to possibly capture the discontinuous lesion appearance heuristic of Crohn's disease, however, this methodology is not tested for accuracy of its ability to do so. Just as the prototype is not indicative of a final product diagnostic tool, the clustering technique we propose is not indicative of a correct way to capture the discontinuous nature of Crohn's disease at this stage in our research. Further testing and clinical trials with physicians having hands-on experience with the prototype is needed to definitively tell if clustering is a necessary component of the backend server process. This type of testing is beyond the scope of this thesis but may be a valid point of future work in this field of study.

Chapter 5.

Conclusion and Future Research

5.1. Discussion of Results

During our analysis of augmentations to identify heuristics in endoscopy capsule videos, we found that the most consistent augmentation technique for increasing the accuracy of heuristic identification was bounding box annotations. Bounding box annotations have been used in a variety of feature extraction research for image processing via works with good results [24], so this result is a reasonable conclusion and given our findings and other published works. Other augmentation techniques did not consistently increase the accuracy of here to get an application for every heuristic tested for, but in some cases specific heuristics did benefit from certain augmentations, while others suffered. We offer some reasons as to why this might be the case.

Firstly, in the case of using meta filters to augment images, ulcer and polyp identification benefited the most from these augmentations, with ulcer accuracy and polyp accuracy reaching 58.5% and 84% respectively. However, erosion and erythema accuracies did not see a significant increase in classification accuracies beyond about 50%. This may be due to the meta filters we tested only changing the color space of the image, specifically increasing contrast and center cropping so the black corners of the frames are mostly removed. Ulcers are characteristically identified by a white area of the tissue, so it is reasonable to propose a change in the color space will highlight the contrast between the white areas of the ulcer with the pink and red unaffected tissue. Polyps, on the other hand, are characterized by protrusions of the wall lining, creating irregular shadows and dark color spaces in the image. Again, it is reasonable to propose that a change in the color space will highlight these irregular dark shadows and color spaces in the image, leading to better classification accuracies. In contrast to these two heuristics, erosions and erythema do not characteristically have white or dark areas they are identified by. Instead, they are more characteristically red in appearance than the pink unaffected wall tissue. While this is a contrast in color, it may not be enough of a contrast for color space augmentations to highlight these two heuristics, leading to our results where ulcers and polyps classification accuracies increase with contrast and cropping, but erosions and erythema do not.

Secondly, when we use bounding box annotations to augment images, this is the first consistent increase we see in our tests where classification accuracy is increased across all heuristics, with ulcers reaching 77.5%, polyps reaching 69%, erythema reaching 55.5%, are erosions reaching 52.5%. While these are not the highest accuracies recorded for certain heuristics, it is the most promising data thickening technique as no heuristic classification accuracy decreases with the increase in others as is the case with segmentation masks and all filters applied. We propose that this is because the bounding box annotations for each heuristic are different in size to the point that it is easier for the network to identify a heuristic based on the number of green annotation pixels in the image. For example, the polyp bounding box annotations are quite large when compared to the bounding boxes for ulcers, as shown in figure 12.

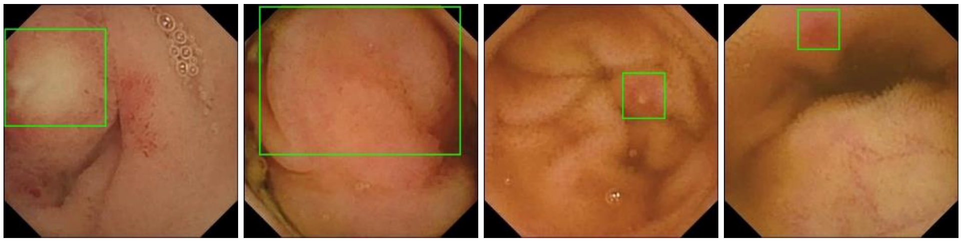

Figure 12: Bounding box annotations for ulcer, polyp, erosion, and erythema from left to right.

Bounding boxes for erythema and erosions are comparable in size, which may explain why the network has a difficult time distinguishing between the two heuristics.

When we use segmentation mask filters and overlay the resulting mask on to the image for analyzing, again we see a consistent increase in classification accuracy during training, with ulcers reaching 66.5%, polyps reaching 71%, erythema reaching 49.5%, and erosions reaching 55.5%. These results are comparable to the other heuristic filter used (bounding boxes) however in this case, classification accuracies do not reach as high a result as they do when using bounding boxes. This may be because the segmentation mask network was only trained on polyp images and corresponding masks, so the training was not enough to identify other heuristics or irregularities in the intestinal or colon tissue. That being said, these results are also not as good as when we use other augmentations, even in the case of polyps which performed the best when using segmentation masks. Here we see polyps reaching an accuracy of 71%, however, when only meta filters were used, polyp classification accuracy reached as high as 84%. Additionally, bounding box annotations still maintain the highest classification accuracies across all other heuristics, most likely due to the reasons stated previously. The segmentation masks may also not be trained on enough classes to identify a variety of irregular tissue. For example, in our studies in identifying Covid-19 using thick data analytics, we used segmentation masks to improve the classification accuracy of our results. In that case, the trained segmentation network used 4 classes (3 heuristics and the background) in its segmentation, whereas in this segmentation network's training set, only 2 classes are used (polyps heuristic and background). With that in mind, we propose that further analysis be conducted on segmentation mask annotation for thick data analytics with datasets that range the full heuristic scope of IBDs. We also conclude that since the segmentation mask annotations did not improve polyp detection greater than when meta filters were used, we cannot assume that training the segmentation network on all heuristics would improve classification accuracies better than other heuristic filters, such as bounding boxes. Therefore, our conclusion that bounding box annotations are the most promising thick data analytics filter is reasonable.

Since augmentations for color spaces and heuristic filters using bounding boxes and segmentation masks all improved classification accuracies independently, we test whether combining these augmentations during preprocessing to see if the improvements to classification accuracy compound into a preprocessing pipeline that can greatly increase a Siamese neural networks ability to learn heuristics from small samples. Our results indicate that although polyps and erosions reached their highest tested classification accuracies of 87.5% and 56.49% respectively using this pipeline, the accuracies for ulcers and erythema suffered, and in the case of ulcer detection, suffered greatly with a reported 55.5% and 53% respectively. This may be because of the proposed reasons, namely for ulcers that are characterized by their white appearance and smaller size, leading to easier identification using color space augmentations and bounding boxes. When the segmentation mask is overlaid onto the image, the white color of the heuristic may be lost in the augmentation if it is not being identified by the segmentation mask, leading to less accurate results in this pipeline. The same can be said for erythema, where the bounding box is difficult to distinguish in terms of size from erosions, and their characteristic red color may be lost in preprocessing if the segmentation mask does not identify those areas. Based on the results of the segmentation mask accuracies, this may be the case for erythema and ulcers which both saw decreases in classification accuracies compared to the results when using bounding boxes. Therefore, it is reasonable to conclude that the pipeline method is not a suitable method for identifying ulcers and erythema but may show promise if the heuristics that require identification are polyps and erosions.

We conclude that bounding box annotations are the best heuristic augmentation method for identifying heuristics associated with Crohn's disease, however our results show that these thick data analytics may have specific use cases when other methods may be more appropriate. For example, if one is developing a network for identifying polyps specifically, then the pipeline method we use in this study would be most suitable. We are more interested in identifying Ulcers over large spaces of a video, so we did not implement this method in our prototype, but it could be implemented in a future network where polyp identification is the main focus. The same can be said about erosion detection, as the pipeline method proved more accurate in their identification, however the highest accuracy achieved after training was 56.49%, so identification of this heuristic will most likely require further study and additional techniques.

5.2. Comparison to Published Work

When compared to other published work, our results in classification are on par with accuracies achieved by traditional convolutional networks, although the goals of some studies are slightly different from ours. For example, we are identifying heuristics indicative of Crohn's disease, to reach a possible diagnosis, not diagnosis for the disease outright in classifying both single diseases and multiple diseases at once. The accuracies of networks that diagnose for diseases of the lung and thorax range from 71% to 88% [11, 12, 13], but they provide no output of the network's reasoning on its reached classification. That is one of the many benefits of using Siamese neural networks and similarity to detect diseases; it can be used as an enhanced diagnosis tool that does not need to be trusted blindly. When compared to studies which aim to segment areas of interest from medical images,

In addition to improving the classification accuracies for heuristics, we manage to improve these accuracies using small batches compared to other published works. For example, the study conducted by Wang et al. [15] on thoracic diseases and the study by Lonseko et al. [16] on endoscopic images were trained and tested on 14,100 to achieve a highest accuracy of 71.34%, and 12,147 images to achieve an accuracy of 93.19% respectively. In contrast, our work which achieved heuristic classification accuracies of 87.5% for polyps with all filters applied, and 77.5% for ulcers with bounding box annotations while only training on 1,974 for the base convolutional neural network, and 36 batches of images triples per test for the Siamese neural network, a significant decrease in training and testing images to achieve similarly classification accuracies due to the decreased training needs of the Siamese neural network.

Thick data analytics is an emerging field with few published works that use comparable techniques to ours in detecting diseases, but some work in the field include ours done on Covid-19 detection in lung CT-scans, and UC severity score using a Siamese neural network and meta filter augmentations. In the case of Covid-19 detection, we used segmentation masks to identify Covid-19 in a similar method to how we segment images in this study. Our results are somewhat comparable to that study, where our implementation of segmentation masks increased the accuracy of heuristic identification in all heuristics except for polyps when compared to results on heuristic identification with no augmentations applied. The result of using segmentation masks to identify Covid-19 also improved accuracy when compared to traditional sequential networks, supporting our results that segmentation masks improve heuristic identification.

When comparing our results to those of the UC severity score, we also see comparable results and conclusions. When studying the viability of using a Siamese neural network to score UC in images, they achieved an accuracy of 76.9%. Although they were not identifying heuristics specifically, we can compare our results of identifying heuristics to their results of identifying UC severity scores. UC is often diagnosed by looking for the presence of ulcers and polyps. Since we achieved, at our highest, a classification accuracy of 87% for polyp detection and 77.5% for ulcer detection, using our heuristic identifying Siamese neural network, it is reasonable to say our results are slightly better at identifying Ulcerative colitis. Although the goal of this paper is detecting Crohn’s disease, the heuristics of these diseases are quite similar, opening the possibility for our network to be used to detect ulcerative colitis in the future.

5.3. Suggestions for Improvements

Based on our results and comparisons to previously published work, we conclude that our Siamese neural network is a reasonable detector for Crohn’s disease by identifying ulcer and polyp heuristics since the classification accuracies of these heuristics were at or above 70%, which is the apparent threshold for a reasonably trained classification network based on related work. However, our network and technique is not without room for improvement. Two apparent improvements that will strengthen the reliability of our Siamese neural network and frontend capabilities are increasing the accuracy of erosion and erythema classification while maintaining or even improving the classification accuracies for ulcers and polyps. This can be done by exploring more heuristic based augmentations for a filter or pipeline of filters which increases all networks proportionally or increases lacking heuristic classification while maintaining heuristic classifications already at acceptable levels of accuracy. For example, in our work, bounding box augmentations increased the accuracy of all heuristics, but erosions and erythema were still lacking, whereas our test pipeline of filters dramatically improved polyps and erythema selection, but at the cost of ulcer and erosion detection. Additional heuristic augmentations that could be researched in future work are key points filters and landmark similarities. Another improvement that could be made is to find what certain heuristic filters work best with identifying what heuristics. In our study, for the reasons stated above, erosion and erythema detection remained difficult to increase accuracy for with our tested augmentations, while ulcer and polyp classification continually got better. We have proposed a few reasons as to why this might be, but a formal study on the reasonings for this and possible remedies would be a worthwhile endeavour in our opinion.

As for the prototype, improvements that could be made is a formal clinical trial to see if such a diagnostic tool using Siamese neural networks is significant in a medical setting. The glaring positive of our prototype as a tool for detecting diseases like Crohn’s disease is that it provides all reasons for the network’s classification to the user, a feature not inherent with traditional convolutional networks where you must, to some degree, blindly trust the result. An additional improvement to the prototype is in the time it takes to process videos at the frame level and rely on the information to the user. Some obvious solutions come to mind like hosting the prototype server on a machine with advanced GPU processing power, but other changes can be made to the network itself, such as decreasing the input size, although this may come at the cost of lost information by loss of pixels, or decreasing the number of layers in the network, however this may reduce the ability of the network to make classifications, as most published work has reached the conclusion that networks with deeper layers make more accurate classifications. Finally, a hope for the prototype and network is to be implemented in the detection of several diseases. Theoretically, with the correct training and heuristic filters, the network could learn to detect a variety of diseases based on heuristics.

Once the network is appropriately trained, all it would need is corresponding similarity images of heuristics for a given disease to be fed in as triplets. With advanced heuristic training and a multitude of similarity images for triplets, the possibilities for disease detection using this framework are quite endless.

5.4. Future Work

In terms of future research, we propose that heuristic identification be further investigated as our work has shown that some heuristic filters work better for identifying certain heuristics, such as bounding boxes being best for identifying ulcers, meta filters being best for identifying polyps, and a combination of filters being best for identifying erythema. Future work should focus on creating a strategy for identifying heuristics in the most optimized manner. We investigated whether an optimized method of heuristic identification involved the pipeline of many augmentation filters, and this showed not to be the most optimal method for all heuristics. An optimized strategy could involve the use of many filters independent of each other based on what heuristic is being identified. This could be the subject of a future study or thesis as identifying these best practices will be lengthy work and could involve other heuristics not studied in this work, as well as augmentation methods not studied in this work, such as key points or landmark similarities. Future work could also involve replicating our results across several IBDs as our focus was mainly on Crohn’s disease. It should be studied further as to whether heuristic identification, specifically bounding boxes, and meta filters, improve classification accuracies for many diseases that can then be implemented into the prototype. With future work focusing on heuristic specific identification augmentations as well as improving classification accuracies across several diseases, our prototype will become extremely robust with the capabilities to identify not only diseases, but disease heuristics at a finite customization level allowing the user to make extremely well-informed diagnoses, without having to blindly trust the network's predictions, allowing for a very acceptable implementation of AI into healthcare diagnoses.

References

[1] Kuenzig, M. Ellen, Douglas G. Manuel, Jessy Donelle, and Eric I. Benchimol. "Life expectancy and health-adjusted life expectancy in people with inflammatory bowel disease." Cmaj 192, no. 45 (2020): E1394-E1402.

[2] Halder W, Laskaratos F-M, El-Mileik H, Coda S, Fox S, Banerjee S, Epstein O. Review: Colon Capsule Endoscopy in Inflammatory Bowel Disease. Diagnostics. 2022; 12(1):149. https://doi.org/10.3390/diagnostics12010149

[3] R. Sathishkumar, K. Kalaiarasan, A. Prabhakaran and M. Aravind, "Detection of Lung Cancer using SVM Classifier and KNN Algorithm," *2019 IEEE International Conference on System, Computation, Automation and Networking (ICSCAN)*, Pondicherry, India, 2019, pp. 1-7, doi: 10.1109/ICSCAN.2019.8878774

[4] L. A. Thamke and M. V. Vaidya, "Classification of Lung Diseases Using a Combination of Texture, Shape and Pixel Value by K-NN Classifier," *2018 2nd International Conference on I-SMAC (IoT in Social, Mobile, Analytics and Cloud) (I-SMAC)I-SMAC (IoT in Social, Mobile, Analytics and Cloud) (I-SMAC), 2018 2nd International Conference on*, Palladam, India, 2018, pp. 235-240, doi: 10.1109/I-SMAC.2018.8653759.

[5] T. Aggarwal, A. Furqan and K. Kalra, "Feature extraction and LDA based classification of lung nodules in chest CT scan images," *2015 International Conference on Advances in Computing, Communications and Informatics (ICACCI)*, Kochi, 2015, pp. 1189-1193, doi: 10.1109/ICACCI.2015.7275773

[6] J. Fiaidhi and S. Mohammed, "Thick Data: A New Qualitative Analytics for Identifying Customer Insights," in IT Professional, vol. 21, no. 3, pp. 4-13, 1 May-June 2019, doi: 10.1109/MITP.2019.2910982.

[7] J. Fiaidhi, "Envisioning Insight-Driven Learning Based on Thick Data Analytics With Focus on Healthcare," in *IEEE Access*, vol. 8, pp. 114998-115004, 2020, doi: 10.1109/ACCESS.2020.2995763

[8] S. Pink, H. Horst, J. Postill, L. Hjorth, T. Lewis and J. Tacchi, Digital Ethnography Principles and Practice, Newbury Park, CA, USA:Sage, 2019.

[9] D. Sawyer, J. Fiaidhi and S. Mohammed, "Few Shot Learning of COVID-19 Classification Based on Sequential and Pretrained Models: A Thick Data Approach," *2021 IEEE 45th Annual Computers, Software, and Applications Conference (COMPSAC)*, 2021, pp. 1832-1836, doi: 10.1109/COMPSAC51774.2021.00276.

[10] J. Fiaidhi, P. Zezos and S. Mohammed, "Thick Data Analytics for Rating Ulcerative Colitis Severity Using Small Endoscopy Image Sample," *2021 IEEE International Conference on Big Data (Big Data)*, 2021, pp. 4687-4693, doi: 10.1109/BigData52589.2021.9671327.

[11] T. Nag, S. S. Rajawat and A. Rana, "Detection of Pneumonia using Chest X-Ray Images and Convolutional Neural Network," *2021 9th International Conference on Reliability, Infocom Technologies and Optimization (Trends and Future Directions) (ICRITO)*, 2021, pp. 1-4, doi: 10.1109/ICRITO51393.2021.9596172.

[12] N. Lee, A. F. Laine, G. Mrquez, J. M. Levsky and J. K. Gohagan, "Potential of Computer-Aided Diagnosis to Improve CT Lung Cancer Screening," in *IEEE Reviews in Biomedical Engineering*, vol. 2, pp. 136-146, 2009, doi: 10.1109/RBME.2009.2034022.

[13] C. Quek, W. Irawan, E.Y.K. Ng, “A novel brain-inspired neural cognitive approach to SARS thermal image analysis,” Expert Systems with Applications 37 (2010) 3040–3054.

[14] R. Zahedinasab and H. Mohseni, "Using deep convolutional neural networks with adaptive activation functions for medical CT brain image Classification," *2018 25th National and 3rd International Iranian Conference on Biomedical Engineering (ICBME)*, 2018, pp. 1-6, doi: 10.1109/ICBME.2018.8703546.

[15] H. Wang, H. Jia, L. Lu and Y. Xia, "Thorax-Net: An Attention Regularized Deep Neural Network for Classification of Thoracic Diseases on Chest Radiography," in *IEEE Journal of Biomedical and Health Informatics*, vol. 24, no. 2, pp. 475-485, Feb. 2020, doi: 10.1109/JBHI.2019.2928369.

[16] Lonseko ZM, Adjei PE, Du W, Luo C, Hu D, Zhu L, Gan T, Rao N. Gastrointestinal Disease Classification in Endoscopic Images Using Attention-Guided Convolutional Neural Networks. *Applied Sciences*. 2021; 11(23):11136. https://doi.org/10.3390/app112311136

[17] Eyal Klang, Ana Grinman, Shelly Soffer, Reuma Margalit Yehuda, Oranit Barzilay, Michal Marianne Amitai, Eli Konen, Shomron Ben-Horin, Rami Eliakim, Yiftach Barash, Uri Kopylov, Automated Detection of Crohn’s Disease Intestinal Strictures on Capsule Endoscopy Images Using Deep Neural Networks, *Journal of Crohn's and Colitis*, Volume 15, Issue 5, May 2021, Pages 749–756, https://doi.org/10.1093/ecco-jcc/jjaa234

[18] Ruan G, Qi J, Cheng Y, Liu R, Zhang B, Zhi M, Chen J, Xiao F, Shen X, Fan L, Li Q, Li N, Qiu Z, Xiao Z, Xu F, Lv L, Chen M, Ying S, Chen L, Tian Y, Li G, Zhang Z, He M, Qiao L, Zhang Z, Chen D, Cao Q, Nian Y and Wei Y (2022) Development and Validation of a Deep Neural Network for Accurate Identification of Endoscopic Images From Patients With Ulcerative Colitis and Crohn’s Disease. Front. Med. 9:854677. doi: 10.3389/fmed.2022.854677

[19] C. P. Sindhu and V. Valsan, "A novel method for automatic detection of inflammatory bowel diseases in wireless capsule endoscopy images," 2017 Fourth International Conference on Signal Processing, Communication and Networking (ICSCN), 2017, pp. 1-6, doi: 10.1109/ICSCN.2017.8085711.

[20] P. Wang, S. M. Krishnan, C. Kugean and M. P. Tjoa, "Classification of endoscopic images based on texture and neural network," *2001 Conference Proceedings of*

the 23rd Annual International Conference of the IEEE Engineering in Medicine and Biology Society, 2001, pp. 3691-3695 vol.4, doi: 10.1109/IEMBS.2001.1019637.

[21] Y. Qi, Y. Guo and Y. Wang, "Image Quality Enhancement Using a Deep Neural Network for Plane Wave Medical Ultrasound Imaging," in *IEEE Transactions on Ultrasonics, Ferroelectrics, and Frequency Control*, vol. 68, no. 4, pp. 926-934, April 2021, doi: 10.1109/TUFFC.2020.3023154.

[22] O. Oktay *et al*., "Anatomically Constrained Neural Networks (ACNNs): Application to Cardiac Image Enhancement and Segmentation," in *IEEE Transactions on Medical Imaging*, vol. 37, no. 2, pp. 384-395, Feb. 2018, doi: 10.1109/TMI.2017.2743464.

[23] Olaf Ronneberger, Philipp Fischer and Thomas Brox, U-Net: Convolutional Networks for Biomedical Image Segmentation, 2015.

[24] M. Rajchl et al., "DeepCut: Object Segmentation From Bounding Box Annotations Using Convolutional Neural Networks," in IEEE Transactions on Medical Imaging, vol. 36, no. 2, pp. 674-683, Feb. 2017, doi: 10.1109/TMI.2016.2621185.

[25] G. Koch, R. Zemel and R. Salakhutdinov, "Siamese neural networks for one-shot image recognition", *ICML Deep Learning Workshop*, vol. 2, 2015.

[26] V. A. Krishna, A. Reddy and D. Nagajyothi, "Signature Recognition using Siamese Neural Networks," *2021 IEEE International Conference on Mobile Networks and Wireless Communications (ICMNWC)*, 2021, pp. 1-4, doi: 10.1109/ICMNWC52512.2021.9688430.

[27] A. Abdalhaleem, B. K. Barakat and J. El-Sana, "Case Study: Fine Writing Style Classification Using Siamese Neural Network," *2018 IEEE 2nd International Workshop on Arabic and Derived Script Analysis and Recognition (ASAR)*, 2018, pp. 62-66, doi: 10.1109/ASAR.2018.8480212.

[28] W. Cui, W. Zhan, J. Yu, C. Sun and Y. Zhang, "Face Recognition via Convolutional Neural Networks and Siamese Neural Networks," *2019 International Conference on Intelligent Computing, Automation and Systems (ICICAS)*, 2019, pp. 746-750, doi: 10.1109/ICICAS48597.2019.00161.

[29] J. Fiaidhi, S. Mohammed and S. Fong, "Orchestration of Thick Data Analytics Based on Conversational Workflows in Healthcare Community of Practice," *2020 IEEE International Conference on Big Data (Big Data)*, 2020, pp. 4859-4865, doi: 10.1109/BigData50022.2020.9377848.

[30] Jenssen, B. H. (2020, March 28). Covid-19 radiology - data collection and preparation for Artificial Intelligence. https://medium.com/@hbjenssen/covid-19-radiology-data-collection-and-preparation-for-artificial-intelligence-4ecece97bb5b

[31] Jinyu Zhao, Yichen Zhang, Xuehai He and Pengtao Xie, "COVID-CT-Dataset: a CT scan dataset about COVID-19", 2020.

[32] Smedsrud, P. H., Gjestang, H., Nedrejord, O. O., Næss, E., Thambawita, V. L., Hicks, S., … Halvorsen, P. (2020, August 2). Kvasir-Capsule, a video capsule endoscopy dataset. https://doi.org/10.31219/osf.io/gr7bn

[33] Borgli, H., Thambawita, V., Smedsrud, P.H. *et al. HyperKvasir*, a comprehensive multi-class image and video dataset for gastrointestinal endoscopy. *Sci Data* 7, 283 (2020). https://doi.org/10.1038/s41597-020-00622-y

[34] Brigato, L., & Iocchi, L. (2021, January). A close look at deep learning with small data. In *2020 25th International Conference on Pattern Recognition (ICPR)* (pp. 2490-2497). IEEE.

[35] Schroff, F., Kalenichenko, D., & Philbin, J. (2015). Facenet: A unified embedding for face recognition and clustering. In *Proceedings of the IEEE conference on computer vision and pattern recognition* (pp. 815-823).