MAUU5900

MASTER THESIS

in

Universal Design of ICT

August 2021

Automated reporting system using deep convolutional neural network in the medical domain

Matrika Subedi

Department of Computer Science
Faculty of Technology, Art, and Design

OSLOMET

Abstract

Nowadays, in the healthcare sector, a massive volume of medical data sources is available. The data is growing at 153 Exabytes in 2013 and an estimated 2,314 exabytes in 2020 (Turner, Gantz et al. 2014). The medical data is composed of the patients' information, medications, follow-up, recommendation, and more other information. One of the medical sector's significant issues is medical experts' need besides their jobs to document and write different medical reports. On the other hand, machine learning plays a vital role in analyzing the large volume of medical data available in the healthcare sector to help diagnose and predict diseases. In the future, deep learning-based methods are considered the most promising machine learning methods in the medical field. To analyze and classify the medical images and present the medical reports based on these images/videos with minimum assistance from the medical staff and with high accuracy and some understandable form. There is a need to apply deep learning methods to produce an automatic documenting and reporting system. This thesis uses a deep convolutional neural network (CNN) to improve the automated reporting and documenting system of the gastrointestinal tract (such as inflammation and colorectal cancer). The report is designed under the principle of universal design, and accessibility procedure is also maintained to generate the report. The thesis also focuses on representing the internal processes of generating medical reports using CNNs.

Keywords: *Machine Learning, Deep Learning, convolutional neural network (CNN), automatic reporting system, and gastrointestinal tract(GI).*

Preface

This report summarizes the Master thesis work taken at the Department of Universal Design of ICT, Oslo Metropolitan University (OsloMet), Oslo, Norway, from 05 January 2020 to 15 August 2021. The workload is equivalent to 60 ECTS. The thesis is carried out under the supervision of Professor Pål Halvorsen and co-supervisor Steven Hicks. I would like to express my gratitude to them for their invaluable and profound guidance throughout the thesis period. Frequent meetings and regular discussions have been the foundation on which the completeness of this thesis work is built. I wish to thank everyone who has directly or indirectly motivated me to complete this thesis successfully. Your love, care, and, most significantly, ability to share knowledge with me is priceless. Finally, I thank my family and friends for their very special place in my life. My classmates have been so fantastic and thank you for all the lovely moments together.

Production note: I use Microsoft Word as the tool for writing this thesis. TensorFlow is used for model design, and Flask API is used for web deployment.

Matrika Subedi
15 August 2021
Oslo, Norway

Table of Contents

List of Figures

List of Tables

CHAPTER 1

INTRODUCTION

1.1 Background

"Machine learning generally represents the changes in devices that integrate duties associated with artificial intelligence (AI)." Such project includes recognition, diagnosis, planning, robotic control, prediction, etc. The "Alterations" can improve the systems still in operation (Nilsson Nils, 1998). Deep learning is a portion of machine learning methods based on artificial neural classification networks. CNN is the network most frequently used during visual image analysis(Schmidhuber, 2015). Presently, machine learning plays a significant role in evaluating the large volume of available medical data in the healthcare sector to identify and estimate disease accurately. Recent studies have often used deep learning in the medical field, for instance, in disease diagnosis.

We will concentrate on this gastrointestinal (GI) endoscopy dissertation, which utilizes some camera systems to identify an anomaly in the digestive tract. It performs a vital role in developing human existence while it begins to break down into nutrients. For instance, there are several diseases; for example, cancer is linked to the GI tract like ulcers, Crohn's cancers, bleeding, and polyps. Early diagnosis of the disease, particularly GI-related cancers, will help individuals sustain and retrieve. The most routine diagnostic technique used to diagnose GI cancer is endoscopy, which requires the doctor's ability to identify possible signs of cancer. So, there was a need to help doctors to perform diagnostic test procedures.

The essential steps of GI disease in the discovery of circumstances as

- Supporting documents
- Practice reviews
- Writing diagnosis

1.2 Research problem

The medical reporting system lacks professionalism and accessible understanding format and capabilities [Zane and Ronda, 2008] to help the healthcare provider save effort and time in writing distinct hospital records. This study utilizes CNN to enhance the automatic investigation and documentation system for the digestive tract.

With this rise in machine learning in the healthcare profession, an automatic and reporting system is needed to evaluate the findings with little help from medical staff and a high degree of quality and some justifiable type (Hicks, Eskeland et al. 2018). Report on patient history in the health industry is crucial. It aims to clarify and analyze various diseases to high precision and enables medical professionals to understand the disorders truly.

There are many challenges of Endoscopy reporting, according to (Can, 2013) which will be discussed later, such as generating reports that are now correlated with insufficient documentation, developing reporting methods for a colonoscopy that are highly varied and often substandard. Also, Reports display significant differences in the completion of various report components. So, computerized documentation and reporting strategies are essential to present prediction with minimal medical support and high accuracy with some tangible reports through machine learning in medicine. This research, therefore, aims to use a CNN to establish a deep learning model, a computerized reporting framework, and a supporting documents template in the intestinal system (GI).

This research will improve the automatic reporting and documenting system of the GI tract using CNN to presents the output of medical reports with minimum assistance from the medical staff and with a high level of accuracy and some of the understandable form which contains multimedia elements such as images or videos. It maintains the universal design principle to design the report. This will help for easy retrieval and processing of medical reports and documents besides reusing data for even teaching and research and giving visual representations of deep neural network layers to increase understanding, trust, and usefulness of disease diagnosis and detection procedures.

1.2.1 Research objectives

1) Studying the existing works shows the advantages and disadvantages of utilizing CNN in the medical industry of automatic reporting procedures.
2) Using CNN in the GI tract to detect anomalies.
3) Generating justifiable examination of identification reports in diagnosing the GI tract to visual effects and process contributes.
4) Improving the existing automatic process's effectiveness to provide medical reports for the GI tract through the
5) Accessibility checks of the web-based report system.

1.2.2 Research Limitations

The GI tract is a broad region and includes several diseases and various types of cancers. Here we took only 15 classes because of the availability of datasets. So, we limited the number of categories to a manageable number. Shortage of medical datasets to be used and analyzed for model training. Also, the existing health datasets need many preprocessed actions to be ready to implement and classify. We used Google Collaboratory to run the program to access the free GPU. Goggle only provide 12hr runtime, so this leads us to runtime problems too.

There are some other limitations to developing and validating deep learning algorithms in the medical field. Data bias is one such constraint. Many algorithms are biased against the training data if they are not thoroughly balanced. Even when cross-validation approaches are used, there is still a bias in favor of the dataset. Since decisions on enhancing the network will be focused on samples from the project's larger dataset.

The lack of sufficient data of the complexity in obtaining the necessary data is a disadvantage in clinical image analysis. Patient confidentiality must be respected, and images collected from medical imaging instruments are expensive per scan. Medical professionals can only access medical photographs; the general public cannot obtain them. These data resource constraints are a problem that will require the collaboration of medical professionals, patients, and scientists to solve. Also, we cannot verify each system's image quality as we won't have access to an illustrated masking dataset for HyperKvasir. For the designing process also it's difficult to maintain all universal design principles in one. Here we faced mainly placeholder problems to make the report accessible.

1.3 Research Methods

There are many types of research methods among them; two types are commonly used:

1. A quantitative approach is used by creating numerical data or converting data into meaningful statistics to solve the problem. For observable results, this technique of constructing factual information and discovering patterns in studies is being used. Methods for gathering quantifiable information include various types of surveys, documentary analysis, and interviews (Apuke, 2017).

2. The primarily qualitative methodology evaluates science, providing helpful information to know the reasons, perspectives, and motivations behind it. It provides some insight or helps to generate ideas or theories on this issue. Non-experimental quantitative work offers a definition of observable data and the discovery of a potential connection of occurrences. In a qualitative study, the member wishes to cautiously monitor all potentially valuable information accurately and objectively using a notebook, audiotapes, drawings, photos, and other appropriate means. Popular strategies involve focus groups, personal interviews, and involvement(Hennink, Hutter, & Bailey, 2020).

Our studies will use a quantitative strategy to reach this project's aim using a design concept for constructing a system (software) to produce more accurate and easy-to-understand hospital records using CNN to evaluate visual image information of the GI tract. The steps for designing the reporting system are as follows:

i) define the system requirements (technical and non-technical requirements)
ii) determine the system specifications
iii) create the experiment steps and implement the system
iv) evaluate the system performance and test the system's version to implement the final plan in a real medical environment.

1.4 Contributions

This research will improve the automatic reporting and documenting system of the GI tract using CNN to presents the output of medical reports with minimum assistance from the medical staff and with a high level of accuracy and some the understandable form which contains multimedia elements such as images or videos. This will help for easy retrieval and processing of medical reports and documents besides reusing data for even teaching and research and giving visual representations of deep neural network layers to increase understanding, trust, and usefulness of disease diagnosis and detection procedures.

We should follow the following rules or requirements while implementing the automatic reporting system of using CNN for the GI tract:

1) The system should be easy to understand by technical and non-technical users (how to use CNN in reporting system).
2) Everyone benefits from universal design, not just the aging population or the disabled.
3) The system should use the images dataset GI tract to produce the medical documentation.
4) The system should be easy to execute and use in real medical environments.
5) The system can improve the performance of the current reporting system effectively.

We have achieved the following objectives:

1) *Studying the existing works and show the advantages and disadvantages of utilizing CNN in the medical industry of automatic reporting procedure.*

This objective is supported by chapter 2 and chapter 3. In chapter 2, we discuss a literature review where we discuss in detail about prior work, advantages, and disadvantages of the CNN model in the medical domain. In chapter 3, we discuss the requirements of a reporting system.

2) *Using CNN for the GI tract to detect anomalies.*

   This objective is supported by chapter 4. Chapter 4 developed the three CNN model, which helps to detect anomalies in the GI tract. Out of the three models, we select DenseNet121 because of its higher accuracy compared to the other two models.

3) *Generating justifiable examination of identification reports in diagnosing the GI tract to visual effects and process contributes.*

   This objective is supported by chapters 4 and chapter 5. In chapter 4, we discuss how to develop the report system using a web application, and in chapter 5, we see the generated output of the report. Here we use the flask to develop the web application.

4) *Improving the existing automatic process's effectiveness to provide medical reports for the GI tract through CNN.*

   This objective is supported by chapters 3 and chapter 4. In chapter 3, we discuss the drawbacks of using CNN and the requirements for the report system. In chapter 4, we developed a new model to detect anomalies in the GI tract.

5) *Accessibility checks of the web-based report system*

   This objective is supported by chapter 5. In chapter 5, we try to consider universal design approaches to make a website and check the accessibility of the generated report. We check our web report with WAVE, which is a web accessibility tool.

1.5 Thesis Outline

Chapter 1: Introduction

This chapter presents the research problem, theoretical background, the research objectives, the research questions, the research limitations, the methodological approach, and the significant points of the thesis contributions.

Chapter 2: Using Deep Convolutional Neural Network in Automatic Reporting for Gastrointestinal Tract.

This chapter discusses the research background, including machine learning (supervised learning, unsupervised learning, active learning, and deep learning, focusing on convolutional neural networks. This literature review chapter is split into two parts; the first part presents related works of using CNN in the medical field. The second part will discuss the current research using CNN in Automatic Reporting for the GI tract as a case study. Also, the design principle of universal design with respect to the website.

Chapter 3: Requirements for An Automatic Medical Reporting System

This chapter presented the automatic reporting system in the healthcare sector, including the Automatic Medical Reporting system. Its advantages and benefits are focusing on the GI reporting system, especially the requirements and recommendations of a GI reporting system.

Chapter 4: Deep Convolutional Neural Network CNN in an Automatic Reporting System for Gastrointestinal Tract.

This chapter discusses the methodology steps of using CNN to enhance the automatic reporting and documenting system for the GI tract trained to the HyperKvasir dataset. This chapter described how to collect the data and different used datasets, preprocessing steps; training and testing the model, various used network architecture, and introducing the proposed system viewing hyperparameters utilized in training. Then, Web Interface was discussed, representing the backend and frontend tools and the automatic generation tool of GI tract reports, and finally, the model performance metrics are examined.

Chapter 5: Results and Discussions

This chapter applies an automatic reporting system of medial documents used the CNN on selected image datasets of the GI tract and discusses results and evaluates the current automated reporting system's improving performance. Accessibility checking of the report based on accessibility guidelines WCAG 2.0.

Chapter 6: Conclusion and Future Works

Finally, this chapter discusses the summary and the conclusion of the thesis and explores future works.

Chapter 2

Literature Review

2.1 Artificial intelligence

Artificial Intelligence now plays a vital role in many areas such as image recognition, robotics, and translation software. Further, to simulate in the medical sector, particularly disease diagnosis and detection. This chapter gives the research background, such as machine learning (supervised, unsupervised, active, and deep learning) concentrating on CNN. This section provides a literature review divided into two aspects: the first part describes the overview of using CNN in the health care profession. The second aspect explains the recent research of using CNN for the GI tract for automatic reporting.

2.1.1 Machine Learning

Machine learning is a portion of artificial intelligence (AI) that allows computer learning to perform through training analyzed with different techniques. The primary machine learning algorithms are considered supervised, unsupervised, and enhanced learning algorithms.

“A computer program is also said to learn from knowledge E with a certain class of assignments T and system performance. P, if its efficiency in T assignments, as evaluated by P, enhances with E knowledge. " (Mitchell, 1997)

2.1.2 Supervised learning

Supervised learning is a group of methodologies that use labeled information to analyze it. By enforcing an iterative method, where the technique anticipates a given sample, its internal weights alter adequately based on how inaccurate its prediction will be. Linear regression and classification methods are the most renowned supervised algorithm. Linear regression is usually used to anticipate, forecast, and detect quantitative data relationships. Classification methods are concentrated on forecasting a qualitative response via statistical analysis and pattern recognition. The most used classification methods are Logistic regression, Decision tree, K-nearest neighbors, Neural network, and support vector machine.

2.1.3 Unsupervised learning

Unsupervised learning is a community of methodologies that use unmarked information to analyze from it. This is designed to detect patterns in data presented instantly and conducting some operations such as clustering. It has traditionally been the most crucial technique of unsupervised machine learning. This algorithm analyses unmarked data from standard practices among data points and instantly groups similar data into clusters. Large, unmarked image datasets could be used for application areas of these methodologies.

Techniques based on unsupervised learning are as follows:

- Clustering tends to help to split the dataset into clusters by correlation instantly. Cluster analysis also overstates the relationship among groups and does not view data points as unique.
- Detection of anomalies: unusual data points can be detected within the data set instantly. It helps in detecting fraudulent transactions or detects outliers throughout data entry due to human error.
- Association mining identifies dataset compilations of items that often take place together.
- Latent variable approaches are frequently used only for preprocessing data, for example, to reduce dataset characteristics.

2.1.4 Reinforcement learning

This machine learning maximizes some payoff in particular circumstances(Arulkumaran, Deisenroth, Brundage, & Bharath, 2017). Several reinforcement applications are business applications, robotics, training systems, data processing, robot motion control, and aircraft control. It ranges directly from supervised learning and not from the behavior of sub-optimal. As an alternative, the importance is based on the discovery of balancing and exploitation of the knowledge base. Dynamic software is used for reinforcement learning. The typical difference between the classical and the reinforcement is that reinforcement learning does not need to know an exact mathematical paradigm and attempts extended Markov Decision Process (MDP) where techniques are unrealistic.

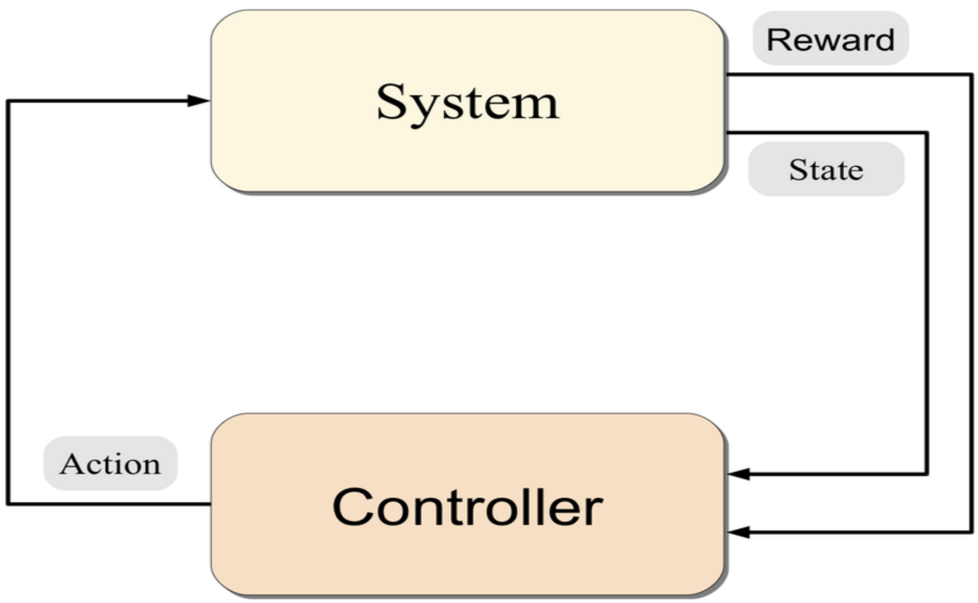


Figure 2. 1 Reinforcement learning scenario (Szepesvári, 2010)

Some elements of reinforcement AI are shown in the above figure, such as

Agent- It is an entity that conducts actions in the circumstances of receiving some recognition.

Environment-A case that an agent has to tackle

Reward-An instant return is provided to an agent when concludes a particular action

State-signifies the current environmental condition.

2.1.5 Active Learning

This learning is permitted to get from a pool of unlabeled instances and the subset of feasible examples to be labeled next, too. The principles that preceded the concept are that machine learning may attain higher accuracy while using fewer training labels to select the required data. During the training process, active learners can ask queried, typically in unlabeled data, to be labeled through what is known as an oracle dynamically.

- Pool-based sampling: This endeavors to experiment with the whole dataset priorly choose or gather the best query. Initially, the active learner is usually trained on a wholly labeled fraction of the knowledge that generates a first model approach used to classify the circumstances that would be most beneficial for the next iteration to be introduced into the training iteration.
- Membership query synthesis: The learner will establish instances of their labeling. This strategy aims to solve challenges with a cold start when it is easy to build a data instance beginning from none.
- Stream-based selective sampling determines whether investigating into single unlabeled admittance is relevant for the label in the dataset. When the approaches are to be trained and determined with data, it adopts quickly whether it is required to view the spot or not.

2.1.6 Deep Learning

It is a broader class of algorithms that employ directly to ANNs and import computers to do what emanates from a human. And it is a vital technology backend for driverless cars, allowing them to detect a stop sign. It is confidential for voice control in various systems such as laptops, phones, and televisions. (Deep reinforcement learning) DRL utilizes deep learning and reinforcement aspects for constructing better algorithms deployed to education, robotics, transportation, video games, finance, and healthcare. It can acquire as input from raw sensors and image signals(François-Lavet, Henderson, Islam, Bellemare, & Pineau, 2018). In this chapter, we focus on deep learning CNN in the form of image classification.

2.1.7 Neural Networks

Neural networks are hypothetical models concentrated with a loose knot on the neural structures that grow animals' brains. Traditionally, this mechanism has paid attention to neural networks' production to generate better machine learning performance outcomes. An artificial neuron is the essential element of neural network evaluation. It may usually comprise millions and hundreds of neurons that every neuron performed to rectify the specific issues. It is significant to know how each neuron operated inside to profit insight into why it works well.

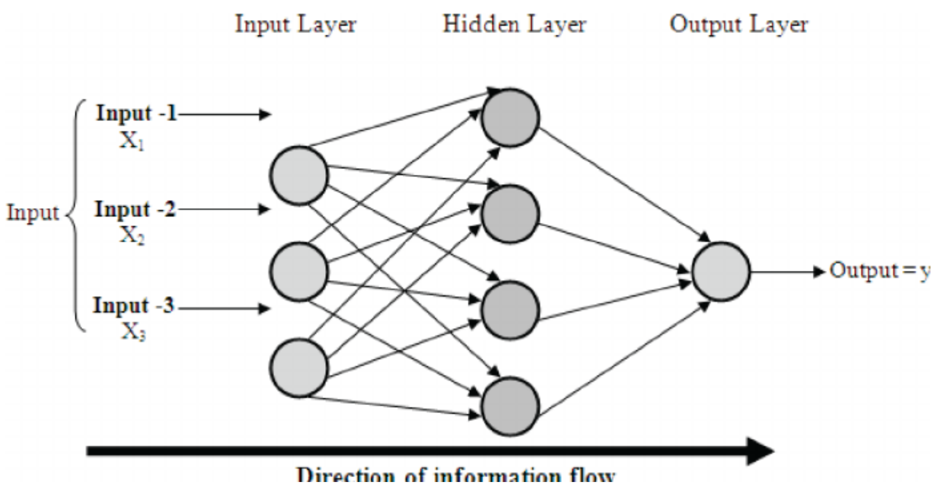


Figure 2. 2 A three-layered multilayered perception on neural network (Kabir & Hasin, 2013)

Figure 2.2 represents a generic neuron. For instance, it took weighted inputs presented through three vectors mark to the neuron. It performs a weighted sum of its information with a further biased term. It has the function of running production in a positive or negative direction. This weighted sum product is then transferred to a process that is anticipated for adding some non-linearity to the outcomes. This function reveals that what fixes the neurons beyond each other. After the activation function is transformed, the result is either sent to the next layer's neuron.

As shown in figure 2.2, the input layer is where data is transformed with the network and goes forward to the first invisible layer. This layer does not have well-known parameters or evaluations. The output layer is a classifier that involves one node per classification—the invisible layer situated between the layers of input and output. The data is always injected into the input layer, transfers through the hidden layers, and result in the output layer.

This puts together the MLP and the feed-forward class of neural networks. We need to measure how inaccurate the predicted network performance is for studying supervised neural networks. This has been done in many ways and is known as the loss function of a network. The technique of calculating the loss in a modern neural network utilizes the cross-entropy algorithm that evaluates the difference between the expected outcome and the actual ground truth.

$$C(x, y) = -\sum_i x_i \log y_i \tag{2.1}$$

Cross-entropy is measured in equation 2.1, where x signifies the predicted result, y denotes the ground truth, and we describe the class. Then optimization function is utilized to reduce the loss; the most depicted algorithm is the gradient descent algorithms and their variants (Hicks et al., 2018).

There are three Gradient Descent as follows,

- Stochastic- It is a method of gradient descent that process per repetition processes one instance of training. Subsequently, after a single iteration, only a single example has been handled to modify each parameter.
- Batch-It is a gradient descent that works with all the training instances in each gradient descent iteration.
- Mini batch- It performs better than the set and stochastic gradient. When the number of training samples improves, they progress in one go for the collection of pieces. Here b and m are instances in which b<m is moved per iteration.

2.1.8 Convolutional Neural Networks

A CNN is a deep learning neural network class utilized for analyzing the visual imaginary. It is a feed-forward neural network that could study in the same means. It uses grid-like topology, creating it particularly adjust during processing data of multiple dimensions like samples taken at regular intervals or the pixel dimensions for an image. It helps identify patterns to recognize faces, objects, and scenes in photos and know directly from image data using practices to detect images and eliminate feature extraction requirements. The application depends on CNN, like computer vision and object recognition, including self-driving vehicles and face recognition apps. And other applications get support from CNN, such as recommender systems, identifying the image and video, image classification, medical image analysis, and natural language processing.

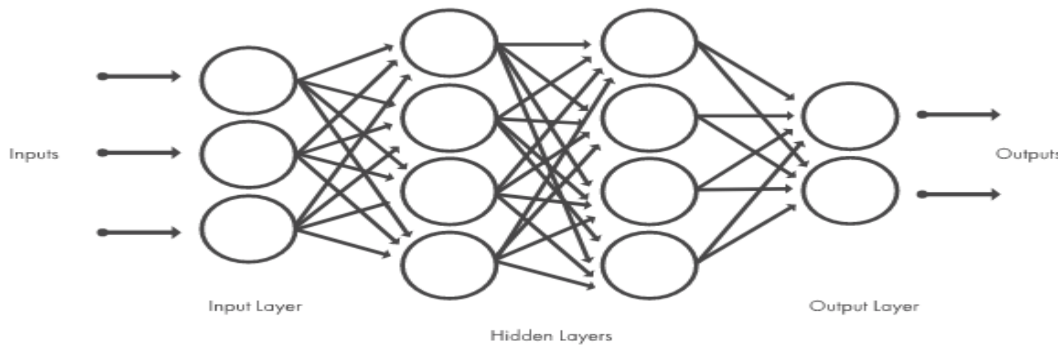


Figure 2. 3 *C*omponents of convolutional neural network

From figure 2.3. CNN comprises an input, output, and many hidden layers in the middle. The most used layers are activation, convolution, and pooling layer.

- Convolution locates the input image using a sequence of convolution filters, which allows some features of the picture.
- The rectified linear unit allows faster and more effective preparation by mapping negative values to zero and holding positive values. It is often known as activation. Meanwhile, only the active features are transmitted to the next layer.
- Pooling shortens efficacy using non-linear downsampling and mitigates the number of parameters the network needs to understand.

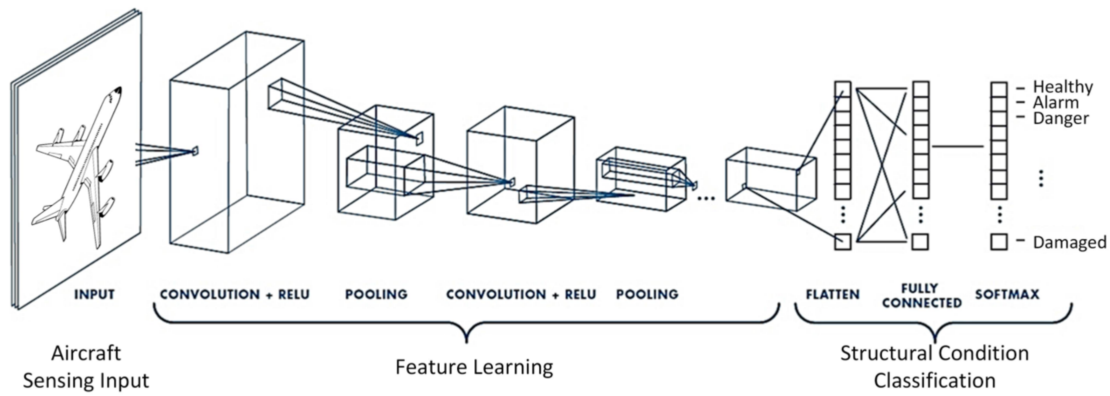


Figure 2. 4 Example of a convolutional neural network (Tabian, Fu, & Sharif Khodaei, 2019)

Figure 2.4 gives an instance with many convolutionary layers of a traditional neural network. Filters are deployed for each training image, and the data from each transformed image is utilized as the actual input to the next layer in various resolutions. Such a process is iterated over tens or hundreds of layers, identifying several characteristics based on each layer. CNN's architecture is then run towards classification after learning features in various layers. The preceding layer of the final layer is a fully connected layer that results in a K-dimension matrix. K is the number of classes that a network can anticipate. Each vector involves the probability of each image being valued for each degree. The final layer of CNN architecture utilizes a classification layer to generate the classification performance.

2.2 The Gastrointestinal –GI Tract

The respiratory system includes the intestinal system recognized as the GI tract—and the liver, pancreas, and gallbladder. The Digestive system is a series of hollow organs decided to join together in a while, trying to twist the tube from the mouth to the anus, as shown in Figure 2.5.

The empty organs that develop the GI tract are the esophagus, the mouth, the stomach, the anus, the small and large intestines. The pancreas, gallbladder, and liver are also the solid organs of the digestive tract. This tract may have various diseases that involve inflammation, infection, and cancer. Colorectal cancer (CRC) is a chronic disease that amounts to roughly 10 percent of all cancer cases. The serious trend with the CRC is that it usually does not have strong signals until it is too late. Thus, it is critical to monitor the GI tract for illness and pre-CRCs periodically.

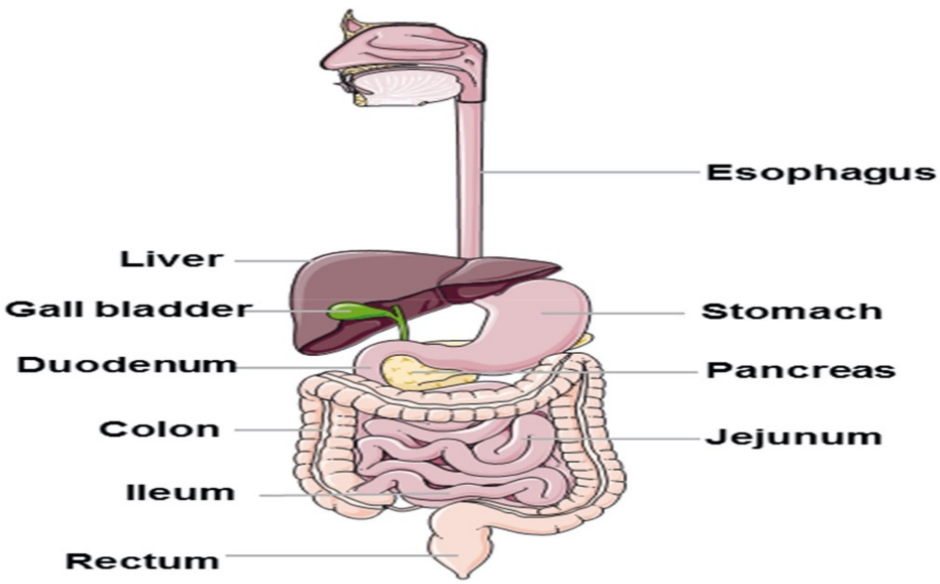


Figure 2. 5 The digestive system (Vertzoni et al., 2019)

The digestive process begins with the starting process using the mouth and food transferred with the esophagus, such as a hallow-like tube, then the stomach observed the food and mixed it using enzymes and acids. The food is then transferred to the small intestine, which combines food with the digestive fluids from the pancreas, intestine, liver and sends the mixture forward for the following process. This process moves the waste products into the large intestine to extract the water, salts, sugars, and vitamins from the indigestible food—finally, the digestive process is completed with the anus, where the remaining products were sent to the anus. GI endoscopy determines the GI tract using an endoscope for detecting various disease and abnormalities shapes. Endoscopes are injected into the organ directly. It is attained using either the mouth –gastroscopy) or anus (colonoscopy), a long elastic tube linked with a tiny camera. The general information is assumed sufficient, but difficulties may happen, even life-bullying, in more complicated cases.

Endoscopies are intrusive, expensive, the happening of high anxiety for many persons. It is often time-consuming and needs one doctor hour and two nurse hours, respectively. Also, colonoscopies have a greater rate of missing because of the inability of endoscopists to detect polyps. This method is termed (video capsule endoscopy) VCE. It is a key to conventional endoscopy scalability that causes inefficiencies. It is a tiny camera bound in a capsule of vitamin size taken through the mouth and transmitted using the GI tract. The tablet was armed with batteries, image sensors, bleeding sensors, wireless transceivers, antennas, and light sources.

2.2.1 Challenges with Endoscopy

The following are some of the challenges of Endoscopy reporting according to (Can, 2013)

1. While narrative reporting was once common, it is now correlated with insufficient documentation, inconsistencies in documenting positive results, relevant negative consequences, and other procedural details.
2. The high workload for the endoscopists; insourcing and Outsourcing (sending patients offsite to an outside endoscopy service for their investigative process).
3. Physicians' reporting methods for colonoscopy are extremely varied and often substandard. Variations in disease classification, such as identifying and quantifying inflammatory processes in colitis, are examples of inconsistent endoscopy documentation.
4. Reports also display significant differences in the completion of various report components. Demographic details, medical history, preparation reliability with visualization, and process understanding are the most frequently low-scoring products. Lesion diagnosis and elimination and sedation practice are two other report components that differ significantly.

2.3 Universal Design, Accessibility, and Usability

Accessibility is creating solutions that serve the needs of all users, including those with and without limitations. It is the correct thing to do for all consumers, not just persons with impairments. A website that is accessible too is considerably more useable for everyone, according to Web developers.

2.3.1 The Process of Universal Design

The following points proposed the process steps of Universal Design (Sheryl, 2021):

1) Indicate the product or context to which universal design should be applied.
2) Define the main users as well as the various characteristics of potential users whom the application is constructed (e.g., teachers, professors, and medical staff with varying gender, age, and race, the local language, learning style, and skills to see, listen, manipulate, read, and write).
3) Individuals with various characteristics (as mention in the previous step) should be considered and included in all aspects of the website development, implementation, and evaluation. Diverse activities, such as the university disability services office, might also help you discover new perspectives.
4) Create new universal design guidelines/standards or choose from existing ones. Combine these with other industry best practices in the sector of the application, for example, the medical sector.
5) Follow a set of guidelines or criteria. To maximize the potential benefits of the website to people with a range of characteristics, implement universal design with best practices inside the sector to the overall design of the website, all subcomponents of the website, as well as all operational activities (e.g., purchasing processes, training programs).
6) Make arrangements for accommodation. Develop procedures for dealing with requests for accommodations (e.g., the procurement of assistive devices, the scheduling of interpreters) from users for whom the website design does not immediately enable access.

7) Train and help. Provide ongoing assistance and development to participants by personalizing it and delivering it (e.g., computer support staff). Discuss the organization's objectives for inclusion and diversity, as well as methods for

ensuring that everyone has a welcome, accessible, and comprehensive experience.

8) Evaluate. Integrate universal design metrics in frequent reviews, evaluate the website with a wide set of users, and make changes depending on their comments. Provide methods for users to provide feedback (e.g., online communications with administrators).

2.3.2 Standards of Universal Design

The web Page accessibility standards are recognized as the Web Content Accessibility Guidelines, or WCAG2.0, which are available through the W3C's Web Accessibility Project. The World Wide Web Consortium (W3C) is a non-profit organization that creates and maintains web standards. The World Wide Web Consortium (WAI) has established criteria for all websites.

The Web Accessibility Initiative (WAI) of the World Wide Web Consortium (W3C) serves as the focal point for establishing Web accessibility rules. Its work is focused on the development of Web Content Accessibility Guidelines (WCAG2.0), which have two main objectives (Matera, 2006):

- Creating perceivable and operable content: this involves employing simple and straightforward language as well as specifying navigation and orientation techniques to allow content access and surfing.
- Providing access alternatives: sites must be structured and programmed in such a way that they may be accessed regardless of the browser tools and systems used, as well as the usage setting.

The first goal is directly related to the concept of Web usability; it can be accomplished by concentrating on usability criteria that improve the efficacy and efficiency of navigational and orienting mechanisms. The second goal can be accomplished by modifying the page mark-up, specifically:

1- Trying to separate presentation from material and navigation, which allows for the presentation of the same content material and navigation systems commands in various presentation modes that are appropriate for different equipment.

2- Adding textual explanations to multimedia content so that it can be viewed via alternate surfing technologies like screen readers for blind users.

3- Creating texts that can be viewed on a variety of hardware platforms. It should, for example, be important to communicate with page content using speech devices, tiny and dark screens, and without the use of pointing devices.

2.3.3 Universal Design Principles

The following are the main principles of the Universal Design according to (Sheryl, 2021):

1) Usage that is appropriate. People of various capacities will find the design helpful and desirable, for example, on the website that is created to be accessible for everybody, including persons who've been blind and use screen reader software.
2) Flexibility in use. Individual tastes and abilities are accommodated by design. For example, on the website might provide doctors the option of reading or listening to an explanation of a medical case.
3) Simple and easy to use. Independent of the user's experience, expertise, language abilities, or existing concentrations, the design is must be simple to use.
4) Information that is visible. Independent of environmental conditions or the participant's sensory abilities, the design layout should efficiently transmit important information to the user. Detailed medical images of the body on the website are an illustration of this approach.
5) Tolerance for mistakes. The design reduces risks and the harmful acts of accidental activities. The website provides advice when a user makes an incorrect choice is an example of this approach.
6) Physical activity is minimal. The design allows for efficient, comfortable, and fatigue-free operation.
7) Size and space for use. Independent of the user's physical appearance, attitude, or movement, adequate size and space are given for this approach, including

reaching, manipulation, and use. This principle has been applied to a flexible work environment created for employees who really are left–hand or right-hand and have a range of other physical features and abilities.

2.3.4 WAI-ARIA

The Accessible Rich Internet Apps Suite, or WAI-ARIA, is a set of guidelines for making Web content and applications more accessible to individuals with impairments. It's notably useful for developing dynamic content and complex user interface controls using HTML, JavaScript, and other related technologies.

The Accessible Rich Internet Apps Suite, or WAI-ARIA, is a set of guidelines for making Web content and applications more accessible to individuals with impairments. It's notably useful for developing dynamic content and complex user interface controls using HTML, JavaScript, and other related technologies. (WAI-ARIA, 2021)

Some Web site function is unavailable to some disabled users without WAI-ARIA, particularly for those who use screen readers or who are unable to operate a mouse. For example, WAI-ARIA solves these accessibility issues by describing how assistive technology might be given functionality. Developers can use WAI-ARIA to make advanced Web apps accessible and useable for disabled individuals.

WAI-ARIA creates a basis for adding attributes to define user interaction features, their relationships, and their present state. WAI-ARIA defines navigation algorithms for identifying areas and common Web structures such as menus, principal content, supplementary content, banner info, and other Web elements. Developers can utilize WAI-ARIA to identify parts of sites and make it easier for keyboard users to travel between them instead of pressing Tab repeatedly.
Web authors can use WAI-ARIA to get the following (WAI-ARIA, 2021):

- Functions such as "menu," and "sliding," are used to describe the sort of widget that is displayed.

- Headers, sections, and tables are examples of roles that characterize the organization of a Web page.
- Widgets have properties that define their current state, such as "confirmed" for a check box and "has popup" for a menu.
- Features to identify active sections of a website that are expected to get improvements and an interruption strategy for those changes—for example, crucial updates may be described in an alert dialogue box, while incidental updates happen on the page.
- A method of allowing keyboard navigating of Web objects or events.

2.3.5 Usability

The usability of a thing refers to how simple it is to use. A device, equipment, method, application software, or website can all be used as the object. Everything that may be used by a person must be usable. Usability has been described in the context of websites and software as the ease with which an average individual can use the software or website to accomplish particular goals.

According to the Nielsen's, usability helps to achieve (Matera, 2006):

- Learnability: the ease with which the website operation and behavior can be learned.
- Productivity: the level of productivity that can be achieved once the user has mastered the website.
- Memorability: the ease of remembering the website functionality so that the casual user can return to the system after a period of non-use without needing to learn again how to use it.
- Few errors: the website's ability to have a minimum error, to support customers making few errors while using the website, and to assist them in recovering quickly if they do make errors.
- User satisfaction: the degree to which the user enjoys using the system.

2.4 Existing work in Using Deep learning CNN for disease diagnosis and disease detection

Much research has been done based on CNN in the medical domain, as explained below. From these existing followers chose one of the most recent trends and proposed that detailed research as methodology:

In (M. Riegler et al., 2016) and in (Pogorelov et al., 2017), the authors presented a multimedia system to automate the videos from the individual GI tract. The framework covered the complete data collection, retrieval, and analysis process, as well as visualization. The framework integrated machine learning filtering, image recognition, and image feature extraction. This system was designed for quick processing so that the doctor can get real-time feedback. The proposed plan had detection and localization accuracy that is as good as current systems, but it can be used in real-time performance.

In (Jha et al., 2021), the authors presented a thorough review of the current Medico GI works. These challenges include establishing a benchmark for various computer vision techniques implemented to multi-class endoscopic images and encouraging new designs to be used effectively in clinics. This work presented the results of 21 teams involved over three years. It included a thorough review of the procedures used by the participants, outlining the challenges and flaws of existing methods and examining their validity for use in medical practice. According to the findings, Participants improved their overall Mathew correlation coefficient (MCC) from 82.68 percent in 2017 to 93.98 percent in 2018 and 95.20 percent in 2019 tests and a substantial increase in computational speed over consecutive seasons.

In (M. Riegler et al., 2017), the authors present a holistic multimedia framework to automate video analysis from GI endoscopy. This holistic view and its real-time efficiency completed the algorithmic GI screening process. The proposed device was scalable so that it could be quickly expanded to evaluate various anomalies, and I made it robust so that it could run in real-time.

In (Esteva et al., 2017), the researchers deployed skin lesions classification using a single CNN, accomplished end-to-end directly from the images. And CNN was trained on the given dataset of clinical images (129,450), comprising 2,032 different diseases. The performance was evaluated based on biopsy-proven medical images against 21 board-certified dentists with two significant use instances for binary classification. The first review demonstrates the most severe tumors, and the second case is skin cancer.

The authors (Abiyev & Ma'aitah, 2018) revealed the availability of chest pathologies classification in chest X-rays using convolution and deep learning methods. The chest disorder was diagnosed using CNN, and for comparison purposes, backpropagation neural networks with supervised learning and competitive neural networks with unsupervised learning were used. All the networks are trained and tested on the chest X-ray dataset; the data were observed from the National Institute of Health- Clinical Center. The dataset consists of 30 805 patients with 112, 120 front-view X-ray pictures. The quality of the networks hired was assessed using performance measures such as error rate, accuracy, and training time.

The researcher (Yadav & Jadhav, 2019) utilized the CNN for classifying pneumonia on a chest X-ray dataset through a linear (support vector machine) SVM classifier. Here, transfer learning is based on the CNN model: Visual geometry groups like InceptionV3, VGG16, and capsule network training. The dataset comprises two kinds: NORMAL (bacteria) and PNEUMONIA(virus), for chest X-ray images. There are 5232 X-ray images, while the testing dataset involves 624 appearances in the training dataset. They used a data preprocessing method known as data augmentation in three aspects. The authors compared the accuracy with a boost as well as without augmentation. They utilized precision, recall, and specificity for evaluating the outcomes of the experiment. The results revealed that data augmentation is an effective way to enhance efficiency, and transfer learning appears to be more relevant on a restricted dataset than SVM. Also, re-trained specified features based on a new target set are necessary for transfer learning.

The researchers (Gamage, Wijesinghe, Chitraranjan, & Perera, 2019) Used a set of in-depth features as a single vector feature by incorporating pre-trained DenseNet-201, VGG-16, and ResNet-18 models as a feature extractor followed by the (global average pooling) GAP layer to anticipate eight class abnormalities of digestive system disorder. Here, the KVASIR dataset contains endoscopic imagery from the GI tract internal part with eight different classes. The authors used preprocessing steps, and the next step is for a group of noticeable CNNs with a global average pooling layer to observe feature vectors. Then, feeding function vectors in a single hidden layer of ANN comprises 128 hidden units with a kernel function of ReLU. The researchers used evaluation metrics to test the outcomes: accuracy, precision, recall, and F1 ranking. The result showed an accuracy of 97%.

In (Vu et al., 2019), the author established an assistance method for diagnosing the (Upper GastroIntestinal Endoscopy) UGIE analysis for labeling stomach anatomical regions. The authors implemented two-step research methods; the first is to use CNN to categorize seven central anatomical areas: gastric body, antrum, fundic, pyloric ring, and lower and higher curvature. The second step is to create a GUI so that endoscopists can quickly describe anatomical locations. The regions are enclosed through 35 images. The images collected had been analyzed through software and then in the form of anatomical areas, and lesions were detected and labeled with specialists' help. Two thousand six hundred sixty-seven images are gathered for 108 patients. Such metrics measure the performance as accuracy metric and diagnostic time. The result revealed that the method decreased the diagnostic time.

In (Arjmand et al., 2020), the researcher showed a fully automated classification method considering the high potential to discriminate four modifications of histological issues. A highly supervised learning strategy with a CNN was used in their present design. The data set contains 64 defined samples and is digitized using a Hamamatsu microscope. In addition to performing linear classification and accuracy, the suggested efficiency was shown to be quantifiable.

In (Sharif et al., 2019), the methodology was used on a mixture of CNN and geometric features. Initially, infection areas are noted from some pictures of WCE with the assistance of a new technology called improved color features. Geometric characteristics have been pointed out from the partitioned region of the disease. The parts are conducted based on special VGG16, VGG19, and Euclidean fisher vectors. Specific elements were fused with geometric feature vectors to choose the best characteristics nourished to the conditional entropy approach. Finally, to detect the feature selection, the researchers used a K-nearest Neighbor. To test the proposed process, they use 5500 WCE images and achieved a classification accuracy of 99.42 % and an accuracy rate of 99.51 %.

In (Xu et al., 2019), the authors have developed multi-task anatomy detection using CNN to evaluate the EGD inspection's effectiveness by incorporating the upper digestive tract detection with ten anatomical illustration classification techniques besides informative video blocks. The proposed method has been used to remove non-informative gastroscopic video frames and define live time and video frames. A sub-branch for categorizing NBI images, informative and non-informative images, was explicitly passed to the classifier. This result detected box also on informative structures, which decreases the (false positive rate) FPR. Evaluating the diagnosis's coherence calculates the video frames on which each anatomical role was efficiently diagnosed. They asked two experienced endoscopists for the detection task to mark 60233 and 40145 gastric images for classification tasks. They used mean average precision (mAP) to measure their performance model with detection of 93.7% and accuracy of 98.7% for the classification.

In (Pannu, Ahuja, Dang, Soni, & Malhi, 2020), the researchers presented a supervised learning ensemble to identify the bleeding for wireless capsule endoscopy images. The researchers utilized augmentation to practice CNN throughout the input phase. The small bowel lesion dataset contains two sets as 3,295 + 600=3895. Significant improvements have been proved through thoughtfully establishing the choices for the CNN layer and optimizing backpropagation after decreasing the color palette utilizing minimal variance quantization. The proposed set got 0.95, and the video dataset got 0.93 % accuracy which is better.

In (Osamu et al., 2020), the authors have enacted CNN and RNN on biopsy histopathology of stomach and colon full side images (WSIs). The designs were developed to spot WSI as non –neoplastic, adenocarcinoma, and adenoma. The researchers used ROC curves to analyze the performance, accomplishing an (area under curves) AUCs of up to 0.97 with 0.99 for gastric adenocarcinoma and adenoma and 0.96 with 0.99 for colonic adenocarcinoma and adenoma in both.

From the existing research, some drawbacks are using deep learning CNN in the medical field, particularly in the GI tract, as follow:

- Increase some consequences to identify some health diseases such as lesion shape, color, texture, size, and irregularity.
- According to (Sharif et al., 2019), Some of the current computer-based methods are utilized in disease detection and diagnosis, but they still generate the wrong anticipation many times.
- Need to deploy for evaluating a more powerful CNN method. Also needs to utilize monitorisation to enhance the explanation of the outcomes (Yadav & Jadhav, 2019).
- The image size must also be optimized to maintain output accuracy and computing time, mostly on the real computer. (Pannu et al., 2020)
- There are several difficulties, therefore, to cope with the previous restriction. In this thesis, we will start exploring the following:
- We use various CNN architectures, such as recurrent neural networks and 3D CNNs, to accommodate injuries consecutively in adjacent sections.
- Also, it needs a way of detecting gastric physiology through narrow-band images.
- Use the Visualization report to enhance the overview of the findings.
- Automated annotation and segmentation techniques must be used to use a cross-modal image and collateral text characteristics obtained from qualified gastroenterologists. (Pannu et al., 2020)

2.5 Using deep convolutional neural network CNN in automatic reporting in the medical domain (GI)

Description of hospital records is essential in the healthcare industry; it provides a very accurate description and evaluation of the various diseases and helps people monitor their conditions. These findings are a useful source for datasets that can be used in study and resource management.

In (Hicks et al., 2018), The authors created a medical reporting framework that promotes accuracy and knowledge of internal processes. They demonstrated a method and explained how it could help generate and comprehend deep neural networks.

As in In (Hicks et al., 2018), (M. Riegler et al., 2018,) the essential elements of the medical colonoscopy report are as follows:

- Private details: age, sex, medical history
- Patient assessment
- Procedure: last colonoscopy, family history, the reason for the examination
- Technical summary: date, time, medication, dosage, duration, bowel preparation
- Conclusions: location, size, morphology, and method for the removal of abnormalities.
- Evaluation based on conclusions
- Follow-up plan: further test results, shifts to drugs, and recommendations

(Jing, Xie, & Xing, 2017) study the automatic generation of diagnostic imaging suggestions to help doctors produce more efficient and precise hospital records. They construct a multi-task learning framework to anticipate the search terms and tags to recognize abnormal regions and document these disorders. Finally, the researchers attempted to create long paragraph reports using the hierarchical LSTM model. They implemented their methodologies to the diagnostic datasets of the Chest X-ray collection (IU-X-ray). They proved the performance of their processes through means of quantitative and qualitative studies.

According to (Kisilev et al., 2015), a novel method for automatic breast radiology documentation has been presented. They construct descriptors and their interactions through structured learning. An SVM was used to coach a collection of images to describe the learned model parameters. When testing new pictures, the output toxicology report is produced in radiological lexicon classifications that assist radiologists in supporting the reasoning to diagnose the system, applying a computer-aided diagnostic (CAD) structure. They used their technique for breast imagery (sonography, modalities, and mammography). Items of 408 sonography images and 203 mammography pictures have been used; each image was matched with a verified diagnostic and radiological lexicon descriptor value system.

In (Zhang, Xie, Xing, McGough, & Yang, 2017), The works have presented MDNet modeling to learned images, generating a solid vision, processing symptom summary images, and providing visualization reports. In addition to the image model, MDNet implemented a language model; the operating system aimed to enhance multi-scale feature sets and improve its use efficiency. The classification algorithm intended to improve the feature representations by reading a report and describing the discriminatory image to map words from a sentence to an image pixel. They have established an optimization technique to develop the overall network. They had used a dataset of images and their radiology tests of bladder cancer pathology. They have used well-known IFAR-10 and IFAR-100 to verify their conceptual approach.

In (Yuan, Liao, Luo, & Luo, 2019), The study gives an encoder with several chest x-ray images to forms and levels 14 common radiographic findings while trying to take consists of additional images by applying cross-view consistency. They enhanced the decoder with explanatory semantics, reinforced the correctness of the deterministic clinical data, trained data to retrieve the most common medical concepts from x-ray images, and obtained a medical conceptual framework based on radiology reports. Then, for each decoding step, those medical concepts were consolidated into a word-level attention mechanism. The Chest X-Ray exploratory database, composed of 224,316 multi-view chest x-ray images of 65,240 healthcare professionals with 14 common radiographic findings, was implemented. IU-RR was used to evaluate the generation of radiology reports.

2.6 Summary

This chapter discussed the background of machine learning; supervised learning, unsupervised learning, active learning, and deep learning, focusing on the convolutional neural network. Then this chapter analyzed the limitations and the current challenges discussed the related works of using deep understanding and convolutional neural network in an automatic reporting system for the medical sector, especially for the GI tract. This chapter also talks about universal design and accessibility principles.

From previous research, the limitations of an automatic reporting system for medical documents, especially for the GI tract as follows:

1. The medical reports are typically long, containing multiple paragraphs.
2. Generating radiology reports need too much time and need extensive expertise.
3. It's difficult to understand complicated medical visual contents and linked them with accurate natural language descriptions.
4. The medical report comprises multiple heterogeneous types of information, results keywords (paragraphs and tags describe the findings).
5. It’s too difficult to identify and detect abnormal regions in medical images and express this in writing reports more challenging.
6. Building and testing automatic reporting on small-scale pathology image datasets is introduced.

In this thesis, we can work on constraints derived from past studies to as will be described in the following chapters:

1. Increase health data analysis and deliver more reliable data.
2. To pay special attention to complexities in the healthcare industry.
3. Implement an automatic reporting system using CNNs in diverse clinical dimensions, e.g., in the GI tract, to use vast databases of medical data reports.

CHAPTER 3

Requirements for An Automatic Medical Reporting System

3.1 Automatic Medical Reporting system

Medical reports are a data-driven method of evaluating the performance of particular processes or functions within a healthcare organization, intending to improve quality, minimizing mistakes, and optimizing healthcare metrics. It's possible to use data to turn indicators into meaningful intelligence, allowing you to spot flaws, recognize strengths, and forecast incidents before they happen, all by using interactive digital analytics. This perfect storm of visual data makes healthcare facilities safer, more efficient, and intelligent in the long run.

Medicare reporting, as a subset of business intelligence in the healthcare sector, gathers data from the five key industries:

- Claims and operating costs analysis
- Metrics related to pharmaceuticals and research and development
- Clinical information gleaned from ongoing patient care, electronic medical records, and clinical studies
- Statistics on patient actions that are important to know
- In an ever-changing world, predictive insights and data can be used to develop preventative measures or interventions to enhance treatment and prescribing processes.

When a person has an appointment with a health care provider or is transported to a hospital, a medical report is an informative report that contains the person's medical history and information. These reports are beneficial to all parties concerned, including health care providers, insurance firms, and even patients, for several reasons. In reality, high-quality medical reports are essential.

A medical report is also essential proof of care provided to patients and particular conditions that affect them. The words medical report, health record, and patient are used synonymously to express the same sense. The following items should be included in medical reports, but this is not an exhaustive list:

- Age, sex, and weight are examples of personal data.
- A brief history of medicine
- Diagnosis data
- Any prescriptions should be recorded.
- Physical and emotional assessment results
- Allergies have a long and illustrious history.
- Hospital admission history
- Treatment background and reaction to treatments
- Health images such as X-rays and lab testing results
- Statements that affirm your strengths and weaknesses

Doctors and nurses have dated and checked some of these documents. Doctors and nurses have dated and checked some of these documents. Vital signs such as heartbeat, breathing rate, temperature, and blood pressure are recorded in daily patient records. These are crucial to track and quantify since they can be used to determine physical functioning and represent essential factors.

Patient reporting allows healthcare professionals to gain insight into a patient's medical background to offer the best possible care. The medical report serves as the foundation for patient care preparation and coordination between patients, physicians, and other professionals involved in the medical situation. These healthcare providers are the ones who keep track of their patient's medical records. Clinicians, Doctors of Medicine, surgeons, dentists, nursing staff, psychologists, pharmacists, and many in the medical specialties are among those who work in this sector. Health reporting should ideally be done by a doctor or medical specialist who is well-versed in the medical illness or has treated the patient for an extended period.

Hard copy data from pharmacy and hematology testing, microbiology protocols, cytology sheets, and notepads from automated analyzers, X-ray images and records, ECG indications, and blood checking such as antibody details and patient blood group are added to the patient report by nursing care units. Once the data has been verified, they have instant access to it. As a result, each information includes the most recent test result operation. Per entry often consists of a sample date and time and the time and date the tests were received.

Admittance notes, on-service documents, treatment plans, postoperative documents, operative documents, treatment notes, delivery information, and discharge notes are all examples of medical reports. These are usually used in acute care and are combined to create a comprehensive record of the patient's medical history, known as a personal medical record.

Electronic reports have now replaced many paper records distributed to all practitioners involved with the patient's health care sooner. While enhanced portability is desirable, it does come with its own set of issues, such as the complexity of keeping medical records secure and the accessibility of patient reports.

3.2 The advantages and benefits of an automatic reporting system

- validating and substantiating patient state disability statements
- assisting in health-care compensation given a way to track down people who have received unique treatments
- identifying therapies that could have resulted in some side effects
- providing proof of injuries and care in the context of workers' insurance
- looking at disease patterns to see if there are any environmental or genetic factors
- keeping track of improvements in your clinical findings, such as tumor progression or edema
- One of the most critical aspects of healthcare reporting and analytics is the opportunity to use historical data to predict possibly deadly medical conditions in

patients before they happen. In the healthcare industry, this subset of predictive analytics is critical for improving care delivery quality and lowering mortality rates.

- Business intelligence and observations in healthcare documents can also assist in monitoring potentially large-scale infectious diseases by combining historical and current metrics with insights to ensure that proper course of action or prevention efforts is taken to monitor or contain the condition.
- Healthcare reports will help make healthcare activities even more cost-effective. Data analysts may assist medical institutions in streamlining their financial planning processes and making changes or substitutions by gathering and reviewing operating indicators associated with the organization's regular healthcare spending in all core aspects, from serving to patient medication.
- Hospital reports enable organizations to combine clinical, economic, and organizational data to assess the efficacy of their various systems, and the health of their clients, and the effectiveness of their healthcare programmers. By having access to these data, a healthcare organization may determine which procedures aren't producing the desired results and make meaningful improvements.
- Data from healthcare reports will help hospitals provide information to patients on an individual level. Depending on their medical records and vital signs, a healthcare facility may offer their patients personalized guidance on maintaining a healthier lifestyle. Not only is this invaluable in terms of delivering superior patient-appropriate care, but it's also a significant step in the right direction in terms of helping to reduce hospital readmission.
- It is possible to reliably measure the results, quality, and efficacy of healthcare workers at the place of supply using healthcare reports. It can also use a medical performance portal and data analytics to provide continuous feed with continuous performance reviews compared to health industry study indicators related to patient health and happiness.
- In addition to measuring professional success transparently, Medicare online data analysis may also assist in identifying any internal bottlenecks or shortfalls that are costing the organization money and obstructing patient care.

3.3 Gastrointestinal reporting system

Endoscopy reports have usually used free text or unstructured words, sometimes accompanied by photo-documentation, similar to medical records in common. On the other hand, free text words avoid substantive data extraction and thus serve as a quality control bottleneck in GI endoscopy. Simple extracted data from endoscopy documents of methodological data, clinical factors, and quality attributes indicators are needed to maintain GI endoscopy and ensure high-quality endoscopy service. It is possible to implement quality-assured and standardized endoscopy testing procedures in daily practice.

As mention in chapter 2, there are some drawbacks are using deep learning CNN in the medical field, particularly in the GI tract, as follow:

- Increase some consequences to identify some health diseases such as lesion shape, color, texture, size, and irregularity.
- According to (Sharif et al., 2019), Some of the current computer-based methods are utilized in disease detection and diagnosis, but they still generate the wrong anticipation many times.
- Need to deploy for evaluating a more powerful CNN method. Also needs to utilize monitorisation to enhance the explanation of the outcomes (Yadav & Jadhav, 2019).
- The image size must also be optimized to maintain output accuracy and computing time, mostly on the real computer(Pannu et al., 2020).
- There are several difficulties, therefore, to cope with the previous restriction. In this thesis, we will start exploring the following:
- We use various CNN architectures, such as recurrent neural networks and 3D CNNs, to accommodate injuries consecutively in adjacent sections.
- Also, it needs a way of detecting gastric physiology through narrow-band images.
- Use the Visualization report to enhance the overview of the findings.
- Automated annotation and segmentation techniques must be used to use a cross-modal image and collateral text characteristics obtained from qualified gastroenterologists(Pannu et al., 2020).

3.4 Requirements of gastrointestinal reporting system

As mention earlier in the methodology section in the requirements part, the followings are requirements and recommendations of the GI report system according to the European Society of Gastrointestinal Endoscopy (ESGE) (Bretthauer et al., 2016):

1. To allow for continuous data tracking, modern endoscopy reporting systems must be digital. At the endoscopist, electronic documentation and recording of text and photographs allow for constant quality control. Electronic reporting also allows for continuous transparency for all those participating in the clinical care of particular patients, such as interdisciplinary teams and auditing of symptoms and adverse reactions. It also allows you to compare digital images from different procedures. Finally, endoscopy reporting software allows for precise tracking and tracing methods, allowing for early identification of possible disinfection and reprocessing defects.
2. Endoscopy reporting systems should preferably be incorporated into different methods to enable data sharing between the endoscopy system and healthcare data systems inside the hospital and between linked hospitals. The primary focus is on endoscopy reporting quality without compromising the collection of organizational and patient data required to track the efficiency and quality of colonoscopy services.
3. Where appropriate, endoscopy reporting systems should use formal terminology (in accordance with validated, standardized terminology), and free text data entry should be avoided. At the conclusion of the endoscopy study, individualized clinical guidelines answering all applicable clinical concerns should be limited to free text. The programmers for endoscopy reporting systems must then generate understandable endoscopy reports for non-endoscopy readers. As a result, clinics must make time available to enable both the learning process of electronic reporting systems.
4. A primary threshold variable should be specified for each treatment included in endoscopy reporting systems, identifying variables as clinical, efficiency, or analysis. Endoscopy reporting systems must also be likely to facilitate the

variable extension required by local clients on a case-by-case scale, and over time, the endoscopy specialty evolves.

5. Endoscopy reporting systems should provide data such as histopathology outcomes, patient safety, client experience, and post-procedure adverse reactions. This can be accomplished by linking the endoscopy reporting system to other systems automatically
6. Endoscopy reporting systems must have the ability to extract data automatically so that predefined reports of clinical results, quality measures, and analysis data can be produced. Local teams should be able to create their personalized data performance reports, ideally.
7. Patient features such as patient gender and age, procedure description, planning, and previous surgery are significant when interpreting performance differences between endoscopists. As a result, endoscopy-reporting systems must offer various output thresholds based on patients and procedure characteristics.
8. Also, endoscopy reporting systems should be configured to allow for changes, especially the addition of new variables, without requiring major software redesigns.

3.5 Need of Universal Design for Report System (Website)

As we mentioned in chapter 2, we used a universal design approach for our report system(website) to achieve:

- Universal design provides access and creates solutions that serve the needs of all users, including those with and without limitations
- Individuals with hearing difficulties require captions to access information delivered through audio. When captions are supplied, English learners may well have a better understanding of the information.

- Persons utilizing screen readers and people with learning difficulties will benefit from a website that is structured with headings, well-organized material, and keyboard navigating. Everyone will find it much easier to understand and more visually appealing.

3.6 Summary

This chapter presented the specified properties and requirements for the automatic reporting system in the healthcare sector, including automated medical reporting system and its advantages and benefits, focusing on the GI reporting system, especially the requirements and recommendations of the GI reporting system. It's Important to follow the requirements of medical reports to improve the GI reporting system. And this shows that we can use CNN to improve the Reporting Generation System for GI Tract.

CHAPTER 4

Deep Convolutional Neural Network in Reporting Generation System for Gastrointestinal Tract

This chapter discusses the methodology of using CNN to enhance the automatic reporting and documenting system for the GI tract trained to the HyperKvasir dataset. This section will describe various used network architecture and other used datasets to introduce the proposed system viewing hyperparameters utilized in training. Then, the model performance metrics are examined.

4.1 Proposed overflow

The motivation of this research methodology is widening the deep learning techniques with the medical domain. It enhances based on the less standardization and quality between GI report generations using an automated system. A complete flow of the proposed approach has been described as follows, and it has three main functionalities:

- The system was designed to assist physicians in making rational choices on the diagnosis of disease identified during examinations, including the diagnosis of disease discovered in the GI tract throughout the colonoscopy.
- Our proposed method generates automatic reports on the automated analysis with images, videos and decreases the time spent on administrative duties following a process theory. This has been described in figure 4.1, where the doctor uses the system to comprehend the neural network analysis and to use this data to diagnose and produce the associated report.
- The current framework could be used by engineers and researchers inventing deep learning architectures such as CNN to obtain a greater understanding of the assessments and reactions of their designs, for example, by comprehending which areas of the image offend the algorithm and when additional pre-processing steps need to be taken.

The proposed work, beginning with an analytical outline of the numerous tools and techniques used to implement the proposed system. Here, most of the time, We would then present and argue about using these techniques, displaying how they impacted the implementation of the project work and on what basis they preferred. With a fundamental understanding of the opportunities behind the proposed work, we provided a systematic glance at each tool included in the system. Beginning with the neural network tool, we demonstrate how it can be used to develop a thorough understanding of CNN's decision-making method via an analysis of CNN's intermediate layers. And ultimately, we look for the report generation tool, which aims to help doctors compose colonoscopy reports through a web interface.

This technique is structured across the architecture of the client-server. The advantage of this configuration is to discharge the computation time of deep learning to a relatively powerful central server, which any user may use without worrying about hardware specifications. In addition, our customer is enacted as a web application that has introduced the benefit of being readily available from every platform that connects a modern web browser. In our case, the client and the server code are segregated into their various directories and workflows, making it very easy to build and run individually. As with one of the most modern software development projects, this methodology is not printed from scratch but also with the assistance of technology, libraries, and frameworks. To achieve a high level of quality and sustainability, we want to set down specific basic rules on which techniques will be used in the growth of figure 4.1. This diagram provides a comprehensive implementation of the study. Starting with the endoscopic procedure, the image information is obtained and sent to the central server. This server is obtained by a medical professional across a web browser where they can execute endoscopic statistical research. Depending on performance, the doctor can generate a web report using the web application and produce a final endoscopic report.

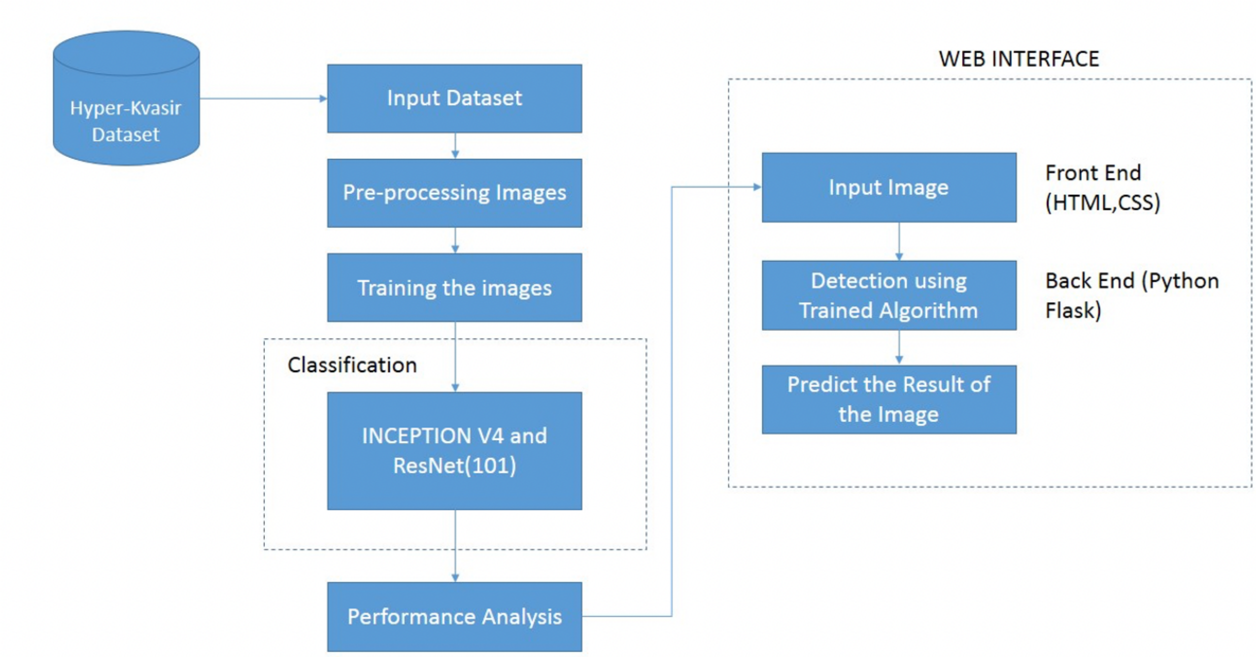


Figure 4. 1 Process Flow diagram

In this proposed system, three kinds of CNN architecture are used, such as VGG16 (Simonyan & Zisserman, 2014), DenseNet121, (K. He, Zhang, Ren, & Sun, 2016), inception (V3) (Szegedy, Vanhoucke, Ioffe, Shlens, & Wojna, 2016). This model has been trained using HyperKvasir datasets for computing the performance metrics with recent tools and techniques.

4.1.1 Data Collection

The dataset was collected from Simula Research Laboratory. Hyper-Kvasir dataset was used (Borgli et al., 2020). In this dataset, we only used labeled images for model training. The data is collected during real gastro- and colonoscopy examinations at a Hospital in Norway and partly marked by experienced GI endoscopists.

In total, the dataset contains 10,662 labeled images stored using JEPG format. The labeled images represent 23 different classes, and each image is found in their class folder. The number of images per class is not balanced, which is a general challenge in the medical field since some findings occur more often than others. So, we used only 15 classes with a minimum of 100 images in their folder for our analysis. HyperKvasir dataset is used for the evaluation of the proposed model. The network is pre-trained based on the ImageNet dataset.

4.1.1.1 ImageNet

It is a massive database of images constructed for image recognition applications. It mainly contains 14 million hands annotated photographs divided into more than 20,000 groups(Deng et al., 2009). A massive number of classes are involved ranges from a various number of images. The proposed system will train CNN on ideas from the ImageNet database used as the base weights during the transfer learning process.

4.1.1.2 HyperKvasir

These are the datasets offered by Simula that contain color images captured from the GI tract. The data were acquired utilizing endoscopic equipment from the Vestre Viken Health Trust (VV) in Norway that each carefully formatted through qualified endoscopists. Here, we have used labeled datasets that contain 10,622 captured images in a JPEG format. The classes that each image belongs to correlate to the file in which it is stored (e.g., the 'polyp' folder appears to contain all the polyp images. The 'Barretts' folder contains all the Barrett images' of the esophagus, etc.). Divides the datasets for training and testing procedures. There are 23 classes, and the levels reveal anatomical landmarks, endoscopic procedures, and pathological identifications in the GI tract. This is the primary dataset we used for training tried to understand the characteristics of various classes in this dataset (Borgli et al., 2020).

4.1.2 Data Preprocessing

Due to the limitation of image numbers, data augmentation was used to get a better result. ImagaDataGenerator is a library inside the TensorFlow library, which helps with data augmentation. The images were normalized by dividing with a maximum pixel 255 to get better convergence in weights parameters. The whole dataset is divided into two parts training and validation datasets. This dataset is split in the ratio of 0.8:0.2 to training and validation. These datasets will be used for the model training purpose.

4.1.3 Training

The pre-construct Keras model is used, which involves weights trained on the ImageNet dataset for training and validation. Then execute transfer learning through re-training the HyperKvasir dataset. Here, we have used the architecture of CNN to build the model. As for each model training process, we will use the pre-construct models with the pre-trained weights to conduct the initial testing. We use simple transfer learning to create a network with the same hyperparameters.

Transfer learning refers to the process of recycling and applying pre-trained weights to various domains. If the volume of data is a constraint, learning can help build the basic features' design. The essential transfer learning element is the re-tooling over the last block of the system, the block accountable for classification. TensorFlow pre-trained model is used with the ImageNet weight. The final step is to fine-tune the prototype. Regarding re-training the classification block, we reassign the different aspects of the program under fine-tuning.

4.1.4 Testing

We take a z-line video which we will get from the endoscopist for analysis for the testing purpose. There will be the patient’s information and a placeholder to upload video/image from the endoscopy in the front end. Then, this information is processed in the backend. After processing it, generate the report, which contains patient information and possible disease they may have.

4.1.5 Architectures

In 1998s, convolution-based neural networks are introduced. Since then, several CNN architectures have been developed, containing input, output, and multiple hidden layers.

The hidden layers comprise convolution layers that concatenate with multiplication to other dot products.

We will use the three; Architecture of VGG16, Architecture of inception (V4), Dense Neural Network (DenseNet121), to set up a system for anomaly detection in GI images and videos. Then compare the results for all the three Architectures using different evaluation metrics.

4.1.5.1 Architecture of VGG16

VGG16 is a CNN architecture utilized for succussing the ILSVR (ImageNet) competition in 2014(Nash, Drummond, & Birbilis, 2018). It is one of the best architectural vision models to date. More than 14 million images labeled for 1000 groups are in the VGG16 dataset. The most notable feature of this architecture is that, rather than using many hyperparameters, they focus on using 3x3 convolution layers with such a stride of one and still use the same padding and max pool layer of a 2x2 stride two filters.

As seen in figure 4.2, the convolution and max pool layers are arranged uniformly in the design. Ultimately, it has 2 FC (top layers connected) followed through a SoftMax for performance. The 16 in VGG16 reflects the fact that it contains 16 layers of different weights. This system is extensive, and it has specifications of about 138 million approximately.

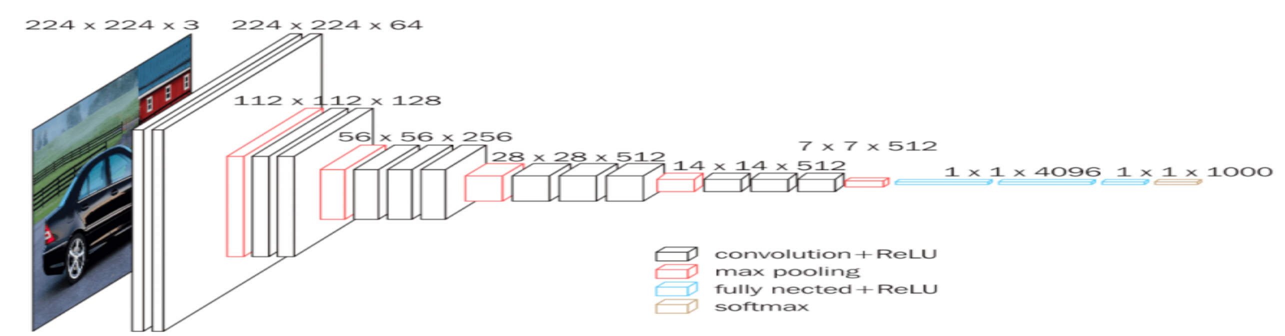


Figure 4. 2 VGG Architecture (Nash, Drummond, & Birbilis, 2018)

4.1.5.2 Architecture of inception (V4)

The Inception architecture is to mitigate the expenses of training massive neural networks by minimizing the parameters (Demir, Yilmaz, & Kose, 2019). This model contains a simple unit called "Inception cell," which performs a series of convolutions on different scales and following aggregates the results shown in figure 4.3. 1x1 convolutions have been used to reduce the depth of the input layer to store data. Here, it acquires a sequence of 1x1, 3x3, and 5x5 filters through each cell, using the input to gain attributes at various levels. But with "same" padding, max pooling is used to maintain the aspects to convolve the output correctly.

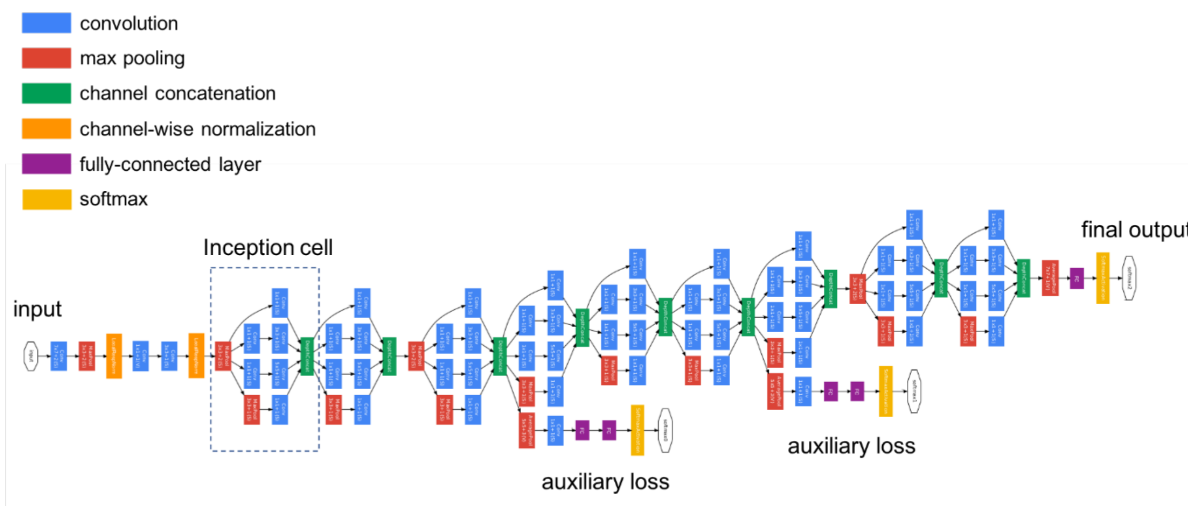


Figure 4. 3 Inception Architecture (Demir, Yilmaz, & Kose, 2019)

4.1.5.3 Residual Neural Network architecture (ResNet101)

The objective of ResNet101 architecture, as shown in figure 4.4, is to coach deep neural networks (Gao & Bie, 2018). This network is thought to understand more complex features and input representations that can improve results, but attaching more layers leads to a negative impact on the results. To address this issue, residual blocks are used as an intermediate block layer to perform a residual function related to the input block. The refining phase is named to adjust the feature map besides better quality characteristics that require different and new feature maps to be taught by each layer. The middle layers will know to progressively modify their weights to zero that no refinement has been required so that the residual block reveals an identity function.

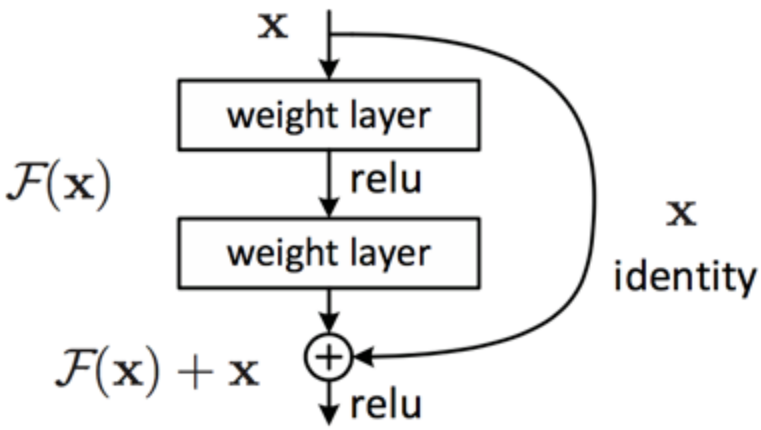


Figure 4. 4 A Residual Unit (Gao & Bie, 2018)

4.1.5.4 Dense network architecture

In this dense network architecture, as shown in figure 4.5, the feature map of each layer within a block to the input of each successive layer. This allows the features of previous layers to be explicitly used by layers inside the network, facilitating the reuse feature. And also enables a user to make tiny quantities of output and reduces the total number of required variables or parameters. The number of filters used as a growth rate through each convolution layer is known as kk because each successive layer would have more filters than the final.

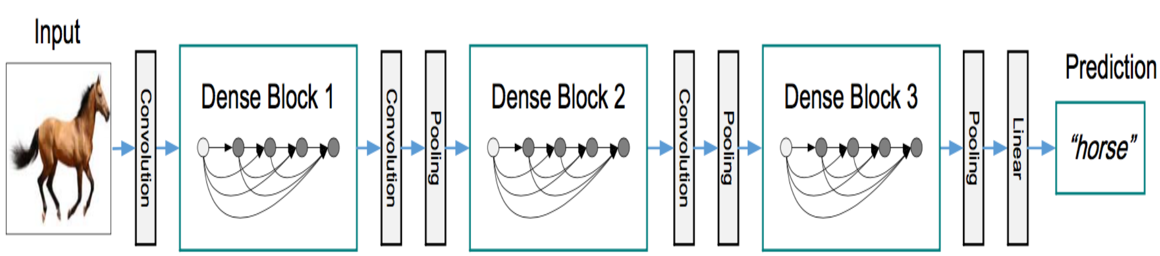


Figure 4. 5 DenseNet Architecture (X. He, Wu, Song, Jiang, & Zheng, 2020)

4.1.6 Hyperparameter selection

The selection of hyperparameters in deep learning did not follow a rigid scientific process. Even though we employ pre-built architectures, we must still choose optimization factors such as learning rate, epochs, batch size, and so on. We tried, and continue to try, to create network processes concerning machine specs like RAM, and to tailor these parameters to specific datasets. Each framework's selection of hyperparameters is shown in the table. We agreed not to spend too much time optimizing the hyperparameters for optimal accuracy because the performance of our models is not the goal of this thesis.

Model	Epochs	Batch Size	Optimizer	Learning Rate
DenseNet121	10	11	SGD	0.001
Inception(V3)	10	11	SGD	0.001
VGG16	10	11	SGD	0.001

Table 4. 1 Hyperparameter selection for each model

4.1.7 Web Interface

4.1.7.1 Front end tools

The front-end tools used in this methodology are HTML5, JavaScript, and (cascading style sheet) CSS. HTML5 is a markup language used for structuring and presenting to create web pages and web applications. Cascading Style Sheets is a style sheet language used for describing the presentation of a document written in a markup language such as HTML, which helps to make the layout of the website.

4.1.7.2 Backend tools

The backend tool for this methodology is python. API is implemented with the help of a micro framework flask based on the Werkzeug toolkit and jinja2. The inserted images/videos are saved in the SQLite database. The image uploaded to the research methods is processed in the SQLite database along with the identifier and the basic meta-information. Identifiers are subsequently used only for retrieval when an assessment is requested, which comes in the form of either categorization or generation of different visualizations. Classification and visualizations are stored together with the respective image as required. CNN models used for assessment show a trend equivalent to those used for multimedia endpoints. You can post, modify, delete, and select designs for use with the API. The endpoints of analysis enable the classification and visualization of the different layers of the chosen neural network. It also sets out the endpoint for the complete study, which combines all aspects of classification and visualization into a single query.

4.1.7.3 Deep learning tools

It is a cost-effective, robust graphics processing unit (- GPUs) that is tailored to deep learning. It uses a python-based host language and necessitates a simulation model so that it is used as an elaborate computational graph in various databases. Most of the deep learning library services come with pre-built and pre-trained models. Keras is the deep learning library used in this methodology. It supports multiple pre-trained CNN models, trying to make exploring with distinct networks. It always offers additional data preparation and generation methods, making it easy to create a neural network. The Keras API provides access to its entire backend. Different monitorization has been formed using the Keras Back-end API (Gulli & Pal, 2017).

4.1.7.4 Report Generation tool

This interface tool, where doctors can configure the recommended report. The report generation application offers basic features like change and image/video addition. The methodology requires a fundamental tool for generating endoscopy reports to promote the supporting documents phase of the completed endoscopy. A screenshot of the tool can be seen in the figure below, which shows the description of the sample endoscopy report generated by the system. At its present state, the report generation application offers basic features, including altering text and adding images to the selected image to the right of the report. We kept the font size and font color suitable according to the design principle.

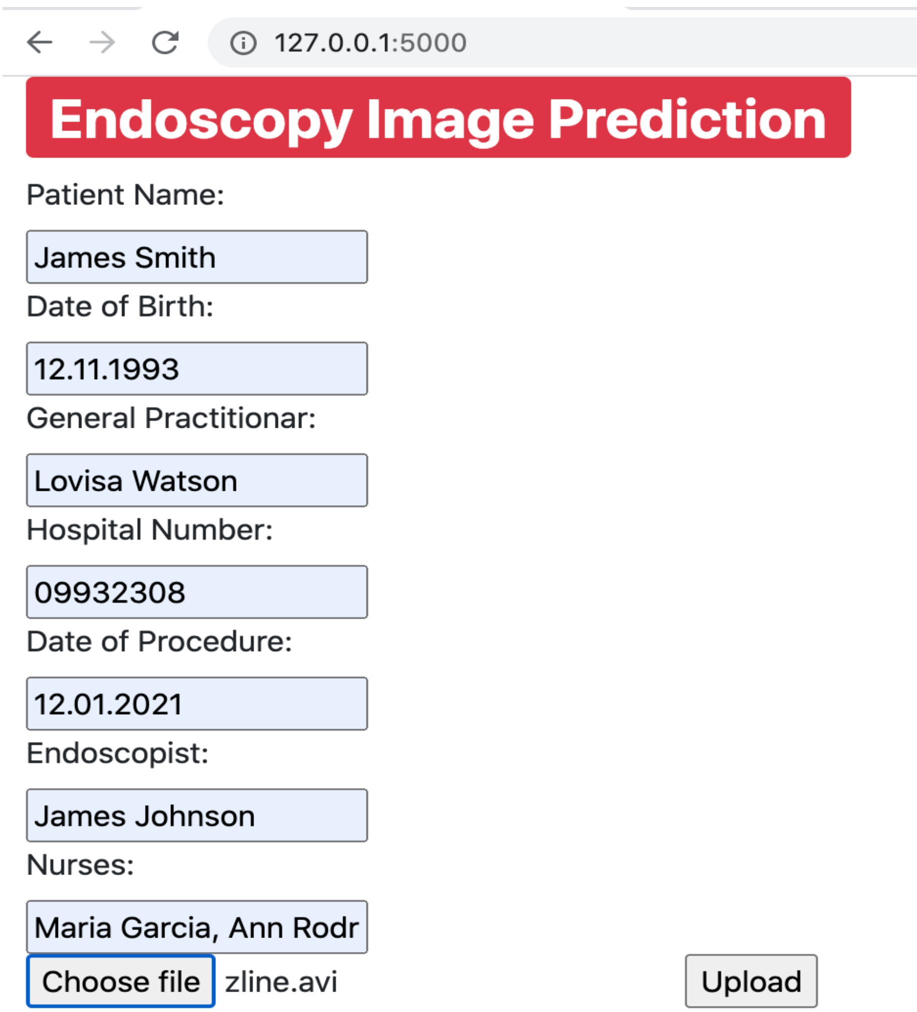


Figure 4. 6 Sample report generators user interface

However, in the case of a detailed report, where several findings have been found that would have to be cataloged, it could take several minutes or more to generate those findings. There is a need for automated multimedia reporting tools to enhance the documentation accuracy, standardize the monitoring by suggesting terminology endorsed by the MST and WEO (World Endoscopy Organization), and reduce the time required to generate a detailed report. As the first iterative process proposed, it concentrated on interpreting the deep analysis of CNN depending on the total assessment of the medical examination, assistance for multiple templates, and a wide range of quality of life (QOL) enhancements, such as assistance for drop-down dialog boxes and drag & drop devices.

4.1.8 Evaluation

4.1.8.1 Confusion matrix

A matrix of confusion is a method for summing up the performance of the classification algorithm. Classification accuracy alone would be misleading if we have a varying sample size for each class or whether we have more than two levels in our dataset. Estimating a confusion matrix can better explain how our classification method will work and what kinds of mistakes it is creating *(Townsend, 1971).*

The confusion matrix below demonstrates the predicted total values and provides identities to the classification pairs: true positive, true negative, false negative, and false positive.

Predicted

Actual

	Positive	Negative
Positive	TP	FN
Negative	FP	TN

Table 4. 2 Confusion Matrix

Four Outcomes of Binary Classification

• True Positives (TP): data points classified as positive that are strictly positive

• False positives (FP): data points labeled as positive that are negative

• True negatives (TN): data points marked as negative that are negative.

• False negatives (FN): data points labeled as negative that are truly positive.

4.1.8.2 Metrics

We analyzed TP, FP, TN, and FN after constructing the confusion matrix. The accuracy, precision, recall, specificity, F1-score, and Matthew correlation coefficient are all performance measures for this method (Hossin & Sulaiman, 2015).

The accuracy level of the methodology's effectiveness is measured by accuracy. It is the most frequent process of measuring the model performance, but this does not provide a complete picture of how well the models are performing. As a result, we frequently employ extra performance indicators to provide a more realistic picture of the model’s ability.

$$\text{Accuracy} = \frac{\mathrm{TP+TN}}{\mathrm{TP+TN+FP+FN}}$$

“Precision signifies how many numbers of predicted positives are true positive out of all predicted positive value.”. This metric is significantly associated with the REC metric, as the two are used combined to indicate classification performance when class results are completely biased.

$$\text{Precision} = \frac{TP}{TP+FP}$$

“Recall evaluates how many numbers of positives predicted precisely out of all actual positive values.” PREC and REC are typically used together, as we noted in the PREC explanation, and provide a more comprehensive view of performance than looking at each indicator separately.

$$\text{Recall} = \frac{\mathrm{TP}}{\mathrm{TP+FN}}$$

"F1 is the average of precision and recall. PREC and REC both are frequently weighted differently".

$$\text{F1-score} = \frac{2*Precision*Recall}{Precision+Recall}$$

Matthews's correlation coefficient, which is a metric that considers all TP, TP, TP, and TP, with a value ranging from 1 to 1 based on data distributions. The MCC, like the F1, offers us a better idea of a model's overall performance than using the previous metrics alone. (Hicks et al., 2018).

$$\text{MCC} = \frac{TP*TN-FP*FN}{sqrt((TP+FN)(TP+FP)(TN+FP)(TN+FN)}$$

4.2 Summary

This chapter discusses the methodology steps of using CNN to enhance the automatic reporting and documenting system for the GI tract trained to the HyperKvasir dataset. This chapter described how to collect the data and different used datasets, preprocessing steps, training and testing the model, various used network architecture, and introducing the proposed system viewing hyperparameters utilized in training. Then, Web Interface was discussed, representing the backend and frontend tools and the automatic generation tool of GI tract reports, and finally, the model performance metrics are examined.

CHAPTER 5

Experiment and Results

This chapter will describe the experiment set to implement CNN classification of GI images and the results of classification and evaluate these results.

5.1 Experiment Configuration

Table 5.1 describes the whole configuration of the experiment, including the used hardware and software. The Google Collaboratory is used to train the model. GPU provided by google is used to train the model. A local mac book with a CPU is used to test the model. From the Hyperkvasir dataset, only 15 classes are used, containing more than 100 images in each class. They are barretts, bbps-0-1, bbps-2-3, cecum, dyed-lifted-polyps, dyed-resection-margins, esophagitis-a, esophagitis-b-d, impacted-stool, polyps, pylorus, retroflex-rectum, retroflex-stomach, ulcerative-colitis-grade-0-1 and z-line. The images were normalized by dividing with a maximum pixel of 255 to get better convergence in weights parameters. The whole dataset is divided into two parts training and validation datasets. This dataset is split in the ratio of 0.8:0.2 to training and validation. These datasets will be used for the model training purpose. After training the model, a sample Z-line video is used for testing purposes.

Level	Category	Name	Version
Hardware	GPU	Tesla K80	
	CPU	2,3 GHz Dual-Core Intel Core i5	
	Memory	8 GB 2133 MHz LPDDR3	
Software	Operating System	Linux/Mac	10.15.7
	Library	Python	3.7.10
		TensorFlow	2.4.1
		Keras	2.4.3
		CUDA	11.2
		Flask	1.1.2

Table 5. 1 System specifications for the machine used for all training and evaluation sessions.

5.2 Evaluation of Classification Results

The goal of this experiment is to implement the CNN classification of GI images. It's essential to evaluate the classification to show how the performance of the classification models suggested on the framework correlating to the actual implementation results. We are using different evaluation metrics for evaluating the effectiveness of the classification model based on the samples of observations as described early and as shown in the table below.

Model	PREC	REC	ACC	MCC	F1
DenseNet121	0.887	0.861	0.916	0.909	0.873
Inception(V3)	0.856	0.818	0.904	0.896	0.836
VGG16	0.819	0.822	0.898	0.890	0.820

Table 5. 2 The evaluation results of all models trained on the Hyperkvasir dataset

Table 5.2 shows the evaluation result of the initial training and evaluation of the proposed architecture trained on the HyperKvasir dataset using hyperparameters. The DenseNet121, Inception(V3), and VGG16 architectures are used for this model evaluation. Each has the performance measures of accuracy, precision, recall, and f1, which calculates the research methodology's efficiency. Among these architectures, DenseNet121 has a higher prediction in all the performance measures. It shows the precision of 0.887, recall of 0.861, the accuracy of 0.916, and F1 of 0.873. Also, the other two architectures, DenseNet, achieved higher accuracy.

5.3 Results Analysis with Confusion matrix

Figures 5.1, 5.2, and 5.3 display the calculated Confusion matrices with each model combined across validation split. Confusion matrices show that some groups are often misclassified with others. It's essential to analyze which classes get distracted from discovering future training changes. It is significant to point out that certain classed were inside the HyperKvasir dataset, which is inherently equivalent to each other. Still, there are also non-natural similarities, such as an imbalance of various art craft types, which may affect the confusion. The confusion metrics are viewed to see what classes need further analysis.

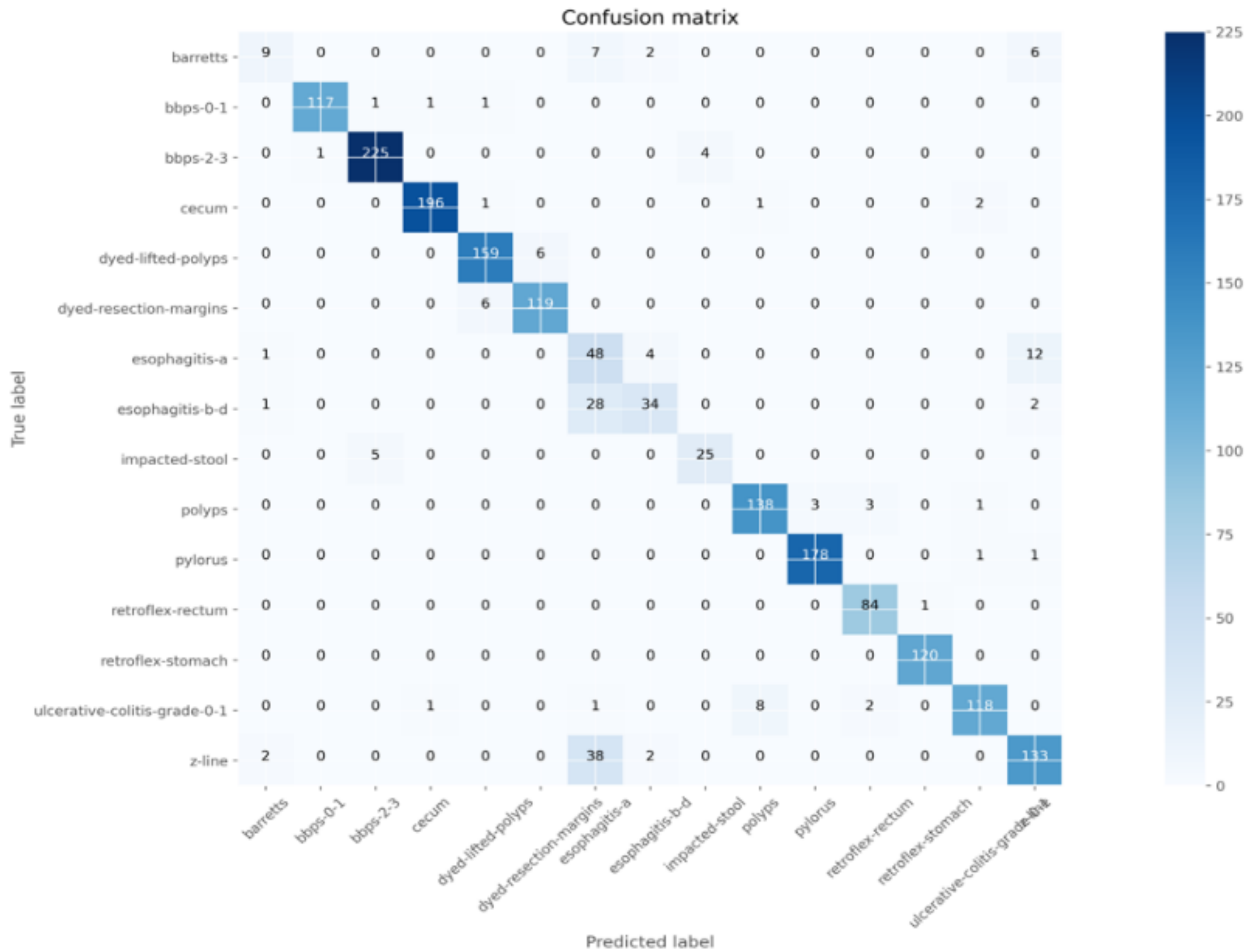


Figure 5. 1 Confusion matrix of DenseNet121

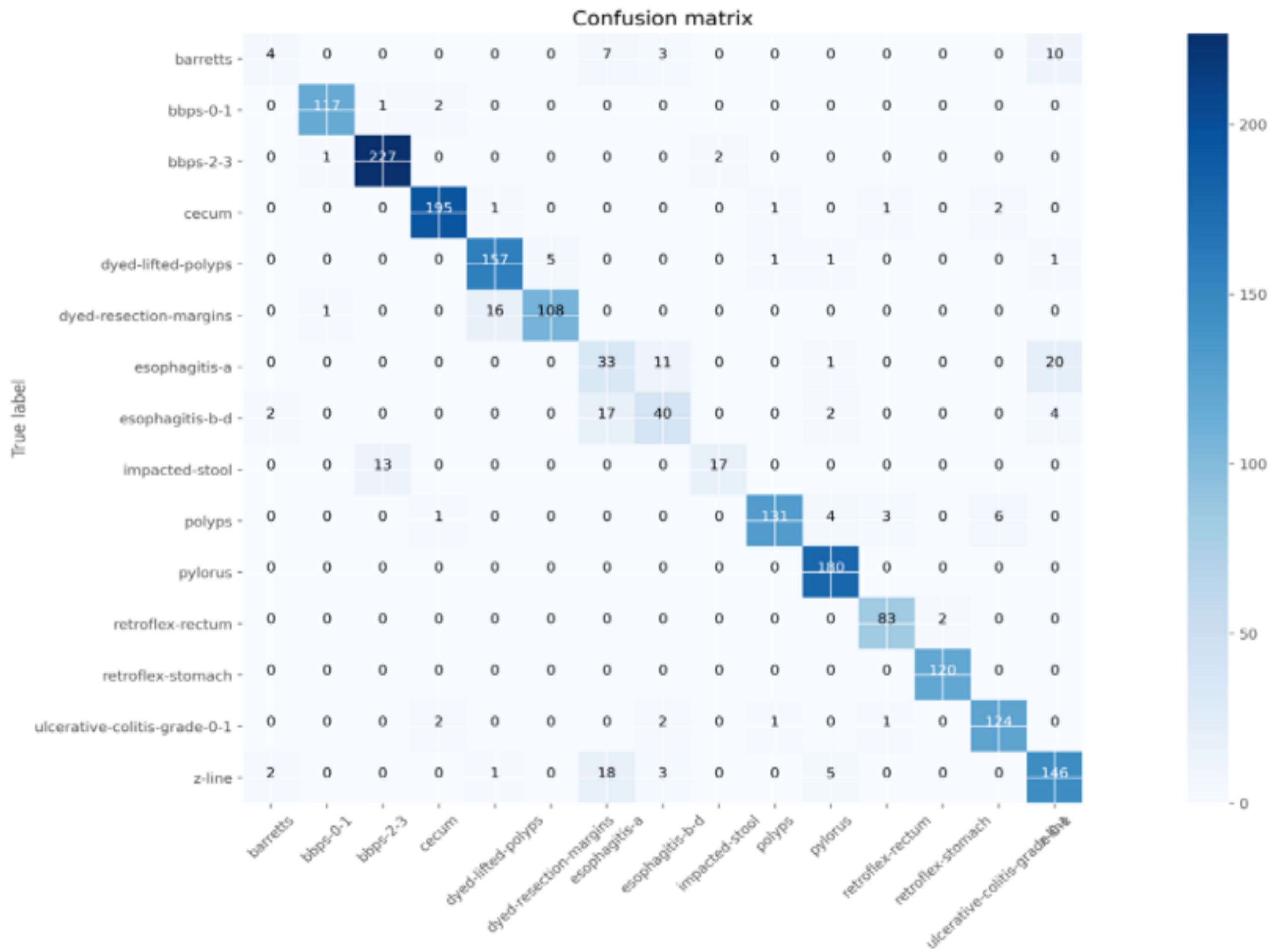


Figure 5. 2 Confusion Matrix of InceptionV3

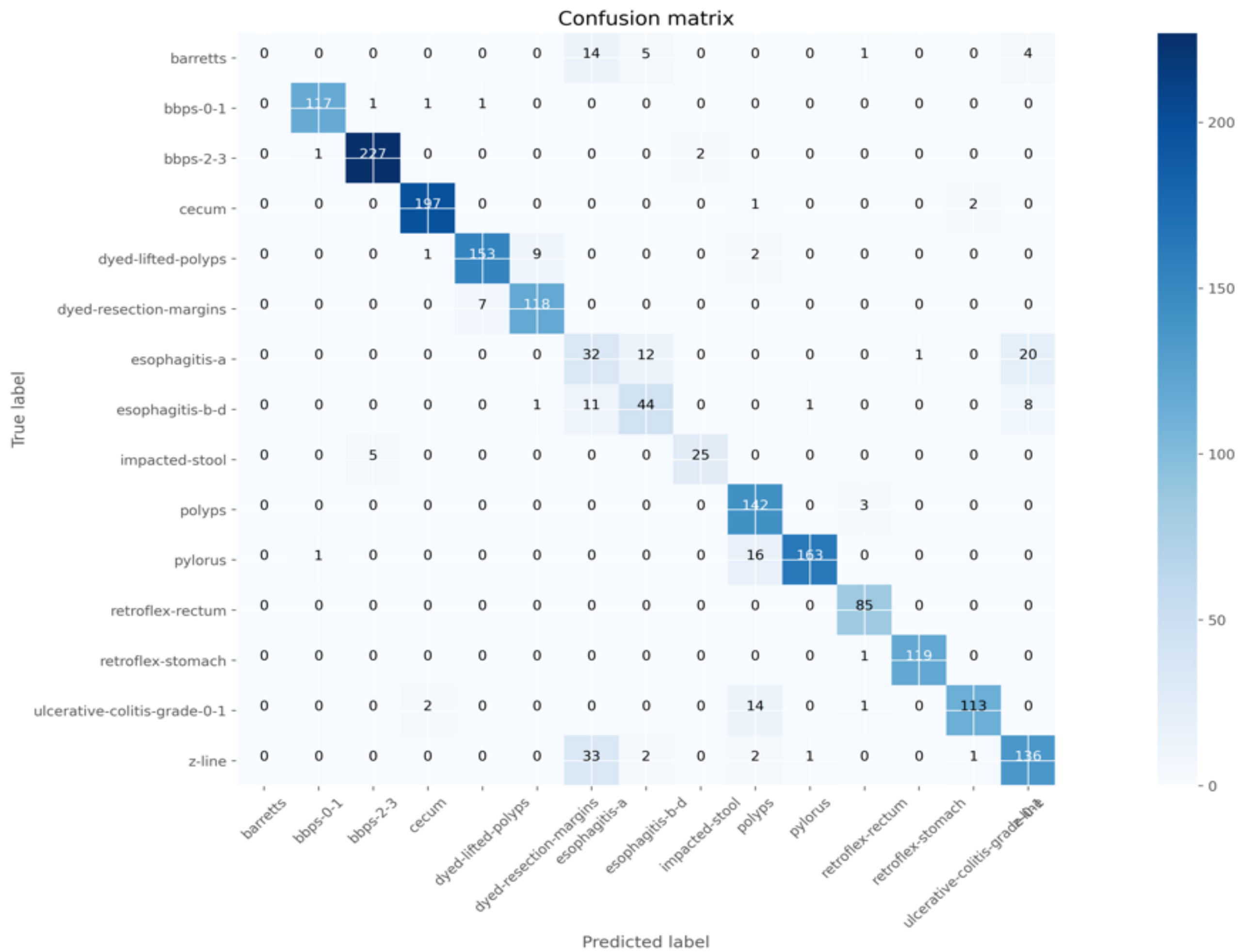


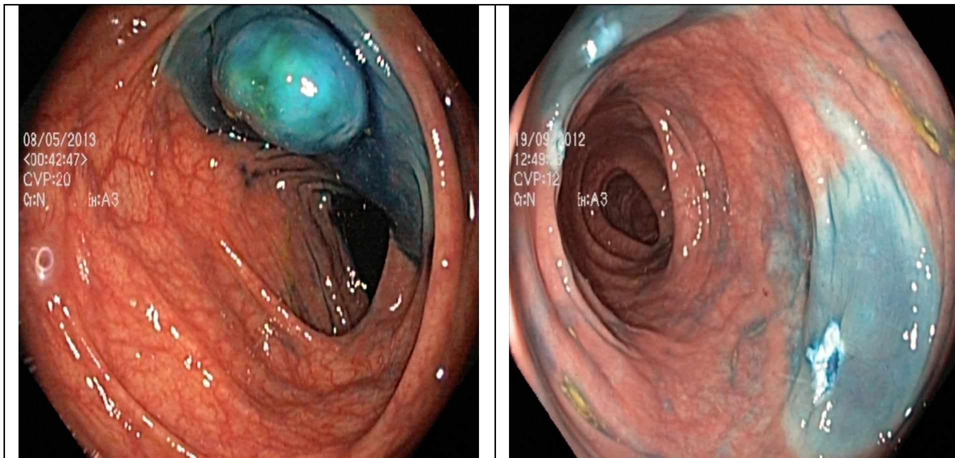
Figure 5. 3 Confusion Matrix of VGG16

This evaluation was primarily carried out using the CNN dissection tools, and we examined which images were misclassified for the target class.

5.4 Misclassification Results

5.4.1 Dyed-lifted polyps to Dyed-resection-margins

In the confusion matrix, we saw that dyed-lifted polyps are misclassified as dyed-resection-margins and vice versa. There are 165 images of dyed-lifted polyps and 125 images of Dyed-resection-margins. The given sample images show a little bit similar pattern that why we can say that module got confused to distinguish between them.



Dyed-lifted polyps	Dyed-resection margins

Figure 5. 4 *Sample image* of dyed-lifted polyps vs. dyed-resection-margins

VGG16: Here, we see very few misclassifications; only 5% of dyed-lifted-polyps are misclassified as dyed-resection margins. In comparison, 6% of dyed-resection margins are misclassified as dyed-lifted polyps.

DenseNet121: Similar to VGG16, DenseNet121 show few misclassifications; only 3 percent of dyed-lifted polyps are misclassified as dyed-resection margins. 5% out of 125 images of dyed-resection -margins are classified as dyed-lifted-polyps.

InceptionV3: The inception model also shows only 5% misclassification of dyed-lifted polyps with respect to Dyed-resection margins. 12% of Dyed-resection-margins is misclassified as dyed-resection-margins.

5.4.2 Esophagitis-a to Esophagitis-b-d

In the confusion matrix, we saw that Esophagitis-a is misclassified as Esophagitis-b-d and vice-versa. There are 65 images of Esophagitis-a and 65 images of Esophagitis-b-d.

VGG16: Here, we see few misclassifications; only 18% of Esophagitis-a is misclassified as Esophagitis-b-d. At the same time, 17% of Esophagitis-b-d is misclassified as Esophagitis-a. In this also the figure are pretty comparable.

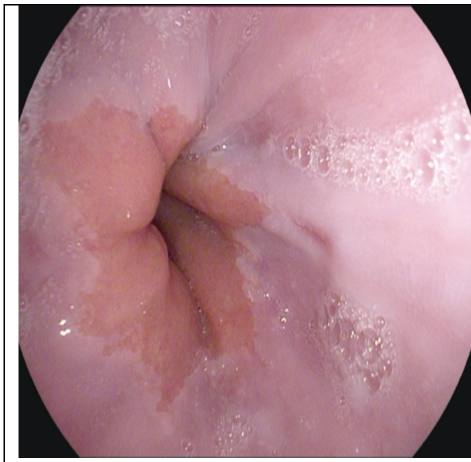
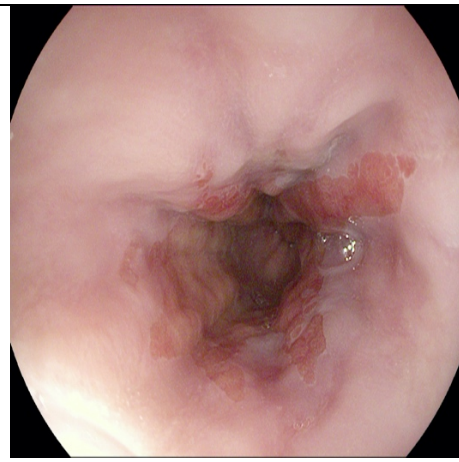
	
Esophagitis-a	Esophagitis-b-d

Figure 5. 5 Sample image of *esophagitis-a vs. esophagitis-b-d*

DenseNet121: Similar to VGG16, DenseNet121 shows few misclassifications; only 6 percent Esophagitis-a is misclassified as Esophagitis-b-d. But here, the percentage of misclassification has increased to 43% of Esophagitis-b-d is misclassified as Esophagitis-a.

InceptionV3: Inception model shows 16% misclassification with respect to Esophagitis-b-d. And 26% of Esophagitis-b-d is misclassified as Esophagitis-a.

5.4.3 Polyps to Pylorus

In the confusion matrix, we saw that polyps are misclassified as pylorus and vice-versa. There are 145 images of polyps and 180 images of the pylorus.

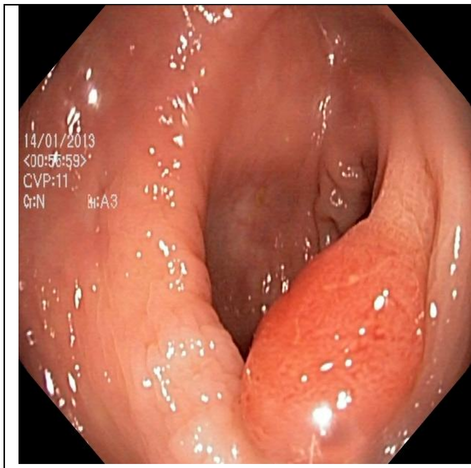
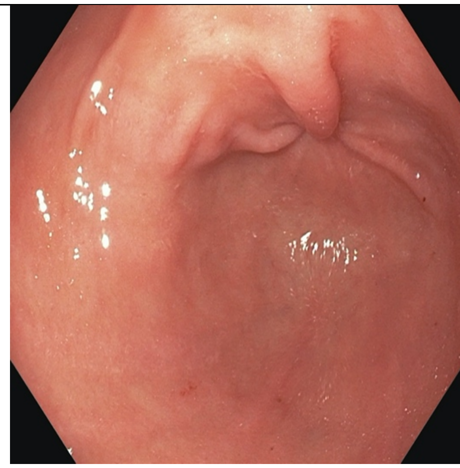
	
Polyps	Pylorus

Figure 5. 6 Sample image of Polyps vs. pylorus

VGG16: Here, we see very few misclassifications; none of the polyp’s image is misclassified as pylorus. And 8% of pylorus images are misclassified as polyps.

DenseNet121: Like VGG16, DenseNet121 also shows few misclassifications; only 2 percent of polyps is classified as pylorus. None of the pylorus is misclassified as polyps.

InceptionV3: The inception model also shows only 2% misclassification with respect to the pylorus. None of the pylorus is misclassified as polyps.

5.4.4 Z-line to Esophagitis-a

In the confusion matrix, we saw that Z-line is misclassified as Esophagitis-a and vice-versa. There are 175 images of z-line and 65 images of Esophagitis-a.

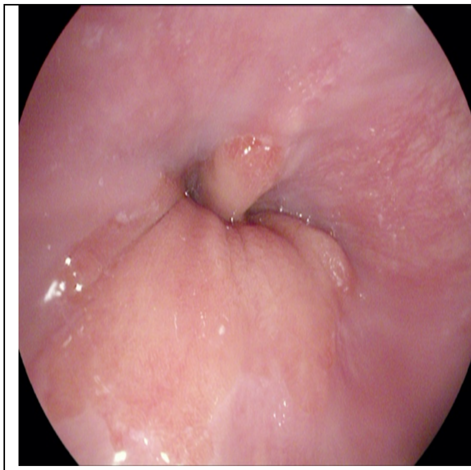
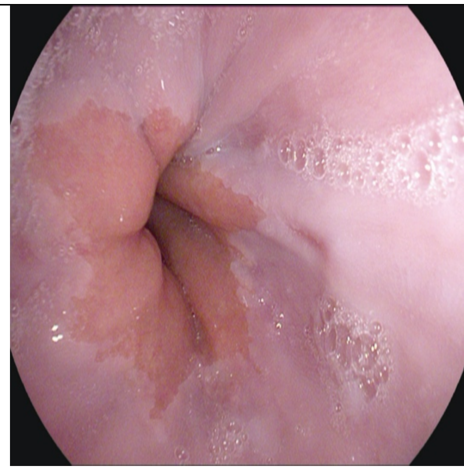
	
Z-line	Esophagitis-a

Figure 5. 7 Sample image of z-line vs. esophagitis-a

VGG16: Here, we see many misclassifications; 18% of zline is misclassified as Esophagitis-a. in contrast, 30% of Esophagitis-a is misclassified as Z-line.

DenseNet121: Similar to VGG16, DenseNet121 show even more misclassifications; 21 percent Z-line is misclassified Esophagitis-a. while 18% of Esophagitis is misclassified as Z-line.

InceptionV3: The inception model also shows only 10% of zline misclassifications to Esophagitis-a. In comparison, 30% of Esophagitis-a is misclassified as zline.

5.4.5 Barretts to Z-line

In the confusion matrix, we saw that Barretts is misclassified as Z-line and vice-versa. There are 24 images of Barretts and 175 images of z-line.

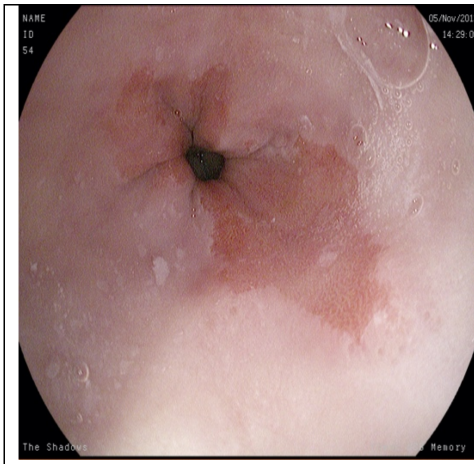
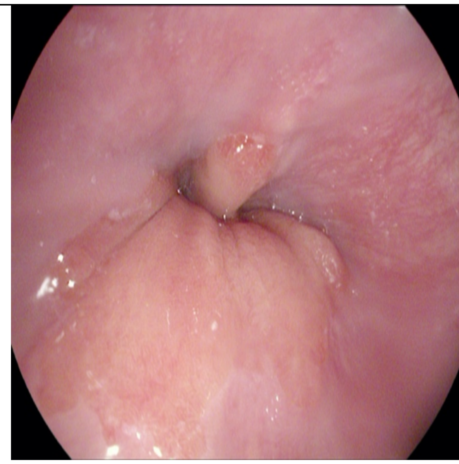
	
Barretts	Z-line

Figure 5. 8 Sample image of barretts vs. zline

VGG16: Here, we see few misclassifications; 16% of Barretts is misclassified as Z-line. None of the Z-line is misclassified as Barretts.

DenseNet121: Like VGG16, DenseNet121 shows few misclassifications; 25 percent of Barretts is misclassified as Z-line. Only 1% of z-line is misclassified as Barretts.

InceptionV3: The inception model also shows only 40% of Barrett’s misclassification with respect to Zline. 1% of Barretts is misclassified as Z-line.

5.4.6 Impacted-stool to bbps-2-3

In the confusion matrix, we saw that Impacted-stool is misclassified as bbps-2-3 and vice-versa. There are 30 images of Impacted-stool and 230 images of bbps-2-3.

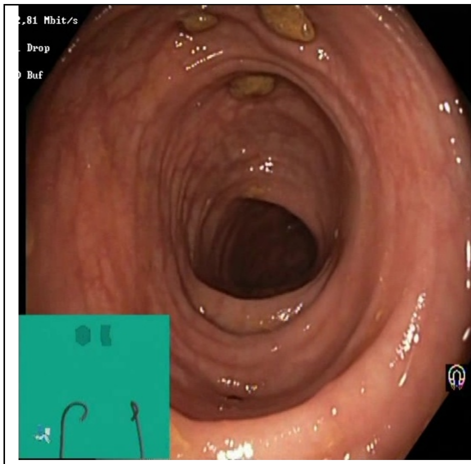
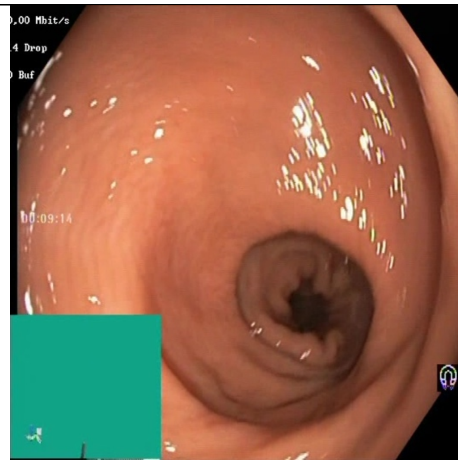
	
Impacted stool	Bbps-2-3

Figure 5. 9 Sample image of impacted-stool vs. bbps-2-3

VGG16: Here, we see many misclassifications; 16% of impacted-stool is misclassified as bbps-2-3. 1% of impacted-stool is misclassified as bbps-2-3.

DenseNet121: Similar to VGG16, DenseNet121 show few misclassifications; 16 percent of impacted-stool is misclassified as bbps-2-3. Only 1 percent of bbps-2-3 images is misclassified as impacted-stool.

InceptionV3: The inception model also shows many misclassifications; 43% impacted-stool is misclassified with respect to bbps-2-3. 1% of bbps-2-3 is misclassified as impacted-stool.

5.5 Accuracy Curves

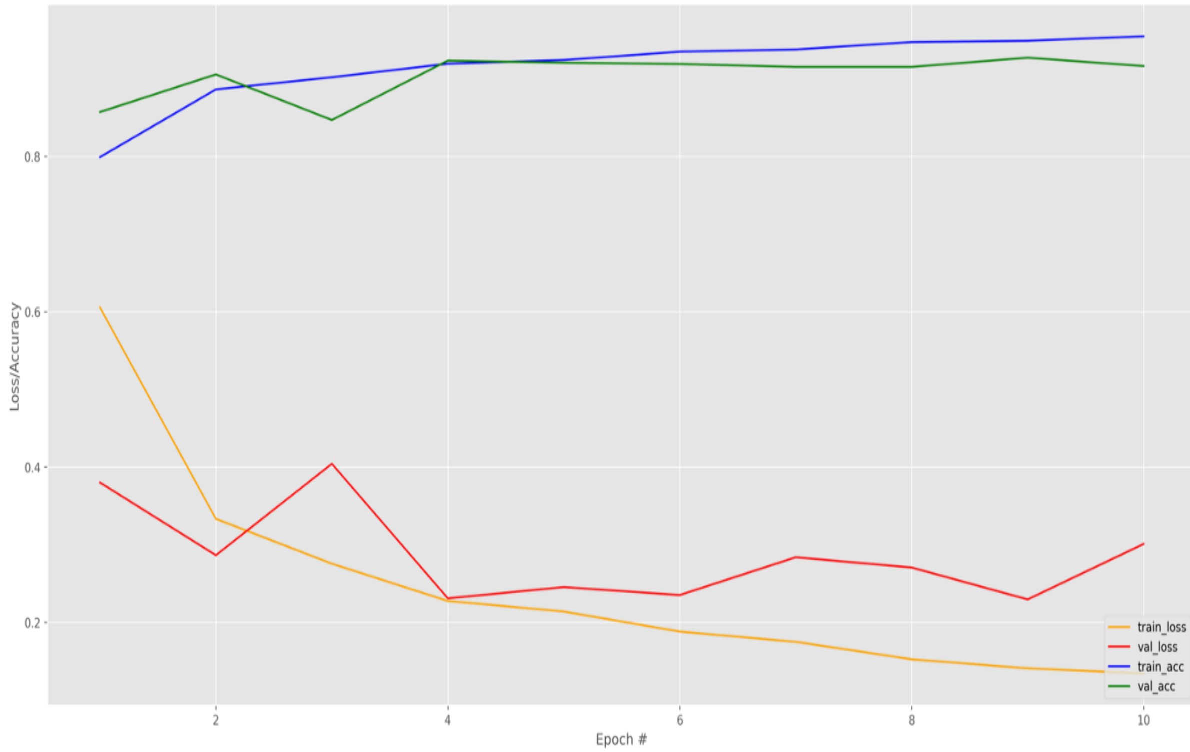


Figure 5. 10 Performance versus epoch on train and validation datasets in Densenet121

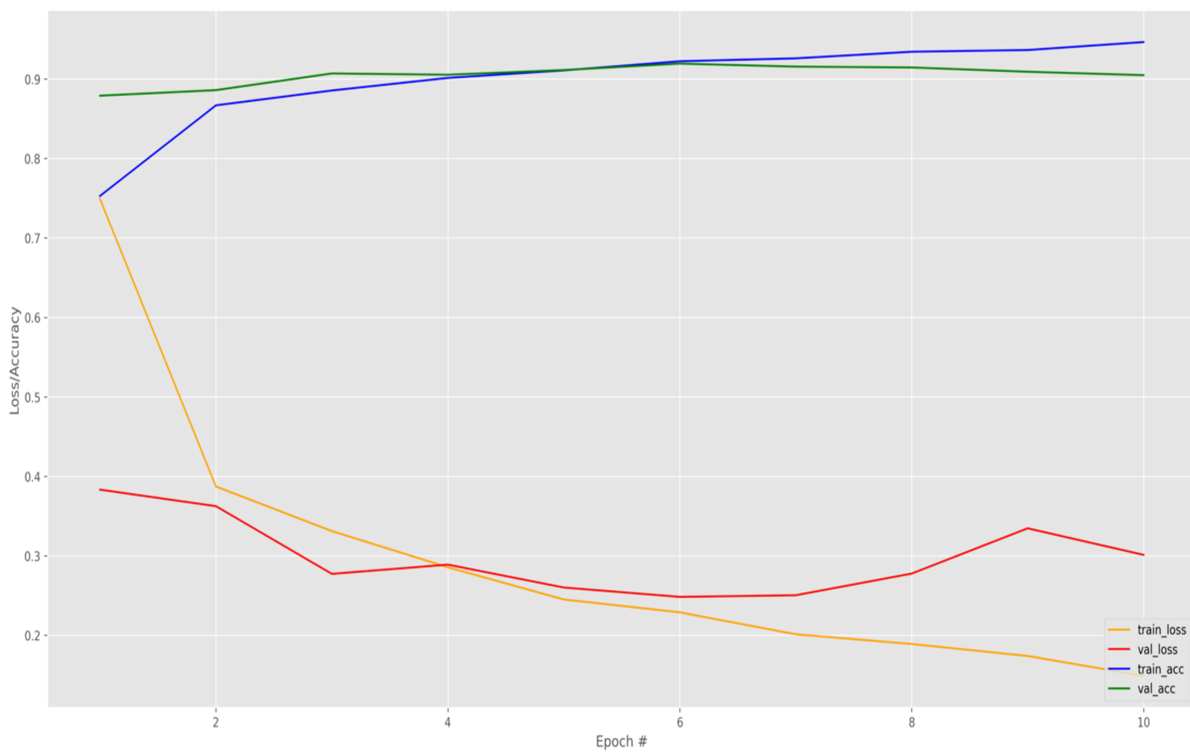


Figure 5. 11 Performance versus epoch on train and validation datasets in inceptionv3

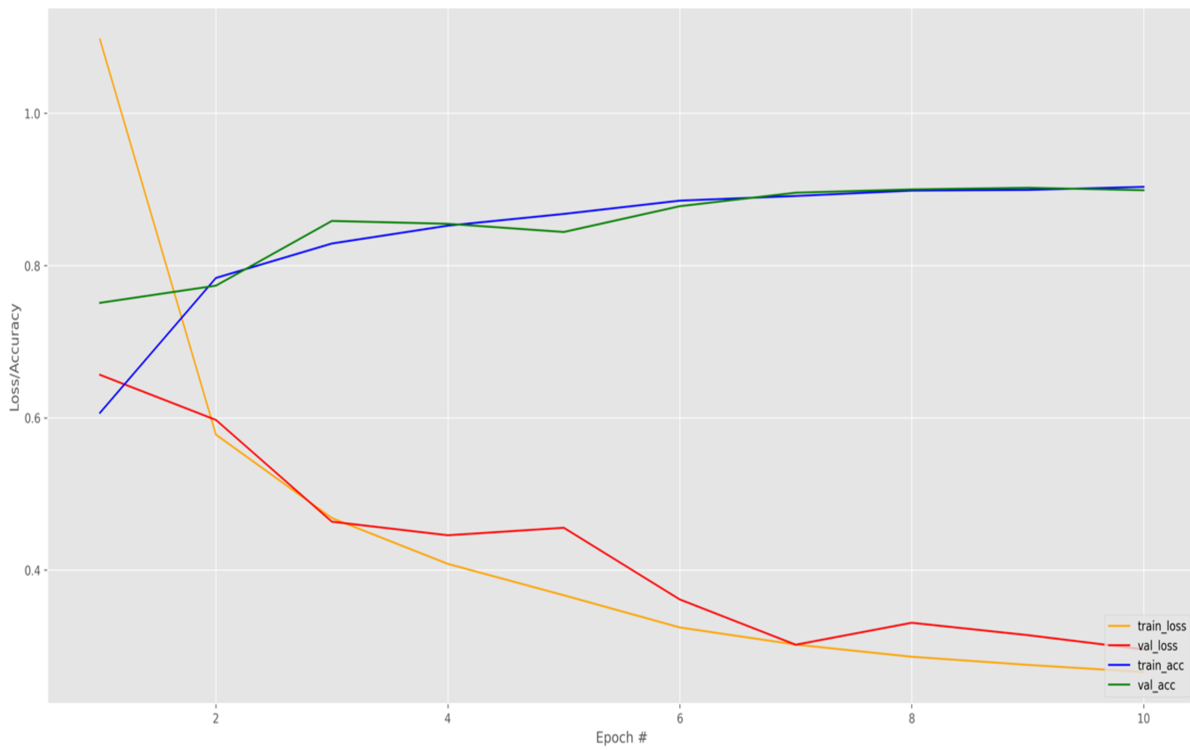


Figure 5. 12 Performance versus epoch on train and validation datasets in vgg16

5.4, 5.5, and 5.6 figures show the performance vs epochs in a different module. Generally, all modules offer increased accuracy when the epochs grow, but loss decreases. In each figure, lines show training loss, validation loss, training accuracy, and validation accuracy. From these figures, we can calculate which module gives the best accuracy at which stage and capture the best weight of the module to improve the accuracy of the model. We only run the model till ten iterations. Here, DensetNet121 shows a little higher accuracy compared to other models.

5.6 Deployment of Image classification model on the flask

From the different metric values, confusion matric, and accuracy curve, we concluded that DenseNet121 gives a better result than any other module. So, we choose this architecture for report generation. We build a web application using flask API. In the final report, we can see patient information and the possible disease they may have. There is some text placeholder section where a doctor can write the review. A doctor can see the pictures on the right side of the report, which helps them analyze. They also collect then information from patient history information. After all, analyze the report's signature and print it, and give it to the patient with a prescription.

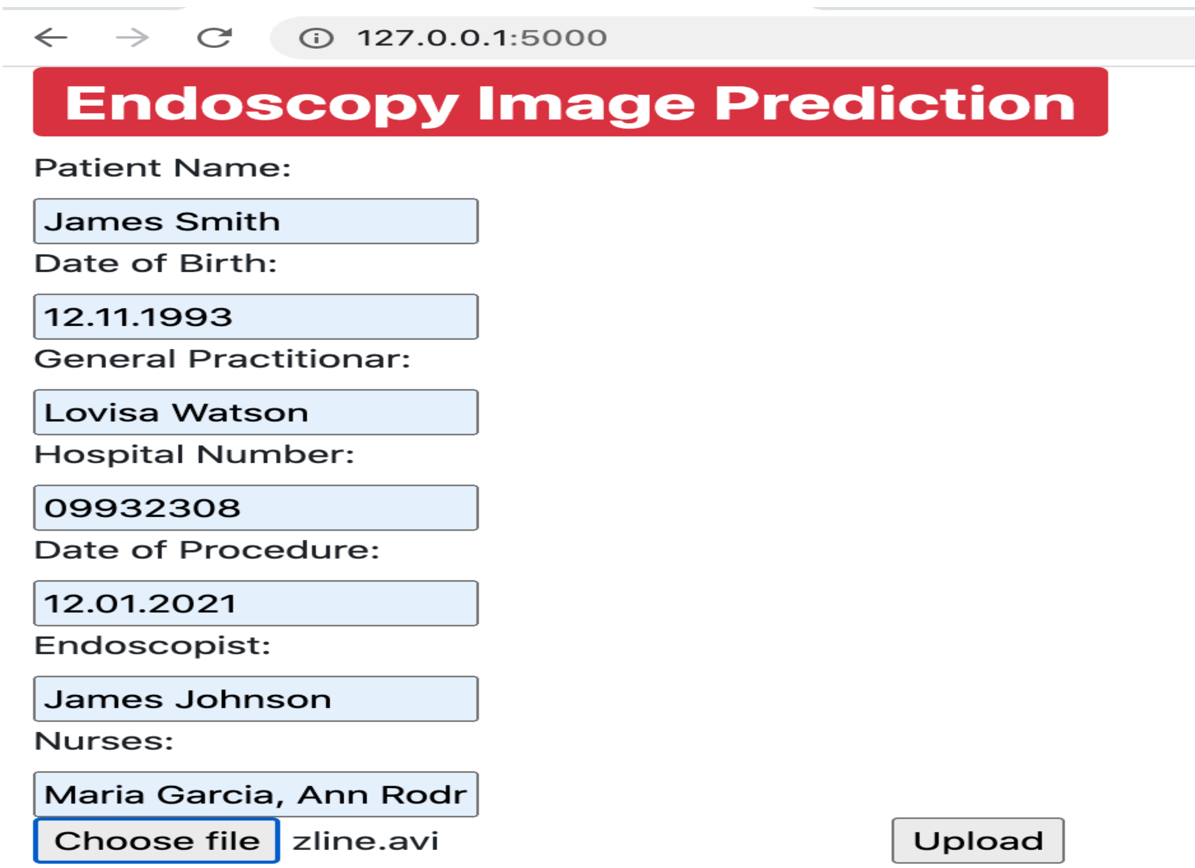


Figure 5. 13 Patient's registration webpage

The above figure shows a webpage is for the registration of patient's information, which contains the endoscopy videos/images. The videos/images are taken from the Endoscopist. Then it is uploaded to the system. Sample images or images in the videos which are uploaded are showed in the figures below.

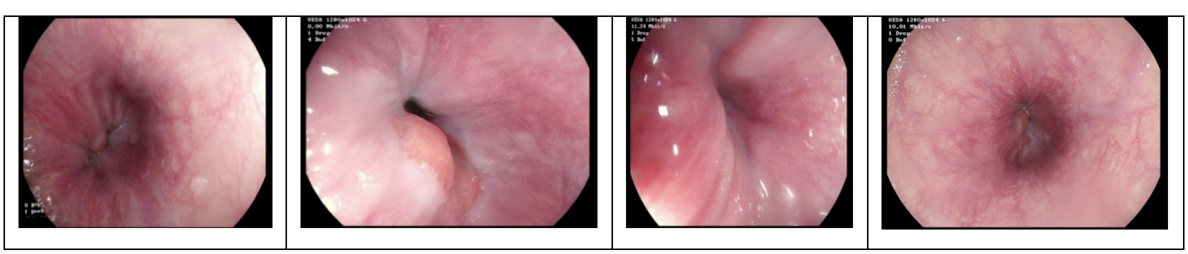


Figure 5. 14 Sample images in the uploaded video

After predicting the model, the endoscopy report will be generated automatically for the GI tract using CNN.

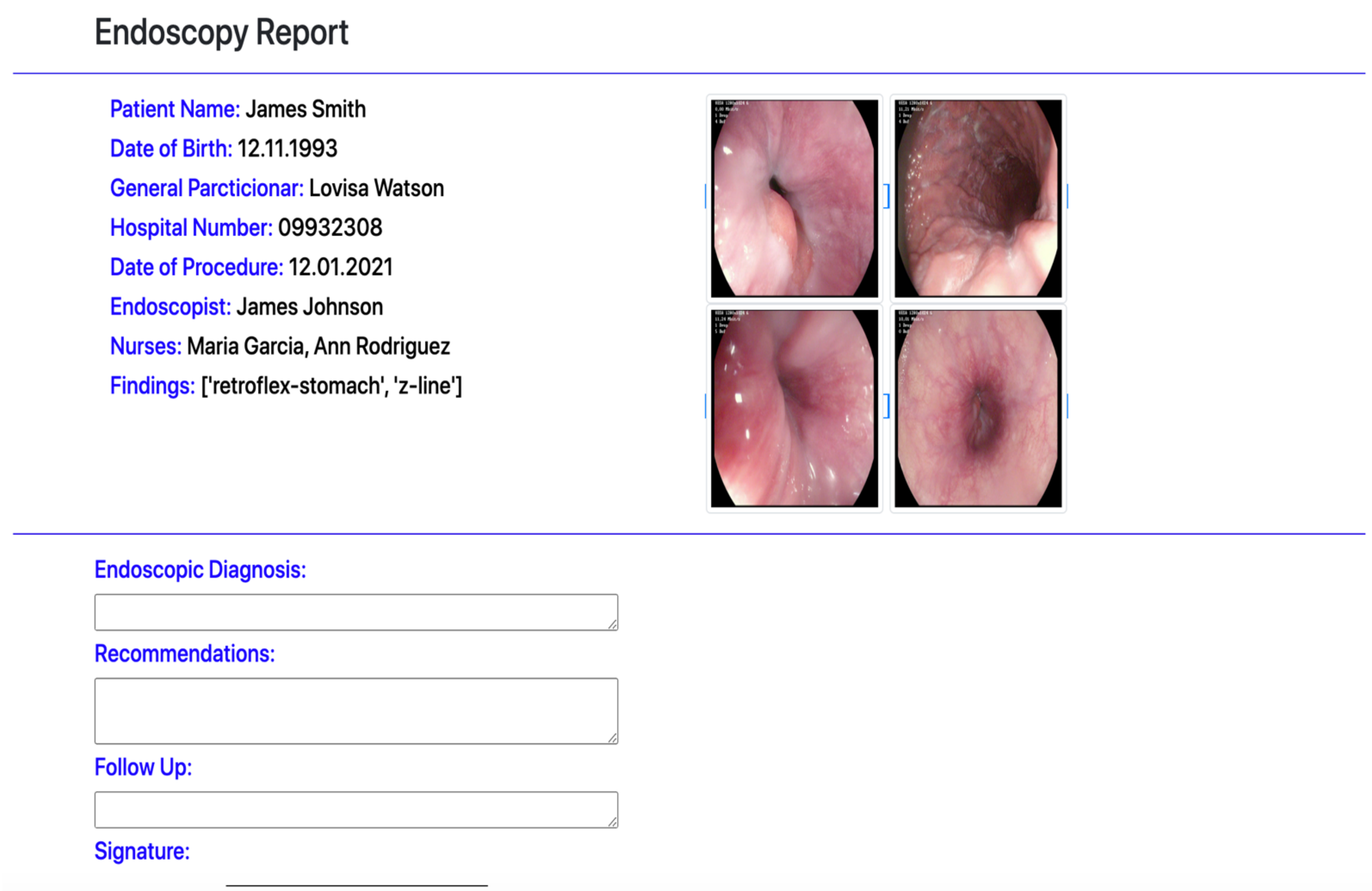


Figure 5. 15 Final report generated from the module

5.7 Universal Design Principles as Considered for Making the Report System (Website)

As we mentioned in chapter 2, we used a universal design approach for our report system (website) to achieve:

- Universal design provides access and creates solutions that serve the needs of all users, including those with and without limitations.
- Everyone benefits from Universal design, not just the disabled and the aging population.
- Individuals with hearing difficulties require captions to access information delivered through audio. When captions are supplied, hearing-impaired learners may well have a better understanding of the information.
- Persons utilizing screen readers and people with learning difficulties will benefit from a website that is structured with headings, well-organized material, and keyboard navigating. Everyone will find it much easier to understand and more visually appealing.
- Greater accuracy in identifying text characters driven by contrast rather than color. Darker text on a lighter background produces higher accuracy than lighter text on a dark background.

We mentioned earlier the principles of universal design in detail. Specifically, we followed the following universal design principles for making our report system (website):

1) **Usage that is appropriate**. People of various capacities will find the design helpful and desirable. For example, our website that created to be accessible for everybody, including persons who've been blind, and use screen reader software. Uses alternative text for images which helps screen readers while explaining images. Font of content size and shape are also considered significant and clear to viewed. Appropriate color contrast also helps everyone, especially visually challenged people.
2) **Flexibility in use**. Individual tastes and abilities are accommodated by design. Our website, for example, provides doctors the option of using images or video to an explanation of a medical case.

3) **Simple and easy to use**. Independent of the user's experience, expertise, language abilities, or existing concentrations, our website provides a simple design to use.
4) **Information that is visible**. Independent of environmental conditions or the participant's sensory abilities, the design layout should efficiently transmit important information to the user. Detailed medical images of the video/anomalies images on our website are an illustration of this approach.
5) **Tolerance for mistakes**. Our website provides advice when a user makes an incorrect choice is an example of this approach. Ask for video or images choices and shows an error if the incorrect format is uploaded.
6) **Physical activity is minimal**. The design allows for efficient, comfortable, and fatigue-free operation. Website is straightforward to handle, so anyone with a bit of understanding of computers can run the application. Shortcuts uploads are there no need to move from keyword to mouse for the simpler task.
7) **Size and space for use**. This principle has been applied to our website for medical staff who really are left-handed/right-handed with a range of other physical features and abilities.

5.8 Accessibility Evaluation of Report

We deployed a web-based application to develop our report system. We use the medical standard format as in (Bretthauer et al., 2016) as a designing principle to generate the report form structure. We make all text very clear and easy to access or understand for everyone. Every person can access this form easily. The font color, font size, and text alignment are in a standard format, making it accessible to everyone. Technical and non-technical doctors can easily use the report system to generate the report.

There are two testing purposes automatic testing and manual testing. Here We use automated testing for our webpage. We check our webpage with WAVE, which is a web accessibility tool. The results show that our website maintains and meets the WCAG 2.1

standard requirements (or guideline factors) as follows.

- **Perceivable**: every element on our website can be interpreted in different ways. A screen reader, for example, will read website content if someone is blind.
- **Operable**: every element on our website can be used in different ways. For example, both mouse and keyboard can be used.
- **Understandable**: every element on our website can be clearly understandable. The error message, for instance, indicates the position of the error as well as how to correct it.
- **Robust**: every element in our website can keep up with the latest hardware and software without cracking.

But the results also show some minor problems as shown in the figure.

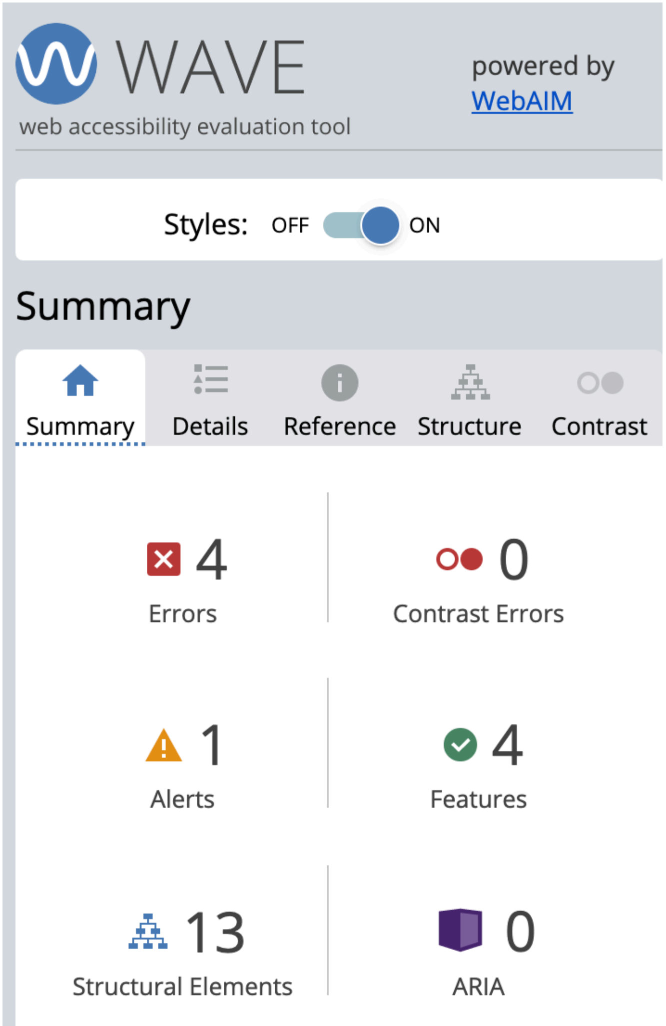


Figure 5. 16 Output generate by WAVE

4 Errors:

only 4 errors were discovered.

3 errors of the 4 errors are for missing form labels, labels are missing because we left the placeholder to be filled by a doctor after print.

And 1 error is for language missing or invalid, and we updated our report and modified this error using <html lang=”en”>

1 Alerts:

1 Alert was found.

1 alert is for no page regions, and we updated our report and eliminated this alert while printing a report.

5.9 Summary

In this chapter, we implemented a CNN system on the GI tract for analyzing different architectures such as VGG16, Inception v3, DenseNet 121, which were trained on HyperKvasir datasets, then compared results according to various evaluation metrics. For the three models, we get excellent classification results using CNN in GI tract, but we have the best results for DenseNet121 with PREC= 0.887, REC =0.861, ACC = 0.916, MCC= 0.909 and F1= 0.873. DenseNet121 gives better results than any other model, so it is selected for report generation purposes. We follow the universal design principle and check accessibility checks using WAVE. From the previous results, this research improved the automatic reporting and documenting system of the GI tract using network CNN to presents the output of medical reports with minimum assistance from the medical staff and with a high level of accuracy and some understandable form which contains multimedia elements such as images or videos. This will help for easy retrieval and processing of medical reports and documents besides reusing data for even teaching and research and giving visual representations of deep neural network layers to increase understanding, trust, and usefulness of disease diagnosis and detection procedures.

CHAPTER 6

Conclusion and Further Work

6.1 Summary

Deep neural networks are used for various situations, varying from automatic disease detection through images. These strategies are rarely understood, however. When addressing minor significance issues, including the automatic differentiation among human GI tracts, this common flaw may not be an issue. But when implementing these methodologies to problems where an error may lead to life-altering implications, we sometimes need some rationale on why these algorithms recommend a conclusive output. As a result, less complicated analytical designs are often recommended, although they may be less effective. With these primary problems, we have researched and developed an automatic endoscopy reporting plan that relies on transparency and knowledge of an in-depth CNN assessment to identify and treat diseases found in the GI tract. This overall improvement is used to create standard endoscopic compliant reports that the user can edit and format. A study is conducted to see how a good understanding of the internal in-depth neural network process can lead to the development of strategies for evaluating the classification of designs using Densenet121. These were done by analyzing models based on CNN architecture trained on an endoscopic image dataset known as HyperKvasir. By analyzing each network, preprocessing methods were extracted that could boost the model efficiency. The result showed the exponential performance of the methodology using the performance metrics. A user-friendly and accessible report is finally generated.

6.2 Contributions

This research improved the automatic reporting and documenting system of the GI tract using CNN, and it presented the output of medical reports with minimum assistance from the medical staff and with a high level of accuracy and some of the understandable form which contains multimedia elements such as images or videos. This would help

for easy retrieval and processing of medical reports and documents besides reusing data for even teaching and research and giving visual representations of deep neural network layers to increase understanding, trust, and usefulness of disease diagnosis and detection procedures.
We followed the rules or requirements while implementing the automatic reporting system of using CNN for the GI Tract:

1. The system should be easy to understand by technical and non-technical users (how to use CNN in reporting system).
2. Everyone benefits from universal design, not just the disabled or the aging population.
3. The system should use the images dataset GI tract to produce the medical documentation.
4. The system should be easy to execute and use in actual medical environments.
5. The system can improve the performance of the current reporting system effectively.

We have achieved the following objectives:

1) *Studying the existing works and show the advantages and disadvantages of utilizing CNN in the medical industry of automatic reporting procedure.*

   This objective is supported by chapter 2 and chapter 3. In chapter 2, we discuss a literature review where we discuss in detail about prior work, advantages, and disadvantages of the CNN model in the medical domain. In chapter 3, we discuss the requirements of a reporting system.

2) *Using CNN for the GI tract to detect anomalies.*

   This objective is supported by chapter 4. Chapter 4 developed the three CNN model, which helps to detect anomalies in the GI tract. Out of the three models, we select DenseNet121 because of its higher accuracy compared tothe other two models.

3) *Generating justifiable examination of identification reports in diagnosing the GI tract to visual effects and process contributes.*

   This objective is supported by chapters 4 and chapter 5. In chapter 4, We discuss how to develop the report system using a web application, and in chapter 5, we see the generated output of the report. Here we use the flask to develop the web application.

4) *Improving the existing automatic process's effectiveness to provide medical reports for the GI tract through CNN.*

   This objective is supported by chapters 3 and chapter 4. In chapter 3, we discuss the drawbacks of using CNN and the requirements for the report system. In chapter 4, we developed a new model to detect anomalies in the GI tract.

5) *Accessibility checks of the web-based report system*

   This objective is supported by chapter 5. In chapter 5, we try to consider all universal design approaches to make a website and check the accessibility of the generated report. We check our web report with WAVE, which is a web accessibility tool.

We implemented a CNN system with different architectures such as VGG16, Inception v3, DenseNet 121 for analyzing the GI tract. Each architecture was trained on HyperKvasir datasets, then compared results according to various evaluation metrics. From the different metric values, confusion metrics, and accuracy curves, DenseNet121 gives better results than any other model, so it is selected for report generation purposes. After that, we designed a user-friendly web-based report system considering the universal design principle and then checked its accessibility using the WAVE web accessibility tool.

6.3 Future works

In the future, we will concentrate more on how to improve the reporting system's effectiveness. We will try to add features including automated text recommendations, support for various reporting templates, and a more comprehensive framework for generating and exporting produced reports to help with this. We are intent in the future to test the proposed reporting system manually and gathering the options of the doctors and the clinics who are going to use the proposed system and the reporting system. We will add more features and extend the universal design principles. Furthermore, to enhance the medical reporting system, it would be promising to more investigation of the experiments using the neural network dissection tool might be helpful for the quality of HyperKvasir. In the future, we will explore other preprocessing steps that could result in increased performance measures if trained through further inspection. In addition to preprocessing steps, it would be promising to erase the text on some of the images, completely removing the green box located in the lower right corner of some of the pictures. Finally, remove all the boundaries by cropping the actual photo without any black borders. Many deep architectures have also demonstrated promising results on image-related challenges. It would be helpful to apply other deep learning models for medical images.

References

Abiyev, R. H., & Ma'aitah, M. K. S. (2018). Deep convolutional neural networks for chest diseases detection. Journal of healthcare engineering, 2018.

Apuke, O. D. (2017). Quantitative research methods: A synopsis approach. Kuwait Chapter of Arabian Journal of Business and Management Review, 33(5471), 1-8.

Arjmand, A., Angelis, C. T., Christou, V., Tzallas, A. T., Tsipouras, M. G., Glavas, E., . . . Giannakeas, N. (2020). Training of Deep Convolutional Neural Networks to Identify Critical Liver Alterations in Histopathology Image Samples. Applied Sciences, 10(1), 42.

Arulkumaran, K., Deisenroth, M. P., Brundage, M., & Bharath, A. A. (2017). Deep reinforcement learning: A brief survey. IEEE Signal Processing Magazine, 34(6), 26-38.

Borgli, H., Thambawita, V., Smedsrud, P. H., Hicks, S., Jha, D., Eskeland, S. L., . . . Nguyen, D. T. D. (2020). HyperKvasir, a comprehensive multi-class image and video dataset for gastrointestinal endoscopy. Scientific Data, 7(1), 1-14.

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

Bretthauer, Michael & Aabakken, Lars & Dekker, Evelien & Kaminski, Michal & Rösch, Thomas & Hultcrantz, Rolf & Suchanek, Stepan & Jover, Rodrigo & Kuipers, Ernst & Bisschops, Raf & Spada, Cristiano & Valori, Roland & Domagk, Dirk & Rees, Colin & Rutter, Matthew. (2016). Reporting systems in gastrointestinal endoscopy: Requirements and standards facilitating quality improvement: European Society of

Gastrointestinal Endoscopy position statement. United European Gastroenterology Journal. 4. 172-176. 10.1177/2050640616629079

Can J Gastroenterol. Endoscopy reporting standards, 2013 May; 27(5): 286–292

Demir, A., Yilmaz, F., & Kose, O. (2019). Early detection of skin cancer using deep learning architectures: resnet-101 and inception-v3. 2019 Medical Technologies Congress (TIPTEKNO).

Deng, J., Dong, W., Socher, R., Li, L.-J., Li, K., & Fei-Fei, L. (2009). Imagenet: A large-scale hierarchical image database. The 2009 IEEE conference on computer vision and pattern recognition.

Esteva, A., Kuprel, B., Novoa, R. A., Ko, J., Swetter, S. M., Blau, H. M., & Thrun, S. (2017). Dermatologist-level classification of skin cancer with deep neural networks. Nature, 542(7639), 115-118.

François-Lavet, V., Henderson, P., Islam, R., Bellemare, M. G., & Pineau, J. (2018). An introduction to deep reinforcement learning. Foundations and Trends® in Machine Learning, 11(3-4), 219-354.

Gamage, C., Wijesinghe, I., Chitraranjan, C., & Perera, I. (2019). GI-Net: anomalies classification in gastrointestinal tract through endoscopic imagery with deep learning. The 2019 Moratuwa Engineering Research Conference (MERCon).

Gao, X., & Bie, H. (2018). Wide & ResNet: An Improved Network for CTR Prediction. The Proceedings of the 2018 International Conference on Algorithms, Computing and Artificial Intelligence.

Gulli, A., & Pal, S. (2017). Deep learning with Keras: Packt Publishing Ltd.

He, K., Zhang, X., Ren, S., & Sun, J. (2016). Deep residual learning for image recognition.

The Proceedings of the IEEE conference on computer vision and pattern recognition.

He, X., Wu, L., Song, F., Jiang, D., & Zheng, G. (2020). Research on Fabric Defect Detection Based on Deep Fusion DenseNet-SSD Network. The Proceedings of the International Conference on Wireless Communication and Sensor Networks.

Hennink, M., Hutter, I., & Bailey, A. (2020). Qualitative research methods: SAGE Publications Limited.

Hicks, S. A. (2018). Mimir: An Automatic Reporting and Reasoning System for Screening of the Gastrointestinal Tract Using Deep Neural Networks (Master's thesis).

Hicks, S. A., Eskeland, S., Lux, M., de Lange, T., Randel, K. R., Jeppsson, M., . . . Riegler, M. (2018). Mimir: an automatic reporting and reasoning system for deep learning based analysis in the medical domain. The Proceedings of the 9th ACM Multimedia Systems Conference.

Hicks, Steven & Riegler, Michael & Pogorelov, Konstantin & Ånonsen, Kim V. & de Lange, Thomas & Johansen, Dag & Jeppsson, Mattis & Randel, Kristin & Eskeland, Sigrun & Halvorsen, Pål. (2018). Dissecting Deep Neural Networks for Better Medical Image Classification and Classification Understanding. 10.1109/CBMS.2018.00070.

Hicks, Steven & Smedsrud, Pia & Riegler, Michael & de Lange, Thomas & Petlund, Andreas & Eskeland, Sigrun & Pogorelov, Konstantin & Schmidt, Peter & Halvorsen, Pål. (2019). 383 DEEP LEARNING FOR AUTOMATIC GENERATION OF ENDOSCOPY REPORTS. Gastrointestinal Endoscopy. 89. AB77. 10.1016/j.gie.2019.04.053.

Hossin, M., & Sulaiman, M. (2015). A review on evaluation metrics for data classification evaluations. International Journal of Data Mining & Knowledge Management Process, 5(2), 1.

Jha et al., (2021). A comprehensive analysis of classification methods in gastrointestinal endoscopy imaging. Medical Image Analysis. 70. 102007. 10.1016/j.media.2021.102007.

Jing, B., Xie, P., & Xing, E. (2017). On the automatic generation of medical imaging reports. arXiv preprint arXiv:1711.08195.

Kabir, G., & Hasin, M. A. A. (2013). Comparative analysis of artificial neural networks and neuro-fuzzy models for multicriteria demand forecasting. International Journal of Fuzzy System Applications (IJFSA), 3(1), 1-24.

Kisilev, P., Walach, E., Barkan, E., Ophir, B., Alpert, S., & Hashoul, S. Y. (2015). From medical image to automatic medical report generation. IBM Journal of Research and Development, 59(2/3), 2: 1-2: 7.

Matera, Maristella & Rizzo, Francesca & Carughi, Giovanni. (2006). Web Usability: Principles and Evaluation Methods. 10.1007/3-540-28218-1_5.

Mitchell, T. M. (1997). Machine learning: McGraw-hill New York.

M. Riegler et al., "EIR — Efficient computer aided diagnosis framework for gastrointestinal endoscopies," 2016 14th International Workshop on Content-Based Multimedia Indexing (CBMI), 2016, pp. 1-6, doi: 10.1109/CBMI.2016.7500257.

M. Riegler et al.,. (2017). From Annotation to Computer-Aided Diagnosis: Detailed Evaluation of a Medical Multimedia System. ACM Transactions on Multimedia Computing, Communications, and Applications. 13. 1-26. 10.1145/3079765.

Nash, W., Drummond, T., & Birbilis, N. (2018). A review of deep learning in the study of materials degradation. npj Materials Degradation, 2(1), 1-12.

Nilsson Nils, J. (1998). Introduction To Machine learning. Robotics Laboratory Department of Computer Science Stanford University, 1(10).

Osamu, I., Fahdi, K., Kei, K., Rambeau, M., Koji, A., & Masayuki, T. (2020). Deep Learning Models for Histopathological Classification of Gastric and Colonic Epithelial Tumours. Scientific Reports (Nature Publisher Group), 10(1).

Pannu, H. S., Ahuja, S., Dang, N., Soni, S., & Malhi, A. K. (2020). Deep learning based image classification for intestinal hemorrhage. MULTIMEDIA TOOLS AND APPLICATIONS.

Pogorelov, Konstantin & Riegler, Michael & Halvorsen, Pål & Eskeland, Sigrun & de Lange, Thomas & Griwodz, Carsten & Randel, Kristin & Stensland, Håkon & Dang Nguyen, Duc Tien & Spampinato, Concetto & Johansen, Dag. (2017). A Holistic Multimedia System for Gastrointestinal Tract Disease Detection. 112-123. 10.1145/3083187.3083189.

Schmidhuber, J. (2015). Deep learning in neural networks: An overview. Neural networks, 61, 85-117.

Sharif, M., Attique Khan, M., Rashid, M., Yasmin, M., Afza, F., & Tanik, U. J. (2019). Deep CNN and geometric features-based gastrointestinal tract diseases detection and classification from wireless capsule endoscopy images. Journal of Experimental & Theoretical Artificial Intelligence, 1-23.

Sheryl B. (2021). Universal Design: Process, Principles, and Applications, University of Washington, 2021.

Simonyan, K., & Zisserman, A. (2014). Very deep convolutional networks for large-scale image recognition. arXiv preprint arXiv:1409.1556.

Szegedy, C., Vanhoucke, V., Ioffe, S., Shlens, J., & Wojna, Z. (2016). Rethinking the inception architecture for computer vision. The Proceedings of the IEEE conference on computer vision and pattern recognition.

Szepesvári, C. (2010). Algorithms for reinforcement learning. Synthesis lectures on artificial intelligence and machine learning, 4(1), 1-103.

Tabian, I., Fu, H., & Sharif Khodaei, Z. (2019). A convolutional neural network for impact detection and characterization of complex composite structures. Sensors, 19(22), 4933.

Tam, Clara, "Machine Learning towards General Medical Image Segmentation" (2020). Electronic Thesis and Dissertation Repository. 6897. https://ir.lib.uwo.ca/etd/6897

Townsend, J. T. (1971). Theoretical analysis of an alphabetic confusion matrix. Perception & Psychophysics, 9(1), 40-50.

Vertzoni, M., Augustijns, P., Grimm, M., Koziolek, M., Lemmens, G., Parrott, N., . . . Van Den Abeele, J. (2019). Impact of regional differences along the gastrointestinal tract of healthy adults on oral drug absorption: an UNGAP review. European Journal of Pharmaceutical Sciences, 134, 153-175.

Vu, H., Manh, X. H., Duc, B. Q., Ha, V. K., Dao, V. H., Nguyen, P. B., . . . Vu, T. H. (2019). Labelling Stomach Anatomical Locations In Upper Gastrointestinal Endoscopic Images Using a CNN. The Proceedings of the Tenth International Symposium on Information and Communication Technology.

https://www.w3.org/WAI/standards-guidelines/aria/

Xu, Z., Tao, Y., Wenfang, Z., Ne, L., Zhengxing, H., Jiquan, L., . . . Jianmin, S. (2019). Upper gastrointestinal anatomy detection with multi-task convolutional neural networks. Healthcare Technology Letters, 6(6), 176-180.

Yadav, S. S., & Jadhav, S. M. (2019). Deep convolutional neural network based medical image classification for disease diagnosis. Journal of Big Data, 6(1), 113.

Yuan, J., Liao, H., Luo, R., & Luo, J. (2019). Automatic Radiology Report Generation based on Multi-view Image Fusion and Medical Concept Enrichment. The International Conference on Medical Image Computing and Computer-Assisted Intervention.

Zhang, Z., Xie, Y., Xing, F., McGough, M., & Yang, L. (2017). Mdnet: A semantically and visually interpretable medical image diagnosis network. The Proceedings of the IEEE conference on computer vision and pattern recognition.

Zane Robinson Wolf; Ronda G. Hughes. Patient Safety and Quality: An Evidence-Based Handbook for Nurses, Chapter 35Error Reporting and Disclosure, Hughes RG, editor. Rockville (MD): Agency for Healthcare Research and Quality (US); 2008 Apr.

Appendix

Source code

The source code for this project is located at
https://github.com/matrikasubedi/Disease_Prediction_in_Gastrointestinal_tract_using_Endoscopy_videos