# Using unsupervised machine learning as a tool for polyp detection in the GI tract

Mathias Kirkerd

Thesis submitted for the degree of
Master of science in Informatics: Technical and Scientific Applications
60 credits

Department of Informatics
Faculty of mathematics and natural sciences

UNIVERSITY OF OSLO

Spring 2019

# Using unsupervised machine learning as a tool for polyp detection in the GI tract

Mathias Kirkerd

Using unsupervised machine learning as a tool for polyp detection in the GI tract

`http://www.duo.uio.no/`

Printed: Reprosentralen, University of Oslo

# Abstract

# Acknowledgements

my cat, if i had one

# Contents

# List of Figures

# List of Tables

# Chapter 1

# Introduction

## 1.1 Background and Motivation

### 1.1.1 Introduction REM

Cancer is, today, the second leading cause of death in the world, only behind cardiovascular diseases.
It is one of the leading causes of mortality worldwide, with approximately 14 million new cases in 2012. It is defined as a disease that has an abnormal cell growth with the potential to spread into other parts of the body.Contrary to normal cells, cancer cells are often invasive, and it will spread if not treated. In contrast to many other diseases cancer does not start from a foreign entity (such as a bacteria or virus), but it is often from a malfunctioning cell that starts dividing rapidly. This can happen when a cell is damaged, by for instance by radiation or other factors that damages the DNA, and the resulting damage causes the cell to uncontrollably divide. Especially in the later part of life everyone has the chance of getting cancer, and in fact everyone does. Our own body is designed to detect and remove cells that are prone to divide uncontrollably. Unfortunately this system is not perfect, and the immune system can in some cases overlook cells that are cancerous.

### 1.1.2 Statistics on cancer REM

The western (or modern) world has been in a battle against cancer, and despite a lot of new cures/innovations it is still one of the deadliest killers in the world. *The most common types of cancer in males are lung cancer, prostate cancer, colorectal cancer and stomach cancer.***stewart2014world**

### 1.1.3 colorectal cancer REM

You can get cancer in every major organ, but some types of cancer are more common than others. For instance cancer in the gastrointestinal tract (GI) is one of the more common places to get cancer. This is just behind x, and it has a mortality rate of x in the first y years. We often call this 5 year survival rate for z. This is the standard way to measure the life expectancy of a patient diagnosed with cancer.

### 1.1.4 polyps REM

The colorectal cancer often starts in polyps. Polyps are, polyps do.

### 1.1.5 preventative matters and early detection REM

*-colonoscopy*
*-mri*
*-pillcam*
A good way to fight cancer is to detect and remove it early, or some times remove areas with a high chance of getting cancer. We classify cancer in to x stages, and the stage the patient are in often determines the chance you have for survival. In general, the earlier you find the cancer, the more likely it is that the patient will survive. And as mentioned above, the colorectal cancer often starts in these polyps. A crucial stage to prevent cancer lies in the early removal of there polyps. Reports shows x about this

*4 stages maybe? *early detection *survival rate

Because of this the ability to find, and remove colorectal polyps is great for preventing cancer in the GI tract.

**colonoscopy/Ontonoscopy** In the most common way to look for polyps in the GI tract is to use a medical team, and perform a colonoscopy or Ontonoscopy colonoscopy is preformed with a camera-stick that is inserted in to the GI tract through the patients anus.
Onoskopy is the same procedure, only the camera is inserted orally.

**Advantages**

- Accuracy: The use of a camera controlled by the doctor gives him/her the opportunity to stop at any anomalies.

- Quick results: Since the doctor is doing the procedure the result is given live.

**Disadvantages**

- Expensive: The cost of the doctor and the nurses needed is often high, especially on a routine check.
- Invasion of privacy: Getting an Colonoscopy or Onoskopy is a

**MRI** MRI (Maggnetic stuff) is the act of taking pictures blabla blabla
MRI (Maggnetic stuff) is the act of taking pictures blabla blabla
MRI (Maggnetic stuff) is the act of taking pictures blabla blabla
**Advantages**

- This is why mri is good
- This is why mri is good

**Disadvantages**

- This is why mri is bad
- This is why mri is bad

**pillcam** In the last 3-4 years there have been testing and development on the pillcam project EIR. Machine learning has, through many of the earlier projects, got the detection rate for the polyps up to x%

**Advantages**

- This is why pillcam is good
- This is why pillcam is good

**Disadvantages**

- This is why cam is bad
- This is why pillcam is bad

### 1.1.6 Simulas contribution to the pillcam project REM

Simulas EIR

* CAD ACD (computer aided diagnosis, Automated computer diagnosis)

## 1.2 Goal / Problem

### 1.2.1 pillcam project has lots of data, can be used to train an unsupervised network REM

The video sequence from the pillcam can last several hours resulting in thousands of images, combined with colonoscopy images we have over 60 000 unlabelled images at our disposal.

### 1.2.2 Use Unsupervised learning as a pre-processing tool REM

The act of finding an algorithm that can enhance the training data. Either through removing artifacts or virtually enhancing resolution.

### 1.2.3 use Unsupervised-NN/GAN for image enhancements so that a NN can train better REM

* Now that we got a lot of tests, why not unsupervised As mentioned, simula research centre has done a lot of testing on the pillcam project.

* We know that we can get some results using a neural network * Can this be done unsupervised? * Can it be done in a fashion that is better than S-ML

REM

## 1.3 Scope and Limitations

### 1.3.1 Use Unsupervised NN to find polyps REM

### 1.3.2 Use Unsupervised NN for pre-processing REM

* Something about earlier research already got far, so the scope is mainly unsupervised deep learning. * (and how to generalise it?) *REMegression

## 1.4 Research method

## 1.5 Related work

## 1.6 Outline

The rest of the thesis is structed as follows:
**Chapter 2 - Background**
*talk about cancer *talk about machine learning. *how to use ML on the pillcam video? **Chapter 3 - Me doing stuff**
**Chapter 4 - Me got and present result**
**Chapter 5 - Me saying result was good A+**

# Chapter 2

# Background

In this chapter we will present the background and motivation of our thesis. We start with our background in medical procedures, looking on how doctors perform colonoscopies, mainly from a gastrointestal perspective. Then will then look at what the objective is for the medical staff, with different anomalies in the GI tract. Then our focus is moved to how doctors use computer aided diagnosis (CAD) today to help with the screening. Lastly we look at current models for CAD made both by simula.

We will then shift our focus to machine learning, and give a breaf introduction in different machine learning methods. Wtih this in mind we will look at neural network, especially convolutional neural networks, and how they work.

Lastly we will combine the need for computer aided diagnosis with the machine learning.

## 2.1 The Medical Background

In the field of medical diagnosis there are allways new and interesting methods beeing researched to help the medical staff when it comes to patient *rate of survival*, and quality of life. Everything from x to y is ways the medical instusty has done to improve the survival rates of their patients. In the last decade *comuters and cameras came to help us.* Another example is the invention and usage of gastro-stick-with-camera.

### 2.1.1 colonoscopy/gastro/procedure

When performin gastronomi we use astik

### 2.1.2 Medical images/data/other

### 2.1.3 Systems in place for detection

### 2.1.4 summary

## 2.2 Machine Learning

We have looked at the challenges that the medical staff has when it comes to detecting polyps, and how it is solved today. But to truly understand how automated systems like works, we need to look at Machine learing.

Machine learning is a very broad term, but can i short be summarised by:

> *A computer program is said to learn from experience E with respect to some class of tasks T and performance measure P, if its performance at tasks in T, as measured by P, improves with the experience E.* **MitchellTomM1997Ml**

Here we have a couple of parameters:
**E** text about e
**T** text about t
**P** text about p

From this we see that the goal of machine learning is to improve some performance P with experience. This is based on how humans, and in our case doctors learn. As the ammount of experience increase, the performance of the task should also increase. With this in mind, we can assume that, given the right ammount and type of data, our machine learning algorithms can solve universially complex problems, given a numerical output.

### 2.2.1 Machine learning types

With basis in the quote from **MitchellTomM1997Ml** , we have a broad definition of what machine leaning can be. As long as we have a model trying to complete a task based on previous experience, it can be called machine learing.

Here we see some of the most famous machine learning alogrithms, and their subsequent subcategory.

**K nearest neighbours**
The K nearest neighbours alorithm was proposed by, and is a method used for both classification and regression.

| Machine Learning | | | | |
|---|---|---|---|---|
| Supervised Learning | | Unsupervised Learning | | Reinforcement Learning |
| Classification | Regression | Clustering | Dimensionality reduction | - |
| Support vector machines | Linear Regression | K means clustering | PCA | SOMething |
| K nearest neighbours | Decision trees | Hidden Markov models | | |
| Neural networks | Neural networks | Neural Networks | | |

Table 2.1: Machine leaning types

**Linear Regression**
How to regress linearly

**Support vector machine**
SVM and 2 class

**Others?**
Other important ones to talk about?

**Neural networks**
NN is future
own chapter

### 2.2.2 How machine learning works

We can start with one of the simplest examples in machine learing: linear regression.
Linear regression, and regression in genera, is a typical task assigned to machine leading, given the simple input and output. In linear regression we want to make a model that can predict a value given an input.

The output, y, from the regression can be calculated with the general formula for a line.

$$y = ax + b \tag{2.1}$$

Or in the machine learing case:

$$y = W^{(1)}x + W^{(2)} \tag{2.2}$$

Where $W^{(1)}$ & $W^{(2)}$ Our goal is to find the optimal value for $W^{(1)}$ and $W^{(2)}$ so that the error between the predicted output data and the actual output data is as small as possible.

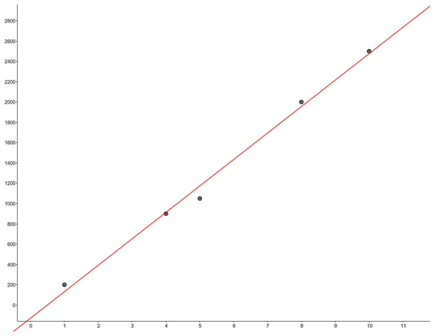


Figure 2.1: Example of linear regression in Geogebra. Here the red line is the best approxmiation of a y value, given an x value.

The most prominent way of calculating this error is to use the mean square error betweet the predicet and actual output of the data.

$$MSE = \frac{1}{2m} \sum_i (\hat{y} - y)_i^2 \tag{2.3}$$

Where $m$ is the number of samples, $y$ is the real output, and $\hat{y}$ is the prediced output. The 2 in the dinominator is just a constant to make derivation of the formula easier.

From this we can intuitivly see that the error tends towards 0 when $\hat{y}$=$y$. We can also note, because of the squaring in the formula, that the error is ony based on L2 distance between $\hat{y}$ and $y$.

Now that we have an error, we need a way to improve it

### 2.2.3 Example with gradient decent

Now that we have a model with an error function, we can see how we would go on to change the weights ($W^{(1\&2)}$) of our model, to get a better result.

Lets start with:

$x = \begin{bmatrix} 1 \\ 2 \\ 3 \end{bmatrix}$ and $y = \begin{bmatrix} 1.5 \\ 2 \\ 2.5 \end{bmatrix}$ with the weights $W^{(1)} = 0$ and $W^{(2)} = 0$

We can first calculate the initial loss of the model given a MSE. Using 2.3 gives us a loss of:

$$\frac{1}{2*3}(1.5^2 + 2^2 + 2.5^2) = 2.08 \tag{2.4}$$

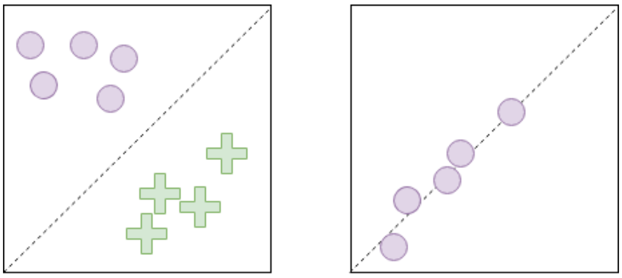


Figure 2.2: Left: Example of binary classification. Right: Example of regression

We will now use gradient decent to estimate

#### 2.2.3.1 Feed forward

#### 2.2.3.2 Loss and gradient decent

### 2.2.4 Supervised & Unsupervised machine learning

We often divide machine learning in to two (diffuse) categories: supervised and unsupervised.

**Supervised learning:** is the act of training with data that has an answer or a label. The learning algorithm can get supervision while training on the task. An example on a supervised task is to recognise handwritten numbers, or differentiate between dogs and cats. The task is supervised if the images comes with the correct label in the data set. These examples are typical classification examples, where the task is to identify the right group to classify the data to A simpler classification assignment is binary classification, where the target is (often) yes or no. Examples for binary classification is if an email is spam or not, is a car Norwegian or International. In the last example the classification changes from binary to multi-class if you sort the cars on every nationality, and not just Norwegian/non-Norwegian.

Another type of supervised learning is regression. This is the act of prediction given prior data. Examples of regression is everything from prediction of stock prices, to house prices in an area, to

**Unsupervised learning:** is the act of training without any supervision, on the sense that we do not give the algorithm the answer to the training data set.

Since we do not have categorised data in unsupervised learning, we often Types of unsupervised learning can for instance be clustering, the act of sorting data based on similarity. An example of this can be if you want to sort plants based on species, or you are detecting anomalies in a dataset. Unsupervised

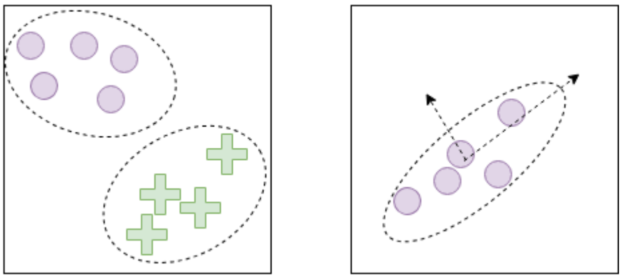


Figure 2.3: Left: Example of binary clustering. Right: Example of principal component analysis

learning can be used for PCA or other dimensionaly reduction methods.

A third method to used unsupervised learning is the adversarial route, where you use machine learning to make similar looking data to the original data set.

In the description of supervised vs unsupervised we looked at a specific branch of machine learning: Classification. Classification is, as the name implies, the task of getting data sorted in to groups of similarity.

- subsfication
- r to the pillcam projression
- transcription/translation
- de-noising /finding missing inputs

## 2.3 Neural Networks

### 2.3.1 How it works

TEXT ABOUT NEURAL NETWORKS TEXT ABOUT NEURAL NETWORKS TEXT ABOUT NEURAL NETWORKS TEXT ABOUT NEURAL NETWORKS

TEXT ABOUT FEED FORWARD TEXT ABOUT FEED FORWARD TEXT ABOUT FEED FORWARD

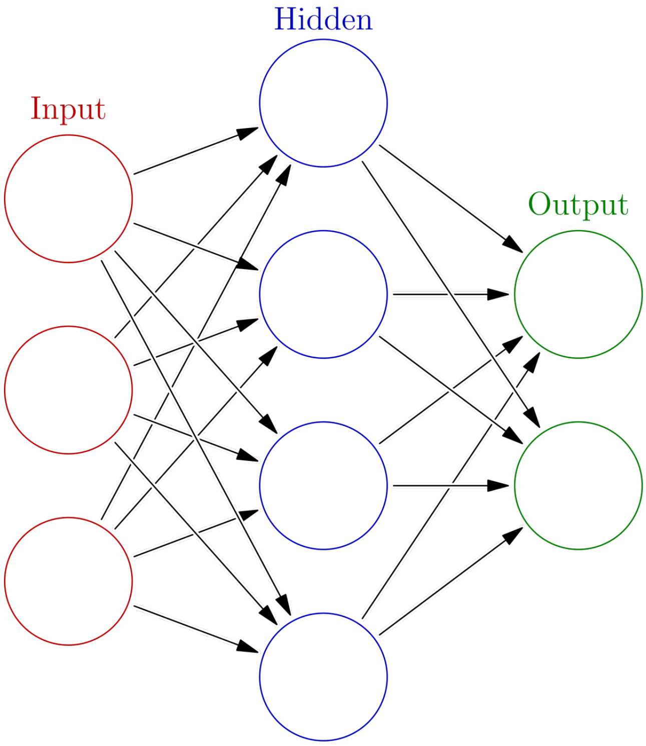


Figure 2.4: THIS IMAGE IS SHAME(LESS)LY taken from the internetz, draw own so the lawyers don't get you!

TEXT ABOUT FEED FORWARD

TEXT ABOUT BACKPROP
TEXT ABOUT BACKPROP
TEXT ABOUT BACKPROP
TEXT ABOUT BACKPROP

### 2.3.2 Convolutional neural networks

### 2.3.3 Advaserial neural networks

**2.3.3.0.1 This is explaining GANS, put me in the right place** Now that we have looked at autoencoders we can take it a step further. generative advaserial models can be used as a generator of new data, and can have som reseblance to autoencoders 2.4.1, especially variational autoencoders

The difference lies in that advaserial networks is based on game theoretic

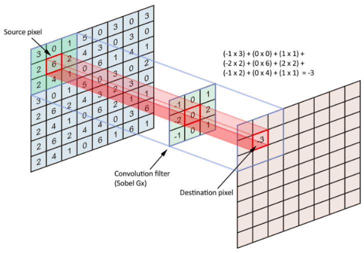

Source pixel
(-1 x 3) + (0 x 0) + (1 x 1) +
(-2 x 2) + (0 x 6) + (2 x 2) +
(-1 x 2) + (0 x 4) + (1 x 1) = -3
Convolution filter
(Sobel Gx)
Destination pixel

Figure 2.5: THIS IMAGE IS SHAME(LESS)LY taken from the internetz, draw own so the lawyers don't get you!

scenarios in which a generator network is compeating agenst an advasery. The generator produces samples $x = g(z;\theta^{(g)})$, where $g$ is the network given the weights $\theta$. Then the discriminator network predicts if a sample is drawn from the dataset or from the generator. More spessific, it gives a probably given by $d(x;\theta^{(d)})$ , and determins if $x$ is from the generator or the data-set. Since we have two networks compeating agenst each other we can look at this as a Zero-sum game with the generators payoff is determined by $v(\theta^{(g)},\theta^{(d)})$, and the discriminators payoff is determined by $-v(\theta^{(g)},\theta^{(d)})$. *$v$ is here a function that is determined by both the sucsess rate of the discriminator and the generator, the most common used is*

$$v(\theta^{(g)},\theta^{(d)}) = \mathbb{E}_{x\sim p_{data}} \log d(x) + \mathbb{E}_{x\sim p_{model}} \log\left(1 - d(x)\right) \tag{2.5}$$

as derived from Goodfellow et al.

Lets look at a gan in detail.

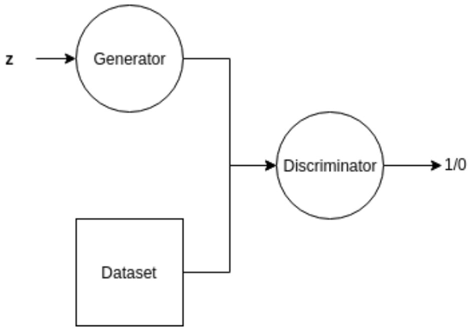


Figure 2.6: The idea behind a GAN. Here the generator saples from a random (Gaussian) distribution and generates samples that the discriminator classifies as real or fake

#### 2.3.3.1 UCNN?

### 2.3.4 Recurrent neural networks

#### 2.3.4.1 LSTM

## 2.4 Models we need to explain at this point (find better tittle)

### 2.4.1 Autoencoders

As we recall from earlier, an autoencoder is a type of neural network that tries to output a recreation of the output.

We can do this by having an encoder, $h = f(x)$, connected to a decoder, $r = g(h)$. An autoencoder has the job to set $g(f(x)) = x$ over the whole input, but in most cases this is not a practical program. We often gives the autoencoder the restriction that it has to map the input through a latent space that has a smaller dimension than the input dataset.
This is called an undercomplete autoencoder.

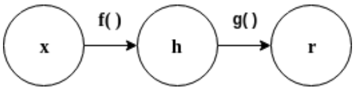


Figure 2.7: The general structure of an autoencoder, mapping **x** through **h** to an output **r**.

As with supervised classifiers we can use gradient decent to optimize the model. This is because we are trying to recreate the input **x** from out output $\widetilde{\mathbf{x}}$

This can simply be done by minimizeing the loss function

$$L(\mathbf{x}, g(f(\mathbf{x}))) \tag{2.6}$$

with for instance MSE with gradient decent.

Now we can transfer this to a more relevant example by making an image as input and use convolutions to reduce the dimensionality in the encoder and increase the dimentionality in the encoder.

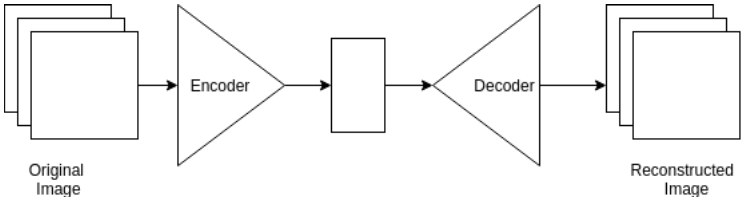


Figure 2.8: Convolutional autoencoder with an RGB image as input, and the reconstructed image as output.

### 2.4.2 Contextencoders

Inpainting can also be done with advaserial models, and using a network trained to do the task of inpainting can be a lot more powerful than using just an autoencoderor the naive methods. A contextencoder is building on the advaserial principle by using a generator/discriminator pair to fill in masked areas in an image.

The concept behind a Contextencoder is to take the whole image as input to an encoder/decoder pair and

### 2.4.3 CC-GANS

HERE IS TEXT ABOUT CCGANS HERE IS TEXT ABOUT CCGANS HERE IS TEXT ABOUT CCGANS HERE IS TEXT ABOUT CCGANS HERE IS TEXT ABOUT CCGANS HERE IS TEXT ABOUT CCGANS HERE IS TEXT ABOUT CCGANS HERE IS TEXT ABOUT CCGANS HERE IS TEXT ABOUT CCGANS HERE IS TEXT ABOUT CCGANS HERE IS TEXT ABOUT CCGANS HERE IS TEXT ABOUT CCGANS HERE IS TEXT ABOUT CCGANS HERE IS TEXT ABOUT CCGANS HERE IS TEXT ABOUT CCGANS HERE IS TEXT ABOUT CCGANS HERE IS TEXT ABOUT CCGANS HERE IS TEXT ABOUT CCGANS HERE IS TEXT ABOUT CCGANS HERE IS TEXT ABOUT CCGANS HERE IS TEXT ABOUT CCGANS HERE IS TEXT ABOUT CCGANS

### 2.4.4 Pixel CNN

HERE is text about pccn HERE is text about pccn HERE is text about pccn HERE is text about pccn HERE is text about pccn HERE is text about pccn HERE is text about pccn HERE is text about pccn HERE is text about pccn HERE is text about pccn HERE is text about pccn HERE is text about pccn HERE is text about pccn HERE is text about pccn HERE is text about pccn HERE is text about pccn HERE is text about pccn HERE is text about pccn HERE is text about pccn HERE is text about pccn HERE is text about pccn HERE is text about pccn

## 2.5 Explain how the ML-methods can be used with the polyps

When you work with machine learning a lot of the job is to make the data as clear as possible.
Imagine that you want to do something simple as reading an analogue clock. The s̈traight forwardẅay to do it is to make a convolutional neural network to look at the dials. This will require a much more complex network compared to if you could convert the angle of the dials to degrees and have that as an input to your model.

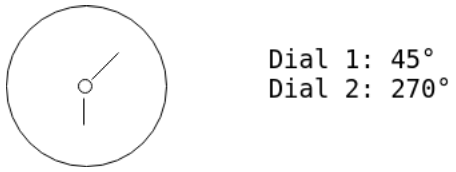


Figure 2.9: A clock needs a more complex network compared to just the degrees

The trick is often to make the data as refined as possible. Further some of the techniques used is described.

## 2.6 The problem at hand

Now that we have the definition of machine learning and the current task, we can focus on the task at hand; finding polyps. In an ideal world[1] we have a Classification problem with only two classes: Non-polyp and polyp.

- SVM
- CNN
- random forests
- knn

[1] Ideal as in the only disease we could get in the GI tract was cancer originating from polyps which looked exactly the same

# Chapter 3

# Methodology

With our background in both machine learning and we can now look at how we want to solve the problems associated with setting up a system for medical diagnosis.

## 3.1 Libraries

In this chapter, we will discuss the foundation of our code, important external libraries, and the setup and execution of our project. We will first discuss the programming language in question, give insight into the reasoning behind it. Then we will look into the framework used for machine learning, and in detail how it implemented in our programming language. Lastly, we will look into the wrapper we use to get a greater level of abstraction over our code, together with custom wrapper functions that are used by our wrapper.

### 3.1.1 python

When doing machine learning, the most popular languages, in no particular order, are: Python, Java, R, C++, and C . Some of these languages, like C and C++, are chosen for their speed, which is often a significant factor in Machine learning. Other languages, like R, is chosen because of its integration into the scientific community long before machine learning became a trend. The last group, consisting of Java and Python has gained popularity because of its already big user base. Python is also the winner when it comes to machine learning because of, like R, its integration into the scientific community. Right now Python is the leading language for machine learning. Driven by this, there is a lot of focus into making it faster, to compete with already fast languages, like the C family.

Python is an interpreted, high-level, general-purpose programming language created in 1991. It is a language that does , and more.

### 3.1.2 tensorflow

Arguably the biggest reason for the success of machine learning in python lies in Tensorflow.Tensorflow is a machine learning package developed by Google in and has since then become the leading framework for machine learning worldwide .

Tensorflow is today a multi-language tool, but it had its origin in python. It is just in later years that other languages has gotten tensorflow support.

### 3.1.3 keras

One of the least attractive things with tensorflow is its unnecessary complexity. Chollet did stuff!

### 3.1.4 Additional packs in keras made by me

## 3.2 Describe code

## 3.3 Describe project

The whole project in

# Chapter 4

# Experiments

## 4.1 Datasets

### 4.1.1 kvasir

Automatic computerised detection of diseases has been an essential focus on the medical industry, but for a long time, a still fairly unexplored research area. As a response to this Simula Research laboratory together with , made the dataset Kvasir. It was originally made to improve medical practice and refine health care systems where similar dataset did not exist.

Kvasir is a dataset containing images from inside the gastrointestinal tract containing three anatomical landmarks, in the form of (A), (B) and (C). Also, the Kvasir dataset includes two categories of images related to endoscopic polyp removal, (D) and (E). And lastly, three classes, (G), (H), and (I), containing images without the anomalies, as a form of baseline. Medical professionals sorted the dataset, and it is made to be used for both single and multi-disease computer-aided detection.

At this time, there are two versions of the kvasir dataset. V1 vs V2 The images are from both from the lower and upper GI tract, Talk about the classes. Talk about the overlays.

### 4.1.2 CVC-356

### 4.1.3 CVC-12k

### 4.1.4 CVC-612

## 4.2 Metrics

When we are measuring preprocessing and classifying, we need a metric to evaluate. In some cases, we want to maximise similarity, and other times minimise error. For the preprocessing, the associated metric can be the mean square error (MSE),

For the transferlerning, the associated metric can either be Validation accuracy or validation loss. Validation accuracy is a number between 0At 100

Validation loss is a measure of how well the training is doing at a given time . The goal for the loss is to reach the value 0, and it can be arbitrarily high.

### 4.2.1 Common metrics

List: Taken from Kvasir site

True positive (TP) The number of correctly identified samples. The number of frames with an endoscopic finding which correctly is identified as a frame with an endoscopic finding.

True negative (TN) The number of correctly identified negative samples, i.e., frames without an endoscopic finding which correctly is identified as a frame without an endoscopic finding.

False positive (FP) The number of wrongly identified samples, i.e., a commonly called a ”false alarm”. Frames without an endoscopic finding which is erroneously identified as a frame with an endoscopic finding.

False negative (FN) The number of wrongly identified negative samples. Frames without an endoscopic finding which erroneously is identified as a frame with an endoscopic finding.

Recall (REC) This metric is also frequently called sensitivity, probability of detection and true positive rate, and it is the ratio of samples that are correctly identified as positive among all existing positive samples.

Precision (PREC) This metric is also frequently called the positive predictive value, and shows the ratio of samples that are correctly identified as positive among the returned samples (the fraction of retrieved samples that are relevant).

Specificity (SPEC) This metric is frequently called the true negative rate, and shows the ratio of negatives that are correctly identified as such (e.g., the fraction of frames without an endoscopic finding are correctly identified as a negative result).

Accuracy (ACC) The percentage of correctly identified true and false samples.

Matthews correlation coefficient (MCC) MCC takes into account true and false positives and negatives, and is a balanced measure even if the classes are of very different sizes.

F1 score (F1) A measure of a test's accuracy by calculating the harmonic mean of the precision and recall.

### 4.2.2 Singleclass vs Multiclass Metrics

The metrics presented are, in general, a solid way to present the validity of a model. However, not all metrics presented is the same when switching between single and multiclass classification. Metrics like Accuracy is designed to work for multiclass classification, given that there is only one way to calculate the score.

$$\frac{sum(diag(covariance_matrix))}{sum(covariance_matrix))} \tag{4.1}$$

The problem with multiclass metrics is more significant in the case for instance Recall and Specificity, where we have multiple ways to calculate the score. We have chosen in this paper to focus on the average of our Multiclass metrics.

In addition to looking at the weighted average of precision and recall, we want to look at specific cases of the classification. In many cases, we have multiple classes, where we are most interested in just one or a handful of the classes shown. For instance, a focus we have in this theis is to give a score on how predictable polyp detection is, and on that case, we want to discuss the True positive rate (TPR) of the polyp detection and not the TPR of the non-polyp classes.

Take for instance the matrix shown in

```
[[10 1]
 [3 12]]
```

Here we can calculate the weighted average recall to be **x**. This can be an interesting observation in itself, but often the first or second True label is much more important relative to the other. In a more practical example: We are more interested in finding areas with polyps when we know they are present, compared knowing there are not a polyp in an area when none are present.

These Metrics becomes a more prominent topic when it comes to inpainting. With inpainting, we take areas with no relevant information and makes it into areas that are similar to the rest of the image. Given that we can inpaint over polyps by mistake, or that we might train our classifiers to not look in certain areas when classifying, we have an interest if also comparing single cases of recall and precision included to the average values.

## 4.3 Setup of experiments

## 4.4 format of experiments

### 4.4.1 Inpaint

### 4.4.2 Classifying

## 4.5 Inpainting Kvasir

### 4.5.1 Black corners

### 4.5.2 Green square

### 4.5.3 Text

### 4.5.4 Combination

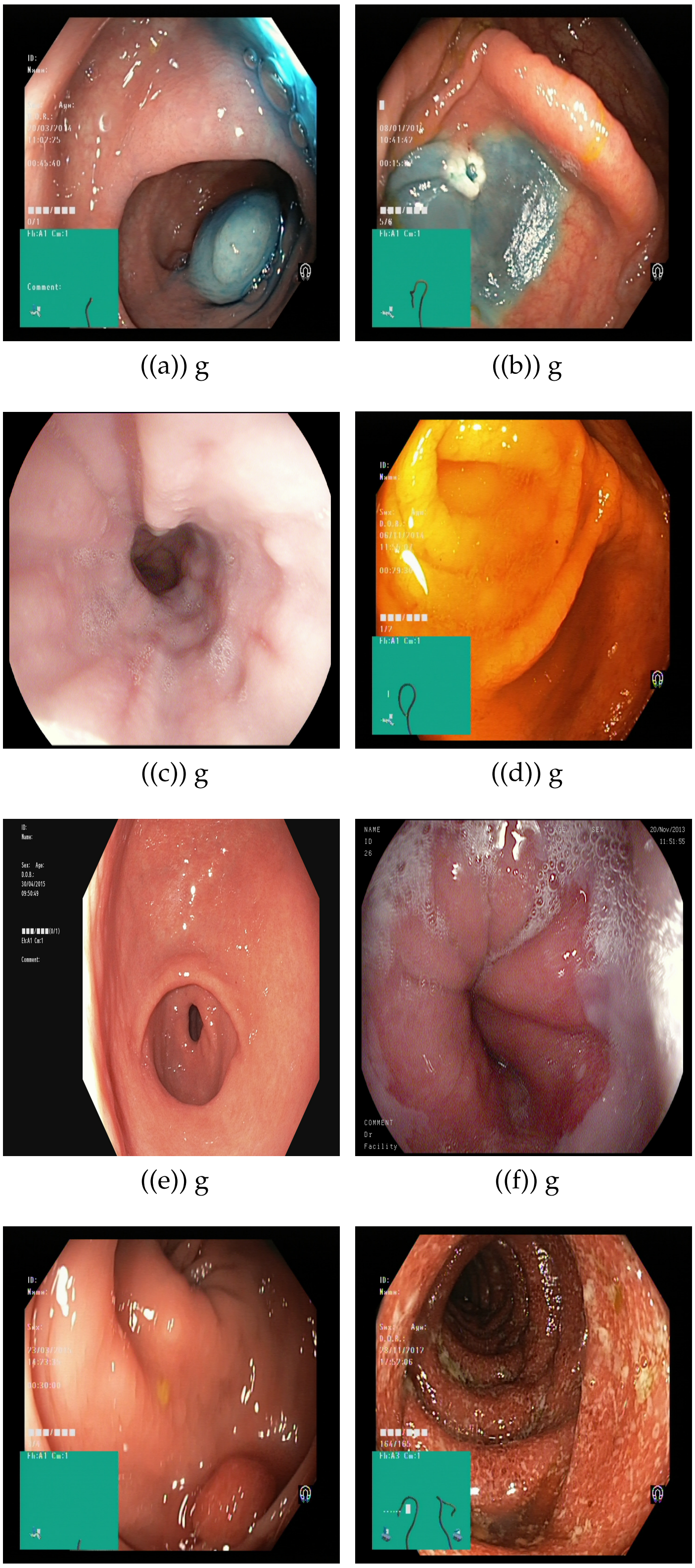
### 4.5.5 Random masking

## 4.6 Kvasir -¿ CVC612

## 4.7 Kvasir -¿ CVC 12k

## 4.8 Kvasir -¿ CVC356

## 4.9 Summary



((a)) g

((b)) g

((c)) g

((d)) g

((e)) g

((f)) g

((g)) g

((h)) g

Figure 4.1: Same image from the z-line with four different inpainting attempts. Each image is re-sized to fit in the figure.

# Chapter 5

# Result and Discussion

# Chapter 6

# Conclusion

# Chapter 7

# Future Work

# Chapter 8

# Appendix