# Introduction

Each year, there are 2.8 million new cases of colorectal, oesophageal, and gastric cancer, with 1.8 million people succumbing to the disease. Endoscopies of the digestive system are the standard of treatment in the medical field. Both gastroscopy and colonoscopy are diagnostic procedures, but gastroscopy focuses on the upper digestive tract (the oesophagus, stomach, and large intestine), whereas colonoscopy focuses on the lower digestive tract (the oesophagus, stomach, and small intestine) (the colon and rectum). Each of these procedures will be supervised using a high-definition, real-time video transmission of the digestive tract. For endoscopic examinations, expensive apparatus and highly trained physicians and nurses are required. Endoscopic removal of precancerous lesions in the colon is necessary to reduce the risk of colorectal cancer. Detection of adenoma is frequently regarded as an indicator of successful colorectal cancer screening. It is highly probable that a physician's ability to diagnose adenomas affects a patient's likelihood of developing colorectal cancer. Based on the findings of an endoscopic examination, multiple physicians may arrive at different diagnoses. The classification of a disease may affect the treatment and monitoring options for a patient. The severity of inflammatory bowel disease (IBD) influences the treatment options available. A dependable, objective, and automated assessment system is required. It has been suggested that automating pathology diagnosis, evaluation, and reporting could help reduce health disparities, enhance service quality, and make better use of the limited resources of the healthcare system. Since endoscopic examinations are performed in real time, it is essential to record all results, positive and negative. If medical report writing could be automated, the quality of care provided could be enhanced. Medical professionals may spend less time on documentation as a result. Treatment and follow-up of patients may benefit from MST's meticulous documentation. To the best of our knowledge, neither an automated nor a standardised method for reporting the results of endoscopic examinations is currently available. Integrating medical imaging with other forms of data analysis and retrieval, as well as artificial intelligence and distributed processing, is a crucial next stage in fundamental computer science and medicine research. Both the United States and the European Union are at the vanguard of the development of medical big data applications, innovation, and healthy competition. As more multimedia research is conducted, it becomes more likely that a relationship between the medical and multimedia disciplines will be discovered. Several studies have examined the viability of using automated algorithms to diagnose polyps and other GI disorders. The paucity of readily available medical data makes it impossible to replicate the findings; the data sets used in the aforementioned study were kept confidential. The Kvasir multi-class picture collection was created by the Vestre Viken Health Trust in Norway to aid in the diagnosis of gastrointestinal disorders. This data set consists of polyps, two novel observations, two classes associated with polyp removal, and three GI anatomical landmarks.

# Dataset

In Norway, the Vestre Viken Health Trust (VV) uses endoscopes. Four facilities within the VV service 470,000 patients. A portion of the training data has already been provided, and more is forthcoming from one of these institutions (the large gastrointestinal department at Baerum Hospital). Moreover, physicians from VV and the Norwegian Cancer Registry provide feedback on each image (CRN). The Cancer Research Network (CRN) provides funding for the disease's investigation. This facility within the South-Eastern Norway Health Authority exists on its own. The Oslo University Hospital Trust manages it, but it is not a medical facility. Nationally, CRN administers cancer monitoring initiatives. By identifying malignancies and precancerous lesions at an early stage, these initiatives aim to reduce cancer-related mortality. The Kvasir dataset contains numerous images depicting various anatomical landmarks, pathology results, and endoscopic GI therapies. Endoscopy specialists have reviewed and commented on these photographs. Various forms of machine learning, deep learning, transfer learning, and image retrieval can utilise numerous images. Included among the anatomical landmarks are the pylorus, cecum, and Z-line. Esophagitis, polyps, and ulcerative colitis are examples of diseases that affect the digestive tract. "the intestinal cecum" In addition, photographic evidence of lesion excision is provided. Images of "dyed and lifted polyps" and "dyed resection margins" are displayed. The images range in size from 720x576 to 1920x1072. The additional training modules use electromagnetic imaging to demonstrate precisely where and how to insert the endoscope into the gastrointestinal tract (Scope Guide, Olympus Europe). This image within a picture may provide context for the imagery. This information was provided because it may be useful for future research. However, it must be approached with care to determine what an endoscopy reveals.

# Problem Statement

Deep learning is currently employed for a vast array of tasks, including forecasting, classification, segmentation, and recommendation. The primary objectives of this notebook are to obtain a high degree of classification precision on the supplied dataset of endoscopic images of the stomach. Deep learning methods contribute to a more precise classification of diseases. The classification of gastrointestinal infections relies heavily on the extraction of numerous characteristics. Deep Learning algorithms reduce the amount of manual labour required for classification purposes, such as feature extraction and data reconstruction. As a result, the implementation of a classification method employing Deep Learning Algorithms, specifically Convolution Neural Network (CNN) and Transfer Learning (TL), will be of great assistance to medical professionals around the world.

# Methodology

CNNs are a type of Deep Learning algorithm that can accept an input image, designate significance (learnable weights and biases), and differentiate between a variety of image attributes and objects. CNNs are sometimes referred to as convolutional neural networks. ConvNet requires significantly less pre-processing effort than other classification methods. ConvNets, with the proper quantity of training, have the potential to learn the model's characteristics and filters, whereas fundamental techniques require the filters to be manually generated.

# "Convolution" is an appropriate word for CNN since it refers to a mathematical function. Convolution is a particular sort of linear operation that multiplies two functions to generate a third function that represents the manner in which the second function modifies the structure of the first function. This Experimental Setup

This distinction was bestowed upon the CNN model as a consequence of its selection to be utilised in the process of image classification. When referring to a convolutional layer with sixteen distinct output channels, the term "Conv. (16)" is used, and it is the most common term. Each convolutional kernel is a matrix with the dimensions of three rows by three columns. The notation "FC (512)" is used to refer to a layer that is entirely connected and has 512 output neurons within a computer programme. There are numerous locations where this sign could be observed. There are numerous other activation function options available, but the Sigmoid, Softmax, and ReLU activation functions are among the most popular and widely used of all the alternatives. Among the numerous alternatives for activation functions, the Sigmoid activation function is by far the most frequently chosen. b, the results of the convolution performed on the first layer of the convolutional network utilising the experimental framework (Conv. (16)). For your convenience, you will also be provided with the residual error in addition to the corresponding PSNR, both of which will be presented in decibels. This is the format that will be used to display each and every result. This information will be sent to you immediately after each conclusion is presented. Using the computer's findings as a reference point for the user. c, the outcomes of the Conv's convolutional analysis; d, the convolutional analysis itself. If you follow the diagonal line that travels through the centre of the distributions of the forecasts, you should be able to locate the majority of accurate projections. Because it passes through the centre of the distributions, this line is diagonal.

In the following paragraphs, an illustration of a model for the categorization of images constructed using deep convolution neural networks will be provided.

## Input

The input consists of N photographs, each of which is marked with one of the K labels applied throughout the classification procedure. Officially, this set is known as the training set, which is also its designation.

The entire dataset, consisting of 8,000 photographs from eight distinct categories, was divided into three categories: training, validation, and testing. Consequently, seventy percent of the data will be utilised for training, twenty percent for validation, and ten percent for testing.

## Evaluation

In addition to predicting the labels that should be given to images that it has never seen before and designating those labels to photographs that it has never seen, the classifier can be used to evaluate the quality of classifications generated by other classifiers. We can determine the accuracy of the classifier by comparing its predictions with the actual labels ascribed to the image. Because the actual classification labels assigned to the image are consistent with the anticipated classification labels generated by the classifier, there is no room for debate regarding the veracity of this assertion. This is a positive development, and it would be to everyone's advantage if this kind of event occurred more frequently in the world.