



Institute of Electrical and Electronics Engineering

3rd year project presentation

Title: Gastrointestinal diseases diagnosis using gastroscopy, KNN and CNN

Student name: BENADJAL Rania

email: benadjelrania@gmail.com

Supervisor: Prof CHERIFI Dalila

Abstract:

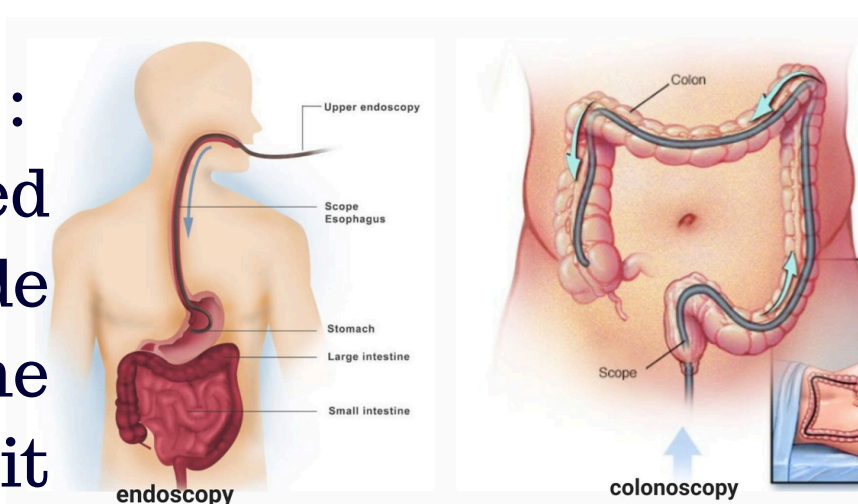
This work presents an implementation of two machine learning models, KNN and CNN, in order to detect and classify gastrointestinal diseases using data of images taken by a gastroscopy method.

1. Introduction:

An integral part of the complex digestive system is the gastrointestinal tract (GIT). It is a network of interconnected organs that facilitates the breakdown of food into essential nutrients to be converted into useful energy and for the growth and repair.

Endoscopy and colonoscopy :

Two imaging techniques used by inserting a camera inside the GIT to look inside the upper, and lower parts of it [2].



2. Methodology:

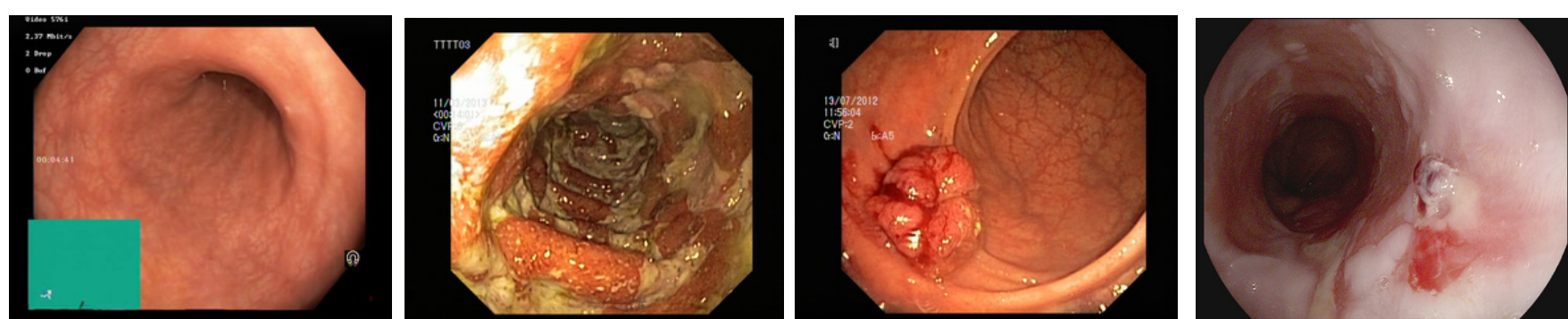
We explored the classification of gastrointestinal diseases using the Kvasir dataset.

Two distinct models, k-Nearest Neighbors (KNN) and Convolutional Neural Networks (CNNs), were rigorously evaluated and optimized to tackle this classification task.

3. Experiments and results:

3.1. Dataset :

computer-aided detection (CAD) for gastrointestinal tract disorders Kvasir is among the earliest, most notable datasets in the GI field. It was collected by the Vestre Viken Health Trust (VV) in Norway. It consists of four different classes.



Normal Ulcerative colitis Polyps Esophagitis

data [1] was firstly pre-processed by resizing it to 128*128 px, and by normalizing the color of each pixel by dividing over 255, with keeping the colors to benefit from the info they give.

3.2 Evaluation metrics:

$$\text{Precision} = \frac{TP}{TP + FP} \quad \text{F1 Score} = \frac{TP}{TP + \frac{1}{2}(FP + FN)}$$

$$\text{Recall} = \frac{TP}{TP + FN} \quad \text{ACC} = \frac{TP + TN}{TP + TN + FP + FN}$$

References:

- [1] KVASIR Dataset, Kvasir. URL: <https://dl.acm.org/doi/abs/10.1145/3083187.3083212>.
 [2] Gastrointestinal polyp detection in endoscopic images using an improved feature extraction method. In: 7 September 2017. URL: <https://link.springer.com/article/10.1007/s13534-017-0048-x>.

3.3 Experiment 1: KNN

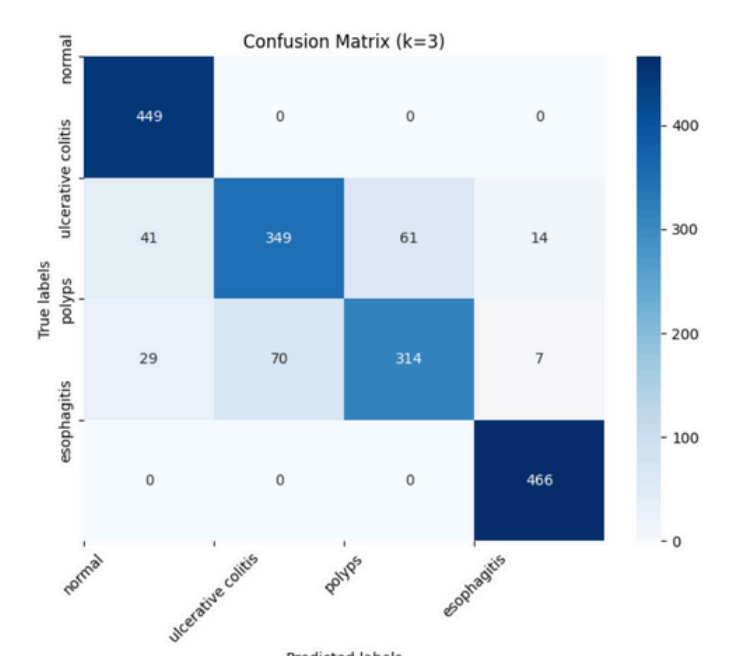
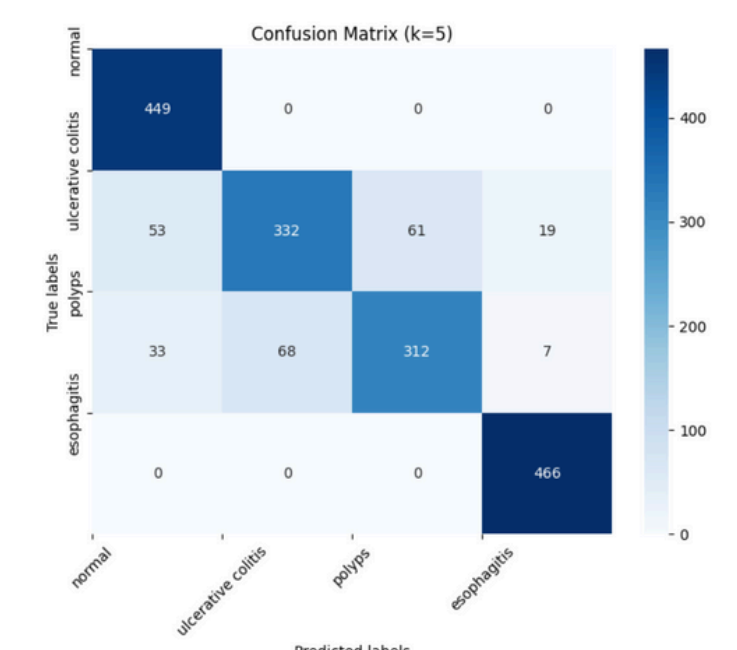
We applied to the dataset the KNN classifier with various values of k (3, 5, and 7).

Result:

Classification Report:				
	precision	recall	f1-score	support
0	0.82	1.00	0.90	449
1	0.83	0.70	0.76	465
2	0.83	0.73	0.78	420
3	0.95	1.00	0.97	466
accuracy			0.86	1800
macro avg	0.86	0.86	0.85	1800
weighted avg	0.86	0.86	0.85	1800

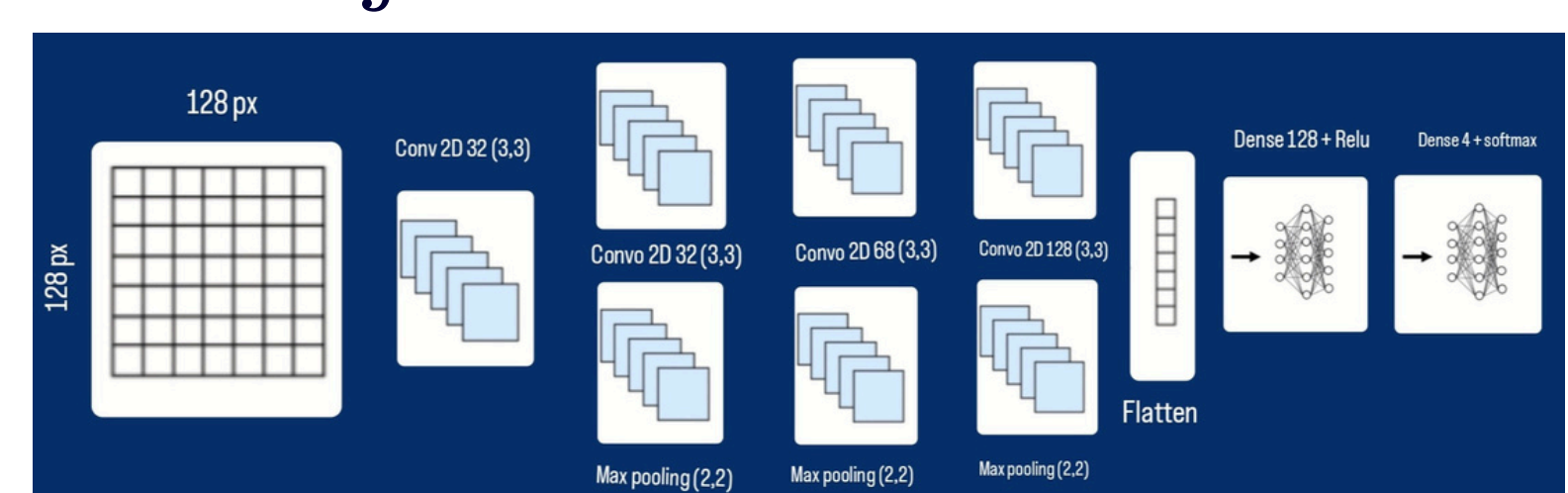
Accuracy: 0.86

- overall accuracy of about 86%
- There is a chance the KNN classifier struggles to distinguish between "ulcerative colitis" and "polyps" and the "ulcerative colitis" and "esophagitis."



3.4 Experiment 2: CNN

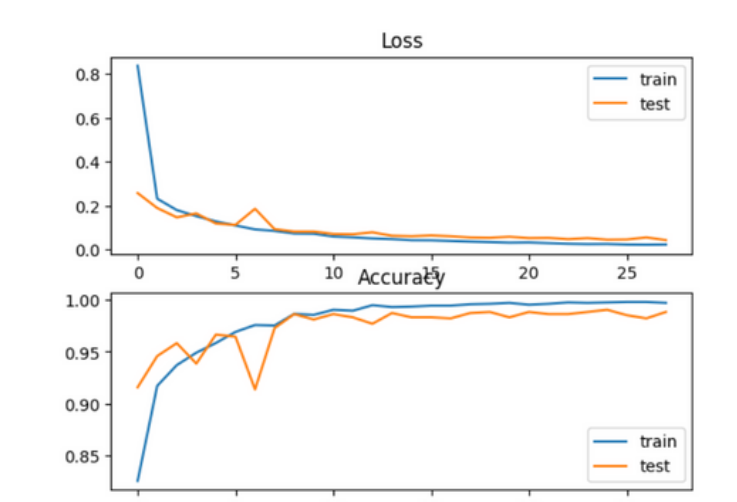
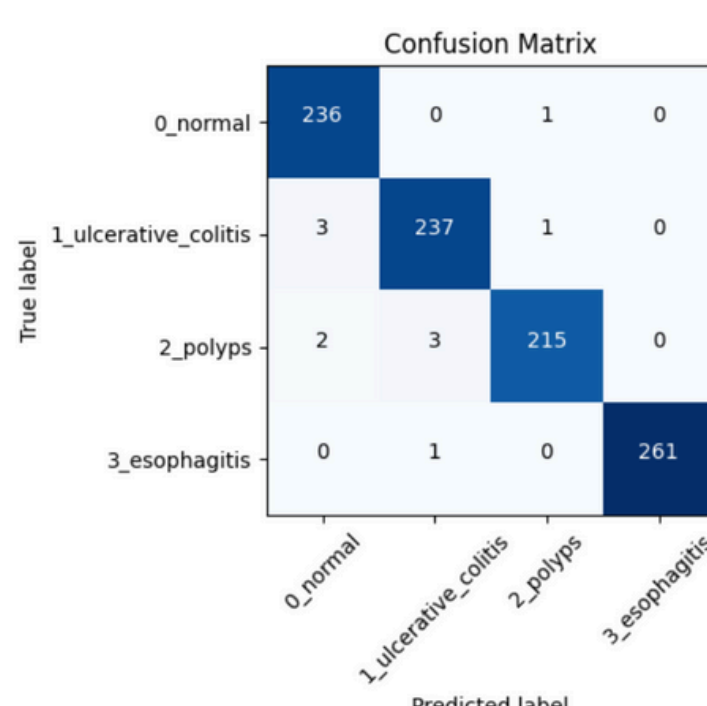
We applied to the dataset the CNN classifier with the following architecture:



Result:

Classification Report:				
	precision	recall	f1-score	support
0	0.99	1.00	0.99	237
1	0.98	0.98	0.98	241
2	0.99	0.98	0.98	220
3	1.00	1.00	1.00	262
accuracy			0.99	960
macro avg	0.99	0.99	0.99	960
weighted avg	0.99	0.99	0.99	960

Accuracy: 0.9895833333333333



- In overall, despite the smaller dataset, the CNN model achieves 98.95% accuracy.
- better classification of the "ulcerative colitis" and "polyps" and the "ulcerative colitis" and "esophagitis." compared to the KNN model.

4. Conclusion:

The CNN model significantly outperforms the KNN model in terms of accuracy and correlation. But in terms of time of learning the KNN model was remarkably faster than the CNN. For further work we can experiment the same data with additional machine learning models.