Endoscopy Image Processing & Classification - Final Report

Sujal
AI20BTECH11020
ai20btech11020@iith.ac.in

Abstract

This project report discusses the use of endoscopy image processing for diagnosing gastrointestinal diseases, focusing on the classification of images taken with the Wireless Capsule Endoscopy (WCE) method. The dataset used in the project has four classes: polyps, ulcerative colitis, esophagitis, and normal. The report presents a dataset of 6000 images which equals the number in classes, and data is around 1.2GB. The pre-processing of the dataset was done using Keras' "ImageDataGenerator." The project compared the accuracy of two pre-trained models, InceptionNetv3 and EfficientNetB2, on the classification task. EfficientNetB2, with an accuracy of 94%, perform better than InceptionNetV3, with an accuracy of 79%. Also, EfficientNetB2 outperformed InceptionNetv3 in terms of requiring less training time and having a smaller model size. The study provides valuable insights into the application of endoscopy image processing for diagnosing gastrointestinal diseases

1. Introduction

This innovative WCE method has revolutionized the examination process, providing better findings without causing any discomfort to the patient. The capsule-like form factor of the WCE method has proved to be advantageous, allowing medical experts to identify abnormalities better and faster. The tiny lighted camera inside the capsule can capture visual images of the GI tract, providing a clearer and more comprehensive view for medical professionals. This way, they can quickly detect any potential signs of infection, inflammation, or cancer, ensuring timely diagnosis and treatment. With the help of WCE, patients no longer have to undergo lengthy and uncomfortable traditional endoscopy procedures, making the entire process more accessible and convenient for everyone. The WCE method is an excellent example of how innovative technology can bring about significant improvements in the field of medicine, benefiting patients and medical professionals alike.

Gastrointestinal (GI) diseases significantly threaten human health, affecting millions of people every year. The conventional endoscopy procedure has been the go-to method for diagnosing GI diseases, but it often causes discomfort and pain to patients. However, recent advancements in medical technology have introduced the Wireless Capsule Endoscopy (WCE) method, which has revolutionized the way medical experts examine the GI tract. The WCE method involves a tiny lighted camera inside a capsule that is swallowed by the patient and can travel through the GI tract, capturing visual images along the way. This new method has proven to be less invasive and more convenient for patients, improving the overall examination process.

One exciting development in the field of endoscopy is the use of image processing techniques to diagnose gastrointestinal diseases. In this project report, we focus on the classification of images taken using the WCE method, which has four classes: polyps, ulcerative colitis, esophagitis, and normal.

2. Literature Review

Following are some papers that helped to build a model on endoscopy images using deep learning AI models.

2.1. Diagnosing gastrointestinal diseases from endoscopy images through a multi-fused CNN

The diagnosis of gastrointestinal (GI) diseases through endoscopy images is a challenging task due to the complexity and variability of the images. In recent years, deep learning techniques have shown great potential in improving the accuracy of this task. In this regard, Montalbo proposed a novel Multi-Fused Residual Convolutional Neural Network (MFuRe-CNN) with Auxiliary Fusing Layers (AuxFL), a Fusion Residual Block (FuRB) both with Alpha Dropouts (αDO) for diagnosing various endoscopic images of GI conditions or diseases. The proposed model handles four cases, including colons with ulcerative colitis, polyps, esophagitis, and a healthy colon, sourced from reliable

databases like the KVASIR and ETIS-Larib Polyp DB. The proposed MFuRe-CNN consists of three state-of-the-art models fused into a single feature extraction pipeline with their partially frozen and truncated layers. This helps propagate robust features and improve diagnostic performance without consuming a hefty fraction of computing cost compared to most existing state-of-the-art models. Additionally, the MFuRe-CNN incorporated with AuxFLs, αDOs, and FuRB have shown a significant contribution in reducing overfitting and performance saturation compared to those without the said components.

Upon evaluation, the proposed model achieved an outstanding 97.5% test accuracy with only 4.8 M parameters and consumed 7.8 GFLOPs during inference, making it more efficient and accurate than most conventionally trained DCNNs. These results demonstrate that the proposed MFuRe-CNN has the potential to improve the diagnosis of the GI tract more cost-efficiently than ensembles and perform better diagnosis than most conventional pre-trained and fine-tuned DCNNs. Overall, this study shows that the proposed MFuRe-CNN is a promising approach for the diagnosis of gastrointestinal diseases from endoscopy images.

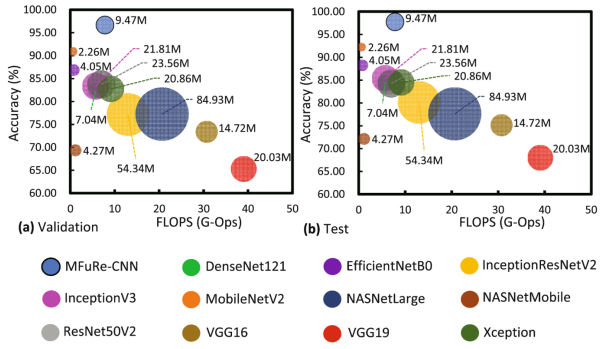

Figure 1. Cost-efficiency to accuracy ratios comparison.

2.2. Endoscopic Image Classification Based on Explainable Deep Learning

It was found that explainable deep learning-based classification is essential in the field of wireless endoscopic imaging. This is because it allows medical experts to understand and validate the decisions made by the model, which can have a significant impact on a patient's health. The study proposed an explainable artificial intelligence method for wireless endoscopic image classification using ResNet-152 combined with Grad-CAM, which resulted in improved model performance with 98.28% training and 93.46% validation accuracy. The use of explainable deep learning in this context can also help in identifying and diagnosing abnormalities and conditions within the body during the endoscopic procedure, improving the diagnostic accuracy of endoscopic examinations, and leading to earlier detection and treatment of diseases. The study also highlighted the limitations, including a lack of updated open-source datasets in wireless endoscopic images and the future directions to improve the model performance. Overall, this literature review provides insights into the importance of explainable deep learning-based classification in the field of wireless endoscopic imaging and the potential benefits it can provide in improving patient care.

3. Dateset and Features

This section describes the dataset used for training, some of its properties and the pre-processing that had to be performed on it.

3.1. Dataset characteristics

For the dataset using the Gastrointestinal Tract or Colon Diseases Image Dataset (KVASIR Dataset) dataset in this project. The KVASIR dataset is a collection of medical images taken from inside the gastrointestinal tract, which are classified into three important anatomical landmarks and three clinically significant findings. The dataset also includes two categories of images related to endoscopic polyp removal. The images are sorted and annotated by medical doctors who are experienced endoscopists. The purpose of this dataset is to provide a resource for research on the computer-aided detection of diseases in the gastrointestinal tract, with the hope that it will encourage researchers from multimedia fields to contribute to the medical domain of detection and retrieval. In the dataset, images are taken using

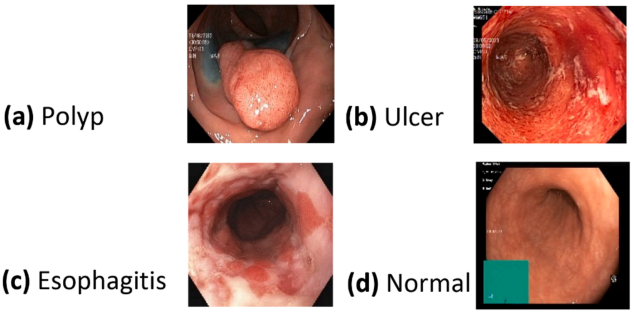

Figure 2. Gastrointestinal conditions: polyp (a), ulcerative colitis (b), esophagitis (c), healthy gastrointestinal tract (d)

the Wireless Capsule Endoscopy (WCE) method. So, there are 6000 images which equals the number in classes and data is around 1.2 GB. So, the dataset is balanced where the training data set has 3200 images and test data set has 800 images, and the validation data has 2000 images. There are four classes in the dataset:

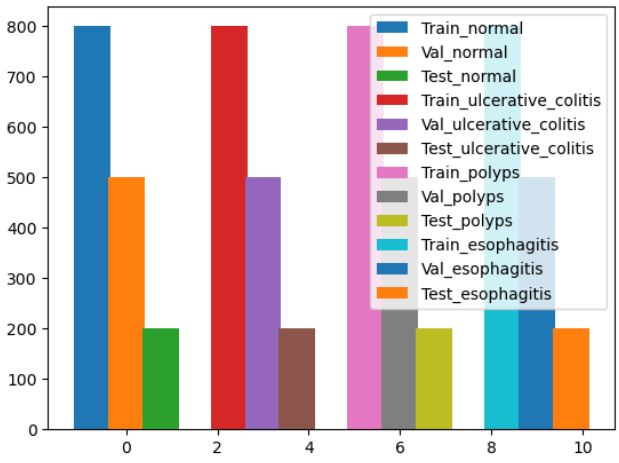

Figure 3. Dataset specifications

(a) **Polyps** A single or cluster of bumpy lesions with distinct surface colour and pattern found in the colon lining. In most cases, polyps do not cause harm or pain.

(b) **Ulcerative colitis** A GI disease that affects the large bowels and is accompanied by mild to severe discomfort. Ulcerative colitis or ulcer usually leaves traces of a white fibrin coating in the GI tract covering the wounds from inflammations.

(c) **Esophagitis** Patients with this finding suffer from inflammations and bleeding of the oesophagus, a passage tube that transports food from the mouth to the stomach.

(d) **Healthy or Normal** Patients a healthy or normal intestinal tract, free from any colour or texture abnormalities compared to the previously mentioned diseases.

3.2. Data Pre-processing and Augmentation

The dataset have 3 directories where each class has 800 images and is split into test, validation and train. Using Keras' "ImageDataGenerator" to apply data augmentation techniques such as rotation, flip, zoom, shear, height shift, and width shift to the images to increase the amount of data available for training and improve the robustness of the model. The images are resized to 224 x 224 x 3 pixels and divided into batches of 16. Three separate generators are created for the training, validation, and testing datasets, each with its own set of image directories and class modes. The output of the code indicates that there are 3200 images in the training set, 2000 images in the validation set, and 800 images in the testing set, each belonging to one of four classes. Table 1 provides the specific values used for each of the data augmentation settings.

Augmentation	Value
Horizontal flip	True
Rotation Range	15
Shear Range	0.2
Zoom Range	0.2
Height shift	0.1
Width shift	0.1

Table 1. Data augmentation settings

4. Architectures

This section describes the methods we used for this classification problem. For both of these methods, we used the images that we pre-processed. We divided the images into train(53.3%), validate(33.3%) and test(13.3%) images.

4.1. Model Implementation

The model is using transfer learning by using pre-trained models on ImageNet. We use this architecture to extract features from an image. InceptionNetv3 and EfficientNetB2 are both deep learning models used for image classification tasks. Moreover, both models use the Adam optimizer because the Adam optimizer adapts the learning rate for each parameter during training based on the first and second moments of the gradients. This means that it adjusts the learning rate for each parameter in a way that depends on the estimated variance of the gradients of that parameter, allowing the optimizer to converge faster and more efficiently. And, record all metrics like accuracy, Precision, Recall and AUC at each epoch.

4.2. InceptionNetv3 Model

First, implementing InceptionNetV3 model as base model. InceptionV3 is an image recognition model shown to attain greater than 78.1 % accuracy on the ImageNet dataset. InceptionNet V3 is a deep convolutional neural network architecture proven effective for feature extraction from images. The network is designed to extract features at multiple scales by using Inception modules that allow for efficient use of computational resources. The Inception modules use multiple convolutional filters with different kernel sizes, which allows the network to extract features at different levels of abstraction.

In addition, InceptionNet V3 has been trained on a large dataset (ImageNet) with a large number of classes, which has allowed the network to learn a wide range of features that are useful for many different image classification tasks.

The architecture appears to use a pre-trained InceptionNet V3 as a feature extractor, which is then followed by a fully connected neural network for classification. The last layer of the InceptionNet V3 is removed and replaced with a flattened layer to convert the 2D features to a 1D vec-

tor. Subsequently, two dense layers are added, along with batch normalization, Gaussian noise, and dropout to reduce overfitting. Lastly, a dense layer with softmax activation is added to output the classification probabilities for the four classes. Additionally, a Gaussian Noise layer is included in the model to minimize overfitting. Architecture can see in Fig 4.

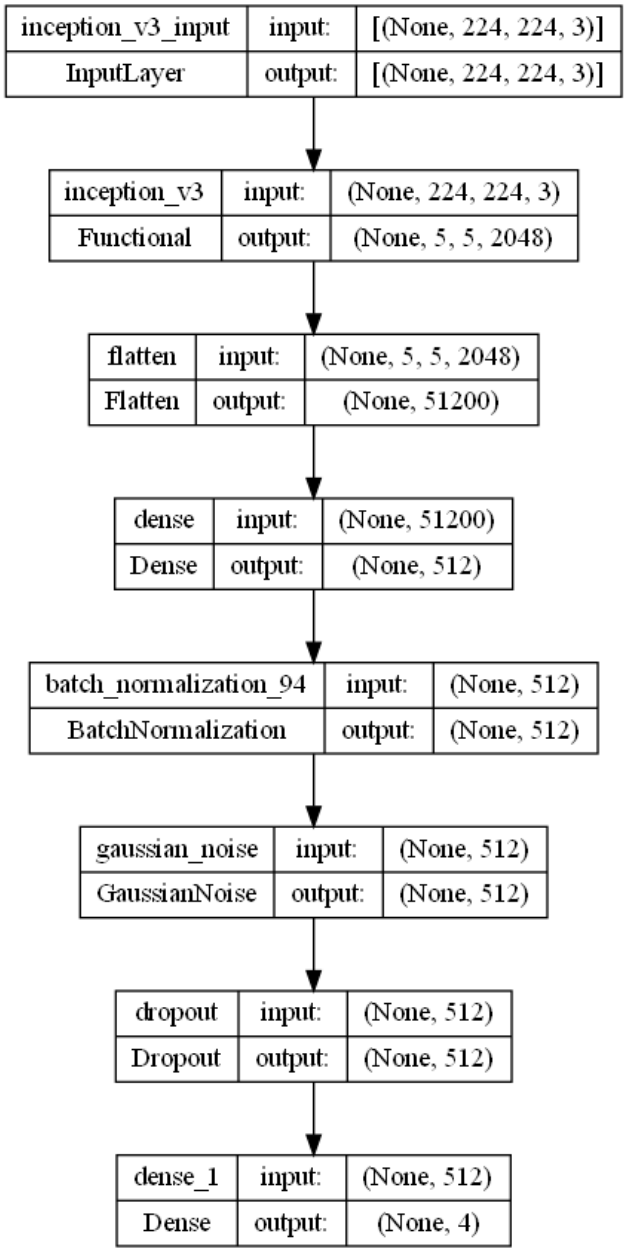

Figure 4. Illustration of InceptionV3Net Architecture

	precision	recall	f1-score
normal cells	0.70	0.97	0.81
ulcerative colitis cells	0.79	0.62	0.70
polyps cells	0.75	0.66	0.70
esophagitis cells	0.96	0.91	0.94
Average	0.80	0.79	0.79

Table 2. Classification Report of inceptionV3 model on Test data

4.3. EfficientNetB2 Model

EfficientNetB2 is a neural network architecture that was designed to have a good balance between model size and accuracy. It is based on a combination of neural network scaling techniques that optimize the performance of the network. The EfficientNetB2 architecture is especially well-suited for feature extraction because it is able to capture complex features in images while maintaining a relatively small number of parameters.

In architecture using the EfficientNetB2 model as the base model for feature extraction. The last layer of the EfficientNetB2 model has been removed and replaced with a global average pooling layer, which reduces the dimensionality of the output. Then, a fully connected dense layer has been added to the model for classification, followed by batch normalization, Gaussian noise, dropout and a final dense layer with 4 output units representing the classes.

Overall, this architecture is a sequential model that consists of the EfficientNetB2 model as the base followed by a few dense layers for classification.

5. Result

We tried the classification of augmented images using the above mentioned models, and we compared the accuracies of these models.

- InceptionV3 model.
  The model had $26,217,988$ trainable parameters and $21,803,808$ non-trainable parameters. Executing the code with 40 epochs, Train Accuracy in first epoch obtained is around 69.50% and that with validation accuracy is around 52.6%. After execution of 40 epochs the final train and validation accuracies obtained were 86.87% and 75.45% respectively with training loss changing from 0.9227 to 0.3518 and validation loss from 1.5151 to 0.6466 as shown in Fig 6.And, also there are plotting recall in Fig 7, precision in Fig 8 and AUC in 9. We can see that model is not get fully trained on data due to resources constraints. We get these results on test data as seen in the confusion matrix in Fig 10 and the classification report in table 3.

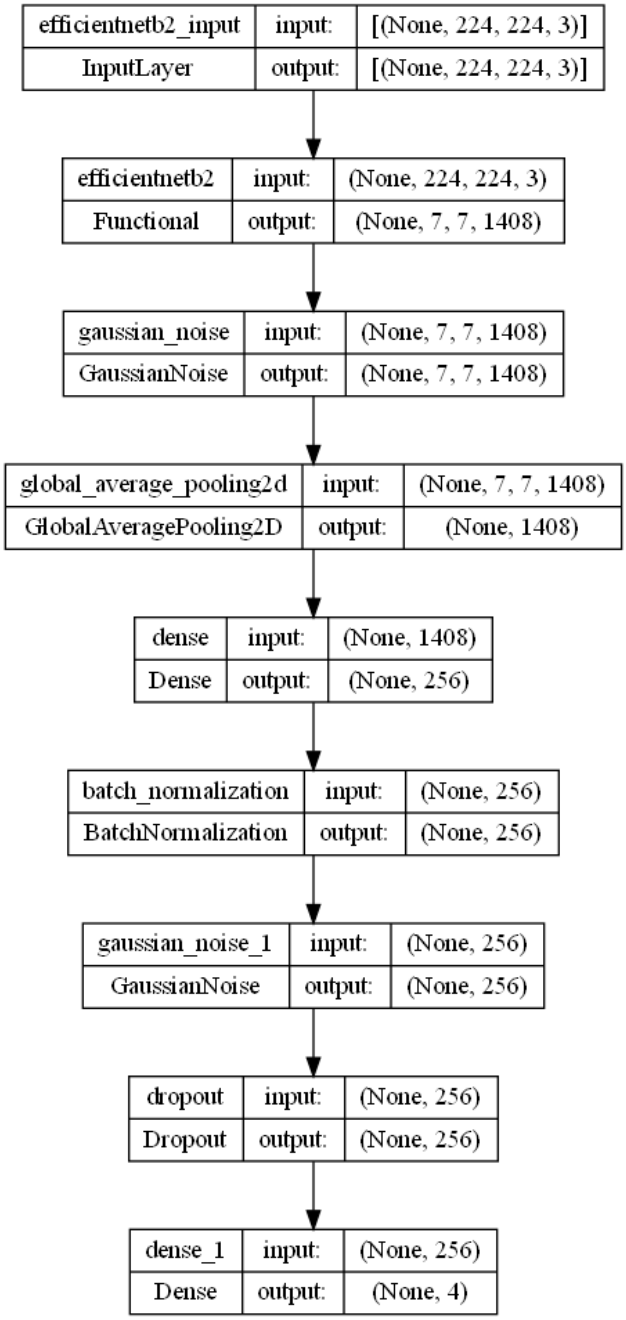

Figure 5. Illustration of EfficientNetB2 Architecture

- Results from EfficientNetB2 model
  It has a total of 8,131,325 parameters, out of which 362,244 are trainable parameters. As executed model for 20 epoch so that it not get over-fit. We can see the accuracy is somewhat stable by the end of 20 epochs. Train Accuracy in first epoch obtained is around 75.63% and that with validation accuracy is around 78.15%. After execution of 20 epochs the final train and validation accuracies obtained were 96.56% and 91.75% respectively with training loss changing from 0.6275 to 0.0951 and validation loss from 0.5601 to 0.2444 as shown in Fig 11.And, also plotting recall in Fig 12, precision in Fig 13 and AUC in 14. We can see that model is get fully trained on data and stop it to 20 to not get over fit. Training it partially reduces the training time on the stake of accuracy. We get these results on test data as see in confusion matrix in Fig 15 and also classification report in table **??**.

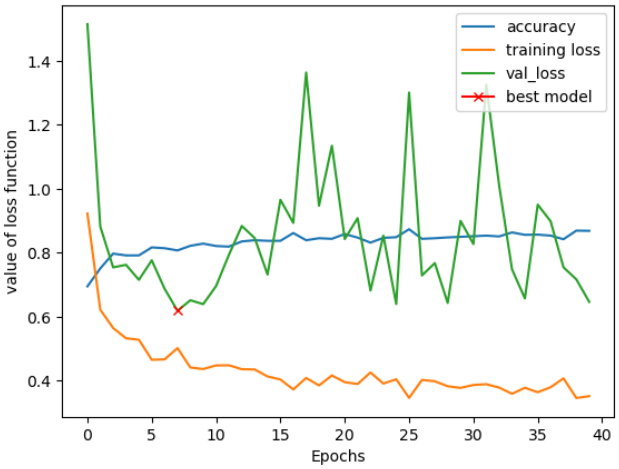

Figure 6. Accuracy & Loss vs Epochs of InceptionV3 model

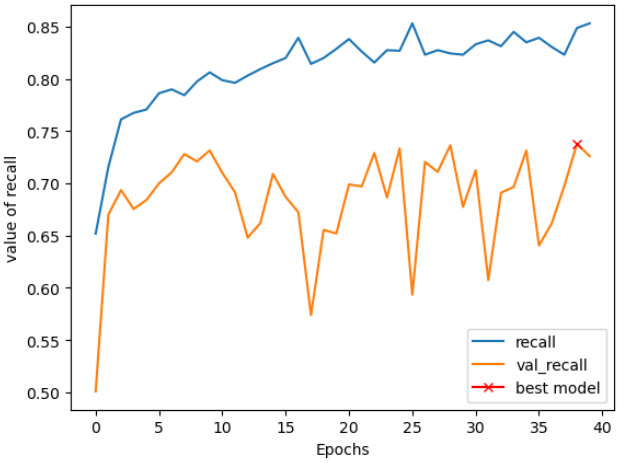

Figure 7. Recall vs Epochs of InceptionV3 model

Here, in both model results macro average and weighted average is same because of taking balanced data. At the end of 40 epochs InceptionNetv3 model produced less accuracy on both training dataset and test dataset compared to that of EfficientNetB2 model.

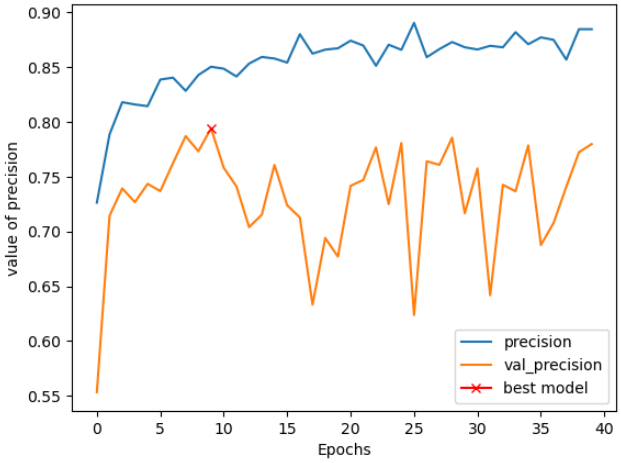

Figure 8. Precision vs Epochs of InceptionV3 model

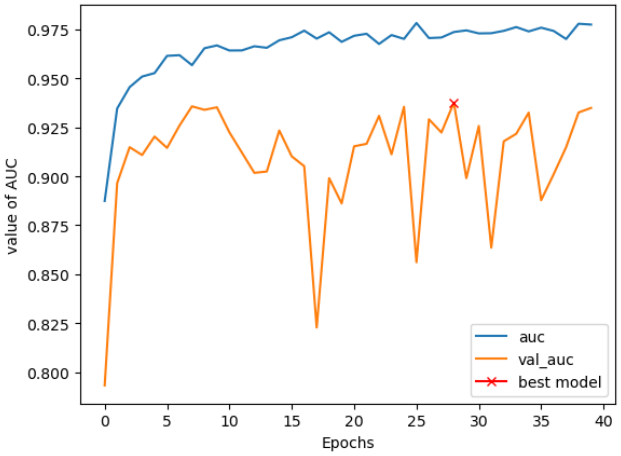

Figure 9. AUC vs Epochs of InceptionV3 model

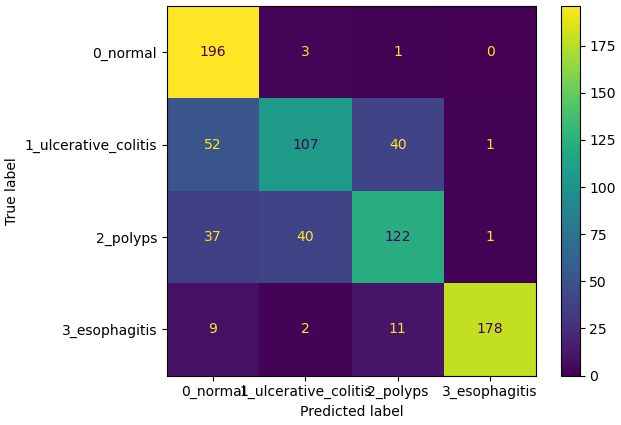

Figure 10. Confusion Matrix of InceptionV3 model on Test data

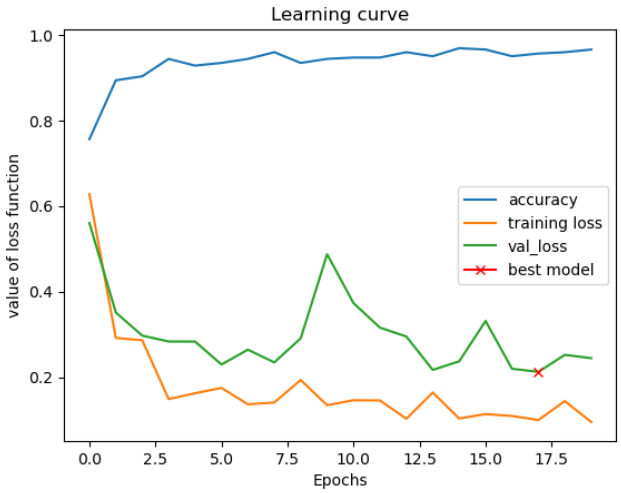

Figure 11. Accuracy & Loss vs Epochs of EfficientNetB2 model

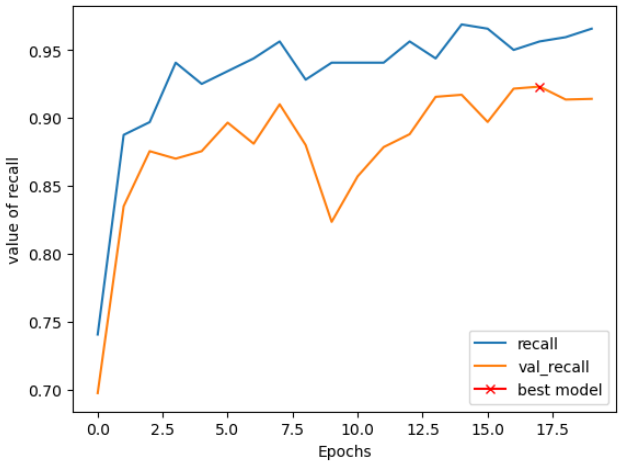

Figure 12. Recall vs Epochs of EfficientNetB2 model

	precision	recall	f1-score
normal cells	0.94	1.00	0.97
ulcerative colitis cells	0.85	0.96	0.90
polyps cells	0.98	0.81	0.88
esophagitis cells	0.99	0.98	0.99
Average	0.94	0.94	0.94

Table 3. Classification Report of EfficientNetB2 model on Test data

6. Conclusion

In this report, we presented our approach to a medical image classification problem using the KVASIR dataset. We described the dataset characteristics, which includes 6000 images of gastrointestinal tract diseases, sorted and annotated by experienced medical doctors. We split the

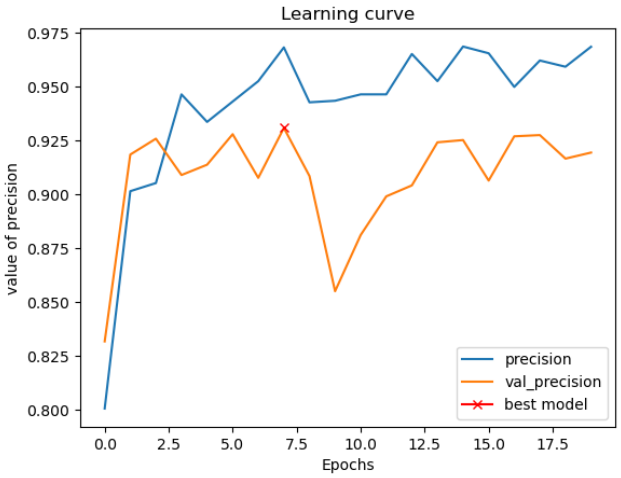

Figure 13. Precision vs Epochs of EfficientNetB2 model

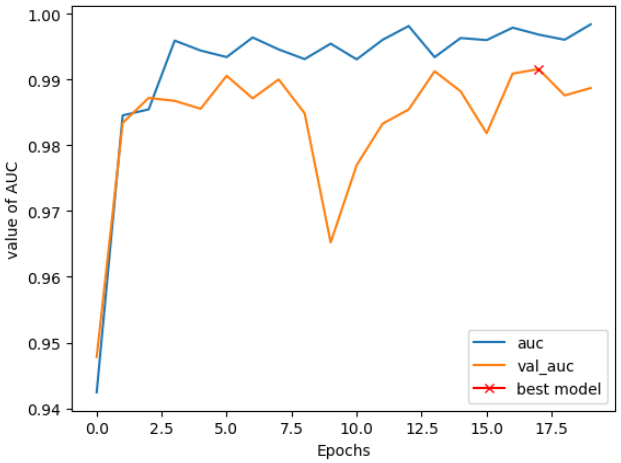

Figure 14. AUC vs Epochs of EfficientNetB2 model

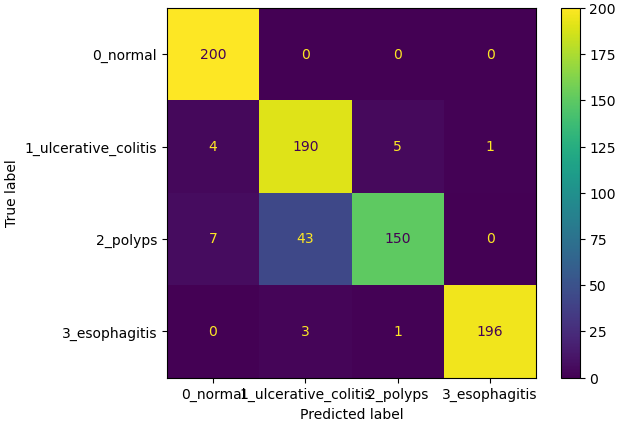

Figure 15. Confusion Matrix of EfficientNetB2 model on Test data

data into train, validation, and test sets and used data augmentation techniques to increase the amount of data available for training and improve model robustness. We used transfer learning with pre-trained models InceptionNetv3 and EfficientNetB2 to extract features from the images, and Adam optimizer to optimize the models. We recorded metrics such as accuracy, precision, recall, and AUC at each epoch. Overall, our approach showed promising results for the classification of gastrointestinal tract diseases, which could contribute to the field of computer-aided detection of diseases in the gastrointestinal tract.

We can see that EfficientNetB2, with an accuracy of 94%, perform better than InceptionNetV3, with an accuracy of 79%. Architecture-wise, InceptionNetv3 uses a combination of convolutional layers, max-pooling layers, and fully connected layers. In contrast, EfficientNetB2 uses a compound scaling method to balance the network's depth, width, and resolution. Here EfficientNetB2 outperforms InceptionNetv3 in terms of accuracy while using fewer parameters Because EfficientNetB2 shows to achieve state-of-the-art results on various image classification benchmarks, including ImageNet. Furthermore, EfficientNetB2 requires less training time than InceptionNetv3 due to its efficient architecture and compound scaling method. In model size, EfficientNetB2 has a smaller model size compared to InceptionNetv3, making it easier to deploy on resource-constrained devices.

7. References

The links to the papers, datasets and libraries we used can be found below:

- KVASIR Dataset
- Diagnosing gastrointestinal diseases from endoscopy images through a multi-fused CNN
- Endoscopic Image Classification Based on Explainable Deep Learning
- Transfer Learning using Tensorflow
- InceptionV3
- EfficientNetV2