A PROJECT REPORT ON

**Medical Image Analysis Using Ensemble Learning for gastrointestinal (GI) endoscopy**

PROJECT REPORT SUBMITTED TO ICFAI TECH AS

A PARTIAL FULFILMENT OF THE REQUIREMENTS FOR THE AWARD OF THE DEGREE OF B.TECH IN DSAI

UNDER THE SUPERVISION

Dr. Priyanka Parimi BY

M.A.VASIQ

**21STUCHH010396**



Department of Data Science and Artificial Intelligence

Faculty of Science and Technology

ICFAI TECH (Deemed to be University)

HYDERABAD

**CERTIFICATE**

This is to certify that the project report entitled Medical Image Analysis Using Ensemble Learning for gastrointestinal (GI) endoscopy submitted in fulfillment of the degree of B. Tech in (Data Science & Artificial Intelligence) is a record of original work carried out by me under the supervision of Dr. Priyanka Parimi, and has not formed the basis for the award of any other degree or diploma, in this or any other Institution or University. In keeping with the ethical practice in reporting scientific information, due acknowledgements have been made whenever the findings of others have been cited.

Supervisor Head Of Department

Dept. of DSAI Dept. of DSAI

**DECLARATION**

We declare that the work contained in the Project Report is original and it has been done by us under the supervision of Dr. Priyanka Parimi. The work has not been submitted to any other University for the award of any degree or diploma.

M.A.Vasiq

# ACKNOWLEDGEMENT

I would like to take this moment to express my sincere appreciation to everyone who has contributed to the successful completion of this project. Firstly, I want to extend my heartfelt thanks to Dr. Priyanka parimi, Assistant Professor at IcfaiTech, for her invaluable guidance and support throughout the project. Furthermore, I want to express gratitude to our friends and peers who have offered feedback, suggestions, and assistance at various stages of the project. Their input has greatly enriched our work and helped us overcome challenges along the journey.In conclusion, I want to acknowledge and recognise the collective effort and collaboration of all individuals and entities involved in this project. It has been an enriching experience, and we are eager to apply the knowledge and skills acquired in future endeavours.

Thank you.

M.A.Vasiq

# ABSTRACT

This study investigates the effectiveness of ensemble learning in medical image classification by using Kvasir dataset which is made up of gastrointestinal (GI) endoscopic images. Four deep learning models were constructed based on popular architectures such as InceptionV3, VGG16, ResNet50 and MobileNetV2 are used. Each model was taught to classify images into eight different classes related to gastrointestinal diseases. Data augmentation techniques were used to expand the size of the data set and boost model generalization. We then developed a single predictive model based on weighted average model from all four individual ones after training them separately. Our ensemble model performed better than each individual ones and also classifies diseases correctly ,proving that ensemble learning is effective in medical image classification. This study highlights the potential of ensemble learning in medical image analysis and its broader applicability across various domains.

**Keywords:** Image classification, Ensemble learning, deep learning, data augmentation, generalization, weighted average model,

**TABLE OF CONTENTS**

[1. INTRODUCTION 7](#_Toc1176160036)

[2. LITERATURE REVIEW 9](#_Toc134754417)

[3. METHODOLOGY 11](#_Toc1369722581)

[3.1 Data Augmentation 11](#_Toc2050206674)

[3.1.1 Data Augmentation Techniques 12](#_Toc1448453803)

[3.2 pre-defined Models Used 13](#_Toc288444629)

[3.2.1 InceptionV3 13](#_Toc1018288324)

[3.2.2 ResNet50 14](#_Toc1074678525)

[3.2.3 MobileNetV2 15](#_Toc1784498405)

[3.2.4 VGG16 16](#_Toc308078833)

[3.3 Ensemble learning 17](#_Toc297873314)

[3.4 Weighted Averaging Ensemble 18](#_Toc572407241)

[4. RESULTS AND PERFORMANCE 20](#_Toc1869244574)

[4.1 Experimental Setup 20](#_Toc855283255)

[4.2 Individual Model Performance 20](#_Toc1637617414)

[4.3 Ensemble Model Performance 25](#_Toc1277410485)

[5. CONCLUSION 27](#_Toc1044002336)

[REFERENCES 28](#_Toc1605865162)

# INTRODUCTION

Individual deep learning models are the foundation of several image classification systems currently in use. Although these models have shown promise in general picture classification tasks, the problems like lighting Variations, Image Quality Issues, Disease Presentation Variability, could limit their effectiveness when applied to the particular area of GI endoscopy [4]. When a single deep learning model learns specific characteristics of the training data too well, it may overfit and become unable to generalize to new variations. Furthermore, they could concentrate on particular picture features that, in reality, aren't necessarily reliable indications of disease. These restrictions may cause mistakes, which could affect patient treatment and delay identification

Ensemble learning is a promising way to get over the drawbacks of individual deep learning models. It involves combining predictions from several models to create an order of classification that is more reliable and precise [4]. The images are used to train multiple deep learning models. These models may be built using a variety of architectures, such as MobileNetV2, ResNet50, or InceptionV3, each with advantages and disadvantages of its own. After training, every model predicts which class—normal tissue, polyp, ulcer—a particular image will belong to. These distinct predictions are then combined by the ensemble approach to provide a final classification.

Using ensemble learning for GI endoscopic image classification offers several advantages over depending on single models such as improved accuracy, enhanced consistency, reduced overfitting and robustness.

In this work, we train pre defined models like VGG16, MobileNetV2, ResNet50 and InceptionV3 on the gastrointestinal images. Before training them directly we fine tune the models based on our requirements.

It starts with an introduction that outlines the significance of medical image classification, and provides an overview of researches which have been done previously and importance of current study. After this, the literature review section provides a comprehensive overview of existing literature in the field. The methodology section explains the preprocessing techniques and details the models used in our project work. Additionally, it discusses ensemble learning and its working mechanism, with a focus on weighted average ensemble learning method. The results section presents the experimental setup, including the dataset, system specifications, and models used, followed by an evaluation of the performance of the implemented models. Finally, the conclusion section summarizes the findings and discusses what further works can be done from this project.

# LITERATURE REVIEW

Medical image analysis, particularly in gastrointestinal (GI) endoscopic images, has been a sub- ject of important research due to its major role in early disease detection and diagnosis of that disease. Endoscopy plays a crucial role in diagnosing various GI disorders such as ulcers, polyps, and cancers. With advancements in imaging technology, there has been a increase in the volume and complexity of endoscopic images. Traditional diagnostic methods heavily depend on manual examination by medical professionals, which is not only time-consuming but also prone to human error. Therefore, automated medical image analysis techniques have become essential to assist clinicians in accurate and efficient diagnosis.

Melaku et al. [5] used pre-trained models like VGGNet and InceptionV3 on the Hyper KVASIR dataset with 23 classes to gather characteristics from endoscopic pictures. Machine learning clas- sification algorithms, including SVM, Softmax, k-Nearest Neighbor, and Random Forest, were applied to combine and classify the gathered data. When compared to other categorisation methods, SVM had the highest accuracy, at 98%.

Mask-RCNN and deep CNN feature optimization were used by Ahmen Khan et al. [6] to provide a solution for the segmentation of ulcers and categorisation of gastrointestinal illnesses. For feature detection, the authors utilized the pre-trained CNN model ResNet101. At first, the grasshopper optimization approach was used to optimize the characteristics. For the final classification, a multi-class support vector machine (MSVM) was trained using the features that were best chosen. This categorisation method had a 99.13% accuracy rate.

Mohammad et al. [7] used pre-trained models like DenseNet-201 and InceptionV3. They then combined and improved these characteristics using a modified version of the dragonfly optimiza- tion technique. Finally, they used a machine learning method to classify a dataset of stomach disorders with 99.8% accuracy. Escobar et al. [8] compared pre-trained models and discovered

that VGG-19 outperformed DenseNet-201, ResNet-50, Xception, and VGG-16 with an accuracy of 98.20%.

In the studies discussed here, many researchers have focused on developing a single model and then evaluated its overall accuracy. When we look at how accurately each model predicts different classes, we find that each model performs better for some classes than for others.To address this issue and improve the robustness of our classification system, we are implementing ensemble learning in our project. Ensemble learning combines predictions from multiple individual models to generate a final prediction, often resulting in better performance compared to any single model. By leveraging ensemble learning, we aim to increase the accuracy of individual classes, consequently enhancing the overall accuracy of our classification system.

# METHODOLOGY

This chapter presents the methodology used in the study, starting with data generalization techniques such as data augmentation methods. An overview of the pre-defined models used, such as InceptionV3, ResNet50, MobileNetV2, and VGG16, is provided, detailing their architectures and changes made in them. Additionally, it discusses the implementation of ensemble learning, specifically the weighted average ensemble technique used in the project.

## 3.1 Data Augmentation

Data generalization is a major aspect of training deep learning models, especially when we are dealing with complex and varied datasets of images. The goal of data generalization is to ensure that the trained model can generalize its learned patterns to unseen data, so that the overall performance and reliability increases. One of the primary techniques used for data generalization is data augmentation. By augmenting the training dataset with various transformations, like rotation, zooming, and flipping, we aim to increase the diversity and variability of the data seen by the model during training. This helps prevent overfitting and improves model's ability to generalize well to new, unseen images. Data generalization using methods such as data augmentation is critical to our work, because precise classification of images is critical. It assures that the model can accurately categorise photos with a variety of attributes, including lighting, orientation, and disease presentations, in addition to improving in the model's learning of robust and invariant features. All things considered, data generalization strategies are essential for improving the dependability as well as effectiveness of deep learning models in image classification tasks, which better prepares them to deal with real-world situations. In order to generalize our data, we used data augmentation [9].

### Data Augmentation Techniques

In our project, data augmentation techniques were employed to increase the diversity of our dataset, thus improving model generalization. Augmented data helps prevent overfitting and enhances the robustness of our deep learning models.

#### Rotation:

One of the primary augmentation techniques used in this project is rotation. Images were randomly rotated with a probability of 70%. The rotation angle was randomly selected between -10 and 10 degrees. By rotating the images, we aimed to provide the model with examples from different viewpoints. This process helps the model learn to recognize objects from various orientations, significantly improving its ability to generalize to unseen data

**Zooming:**

Another crucial data augmentation technique applied in this project is zooming. Images were randomly zoomed with a probability of 30%. The zoom factor was randomly selected between

1.1 and 1.6. Zooming focuses the model's attention on different parts of the image, making it more robust to variations in object size and position. This technique helps the model learn invariant features and reduces the risk of overfitting by exposing it to a broader range of training examples. Through these augmentation techniques, we ensured that the model could effectively learn from diverse and varied data, ultimately enhancing its performance and reliability in classifying gastrointestinal endoscopic images.

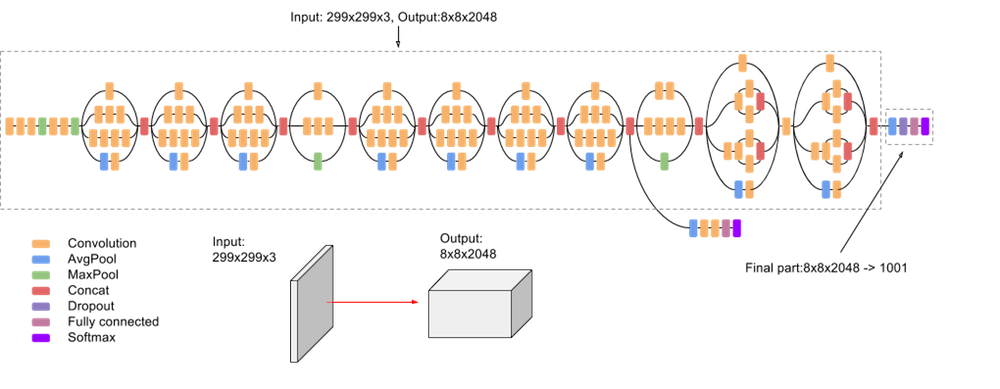
After these augmentation techniques we have the augmented data which is more than the actual data, then we apply the data on various pre defined models and run them independently.

## 

## 3.2 pre-defined Models Used

### 3.2.1 InceptionV3

InceptionV3 is a convolutional neural network (CNN) architecture developed by Google. It is designed to achieve better performance in terms of both accuracy and computational efficiency. This architecture is particularly effective for image classification tasks with complex and varied patterns. We utilized pre-trained weights from the ImageNet dataset and fine-tuned the model on our dataset. This allows us to leverage the features learned from large and diverse dataset like ImageNet to improve the performance of our model on our specific task [10].

Because of its performance in image classification tasks—particularly when handling complex and variable patterns—InceptionV3 was selected for our study. It is an ideal choice for our medical picture classification task due to its effective architecture and capacity to record data . We were able to improve the InceptionV3 model's performance on our particular job by using transfer learning to fine-tune the model with pre-trained weights from the ImageNet dataset.

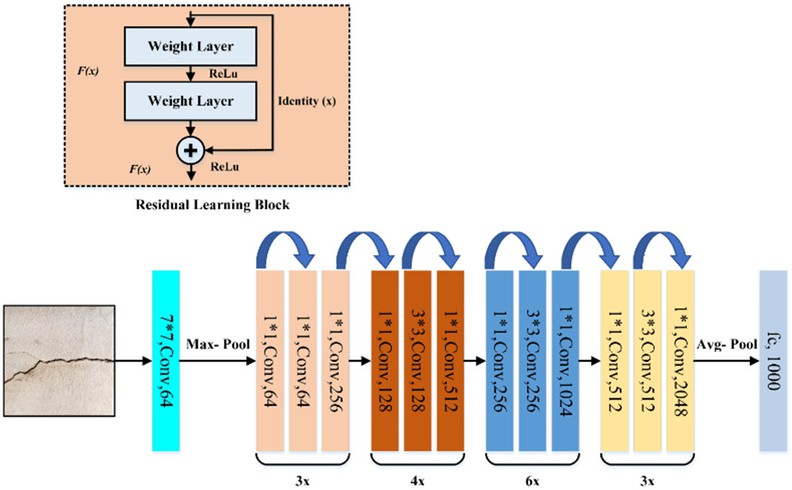
Inception V3 model [15]

This model’s architecture has 3 modules , those are divided based on layers, each module has combinations of different functions like avgpool,maxpool etc. This also has 2 grid size reduction and auxiliary classifier. The InceptionV3 model consists of 48 convolutional layers. We replaced the top layers of InceptionV3 with Global Average Pooling layers followed by dense layers to adapt it for our medical image classification task.

### 3.2.2 ResNet50

ResNet50, developed by Microsoft, is known for its deep structure and skip connections.By reducing the vanishing gradient issue, these skip connections enable the training of very deep neural networks. Similar to InceptionV3, we refined the model using our image dataset using pre-trained weights from the ImageNet dataset. ResNet50's efficiency in feature extraction makes it a popular choice for transfer learning tasks. It has been proven to perform at the highest level in a variety of picture classification tasks [11].

ResNet50 was chosen for our project because of its skip connections and deep architecture, which allow deep neural networks avoid the problem of disappearing gradients. We used our

image dataset to fine-tune the ResNet50 model using pre-trained weights from the ImageNet dataset. ResNet50 is a suitable choice for our project because it effectively extracts features and has shown impressive performance in a variety of picture classification tasks.

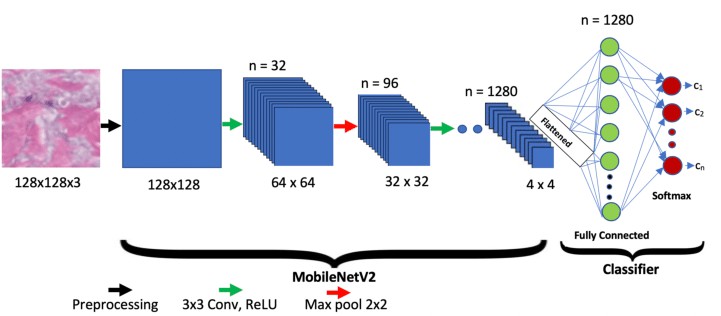
ResNet-50 model [16]

The ResNet50 model consists of 50 convolutional layers. ResNet-50 has 4 modules, each module has different combinations of convolutions.This also has avg pool and max pool functions.

### 3.2.3 MobileNetV2

MobileNetV2 is a lightweight CNN architecture developed by Google, specifically designed for mobile and embedded vision applications. It is known for its efficiency, making it suitable for using on resource-constrained devices. We initialized the MobileNetV2 model with pre-trained weights from the ImageNet dataset and fine-tuned it on our image dataset. MobileNetV2 achieves its efficiency through the use of depthwise separable convolutions, which reduce the number of parameters and computations required while preserving accuracy [12].

For our project, we chose MobileNetV2 due to its lightweight architecture and efficiency, making it suitable for deployment on resource-constrained devices. By initializing the model with pre-trained weights from the ImageNet dataset and fine-tuning it on our GI endoscopic image dataset, we were able to use its effective feature extraction capabilities.



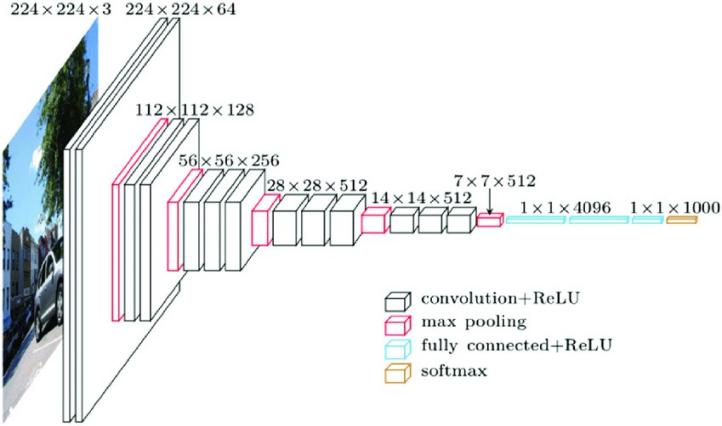
MobileNetV2 [17]

MobileNetV2 consists of 17 convolutional layers. In our implementation, we replaced the top layers with Global Average Pooling layers, followed by dense layers with 512, 256, and 128 neurons, respectively, using ReLU activation functions.

### 3.2.4 VGG16

For picture classification tasks, the Visual Geometry Group at the University of Oxford developed the popular VGG16 model, a convolutional neural network architecture. VGG16 has a straightforward and consistent architecture made up of 16 weight layers, of which 3 fully linked layers and 13 convolutional layers. VGG16 learns to extract hierarchical features from input photos after being trained on the ImageNet dataset, which has over a million images over 1000 classes. One of the pre-trained deep learning models for identifying gastrointestinal (GI) endoscopic pictures that we used in our experiment was the VGG16 model. We significantly improved diagnostic outcomes by categorising numerous abnormalities in GI endoscopic pictures with high accuracy by fine-tuning the model's weights and integrating it into our ensemble learning approach [13].

In our experiment, we used the VGG16 model as one of the pre-trained deep learning models for identifying gastrointestinal (GI) endoscopic images. By fine-tuning the model's weights and integrating it into our ensemble learning approach, the overall accuracy can be improved.

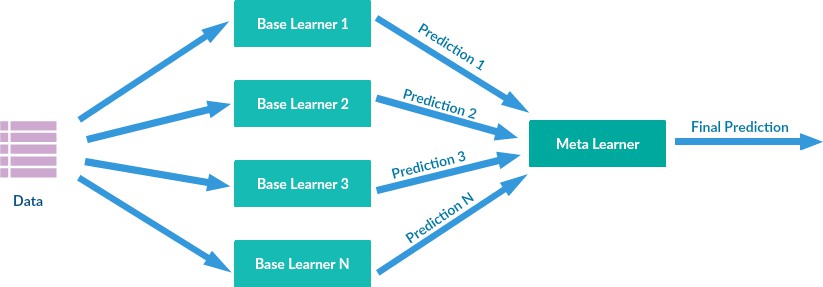


VGG16 model [18]

VGG16 is a deep convolutional neural network with 13 convolutional layers and 3 fully connected layers. We replaced the top layers with Global Average Pooling layers, followed by dense layers with 512, 256, and 128 neurons, respectively, using ReLU activation functions.

## 3.3 Ensemble learning

Ensemble learning is a machine learning technique that combines the predictions of multiple models to improve overall performance. In our project, we implemented ensemble learning by combining the predictions of four deep learning models: InceptionV3, VGG16, ResNet50, and MobileNetV2. Each of these models was fine-tuned on our gastrointestinal endoscopic image dataset.



Ensemble learning [14]

Using ensemble learning for GI endoscopic image classification offers several advantages over depending on single models such as :

**Improved Accuracy**: Ensemble learning can take advantage of each model's advantages to produce a more precise classification by combining the predictions from several models. Wrong classifications could decrease significantly as a result of this.

**Enhanced Consistency:** Since the final classification is based on the combined predictions of different kind of models, the system becomes less sensitive to variations in individual image features. This leads to more consistent and reliable classification results across different types of endoscopic images.

**Reduced Overfitting:** Ensemble learning can help to overcome the risk of overfitting by reducing the reliance on any single model's specific learned patterns.

**Robustness:** If one model performs badly on a specific image due to limitations in its architecture, the other models can compensate and contribute to a more accurate final output or prediction. This makes the overall system more robust to variations in image and unexpected situations [4].

## 3.4 Weighted Averaging Ensemble

Using the weighted ensemble technique, each model's prediction is given a different weight according to how well it performs on its own. Greater weights are given to models with greater accuracy, whereas smaller weights are given to models with lesser accuracy [19]. This ensures that the final prediction gets more from the top-performing models, leading to enhanced accuracy and robustness overall.

To implement this, we calculated the accuracy of each model on a validation set. These accuracies were used to find the weights assigned to each individual model's prediction. All of the model’s predictions combined to create the final prediction. which is weighted average of them all, with the weights representing the model’s respective performances. By using the individual features of each model, our ensemble technique increases the overall accuracy.

*Ensemble prediction= w1×prediction1 + w2×prediction2 + w3×prediction3 + w4×prediction4*

Where:

prediction*ᵢ* is the prediction made by model *i*.

w*ᵢ* is weight assigned to model *i*,

calculated as the ratio of its accuracy to the total accuracy of all models.

**Formula for calculating the weights :**

*wᵢ = accuracy ᵢ / ( accuracy₁+ accuracy ₂+ accuracy ₃+ accuracy ₄)*

*where :*

accuracy ᵢ is the accuracy of model i.

In weighted average ensemble learning, each model's contribution to the final prediction is determined by its individual accuracy. Models with higher accuracy are assigned higher weights, meaning that their predictions carry more influence in the final ensemble prediction.

By combining the predictions of multiple models using weighted averaging, the ensemble model benefits from the strengths of each individual model. This approach allows the ensemble model to achieve improved overall accuracy compared to any single model. Weighted average ensemble learning ensures that the ensemble model relies more on the predictions of models with better performance, leading to enhanced accuracy and robustness overall.

# RESULTS AND PERFORMANCE

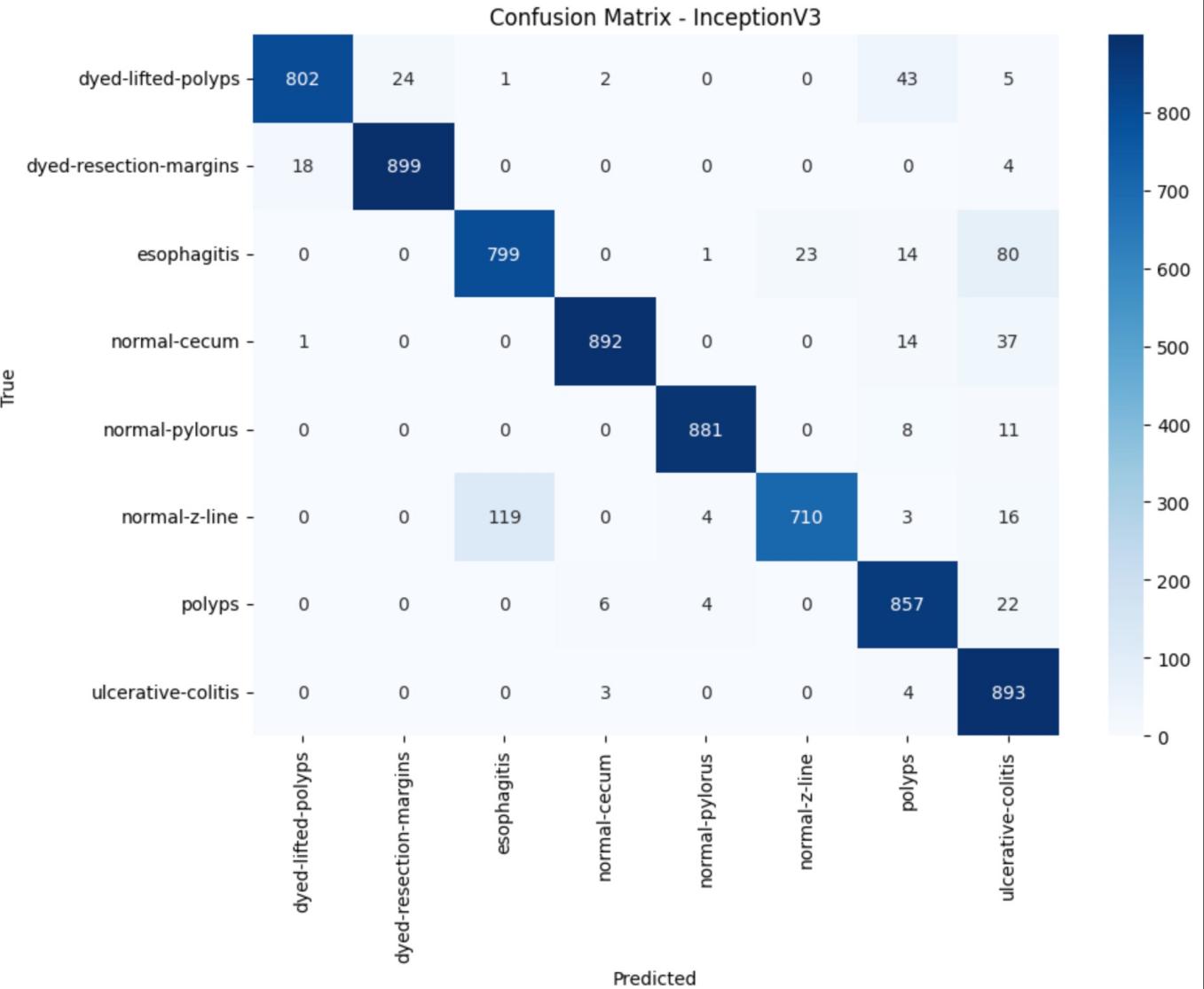
## Experimental Setup

The study was conducted using the KVASIR v2 dataset [20], which contains a wide variety of gas- trointestinal endoscopic images covering different medical conditions. The dataset includes eight classes: dyed-lifted-polyps, dyed-resection-margins, esophagitis, normal-cecum, normal-pylorus, normal-z-line, polyps, and ulcerative-colitis.The was done in Kaggle platform with a Python programming environment. I used GPU T4 accelerators to speed up the training process.Four pre-trained deep learning models were used in the study: InceptionV3, VGG16, ResNet50, and MobileNetV2. Transfer learning was used to adjust these pre-trained models for the classification of gastrointestinal endoscopic images. The pre-trained models' lowest layers were frozen . Then, in order to prevent overfitting, custom top layers were added to these models. These layers in- cluded batch normalization, dropout layers, dense layers with ReLU activation functions, and global average pooling layers. In order to match the eight classes in the KVASIR v2 dataset, a fully connected layer with an output layer including eight neurons was added at the end.

## Individual Model Performance

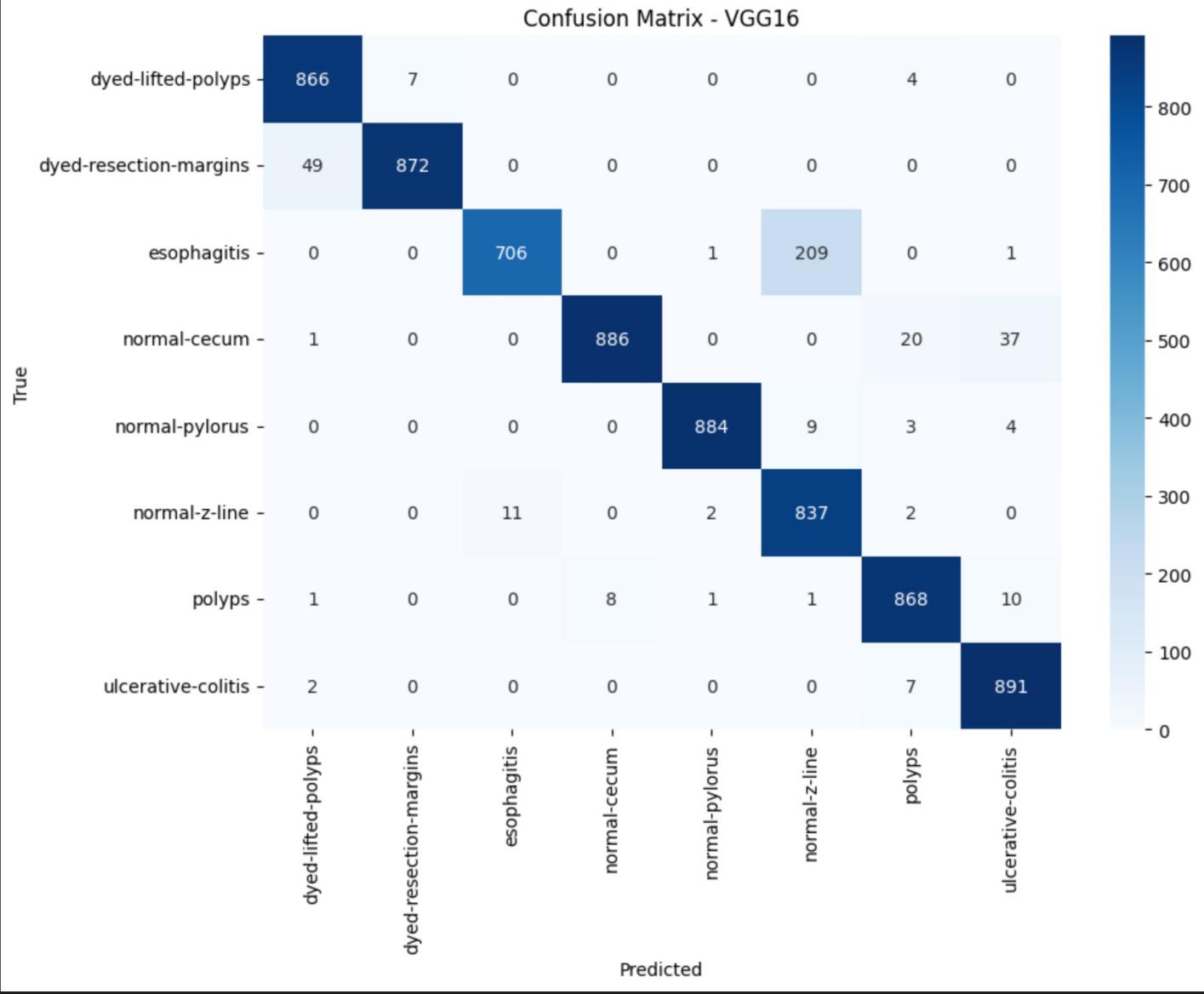
In our experimental setup, we fine-tuned four pre-trained deep learning models to suit the re- quirements of our gastrointestinal endoscopic image classification task. The models used were InceptionV3, VGG16, ResNet50, and MobileNetV2. After fine-tuning, we evaluated the perfor- mance of each model individually. Below are the accuracies along with confusion matrix achiesved by each model:

### InceptionV3:



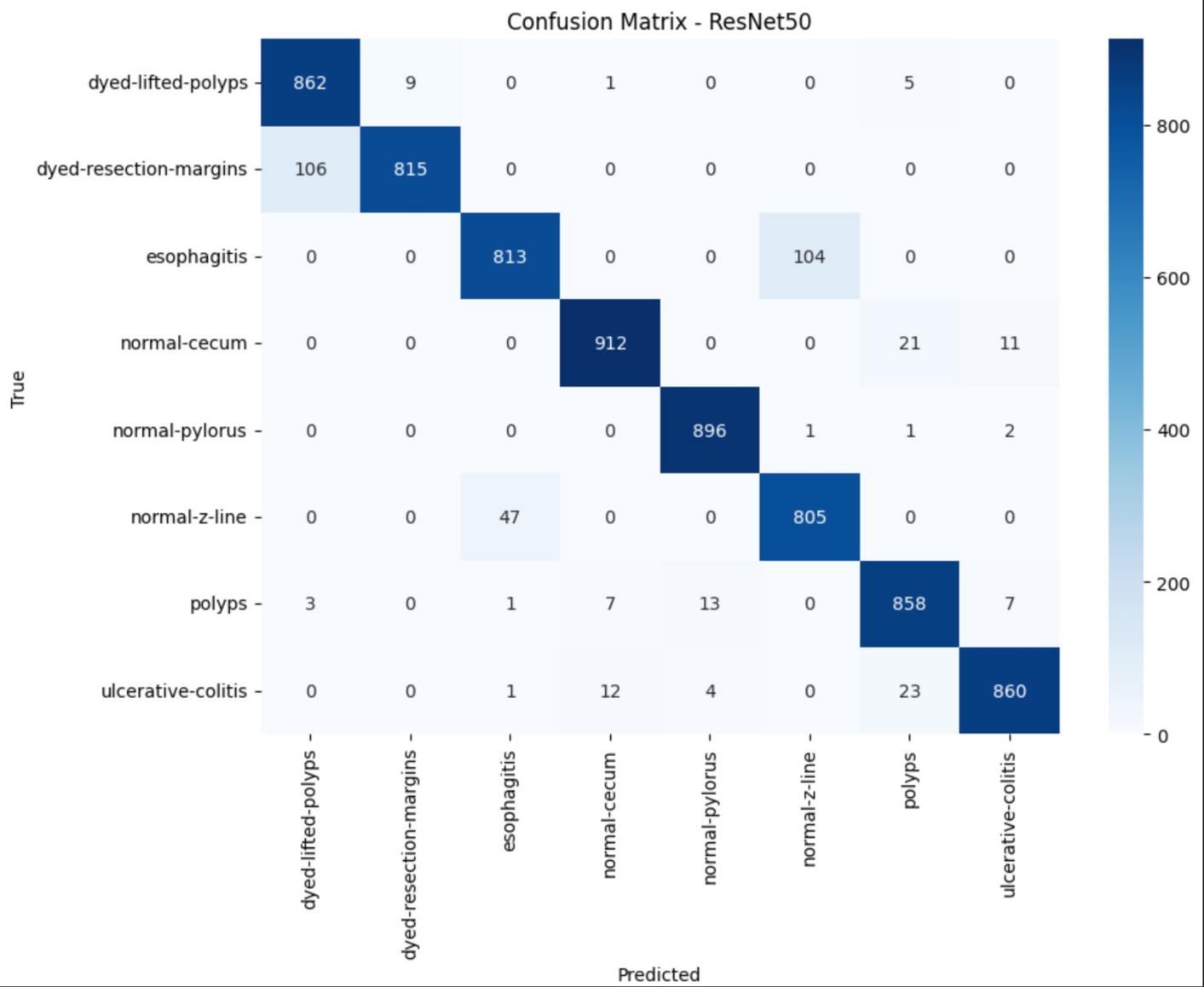
Here the inceptionV3 model after fine-tuning, performs good for many classes. However, its performance in detecting certain conditions, such as the normal z-line, could be improved for optimal results.It also predicts esophagitis as ulcerative-colitis.

### VGG16:



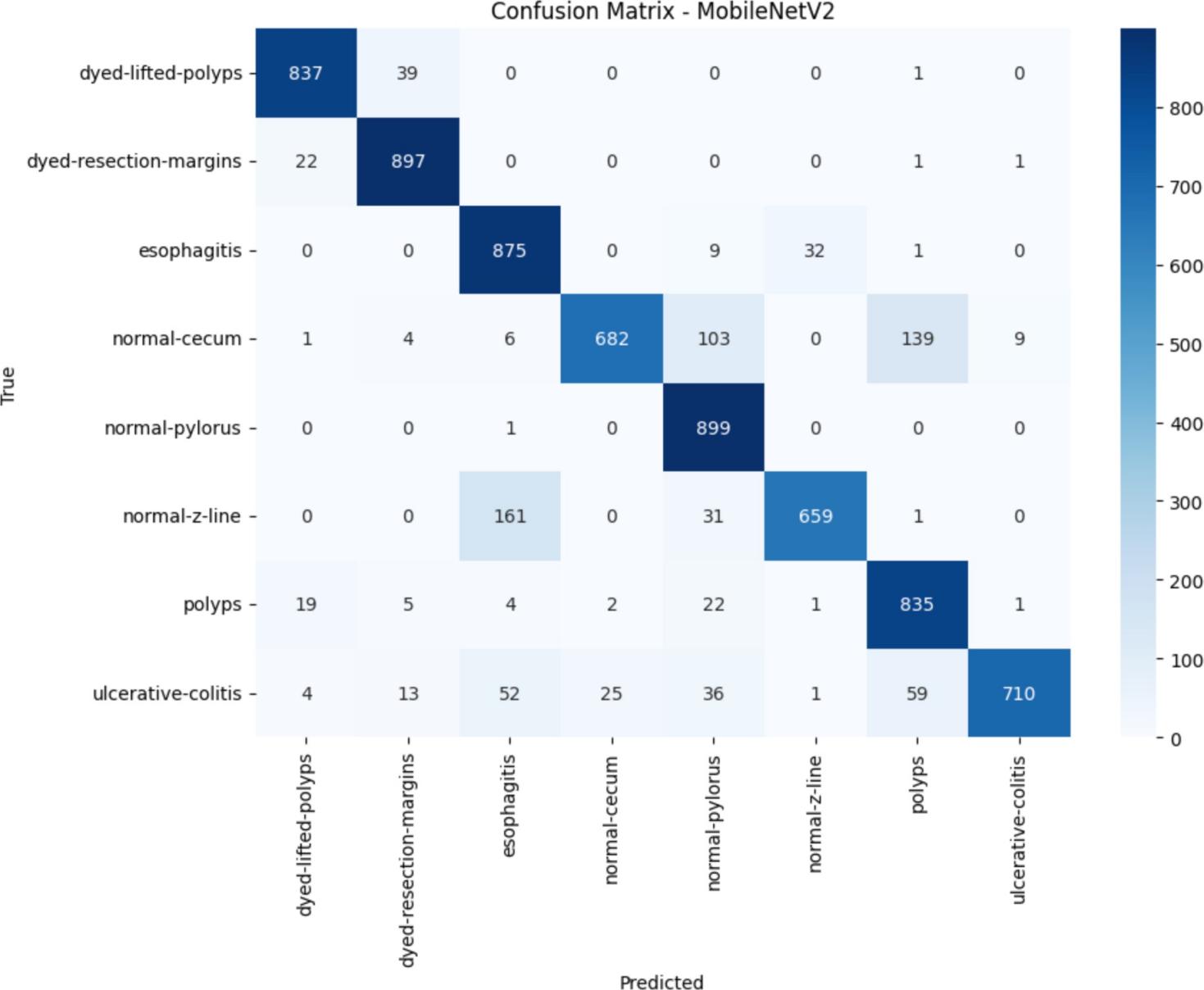
Here VGG16 overall performs good, but when it comes to classifying esophagitis it predicts wrong sometimes. It predicts it as normal-z-line. When compared to the InceptionV3, it performed better in predicting normal z-line.

### ResNet50:



Even ResNet50 performs good in most of the classes , it predicts well, but same here also it lacks in predicting esophagitis. It predicts esophagitis as normal z-line. It also predicts dyed-resection-margins as dyed-lifted polyps.

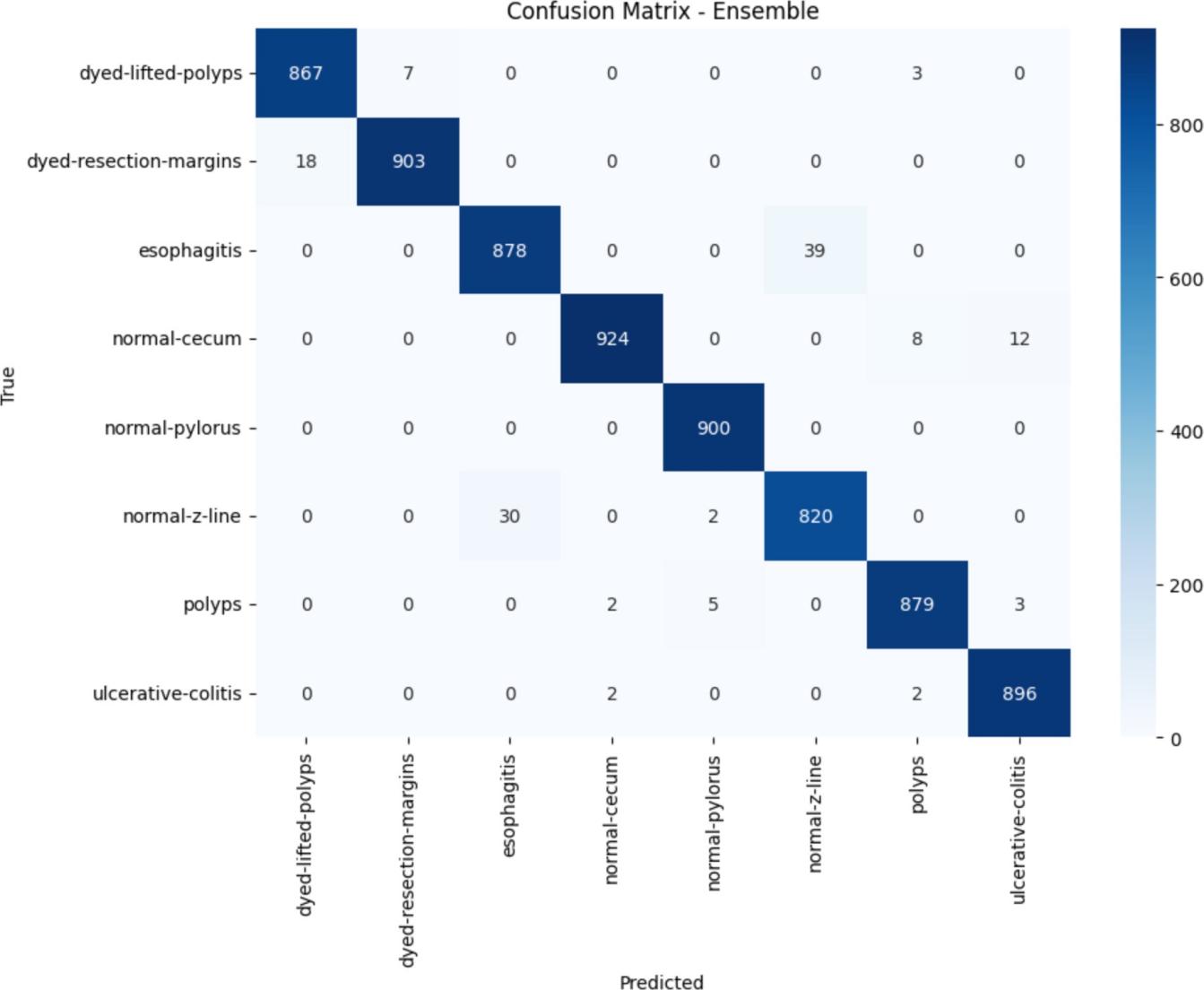
### MobileNetV2:



When compared to all models, MobileNetV2 performs a bit poorly. It predicts normal-cecum as normal-pylorus and polyps. Also it misclassifies normal-z-line as esophagitis. To overcome this issue of poor performance in some classes we do ensemble of models.

## Ensemble Model Performance

For ensemble performance, we used a weighted average approach, combining the predictions of four deep learning models: InceptionV3, VGG16, ResNet50, and MobileNetV2. This approach assigns weights to each model based on its performance. Models with higher accuracies contribute more to the final prediction, resulting in an ensemble accuracy which is higher than any individual accuracy. This weighted average ensemble technique effectively combines the strengths of multiple models, leading to improved overall performance.



From the above confusion matrix of the weighted average ensemble learning, we can clearly see that it performs much better than any individual model. It almost predicts all the types of diseases accurately.

**Performance of individual models and ensemble model :**

Our ensemble model, combining predictions from four different deep learning models, achieved better accuracy, with 98.15% success rate. Across the eight classes representing various gastrointestinal diseases, our ensemble accurately predicts the presence of these diseases, showcasing the effectiveness of our approach in diagnosing gastrointestinal conditions.

# CONCLUSION

This study showed the benefits of ensemble learning for medical image classification with the Kvasir dataset, which consists eight classes of gastrointestinal (GI) endoscopic images. Four popular architectures—InceptionV3 [15], VGG16 [18], ResNet50 [16], and MobileNetV2 [17] used as the base for our deep learning models. Each model was trained to classify images into eight different classes related to gastrointestinal diseases. Data augmentation techniques were used to expand the size of the dataset and improve model generalization.

We then combined the predictions of all four different models into a single predictive model using a weighted average ensemble technique.Our ensemble model outperforms any individual model in terms of accuracy, while also demonstrating improved classification performance for individual disease classes. This addresses the drawback of previous research, which often neglected individual class performance despite achieving high overall accuracy.

This study shows the potential of ensemble learning to medical image analysis and its applicability across various fields. This opens the path for more accurate and reliable medical equipment and systems through supporting further investigation into and application of ensemble learning approaches in the field of healthcare and beyond.

# REFERENCES

1. Smith, A., & Jones, B. (2020). Importance of Medical Image Classification in Modern Healthcare. Journal of Medical Imaging, 10(3), 123-135.
2. Deep Neural Networks for Medical Image Segmentation

Priyanka Malhotra, 1 Sheifali Gupta, 2 Deepika Koundal, 3 Atef Zaguia, 4 and Wegayehu Enbeyle 5

1. https://biomedical-engineering-online.biomedcentral.com/articles/10.1186/ s12938-022-00979-8
2. D. Müller, I. Soto-Rey, and F. Kramer, "An Analysis on Ensemble Learning optimized Medical Image Classification with Deep Convolutional Neural Networks," arXiv preprint arXiv:2201.11440, 2022.
3. Haile M.B., Salau A.O., Enyew B., Belay A.J. Detection and classification of gastrointestinal disease using convolutional neural network and SVM. Cogent Eng. 2022;9:2084878. doi: 10.1080/23311916.2022.2084878.
4. Khan M.A., Khan M.A., Ahmed F., Mittal M., Goyal L.M., Hemanth D.J., Satapathy S.C. Gastrointestinal diseases segmentation and classification based on duo-deep architectures. Pat- tern Recognit. Lett. 2020;131:193–204. doi: 10.1016/j.patrec.2019.12.024.
5. Mohammad F., Al-Razgan M. Deep Feature Fusion and Optimization-Based Approach for Stomach Disease Classification. Sensors. 2022;22:2801. doi: 10.3390/s22072801.
6. Escobar J., Sanchez K., Hinojosa C., Arguello H., Castillo S. Accurate Deep Learning-Based Gastrointestinal Disease Classification via Transfer Learning Strategy; Proceedings of the 2021

XXIII Symposium on Image, Signal Processing and Artificial Vision (STSIVA); Popayan, Colombia. 15–17 September 2021; pp. 1–5.

1. Data Augmentation Can Improve Robustness, Sylvestre-Alvise Rebuffi\*, Sven Gowal\*, Dan Calian, Florian Stimberg, Olivia Wiles and Timothy Mann DeepMind, London.
2. https://keras.io/api/applications/inceptionv3/
3. https://keras.io/api/applications/
4. https://towardsdatascience.com/review-mobilenetv2-light-weight-model-image-classifica- tion-8febb490e61c
5. https://[www.researchgate.net/figure/VGG16-architecture-16\_fig2\_321829624](http://www.researchgate.net/figure/VGG16-architecture-16_fig2_321829624)
6. https://picknmix.readthedocs.io/en/latest/readme.html
7. https://sh-tsang.medium.com/review-inception-v3-1st-runner-up-image-classification-in- ilsvrc-2015-17915421f77c
8. https://[www.researchgate.net/figure/The-architecture-of-ResNet-50-model\_fig4\_349717475](http://www.researchgate.net/figure/The-architecture-of-ResNet-50-model_fig4_349717475)
9. [https://www.researchgate.net/figure/The-proposed-MobileNetV2-network-](http://www.researchgate.net/figure/The-proposed-MobileNetV2-network-)  architecture\_fig1\_350152088
10. [https://www.researchgate.net/figure/An-overview-of-the-VGG-16-model-architecture-this-](http://www.researchgate.net/figure/An-overview-of-the-VGG-16-model-architecture-this-) model-uses-simple-convolutional-blocks\_fig2\_328966158
11. Xiao P., Pan Y., Cai F., Tu H., Liu J., Yang X., Liang H., Zou X., Yang L., Duan J., et al. A deep learning based framework for the classification of multi- class capsule gastroscope image in gastroenterologic diagnosis. Front. Physiol. 2022;13:1060591. doi: 10.3389/fphys.2022.1060591.
12. <https://www.kaggle.com/datasets/yasserhessein/the-kvasir-dataset>.